Commissioning of the CMS Muon Detector and Development of Generic Search Strategies for New Physics

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Physiker

Philipp Alexander Biallass

aus Duisburg

Berichter: Univ.-Prof. Dr. Thomas Hebbeker
Univ.-Prof. Dr. Christopher Wiebusch

Tag der mündlichen Prüfung: 27.03.2009

Contents

Zusammenfassung

Titel der Arbeit: Inbetriebnahme des CMS Myon Detektors und Entwicklung einer generischen Suchstrategie nach neuer Physik

Der Nachweis und die Rekonstruktion kosmischer Myonen sind essentiell für die Inbetriebnahme und das Alignment des Compact Muon Solenoid Experimentes (CMS), sowohl in der Zeit vor als auch während der ersten Proton-Proton Kollisionen. In diesem Zusammenhang war der Magnet Test/Cosmic Challenge (MTCC) eine Art Generalprobe für die Detektor-Hardware und die Rekonstruktions-Software. Es wurden hierfür umfassende Datennahmen kosmischer Myonen durchgeführt und auch erstmals das enorme Magnetfeld angeschaltet. Ein zentraler Punkt einer physikalischen Datenanalyse ist immer auch der Vergleich mit simulierten Ereignissen. Im ersten Teil dieser Arbeit wird daher ein neuer kosmischer Myon-Generator vorgestellt, CMSCGEN. Seine Richtigkeit wird an Hand von Vergleichen mit Daten vom MTCC durchgeführt und es werden hierfür erstmals Daten-Monte Carlo Vergleiche mit dem zentralen Myon System gezeigt, angefangen von einzelnen Zelltreffern bis hin zu ersten rekonstruierten Spurmessungen mit Impulsvorhersage.

Da Leptonen (Elektronen, Myonen) sehr klare Signaturen für Signale neuer Physik haben, sind diese Detektortest- und Alignment-Strategien auch unerlässlich für die meisten Physik-Analysen. Im zweiten Teil dieser Arbeit wird eine modellunanhängige Suchstrategie für Signale neuer Physik in CMS vorgestellt, ein Ansatz der Ereignisse mit Leptonen benutzt und nur die Kenntnis des Standard Modell Untergrundes voraussetzt. Durch diese Vermeidung einer theoretischen Voreingenommenheit ist man so sensitiv auf eine Vielzahl von Modellen, inklusive solcher noch nicht erfundener Szenarien. In dieser Machbarkeitsstudie werden die Ereignisse geordnet anhand des Teilcheninhaltes (Myonen, Elektronen, Photonen, Jets, fehlende Energie) und in sogenannten Ereignisklassen gruppiert. Ein umfassender Scan verschiedener kinematischer Verteilungen wird durchgeführt, in dem signifikante Abweichungen vom SM Monte Carlo identifiziert werden. Hierbei wird die Wichtigkeit der systematischen Unsicherheiten betont, die sorgfältig im Suchalgorithmus berücksichtigt und eingebaut werden. Mehrere theoretische SM Erweiterungen wie z.B. Supersymmetrie und neue schwere Eichbosonen werden untersucht, genauso wie mögliche Detektor-Effekte und Generator-Schwierigkeiten. All diese Beispiele demonstrieren die weiten Anwendungsmöglichkeiten dieses Such-Algorithmus und sie zeigen auf, wie dieser Ansatz die eher traditionellen Modell-spezifischen Suchen ergänzt.

Preface and Abstract

In the history of mankind major scientific progress has often been linked to important technological developments changing the face of the earth. The exploration of electricity and magnetism lead to the invention of light bulbs and electric transformators which both can be regarded as foundations of our modern life style. In some cases the link between fundamental physics progress and benefit for the whole mankind is not as clear as in the examples just mentioned. While the theories of special relativity and quantum mechanics revolutionized our understanding of the very basic laws of nature there was no foreseeable technological and economical benefit at the time of their invention. Still without these theories modern satellite communication and semiconductor technology, essential for computers and thus the internet, would not be possible. In the past centuries the understanding of the most fundamental laws of nature, the various particle species and their interactions via forces, has reached an unprecedented level of precision. The Standard Model (SM) of particle physics is able to describe the complexity of nature with impressive accuracy by exploiting the principles of symmetry. It has endured numerous tests performed by probably the most ambitious experiments ever built by humans. Thus one might argue that physics reached a dead end with nothing fundamental more to discover.

Taking a closer look at the theory of Quantum Field Theory, which is the basis of the Standard Model, it becomes clear that this is surely not the case. The Standard Model intrinsically breaks down at sufficiently high energies, and one of its components (the Higgs particle) has even not been found yet. On the other hand some other particles well known and probed since many years, the neutrinos, cannot be included easily in the framework of the SM. A maybe even graver indication for physics beyond the Standard Model is given by Cosmology. Analyzing the matter content of the universe it becomes clear that only a tiny fraction (4%) consists of matter described by the SM. A much larger part (20%) is made of new species of particles, so called dark matter. And the remaining 76% are given by a totally unknown ingredient, the dark energy. These numbers reflect that our understanding of nature and the universe is still very limited, but this also indicates that new physics' revolutions are only waiting to change our perception of the world and maybe also lead to new dimensions of technology.

In this context a large experiment is currently about to start at the European Organization for Nuclear Research (CERN) in Geneva: The Large Hadron Collider (LHC). In this particle accelerator two protons will collide with an energy of 14 TeV, creating conditions that existed 10^{-12} s after the Big Bang which have never been investigated in laboratory experiments so far. Various theoretical considerations motivate that physics beyond the Standard Model should appear at these energies. Thus there is great anticipation and belief that the LHC will open up the door to a new era in physics and pave the way to some future "Theory of Everything".

One of the four detectors aiming to contribute to this important step is the Compact Muon

Solenoid experiment (CMS). After years of design, construction and tests this ambitious project is now waiting for the first data from proton-proton collisions to arrive, thanks to the hard work of thousands of physicists and technicians from all around the world. Covering only a tiny fraction of this enormous effort this thesis in a way reflects the important steps in setting up the CMS experiment during the past years:

After the first major parts of CMS have been constructed and later assembled at CERN these have to be tested and conjoined to a single unit. In this context the commissioning of the various detector components plays an important role and some initial large scale test of CMS, the Magnet Test and Cosmic Challenge (MTCC), will be discussed in the first half of this thesis. After the construction phase of CMS the preparations for first collision data become more and more important. Thus the focus shifts from detector commissioning to physics commissioning, developing data analyses to look at the physics content of the collision events. Initially this means re-measuring what has already been established by former experiments and thus testing if the detector works as expected. Eventually this will pass into the search for first hints of new physics, which of course is the main aim of CMS. Both aspects are important during the start-up phase of the experiment and are addressed in the second part of this thesis which presents a generic data analysis looking for deviations from the Standard Model expectation.

Analysis Summary

The detection and reconstruction of cosmic muon rays is important for the commissioning phase and alignment of the Compact Muon Solenoid experiment (CMS), in particular during the early phases of operation with physics collisions. In this context the Magnet Test/Cosmic Challenge (MTCC) with its comprehensive cosmic data taking periods including the presence of the magnetic field has been like a dress rehearsal of detector hardware and software for the upcoming start-up of the CMS detector. In addition to data taking also the comparison with simulated events is a crucial part of physics analyses. The first part of this thesis introduces a new cosmic muon generator, CMSCGEN, and it presents its validation by comparing its predictions with data from MTCC. As an example, results from a reconstruction study using the barrel muon system are shown, comparing data and Monte Carlo prediction at the level of single chambers up to reconstructed tracks including momentum measurements.

Since leptons (electrons, muons) constitute very clean signatures for signals of new physics these commissioning and alignment procedures are also vital to most physics analyses. In the second part of this thesis a model independent search approach for new physics within CMS is presented, utilizing events with leptons and relying only on the knowledge of the Standard Model simulation. Due to the absence of a theoretical bias this approach is sensitive to a variety of models, including those not yet thought of. Within this feasibility study events are classified according to their particle content (muons, electrons, photons, jets, missing energy) into so called event classes. A broad scan of various distributions is performed, identifying significant deviations from the SM Monte Carlo simulation. The importance of systematic uncertainties is outlined, which are taken into account rigorously within the search algorithm. Several theoretical models such as Supersymmetry and new heavy gauge bosons as well as possible detector effects and generator issues have been fed into the search algorithm as benchmark scenarios to demonstrate its capabilities. Thus this approach supplements the traditional model-driven searches.

Chapter 1

Theoretical Foundations

1.1 Theoretical framework of particles and interactions

1.1.1 Notations and conventions

The thesis presented here uses "God-given" units, where

$$\hbar = c = 1 \ . \tag{1.1}$$

In this system (natural units),

$$[\text{length}] = [\text{time}] = [\text{energy}]^{-1} = [\text{mass}]^{-1} \ . \tag{1.2}$$

The mass (m) of a particle is therefore equal to its rest energy (mc^2), and also to its inverse Compton wavelength $(mc/\hbar)$. For example,

$$m_{\text{electron}} = 9.109 \times 10^{-28} \ \text{g} = 0.511 \ \text{MeV} = (3.862 \times 10^{-11} \ \text{cm})^{-1} \ . \tag{1.3}$$

This text uses the common conventions for relativity, where the metric tensor $g_{\mu\nu} = g^{\mu\nu}$ runs with Greek indices over 0, 1, 2, 3 or t, x, y, z. γ^μ are the gamma matrices with $\mu = 0, ..., 3$. Repeated indices are summed in all cases. As a consequence, for a particle with mass m holds

$$p^2 = p^\mu p_\mu = E^2 - |\mathbf{p}|^2 = m^2 \ . \tag{1.4}$$

Throughout the text, particles and anti-particles are abbreviated by the name of the corresponding particle, and indices specifying the particle's charge are often omitted. Thus, the reaction $Z^0 \to \mu^+\mu^-$ is described by writing $Z \to \mu\mu$.

1.1.2   The Standard Model

The LHC is designed to seek deviations from the Standard Model (SM), a theoretical framework describing nature which has been tested with enormous accuracy and success in the past. One could question the value of an experiment which seeks to find alternatives to a model so demonstrably trustworthy. To understand the reasons for believing that a theory beyond the Standard Model exists, the principles and main features of the SM should first be introduced. As the mathematical formalism of the SM, *Quantum Field Theory* (QFT), is extremely complex, this Section is meant only to address a few of the basics and to outline the formalism. Details can be found in the literature, e.g. [1].

Quantum Field Theory is a relativistic field theory which can be regarded as an expansion of Quantum Mechanics. In contrast to every day life, matter is no longer characterized by particles, but for every particle a corresponding field is postulated. A field is a physical object which can emerge everywhere in space-time, e.g. a scalar field $\Phi(\vec{x}, t)$. With the help of the field concept, commonly known from "classical" Electrodynamics, many problems of relativistic Quantum Mechanics can be solved.

After some simple steps (but without providing details here), QFT uncovers numerous milestones: (1) Requiring a *continuous symmetry* (e.g. a global $U(1)$) of the quantized field inevitably leads to the existence of *anti-particles* with opposite quantum number (e.g. electrical charge); (2) the spectrum of states has positive energy $E \geq 0$; (3) causality is preserved. This causality is very important as it leads to the conclusion that QFT is "local", i.e. a measurement of the field at the space-time-point x does not affect a measurement at y.

Most concepts from Quantum Mechanics, like the construction of a Lagrangian $\mathcal{L}$, a Hamiltonian $\mathcal{H}$, and the principle of minimal action $\delta S = 0$, are retained. From this well known formalism, equations of motion for non-interacting particles can be derived, e.g. the Dirac-equation describing the spinor field Ψ (spin-$\frac{1}{2}$ fermion):

$$(i\partial_\mu \gamma^\mu - m)\Psi = 0 \ . \tag{1.5}$$

Until now only Lorentz-symmetry has been assumed, expressed by the invariance under Poincaré transformations. This philosophy of symmetries determines the structure and beauty of the Standard Model: One postulates fields (e.g. spinor Ψ or vector field A_μ) and symmetries, and one constructs the most global Lagrangian which is compatible with these symmetries. In addition to the Poincaré symmetry, *local symmetries* (e.g. $SU(2)$) are demanded, defined by:

$$\Psi_b \longrightarrow [e^{i\varepsilon(x)^A T^A}]_{ab}\Psi_b \quad \text{with} \quad A = 1, .., D \ . \tag{1.6}$$

Here $\varepsilon(x)^A$ are real parameters and T^A are the generators of the transformation. D is the dimension of the corresponding symmetry group (e.g. $D = 1$ for $U(1)$) and a, b are the *inner* quantum numbers. These local symmetries are called inner symmetries, and their quantum numbers are discrete. The inner indices are traits of a particle which add to the spin and the mass. An example is the $SU(3)$ symmetry which defines the strong force and the Quantum Chromo Dynamics (QCD). Its eight generators correspond to the eight gluon fields G_μ^A, and the inner degrees of freedom are the three colors r, g, b.

At the point where local symmetries are assumed, the theory of free fields transforms into a theory of interacting fields. Expressions in the Lagrangian like $\overline{\Psi} i \partial_\mu \gamma^\mu \Psi$, where $\overline{\Psi} = \Psi^+ \gamma^0$, are not invariant under the local symmetry transformation. *Gauge fields* A_μ^A must be introduced in addition, fields with a transformation behavior which ensures the invariance of the Lagrangian. In order to ensure invariance of all terms of $\mathcal{L}$, the covariant derivative is defined:

$$\partial \longrightarrow \partial - ig A_\mu^A T^A \ . \tag{1.7}$$

As a consequence, terms in $\mathcal{L}$ with combinations of matter fields like Ψ and gauge fields like A_μ^A are generated, representing the interaction between the fields with the coupling constant g. These vector fields correspond to spin 1 particles which mediate the forces. In the Standard Model, three fundamental forces are realized and can be described using the formalism of symmetry and gauge field:

- $SU(3)$ strong force $\longrightarrow$ eight gluon fields G_μ^A

- $SU(2) \times U(1)_Y$ electroweak force $\longrightarrow$ three W-boson fields W_μ^A and one hypercharge boson field B_μ

The vector bosons which can be observed in nature (W, Z, γ) are linear combinations of these electroweak fields, e.g. for the photon field A_μ which mediates the electromagnetic force with the couplings g and g':

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}} \left(g' W_\mu^3 + g B_\mu \right) \ . \tag{1.8}$$

On the other hand, the matter fields represent the fermions which interact with one another:

- Quarks: strong, weak and electromagnetic interaction

- Leptons: weak and electromagnetic interaction

- Neutrinos: weak interaction

All fermions of the Standard Model are summarized in Table 1.1.

	I	II	II	Charge
Leptons	e	μ	τ	-1
	ν_e	ν_μ	ν_τ	0
Quarks	u	c	t	+2/3
	d	s	b	-1/3

Table 1.1: *The three generations of matter*

Each of these matter fields is divided into three families, e.g. electron/muon/tau. In the SM there are only three families, but no theoretical explanation of this number is given. With respect to the electroweak interaction the fermionic matter fields can be grouped into $SU(2)$ doublets containing the left-handed particles and $SU(2)$ singlets for the right-handed ones:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, ... \quad \text{and} \quad e_R, u_R, d_R, ... \qquad . \tag{1.9}$$

Note that assuming a vanishing neutrino mass only left-handed neutrinos are included in the SM (in conflict with observations from neutrino-oscillations [2] which indicate $m_\nu \neq 0$). This grouping is needed in order to be consistent with the experimental observation of parity violation of the weak force. Thus only left-handed fermions can interact weakly while the $U(1)_Y$ gauge field can also couple to the right-handed ones. The weak force can mediate transformations between and within the $SU(2)$ doublets, i.e. transitions from one family to another and within a family.

The final Lagrangian of the SM can be written as:

$$\mathcal{L}_{SM} = \mathcal{L}_{Kin} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \ . \tag{1.10}$$

The kinetic term $\mathcal{L}_{Kin}$ includes the free fields and all interactions between different fields, i.e. interactions between the fermion fields and gauge boson fields, and combinations of gauge boson fields only (e.g. possible self-interactions). $\mathcal{L}_{Higgs}$ accounts for a complex scalar field ϕ which fulfills all required symmetries. This Higgs field will play an important role in the formalism of *electroweak-symmetry-breaking* discussed at the end of this section. $\mathcal{L}_{SM}$ does not include any terms of the form $m^2 \overline{\Psi}\Psi$ as they violate gauge-invariance. As a consequence, all gauge bosons and all fermions are without mass. With the help of symmetry-breaking the mediators of the weak force, $W^\pm$ and Z^0, can get a proper mass term. The term $\mathcal{L}_{Yukawa}$ is then added to $\mathcal{L}_{SM}$ since its combinations of Higgs field and fermion fields can give rise to all fermion masses.

The goal of QFT is to calculate cross sections of particle interactions, i.e. to describe the dynamics of the theory. These dynamics are defined by the structure of the postulated Lagrangian, i.e. the allowed interactions between fields and the corresponding coefficients in $\mathcal{L}_{SM}$. The computation can be performed by applying "Feynman rules" to a certain scatter process $ij \rightarrow kl$. The interaction is illustrated by Feynman graphs, see "Bhabba scattering" ($e^+e^- \rightarrow e^+e^-$) as an example in Figure 1.1. These graphs are calculation rules for the specific process: The outer lines represent the initial- and final state particles; the inner lines represent the Feynman propagators; and the vertices combine inner and outer lines. The number of lines which end at a vertex is defined by the structure of the Lagrangian, i.e. which fields interact. Each vertex has a certain factor which includes the coupling constant of the specific force.
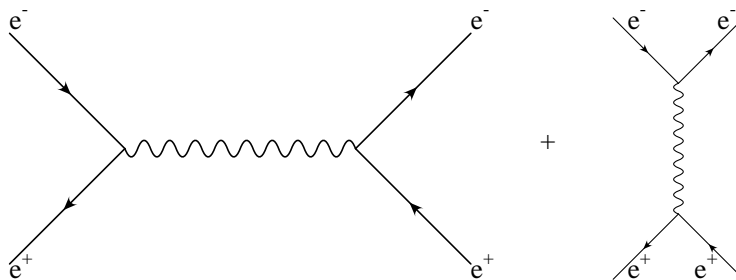
Figure 1.1: *The two Feynman diagrams contributing to Bhabba scattering $e^+e^- \rightarrow e^+e^-$*

As Figure 1.1 shows, more than one graph can contribute to a single process, as only initial and final state are well defined. All possible inner lines are only subjected to the structure of $\mathcal{L}_{SM}$. As a consequence, *loop corrections* must be implemented. Loops are additional inner lines which can cause the cross sections to become divergent. This non-physical behavior can only be solved by the formalism of *renormalization*.

Up to now a local and lorentz-invariant theory with $\mathcal{L}(\phi_0; M_0, \lambda_0)$ has been described, where ϕ_0 is an arbitrary field, M_0 its mass parameter and λ_0 the coupling constant. If a "cut-off" Λ is defined which restricts the momenta of loops, then relations between the physical parameters (M, λ) and the "naked" parameters (index 0) can be obtained:

$$M = M(M_0, \lambda_0) \quad \text{and} \quad \lambda = \lambda(M_0, \lambda_0) \ . \tag{1.11}$$

These relations are divergent in Λ. The cut-off parameter has the meaning of an upper energy limit. The theory, i.e. the Standard Model, is only valid up to this energy regime. If $\mathcal{L}(\phi_0; M_0, \lambda_0)$ is a reasonable theory, then predictions made by the theory should not depend on the renormalization scale Λ, i.e. they should be finite for $\Lambda \to \infty$. The divergence of $M = M(M_0, \lambda_0)$ is no problem as M_0 and λ_0 are hypothetical parameters which cannot be measured by any physical apparatus.

Again, a complex formalism of renormalization exists which cannot be explained in detail at this point. The essence is that besides the postulation of certain symmetries of a Lagrangian, a valid theory has to be renormalizable. This means that with the help of a *rescaling of parameters and fields*, all cross sections have to become finite. This requirement strongly restricts the multitude of possible terms contributing to $\mathcal{L}$. With the help of these complex techniques, it can be proved that the Standard Model is a renormalizable theory. The strategy of constructing the Standard Model (symmetries and renormalization) also holds for most extensions of the SM, e.g. Supersymmetry, and thus represents the core of QFT.

One might think that the scheme of loop corrections and renormalization is only an abstract issue which does not affect physical observables. But it does affect them strongly. The rescaling of parameters and fields leads to a *scale dependence* of the couplings, e.g. for QCD:

$$\lambda(\mu_2) = \frac{\lambda(\mu_1)}{1 - \beta_0 \lambda(\mu_1) \ln(\frac{\mu_2^2}{\mu_1^2})} \quad \text{with} \quad \beta_0 < 0 \ . \tag{1.12}$$

Here, β_0 is a constant parameter and μ_2 is the energy scale where the coupling is tested. The other scale parameter μ_1 defines a fixed energy scale, and it is related to the Landau pole μ^* via:

$$\mu^* = \mu_1 \exp \frac{1}{2\beta_0 \lambda(\mu_1)} \ . \tag{1.13}$$

The Landau pole marks the energy region where the interaction becomes non-perturbartive. Equation 1.12 contains two important physical statements. The first is that $\lambda(\mu) \to 0$ for $\mu \to \infty$. This fact is called "asymptotic freedom" of QCD and means, that for very large energy scales quarks can be regarded as free particles which no longer interact strongly. Thus with the help of electrons and a large momentum transfer q^2 the electromagnetic substructure of hadrons can be probed (see experiments at Hera in Hamburg).

The second statement regards the μ^*. By measuring λ at a certain scale, μ^* can be calculated using the above expression. The result is that μ^* amounts to several hundred MeV, the typical scale of hadron masses. For $\mu \lesssim 1$ GeV the coupling becomes so large that free quarks cannot be observed. This is called the "confinement of quarks and gluons".

In Figure 1.2 the scale dependence of the forces is shown. In addition to the different slopes, one can see that there is a point where the electromagnetic and the weak force merge into a single
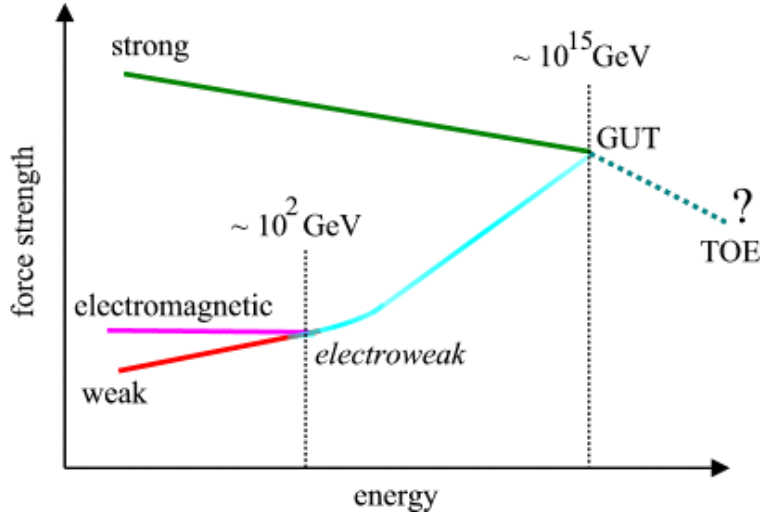
Figure 1.2: *Scale dependence and unification of forces [3].*

force, the *electroweak* force. This represents another important issue in QFT: the "spontaneous breaking of a symmetry". The Higgs-mechanism, which was mentioned earlier and stated as the origin of mass, is closely linked to this unification of forces.

The Higgs-mechanism starts with the task of defining the ground state of the Standard Model. The Higgs field ϕ in $\mathcal{L}_{Higgs}$ is the only candidate for such a ground state. Even though the spectrum of the possible ground states is degenerate, nature realizes a specific state ϕ_0. Whereas the Lagrangian is not affected by this particular choice, the state itself is no longer invariant under the electroweak symmetry transformation $SU(2) \times U(1)_Y$, the symmetry is said to be "broken spontaneously". The $SU(2)$ doublet of the Higgs field can be written as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} . \tag{1.14}$$

Here v represents the constant value of the minimum and $H(x)$ is the dynamical Higgs field. If this expression is inserted into the Lagrangian $\mathcal{L}_{SM}$, the parts with couplings between the Higgs field and the gauge bosons generate mass terms. Within $\mathcal{L}_{Yukawa}$ terms of the form $\sim \overline{\Psi}_L \phi \Psi_R$ can be added "ad hoc" to $\mathcal{L}_{SM}$, separately for each fermion family with $\Psi_L = SU(2)$ doublet and $\Psi_R = SU(2)$ singlet. Using the Higgs ground state this combination of left- and right-handed fermions produces Dirac mass terms $\sim v\,\overline{e}_L e_R$ (i.e. lepton and quark masses), as well as fermion-Higgs interactions $\sim H(x)\,\overline{e}_L e_R$ (with a coupling determined by v and the measured fermion masses). In this way the experimentally observed massive gauge bosons and massive fermions can be accomodated within QFT and the Standard Model.

The ground state of the Higgs field introduces a scale for the theory and determines the mass of the Higgs particle. If we take $v \approx 250$ GeV (derived from $m_W = \frac{ev}{2\sin\theta_W}$ where e = electron charge, m_W = mass of W-boson and θ_W = weak-mixing angle), then at scales much higher than the scale of the ground state, v can be neglected and thus all mass terms are negligible. Without these mass terms W and Z become massless, and the unified electroweak force emerges.

The description of the Higgs-mechanism concludes this overview of the Standard Model and its founding principles. As the energy scale of the electroweak symmetry breaking is near the TeV scale where current experiments are running, there is anticipation that very interesting physics can be observed in this unexplored regime. The Higgs particle itself should have a mass of this magnitude, and extensions of the Standard Model could be verified (e.g. Supersymmetry). Using the techniques described above results in the construction of even higher symmetries, corresponding to Lagrangians with renormalizable fields and thus new theories. These theories are designed to solve problems of the SM, some of these problems and details of possible extensions of the SM will be discussed in the next section.

1.2 Extensions of the Standard Model

1.2.1 Problems of the Standard Model

Before discussing similarities and differences of some typical theories beyond the Standard Model one should specify more precisely the open questions which cannot be answered by the SM alone (details for example in [4] [5]) :

a) Hierarchy problem

b) Cold dark matter

c) Unified theoretical description of gauge forces (+gravity)

The hierarchy problem refers to the two major scales observed in nature; the scale of SM physics with the Higgs vacuum value of $v \approx 10^2$ GeV and the Planck scale of gravity with $M_{Pl} \approx 10^{19}$ GeV. As introduced in the previous section also the Higgs mass parameter depends on some scale Λ, thus loop corrections have to be considered. It turns out that these lead to quadratic divergences in Λ, thus to a huge correction of the physics predictions of the theory if $\Lambda \sim M_{Pl}$, unless an unnatural fine tuning is performed. Here the postulation of additional symmetries can solve the divergence. These theories beyond the SM predict additional particle species which can also contribute to the loop corrections and thus cancel the large radiative corrections (e.g. fermion-boson Supersymmetry or Little Higgs models assuming new heavy gauge bosons). For the cancellations to work at least in Supersymmetry these new particles should not be heavier than about 1 TeV, so ideally for LHC.

Results from the observation of the Cosmic Microwave Background (WMAP) and of the galactic rotation curve indicate that the majority of the matter content in the universe (80% of the matter) is of unknown nature. If made up of particles these should carry no electrical charge, be massive, stable and interact only weakly. It is a main goal of theories beyond the SM to provide candidate particles for dark matter. In the case of Supersymmetry a conserved multiplicative quantum number, R parity, is postulated solely for producing a stable lightest SUSY particle, the LSP. It turns out that assuming such a LSP around 10^2 GeV (thus again in the reach of LHC) is at the right level to generate the observed relic density of cold dark matter in the universe.

Finally theories beyond the SM promise to pave the path to a unified description of forces. As discussed in the previous section the couplings of the SM depend on the scale at which they are probed, e.g. via collisions. Extrapolating the scale to high values in the Standard Model the three couplings of the electroweak and strong interactions miss each other by orders of magnitude. This can again be changed by postulating new fields within a theory extension; new loop contributions can bend the scale behavior of the couplings and a unification seems possible, e.g. in Supersymmetry.

In addition to this Grand Unification (GUT) of course also a quantum field theory of gravity is still a dream to come true. Elaborating the connection between the forces of the Standard Model and gravity is also an important aspect, finally leading to a "Theory of Everything". Such an inclusion of gravity could also lead to a deeper understanding of the cosmological constant, possibly explaining the mysterious dark energy content of the universe. Some SM extensions as Extra-Dimensions or Supersymmetry make first progress in this direction.

Note that these three examples are not the only open questions of the SM. There are others which also cannot be answered by most of the new physics models, e.g. why are there 3 families of matter and what is the origin of the 18 "nature given" parameters of the Standard Model. Supersymmetry even worsens the situation with a multitude of new unconstrained parameters. Still there is hope that once discovered and studied at LHC such a new theory might lead to new ideas possibly answering more questions than originally imagined.

Also it should be stressed that there is no real proof that these three problems of the SM require new physics. They are only facts that look unsatisfactory and worrying, but there are quite some loop holes for the SM to be valid up to the Planck scale without complicated extensions. Maybe fine-tuning is not a problem but just reflects our lack of understanding how nature works; maybe cold dark matter consists of very heavy particles (e.g. Axions) which can be easily added to the SM; maybe there is no unification of forces at large scales.

The analysis presented in the second part of this thesis deals with a generic search for deviations from the Standard Model. Thus no a priori preference is given to a certain extension of the SM, and some common aspects of models beyond the SM will be discussed in the following. Still there is also the possibility, even if disliked, of SM-only physics also above 1 TeV, as mentioned before. This is compatible with the search Ansatz since this results in a good agreement between data and SM expectation.

1.2.2   Common models of new physics

In the previous section several reasons why to believe in new physics around 1 TeV have been summarized. In the past 25 years of theory a number of models have been predicated on the idea of extensions of the SM (see e.g. [6]), many of them trying to address at least some of problems mentioned earlier. Especially the hierarchy problem can be used to motivate the benefits of most of these extensions:

Exotic Models

In models like Technicolor [7] the hierarchy problem is solved by avoiding the divergences connected to the SM Higgs boson. Inspired by Quantum Chromo Dynamics one gets a composite Higgs Boson created through new strong forces and a ratio Λ_{TC}/M_{Pl} which is naturally small, so no fine-tuning is needed. The W and Z bosons acquire their masses using this new strong dynamics instead of electroweak symmetry breaking. In a similar way also in Little Higgs models [8] the Higgs boson is a composite field caused by some new strong dynamics and it would again be light in a natural way. Here the SM is embedded inside a larger group with an enlarged symmetry and the Higgs is a pseudo Goldstone boson arising from the breaking of this global symmetry.

Many of these models are tightly constrained by electroweak precision measurements. In order to escape from large electroweak loop corrections incompatible with the measurements one can postulate "repeaters" of SM particles like W- and Z-boson, Higgs and top-quark. So using a combination of a new strong dynamics plus new heavy particles might be sufficient to keep the Higgs light enough, thus solve the hierarchy problem.

Another possibility would be the existence of a quark and lepton substructure [9], accompanied by some new strong dynamics (mostly approximated by contact interactions). This would result in excited leptons and quark states with spectacular experimental signatures.

More focused on the problem of a unified theory (GUT) some models predict the existence of Leptoquarks [10] which can transform leptons into quarks, thus establishing a symmetry between both fermionic species. Considering the Lagrangian of such models many similarities to SUSY can be found.

Extra dimensions

Another type of models can be summarized by Extra Dimensions. In contrast to the models discussed before Extra Dimensions address several problems of the SM, not just the hierarchy problem. These models mostly originate from Kaluza-Klein theories attempting to unify gravity and electromagnetism in a five dimensional theory. Modern variants often also combine ideas from string theory since also there extra space dimensions appear. Thus there is some hope that these theories could contribute to item c), possibly unifying the description of gravity and the SM. Still unlike in Supersymmetry there is no real indication of any unification of electroweak and strong couplings.

Some Extra Dimension models also produce candidate particles for the cold dark matter, i.e. item b).

Still they can be motivated best using the hierarchy problem: Using different techniques Extra Dimensions are able to bring the effective Planck scale down to 1 TeV, thus avoiding fine-tuning problems. In this case the SM is only an effective theory and new physics should take over at a scale λ comparable to the Higgs mass. Thus one does no longer have the problem of divergences in the loops of the Higgs mass since no integration up to M_{Pl} is needed.

One type of model predicts Large Extra Dimensions (ADD) [11]. Here several macroscopic extra dimensions are realized in which only gravity can propagate. Considering the Einstein

Lagrangian in $D + 1$ dimensions one gets the following relation between the Planck scale $M_{Pl,3}$ of our world and the fundamental one of the full theory $M_{Pl,D}$:

$$M_{Pl,3}^2 = M_{Pl,D}^{D-2} \cdot V_{D-3} \qquad , \qquad (1.15)$$

where V_{D-3} denotes the volume of the extra dimensions. Thus if this volume and/or D are large enough one can get a large $M_{Pl,3}$ with a fundamental scale $M_{Pl,D}$ only around TeV. The problem of explaining the huge hierarchy between Planck scale and the weak scale is translated into the problem of explaining why the extra dimensions are stabilized at such large radii relative to the weak scale.

A second type of models is called Warped Extra Dimensions (Randall-Sundrum) [12]. Here only a single extra dimension is needed which is strongly curved by a large negative cosmological constant (AdS space). The "weakness" of gravity, i.e. the largeness of M_{Pl}, can be explained by the $5th$ dimension separating a "Planck brane" and a "Weak brane". While gravity is localized and strong on the first brane there is an exponential suppression (warp factor)

$$e^{-M_{Pl} \cdot x_5} \qquad , \qquad (1.16)$$

where x_5 is a measure for the distance between the two branes in the $5th$ dimension. Thus there also is an exponential suppression of its strength on the weak brane (SM). In this way all mass parameters starting at $\sim M_{Pl}$ are rescaled down to 1 TeV. Thus again the problem of understanding the hierarchy of Planck and weak scale is translated into understanding the distance x_5 between the two branes.

Most models of Extra Dimensions result in signatures detectable at the LHC, some of them with spectacular deviations from the SM. There are towers of repeaters in most cases (Kaluza-Klein excitations) which result in a spectrum of many new heavy mass Eigenstates. In the case of Warped Extra Dimensions these should be a number of resonances easily resolved while for ADD models they rather result in an overall excess in the high-p_T tails. Still under special conditions Large Extra Dimension models even predict spectacular scenarios such as Mini Black Holes decaying into a high multiplicity of Standard Model particles.

Supersymmetry

Finally yet another extension of the SM will be discussed, one which is considered by a large fraction of theoretical and experimental physicists to be the most attractive and elegant solution to the problems discussed above: Supersymmetry [13].

In this theory our 4-dimensional space-time is enlarged by additional fermionic coordinates (extra dimensions of SUSY algebra) $\theta_{1,2}$, $\overline{\theta}_{1,2}$. Shifts and gauge transformations in the superspace lead to the construction of *chiral matter superfields* and *gauge superfields*. These superfields represent the maximal symmetry concept including bosonic and fermionic degrees of freedom, pairing fermions and bosons as shown in Table 1.2:

Thus the particle content of the Standard Model is doubled leading to an equal number of fermionic and bosonic degrees of freedom in each multiplet. The transformation from one type to another within a supersymmetric multiplet can be realized using the Q operator:

SM	spin	SUSY	spin
lepton, quark	1/2	slepton, squark	0
gauge boson, Higgs	1, 0	gaugino, Higgsino	1/2
gluon	1	gluino	1/2
graviton	2	gravitino	3/2

Table 1.2: *Supersymmetric pairing of fermions and bosons in superfield multiplets.*

$$Q \, |F> \; = \; |B> \qquad . \tag{1.17}$$

This operation also implies that in a supersymmetric world the masses of the particles and the sparticles should be equal. Further theoretical considerations lead to the fact that the supersymmetric ground state has a vanishing energy, thus also a vanishing cosmological constant. Both points strongly indicate that SUSY is broken in nature since one can observe a cosmological constant and even more important, since SUSY has not been discovered yet. Thus the sparticle masses have to be much larger then their SM partners, current limits indicating at least $O(100\,\mathrm{GeV})$ [14]. The possible implications of SUSY with respect to the hierarchy problem will be discussed in the following. In the context of SUSY breaking in order to maintain the cancellation of the quadratic divergences, i.e. solve the hierarchy problem, these breaking terms have to be soft and arise within a hidden sector. Messenger particles are needed, e.g. gravity or gauge fields, to transfer the SUSY breaking to our world.

In order to motivate Supersymmetry again the three problems of the Standard Model mentioned above can be used. Unlike in the previous models SUSY comes up with a solution or at least paves the path to a solution of all three problems. This is probably the main reason why so many people "believe" in SUSY and why past experiments like SPS, LEP or TEVATRON have all looked for hints of SUSY:

Problem a)

Considering the loop corrections to the bare Higgs mass m_0, the W-boson leads to corrections

$$\delta m_W^2 \sim -g_W^2[\lambda^2 + m_W^2] \qquad ,$$

where g_W refers to the coupling and λ is the upper limit of the loop integration. Not considering the techniques of renormalization, but interpreting the loop as a physical reality this represents the scale up to which the SM is valid, e.g. M_{Pl}. It has already been stated that extreme fine-tuning is needed in order to stabilize the Higgs mass m_H below 1 TeV. At this point the duplication of the SM particle spectrum becomes important. In addition to the SM particle loops also all supersymmetric partners contribute. Since there is an equal numbers of fermions and bosons in each supersymmetric multiplet and since the transition from a bosonic loop to a fermionic loop leads to a sign flip, the radiative corrections and thus the quadratic divergences cancel:

$$\delta m_{\widetilde{W}}^2 \sim +g_{\widetilde{W}}^2[\lambda^2 + m_{\widetilde{W}}^2] \qquad .$$

Taking a look at the expression it is clear that this solution to the hierarchy problem only works in case of identical couplings, the SUSY Yukawa coupling $g_{\widetilde{W}}$ and the SM gauge coupling

g_W. Also in order to sustain $m_H \sim 1$ TeV the SUSY sparticles must not be too heavy. Both aspects are fulfilled per construction in a supersymmetric world, but they also have to be preserved after SUSY breaking in order to be consistent with the precision measurements of the SM.

Problem b)

In terms of a candidate for cold dark matter SUSY also looks promising. In this context an additional global symmetry is postulated which leads to a new conserved multiplicative quantum number:

$$\text{R-parity} = \left\{ \begin{array}{l} +1 \text{ for SM particles} \\ -1 \text{ for SUSY particles} \end{array} \right. .$$

This conserved global charge has considerable consequences:

- SUSY particles ($R = -1$) must be produced in pairs.

- SUSY particles ($R = -1$) cannot decay into SM particles alone.

- The lightest SUSY particle must be stable.

The second point leads to the *cascade decays* which determine most experimental signatures. The first point permits to observe resonance peaks in SUSY events which would facilitate mass determinations of the sparticles. Instead of this end-point determination within mass edges can be used to constrain the SUSY spectra. Finally the last point implies that SUSY scenarios where R-parity is conserved can provide a neutral and colorless dark matter candidate, e.g. the lightest neutralino or the gravitino. Theoretical consideration assuming a weak coupling and a mass around 100 GeV generate a relic density just at the right level to explain current cosmological observations (WMAP [15]).

Problem c)

It has already been seen in Equation 1.12 that the three couplings of the SM depend effectively on the momentum transfer in a collision. The higher the momentum transfer the more loop effects can be probed. While in the SM there is no single crossing point of its three couplings, solely by assuming all additional supersymmetric partners the evolution of the couplings is modified and they meet with a precision at the percent level near 10^{16} GeV. One should note that this implication of SUSY is independent of its concrete realization, i.e. the exact SUSY mass spectrum (provided the sparticles are not too heavy $m_{SUSY} \lesssim 10^3$ GeV). The evolution is only driven by the quantum numbers of its constituents.

In addition to this Grand Unification of couplings, which is considered to be esthetically important for a unified theoretical description of forces and matter, SUSY also has some first ideas to describe gravity interactions in a quantum field theoretical way. Without going into detail, *local* Supersymmetry transformations lead to the introduction of a spin 3/2 and a spin 2 field which can be interpreted as the gravitino and the graviton. One should note that this is not a perfect, i.e. renormalizable, theory yet, but hope is justified that SUSY might be a first step towards some superstring theory including gravity in the theoretical description of our universe.

After introducing superfields and after discussing the virtues of Supersymmetry now a Minimal Supersymmetric Standard Model (MSSM) will be presented, designed to be compatible with current experimental measurements and manageable for exploring the phenomenology at the LHC. One should note that this a not really a theory but rather a scenario assuming some well or not so well motivated parameter reduction scheme. In addition to this the SUSY breaking mechanism still remains unclear at the moment and different possibilities lead to vastly different predictions. Many experimentalists as well as theoretical physicists do not really believe that the MSSM or models including a breaking mechanism like mSUGRA or GMSB will accurately describe nature as observed by LHC. Thus there is a danger of potentially misleading predictions of LHC phenomenology. The analysis strategy presented in the second part of this thesis does not focus on some specific SUSY model, but aims to spot deviations in the data from the SM expectation in a generic way. While using some SUSY benchmark points as toy inputs to the algorithm, these specific models used should only be understood as examples.

Minimal Supersymmetric Standard Model

The MSSM represents the minimal extension of the Standard Model which is consistent with current experimental data and able to provide masses to all observed particles. In addition to simply doubling the particle content of the SM, see Table 1.2, additional Higgs fields are required. Still also this "minimal" extension leads to more than 100 new free parameters, consisting of masses, mixing angles and CP-phases. Only via the requirement of several physical assumptions these can be reduced drastically. Before discussing this one should first summarize the particle content of the MSSM.

As mentioned earlier in addition to the 4 degrees of freedom of the SM Higgs doublet an additional independent Higgs doublet field again with 4 degrees of freedom is needed, leading to two independent scalar fields H_u and H_d. Supersymmetric theories require this additional field in order to provide masses to up- and down-type particles.

Also these theories provide a more natural way to explain the Higgs mechanism since a Higgs potential consisting of H_u and H_d leads to a non-vanishing vacuum expectation value which breaks the electroweak symmetry. This mechanism requires the masses of H_u and H_d to be different from each other, i.e. a soft SUSY-breaking, which is naturally caused by the radiative corrections via top/bottom loops. Thus a close connection between gauge symmetry breaking and SUSY-breaking is established.

From the 8 degrees of freedom which are left after symmetry breaking 3 are absorbed as Goldstone bosons by the electroweak bosons and thus provide their masses. The remaining 5 physical states mix to two charged Higgs bosons $H^\pm$, one pseudo-scalar A (CP-odd) and two scalars h and H (CP-even). Considering radiative corrections it becomes evident that the lightest of these, the h, should be lighter than 140 GeV. Also considering the couplings, h behaves similar to the SM Higgs while H and A prefer to couple to down-type particles. In general the production cross sections of the MSSM Higgs particles are significantly enhanced compared to the SM Higgs, thus facilitating possible discoveries.

In the corresponding SUSY sector of gauginos and higgsinos the particle content is determined by flavor Eigenstate mixing into mass Eigenstates: The charged winos $\widetilde{W}^\pm$ and the charged higgsinos $\widetilde{H}^\pm$ mix into two charginos $\widetilde{\chi}^\pm_{1,2}$. Then the neutral gauginos and higgsinos $\widetilde{W}^3$, $\widetilde{B}$, $\widetilde{H}^0_1$ and $\widetilde{H}^0_2$ mix into four neutralinos $\widetilde{\chi}^0_{1,2,3,4}$, ordered by mass.

Since for spin 1/2 particles one has to consider left- and right-handed states the squark and slepton sector includes e.g. $\widetilde{e}^{\pm}_{L,R}$ or $\widetilde{u}^{\pm}_{L,R}$. The scalar states L, R mix again forming mass Eigenstates $1, 2$ where the mixing is especially relevant in the third generation, i.e. for τ, b and t.

In addition to this the eight gluons g are supplemented by eight gluinos $\widetilde{g}$.

Since the LHC is a hadron collider the strong interaction leads to large production cross sections of squarks and gluinos. Irrelevant of the mass hierarchy of $m(\widetilde{g})$ and $m(\widetilde{q})$ the cascade decays into quarks and $\widetilde{\chi}^0_1$ lead to signatures of several hadronic jets plus missing transverse energy. In addition to this the decays are a rich source for non-colored particles, e.g. via $\widetilde{q} \to q + \widetilde{\chi}^0_2$, leading to additional leptons in the event arising from the subsequent neutralino decay. The combination of the two leptons arising from $\widetilde{\chi}^0_2 \to l^+ \widetilde{l}^-_R \to l^+ + l^- + \widetilde{\chi}^0_1$ can be used to constrain mass differences of $\widetilde{\chi}^0_1$, $\widetilde{\chi}^0_2$ and the slepton.

After introducing the particle content of the MSSM and after discussing some prominent signatures one still has to reduce the number of free parameters in order to be able to theoretically describe the phenomenology of supersymmetric models and to compute their mass spectra. This parameter reduction is driven by the expectation that SUSY represents the fundamental theory at the GUT/Planck scale, incorporating maximal symmetry and thus simplicity and opening the gate to a description of gravity.

A well-know example of a possible SUSY-breaking mechanism fulfilling these requirements is the *minimal supergravity mSUGRA* [16]. Here the SUSY breaking takes places in a hidden sector and is transferred by gravitational interaction to the "visible Eigenworld". In addition one requires CP-violation compatible with the measurements, suppression of flavor changing neutral currents to stabilize the proton and a natural generation of the Higgs mechanism via radiative symmetry breaking. The final assumption is the universality at Λ_{GUT} and one is left with just 4 parameters and one sign:

- common gaugino mass $m_{1/2}$ for all gauginos
- common scaler mass m_0 for all sfermions
- universal trilinear coupling A_0
- ratio of the two Higgs vacuum expectation values $tan\beta = \frac{v_u}{v_d}$
- sign of the Higgsino mass parameter sgn(μ)

Note that the latter two parameters are defined at the electroweak scale.

As already discussed Supersymmetry provides a way to allow a unification of couplings. Using the extrapolation mechanism of the renormalization group the three couplings $\alpha_i(Q^2)$ can be combined to $\alpha_U(\Lambda^2_{GUT})$. In a similar way for example the unification of the gaugino masses can be described:

$$M_i(Q^2) = M_{1/2}(Q^2)\frac{\alpha_i(Q^2)}{\alpha_U(\Lambda^2_{GUT})} \qquad . \tag{1.18}$$

Thus using this renormalization group extrapolation the fundamental SUSY theory at the GUT scale can be reconstructed, providing the sparticle masses have been determined in a coherent analysis of LHC data and data from a future linear collider. Note that the interplay between
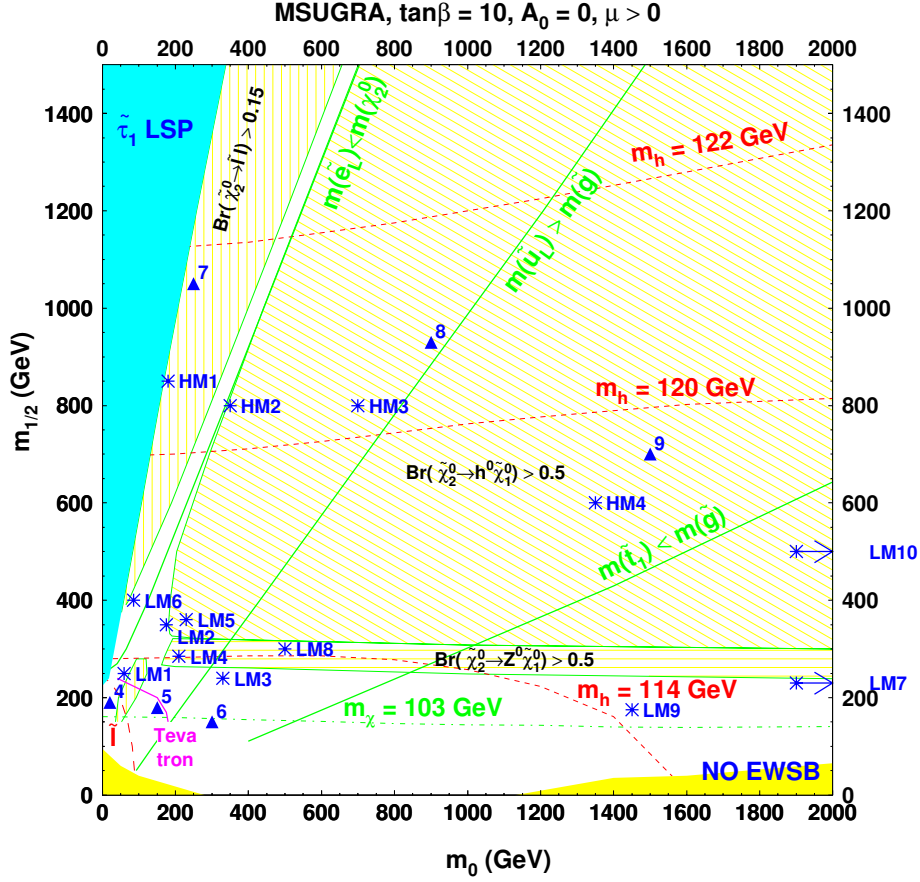
Figure 1.3: *mSUGRA benchmark points in m_0 versus $m_{1/2}$ plane, assuming $tan\beta = 10$, $A_0 = 0$ and $\mu > 0$ [17].*

these two experiments is essential to really measure all SUSY parameters. These measured parameters can also provide information about the underlying microscopic breaking mechanism, be it mSUGRA or gauge mediation (GMSB).

Still, also these 5 parameters lead to a vast possible parameter space, with sparticle spectra strongly depending on this choice. In order to facilitate the comparison of experimental feasibility studies physicists have agreed on certain benchmark points, picking exemplary parameter sets which still represent important phenomenological use cases. Thus one can optimize search strategies depending on the benchmark point in question and determine promising experimental signatures. Figure 1.3 displays the m_0 versus $m_{1/2}$ plane including several benchmark points, LM for low mass and HM for high mass. The shaded areas are theoretically excluded since either $\widetilde{\tau}_1$ would be the LSP (no cold dark matter candidate) or no electroweak symmetry breaking would occur. The red lines refer to the Tevatron exclusion limits and the LEP Higgs bounds. One should note that the LHC experiment really covers a large part of this phase space, thus hopefully it should be possible to either discover SUSY or rule out a vast number of SUSY models.

1.3 Hadron collider physics

At the LHC protons collide with protons. Both are not point like particles, but they consist of several partons. The simple picture of three quarks (uud) building the proton has to be replaced by the complex formalism of a proton's *structure functions*. In the simplified quark-parton model one of the structure functions, F_2, can be written as:

$$F_2 = \sum_i q_i^2 x [f_i(x, Q^2) + \bar{f}_i(x, Q^2)] \quad . \tag{1.19}$$

Here the sum includes all quark flavors $i = u, d, ... $; q_i is the electric charge of the quark and Q^2 is the momentum transfer of the interaction. The structure functions define the composition of the proton:

The variable x is called "Bjorken x" and partitions the four-momentum of the incoming proton P^μ. Each parton gets a fraction resulting in the four-momentum k^μ, i.e. $k^\mu = xP^\mu$. In deep inelastic scattering, only two partons interact. As each of the partons has a certain fraction x of the total momentum P^μ, the entire center of mass energy of the accelerator is not available for the reaction. This is the reason why the laboratory frame (detector) is not the rest system, and most events are boosted along the z-axis.

$f_i(x, Q^2)$ is the parton distribution function (pdf), i.e. the number of quarks with flavor i which have a relative momentum between x and $x+\mathrm{d}x$. What Equation 1.19 means is that a proton consists of a combination of all quarks and anti-quarks. In general, a distinction is drawn between "valence quarks" (uud for proton) and "sea quarks", which are both part of a hadron. The existence of sea quarks, e.g. a virtual strange quark in a proton, can be explained by the presence of gluons keeping the hadron together. Besides the exchange of gluons between two valence quarks, the radiation of gluons converting into a quark anti-quark pair is also possible. As a consequence, anti-quarks and heavy quarks can be created, even though their distribution function for large x is naturally much smaller than, for example, the one of a u-quark, as they are the consequence of radiation processes.

In addition to the momentum given to the sea quarks, experimental data show that the gluons take about half of the initial momentum P^μ. As described by QCD there are quark-gluon interactions within the proton, thus also a gluon distribution $g(x, Q)$ has to be included into the parton model. Figure 1.4 summarizes the different shares of the components of the proton (valence quark density, sea quark density and gluon density). One can see that for small x, the gluons and the sea quarks dominate.

In Equation 1.19, the structure function and all parton densities are not only a function of x, but they also depend on the momentum transfer Q^2. This "scaling violation" results in a structure function F_2 which is shifted towards smaller x-values as $|Q^2|$ rises. If the momentum transfer rises, gluon radiation is enhanced, and the proton is thus dominated by gluons and sea quarks. At a TeV collider, valence quarks are no longer the main contributor to the total cross section of inelastic scattering.

These pdf's are all determined by experiments (e.g. HERA) as theoretical QCD calculations are very complicated. Monte Carlo simulations need these parton distribution functions as an input, so different parameterizations of the experimental data are computed. Effects of different
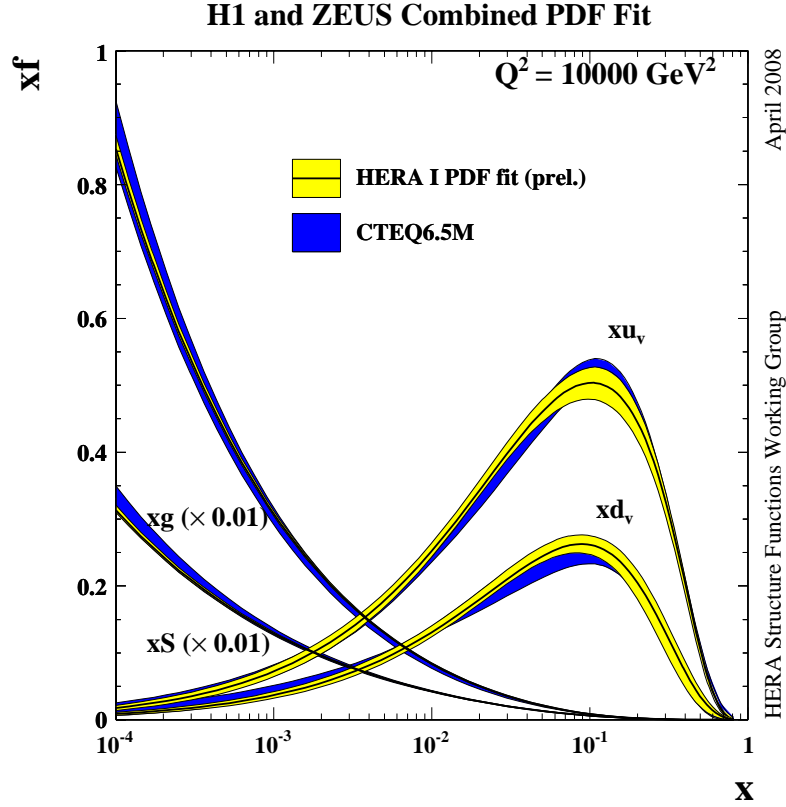
H1 and ZEUS Combined PDF Fit

Figure 1.4: *Parton distribution functions f of the proton times momentum fraction x, H1 and ZEUS combined fits of the parton densities of the proton compared to the CTEQ parameterization. Valence quarks (u_v, d_v), sea quarks (S) and gluon distribution (g) are plotted as a function of the momentum fraction x [18].*

parameterizations can be found in Figure 1.4 when comparing the HERA fit to the CTEQ parameterization. The cross section of a specific process depends on the assumption of a certain pdf, so an inherent systematic uncertainty is attached to the choice of a pdf (see Section II.5.1). Parton distribution function and differential cross section of a process $p + p \rightarrow F$ are connected by:

$$\frac{\mathrm{d}\sigma_F(\sqrt{s}, Q^2)}{\mathrm{d}Q^2} = \sum_{i,j} \int \mathrm{d}x_i \mathrm{d}x_j \, f_i(x_i, Q^2) f_j(x_j, Q^2) \frac{\mathrm{d}\sigma_F^{ij}(x_i, x_j, Q^2)}{\mathrm{d}Q^2} \ . \tag{1.20}$$

Here, F represents an arbitrary final state and $\sqrt{s}$ is the center of mass energy of the p p system. One has to include all possible processes $parton_i + parton_j \rightarrow F$. All these contributions are incorporated in the differential cross section of the individual process $\sigma_F^{ij}(x_i, x_j, Q^2)$, then these are weighted according to the specific parton distribution functions $f_{i,j}$.

After the principles of cross section determination at a hadron collider have been introduced, the total cross sections of different process like $b\bar{b}$-production or W-production have to be discussed. Figure 1.5 shows the total cross sections of various Standard Model processes as a function of the center of mass energy. Arrows mark the energy of different facilities like Tevatron or LHC.

Considering LHC as an example, one can see that QCD multi-jet processes like $b\bar{b}$-production with $\sigma \approx 500\,\mu$b are orders of magnitude larger than the processes important for many analyses, e.g. $W \to e\nu$ with $\sigma \approx 20$ nb. And signatures of new physics are mostly even orders of magnitude below electroweak processes like W-boson production, which underlines that finding these rare events is not a trivial task. Most of the events created at a hadron collider are QCD multi-jets without any leptons in the final state. The triggers of an experiment must be able to filter these events as the amount of stored data would exceed the available resources. As QCD processes dominate, a physics analysis must deal with a large background of all-jet events, mimicking signatures important for the specific analysis (see also Section II.5.2).
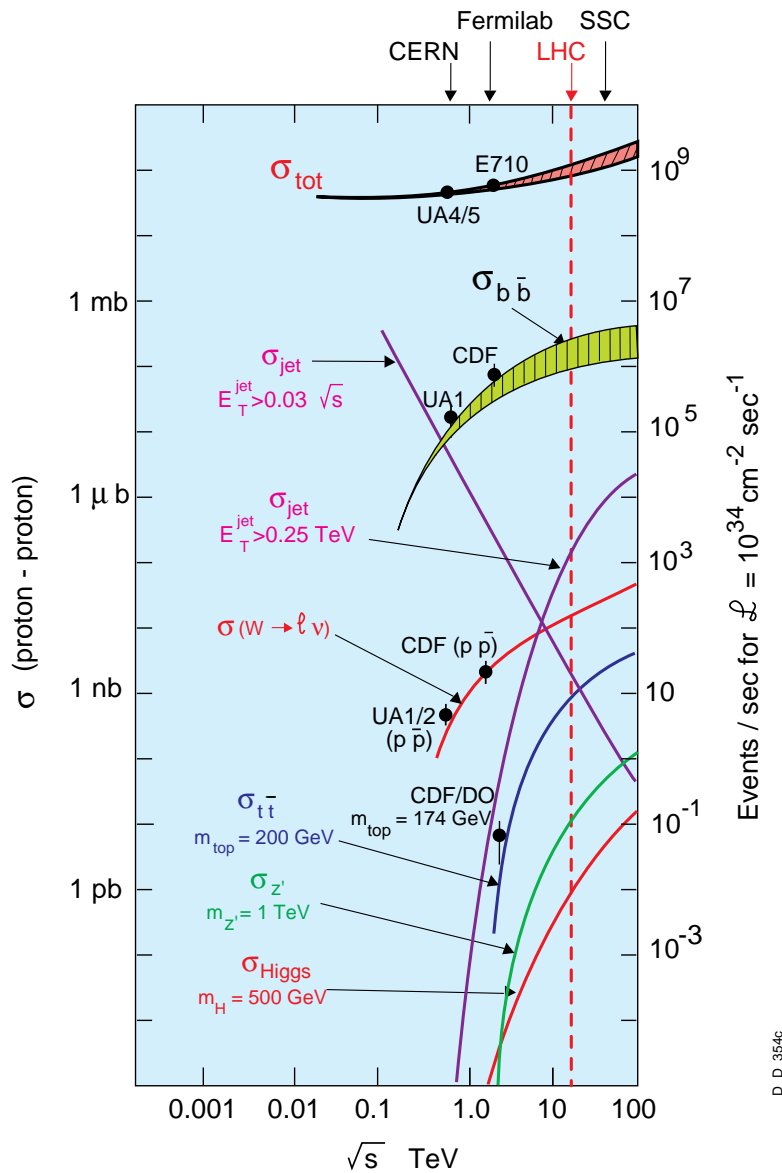

Figure 1.5: *Different total cross sections as a function of the center of mass energy $\sqrt{s}$ of a hadron collider [19].*

1.4   Important quantities of high energy physics

This section introduces certain fundamental quantities which characterize the physics at hadron colliders.

1.4.1   The coordinate system and kinematic quantities

First of all a coordinate system has to be chosen to describe the kinematics of the initial and final state particles. Therefore the nominal interaction point, i.e. the geometric center of the detector, is located at the origin, and the z-axis runs parallel to the beam. The two other axes x and y span the transverse plane which is very important for the measurement of the transverse momentum p_T. In order to obtain complete momentum information, the polar angle θ must be measured.

The direction of particles is defined by two variables: The azimuthal angle φ and the pseudo-rapidity η. The pseudo-rapidity is defined only in the high energy limit where rest masses can be neglected. Here, a connection between η and the angle θ in the rest system of the detector can be utilized:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \; . \tag{1.21}$$

For particles escaping perpendicular to the beam pipe ($\theta = 90°$) the pseudo-rapidity is zero. Towards $+z$ the pseudo-rapidity is positive and rising infinitely as θ becomes smaller. The same behavior can be found towards $-z$ with the difference that the pseudo-rapidity is negative.

The choice of η is convenient as differences in the pseudo-rapidity $\Delta\eta$ are Lorentz-invariants. As partons with different momenta along z interact with each other, each event is subjected to a Lorentz-boost, so invariant quantities are needed. Differences in the azimuthal angle $\Delta\varphi$ are also invariant, so a variable ΔR defining the spatial distance of two detector objects can be constructed:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} \; . \tag{1.22}$$

1.4.2   The kinematic variables p_T and MET

At a hadron collider particles with a substructure (protons) are accelerated. As a consequence all events are boosted in the $\pm z$-direction. The transverse component of the momentum, p_T, is not affected by this boost, so events are characterized best by the transverse momenta of the particles observed by the detector:

$$p_T = \sqrt{(p_x)^2 + (p_y)^2} = p \cdot \sin\theta \; . \tag{1.23}$$

As the center of mass energy is limited, objects with very large transverse momentum have only little momentum along the $z-axis$. They are thus likely to have a small pseudo-rapidity η and

can be measured within the well instrumented central region of the detector.

For all physical objects detected, p_x and p_y can be determined. As the colliding proton beams circle along z, no transverse momentum is present prior to the collision. The law of momentum conservation demands that also after the interaction, no overall p_T is allowed to remain. Momentum balance implies:

$$\left(\sum_{particles} p_{x,y}\right)^{before\ interaction} = \left(\sum_{particles} p_{x,y}\right)^{after\ interaction} = 0 \ . \qquad (1.24)$$

Extending the momentum balance also to the longitudinal z-component of the momentum is not possible as the detector is not hermetic. After the hard scattering, some fragments of the protons can remain undetected while escaping along the beam pipe.

In addition to particles flying in the very forward region of the detector, particles not interacting with the detector material are also not part of the energy balance in p_x and p_y. Examples for this would be neutrinos or non-SM particles which do not interaction with the detector material. With the help of momentum conservation, information about the transverse component of the Missing Transverse Energy (MET or $\not{E}_T$) can be found:

$$\left(\sum_{particles} p_{x,y}\right)^{after\ interaction} + MET_{x,y} = 0 \qquad (1.25)$$

$$\implies MET = \sqrt{(MET_x)^2 + (MET_y)^2} \ . \qquad (1.26)$$

1.4.3   Luminosity and cross section

The most important properties of a particle accelerator are its center of mass energy $\sqrt{s}$ and its instantaneous luminosity l. The reason for this is a simple connection between the event rate $R = \frac{dN}{dt}$, the cross section of a certain interaction σ and the instantaneous luminosity l:

$$R = \sigma \cdot l \ . \qquad (1.27)$$

Event rates are what a detector measures. As the cross section (unit [1 pb=10^{-36} cm^2]) of a specific process is a function of the available energy $\sqrt{s}$, event rates are both affected by the amount of particle acceleration and the luminosity.

In principle, luminosity characterizes the intensity of the two beams heading for each other and their mutual penetration. In order to enhance the probability of a hard process, many particles should meet at an area as tiny as possible. Suppose two colliding proton beams are divided into B separate bunches with a Gaussian bunch profile, each containing n_p particles. The profile of the beam cycling with a frequency f can be described by its widths σ_x and σ_y; luminosity per interaction point is given by:

$$l = f \cdot \frac{B \ n_p \ n_p}{4\pi \ \sigma_x \ \sigma_y} \ . \tag{1.28}$$

The beam optics of the accelerator make sure that the profile of the beam is minimized within the detector using focusing magnets, thus leading to high luminosity and high event rates. For the LHC an additional geometrical reduction factor has to be considered since the proton bunches meet with a certain crossing angle in order to minimize electromagnetic beam-beam interactions. As data are recorded during a certain amount of time $[t_1, t_2]$, Equation 1.27 has to be integrated. The quantity $\mathcal{L} = \int_{t_1}^{t_2} l \ \mathrm{d}t$ is called integrated luminosity (unit $[1 \ \mu b^{-1} = 10^{30} \ cm^{-2}]$) and is a measure of the amount of events seen by the experiment.

Chapter 2

LHC and the CMS-Detector

2.1 The Large Hadron Collider

As mentioned in the Preface the LHC at CERN, Geneva, is about to start its operational mode within the next months. This will be the starting signal for an exciting exploration of physics at a new energy range, e.g. studying the validity of the Higgs mechanism and searching for phenomena beyond the SM. This two-ring superconducting proton-proton collider is now in its final installation and commissioning phase, see Figure 2.1 and [20] for details. Displayed is the former large electron positron (LEP) tunnel with a circumference of 27 km and approximately 100 m below the ground which now hosts the new machine with its design center-of-mass energy of 14 TeV. One can also see the various underground sites (Point 1-8) which either sustain accelerator infrastructures or large experimental caverns. At Point 1 and Point 5 one can find the two large multi-purpose detectors ATLAS and CMS (including the small angle scattering experiment TOTEM), designed for probing physics at the Tera-scale. The LHCb experiment at Point 8 focuses more on precision measurements of the CP-violation and quark mixing while ALICE at Point 2 will look at heavy ion collisions in some future operational mode of LHC. Then this figure emphasizes another important fact: At the date of this figure (August 2008) the LHC beam vacuum system has been installed completely and all sectors have been cooled down to the operation temperature of 1.9 K, see blue color code. Thus the machine is looking forward to the first proton injections to occur. At 10h28, 10 September 2008, this waiting was over: The first beam in the Large Hadron Collider was successfully steered around the full 27 km with both CMS and ATLAS successfully measuring detector signals from the cycling beam. Unfortunately the proclamation of success was muted only shortly after when on 19 September 2008 a serious incident occured during commissioning (without beam) of the final LHC sector (sector 3-4), resulting in a large helium leak into the tunnel and a major damage of various components of the sector. Caused by a faulty electrical connection between two magnets, this lead to a major repair and investigation campain, bringing the date for restart of the accelerator complex to summer 2009. LHC beams will then follow.

In order to provide p-p collisions with an unprecedented design luminosity of 10^{34} cm^{-2}s^{-1} and a center-of-mass energy of 14 TeV various pre-acceleration steps have to be performed. Starting at the proton source and via a linear accelerator LINAC2 the beams are first injected into the existing 25 GeV Proton Synchrotron PS which also generates the LHC bunch structure (25 ns
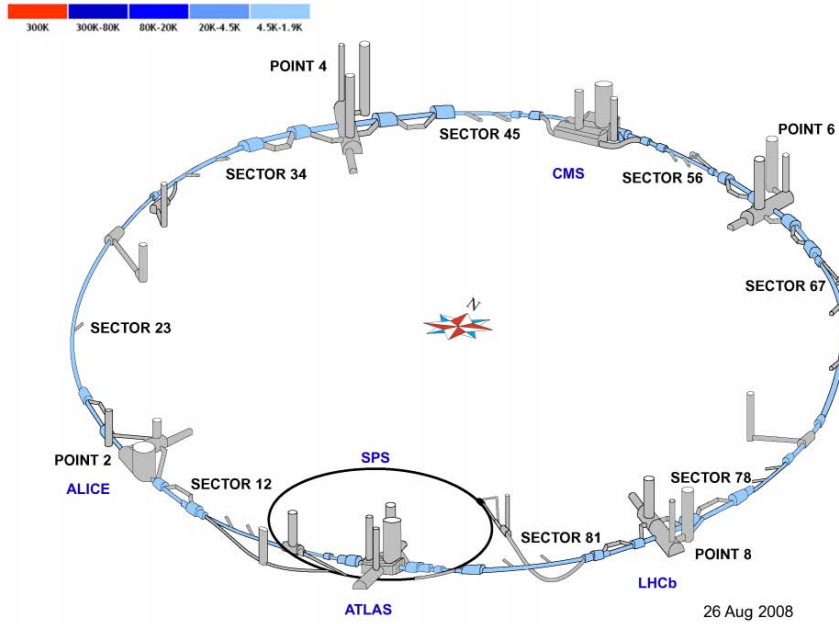
Figure 2.1: *LHC layout including all infrastructure and experimental sites. The color coding refers to the current cool-down temperature of various sectors [21].*

or 75 ns spacings) and then into the Super Proton Synchrotron SPS where they reach an energy of 450 GeV. After the two LHC rings are filled, the machine is ramped to its nominal energy of 7 TeV over about 28 min. In order to operate at such high energies and with such intense beams many innovative features have to be incorporated and challenges have to be faced. The high dipole field of 8.3 T can only be achieved using superconducting (NbTi) magnets with super-fluid helium cooling. In addition to this the old LEP tunnel with a diameter of only 3.8 m provides only limited space, therefore the two rings are incorporated into a single magnetic structure with two sets of coils in a common yoke and cryostat. Now some main features and design parameters of the LHC will be discussed in detail, some performance parameters are also summarized in Figure 2.2

The LHC machine consists of eight arcs with a length of 2.5 km each and eight straight sections of approximately 530 m. The straight sections host the four experiments (two of them also contain the injection systems), one point with two radio frequency systems operating at 400 MHz for the two independent beams, two collimation systems for removing off-momentum and beam halo particles and one point with two beam abort systems. With an energy stored in each beam of more than 350 MJ the safety systems of collimation and external beam dumps have to be highly reliable in order to protect the superconducting machine and the experiments.
The arcs on the other hand consist of 23 regular periods of six dipoles and two short straight sections each. Between the eight arcs and eight straight sections two perturbed periods with only four dipoles are used (dispersion suppressor), resulting in a total of 1232 dipoles needed for the complete LHC. The short straight sections within one arc are used to control the beam optics using the main quadrupoles, correction sextupoles and lattice correctors. At the high luminosity experiments ATLAS and CMS the beam-focusing is achieved using a quadrupole triplet assembly.

Circumference	26.7 km
Beam energy at collision	7 TeV
Beam energy at injection	0.45 TeV
Dipole field at 7 TeV	8.33 T
Luminosity	$10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$
Beam current	0.56 A
Protons per bunch	1.1×10^{11}
Number of bunches	2808
Nominal bunch spacing	24.95 ns
Normalized emittance	$3.75\,\mu$m
Total crossing angle	$300\,\mu$rad
Energy loss per turn	6.7 keV
Critical synchrotron energy	44.1 eV
Radiated power per beam	3.8 kW
Stored energy per beam	350 MJ
Stored energy in magnets	11 GJ
Operating temperature	1.9 K

Figure 2.2: *LHC performance parameters.*

One of the biggest engineering challenges within the LHC project has been the design of the superconducting dipole magnets. With two separate beam lines within a single cryogenic system operating at 1.9 K a magnetic field strong enough to bend the 7 TeV protons can be achieved, resulting in the complex dipole cross section as displayed in Figure 2.3. Using in total 100 t of super-fluid helium enough cooling can be provided to operate the classical and affordable NbTi-superconductor below its critical current density and still achieve a dipole field of above 8 T. On the other hand operation of the superconductor at such low temperatures decreases its heat capacity considerably, making the magnets thus very sensitive to quenches. The two dipoles are both embedded within a common iron yoke, see Figure 2.4 for the magnetic flux within the whole unit. Due to the opposite field directions of the two apertures there is no saturation in the central part of the yoke. The dipole coil itself is wound in two layers in six blocks separated by chopper wedges, an optimized geometry resulting in a pure dipole field. This exact geometry has to be kept constant during the whole series production of the magnets, requesting tight quality controls throughout the whole manufacturing sequence.

In terms of cryogenics the super-fluid helium has the advantage of permeating the smallest cracks which leads to direct contacts between superconductor and fluid. Its high thermal conductivity and large specific heat helps to absorb and dissipate unwanted thermal loads outside the coil. In total eight cryogenic plants are distributed along the LHC. Using a single plant one sector can be cooled down within two weeks.

In terms of radio-frequency acceleration systems two sets of superconducting cavities, providing 16 MV, are enough to ramp up the energy and compensate for the synchrotron radiation losses. In order to ensure an adequate beam lifetime and superconductivity of the dipoles the vacuum system has to minimize the effect of heat sources such as synchrotron light, nuclear scatterings and energy dissipated by the development of electron clouds. In addition to providing a pure vacuum where remaining gas condensates on the cooled vacuum chambers, a dedicated beam screen protects the cold bore of the magnets from the heat of the processes mentioned.

Especially the formation of electron clouds constitutes a new effect in hadron machines and is an
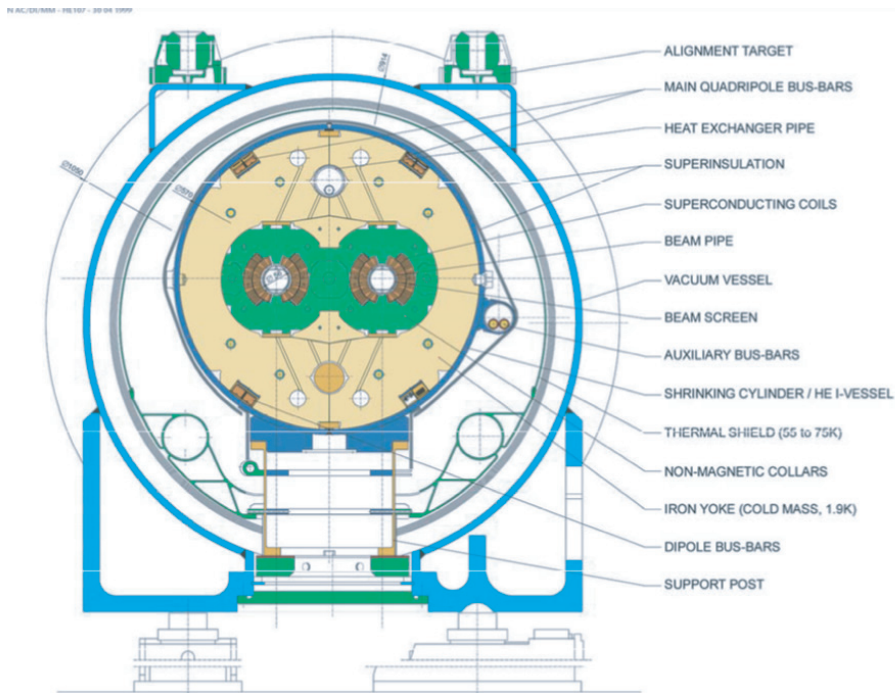
Figure 2.3: *LHC dipole cross section including support structures and vacuum vessel [20].*
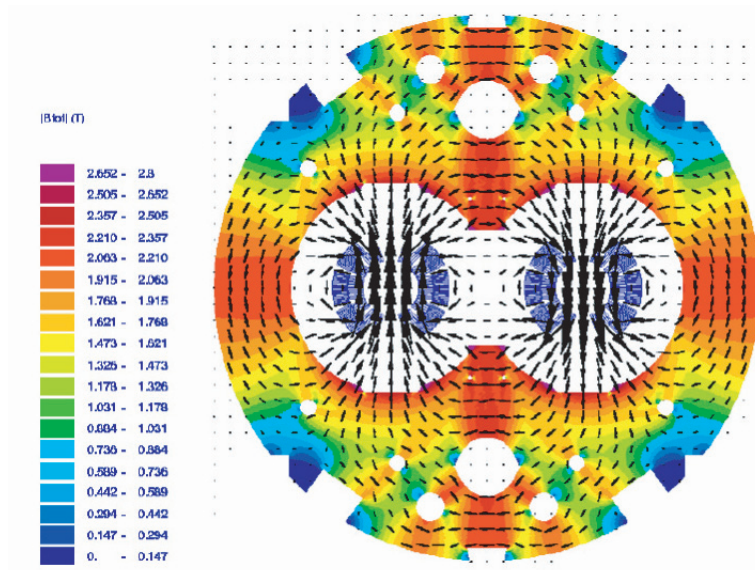
Figure 2.4: *Magnetic flux within iron yoke and beam pipes of the LHC dipole [20].*

interesting accelerator physics issue. Via ionization of residual gas molecules or via synchrotron radiation electrons can accumulate within the vacuum chambers. These electron get a certain momentum transfer induced by the passing proton beam and can thus produce secondaries when interacting with the beam screen. In case of a resonance effect caused by a small bunch separation this electron cloud can grow exponentially and thus limit the performance of the storage rings. As a consequence during early commissioning and first machine operation a 75 ns bunch spacing is foreseen.

In addition to these beam-beam interactions, intra-beam scattering and other particle orbit

instabilities lead to a finite lifetime of the proton beams which can hold a maximum of 2808 proton bunches with $1.1 \cdot 10^{11}$ protons each. With expected lifetimes of about 15 h per fill a maximum integrated luminosity of ≈ 100 fb^{-1} per year can be achieved. For the initial operation though a much lower luminosity ($l < 10^{31}$ cm^{-2}s^{-1}) is planned, colliding only 43 or 156 bunches with less protons per beam, using the wider spacing configuration and, most importantly, at a lower center-of-mass energy of 10 TeV. Still it is expected to finally reach the LHC design parameters by increasing the complexity of the machine operation step by step.

2.2 The CMS detector

The name **C**ompact **M**uon **S**olenoid nicely summarizes the main features of this LHC detector [23][22]: A multi-purpose detector with a compact design determined by an iron return yoke containing a highly performant muon system on the one hand, and a superconducting solenoid coil with 4 T enclosing the calorimeter and inner tracker on the other hand. Its dimensions also underline the idea of compactness: With a length of 21.6 m and a diameter of 14.6 m it is considerably smaller than its counter-experiment ATLAS, but with a weight of around 12500 t much heavier though. Since by now CMS has completed its transition from technical drawings to reality, Figure 2.5 and Figure 2.6 show photographs of the almost completed experiment and illustrate the main design features of CMS.

Figure 2.5 shows the φ-slice view of the barrel parts of CMS. One can nicely see the shell-like structure of the sub-detector components around the nominal interaction point. Closest to the beam line is the silicon tracker, the largest ever built. Outside a ring of PbWO$_4$ crystals guarantees a precise energy measurement of electromagnetically interacting particles. Followed by this the hadronic calorimeter makes for several nuclear interaction lengths to absorb and measure the energy of the remaining particles. This inner core of instrumented detector material is surrounded by the huge superconducting solenoid coil, an engineering masterpiece. The only type of SM particles able to pass through this compact arrangement of material are muons and neutrinos. Thus CMS is completed by several stations of muon chambers which are embedded in the massive iron return yoke needed to guide the magnetic flux.

In addition to the configuration in φ CMS is of course also extended in the z-direction in order to provide a nearly hermetic solid angle coverage. Figure 2.6 illustrates the main design features of CMS in this context: There are 5 barrel wheels and two endcaps, each consisting of 3 disks. One can also see in the picture that these sub-structures can be moved apart from each other, thus facilitating access to the various detector components needed for maintenance. This wheel structure is also related to the assembly strategy of the CMS experiment: All sub-detectors are pre-assembled on the surface and can then be lowered as a whole or in large slices via the access shaft (90 m below the ground). Figure 2.6 was taken during the lowering of the silicon tracker, but also whole wheels had to be lowered in a similar way. The advantages of this assembly strategy are evident looking at the pictures: The CMS detector hardly fits into the cavern, the balconies with detector infrastructure almost touch the outer muon chambers. Such limited spacing would complicate any mounting or the commissioning of detector parts. Thus CMS was pre-assembled in a large surface hall where many people could easily work in parallel on various parts of the detector without affecting each other.

Now the various detector components will be discussed in detail, see [22] and [23] for additional

cavern infrastructure

silicon tracker

electromagnetic
calorimeter

hadronic
calorimeter

solenoid coil

iron return yoke   muon system

silicon tracker
being lowered

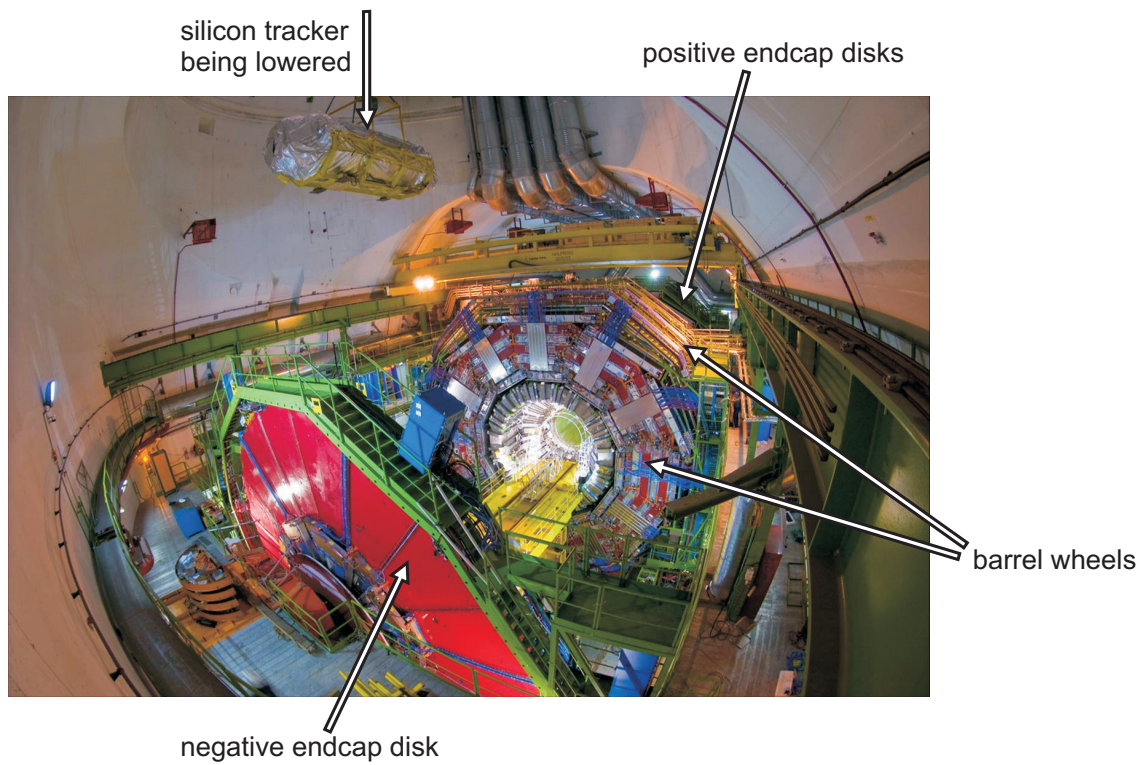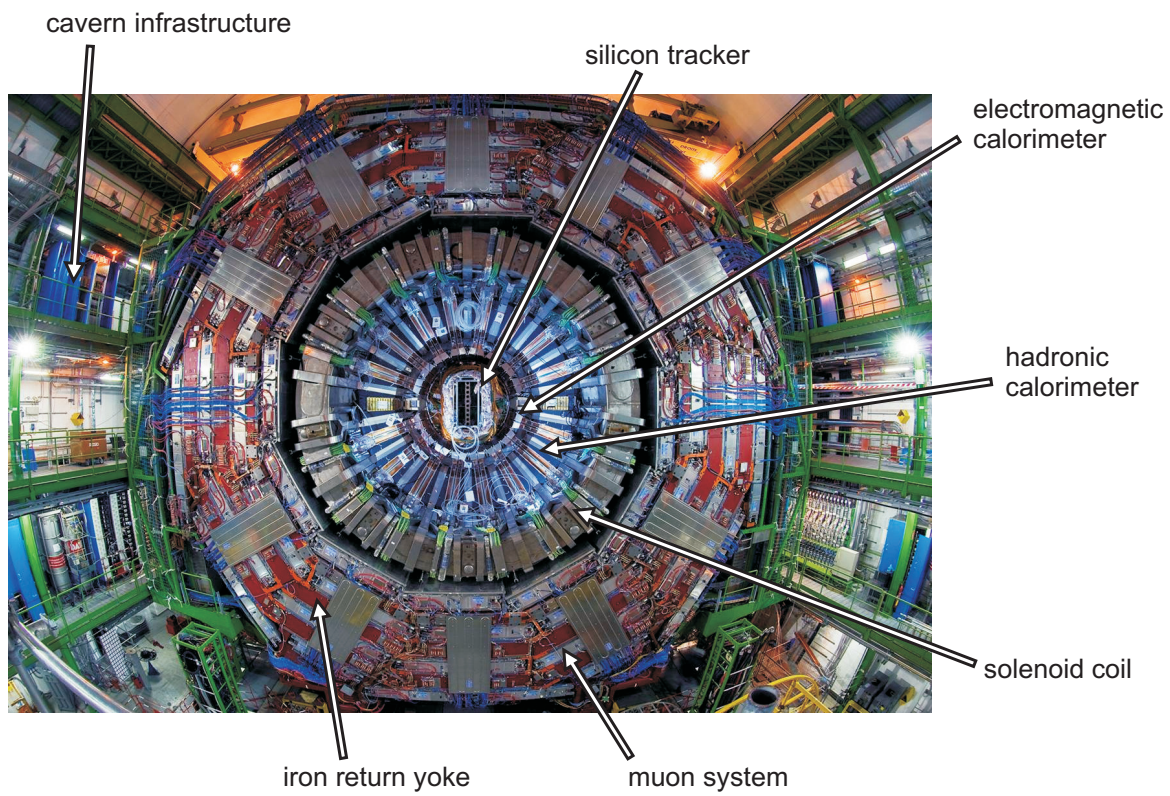positive endcap disks

barrel wheels

negative endcap disk

Figure 2.6: *Barrel and endcap parts of CMS detector in underground cavern, silicon tracker being lowered through access shaft [24].*
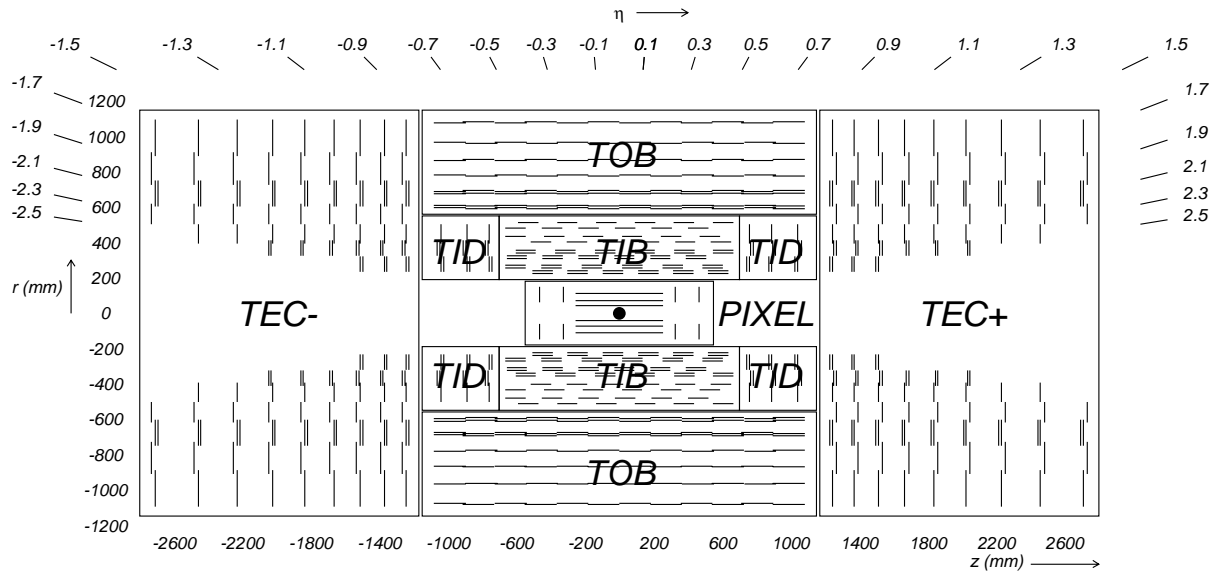
Figure 2.7: *Schematic view in r- and z-coordinates of all detector parts of the inner tracking system [23].*

information. The performance in terms of efficiency and resolution of the various sub-detectors will also be discussed in Section II.3.1 in the context of the analysis presented in the second part of this thesis. In general all sub-detectors have to be designed with high granularity and good time resolution/synchronization in order to cope with the huge particle densities at $l = 10^{34}$ cm^{-2}s^{-1} and the small bunch spacings of 25 ns. In addition to this radiation hardness is important especially for the parts close to the beamline.

2.2.1   The inner tracking system

The requirements of granularity, speed and radiation hardness lead to a tracker design entirely based on silicon detector technology. At the LHC design luminosity there will be on average about 1000 particles from more than 20 overlapping proton-proton interactions, and this harsh environment is repeated every 25 ns. Since the hit rate density decreases with the radius the inner tracking system is composed of silicon detectors with variable sensor sizes, ranging from pixel sensors with 100×150 µm^2 for radii below 10 cm up to silicon strip sensors with 10 cm $\times$ 80 µm and even larger pitches for radii above 20 cm. This occupancy-driven design also satisfies the requirements on position resolution. In this way a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collision is provided, including accurate secondary vertex reconstruction with the help of the pixel detectors.

Figure 2.7 illustrates the overall tracker design. Closest to the beam line three barrel pixel layers and two pixel disks can be found. For radii above 20 cm the silicon strip tracker consists of ten barrel layers and is completed in both endcaps by three plus nine disks (inner disks and endcap). This results in a total length of 5.8 m, a diameter of 2.5 m and a pseudo-rapidity coverage $|\eta| < 2.5$. With about 200 m^2 of active silicon area this is by far the largest silicon tracker ever built.

On the other hand the LHC operation conditions constitute a number of problems and challenges. The intense particle fluxes cause severe radiation damage to the tracking system. In order to guarantee an expected lifetime of ten years several effects had to be considered in the design of a radiation tolerant tracker. Non-ionizing energy loss causes modifications in the silicon crystal lattice. This then results in an increase of the leakage current, a change of doping from n- to p-type and a general reduction of the signal by roughly 10% after ten years of LHC. The change of doping can be compensated by an adjustment of the depletion voltage. The loss in signal magnitude has to be compensated by the design of the silicon sensors and the read-out-electronics, assuring a signal-to-noise ratio of 10 : 1 or better over the full lifetime. The increased detector leakage current can lead to a dangerous effect called thermal runaway. Thus effective cooling is needed in order to avoid this, resulting in a low operating temperature of the whole tracker at $-10°$C at the beginning and $-27°$C after many years of operation.

It is clear that this leads to a rather complex cooling system, using liquid C_4F_{16} for refrigeration. In addition to this the silicon modules themselves, all the needed support structures (made of carbon fibers) and the readout electronics add up to a considerable amount of material budget, up to 1.8 X_0 in units of radiation lengths at $|\eta| \approx 1.4$. Thus multiple scattering, Bremsstrahlung, photon conversion and nuclear interactions limit the performance of the CMS tracker.

One can summarize its performance in the following way: First the pixel detector delivers three high precision space points on each charged particle trajectory. Since several rings and layers of the silicon tracker carry a second micro-strip detector module which is mounted back-to-back with a stereo angle, at least ≈ 9 hits can be measured with the strip detectors, with at least ≈ 4 of them being two-dimensional. With single point resolutions varying between 23 µm and 53 µm (15 µm for the pixel using charge interpolation) this results in a transverse momentum resolution around $1 - 2$% for 100 GeV tracks and $|\eta| < 1.6$ (assuming $B = 4$ T). With the help of the pixel layers a transverse impact parameter resolution around 10 µm for high-p_T tracks can be achieved. For muons the reconstruction efficiency is about 99% over most of the acceptance, while for pions and hadrons it is considerably lower due to interactions with the tracker material.

Figure 2.8 shows a photograph of one completed tracker endcap (TEC). One can see the carbon support structure and the silicon sensors embedded within. The disks consist of several modular elements (petals) and the consecutive petals are rotated by 180° with respect to each other in order to ensure non-zero overlap of the active sensor material.

An important task for a successful operation of the tracker is the alignment. In total the determination of three translational and three rotational parameters for each of the 15148 tracker modules is needed. This can be realized by measuring the assembly precision of the various tracker components, thus leading to an initial knowledge of the geometry. Then there is the Laser Alignment System which uses infrared laser beams and monitors the position of the tracker substructures on a regular basis and with a precision around 100 µm. Finally alignment with tracks, utilizing various dedicated algorithms to solve the huge system of linear equations, should help to determine the ultimate alignment constants. Here tracks from collision events as well as beam halo or cosmic muons are important.

Then it should be noted that the tracker itself is not part of the lowest level of trigger decision
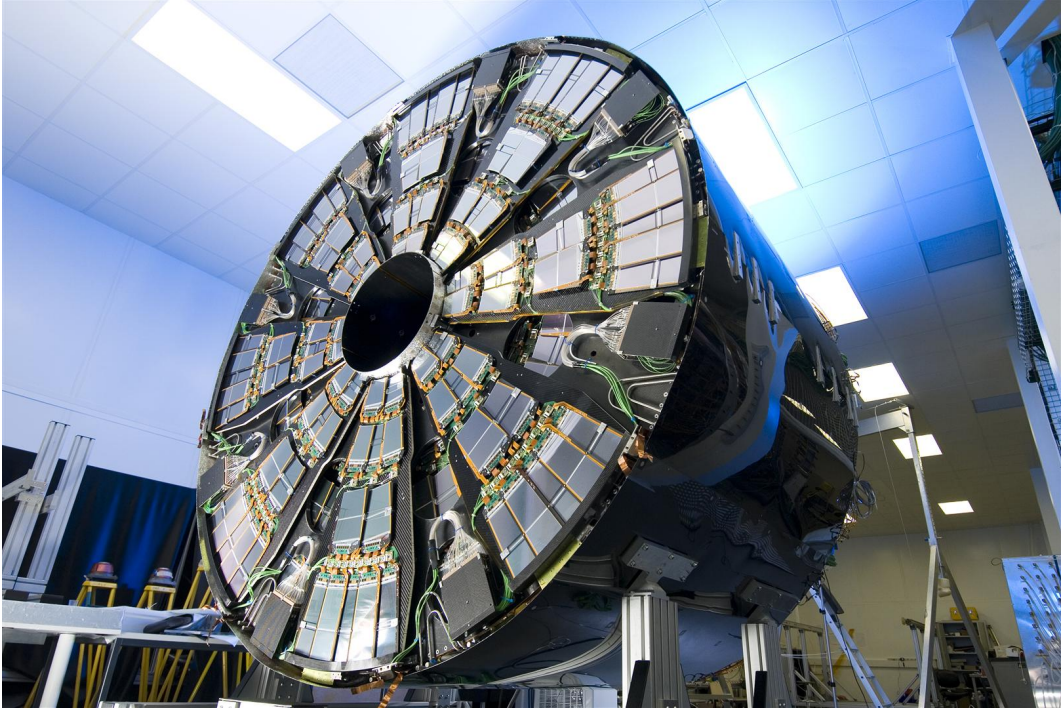
Figure 2.8: *Front view of silicon tracker endcap TEC, petals containing silicon sensors can be clearly seen [24].*

(L1). Thus only using external triggering sources the information of the whole tracker is read out. Still for the high level trigger the tracking information is heavily used in order to reduce the rate of events which can be permanently stored (≈ 100 Hz).

2.2.2 The electromagnetic calorimeter

The idea of compactness while preserving excellent energy resolution, a driving design criterion for CMS, can be nicely seen in the electromagnetic calorimeter (ECAL). The usage of lead tungstate (PbWO$_4$) crystals enables a hermetic and homogeneous calorimeter with fine granularity. This is caused by the characteristics of the crystals which have a high density (8.28 g/cm^3), short radiation length (0.89 cm) and a small Molière radius (2.2 cm). In order to contain the full electromagnetic shower within a single mono-crystal a cross section of 22×22 mm^2 for the front face and 26×26 mm^2 at the rear face is chosen (28.6×28.6 mm^2 for the front face and 30×30 mm^2 at the rear face in the endcap). The short radiation length leads to a compact design with a crystal length of only 23 cm ($\widehat{=}25.8\ X_0$) in the barrel and 22 cm ($\widehat{=}24.7\ X_0$) in the endcaps.

The overall ECAL layout can be seen in Figure 2.9: 61200 crystals are mounted in the central part of the detector (EB) at a radius of 1.29 m, covering $|\eta| < 1.479$ and with a tapered shape slightly varying with position in η. The barrel granularity is 360-fold in φ and (2×85)-fold in η. The crystals are grouped into four types of modules, according to the position in η. These four modules are combined into supermodules which can be attached as whole units to the front face of the HCAL. In total 18 supermodules, each covering $20°$ in φ, form a half barrel.

The endcap part (EE) consists of 7324 identically shaped crystals for each endcap, residing at a longitudinal distance of 3.15 m from the nominal interaction point and covering $1.479 < |\eta| < 3.0$.
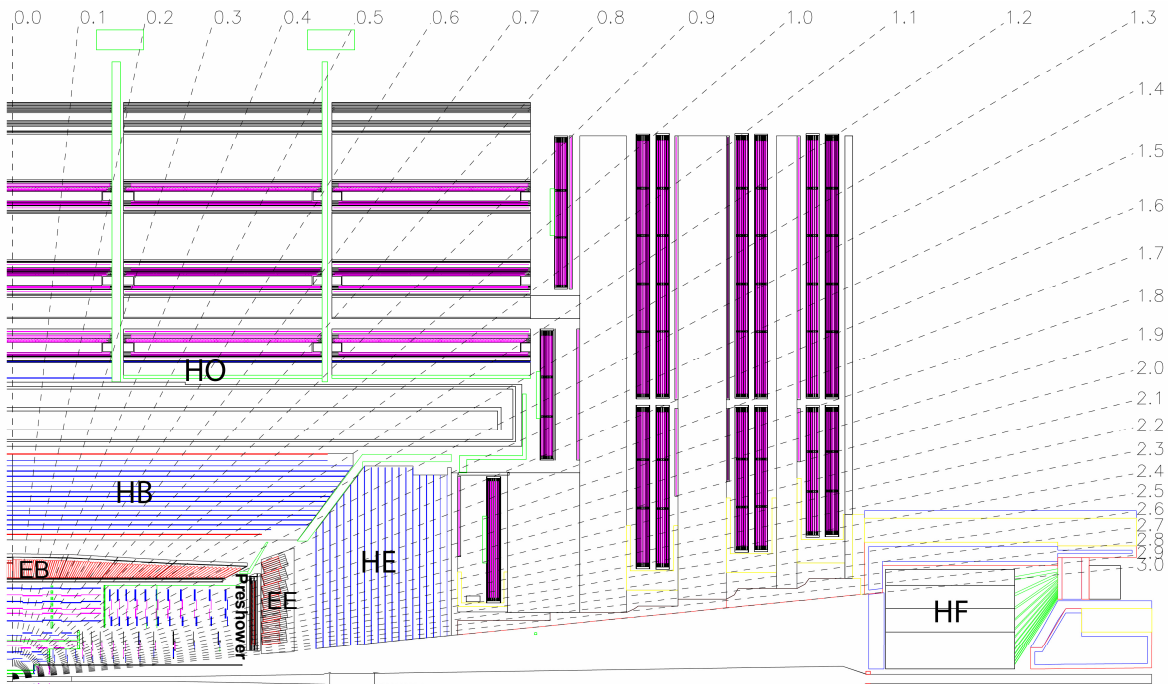
Figure 2.9: *Schematic $r - z$ view of the CMS detector, electromagnetic and hadronic calorimeter detectors are marked [23]. The dotted lines indicate the η-coordinates.*

The crystals are grouped into mechanical units of 5×5 called supercrystals, and these are assembled in two endcap halves called "Dees". Figure 2.10 shows a completed Dee shortly before installation into CMS. The groupings of supercrystals as well as the individual crystals can be nicely seen.

Both barrel and endcap crystals are mounted in a quasi-projective geometry, thus their axes make a small angle (2-8°) with respect to a particle trajectory from the nominal interaction point. In this way the energy loss caused by the cracks between the crystals (< 0.5 mm) is minimized.

As discussed earlier the LHC environment requires high granularity, speed and radiation hardness. The fine segmentation of the ECAL already provides high granularity while ensuring a good energy resolution. The scintillation properties of $PbWO_4$ lead to a fast readout of the light yield where about 80% of the light (broad maximum at $420 - 430$ nm) is produced within the LHC bunch spacing of 25 ns. In the context of light output the issue of temperature stability is important since with increasing temperature the number of scintillation photons emitted decreases ($-2.1\%\ °C^{-1}$ at $18\ °C$). Using water cooling the nominal operating temperature of the ECAL is $18.00 \pm 0.05\ °C$, including a thermal screen with respect to the tracker operating at $-10°C$.

The radiation levels and particle fluxes expected for the duration of the CMS experiment lead to a wavelength-dependent loss of light transmission in the crystals, caused by the formation of absorption bands due to imperfections of the crystal lattice. This aging effect can be traced and corrected for by monitoring the optical transparency. The evolution of the transparency is measured using laser pulses injected via optical fibers and normalizing this with respect to a measurement of the pulse using silicon PN photodiodes. One should emphasize that in general a dose-rate dependent equilibrium between damage and recovery at $18\ °C$ leads to a constant
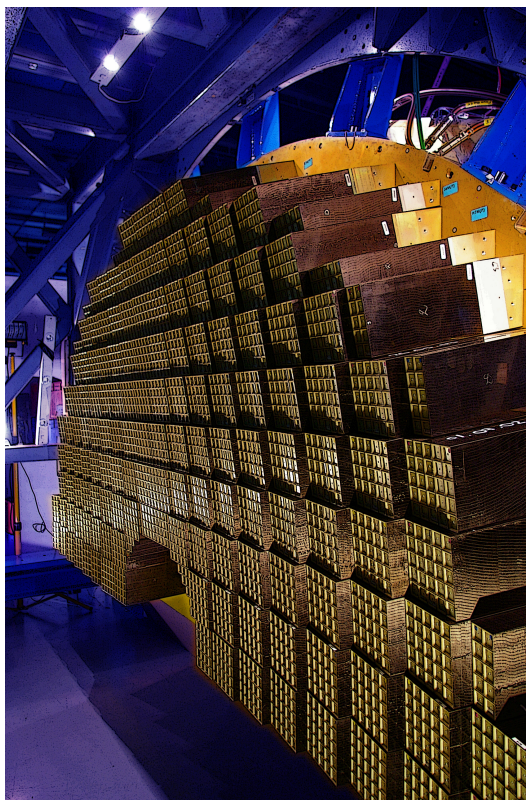
Figure 2.10: *One half of the completed ECAL endcap (one Dee out of four) shortly before installation into CMS [24].*

optical transparency, which is superimposed by a cyclic modulation caused by the varying conditions between LHC collision runs and machine refills. Both, the constant change as well as the cyclic modulation needs to be measured and corrected for in oder to ensure optimal resolution performance.

The scintillation light produced by the crystals then has to be converted into an electronic signal using photodetectors. Also here the LHC environment leads to the requirements of speed, radiation hardness and insensitivity to the strong magnetic field. Since in the barrel region the particle fluxes are lower and the magnetic field is higher, avalanche photodiodes were chosen. In the endcaps where the level of radiation is considerably higher vacuum phototriodes are used, i.e. photomultipliers with a single gain stage. Given the small light yield of the crystals both types of detectors need to have a sufficient quantum efficiency and gain in order to efficiently amplify the signal. While the avalanche photodiodes are superior in terms of gain and quantum efficiency (quantum efficiency = 75% for avalanche photodiodes and = 22% for vacuum phototriodes, gain = 50 for avalanche photodiodes and = 10 for vacuum phototriodes) this is compensated by the larger surface coverage on the back face of the crystals of the phototriodes (active area $= 2 \times (5 \times 5 \, \text{mm}^2)$ for avalanche photodiodes and $280 \, \text{mm}^2$ for vacuum phototriodes). Both detector types are insensitive to particles traversing them (nuclear counter effect). While for the avalanche photodiodes special care has to be taken to provide a very stable power supply and thus stable gains, in the case of vacuum phototriodes a light pulsar system delivers a constant background rate since the anode sensitivity depends on the anode current.

An additional component of the ECAL is the preshower detector in the endcaps, see Figure 2.9, covering $1.653 < |\eta| < 2.6$. With two layers of lead radiators and silicon strip sensors this is a sampling calorimeter measuring the deposited energy and transverse shower profiles. The high granularity of the silicon sensors (1.9 mm pitch) contributes to the discrimination between neutral pions, minimum ionizing particles and photons/electrons. The material thickness of the preshower corresponds to 2 X_0 in the first plane and 1 X_0 in the second absorber, adding up to a total thickness of the preshower detector of 20 cm.

The performance of the ECAL in terms of energy resolution can be summarized in Equation 2.1:

$$\frac{\sigma(E)}{E} = \sqrt{\frac{S^2}{E} + \frac{N^2}{E^2} + C^2} \ . \tag{2.1}$$

The first term S is the stochastic term and consists of three contributions: event-to-event fluctuations in the lateral shower containment, a photo-statistics contribution caused by the readout using photodetectors and fluctuations of the energy deposit within the preshower absorbers. In total the stochastic term is measured to be $S = 2.8\%$ during a 2004 CERN test beam.

The second term, the noise term N, accounts for electronics noise, digitization noise and pileup noise. While the origin of electronics and digitization noise is clear, the pileup noise arises from the fact that the signals from the preamplifier will extend over several LHC bunch crossings, thus additional particles can cause overlapping signals. The noise term contribution is measured to be $N = 0.12$.

The third term C is called constant term, some main contributions are non-uniformities of the longitudinal light collection, channel-to-channel calibration errors and the effect of rear leakage for late developing showers. In total test beam measurements estimate $C = 0.30\%$.

For 100 GeV electrons this would correspond to a relative energy resolution of only 0.4%. However this value can worsen to a few percent in case the electron emits Bremsstrahlung on its passage through the material before the ECAL.

2.2.3   The hadronic calorimeter

Radially restricted by the ECAL and the solenoid coil ($R = 1.77 - 2.95$ m) resides the hadronic calorimeter (HCAL). Designed to identify and measure the energy of hadronic jets it has a nearly hermetic pseudo-rapidity coverage (up to $|\eta| = 5.2$) in order to provide a good missing transverse energy determination. As shown in Figure 2.9 the HCAL consists of several detector components which will be discussed in the following: The hadron barrel (HB, $|\eta| < 1.3$) supplemented by the hadron outer (HO, outside the coil), the hadron endcaps (HE, $|\eta| < 3$) and the forward calorimeter (HF, $|\eta| < 5.2$) placed at 11.2 m from the interaction point.

Similar to the ECAL barrel design the HB is composed of two half-barrels. Each of them is segmented into 18 wedges, covering $\varphi = 20°$ each, and segmented into 16 η sectors. The HB is a sampling calorimeter where flat brass absorber plates ($50 - 56.5$ mm thick) alternate with 17 layers of plastic scintillator tiles with embedded wavelength shifting fibers. Each wedge is again subdivided into four φ-sectors which are arranged in a staggered geometry in order to avoid
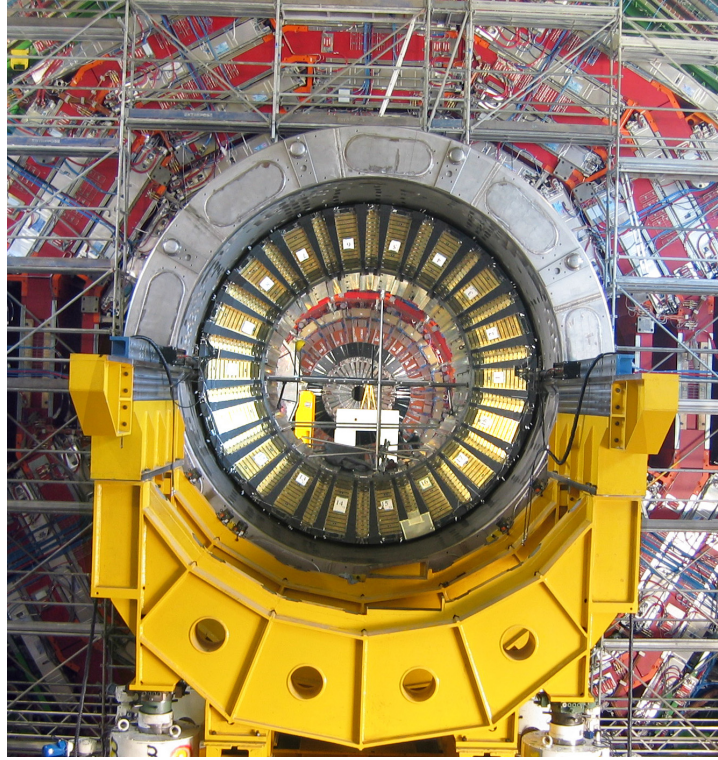
Figure 2.11: *Barrel hadronic calorimeter (HB) after installation inside the solenoid coil, the wedges made of brass plates can be seen nicely [24].*

projective dead material. Figure 2.11 shows a picture of one half-barrel after installation inside the CMS coil. The wedge structure as well as the individual layers of brass plates can be seen. This setup corresponds to an absorber thickness of 5.82 interaction lengths (λ_I) at $|\eta| = 0$ and $10.6\lambda_I$ at $|\eta| = 1.3$.

The HB has a tower-like readout, thus all the light collected by the various scintillator tiles of one segment ($\Delta\varphi = 5°$ and $\Delta\eta = 0.087$) is directed via fibers to a hybrid photodiode (gain ≈ 2000), thus neglecting any longitudinal segmentation information.

For calibration purposes radioactive sources as well as ultraviolet laser light are utilized to monitor the scintillator tiles.

In order to supplement the containment of hadron showers in the barrel part an outer calorimeter is placed outside the solenoid coil, operating as a tail catcher for late starting showers. The HO is constrained by the geometry of the muon system, thus the 5-fold wheel structure and the 12 sectors are reflected in the setup of the HO layers. For all wheels the sensitive layers are placed in front of the first muon station and thus in front of the iron return yoke. While for the central wheel YB0 two layers of scintillator tiles surround a 19.5 cm piece of iron for the other wheels a single layer is used with only the coil as an additional absorber. As concerns segmentation and readout the HO reflects the tower structure of the HB, thus the units of tiles which are combined roughly correspond to $\Delta\varphi = 5°$ and $\Delta\eta = 0.087$. As the main physics impact the HO is expected to reduce the effect of energy leakage which leads to energy measurements that are more Gaussian in nature and as a direct consequence to a considerable improvement of the missing transverse energy measurement.

Two endcap plugs cover the range $1.3 < |\eta| < 3$, its basic functionality is similar to the barrel part. Non-magnetic brass plates act as absorber material while 18 layers of trapezoidal-shaped scintillator tiles produce light which is transmitted by optical fibers to hybrid photodiodes. Including the ECAL the total length of the HE is about 10 interaction lengths (λ_I). Similar to the HB also in the endcaps a staggered geometry is used to avoid projective dead material. In addition cracks in the transition region between HB and HE are minimized. A particular engineering challenge is caused by the fact that the endcaps are attached to the muon endcap yoke as shown in Figure 2.15. The large weight of about 300 t has to be carried by only a small amount of support structure, with most of the space between HE and iron yoke disk (YE1) used by muon cathode strip chambers and the ECAL endcaps being attached to the face of the HE.

As concerns segmentation the HE consists of 14 η-towers and 36 wedges in φ. More specifically the granularity of the HE is $\Delta\varphi \times \Delta\eta = 5° \times 0.087$ for $|\eta| < 1.6$ and $\Delta\varphi \times \Delta\eta = 10° \times 0.17$ for larger $|\eta|$. The longitudinal segmentation of the HE is finer than in the barrel part since for most towers 2 divisions in depth are read-out separately, the towers nearest to the beam line even have 3 read-outs. In this way the HE can also partly act as an electromagnetic calorimeter, a circumstance useful since the completion of the ECAL endcaps in time with LHC start-up has not been certain all the time.

The design of the forward calorimeter (HF) is quite different from the rest of the HCAL and reflects the hostile environment at 11.2 m from the interaction point at a distance between $R = 125 - 1300$ mm from the beam-line. On average, an energy of 760 GeV per proton-proton interaction is deposited into the two forward calorimeters, thus radiation hardness is a considerable challenge. Quartz fibers were chosen as the active medium and are embedded within 165 cm of steel absorber structure ($\approx 10\lambda_I$). Within the fibers charged shower particles can generate Cherenkov light, thus measuring mainly the electromagnetic component of showers. Surrounded by steel, lead and polyethylene shielding the light is detected by photomultiplier tubes. As concerns radiation hardness the Quartz fibers are expected to suffer from a loss of optical transmission by 50% after 10 years of LHC running.

18 wedges make up one HF calorimeter and bundles of fibers define a tower segmentation of $\Delta\varphi \times \Delta\eta = 10° \times 0.175$. In longitude the detector is divided into two segments. Half of the fibers run over the full depth of the absorber parallel to the beamline. The other half starts only at a depth of 25 cm. In this way one can distinguish electromagnetic from hadronic showers (note that ECAL only extends to $|\eta| < 3.0$) since hadrons deposit equal amounts of energy in both segments while electrons and photons are stopped mostly within the first 25 cm.

Another important task of the HF will be the CMS luminosity measurement on a bunch-by-bunch basis and to provide overall normalization for physics analyses. A determination of the luminosity can be derived from the mean number of interactions per bunch crossing and the average transverse energy per tower, which can both be measured using the HF (see also Section 2.2.7).

Due to the fact that both calorimetric compartments, the ECAL and the HCAL, are non-compensating also the calorimetric response to the deposited energy is different for electrons and hadrons. This leads to an energy dependent response and to a degradation in the jet resolution. The non-corrected energy resolution of the full ECAL+HCAL barrel calorimeter, as measured by test beam, can be parameterized as [22]

$$\frac{\sigma(E)}{E} = \sqrt{\frac{(120\%)^2}{E} + (6.9\%)^2} \; . \tag{2.2}$$

Using clustering techniques that utilize the separate energy measurements of the ECAL and HCAL the stochastic term can be significantly improved to $S = 70\%$.

2.2.4   The superconducting magnet

A distinctive design feature and engineering challenge of CMS is the solenoid coil, the biggest superconducting solenoid ever built, with a stored energy-over-mass ratio well beyond the values of previous detector magnets. With a diameter of 6 m and a length of 12.5 m the dimensions are large enough to contain all calorimeter and tracking parts of CMS, providing a homogeneous magnetic field of up to 4 T. The flux is returned through a 10000 t yoke consisting of 5 barrel wheels (YB) and 6 endcap disks (YE), partly saturated at 1.8 T. Using air pads these heavy elements can be moved apart, thus facilitating the assembly of and future access to the sub-detectors. Figure 2.12 shows a picture of the cold mass being inserted into the outer vacuum tank before the central wheel (YB0) was lowered to the underground cavern. Several bare yoke elements (red) can also be seen.

The 220 t cold mass of the coil resides in a vacuum vessel including some thermal shielding. Using liquid helium the superconductor is operated at 4.6 K, well below the critical temperature of the NbTi superconductor used for the solenoid. In order to allow for a current strong enough to generate the 4 T magnetic field the winding is composed of 4 layers. The conductor itself is mechanically reinforced with aluminum and in this way also constitutes a large part of the structural material. The ratio between the energy stored in the magnetic field (350 MJ in total) and the cold mass of 220 t is so high (1.6 kJ/kg) that strong magnetic forces and large mechanical deformations (0.15%) have to be tolerated. Due to the self-supporting conductor in addition to the support cylindrical mandrel the radial extent of the coil can be kept small ($\Delta R/R \sim 0.1$), thus minimizing effects of multiple scattering and particle absorption.

Any unexpected superconductive-to-resistive transition in the conductor, a so-called quench, constitutes a serious danger to the magnet. Here the mutual coupling to the external mandrel plays an important role: Eddy currents are induced in the mandrel and uniformly heat up the whole coil above the critical temperature. In this way local overheating and thus thermal stresses inside the windings are minimized. In case of a quench or other technical problems a fast discharge is triggered, extracting most of the magnetic energy by thermal dissipation in the so-called dump resistor positioned outdoors. A slower and more controlled current dumping is the slow discharge where the coil stays in the superconducting state as the heat load is fully absorbed by the cooling refrigerator.

In summer 2006 a large operating test of the magnet has been performed, the Magnet Test and Cosmic Challenge (MTCC, see also Chapter I.1 in the first part of this thesis). This test took place in the CMS assembly hall above ground and included the successful cool-down of the magnet, a continuous operation at the design field of 4 T and a detailed mapping of the field. In addition to this various other vital measurements were performed, e.g. the electrical, magnetic, thermal and mechanical behaviors during charge and discharge cycles, the magnetic misalignment of the coil or the mechanical displacements of the yoke in the context of the strong magnetic forces present.
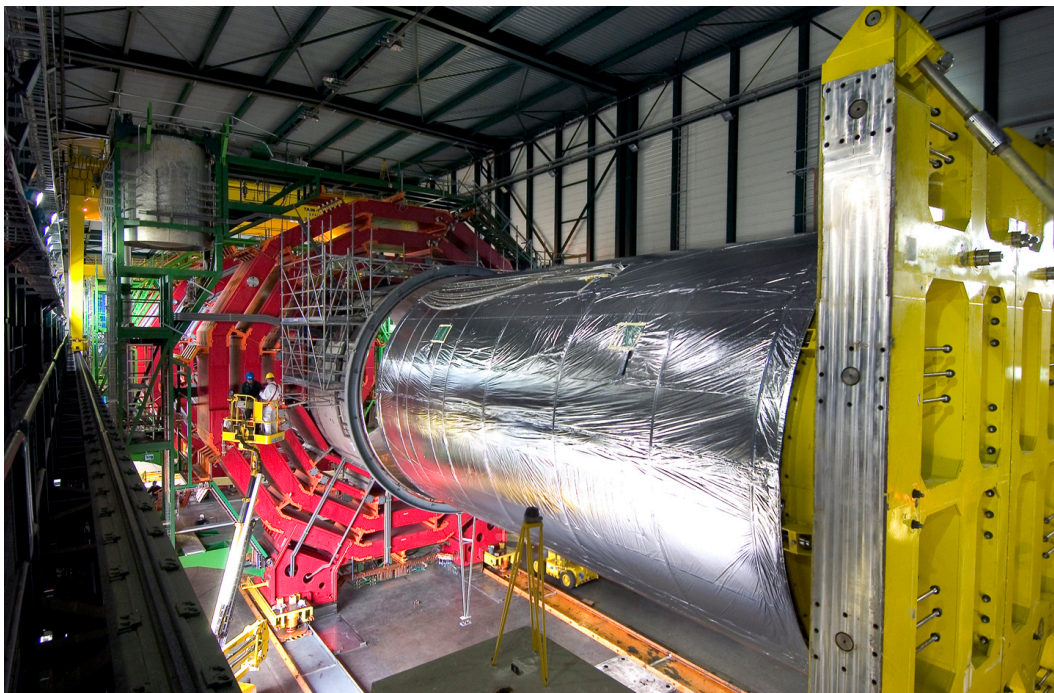
Figure 2.12: *Cold mass of the solenoid coil being inserted into the outer vacuum tank [24].*

2.2.5   The muon system

As implied by the experiment's middle name, the identification and momentum measurement of muons is of central importance to CMS. Due to the considerable amount of material before the muon system (calorimeter and solenoid coil) and within the system itself (iron return yoke), even at the highest LHC luminosity muon detection promises to guarantee clean signatures, be it the gold plated Higgs discovery channel $H \to 4\mu$ or new physic's signals involving muons. Thus a muon system with good muon momentum resolution, good trigger capability and wide angular coverage is needed. Because of the sheer size of CMS these muon detectors have to be inexpensive, reliable and robust. Three types of gaseous muon chambers were chosen: For measuring the momentum drift tubes (DT) are used in the barrel wheels, while in the endcap disks cathode strip chambers (CSC) cope with the higher background rates. The main purpose of the resistive plate chambers (RPC) in both barrel and endcap regions is to provide a fast and independent trigger, in addition to the self-triggering capabilities of the DTs and CSCs. All three sub-systems will be discussed in the following, but particular focus will be placed on the DT chambers since they play a major role in the first part of this thesis (Chapter I.1 – Chapter I.6).

To summarize the performance of the muon system a global reconstruction efficiency, thus tracker combined with muon system, between $95 - 99\%$ can be seen in almost the whole pseudo-rapidity coverage ($|\eta| < 2.4$) and a muon momentum resolution of the standalone muon system of about 9% for small values of $|\eta|$ and $p < 200$ GeV (combined fit with tracker gives an improvement by an order of magnitude).

The barrel part of the muon system ($|\eta| < 1.2$) consists of the 5 wheels of the CMS detector. For each wheel 4 types of drift tube chambers, called stations, form concentric cylinders around the beam line. Three iron layers are embedded within the yoke, thus the first muon station, MB1, is

placed before, the next two, MB2 and MB3, are between iron layers and the last station, MB4, is placed outside. This choice of multiple stations ensures a reliable identification and measurement of a muon. The combination of the 4 stations forms a sector in φ, in total 12 sectors make up the concentric cylinder, see Figure I.1.1.

The basic detection element of a drift tube chamber is a 42 mm × 13 mm drift cell (wire-length ≈ 2.5 m, depending on chamber type) filled with an Ar/CO_2 gas mixture, see Figure 2.13. Four layers of staggered drift cells form a group, called a Superlayer (SL), with three Superlayers making up a chamber, in total 12 layers of drift cells. Redundancy is obtained by having several layers of separated drift cells per station. The bending of the muon trajectory in the r-φ plane of CMS is measured by the two ϕ-SL while the third, the θ-SL, determines the perpendicular coordinate in the r-z plane. A honeycomb panel of 128 mm thickness placed between the θ-SL and one ϕ-SL gives rigidity and a bigger lever arm to the ensemble. The only exception to this scheme are the outermost MB4 chambers containing only two ϕ-Superlayers. Both units, a single drift cell as well as a whole DT chamber, are displayed in Figure 2.13. Some installed DT chambers can also be seen in the photograph of Figure 2.15.

The transverse dimension of the drift cell, i.e. the maximum drift time of 380 ns, was chosen to limit the number of active channels while also producing a negligible occupancy in a single cell. In total there are about 172000 sensitive DT wires in CMS. The relatively thick tube walls help to decouple one cell from another and thus protect against the electromagnetic debris accompanying the muon itself. As shown in Figure 2.13 the cell itself consists of a rather complex setup of anode wire, cathode strips and field shaping electrode strips which are all set to different values of voltages. With the help of this cell optics a linear relationship between the distance from the wire of the crossing track and the drift-time of the electrons along the entire drift path is maintained to a good extent (deviation from linearity of the space-time relation in each drift cell are less than $100 - 150$ µm). These non-linearities can also be seen in the characteristics of the isochrones, i.e. equipotential lines of identical drift time. Especially for large track angles deviations can be expected. In addition to this some chambers (especially MB1) are placed within considerable magnetic stray fields (up to 0.8 T), leading to a distortion of the drift lines and thus to a deviation from the linearity of the space-time relation. In general the applied voltages and the gas mixture lead to a cell operation at a drift velocity saturation of ≈ 54 µm/ns and a gas gain of 10^5.

The drift cells of each layer within a SL are offset by a half-cell width with respect to their neighboring layer to eliminate dead spots in the efficiency. This arrangement also provides a convenient way to measure the muon time with excellent time resolution (a few nanoseconds), using simple meantimer circuits. In this way the cell hits can be assigned to the correct bunch-crossing in an efficient and standalone way. This is important since the maximum drift time of the electrons is large compared to the 25 ns bunch spacings. The target chamber resolution for 8 track points measured by both ϕ-SLs is 100 µm, while a single wire resolution is around 250 µm. Thus in addition to the requirements of linear space-time relation also the mechanical precision of the construction of a chamber is of major importance and had to be retained throughout the whole process of chamber mass production.

The high voltage services and front-end electronics (amplification, thresholds) are located at opposite ends of the wires, whereas the gas enclosures are on both sides of the wire. In order to avoid contamination by nitrogen and oxygen, which changes the drift velocity, the Superlayers are gas tight. The read-out and trigger electronics that collect the full chamber information are
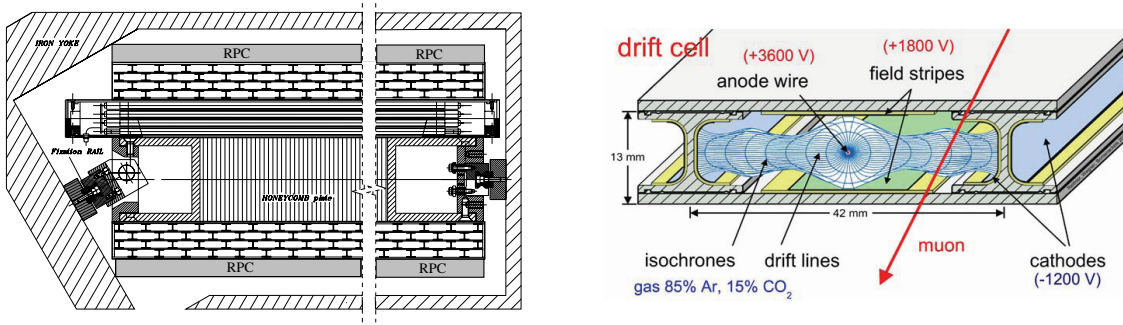
Figure 2.13: *Left: A DT chamber in position inside the iron yoke, mounted using the honeycomb plate [23]. One can see the cell structure of the two ϕ-Superlayers which sandwich the θ-Superlayer with wires perpendicular to it. The RPCs are glued to the bottom and/or top faces of the DT chambers, depending on chamber type.*
Right: Sketch of a cell showing drift lines and isochrones, the voltages applied to the wires, the stripes and the cathodes are also indicated.

contained in the so-called "minicrates", an aluminum structure attached to the honeycomb of the DT. These minicrates also incorporate the time digitization as part of the Read-Out Boards (ROB), using a high performance TDC with a time bin of 25/32 ns.

One major challenge of the mass production of the DT chambers is to build and assemble them with a constant level of high quality. In order to ensure this each chamber has to undergo various testing steps at various production stages. First of all the chambers are extensively tested at the production sites, e.g. in Aachen where all MB1 chambers have been built, before delivery to CERN. At CERN the DTs are equipped with additional service items and then tested with a high voltage long term test, a gas tightness test and an extensive period of cosmic muon ray data taking. In addition to this various alignment surveys are done in order to monitor the position of the whole chamber as well as individual wires with respect to the CMS coordinate system. After this the chambers are finally installed into the iron pockets of the CMS yoke.
Besides the tests at the stages of production and installation also dedicated test beams, chamber commissioning after installation in CMS and the Magnet Test and Cosmic Challenge (see also Chapter I.1 – Chapter I.6) have been conducted to analyze the performance of the chamber and trigger. Chamber efficiencies, trigger properties and bunch crossing assignment, cell resolutions, calibration properties, noise measurements as well as deviations from the space-time linearity in the presence of a magnetic field have been studied to great extent. In all cases the design goals of the DTs have been met and a successful operation under LHC conditions is ensured.

The endcap disks consist of 3 layers of iron which are instrumented with 468 cathode strip chambers. Similar to the DT naming convention these iron layers define 4 stations, ME1-ME4. The chambers are trapezoidal and cover either 10° or 20° in φ, forming a ring of chambers around the beam line. In the first station there are three rings of chambers, for ME2 and ME3 there are two rings and placed outside the return yoke only a single ring of ME4 chambers is present. This layout of the muon endcap system is also shown in Figure 2.14 and the rings of CSCs can be seen in the photograph of Figure 2.15. The endcap muon system provides a pseudo-rapidity coverage up to $|\eta| < 2.4$ with $3 - 4$ CSCs crossed by a muon or both DTs and CSCs hit in the endcap-barrel overlap region $(0.9 < |\eta| < 1.2)$.
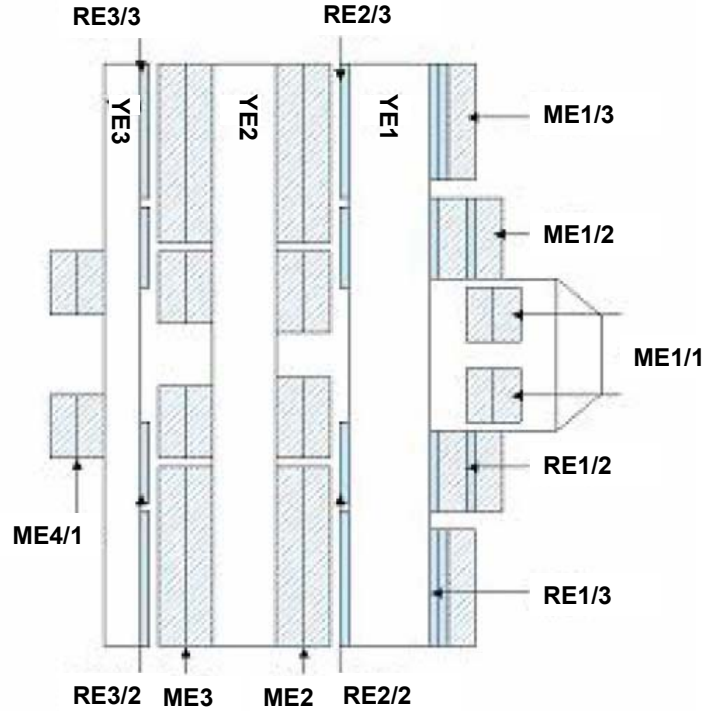
Figure 2.14: *Schematic layout of the CMS endcap for the initial muon system [23]. CSCs are labeled with ME and RPCs with RE, numbers correspond to station and ring affiliation.*

The CSCs are multi-wire proportional chambers comprised of 6 anode wire planes interleaved with 7 cathode panels. Wires run azimuthally and define a track's radial coordinate. Strips are milled on cathode panels and run lengthwise at constant $\Delta\varphi$ width. The chambers are up to 3.4 m long and up to 1.5 m wide.

The wire spacing of about 3.5 mm and the grouping of 16 wires per anode read-out channel defines the resolution in the global $r - z$ coordinate (≈ 5 cm). On the other hand the bending due to the magnetic field which provides momentum information is determined measuring the global φ angle, thus the muon coordinate along the wires. Here the resolution is determined by the strip pitch, which varies from 8.4 mm at the narrow chamber end to 16 mm at the wide end, and by the gas gap between anode wires and cathode panels, which is 9.5 mm. The avalanche produced by the primary ionization near the wire induces charges on the cathode plane. Since the induced charge is shared among $3 - 4$ strips one can interpolate and thus determine the position along the wire with a precision up to 150 μm. Due to differences in the mechanical design the inner ring of ME1 can achieve resolutions up to 75 μm and since these chambers are inside the solenoid the wires a tilted by the Lorentz angle to account for the shift of the drifting electrons caused by the strong magnetic field. In all cases a gas mixture of $40\%\mathrm{Ar} + 50\%\mathrm{CO_2} + 10\%\mathrm{CF_4}$ is used.

The resolution figures illustrate that the CSCs can provide precise muon momentum measurement. In addition to this excellent triggering properties are of major importance, a design goal is 99% efficiency in finding a track and assigning it to the correct bunch crossing. These specifications have to be seen in the context of non-uniform magnetic fields and high background rates, especially caused by the neutron induced photon background, which takes place in the endcap regions of CMS. Moreover the CSCs are very robust detectors which do not require precise gas,
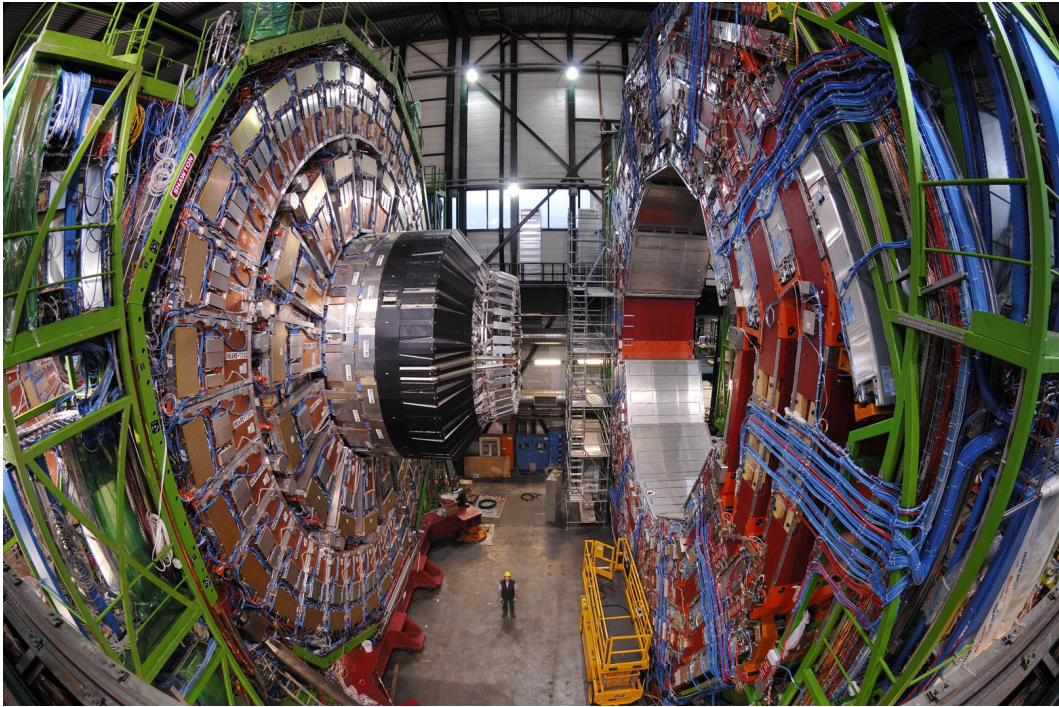
Figure 2.15: *Endcap (left) and barrel (right) wheels close to each other [24]. The instrumentation with CSC as well as with DT chambers can be seen, the hadron endcap calorimeter (center) being mounted on the endcap iron yoke.*

temperature or pressure control.

The triggering properties and efficient rejection against spurious single plane hits are realized using pattern recognition of track-like hits in multiple planes. Local Charged Track candidates (LCT) are formed both by the anode wires as well as by the cathode strips and the information is passed on to the Level 1 trigger decision. The narrow timing distribution ($\sigma < 5$ ns) of the 2nd or 3rd hit in an anode LCT pattern can also be used for reliable bunch crossing tagging.

Similar to the DT chambers also in the CSC system various tests conducted in high-energy muon beams at CERN, with cosmic ray muons in a lab or in situ after installation, and at the Gamma Irradiation Facility (GIF) showed that all design performance parameters are met and that the CSCs have proven to be reliable in operation.

In order to supplement the self-triggering capabilities of the DTs and CSCs resistive plate chambers (RPC) are added to the barrel as well as to the endcap part in order to ensure an efficient identification of candidate tracks and correctly assign a bunch crossing. Six layers of RPC chambers are embedded in the barrel iron yoke, two are located in each of the first and second muon stations and one in each of the 2 last stations, see also Figure 2.16. In this way also low p_T muons which may stop inside the yoke can be reconstructed on the basis of the first 4 layers. In the endcap region, in the first phase of CMS, only 3 layers up to $|\eta| < 1.6$ are installed (before, in between and after the first two iron disks). This layout of the RPCs in the endcap system is also shown in Figure 2.14. The geometrical design of the chambers follows closely that of the DTs and CSCs, thus they are rectangular in the barrel and trapezoidal in the endcap.

The detection principle is the same in both RPC types, they are gaseous parallel-plate detectors
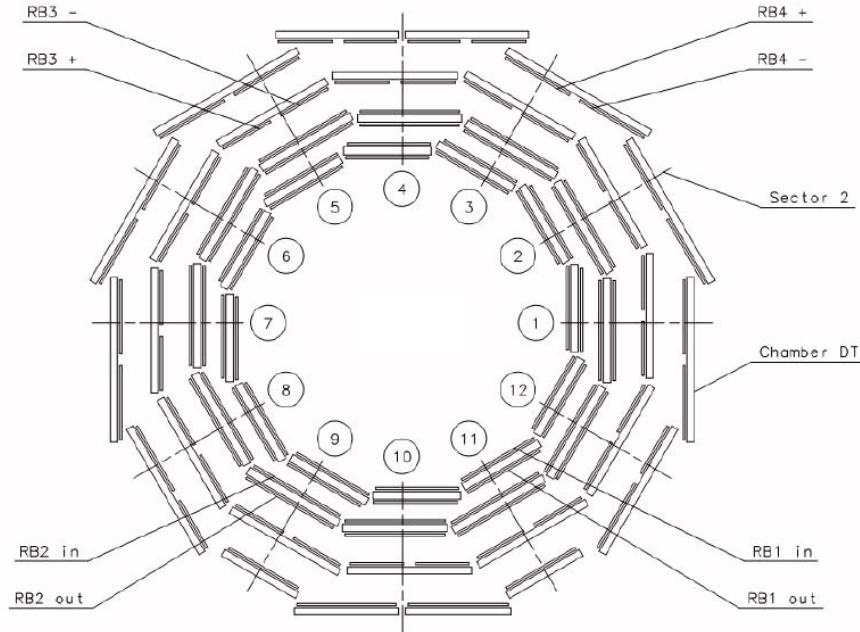
Figure 2.16: *Schematic layout of one barrel wheel [23]. RPCs are labeled with RB and are mounted on the DT chambers.*

which have a 2 mm double gap and read-out strips in between. Operated in avalanche mode the total signal induced is the sum of the two gaps, allowing thus to operate them at lower voltage (≈ 10 kV) and with higher efficiency compared to a single gap design. In the barrel the strips run along the beam line and provide thus information on the φ-angle, while in the endcap they run radially accordingly. In both cases the strip width increases to preserve projectivity and covers $5/16°$ in φ.

The triggering capabilities of the RPCs can be seen best mentioning their excellent time resolution of around 1 ns, enabling a reliable bunch crossing assignment. In addition to this the Pattern Comparator Trigger (PACT) logic compares strip signals of the RPC muon stations to predefined patterns in order to assign p_T and electric charge, after having established a certain number of coincident hits in the planes.

Similar to the inner tracker also in the muon spectrometer considerable alignment is needed in order to precisely measure momenta up to 1 TeV. The barrel and endcap muon detectors are continuously monitored using an opto-mechanical system with LEDs and laser beams. The local system of the barrel measures the relative position of all DT chambers, while in the endcap it is only 1/6 of the CSCs. In addition to this a link system relates the muon and central tracker systems, capable of a precision of a few hundred µm in $r - \varphi$. To achieve even better precision also in the muon system, alignment algorithms based on muon tracks will be utilized.

In addition to this precise survey and photogrammetry measurements provide an initial geometry (position and orientation of each muon chamber in the different yoke structures) which absorbs installation tolerances and static steel deformations.

The alignment system was tested successfully during the MTCC and has proven to operate under magnetic fields with the designed precision. Furthermore main features of the yoke displacements and deformations with and without magnetic field were studied.

2.2.6 Trigger

At the LHC proton-proton collisions will take place with a crossing frequency of 40 MHz, and up to 20 collisions can occur at the same bunch crossing at the design luminosity of 10^{34} cm^{-2}s^{-1}. Even with modern computer and data storage techniques it is impossible to store all this information, thus a drastic rate reduction has to be achieved by the trigger system selecting only events with interesting signatures. CMS uses a two-stage trigger system [23]: The Level-1 (L1) Trigger consists of custom-designed, largely programmable electronics and achieves a reduction up to 100 kHz. The events which pass this decision are processed by the High-Level Trigger (HLT), which is a software system implemented in a processor filter farm and defines the rate of events for offline analysis on the order of 100 Hz. This software trigger has access to the complete read-out data and can therefore perform complex calculations, therefore HLT algorithms will evolve with time and experience and can hardly be summarized at this stage.

The L1 stage on the other hand has a clear setup, combining local, regional and global components. These three components exist both for calorimeter triggers as well as for muon triggers and have a total L1 latency of 3.2 µs. A L1 stage using inner tracker information is only scheduled for the LHC upgrade SLHC.

In the calorimeter trigger the Local Triggers consist of trigger towers ($\Delta\varphi \times \Delta\eta = 0.087 \times 0.087$ for $|\eta| < 1.74$) which sum up the transverse energies of the ECAL crystals and the HCAL read-out towers. The Regional Trigger then combines various towers and determines regional candidate electrons/photons, transverse energy sums and isolation information. Finally the Global Calorimeter Trigger ranks all trigger objects (jets, total transverse and missing transverse energy, scalar transverse energy sum, e/γ candidates) across the entire detector using energy and quality criteria.

The muon triggers make use of all three sub-detectors DT, CSC and RPC. In the Local Trigger the DTs provide track-segments in φ and hit patterns in η, while the CSC deliver 3-dimensional track segments. Both also assign a bunch crossing to the event. The Regional Trigger Track Finders joins segments and forms track candidates assigning momentum, position, charge and quality. The RPCs with their excellent time resolution are particularly important for the bunch crossing assignment, but they also deliver track candidates. The Global Trigger then combines the information of all three sub-systems, providing improved momentum resolution, higher efficiency and better background suppression (ghost segments) compared to the standalone systems. Up to four muon candidates from the barrel as well as the endcap part can be received per bunch crossing with a start-up pseudo-rapidity coverage of $|\eta| \leq 2.1$ (trigger electronics upgrade planned for $|\eta| \leq 2.4$).

In addition to the CMS trigger system the events have to pass through the Data Acquisition (DAQ) chain and are then stored, transfered and manipulated by the CMS offline computing system. Finally the data and simulated samples are distributed worldwide via the Worldwide LHC Computing Grid (WLCG) project [25] and can be processed in various computer centers with hierarchical architecture (Tier). Both aspects, DAQ and computing, are considerably complex and important but are beyond the scope of this thesis, see e.g. [23] for details.

2.2.7 Luminosity measurement

Another important task in order to study the physics of proton-proton collisions is the determination of the luminosity [22]. As one can see in Equation 1.27, if the measured rate R and the cross section σ are known one can determine the luminosity l. In addition to this overall luminosity normalization which is important for physics analyses, one also needs to monitor the LHC beam performance by providing average luminosity measurements in real time.

For the latter point multiple techniques capable of providing suitable luminosity information in real time have been identified. One technique employs signals from the forward hadronic calorimeter (HF) while another, called the Pixel Luminosity Telescope (PLT), uses a set of purpose-built particle tracking telescopes based on single-crystal diamond pixel detectors (which will be included at a later stage of CMS). The design goal for these real time measurements is a statistical accuracy of 1% in 0.1 s. For the HF case a technique called "zero counting" is utilized. Here one simply counts the number of bunch crossings where no interaction occured, thus where the towers in the HF show only signals below a certain threshold. Since the number of interactions in a given bunch crossing is distributed according to Poisson statistics one can derive from this the mean number of interactions μ. Then given a known total cross section σ_{tot} and a known effective bunch crossing rate f_{BX}, determined by the beam parameters, one can determine the instantaneous luminosity l:

$$\mu = \frac{\sigma_{tot}\, l}{f_{BX}} \qquad (2.3)$$

Note that this technique needs the knowledge of the total cross section as an input and its determination will be discussed in the context of overall normalization below. Also note that in case of high luminosities and thus very large μ the zero counting can no longer be performed using all the towers of the HF ("zero starvation"), but one can still consider only single physical towers where the mean number of interactions is small.

The second technique using the PLT determines the number of particles traversing the telescope by forming coincidences of several detector planes. The amount of coincidences is a direct measure of the luminosity, thus the relative bunch-by-bunch luminosity can be monitored. In addition to this the PLT also provides a real-time measurement of the relative location of the interaction point centroid.

As mentioned earlier physics analyses also need the absolute luminosity. Here a precision around 5% or better including systematic uncertainties is aimed for. The TOTEM experiment [26] will measure the total pp cross section and study elastic and diffractive processes at the LHC. Elements of TOTEM will be situated in the far forward regions of CMS. Of particular relevance to the CMS luminosity determination will be a precision measurement of the total pp cross section using the so-called luminosity independent method. This approach uses the optical theorem, which holds that the total cross section can be related to the elastic cross section which can be measured using roman pots.

In addition to this the production rates for W's and Z's provide a "standard candle" with considerable statistics at the LHC. Here the challenge lies in the uncertainties associated with the theoretical estimates of the cross sections (parton density functions) and controlling the detector acceptance and efficiencies adequately.

Part I

Commissioning of the CMS Muon Detector

Chapter I.1

Commissioning with Cosmic Ray Muons

I.1.1 Introduction and motivation

Prior to the first cycling proton beams in September 2008 LHC and CMS have undergone a long period of design, construction and tests of all their components. Of course in parallel also sensitivity studies about the physics reach have been performed and are an important preparation for the first data. Still, especially in the last years before start-up of the LHC, tests and studies with a closer connection to the actual detector hardware are essential for the success of the whole experiment. During the assembly and the commissioning of different parts of the Compact Muon Solenoid (CMS) experiment, cosmic muons are the only available source of high energy particles that can be used for tests. As CMS has been pre-assembled on the surface, the rate of cosmic muons traversing the detector was high enough to enable high statistics tests during a relatively short period of time. Thus cosmic muons are an ideal source for extensive checks of detector hardware, readout electronics, data acquisition and even reconstruction algorithms.

The test of the CMS magnet in summer 2006 provided a unique opportunity to test also the CMS detector since the iron return yoke, and thus the entire detector, had to be closed to allow the operation of the magnet. A dedicated cosmic data taking period – the Magnet Test and Cosmic Challenge (MTCC) – has taken place between June and November 2006 using subsets of the CMS sub-detectors. Apart from integration issues, recording of cosmic muon tracks with and without magnetic field was a goal. This test accomplished to record more than 230 million events, to run the detector in stable mode for more than 2 months and to achieve the nominal $B = 4$ T magnetic field. The Cosmic Challenge can be regarded as a very promising dress rehearsal for the upcoming start up of the Large Hadron Collider (LHC).

Another important aspect of the Cosmic Challenge is the coincidence in time with the availability of the new reconstruction software package in CMS, CMSSW [27]. The Cosmic Challenge also offered an extensive test of this important new development. It gave the opportunity to gain experience handling real data and to investigate the performance of the reconstruction algorithms using data.

This study [28][29] completes the full chain from data taking to the final data analysis by comparing cosmic data recorded with the CMS detector to predictions obtained from simulation. In order to achieve this goal a new cosmic muon generator, CMSCGEN, has been developed and included in the CMSSW-packages. Two dedicated high statistics samples, with and without magnetic field, have been produced in order to validate the Monte Carlo prediction of the generator and the subsequent detector simulation, comparing them with data recorded during the MTCC. Results of these comparisons and general reconstruction studies are presented here. The focus is on the reconstruction of cosmic muons using the Drift Tube chambers of the CMS barrel muon system, see also Section 2.2.5. The gained experience is also of great interest for the future as cosmic muons are an important instrument of monitoring and aligning the CMS detector during LHC operation.

I.1.2 The MTCC setup and the CMS muon system

During the MTCC, parts of all CMS sub-detectors have been operated and read out simultaneously, ranging from inner tracker modules to some crystals of the Electromagnetic Calorimeter and sectors of the Hadronic Calorimeter, up to a considerable number of muon barrel and endcap detectors.

For the muon system, detector installation and commissioning was ongoing and a subset of these detectors sufficient to achieve the goals has been operated for the MTCC, including final cabling and final electronics. From the installed muon barrel detectors – Drift Tube chambers (DT) and Resistive Plate chambers (RPC) – three instrumented sectors were read out in the two positive barrel wheels, in wheel YB+2 the bottom sector 10 along with its adjacent sector 11 and in wheel YB+1 only sector 10 (see Fig. I.1.1). This accounts for 14 DT chambers of type MB1/MB2/MB3/MB4 and 21 RPC chambers (note that station 4 in sector 10 is split in two chambers), corresponding to about 10000 Drift Tube channels.

In the MTCC a subset of the final muon readout and trigger electronics [30] has been operated for triggering and data taking – that is a ReadOut Server (ROS) and a Trigger Sector Collector (TSC) per sector, three of each in total, installed on the towers near the wheels. Connected via 130 m long optical fibers, three Drift Tube Track Finders (DTTF) were used in the counting room, along with a Wedge Sorter (WS) and a Barrel Sorter (BS) in order to generate the Level 1 trigger information provided to the CMS trigger system. Apart from the DT system, also the forward muon system and the barrel RPC could provide a first level trigger, including multiple triggers for example of barrel DT and RPC detectors. The trigger input was processed and distributed to CMS by the Local Trigger Control (LTC) which assigned a trigger bit for the sub-detector generating the trigger. This information was part of the data stream and could be used to select events triggered by, for example, Drift Tube chambers.

The muon barrel trigger electronics are designed to select high p_T muons originating at the interaction point of pp-collisions. In the MTCC, where cosmic ray muons mainly from above are traversing the detector, the trigger logic had to be modified. The angular acceptance per chamber was maximal in order to allow for muons arriving from all directions and not only from the interaction point. The on-chamber trigger electronics required either two high quality track segments (4 hits) in both ϕ-SLs or the combination of one high quality and a low quality segment (3 hits) in both ϕ-SLs (HH and HL trigger configuration). Both ϕ track segments are correlated

by the TRAck COrrelator (TRACO) which is a key element of the trigger electronics mounted directly on the chamber. The DTTF, WS and BS did not select muons based on their momentum and quality, as it is designed to be at LHC operation, but simply issued a trigger when two Drift Tube chambers recorded a track segment.
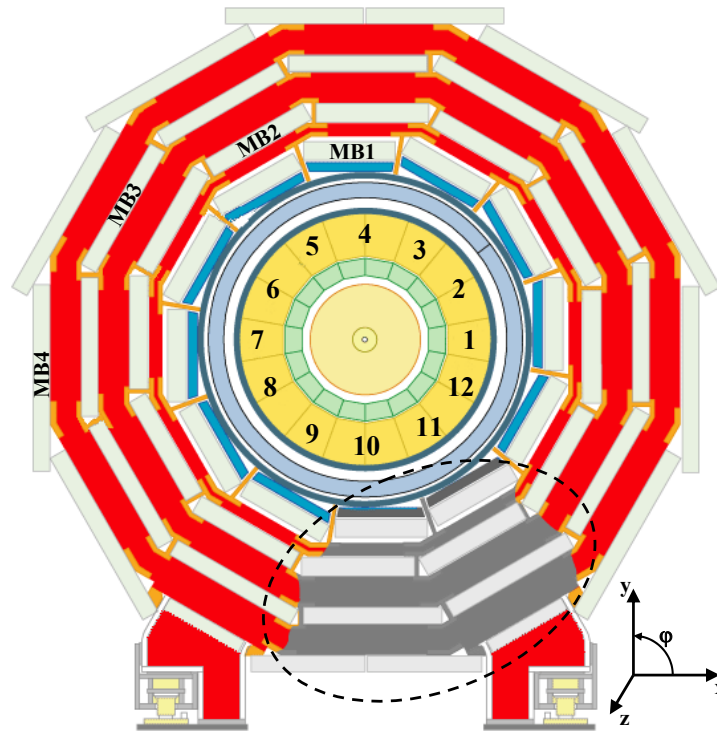

Figure I.1.1: *The MTCC exploits in the barrel region three sectors in the wheels YB+1 (sector 10) and YB+2 (sectors 10, 11) instrumented with Drift Tube chambers and Resistive Plate chambers. Here z in the global CMS coordinate system is along the beam line.*

Chapter I.2

Simulating Cosmic Ray Muons

I.2.1 CMSCGEN – A new cosmic Monte Carlo generator for CMS

This new cosmic generator [28][29] has been included in the CMSSW software package [27] since version 0_7_0, which enables the user to simulate cosmic muons for a detector on surface as well as in the underground cavern. It has replaced a previous generator [31], re-using a large part of the infrastructure from the old generator but improving some of the extrapolation features and its performance when compared to real cosmic data. In particular the dependence on the incident angle is now more accurate and the estimate of rates more realistic. The changes made are almost transparent to the user as only the physical input of the generator has changed. A first study using the old generator can be found in [32][33]. Note that after this study using data from the Cosmic Challenge has been completed, the generator has been upgraded even further, see [34] for details.

As an overview, cosmic muons arrive at the surface of the earth (assumed locally flat) with an energy dependence of $\mathrm{d}N/\mathrm{d}E \propto E_\mu^{-2.7}$, their angle off the vertical axis can be described by $\mathrm{d}N/\mathrm{d}\theta_y \propto \cos(\theta_y)\sin(\theta_y)$, and their angle in the horizontal plane ϕ_{xz} is uniformly distributed (see Fig. I.2.1 for definition of angles). In reality energy E and incident angle θ_y are correlated and the energy dependence becomes even steeper with increasing energy, see [35].

In order to model a realistic cosmic muon spectrum, dedicated parameterizations of energy dependence and incident angle have been used, also accounting for an energy dependence of the incident angle. These parameterizations have been adopted from L3CGEN [36], a cosmic muon generator developed for the L3+Cosmics experiment [37]. The original parameterizations have been obtained from the CORSIKA program version 5.20 [38], a program widely used for the simulation of extensive air showers, and have now been included in the CMSSW-package of CMSCGEN. For detailed information on the parameterizations see [34][36] and Appendix A. Note that the results of the L3+Cosmics experiment also represent an initial validation of the parameterizations implemented in the generator since they have been successfully used for several studies, e.g. [39] and [40]. In addition to this a quick check at the generator level can be performed, determining the spectral index of the energy distribution at 100 GeV for vertical cosmics: Using

CMSCGEN one can obtain a dependency $dN/dE \propto E^{-3.09\pm0.05}$ which is compatible with the value of the exponent 3.11 ± 0.03 from [41].

Cosmic muons at sea level have a charge ratio $N_{\mu+}/N_{\mu-}$ which is bigger than one. This excess of positive muons is caused by the interactions between the primary cosmic radiation arriving at the earth, esentially protons, and the nuclei of the atmosphere, protons and neutrons. Since positive mesons (pions and kaons) can be produced both by $p+p$ and $p+n$ collisions, but negative mesons only via $p+n$ there is an inherent charge assymmetry which is transfered to the decay products of the mesons, the muons. Cascades and multiple interactions in the atmosphere produce additional mesons, but now with approximately equal charge fractions. Thus the initial surplus is reduced, so that in the end a ratio not far from unity can be observed. As the correlation with energy and angle is not strong, an average of $N_{\mu+}/N_{\mu-} = 1.33$ is used in the generator (1.28 for the generator upgrade [34]).

Another property in which cosmic muons differ from muons produced during pp-collisions is the timing uncertainty. As they arrive at random times and are not bunched, the exact arrival time at different sub-detectors is not known. Still the CMS detector is read out in a clock cycle of 25 ns, so that the starting time of the cosmic muons is distributed uniformly between -12.5 ns and 12.5 ns. The generated starting time of the cosmic muons is chosen randomly in this interval in order to account for the timing jitter.

In the process of generation the cosmic muons are produced on a flat surface, assuming the properties of cosmic muons at an altitude of 470 m above sea level (corresponding to the surface of the L3 site, CMS surface is very similar with 509 m. In the later generator upgrade [34] the proper 509 m are used). They are then extrapolated as straight lines either to the underground cavern 90 m below the ground (including the simulation of energy loss in rock, wall and air) or to a detector setup on the surface without any energy loss during propagation. Only the cosmic muons hitting a cylinder with $R = 8$ m and $Z = \pm15$ m around CMS are selected and passed on to the rest of the simulation chain, using the default detector simulation inside CMSSW. The vertex of the cosmic muon is placed on this cylinder and its momentum vector contains the generated angles and energy of the cosmic muon.
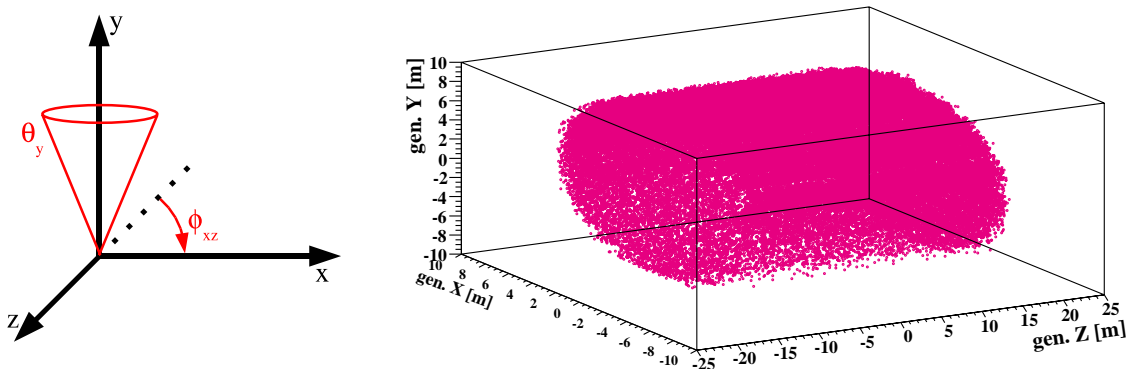
Figure I.2.1: *Left: Definition of generator angles which can be specified when configuring the cosmic muon generator. The coordinate system represents the global CMS one, thus the z-axis is parallel to the beam line.*
Right: The position of the generated vertices of 1 million cosmic muons is shown. The cosmic muons are on the surface of a cylinder around CMS which is the starting point before they are traced through the material and magnetic field by the detector simulation.

I.2.2 Kinematical range

Using the parameterizations obtained from CORSIKA it is now possible to simulate cosmic muons down to lower energies than before and with higher accuracy. The user can choose within the energy range of 2 GeV − 10000 GeV and incident angles up to $\theta_y = 88°$ (3 GeV − 3000 GeV and $\theta_y < 84°$ for the later generator upgrade [34]). Note that the angles above 75° represent only extrapolations of the parameterization, thus the results are expected to have slightly larger errors. In general an uncertainty of 5% between used parameterizations and original predictions from CORSIKA can be expected in the energy range between 10 GeV − 1000 GeV. Note that below 10 GeV geomagnetic effects and solar influences play a significant role (see [41]) and further increase the uncertainty. Below 2 GeV the energy and angular dependencies of cosmic muons change dramatically and new physical interaction processes set in. Thus the simulation of even smaller energies cannot be described by this generator and much more complicated parameterizations must be developed. Still, a 2 GeV cut-off is highly sufficient to model the cosmic muon spectrum expected during MTCC on surface or during LHC running in the underground cavern.

Fig. I.2.1 shows the definition of the angles which can be chosen to model a certain cosmic muon spectrum. As an example for the generated distributions, the output of 1 million cosmics generated can be seen in Fig. I.2.1 and Fig. I.2.2, using the full kinematical range and assuming a CMS detector on the surface. The first plot shows the position of the generated cosmic vertices which reside on the cylindrical surface around CMS. The other plot depicts the kinematical properties of the cosmic muons. One can nicely see the steep energy dependence and the suppression of small incident angles due to a vanishing solid angle for almost vertical cosmics $\theta_y = 0°$. The modulation in ϕ_{xz} can be explained with the cylindrical surface of the CMS detector, which is the starting point for the cosmic muons. The minima at 0° and 180° correspond to the endcaps facing the cosmics while the maxima refer to the transition between endcap and barrel view. Note that for vertical cosmics no modulation is expected.
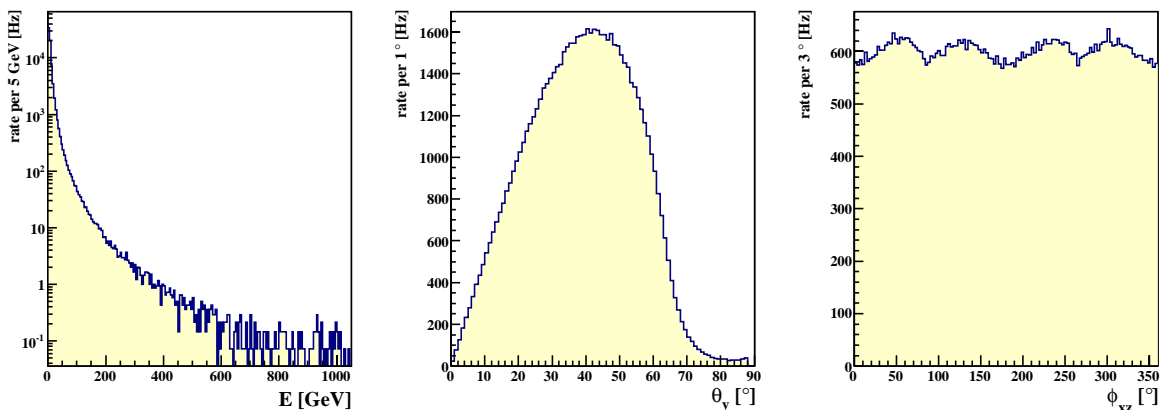

Figure I.2.2: *The kinematical properties of 1 million cosmic muons generated with CMSCGEN. The full kinematical range of the generator is used with an energy of $E = 2$ GeV − 10000 GeV (left), an incident angle of $\theta_y = 0° − 88°$ (middle) and full ϕ_{xz} coverage (right).*

I.2.3 Generator level preselection using straight line extrapolations

As explained above the cosmic muons are generated at a plane surface and are then extrapolated as straight lines to the CMS detector. This simple extrapolation mechanism offers the opportunity to include simple preselections at the generator level. First of all, as a generator option, it is possible to describe the energy loss in the rock above the CMS underground cavern. In this context a map of the material budget above the CMS detector is included, e.g. rock, concrete or the access shaft. The cosmic muons are tracked through this material and energy loss dE/dx is accounted for using first principles (known material properties and interactions with muons). Disabling the energy loss means simulating cosmic muons for a CMS detector on surface. In addition to this, in general two types of generator preselection can be chosen:

1. Change the dimension of the whole detector volume

2. Change the dimension of the target volume

Using option 1. means that the cylindrical surface around CMS on which the cosmic vertices are located can be shrunken. As a consequence cosmic muons which are generated on a plane are extrapolated to this newly defined cylinder. As the vertices are passed on to the detector simulation, this means that the cosmic muons may start within the CMS detector. Note that no energy loss in the material of the detector and no influence of the magnetic field outside this cylinder is accounted for as the extrapolation uses simple geometrical predictions. A good example for this feature is the *TrackerOnly* option which can be chosen in the generator. In this particular case the starting vertex of the cosmic muons within CMSSW is a cylinder around the tracker with $R = 1.1$ m and $Z = \pm 1.1$ m. This option is very useful to create a sample dedicated for comparisons with the cosmic assembly tests of the tracker-setup before installation in the CMS detector, like the TIB/TID Slice Test.

The second option controls some target volume inside CMS which can be defined within the generator. This means that the vertices of the cosmic muons are on the default cylinder around CMS, but still the muons are extrapolated even further according to their angles and only those hitting the target volume are saved and passed on to the time consuming detector simulation and digitization. Again, no detector material or magnetic field effects are accounted for during extrapolation. As a consequence a target volume too small can introduce a bias, especially due to bending caused by the magnetic field in the $x-y$ plane. An additional option (called *MTCCHalf*) is to cut the default target volume (i.e. full CMS detector) in two halves in z which ensures that all simulated cosmic muons penetrate the positive wheels and end-cap. This option has been included in order to increase the statistics of the simulated sample for the MTCC. Note that nevertheless during detector simulation all material and magnetic field in the negative half of CMS are included and no bias is introduced.

I.2.4 Normalization of the cosmic muon flux

Another important improvement with respect to the previous generator is the absolute normalization of the cosmic muon flux which allows to predict cosmic muon rates as seen by the CMS detector. As a first step the number of cosmic muons which are originally generated at the plane surface is set in relation to a known reference flux. Using the average of various measurements, see [41], the muon flux is normalized with respect to the amount of vertical cosmics at 100 GeV and per solid angle:

$$\Phi_{norm} = \frac{\mathrm{d}N}{\mathrm{d}A \; \mathrm{d}\Omega \; \mathrm{d}E \; \mathrm{d}t} = (2.59 \pm 0.18) \cdot 10^{-3} \; \; \mathrm{m}^{-2} \; \mathrm{sr}^{-1} \; \mathrm{GeV}^{-1} \; \mathrm{s}^{-1} \qquad . \qquad (\mathrm{I.2.1})$$

Of course this is only a suggested number to calibrate the expected rates, one can easily include different reference fluxes obtained from different measurements or simulations. The uncertainty of this flux directly translates into the systematic error of any rate estimate.

As a final step one has to correct for the selection efficiency of the generator as not all generated cosmic muons might actually hit the target cylinder around the CMS detector. The generator provides the rate estimate for the full target volume as specified during generation, including statistical and systematic errors. As an example the following predicted rate correspond to the MTCC setup of the Monte Carlo sample, i.e. detector at surface and cosmics have to hit the positive z-half of CMS:

$$\mathrm{rate}_{\mathrm{MTCC\text{-}setup}} = 15873 \pm 127 \; (stat.) \pm 1103 \; (syst.) \; \; \mathrm{Hz} \qquad . \qquad (\mathrm{I.2.2})$$

This number refers to the complete half of the detector. For practical purposes of course only small parts of the detector are read out and only rate predictions for these sub-parts can be compared to measured rates. In order to get rate estimates for a desired subpart of the detector, one simply has to compute

$$\mathrm{rate}_{\mathrm{sub\text{-}detector}} = \mathrm{N_{hit}} \cdot \frac{15873}{\mathrm{N_{events}}} \; \; \mathrm{Hz} \qquad , \qquad (\mathrm{I.2.3})$$

where $\mathrm{N_{hit}}$ is the number of simulated events with hits in a certain sub-detector and $\mathrm{N_{events}}$ is the number of all events in the Monte Carlo sample analyzed, thus all events hitting the MTCC-half of CMS.

Chapter I.3

Data and Monte Carlo Samples Used

Data samples

For this study the data taken during run 4320 and run 4406 were used. Both runs took place at the end of MTCC phase II in October 2006 and thus are part of the very final days of the Cosmic Challenge (after phase I the tracker modules have been replaced by instruments to map the magnetic field). After a month of data taking, at this late stage many experiences and improvements in operating the magnet and the muon system have been obtained. Thus the quality of the data can be expected to be better compared to earlier runs. Of course the results of this study represent only the final step and many runs from MTCC phase I and MTCC phase II have been looked at in order to develop the analysis properly. In the following the characteristics of both data runs are summarized.

<u>**Run 4320:**</u>

- magnetic field $B = 0$ T

- active detectors:
  14 DT chambers (MB1, MB2, MB3, MB4 in sector 10 YB+1 / YB+2;
  MB1, MB2, MB3, MB4 in sector 11 YB+2; MB4 in sector 14 YB+1 / YB+2)
  and 36 CSC chambers

- number of events: $1, 844, 000$

- number of DT triggers: $498, 000$

- 6 DT chambers used by trigger: MB2, MB3 in sector 10 YB+1 / YB+2;
  MB2, MB3 in sector 11 YB+2

<u>**Run 4406:**</u>

- magnetic field $B = 4$ T

- active detectors:
  14 DT chambers (MB1, MB2, MB3, MB4 in sector 10 YB+1 / YB+2;
  MB1, MB2, MB3, MB4 in sector 11 YB+2; MB4 in sector 14 YB+1 / YB+2)
  and 36 CSC chambers

- number of events: $1,826,000$

- number of DT triggers: $420,000$

- 6 DT chambers used by trigger: MB2, MB3 in sector 10 YB+1 / YB+2;
  MB2, MB3 in sector 11 YB+2

These two runs represent a good choice to compare real MTCC data with dedicated simulation samples as it is only possible to generate samples either with $B = 0$ T or with the nominal field of $B = 4$ T. No magnetic field map is available for other values of the magnetic field. The trigger conditions in both data runs are identical, such that the results with and without magnetic field can be compared. Only events where at least the DT trigger has fired are considered in this study, for both, data and simulated events. While for data it is inherent to the data taking, for simulated events the DT trigger conditions were modeled accordingly as described in Sec. I.5. To select these events in the data the LTC (Local Trigger Control) trigger-bit was used which defines if the DT system has fired. In total for each run 4×10^5 events (triggered either by DT or CSC or by both) have been processed in this study. By requiring that at least the DT trigger has fired this total number of events is reduced to a statistics of $108,253$ inclusive DT-triggered events for the B=off run and $91,952$ inclusive DT-triggered events for the B=on run.

Simulated samples

There are two dedicated Monte Carlo samples with large statistics available: $997,500$ events with magnetic field $B = 0$ T and $997,500$ events with $B = 4$ T.

These include generation of the cosmic muons, simulation using the GEANT [42] based detector simulation within CMSSW and finally digitization of the signal of various sub-detectors active during the Cosmic Challenge. The production was performed using CMSSW_0_9_1 and the official production tools ProdAgent, using the computing resources of the GRID Tier 2 center Aachen/DESY. The 140 GB files for each dataset have been registered in the global DBS/DLS database and are available using the GRID analysis tool CRAB [43].

These samples have some special features which restricts them to be used only when comparing with MTCC-data:

- In the generator the option MTCC-half is used (see Sec. I.2.1), meaning that all cosmic muons actually hit the CMS-half with $z > 0$ m (z-axis is along the beam line). Still on their way trough the detector they may also cross parts of the other half.

- A special MTCC-like geometry is used during simulation, where the pixel detector is not present and only a fraction of the silicon strip tracker modules are included, while sub-detectors such as HCAL and muon system are complete. Since only parts of the muon system are actually read out in the MTCC, the other muon chambers have to be skipped before the reconstruction step.

- For the MTCC the detector is situated in the surface hall, and hence, energy loss in the rock above the cavern is not simulated.

- During the MTCC some sub-detectors are operated in special modes optimized for the detection of cosmics. So also in the simulation these special running-modes have to be accounted for, which leads to changes in the digitization with respect to pp-collisions (e.g. inner-tracker).

- There is a lower energy-cutoff in the Monte Carlo samples in order to increase the statistics of cosmic muons reaching the bottom part of the detector, so

$$E_{generated} > 7 \text{ GeV} \qquad .$$

This value has been determined optimizing the amount of events likely to fire a DT trigger in the MTCC setup. Still the lower energy threshold has to be kept in mind when comparing the simulated predictions to data recorded during MTCC.

Note that only $\approx 1.5\%$ of the 1 million generated events finally satisfy the acceptance and trigger conditions of the MTCC-setup and thus can be compared to data.

Chapter I.4

Calibration and Synchronization

In the case of the barrel DT chambers, muon reconstruction is based on the measured electron drift-time. In that sense calibration of the chambers aims at synchronizing them such that the muon's arrival time at a given detector is common for all stations. This calibration of drift-times has to be performed in the case of pp collisions as well, so the planned strategies can be tested using cosmic muons. In the case of cosmics, there are some additional complications: Cosmic muons arrive at random times at the detector and their kinematic properties are less constrained than muons coming from a nominal interaction point. The necessary steps of synchronization for data and generated events will be discussed in the following. All of them are carried out using the official calibration tools [44] within the DTCalibration package of CMSSW.

I.4.1 Calibration of the data

Timing differences in the arrival time of DT signals appear at different levels with varying granularities. Firstly, signals from the individual wires arrive with a slight variation of ± 5 ns at the TDC, caused by small differences in the chamber electronics and the different cable lengths. Using test pulses a so-called "T0"-offset for each cell can be determined such that only the real drift-times and the signal propagation along the wire can cause timing differences. For the reconstruction, these corrections are read from a SQLite [45] database and the digitized drift-times are shifted accordingly.

While the first correction is intrinsic to the chamber itself and does not change after chamber installation, the next one is a mixture of different contributions. First this offset largely depends on the cable length and latencies between the chamber and the off-detector electronics. The latencies have been modified during the MTCC data-taking in order to synchronize the subsystems with respect to each other. Then the offset depends on the time when the trigger is fired and which bunch crossing is assigned to the event. In addition to this the cosmic muon has a time-of-flight from one chamber to another, depending on the place and angle where it crosses the MTCC setup. This place and angle varies from event to event but is only averaged in this calibration, see Sect. I.5 for event-based correction. These effects all together are reflected in a constant offset called "tTrig" which is determined for each Superlayer. This is done using a fit routine which identifies the starting point of the drift-time spectrum when all cells of a Super-

layer are added up (see Figure I.4.1 for an example of a drift-time spectrum). The derivation of a Gaussian is matched to the rising edge of the spectrum. Using mean and σ of the Gaussian the starting point of the drift-time spectrum is determined. Details of the calibration routines can be found in [44]. Again, a SQLite database is used to save these offsets and shift all timeboxes to zero.

To summarize, the applied calibrations to ensure a common starting point for the drift-time determination are:

$$t_{\text{drift}} = t_{\text{measured}} - T_0 \text{ (per cell)} - t_{\text{tTrig}} \text{ (per SL)} \qquad . \qquad \text{(I.4.1)}$$

Noisy cells are selected and discarded prior to the tTrig-fit since they interfere with the calibration routine. In the reconstruction noisy cells are included in order to avoid any bias.

I.4.2   Calibration of Monte Carlo events

As described in Sect. I.2.1, the timing uncertainty of 25 ns in the cosmic muon arrival is introduced during the generation of the Monte Carlo samples. In addition to this intrinsic uncertainty, also simulated cosmic muons have different timing offsets in each DT chamber. While real cosmic muons suffer from time-of-flight differences with respect to the moment the trigger fired and from hardware latencies, simulated cosmics obtain a certain time-of-flight caused by the distance between generated vertex and each traversed chamber. In addition to this, certain offsets per chamber are subtracted during default digitization, assuming the muon to originate from the nominal interaction point. This yields a situation comparable to the real data and the same calibration tool can be used in order to determine the "tTrig" from each Superlayer. Since data and Monte Carlo events are calibrated in the same way, effects caused by the used routines should be observable in data as well as Monte Carlo prediction. No "T0"-correction per cell is needed here since no electronic latencies between different wires are included in the simulation.

Figure I.4.1 shows the comparison of the drift-time spectrum for simulated cosmic muons and data with magnetic field, after the calibration has been performed. Here all drift cells within Superlayer $\phi 1$ of MB1 in sector 10, YB+2, are summed up, corresponding to the granularity of the applied calibration routine. Data are normalized to the simulated events in order to reflect the smaller statistics of the Monte Carlo sample. After proper synchronization the observed spectrum of drift-times ranges from almost 0 ns (for muons passing close to the wire) to 380 ns (or 486 Time to Digital Converter counts, one TDC count being the "hardware unit"). In order to suppress effects of noise and after-pulses, only cell hits contributing to reconstructed objects, in this case a 2D-segment built from hits of either ϕ- or θ-Superlayer (see Sect. I.5), are used. Measured data and simulation agree well and the general shape of the timebox is as expected. This fact is also important since at the location of this specific chamber a relatively large magnetic fringe field of up to ≈ 0.8 T is present. In certain areas of the chamber with such a magnetic field the drift times are increased by about 10 ns. This effect is also included in the digitization of the simulated cell hits. Still, at the level of drift-time spectra for a complete Superlayer such an effect is not visible and the detector simulation seems to describe the conditions within the chamber well.
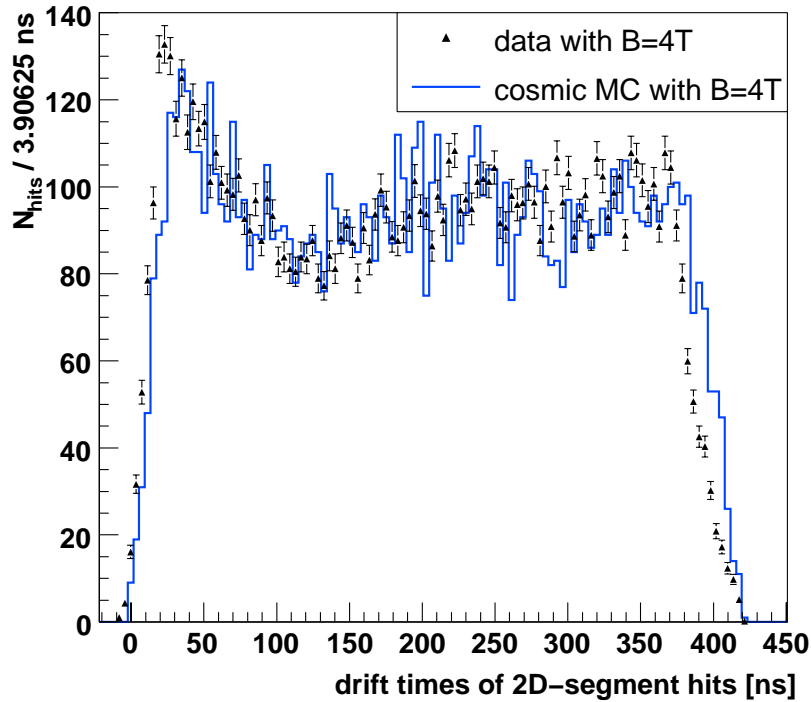
Figure I.4.1: *In order to verify the signal creation and the signal calibration in the cosmic muon simulation, a drift-time spectrum from simulation (continuous line) is compared to drift-times as measured with real cosmic ray muons from MTCC (triangles) for hits contributing to a 2D-segment. For this study data and simulated events with magnetic field are regarded after calibration has been performed, here for Superlayer $\phi 1$ of the horizontal MB1 of sector 10 in YB+2, corresponding to the granularity of the calibration routine (data normalized to MC).*

Note that the calibration routines described above still cause some remaining timing uncertainties. It is clear that the inherent cosmic jitter of 25 ns and the propagation time of the signal along the wire remain. In addition to this the constant tTrig-offset for each Superlayer is only the result of an average of many cells and many events. The obtained time-of-flight corrections are correlated with the angular distribution of the incoming cosmics and with the geometrical acceptance of the MTCC setup. The exact sum of all offsets is specific for each event and thus can only be extracted using further information obtained during reconstruction. Such an event based correction will be introduced in Sect. I.5. In this way almost the optimal resolution of the drift cells can be obtained, also when using non-bunched cosmic data.

Chapter I.5

Reconstruction of Cosmic Ray Muons

As in the case of calibration, also during reconstruction of cosmic muons some changes and optimizations have to be performed with respect to the default code designed to measure muons coming from a central vertex during bunched pp-collisions. While the timing uncertainty of the cosmic muons only worsens the resolution of the Drift Tube chambers, the large incident angles of the cosmic muons and the fact that they penetrate the detector from all directions causes event topologies which the default reconstruction code is not designed to handle. Therefore all vertex constraints normally used to enhance the measurement have been disabled and a dedicated cosmic track-reconstruction algorithm has been used. Also the trigger-setting has been arranged such that events with larger angles can be recorded and the acceptance is only limited by the angular reach of the trigger-electronics. One cannot expect the efficiency of the trigger and the reconstruction to be equal for all event topologies. But still these modifications enable to measure almost the natural properties of the incoming cosmic ray muons, with only little bias and geometrical constraints.

In the case of data from the Cosmic Challenge, CMSSW_1_1_0 has been used to process the recorded events, while for the simulated events a previous version, 1_0_6, has been chosen. The reason for this difference is that various slight changes in the geometry have been performed between both versions. So for real data the corrected geometry is used while for simulation the one compatible with the version the sample has been created with has to be applied. Despite the differences in geometry which do not affect the physical output of this study the rest of the reconstruction chain is identical so that data and Monte Carlo prediction can be compared.

In the following the different reconstruction steps including the modifications used with respect to muons from LHC running are discussed, ranging from single cell hit reconstruction to complex track-fits. In general all 14 DT chambers active during MTCC have been used for local reconstruction as well as track reconstruction.

I.5.1   Local reconstruction at chamber level

Drift-times are the measured quantities used in the barrel muon system. Given the intrinsic timing uncertainties of cosmic muons a linear time-to-drift relation is used in this study. This simplification assumes a perfectly homogeneous electrical field inside the cell volume which is an approximation. Influences of the magnetic field inside the muon chambers and non-perpendicular

angles of the muons are neglected in this approach. Still these effects are of second order and especially for perpendicular incidence the differences between linear time-to-drift and a more detailed parametrization of the cell behavior are small [46]. Using the linear approximation it is possible to reconstruct the one-dimensional position of the traversing muon:

$$x_{\text{hit}} = v_{\text{drift}} \cdot t_{\text{drift}} \qquad , \qquad \qquad \text{(I.5.1)}$$

where a drift velocity of $v_{\text{drift}} = 54.3$ µm/ns is applied.

In the case of the simulated events, an additional step is performed during cell hit reconstruction. Since the MTCC-geometry used contains more muon chambers than were actually read-out with the real setup, only cell hits belonging to the actual MTCC-setup are selected and passed on to the rest of the reconstruction chain.

Still, the left-right ambiguity with respect to the wire cannot be solved at the stage of single hit reconstruction. Different cells in different layers have to be combined to form 2D-segments within one Superlayer. The name implies that these segments contain a two-dimensional information, bearing in mind that the wires of a ϕ-Superlayer run perpendicular with respect to a θ-Superlayer. The segment reconstruction performs a linear fit combining the cell hits of the 4 contributing layers. Using pattern recognition and χ^2-minimization the best possible combination of cells can be determined and this angular information solves the left-right ambiguity.
The default CMSSW segment reconstruction includes vertex constraints for ϕ- as well as θ-Superlayers. These have been disabled so that all possible angles of the segments can be measured, given that the trigger has accepted the event. As a consequence in some topologies the left-right ambiguity can no longer be solved uniquely simply by choosing the segment candidate with the smallest χ^2. This can lead to segments with unphysical angles even larger than the acceptance of the trigger. In addition to these misreconstructed segments there is also a candidate which corresponds to the true path of the cosmic muon. In this case both candidates are kept and the decision which one is correct is passed over to the following track-reconstruction step. Using the information of several DT chambers, wrongly measured segments can easily be discarded.

As introduced in Sect. 2.2.5 a Drift Tube chamber consists of 2 ϕ-Superlayers which are staggered around a single θ-Superlayer. Since the orientation of the wires in each Superlayer type is perpendicular to each other the local x-z coordinates can be measured in the ϕ-SL (φ-plane) while the local y-z coordinates are provided by the θ-SL (ϑ-plane), see left illustration in Figure I.5.1. The angles φ and ϑ are in this case the normal spherical coordinates of the global CMS coordinate system (z along the beam line). Proper spherical angles are noted with φ while for the subdetector names ϕ is used. The sectors of the muon system are orientated cylindrically around the global z-axis. Thus for muons coming from the nominal interaction point the angles measured by an individual Superlayer, which are projections in the local reference frame only, almost correspond to the global φ- and ϑ-angles.

The final step in local chamber reconstruction is the combination of the 2-dimensional information of each Superlayer type in order to obtain 3-dimensional chamber segments. For the θ-Superlayer one can simply copy the segment information obtained before. Any further updates of the contributing hits will be discussed below in the event based correction. For the ϕ-view the 2 Superlayers are combined to form a "Superϕ-segment", using all 8 contributing layers. Here the

segment is refitted using the original hits of both 2-dimensional segments, thus avoiding any bias from the previous reconstruction steps. Finally θ- and ϕ-views can be combined to form a 3-dimensional chamber segment. Of course different segment candidates can be present and different combinations of them are possible. In order to limit the combinatorics and the amount of chamber segments per event, a cleaning algorithm based on the number of hits and the χ^2 of the fit is applied in order to reject poor segments. Further details on segment reconstruction can be found in [22].

Figure I.5.1 shows an example of chamber reconstruction in a data event where the nominal magnetic field of 4 T is present. One can see the r-φ view of the CMS-detector. The red lines correspond to the Superϕ-segments of the DT chambers. Even without complete track-reconstruction the combination of the 4 chamber segments shows a cosmic muon going through the bottom sector 10 of the detector, being bent by the magnetic field according to its momentum and charge. The green horizontal bars correspond to segments found in the orthogonal r-z view.
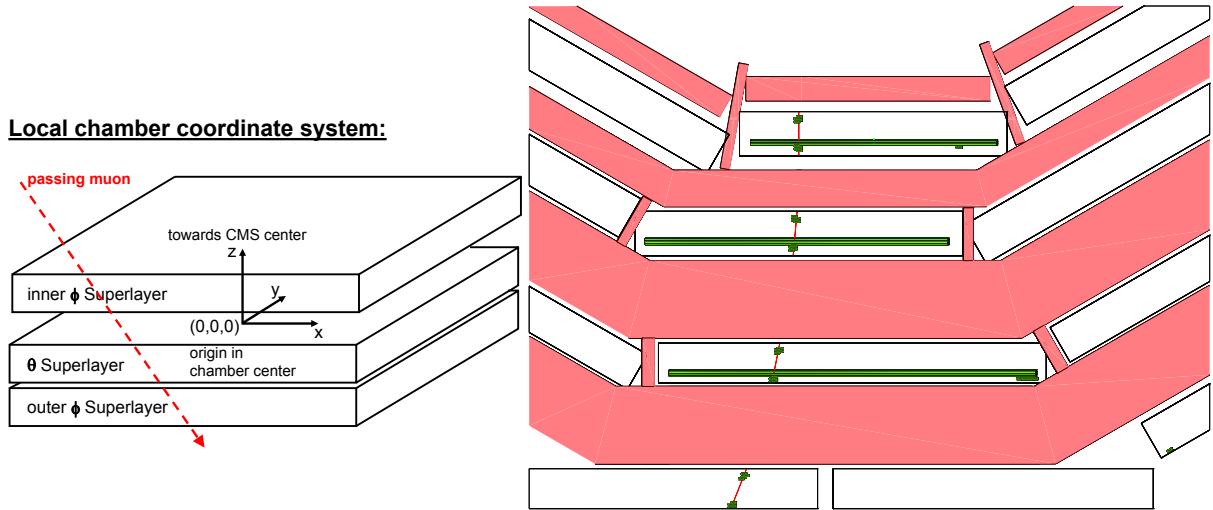
Figure I.5.1: *Left: Definition of the local DT chamber reference frame with its origin in the center of the chamber (local y is along global z).*
Right: Iguana [47] event display from MTCC data with magnetic field of $B = 4$ T , here the r-φ view of the CMS detector is shown. The red lines indicate reconstructed chamber segments, the greenish rectangles correspond to individual cells with a hit. The θ-Superlayers do not provide information for this projection, thus their measurements are displayed as green bars only (corresponding to the θ-wires). The bending due to the magnetic field can easily be seen.

I.5.2   Event based correction

As discussed in Sect. I.4 the Drift Tube chambers have been synchronized so far assuming a constant time offset. This "tTrig" accounts only for an average latency which can be subtracted from all measured drift-times. Still the uncertainty caused by the random arrival time of the cosmics can only be corrected if the additional information of reconstructed segments is available.

The default segment reconstruction performs a simple linear fit, thus determining a space point and the 2-dimensional direction of the cosmic muon track for each Superlayer. In order to increase the precision of the measurement for cosmic muons a dedicated refitting-algorithm has been developed within CMSSW, for detailed information see [48].

The basic idea is to modify the linear time-to-drift relation, assuming that for each single event all hits contributing to the segment have the same latency-offset and see the same drift-velocity:

$$x_{\mathrm{hit}} = v_{\mathrm{drift}} \cdot t_{\mathrm{drift}} \tag{I.5.2}$$
$$= (v_{0\ \mathrm{drift}} + \delta v_{\mathrm{drift}}) \cdot (t_{0\ \mathrm{drift}} + \delta t_{\mathrm{drift}}) \qquad . \tag{I.5.3}$$

The higher order terms $\delta v_{\mathrm{drift}} \cdot \delta t_{\mathrm{drift}}$ can be neglected in the following. In the expression $t_{0\ \mathrm{drift}}$, corresponding to Equation I.4.1, is already corrected for the constant tTrig-offset and is different for each hit. The value $\delta t_{\mathrm{drift}}$ is different for each event (or segment if more than one are present), depending on the distance of the track from the readout electronics since the signal propagation along the wire can take up to 7 ns. More important, the arrival time of the cosmic muon with respect to the 40 MHz clock adds to this correction. On the other hand $\delta v_{\mathrm{drift}}$ fine-tunes the used drift-velocity for each segment since this can be affected by the quality of the gas inside each chamber and by the intrinsic influence of the magnetic fringe-field inside the chambers, see [49]. If the angles of the incoming cosmic muons are not too large the assumption that the correction in drift-times and drift-velocity are the same for all hits contributing to one segment is well satisfied.

Using the event based correction a 4 parameter fit of the segment is performed if at least 5 hits are available (in the case of combination of the two ϕ-Superlayers; parameters are x-position, angle, $\delta v_{\mathrm{drift}}$ and $\delta t_{\mathrm{drift}}$). By trying to minimize the residuals, i.e. distance between measured hit-position and hit-position predicted by the segment-algorithm, a system of 4 linear equations is solved. This determines the optimal choice of the x-position and angle of the segment, of $\delta v_{\mathrm{drift}}$ and of $\delta t_{\mathrm{drift}}$. If there are less than 5 hits, e.g. for a θ-Superlayer, only a 3 parameter fit is applied, using an unchanged $v_{0\ \mathrm{drift}}$. In both cases the individual cell hits are updated according to the changes in drift-time and drift-velocity and the segment is refitted again, resulting in the optimal measurement of segment-direction and segment-position.

I.5.3   Modeling the DT-trigger setup for the Monte Carlo events

The Monte Carlo sample generated to be compared to data from the Cosmic Challenge includes the full detector simulation, so the reconstruction should perform similar in data and in Monte Carlo events. However, a trigger simulation corresponding to the special MTCC trigger setup could not be included, since it was not available at the time. Efforts have been made to account "manually" for the trigger efficiency.

First the geometrical acceptance of the Drift Tube chamber is accounted for. As mentioned in Sect. I.3 the two chambers MB2 and MB3 of all three active sectors were included in the trigger conditions of both runs. The Drift Tube trigger during MTCC requires a coincidence of two chambers, while each chamber has to detect a correlated signal of both ϕ-Superlayers. This yields the following "pseudo-trigger" condition which have been applied both in the data and in

the Monte Carlo events (to get comparable samples):

<u>pseudo-trigger:</u>   events with ≥ 7 hits combining both ϕ-Superlayers in at least 2 chambers from all MB2/MB3 of sector 10 YB+1/YB+2, sector 11 YB+2

It is clear that an event passing this criterion is within the geometrical acceptance of the chamber since it created multiple cell hits. Still, this does not mean that the event was indeed triggered by the real MTCC setup since the on-chamber trigger electronics has a certain angular acceptance and trigger efficiency which depends on the projection of the global φ-angle in the local x-z plane, called φ_{SL}. For large angles one expects the trigger-efficiency to drop since the hardware is designed to record muons originating from the nominal vertex (which have small angles). By computing the ratio of $N_{\mathrm{data}}/N_{\mathrm{MC}}$ as a function of φ_{SL} one can use the simulated events to extract the shape of the trigger-curve from the MTCC data (at this stage only shape, not an absolute scale). Here data and Monte Carlo prediction are scaled to each other only in the central region at small angles ($|\varphi_{\mathrm{SL}}| < 10°$) since an efficiency close to 100% is expected here. Afterwards the obtained shape can be used to discard a fraction of the generated events according to this trigger-efficiency.

The result of this trigger study can be seen in the left plot of Fig. I.5.2, here for φ_{SL} of MB3 in sector 10 YB+1. In order to avoid the effect of bending the samples without magnetic field are used. For this horizontal chamber a vanishing angle corresponds to a vertically incoming cosmic muon. One can nicely see a plateau and then a rapid drop in efficiency at larger angles. Efficiencies larger than 100% may appear if a bin contains more data than simulated events since both are normalized with respect to each other. This can also be compared to the expected trigger-efficiency obtained from trigger emulation studies, see dotted line "correlated" at the right plot in Fig. I.5.2.

For the simulated cosmic events the shape is approximated by straight lines, with the minima at $\pm 50°$ and the plateau between $\pm 30°$. The simulated events are corrected for this efficiency shape, discarding events accordingly. Without accounting for the trigger-efficiency the tails at large φ-angles would be more prominent in the Monte Carlo prediction than in the data, also when looking at track-reconstruction.

I.5.4   Cosmic track-reconstruction combining chambers

A dedicated cosmic track-reconstruction algorithm has been developed within CMSSW, a package called CosmicMuonSeed and CosmicMuonProducer (see [50] for detailed information). The motivation has been to provide a tool optimized for the cosmic data taking periods during Cosmic Challenge and during detector alignment in the early runs of LHC pp-collisions. In both cases cosmic muons are an important particle source to commission the detector and use tracks to align its components. Again, the focus is on the track-reconstruction using DT chambers as an example. Reconstructing "global" muon tracks is beyond the scope of this study, even though incorporating information from tracker or end-cap muon chambers is also possible.
This new algorithm assumes muons to come from "outside" the detector and it excludes the implicit vertex constraints which are used in many steps of the standard tracking software. As
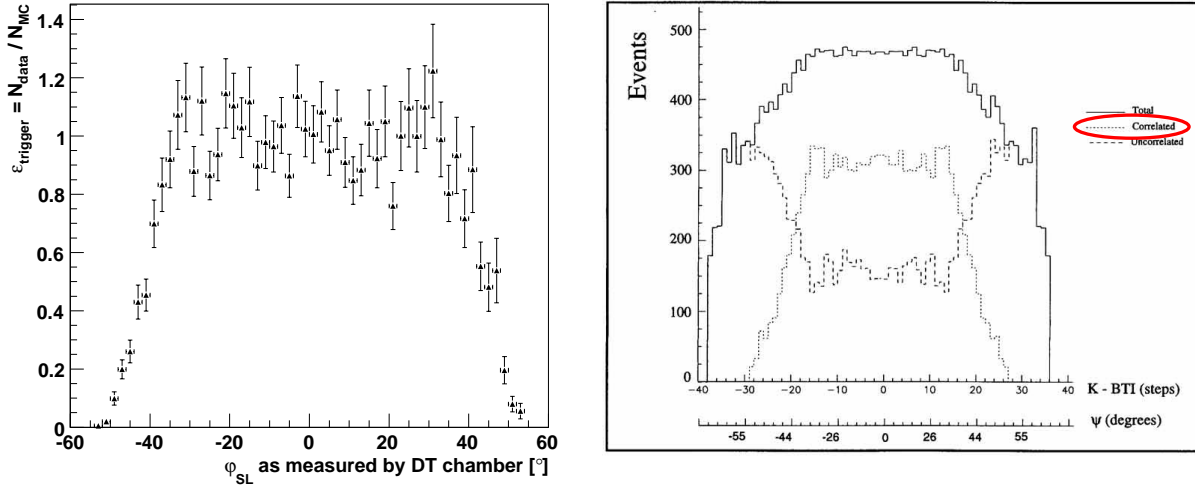
Figure I.5.2: *Left: Trigger-efficiency-shape from MTCC data and MC in comparison, here for φ_{SL} of MB3 in sector 10 YB+1 using data and Monte Carlo samples without magnetic field. Right: Simulated response of the track correlation algorithm (TRACO) as a function of φ_{SL} angle, muons were generated flat in angle (from Muon TDR [51]). Correlated means that the segments of both ϕ-Superlayers match.*

a starting point the 3-dimensional chamber segments of the local reconstruction are used for seed-finding. Since all vertex constraints have been already disabled there and since the actual direction of the segment is used, it is possible to efficiently find muon tracks passing the MTCC setup from various directions.

The next step consists of the actual track-fit. Starting from the seed and using a Kalman-filter technique [52] the segments of the different Drift Tube chambers are combined. The general procedure is to use the measured segment-position and -direction and extrapolate to the next muon station, assuming a certain momentum estimate. This results in a predicted state in which the influences of energy loss in the material, magnetic field and multiple scattering are included. This predicted state can then be combined with the measured values, forming a weighted average of both measurement and prediction.

In this study a forward-fitter is used combining the 3-dimensional chambers segments, starting from inside-out (MB1 to MB4). Afterwards a backward-fitter is applied with a finer granularity, this time using the measurements of individual cell hits. The propagator used so far in the algorithm is the Stepping-Helix-Propagator which uses a simple model of the detector geometry/magnetic field and a simple accounting for energy loss and multiple scattering. In the end it is possible to access the track parameters, i.e. position/direction/energy, at the innermost and outermost measurement, e.g. MB1 and MB3. The common tracking angles used are defined in Fig. I.5.3. A schematic drawing of the track-fitting procedure without the presence of a magnetic field is also shown.

Global CMS coordinate system:

track-fit (CosmicReco-muon)

φ_{track}

ϑ_{track}

$\vartheta_{\text{track}} = 0°$ means
track along the beamline (z)

MB1 — innermost measurement

θ-Superlayer
MB2 — Φ-Superlayer

MB3 — outermost measurement

2D-segment in global y-z plane
combined 2D-segment in global x-y plane
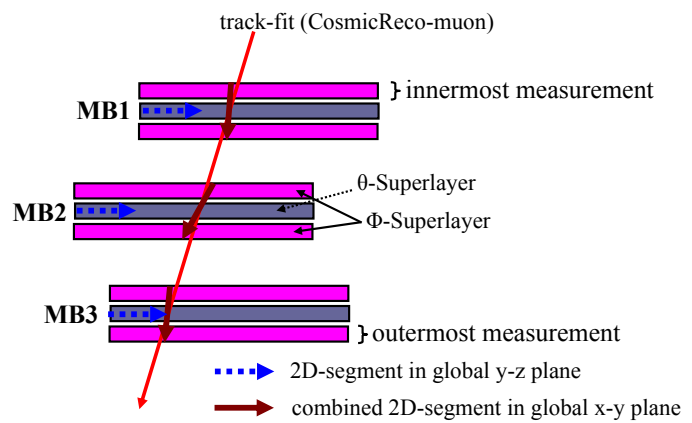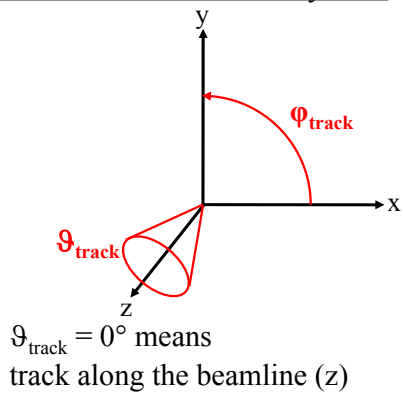
Figure I.5.3: *Left: Definition of the angles used during track-reconstruction within the global CMS coordinate system.*
Right: Illustration of the combination of several DT chamber measurements resulting in a track-fit. In this projection only the ϕ-Superlayers contribute to the measurement. Since no magnetic field is present the final track is a straight line and the track-parameters can be accessed at the innermost and outermost point.

Chapter I.6

Results and Summary

In this section several results will be presented which illustrate this first data-Monte Carlo comparison using the barrel muon system of CMS. The output of this study aims at different aspects. First of all the comparison with the real world, i.e. data coming from the detector, serves as an important validation of the cosmic muon generator. If general properties such as angular distributions or energy of the cosmic muons measured by the detector match in data and in simulation, then this means that the assumed parameterizations which are the input to CMSCGEN can describe nature to a satisfying degree. Even though for many results of the L3+Cosmics experiment these parameterizations have already been used (e.g. [39], [40]), it is important to study their correctness also with the final code used in CMSSW.

Then, of course, the intrinsic cosmic distributions are not directly measured by the detector. The observed distributions are strongly influenced by the geometrical acceptance of the MTCC setup, the efficiency of the hardware and by the physical processes which take place inside the muon chambers. Thus the next important check when comparing to measured data is the correctness of the detector simulation and the modeling of the different sub-detector parts. One has to bear in mind that in the beginning of LHC-running CMS will undergo an extensive period of detector calibration and understanding its performance, which will result in a more realistic tuning of the simulation. Even though the results presented here use cosmic muons only, still they give first hints of how well the data can be described by the simulation and they give cross-checks to spot bugs in the software from early on. Another advantage of using simulated events is that one can access the generated properties of the particle and thus can compare in detail the differences between reconstructed and expected values. So to compare distributions in data and Monte Carlo is one thing, but in order to conclude that the reconstruction software is working properly and to judge its performance, accessing the generated information of particles is essential.

Finally simply the fact that an entire analysis chain has been carried out is a very important achievement. Numerous steps have to be done, starting from the operation of the detector and the data-taking, transferring them to the computing sites, producing a Monte Carlo sample with comparable conditions and analyzing both data and Monte Carlo events with the official reconstruction software. Many problems have to be solved along the way and many lessons can be learned. Of course the plots which will be shown do not reflect the full picture, they are the result of the work of countless people at different levels of the experiment and they give proof of the fact that CMS is ready for physics analyses with LHC-data.

Normalization of plots with data-MC comparisons:

In all the following results the data are normalized to the simulated events in order to reflect the smaller statistics available for the MC sample. In this way the reader can easily determine the statistical error of the MC estimate by simply taking the square root. The statistical errors of the measurement are assigned to the data points and scaled according to the proper ratio N_{data}/N_{MC}. This ratio is also stated in the text for most plots.

I.6.1   Results from local reconstruction

As introduced in Sect. I.5 local chamber reconstruction combines the different cell measurements of all layers of a single Drift Tube chamber. Since the θ-Superlayer and the two ϕ-Superlayers can be combined to provide a 3-dimensional chamber segment, even at this level the angular properties of data and simulated events can be tested. For all the following results events have been selected where only a single chamber segment has been found per chamber. This reduces the effect of multiple segment candidates due to noise hits and to different combinatorial possibilities given the lack of a vertex constraint. In addition to this the segment is required to contain φ- as well as ϑ-information.

A very characteristic quantity is the incident angle θ_y of the cosmic muons in the global CMS coordinate system. The definition of the angle has been presented in Fig. I.2.1 and the generated angle has been already shown in Fig. I.2.2. This now can be compared to the reconstructed incident angle as measured by the horizontal MB3 in sector 10 of wheel YB+2, see Fig. I.6.1. This reconstructed angle is now defined in the local chamber coordinate system, but since the chamber in question belongs to a horizontal sector it roughly corresponds also to the global θ_y. The plot shows the data with triangles in comparison with the prediction of the simulated sample as the solid line. The distribution of the incident angle of the cosmic muons with and without magnetic field can be seen in the right/left plot respectively. The data are normalized to the simulated events in order to reflect the smaller statistics available ($N_{data}/N_{MC} = 5.9$ for B_{off}, $N_{data}/N_{MC} = 7.0$ for B_{on}).

First of all these two plots illustrate that it is possible to reconstruct the direction of cosmic muons correctly with and without magnetic field. Secondly the agreement between data and MC prediction is very good, the general shape of the distribution is reproduced nicely and looks very similar to the generator output, see Fig. I.2.2. This means the assumed parameterizations of the generator CMSCGEN are correct and that the detector simulation works well. Compared to the generator information the reconstructed angle in data as well as Monte Carlo drops much sharper above $\approx 30°$. This reflects the limited acceptance of the chamber geometry and the trigger-setup. Since DT chambers are designed to measure muons coming from the nominal interaction point they are not optimized for large angles. Without modeling of the trigger as described in Sect. I.5 the agreement between data and Monte Carlo prediction at large angles would be much worse. Still the slight excess of simulated events around $50°$ could be due to the imperfect description of the actual trigger. Another important aspect is the comparison with and without magnetic field. One can see that the general shape of the distribution is only slightly affected by the presence of the 4 T magnetic field, the right plot appears only a bit sharper.

More important is the reduction (20%) of reconstructed muons in the presence of the magnetic field, the reason being that cosmic muons are deflected on their way through the detector so that they no longer reach the MB3 of the bottom sector.
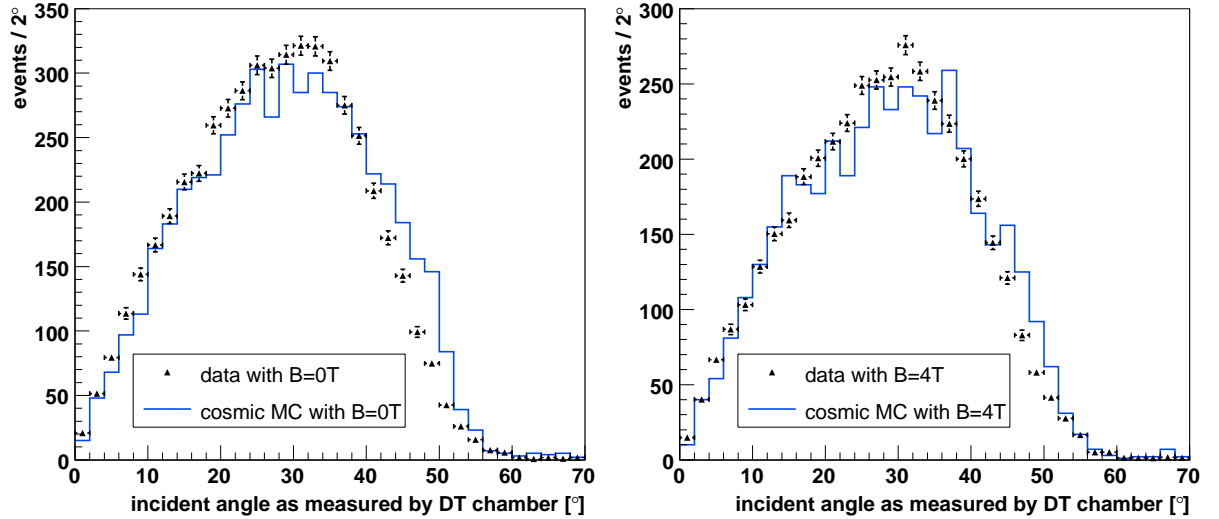

Figure I.6.1: *Incident angle of the cosmic muons as measured by local chamber reconstruction, here for horizontal MB3 in sector 10, YB+2. Data and MC prediction can be compared, without magnetic field (left) and with magnetic field (right). In both cases the data are normalized to the simulated events (continuous line).*

Correlated with the angular dependencies of the incoming cosmic muons is also the number of muons crossing two wheels or crossing two sectors. These specific events are important for alignment studies since they contain important information to determine the relative position of sectors and wheels with respect to each other. In general each of the 3 sectors read-out during the Cosmic Challenge sees $\approx 1/3$ of the triggered events, in the data as well as in the Monte Carlo sample. So the sectors are almost equally illuminated. From all events with magnetic field fulfilling the trigger conditions only 7% in the data and 11% in the Monte Carlo prediction have 2-dimensional segments in both wheels. Only 8% of the data and 9% of the Monte Carlo events cross the two sectors 10 and 11, not distinguishing between wheels. Only this fraction of the data can be used to contribute to the global alignment of wheels and sectors.

In order to judge the correctness of the measurement and the performance of the segment-reconstruction, residuals can be studied. A residual is defined as the distance between the predicted hit position when extrapolating the segment direction into a cell and the original hit position. This quantity is very sensitive to the intrinsic cell resolution and strongly depends on the correctness of the calibration of data and Monte Carlo events, see Sect. I.4. Figure I.6.2 shows the residuals of a specific layer for a single chamber, layer 2 of Superlayer $\phi1$ in chamber MB2 Sector 10 YB+1. In this case the cell for which the residual is measured is included in the segment fit. Data and simulated events with and without magnetic field are shown after the following selection cuts for a segment have been applied: The directional information is taken from a "Superϕ-segment" which is built using 8 hits. In addition to this an angular cut

of $|\varphi_{SL}| < 10°$ is applied. This ensures that the incoming cosmic muons have small angles since it is known that the linear time-to-drift relation is comparable to the more sophisticated parameterization only in this regime [46].
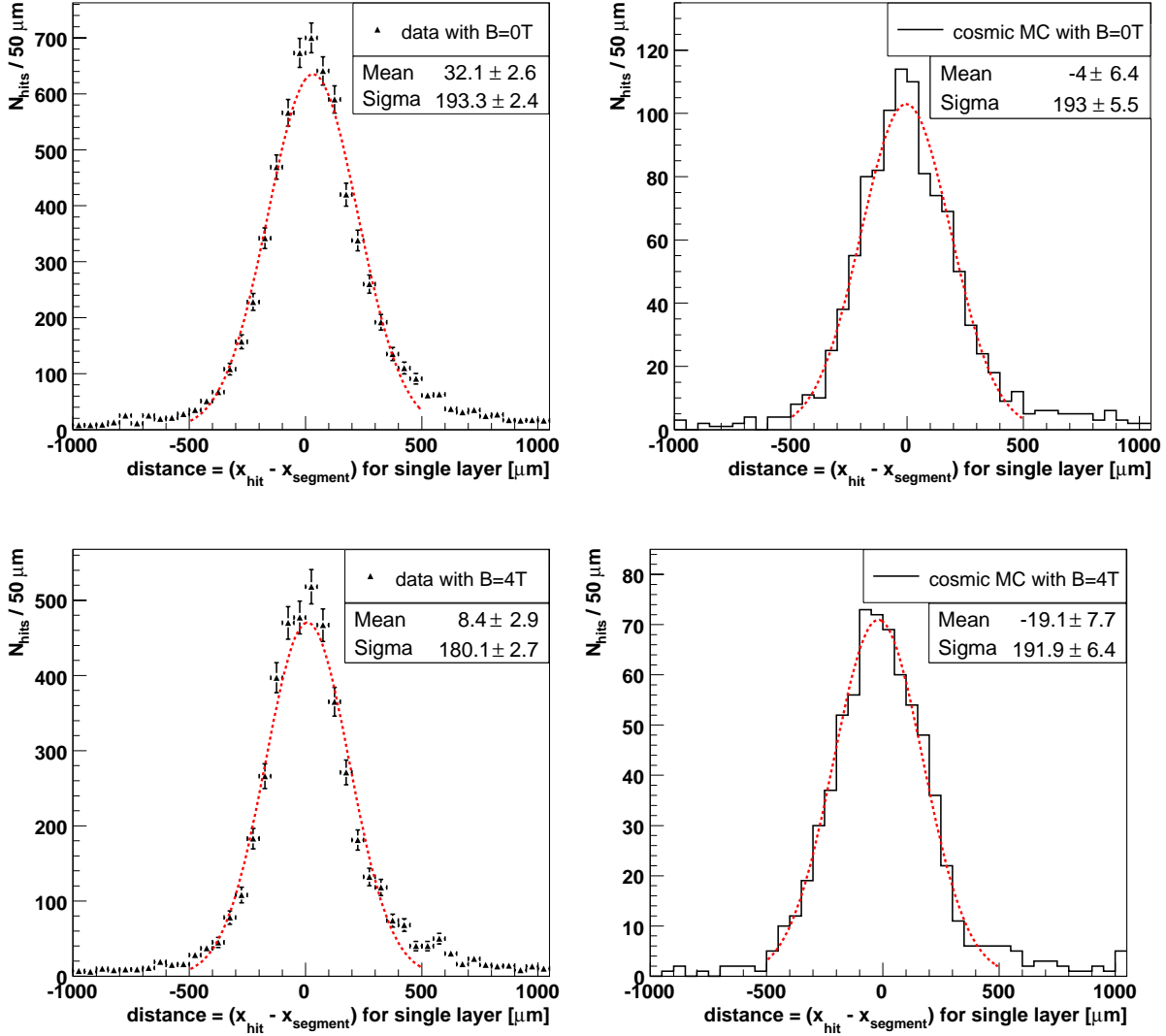
Figure I.6.2: *The residuals for a single layer (layer 2, SL ϕ1, MB2, Sec 10, YB+1) are shown including a Gaussian fit. The left plot corresponds to cosmic data (triangles) and the right plots to the simulated sample (continuous line); the top row displays data and MC without magnetic field and the bottom row the case with magnetic field. Note that only segments which origin from the combination of the two ϕ-SLs with 8 hits and which have $|\varphi_{SL}| < 10°$ are used.*

All four plots are compatible with a Gaussian distribution which is centered around zero and has a sigma around 190 μm. This shows that the calibration of both data and Monte Carlo samples is correct and that the reconstruction algorithm works well in all cases. Since the sigma of data and Monte Carlo distributions are comparable the detector simulation seems to model well the uncertainties of the real measurement. Note that the small sigma also underlines the importance of the event based calibration correction as introduced in Sect. I.5. Without this

optimization the 25 ns jitter of the cosmics results in widths of the residuals of approximately $500\ \mu\text{m} = \sqrt{(54\ \mu\text{m/ns} \cdot \frac{25\ \text{ns}}{\sqrt{12}})^2 + (250\ \mu\text{m})^2}$, where $250\ \mu\text{m}$ is the intrinsic design cell resolution [51] and $\sqrt{12}$ accounts for the uniform distribution of the arrival times.

Note that the observed sigma of $180\ \mu\text{m}$ in the data with magnetic field does not directly correspond to the single cell hit resolution. Even though the residuals are shown only for a single layer the probed cell is still included into the segment fit and into the event based correction fit. In order to really compare the actual measurement with the predicted hit position, layer 2 can also be excluded from both fits. Thus segment direction and timing correction are determined by using only the remaining 7 hits. Still all hits are finally updated using the tTrig and v_{drift} output of the fit. Table I.6.1 summarizes the sigmas of the residuals under different conditions in data and in simulated events with magnetic field. Using the observed values with and without the probed cell included in the fits one can estimate the true cell hit resolution of the Drift Tube chamber. In both cases, with and without the hit included, the remaining uncertainty of the segment fit causes the width of the residual to differ from the true single cell hit resolution. The conversion from residual width to cell resolution can be achieved using correction factors. These correction factors have been derived from a toy Monte Carlo performing a linear fit through 7 or 8 hits with their positions smeared by the design resolution of $250\ \mu\text{m}$:

The correction factor is 0.87 for the residual with the hit not in the fit and 1.17 for the residual with all 8 hits used for the fits. For the data this results in $\sigma_{\text{cell}} = 238 \pm 5\ \mu\text{m}$ and $\sigma_{\text{cell}} = 211 \pm 4\ \mu\text{m}$ respectively, which leads to the weighted average of $\overline{\sigma}_{\text{cell}} \approx 222\ \mu\text{m}$. In the Monte Carlo prediction one obtains a very similar value, $\overline{\sigma}_{\text{cell}}(\text{MC}) \approx 247\ \mu\text{m}$. To summarize, all this means that using the event based correction also with cosmics one can reach the design cell hit resolution.

	hit not included in calibration- and segment-fit	hit included in both fits	extracted cell hit resolution
data with $|\varphi_{\text{SL}}| < 10°$ and $N_{\text{hits}} = 8$	$\sigma = 274 \pm 6\ \mu\text{m}$	$\sigma = 180 \pm 3\ \mu\text{m}$	$\sigma_{\text{cell}} \approx 222\ \mu\text{m}$
MC with $|\varphi_{\text{SL}}| < 10°$ and $N_{\text{hits}} = 8$	$\sigma = 334 \pm 11\ \mu\text{m}$	$\sigma = 192 \pm 6\ \mu\text{m}$	$\sigma_{\text{cell}} \approx 247\ \mu\text{m}$

Table I.6.1: *This Table summarizes the widths of the residuals for a single layer. The distinction between cosmic data and prediction from simulation with magnetic field is made, as well as the comparison whether the probed layer is included in the calibration- and segment-fit or not. Note that the true cell hit resolution is expected to be in-between, see last column (obtained by applying certain correction factors).*

Besides the performance of the segment-reconstruction in terms of resolution, Fig. I.6.3 displays the distribution of the number of hits contributing to a segment. The amount of hits used to create a combined Superϕ-segment and a θ-segment is shown for data and simulation in comparison. One can see that hardware and software show very good results since in most cases 8 hits and 4 hits respectively can be measured by the chamber and are combined to form a segment. As one would expect in the simulation the efficiency for this ideal case is a bit higher than in the data, where one or more hits are lost more often or layers with multiple hits

may contribute. These multiple hits can arise from secondaries or noise. Still the measured and the predicted performance is quite comparable. Also note that for all events the trigger conditions as defined earlier are required (see Sect. I.5). This explains why in the case of the Superϕ-combination only a small number of segments with < 7 hits can be found.
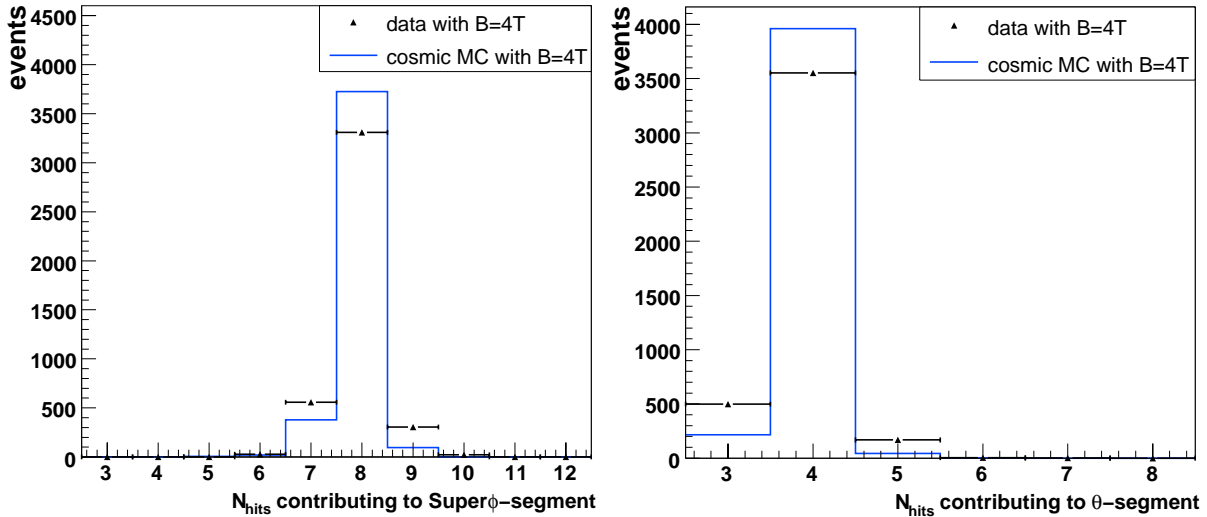

Figure I.6.3: *In order to study the performance of the local reconstruction the number of hits contributing to the chamber segments are displayed here, for data and simulated events in comparison as seen by MB3 in Sector 10 of YB+2 with magnetic field (data normalized to MC statistics). The left plot corresponds to the combined Superϕ-segment and the right one to the measurement of the single θ-Superlayer.*

I.6.2   Results from cosmic track-reconstruction

As introduced in Sect. I.5 the different chamber measurements can be combined performing a track-fit. The track parameters such as angle and momentum of the cosmic muon can be accessed at the innermost muon station after both forward- and backward-fit have been applied. As always only events fulfilling the trigger conditions are analyzed as discussed in Sect. I.5. No further selection cuts have been applied in order to select only high quality tracks. The tracks are simply required to fulfill some loose χ^2-cuts on the compatibility of including new hits in the fit (30000 for forward-fit using 3D-segments, 100 for backward-fit using cell hits, $N_{dof} = 1$) and that at least 2 Drift Tube chambers contribute. Comparing distributions in data and in simulation again the Monte Carlo prediction can be validated and by looking at track-parameters the performance of the tracking software can be tested. In general the results from the Cosmic Challenge represent the first time where real tracks can be measured with the CMS detector and can even be reconstructed, including the influence of the magnetic field and using the official software packages.

The first important number is the average efficiency of the track-reconstruction: In the data as well as in the simulated events with magnetic field the CosmicMuonProducer can find tracks

with an average efficiency of

$$\varepsilon \; (\text{track-reconstruction}) \approx 97\% \tag{I.6.1}$$

for the triggered events. Only less than 1‰ of the events in data and in simulation have more than one cosmic track, caused by noise, segment ambiguities or by two real cosmics in coincidence. Thus so-called "ghost tracks" are very rare. These efficiencies demonstrate impressively the good performance of the cosmic track finder.

Figures I.6.4 and I.6.5 show the global tracking angles φ_{track} and ϑ_{track} for the innermost measurement as defined in Fig. I.5.3. Again the data are normalized to the lower Monte Carlo statistics ($N_{data}/N_{MC} = 6.2$ for B_{off}, $N_{data}/N_{MC} = 7.2$ for B_{on}) and the dropping of the trigger-efficiency at large angles is included in the simulation. Just like in the case of local reconstruction also at the tracking-level the simulation nicely reproduces the shape of the measured distributions, both with and without magnetic field. Also the physical content of the plots stays within expectations: The φ_{track}-distribution is peaked around $-90°$ which corresponds to a vertical cosmic muon coming from above. Also in ϑ_{track} a peak around $90°$ can be observed in the measured and simulated events, again since vertical cosmic muons are expected to dominate. Due to the specific acceptance of the MTCC-Drift Tube setup a non-Gaussian and asymmetric behavior can be seen. Especially in φ the staggering of the chambers relative to each other and the fact that the inclined sector 11 is active leads to the patterns in this figure. In the φ-plane the bending of the magnetic field takes place and causes some differences between the left and the right plot. But just as in the case of local reconstruction, the biggest effect of the magnetic field is the drop in statistics as less cosmic muons reach the bottom sectors and form a track there.

Figure I.6.4: *The angle φ_{track} for the innermost measurement of the cosmic muon track is shown here, for simulation (continuous line) and data (triangles, normalized to MC statistics) in comparison. The left plot corresponds to the case without magnetic field, the right plot to 4 T within the solenoid.*

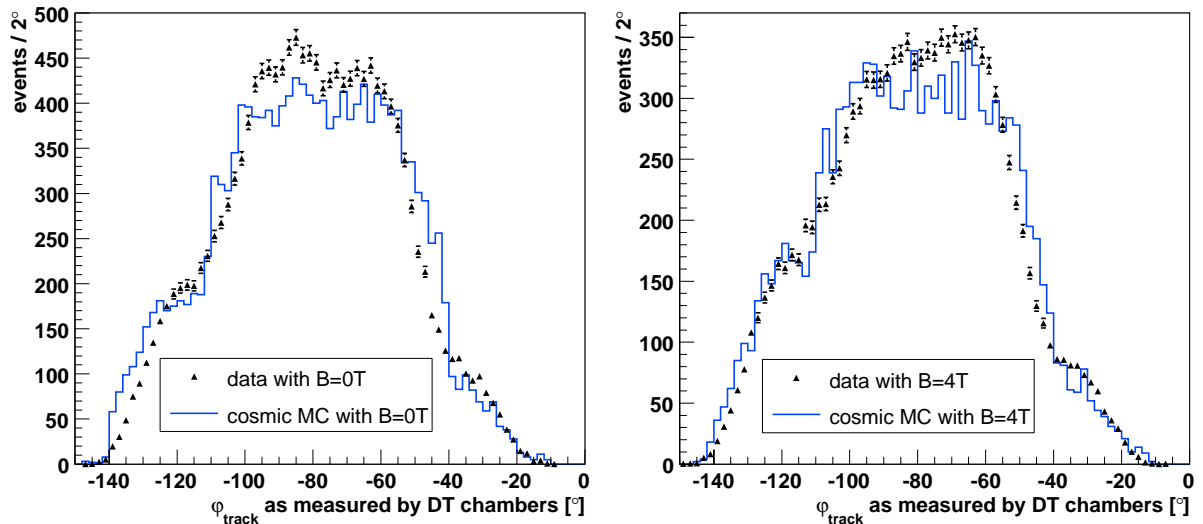**Figure I.6.5:** *The angle ϑ_{track} for the innermost measurement of the cosmic muon track is shown here, for simulation (continuous line) and data (triangles, normalized to MC statistics) in comparison. The left plot corresponds to the case without magnetic field, the right plot to 4 T within the solenoid.*
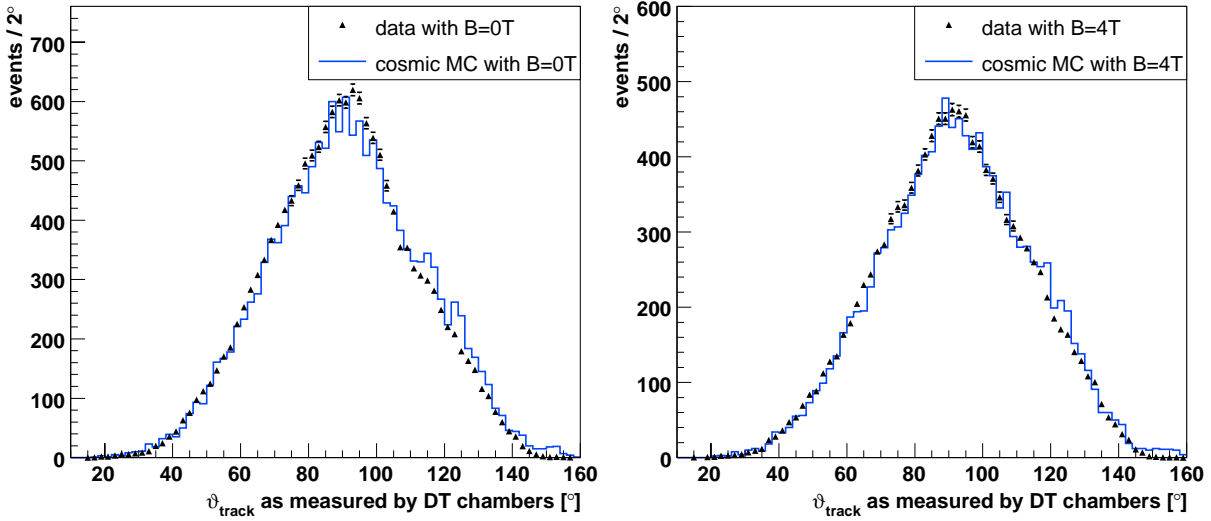
In order to test the correctness of the track-reconstruction the measured angles can be compared with the information obtained from simulation. During the detector simulation the muons are propagated through the detector material, accounting for the interaction with the magnetic field and with the hardware material. The angle at the innermost tracking-measurement can be compared to the "true" angle of the simulated muon inside the same cell. This is shown in the left plot of Fig. I.6.6 where the difference between simulated and reconstructed φ_{track} is displayed for the Monte Carlo sample with magnetic field. Since measurement and true value are compared at the same place the magnetic field or multiple scattering do not affect this difference. One can see a narrow Gaussian distribution which is centered at zero, thus in most cases the cosmic muon is measured correctly. The width amounts to only $0.1°$ which means that the cosmic tracks are reconstructed very precisely.

The right plot of Fig. I.6.6 compares the amount of cell hits contributing to the track-fit in data and in Monte Carlo events. The prominent peaks correspond to 24 hits, 32 hits, 36 hits and 44 hits; MB1, MB2 and MB3 can contribute with 12 hits at most while MB4 only contains 8 layers, thus the combination of the different chambers explains this pattern. Since a cosmic muon crosses the detector setup coming from all directions not all four muon stations are hit in each event. Comparing the data with the Monte Carlo prediction one can see that the ideal case of two, three or four crossed Drift Tube chambers with all layers hit is present less often in the real measured tracks. Hits may get lost or additional hits due to noise may contribute. Note that alignment corrections have not been applied in the data yet and thus displaced hits might not match with the residual track fit.

In Fig. I.6.7 one can see how the different muon stations contribute to the track-finding when a magnetic field is present. Displayed here are the number of cell hits used for the backward-fit. The different bins in x correspond to the different chambers MB1-MB4, one group of bins consists
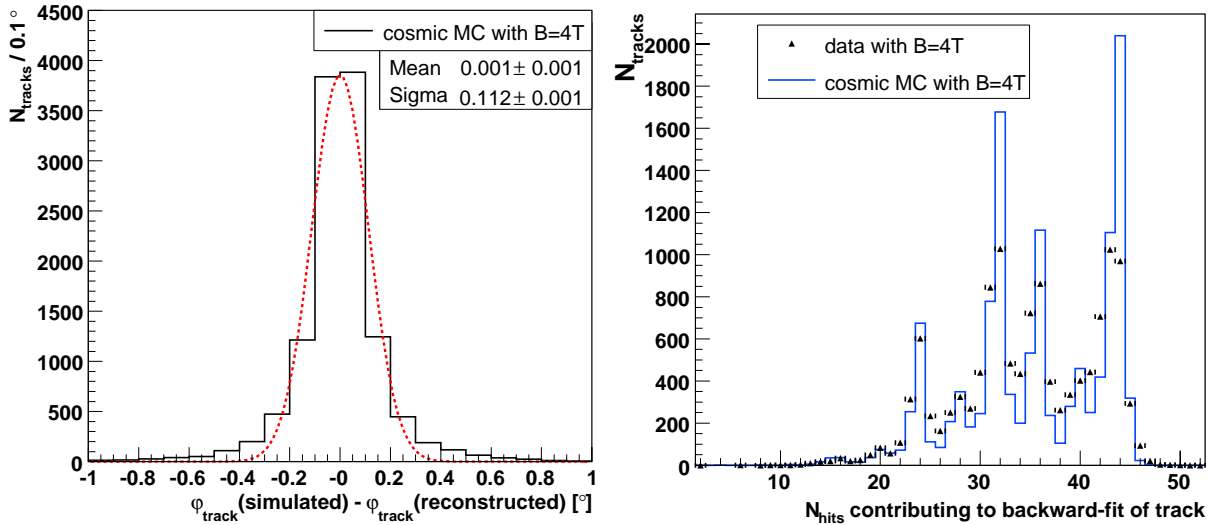
Figure I.6.6: *Here the performance of the cosmic track-reconstruction is studied. The left plot shows the difference between simulated and reconstructed φ_{track}-angle at the innermost measurement for simulated events with magnetic field; a Gaussian fit is performed. The simulated information is taken from the detector simulation at the position of the innermost track measurement. The right plot compares the amount of hits which contribute to the final track-fit in data and simulated events with magnetic field (data normalized to MC statistics).*

of sector 10 in YB+1, the next of sector 10 in YB+2 and the last peak of sector 11 in YB+2. Note that in both sectors 10 two MB4 are present (the second MB4 is assigned to sector 14) which explains the two peaks standing apart. In general the simulation describes the distribution of the hits as seen in the data quite accurately. Consistent with the numbers from local reconstruction also here an equal amount of tracks is found in each sector of the MTCC-setup. The cosmic track-reconstruction is also quite efficient for inclined sectors such as sector 11. In most events hits from MB2 and MB3 are used since these two chambers are required to contain signals by the trigger. This underlines again the importance of modeling the trigger-conditions also in the simulated events.

The results so far demonstrate that it is possible to reconstruct cosmic muon tracks, with and without the 4 T magnetic field. Of course this suggests to also investigate the design motivation why the CMS detector has such a strong magnetic field at all: The transverse momenta of the particles are to be measured. Figure I.6.8 addresses this issue since the p_T of the cosmic muons as seen at the innermost point of the track is shown here. The upper plot displays the low-momentum region and the lower plot refers to cosmic muons with higher p_T, in both distributions the data are normalized to the simulated events only in the range visible in the plots ($N_{data}/N_{MC} = 7.0$ for upper plot, $N_{data}/N_{MC} = 10.1$ for lower plot).
The shape of the distribution is as expected since cosmic muons follow a steep declining energy dependence (see Sect. I.2.1). The agreement between data and Monte Carlo prediction is reasonable. Data and MC distributions have their maximum at some minimal momentum of $2 - 4$ GeV; muons with even smaller momentum do not cross enough active detector material to form a track candidate. In general the shape of the distribution is reproduced nicely by the
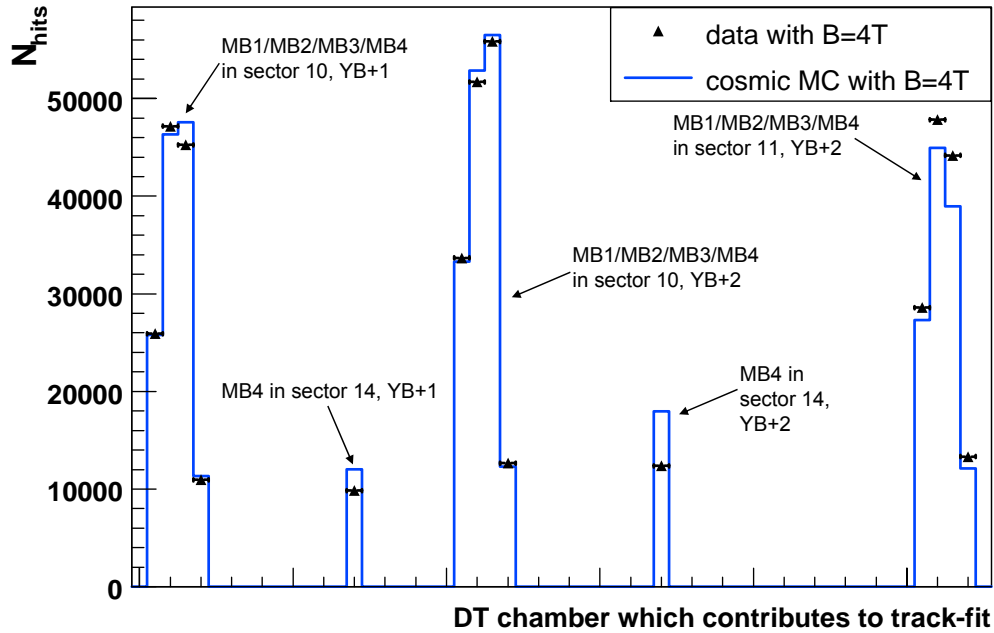
Figure I.6.7: *The plot illustrates how often the different DT chambers of the MTCC-setup contribute with hits to the track-fit. Each bin in x which has an entry corresponds to a specific chamber. Data and Monte Carlo prediction with magnetic field are compared to each other (data normalized to MC statistics). One can see that in general all 3 sectors are illuminated equally.*

simulation. Also the high-p_T tail is described properly, normalizing only to the events above 30 GeV. Both plots show that the energy dependencies assumed in the cosmic generator are correct and that the track-reconstruction performs similar in data and simulation. Looking at the upper plot the Monte Carlo curve appears to be narrower than the measured one, leading to a MC excess at small momenta and less populated high-p_T tail. This could be caused either by the fact that no alignment corrections have been applied in the data or by an overestimation of the assumed resolutions in the detector simulation, as well as a miscalibrated momentum scale. Small momenta are also affected by the lower energy cut of 7 GeV during generation. In the data some cosmic muons with smaller momenta might also produce track-candidates. Also, as mentioned before, the trigger conditions are only modeled approximately in the simulation. All these points, as well as the intrinsic uncertainty of the cosmic generator, make clear that there are differences to be expected between data and Monte Carlo which might affect the level of agreement in the p_T-distribution. Trigger and 7 GeV cut have especially an impact on the low momentum region (upper plot). At high-p_T the trigger-efficiency should become flat in energy and an energy cut in the generation has only little influence. Thus in the lower plot missing alignment corrections and the unknown momentum-resolutions and absolute momentum-scale dominate the uncertainties of the measurement.
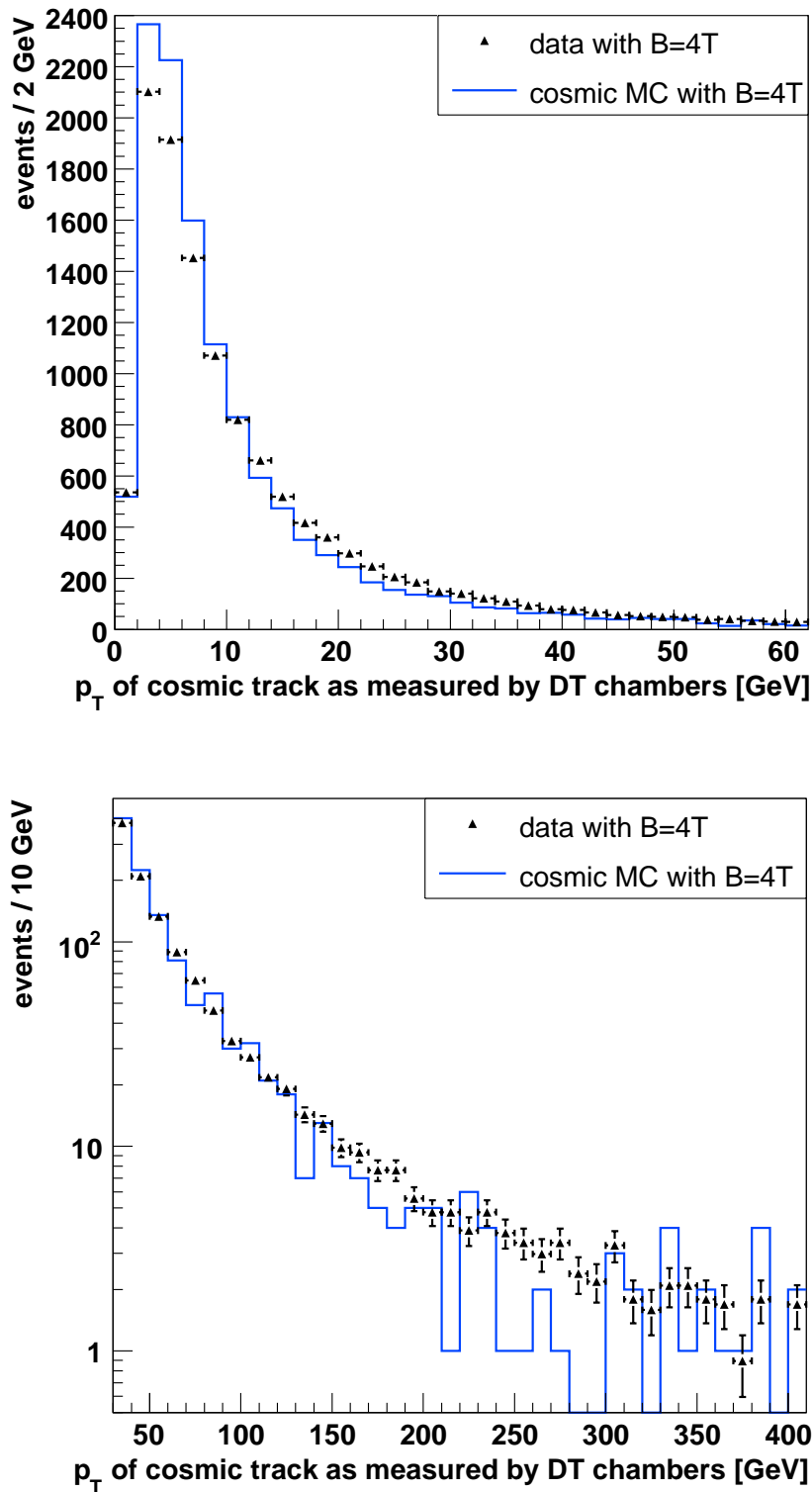
Figure I.6.8: *The p_T-distribution at the innermost measurement of the cosmic muon track is shown here, for simulation (continuous line) and data (triangles, normalized to MC statistics in the range displayed). Note that in the low-p_T region (upper plot) differences might arise from the fact that in simulation a lower threshold of $E_{gen} > 7$ GeV is used and no full trigger-simulation is run. In the high-p_T tail (lower plot) the important fact is that no alignment-corrections have been applied to the data and that a calibration of the momentum-resolution and absolute momentum-scale is missing.*

The next step is to investigate the quality of the p_T-measurement. This can be easily done using the Monte Carlo sample since the simulated information can be compared to the reconstructed one. The left plot of Fig. I.6.9 shows the difference between simulated and reconstructed p_T. Again the simulated hit-information at the position of the innermost track-measurement is used to obtain the "true" p_T-value. A Gaussian fit is performed and both mean and sigma show that when using cosmics the momentum can be estimated from the track.

An assessment of the performance of the momentum measurement can be extracted from the right plot of Fig. I.6.9. Since the measured quantity is the sagitta s of the track where $s \propto 1/p_T$, and since $\frac{\Delta(1/p_T)}{1/p_T} = \frac{\Delta p_T}{p_T}$, the sigma of the Gaussian fit shown corresponds to the p_T-resolution. So using the MTCC-setup and the cosmic track finder the simulation predicts a relative p_T-resolution of 24%. This value can be compared to the results of Standalone muon reconstruction, i.e. muon system only, as stated in the Technical Design Report [22]: Here a p_T-resolution of $\approx 10\%$ is expected for muons from pp-collisions in the range from $5 - 100$ GeV and measured by the barrel muon system. The difference can be explained by the peculiarities of cosmic muons, i.e. their timing uncertainty and the lack of a vertex constraint.
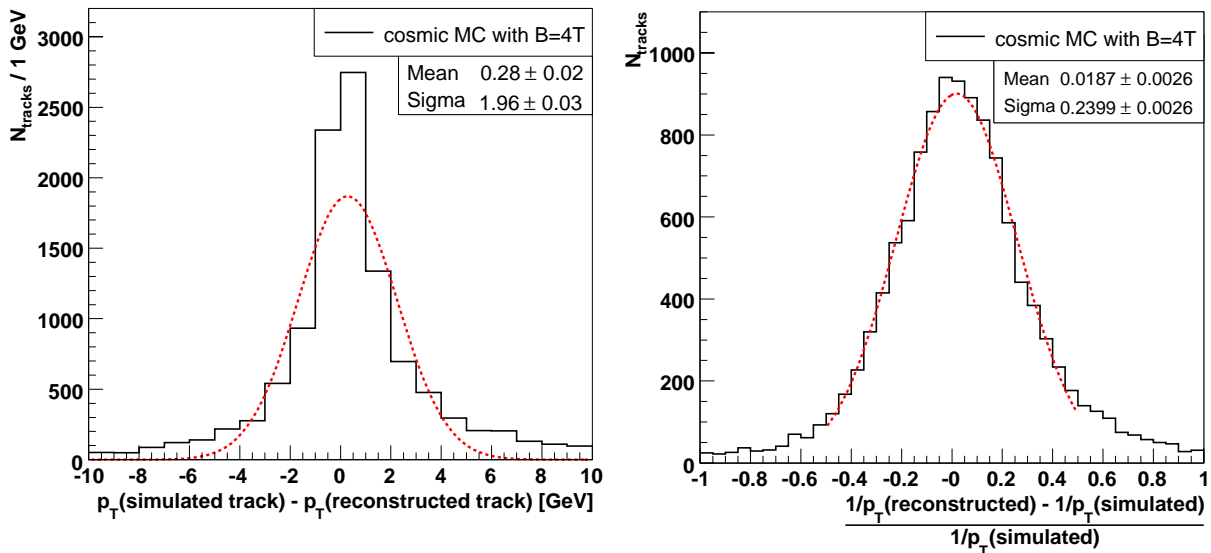
Figure I.6.9: *The quality of the p_T-measurement is studied here, using the simulated sample with magnetic field. The left plot shows the difference between simulated and reconstructed p_T-value at the innermost measurement, a Gaussian fit is performed. The simulated information is taken from the detector simulation at the position of the innermost track measurement. In the right plot the reconstructed and simulated information is used to compute $\Delta(\frac{1}{p_T})/\frac{1}{p_T}$. The widths of the Gaussian fit corresponds to the p_T-resolution when measuring cosmic muons with the DT chambers only, without any vertex constraint applied.*

I.6.3 Multiple scattering, charge ratio and rate estimates

I.6.3.1 The effect of multiple scattering

At low momenta the resolution of the track-measurement is mainly limited by the effect of multiple scattering of the cosmic muons inside the iron of the magnetic flux return yoke and the hadronic calorimeter. Thus an important check of the precision of the track-reconstruction is to test whether the observed measurements are compatible with the uncertainties caused by multiple scattering.

In this context Fig. I.6.10 compares the generated angle ϑ_{track} with the reconstructed one. As the cosmic muons are generated on a cylinder outside the CMS detector and are mainly reconstructed in the bottom sectors 10 and 11 of the MTCC-setup (MB1/MB2 since innermost measurement), the cosmic muons have to pass through a well defined material volume of ≈ 330 radiation lengths [22]. This yields a deviation in the angle due to multiple scattering of

$$\Delta\vartheta_{track} \approx \frac{13.6 \text{ MeV}}{5 \text{ GeV}}\sqrt{330} = 2.8° \qquad , \qquad (\text{I.6.2})$$

see [6]. Since the generated cosmic muons have a minimal energy of 7 GeV and since they loose further energy on their way through the detector material, 5 GeV is assumed in this rough estimate. According to [6] $\Delta\vartheta_{track}$ can be approximated by a simple Gaussian using the central 98% of the distribution. In the left plot of Fig. I.6.10 a double Gaussian fit, i.e. a linear sum of two Gaussians, is performed on the difference of generated and reconstructed angle. This fit accounts for the larger tails which are typical of multiple scattering processes. Note that in the ϑ-plane the cosmic muons are not affected by any bending due to the magnetic field, thus the distribution has a mean centered around zero. The width of the narrower Gaussian is $\approx 2.8°$ which means that the results of the track algorithm are well compatible with the limitations given by multiple scattering. By shrinking the fit range one can also succeed in matching better with the very inner peak of the distribution. In this case the requirement that the inner Gaussian roughly corresponds to the central 98% of the distribution does no longer hold for sure. Thus the resulting inner width can no longer be compared to the expected $\approx 2.8°$ even though the fit would match better to the data.

Within one sector of the MTCC-setup one can also investigate the multiple scattering effects between two fixed muon stations, e.g. between MB1 and MB3. Again in the ϑ-plane only the material of ≈ 50 radiation lengths [22] can cause deviations in the angles, not the magnetic field. This results in an expected deviation in the angle of

$$\Delta\vartheta_{track} \approx \frac{13.6 \text{ MeV}}{5 \text{ GeV}}\sqrt{50} = 1.1° \qquad , \qquad (\text{I.6.3})$$

assuming a mean cosmic energy of 5 GeV. This expectation can now be compared with the reconstructed difference as seen in data and in simulated events. In addition to this the measurements from local chamber reconstruction, i.e. before any track-fit has been performed, can be compared to the results obtained from track-reconstruction. All these studies can be seen in Fig. I.6.11, here for the samples with magnetic field. The left column corresponds to the data,
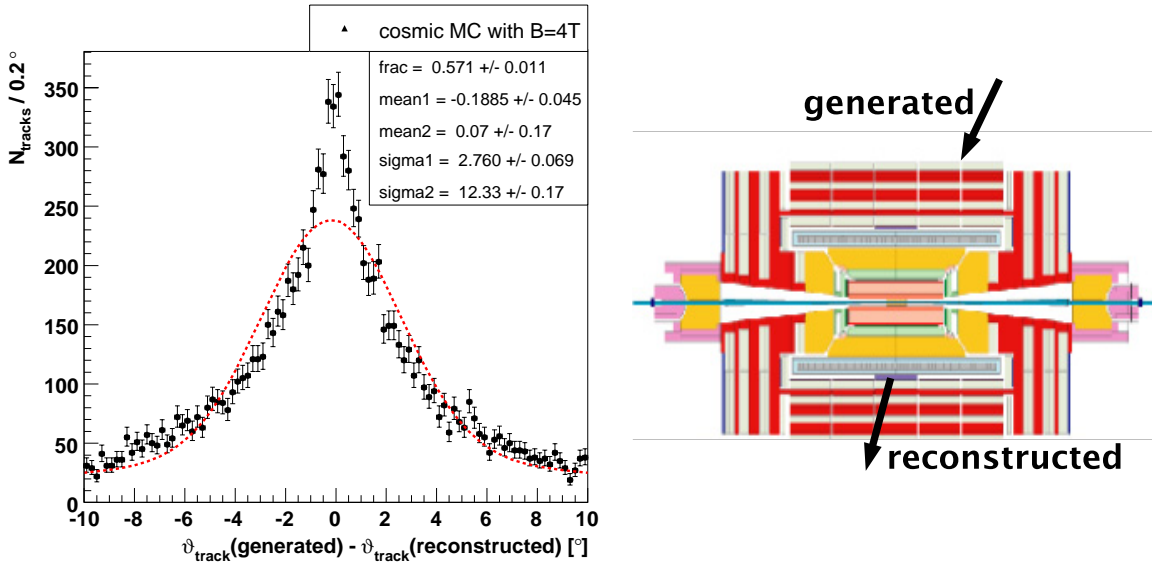
Figure I.6.10: *Left: The difference between generated ϑ_{track} and the one reconstructed at the innermost measurement is shown, using simulated cosmics with magnetic field. A deviation in the angle is expected from multiple scattering effects and a double Gaussian fit is performed to account for the long tails.*
Right: Illustration of where generated and reconstructed angles are located.

the right depicts the Monte Carlo expectation. The two upper plots show the difference of the ϑ_{track}-angles as observed by the chamber segments of local reconstruction in MB1 and MB3. Again a double Gaussian-fit has been performed to account for the long multiple scattering tails. In this way the inner Gaussian roughly corresponds to the central 98% of the distribution [6]. The sigma of the narrower Gaussian of both data and Monte Carlo are around 1.8°, so only slightly larger than the estimated 1.1°. The two bottom plots now represent the measurements after the track-fit has been performed (innermost measurement=MB1, outermost measurement=MB3). Note that the scale on the x-axis has shrunken since the distributions are much narrower. Both data and Monte Carlo prediction have $\sigma_1 \approx 0.7°$, thus the fitting procedure largely improves the measurement. In general the performance of the track-reconstruction seems to be very comparable in data and in simulated events, only the bottom data distribution reveals a slight offset in the mean of the narrower Gaussian, not reproduced by the detector simulation. Again influences of the missing chamber alignment could explain this.

I.6.3.2 Measuring the charge of cosmic muons

As the p_T-measurement is based on the curvature of the cosmic muon track caused by the magnetic field present in the iron return yoke, the charge of the muon can be determined from the sign of the curvature. In the simulated events the knowledge of the "true" charge can be used in order to estimate the charge misidentification probability. Averaged over all momenta and angles of the generated cosmic muons the Monte Carlo sample predicts a probability of

$$p \text{ (charge misidentification)} = 0.036 \pm 0.002 \qquad . \qquad (I.6.4)$$
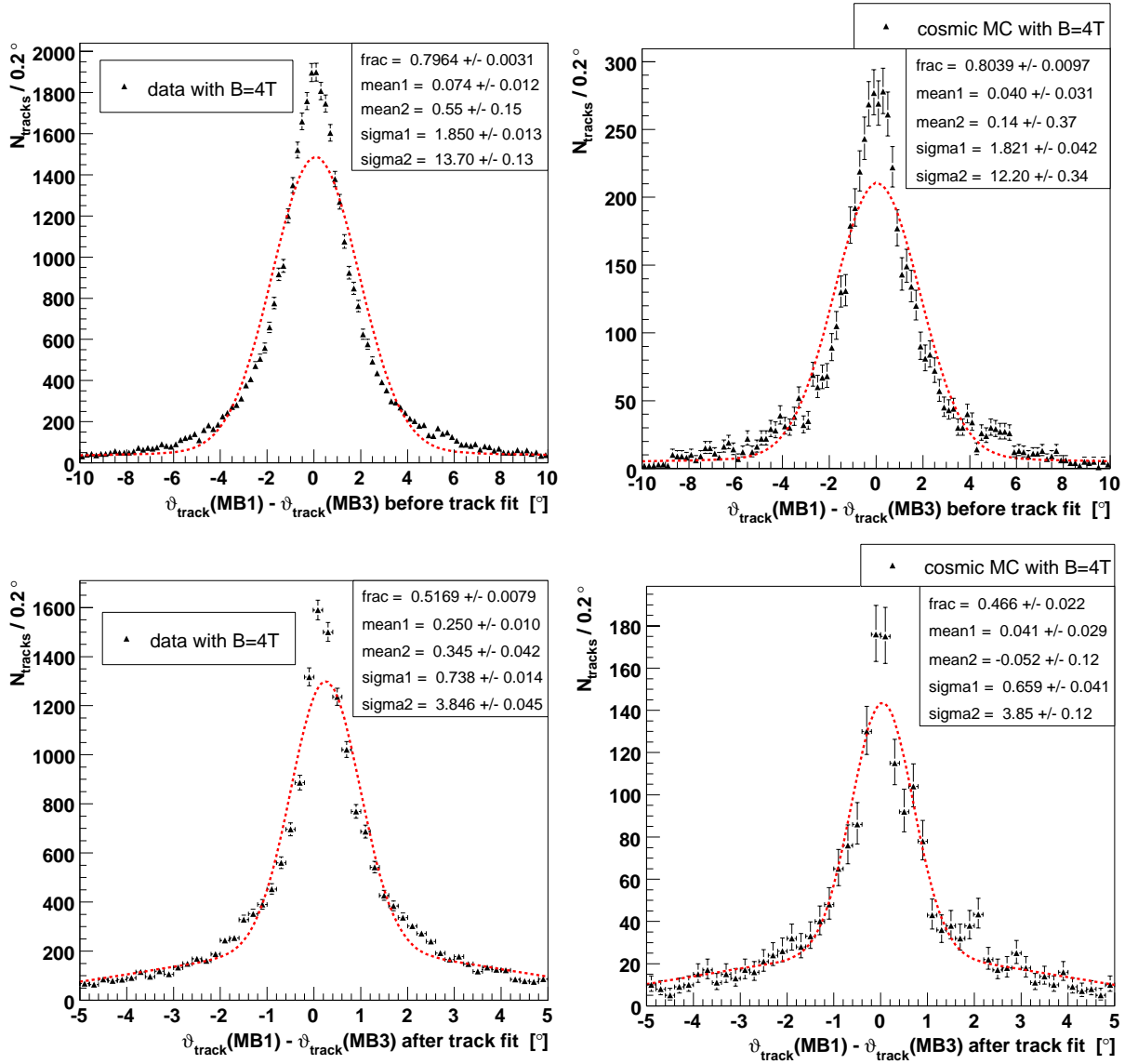
Figure I.6.11: *Differences in ϑ_{track} due to multiple scattering are investigated at two distinct places (between MB1 and MB3). The left column refers to data with magnetic field and the right to the MC prediction. The two upper plots display the difference in the angle at chamber reconstruction level, thus the difference for the segments before a track-fit is performed, and the bottom plots the final track-angles where innermost and outermost measurement correspond to MB1 and MB3 respectively. In all cases a double Gaussian fit is performed accounting for the long tails.*

In [22] a value of 0.001 is stated for 10 GeV standalone-muons from pp-collisions, here using a vertex constraint. Still the 4% observed with cosmic muons is a promising results, especially as additional track-quality cuts can further improve this value. Note that in order to obtain these results a small fix in the CosmicMuonProducer code of the CMSSW version used had to be performed which eliminates unphysical charge-flips during reconstruction.

This low charge misidentification probability shows that in principle the charge of the cosmic muons can be measured correctly. Before presenting the results one should remember the nat-

ural charge ratio of cosmic muons $N_{\mu+}/N_{\mu-} \approx 1.3$, see Sect. I.2.1. In this part of the study only tracks reconstructed in the two horizontal sectors 10 are used in order to ensure a more symmetrical detector setup.

The results of the charge measurements and the corresponding expected values are summarized in Tab. I.6.2, again averaged over all occurring momenta and angles. Note that for the reconstructed tracks the ratio is corrected for the charge misidentification probability of $\varepsilon = 3.6\%$ mentioned above, assuming

$$N^+(\text{true}) = N^+(\text{meas}) \cdot (1 - \varepsilon) + N^-(\text{meas}) \cdot \varepsilon \qquad . \qquad (\text{I.6.5})$$

The errors are statistical only.

	data events	Monte Carlo events
$\frac{N_{\mu+}}{N_{\mu-}}$(generated) for all events		1.33
$\frac{N_{\mu+}}{N_{\mu-}}$(generated) for events with track		1.32 ± 0.03
$\frac{N_{\mu+}}{N_{\mu-}}$(reconstructed) for events with track	1.35 ± 0.01	1.35 ± 0.03
$\frac{N_{\mu+}}{N_{\mu-}}$ from independent measurements for vertical 10 GeV muons only [41]	1.27 ± 0.01	

Table I.6.2: *Summary of generated, reconstructed and measured values of the cosmic muon charge ratio. The reconstructed values in simulated and measured events with magnetic field are corrected for a charge misidentification probability of 3.6% and can be compared to each other (line 3), while line 2 states the "true charge ratio" of the reconstructed tracks within the simulation.*

One can see that in the simulation the charge ratio which is an output of the cosmic generator and the ratio obtained by the track-reconstruction, using sector 10 only, are well within their errors. Also the ratio reconstructed in the data is very similar to the one obtained with the Monte Carlo sample (line 3). Still the measured ratio of 1.35 is higher than the one expected from previous measurements (1.27 for vertical 10 GeV muons only). This could be due to the specific (asymmetric) acceptance of the MTCC-setup and the influence of the magnetic field inside the CMS detector which might enhance a particular charge. In this context averaging over all angles and momenta does also affect the reconstructed charge ratio. In general, all values are compatible with a charge ratio of ≈ 1.3.

A dedicated analysis focusing on the measurement of the charge ratio using MTCC data can be found in [53], obtaining on average $\frac{N_{\mu+}}{N_{\mu-}} = 1.282 \pm 0.008$.

I.6.3.3   Comparison of measured and predicted rates

Another task of CMSCGEN is the prediction of rates of cosmic muons arriving at certain sub-detector parts (see Sect. I.2.1), for example to estimate the statistics potentially available for detector alignment. Cosmic ray muons may also constitute a background in the search for rare physics processes during pp-collisions. The predicted rates can be compared to the trigger rates measured with the Drift Tube MTCC setup. Such comparisons help to validate the Monte Carlo prediction and qualify the performance of the trigger setup used during the Cosmic Challenge.

The DT trigger-rates during MTCC data taking can be deduced from the Run Summary Web Page. The efficiency of the trigger may vary at the level of a few percent thus adding a systematic uncertainty. Also the predicted rates have some uncertainty, largely caused by the modeling of the trigger conditions as described in Sect. I.5. While the DT trigger efficiency has been determined to be $\approx 90\%$ with a bunched muon testbeam [54], with cosmic muons the average DT trigger efficiency has been measured to be only $\varepsilon_{\text{trigger}} = 45 \pm 5\%$ [55], varying from run to run. The low efficiency is attributed to the non-synchronization of cosmic muons with the trigger clock. This causes an increased probability of poor quality triggers and of muons assigned to a neighboring bunch crossing. This trigger efficiency is a rough estimate for the trigger configuration used in the two runs analyzed for this study. It has been applied as a constant factor to correct the predicted Monte Carlo rates accordingly.

The comparison between the DT trigger rates as measured in the two particular MTCC runs used in this analysis and the ones predicted by simulation (including the modeling of the trigger conditions as discussed before) are summarized in Tab. I.6.3. The DT trigger for these two runs included only MB2 and MB3 stations in the three MTCC sectors. The third column states the original predicted rates, the errors include statistical and some systematic contributions, following Equations I.2.2 and I.2.3. The second column of the table refers to the predicted rates after correcting for a trigger efficiency of $45 \pm 5\%$, the total error accounts for the uncertainties of the predicted rates and for the uncertainty of the trigger efficiency. For the data (first column) a 10% uncertainty is assumed corresponding to the relative uncertainty of the trigger efficiency.

	measured rates	predicted rates $\times \varepsilon_{\text{trigger}}$	predicted rates
$B = 0$ T	90 ± 9 Hz	118 ± 15 (tot.) Hz	262 ± 2 (stat.) ± 18 (syst.) Hz
$B = 4$ T	70 ± 7 Hz	91 ± 12 (tot.) Hz	202 ± 2 (stat.) ± 14 (syst.) Hz

Table I.6.3: *Comparison of the cosmic muon trigger rates as predicted by CMSCGEN with the trigger rates measured by the Drift Tube MTCC trigger configuration with MB2 and MB3 stations participating in the DT trigger. The rates without magnetic field (first line) are slightly higher than with magnetic field (second line) due to bending of low momentum muons out of the acceptance. Rates in the second column are already corrected for an assumed trigger-efficiency of 45%, the last column states the original prediction.*

In general, data and Monte Carlo prediction (corrected for $\varepsilon_{\text{trigger}}$) have a value consistent with 100 Hz, thus the simulation of the cosmic muon rates gives results in the correct order of magnitude. Both also show the trend towards lower trigger rates with the magnetic field being present. The measured rates are lower by a factor of 0.8, but they still agree within errors ($\approx 2\sigma$). This could be related to the modeling of the trigger conditions rather than using a detailed simulation of the specific MTCC trigger setup which had to be modified with respect to its design for pp-collisions. The fact, that the predicted trigger rates without the efficiency correction are much higher than the measured ones, is yet another indication that the measured low efficiency is correct for the case of cosmic muons and support the identified reason related to the timing uncertainty of the incoming cosmic muons. In summary, the cosmic simulation is well suited to predict rates quite precisely, allowing reliable cosmic rate estimates for future tests and studies.

I.6.4   Summary and conclusions

In summary, at all levels of reconstruction the comparison of cosmic muons from MTCC with the predictions made by the cosmic generator CMSCGEN shows very promising results. This demonstrates that the assumed angular- and energy-dependencies within the cosmic generator are correct, and that the detector simulation within the CMS software can describe the data quite precisely.

The analysis of the MTCC data show that the CMS experiment can indeed be operated in a stable mode, including a magnetic field. Physics objects up to tracks can be reconstructed correctly in MTCC data as well as Monte Carlo events using the barrel muon system. The extraction of cosmic muon properties from simulated data allows to study the performance of reconstruction algorithms and their measurement precision. These results meet or even exceed the expectations for non-bunched muons so far. Moreover, cosmic muons are anticipated to be used for synchronization and alignment purposes in the future. All this underlines that cosmic muons are the ideal source to study the detector performance during the commissioning of CMS and the early days of LHC-running, ranging from basic operation tests up to high definition objects such as tracks.

Part II

Generic Searches for New Physics

Chapter II.1

The Concept of Model Independent Searches

After having discussed the important aspects of commissioning of the CMS detector in the first part of this thesis the focus will now move on to the preparation of physics analyses. Since also these are rather complex, one cannot start designing them when first data arrive, but they have to be prepared well in advance. These analyses running on the first data will initially continue to commission the detector, trying to understand and control its physics output. In addition to this a physics analysis utilizing the muon detectors benefits from the experience and knowledge gained during commissioning with cosmic ray muons. Of course the aim is to overcome these initial problems as fast as possible in order to be ready for exciting physics measurements.

II.1.1   Introductory remarks

The highly anticipated start-up of the Large Hadron Collider (LHC) promises to enter an unknown territory in particle physics. Clearly the hope is that a door to the regime of new physics can be opened, providing answers to some long outstanding questions on the fundamental principles of the universe. Still it is not clear how effects of new physics will look like in the first data and which theory beyond the Standard Model (SM) will describe them. In order not to miss any unexpected signal it might be desirable to systematically analyze the data with as little bias as possible.

In the context of a generic search for new physics in early LHC data a special algorithm called "MUSiC" (Model Unspecific Search in CMS)[1] has been developed [56][57][58]. Details of the strategy, the algorithm and some exemplary results on the sensitivity to various models beyond the SM will be presented in the second part of this thesis, investigating expectations from simulated events assuming $1\,\mathrm{fb}^{-1}$ of data at a center of mass energy of $14\,\mathrm{TeV}$.

It should be made clear right from the start that it is not the aim of MUSiC to repeat, nor to compete with, the work of the many dedicated analyses already present in CMS. It has to be regarded as an alternative approach with advantages as well as drawbacks, thus it is a complementary addition to the CMS physics program. Also the obvious interplay between such a

---

[1] work of this analysis in close collaboration with [59]

generic search and more specific analyses will be discussed below. Ultimately significant deviations found by MUSiC need to be taken with some care and have to be interpreted by physicists. The suspicious final state(s) have to be investigated in detail in order to determine the reason of the deviation: May it be a detector effect, a lack of understanding of the event generation and simulation or truly a new physics signal. Here crosschecks from more specific analyses are likely and desired. Thus to summarize:

- MUSiC is a *global physics monitor*, sending "alarms" in case of interesting deviations

- the conclusion that this deviation is a discovery cannot be drawn by MUSiC alone but only in cooperation with more dedicated studies

- MUSiC can help to improve the understanding of detector and SM backgrounds and contribute to the MC-tuning

- MUSiC has a rather large coverage of new physics, but for some signals it is likely to be less sensitive compared to dedicated analyses in a specific channel.

II.1.2   The concept of model independent searches

Model independent searches aim to analyze a large fraction of data, systematically scanning them for deviations from the Standard Model. Therefore the selection cuts are not optimized for any expected new physics signal; however, the quality of the measurement is ensured by selecting well-measured and well-understood physics objects such as isolated high-p_T leptons. Similar strategies have already been applied successfully at other accelerator experiments, see e.g. [60–63].

From the historical point of view the concept presented here is following the principal ideas of a similar strategy at L3 (1998) [63]. In terms of the search algorithm the main guide came from the H1 approach [61] and the concept has already been exercised at DØ with collision data [64]. In contrast to the VISTA approach [62] one does not rely on a complex self-correction model trying to tune the simulation according to the data, but one wants to benefit from the many detailed studies on particle identification and detector efficiencies. The approach is to use the current detector knowledge as input and then learn from the results of a global data-Monte Carlo comparison. In this way it is an excellent monitor to detect improvements or new discrepancies for high-p_T processes, thus cooperation with dedicated studies and contributions to the tuning of the simulation are evident.

The motivation for such an analysis strategy is obvious: The LHC will enter an unknown territory and there are multiple reasons why new physics is expected to appear. Unlike in experiments of the past there is an enormous number of predictions from theory of how exactly these new physics models will look like. Following the saying "expect the unexpected", a model independent search tries to cover a wide range of the phase space and is not limited to a specific topology. In this way it should be sensitive more exotic models with spectacular final states such as mini black holes, give a consistent picture of the various channels where a SUSY signal could contribute, or quickly discover discrepancies caused by detector effects or effects not properly accounted for in the Monte Carlo simulation.

However, such an ambitious strategy also has its drawbacks: For some signals it could be less sensitive than a model-driven analysis. In addition, since a variety of final states are investigated,

one has to rely more on the background predictions made by Monte Carlo generators. While some SM backgrounds may be reasonably well predicted in particular thanks to the availability of Monte Carlo programs based on matrix element calculations like ALPGEN or MC@NLO, there are others where this is likely not to be the case. As one obvious example, the QCD multi-jet background will have to be estimated from the data in such a generic search.

The variety of final states and distributions examined also leads to a fundamental and irreducible disadvantage: A statistical penalty factor has to be included for looking at many places. This issue is accounted for carefully within the search algorithm, see Chapter II.4. To summarize, it leads to a loss in significance in case of deviations found in the data.

As mentioned before each single final state cannot be studied in such a great detail as within a model-driven analysis, and not all effects can be understood and well modeled by the Monte Carlo expectation. There is a clear trade-off between trying to cover a large amount of data and describing all of it properly. The key issue is to estimate and implement uncertainties in the correct order of magnitude such that problematic areas of the phase space have a large error assigned. In this way only indications of new physics, unexpected detector effects or the insufficient knowledge on SM processes at $14\,\mathrm{TeV}$ should remain as significant deviations. In addition to this the physics interpretation of the deviations cannot only be done in a generic way: as in all analysis one has to prove, that the deviations are caused by new physics and not some poorly understood effect within the data or the MC. In the end the physicists looking at the objective results have to add some subjective knowledge or intuition, e.g. a deviation in a 10 jet channel is not really a surprise since Monte Carlo generators are not expected to model such extreme topologies correctly. Even if one could conclude that the deviations are caused by physics beyond the Standard Model there still remains the question known as the inverse LHC problem: What is the underlying theory causing this signal?

Before introducing the concept of the MUSiC approach the following "guidelines" for this analysis are set:

- Robustness: focus on well-understood physics objects, i.e. high p_{T}, central $|\eta|$ and solid object identification

- Model independence: no optimization of selection cuts with respect to some expected signal

- Simplicity: the steps of algorithm should be easy to follow, standard statistical estimators and methods should be preferred

- Completeness: include any possible systematic differences between data and Monte Carlo prediction in the search algorithm

- Expect new physics to contribute predominately to a single channel (heavy resonance like W'), but be prepared for the case where signals cause deviations in numerous final states (SUSY).

The concept of the MUSiC approach can be divided into 3 steps:

I. Focus on leptonic events

In order to have a well defined trigger stream and in order to reduce the QCD multi-jet background the analyzed phase space is restricted to events which contain **at least one charged lepton** (electron or muon). Selecting events with leptons also fits to the guideline of well understood objects mentioned above.

II. The concept of event classes

The events are then classified into so called **event classes** which group events according to their final state topology, see Figure II.1.1. These classes can be *exclusive* or *inclusive*. Each event class is defined by the amount of physics objects in the event, e.g. 1μ 3jet. In the exclusive case the exact number of particles is required (e.g. 1μ $0e$ 0γ 2jet $\not{E}_\mathrm{T}$), while the inclusive classes require only a minimal number of particles (e.g. 1μ 3jet $+X$, so at least one muon and at least 3 jets). The latter are denoted with the suffix "$+X$". Given the complex decay chains of e.g. SUSY events inclusive classes might be useful. Also for final states with many jets it might be desirable to combine events with e.g. ≥ 5 jets into a single inclusive class since Monte Carlo generators are not expected to model the kinematics of the 8th or 9th jet correctly anyway.

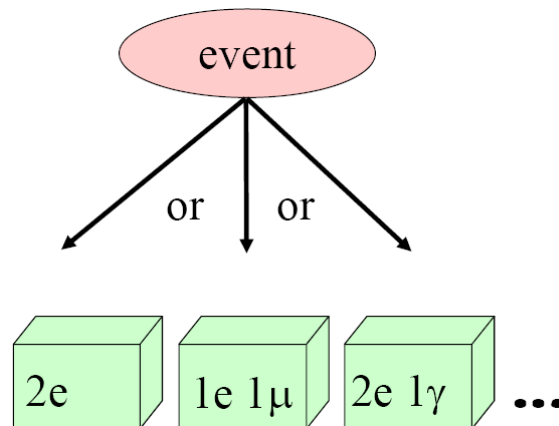**Figure II.1.1:** *Illustration of how events are distributed into event classes, here for the exclusive case.*

The following **physics objects** measured by the CMS detector are considered, leading to approximately 300-400 event classes (exclusive and inclusive each) to be considered:

- muons (μ)
- electrons (e)
- photons (γ)
- hadronic jets (jet)
- missing transverse energy ($\not{E}_\mathrm{T}$).

The LHC is designed to probe the high-energy frontier, thus the analysis assumes that new physics will appear in events with high-p_T objects. Selection cuts are desired to remain as simple as possible. Similar strategies are useful for any start-up physics study.

Tau-leptons are not included so far, only by decaying into electron or muon they can enter the selection. Once a τ-identification is well-studied and well-controlled with the first data one could imagine to include also tau-leptons as individual physics objects. A similar argumentation also holds for the inclusion of b-tagging, adding b-jets as physics objects. In addition to this taking into account the charges of the leptons would be a desirable extension in the future, thus distinguishing between same-sign and opposite-sign combinations.

III. Searching for deviations

Deviations from the Standard Model which might indicate new physics can be quantified by comparing the events of each event class as seen in the data to the expectation of simulated SM samples. Variables sensitive to new physics which are analyzed systematically within each MUSiC event class are:

- the **total cross section**, i.e. number of events per class

- **kinematic distributions** of an event class such as the scalar sum $\sum p_T$ of the transverse momenta of all its physics objects, the invariant mass M_{inv} or transverse invariant mass M_T in case of event classes containing $\not{E}_T$. In addition the $\not{E}_T$ distribution is investigated separately for all event classes which contain missing transverse energy.

The systematic analysis of all event classes is performed by the search algorithm explained below. Some representative interpretations of the results/output of this algorithm will be discussed in the end.

In this context the examination of the total cross section will not be discussed explicitly since for the MC scenarios studied the kinematic distributions are more sensitive to deviations. Still the search algorithm also examines the sum of all bins of a distribution, thus the total cross section, but choosing only parts of the bins usually gives better results. For first data though also the total cross section will be examined seperately in a first step of the analysis before running any search algorithm. Comparing the predicted and measured number of events per class is a very convenient and quick test for interesting deviations.

One should note that in the investigation of the kinematic distributions the MC prediction is scaled according to the assumed luminosity and cross sections. Normalizing both distributions to unity has advantages as well as disadvantages and will be studied in the future.

Chapter II.2

The Technical Setup

II.2.1 CMS software and simulated samples

The CMS software framework is used in order to process the simulated samples and to reconstruct the physics objects, using version CMSSW_1_6_8 [27]. Various reconstruction tools are utilized and will be discussed below in the context of the selection cuts. All samples are generated with the full detector simulation and originate from the MC production during the Computing, Software and Analysis Challenge 2007 (CSA07). These samples have been generated and simulated using version CMSSW_1_4_6. For the digitization, the trigger-simulation and the reconstruction version CMSSW_1_6_7 is utilized. Finally the physics objects reconstructed with a perfect CMS detector are modified assuming the 100 pb^{-1} scenario for detector alignment and calibration. This scenario corresponds to the expected detector knowledge after 100 pb^{-1} of data used within the alignment and calibration algorithms.

As a part of this computing exercise the bulk of the Standard Model backgrounds has been merged into special combined datasets, so-called "soups". The soup "Chowder" contains the W+jets, Z+jets and $t\bar{t}$ +jets backgrounds simulated with the ALPGEN [65] generator (together with PYTHIA [66] for parton showering). The lepton enriched QCD, bottomonia and charmonia events, simulated using PYTHIA, are contained in the "Stew" soup. Finally the "Gumbo" soup contains the QCD multi-jets and photons+jets backgrounds, as well as minimum bias events.

As another step of the exercise these datasets have been reprocessed and split up according to trigger requirements and additional selection criteria. This process is known as "skimming". Similar steps also have to be performed with data taken with the CMS detector: events will be classified according to the fulfilled trigger bits and will be stored within special data streams. Additionally, in order to reduce the amount of events to be analyzed within a certain analysis the data volume is shrunken within the skimming step. In a final step these samples can be accessed and analyzed via the Worldwide LHC Computing Grid (WLCG) [25] using the job submission tool CRAB [43].

The complete soups Chowder, Stew and Gumbo have been processed, using the susyHLTElectron skim (logical "OR" of various electron HLT triggers) and the susyHLTMuon skim ("OR" of various muon HLT triggers). To give an estimate of the sizes: the muon skim contains more than five million events, while the total number of events in the CSA07 exercise exceeds 100 million.

In addition, various so-called CSA07 "signal" samples have been processed in order to supplement the Soup-SM-cocktail (SM processes included in "signal" production for technical reasons):

- ALPGEN $W + N_{\text{jets}}$ with $N = 0, .., 5$ $(300 < p_{\text{T}}(W) < 1600 \,\text{GeV})$

- ALPGEN $Z + N_{\text{jets}}$ with $N = 0, .., 5$ $(300 < p_{\text{T}}(Z) < 1600 \,\text{GeV})$

- PYTHIA $Z \rightarrow 2l$ $(l = e, \mu)$ with $M_{ll} > 200 \,\text{GeV}$
  (the soup samples include a cut on $M_{ll} < 200 \,\text{GeV}$)

- PYTHIA WW, WZ and ZZ diboson samples

- SOFTSUSY [67] LM4 Supersymmetry signal samples

- PYTHIA 1st generation leptoquark $(m_{\text{LQ}} = 400 \,\text{GeV})$ signal sample

For the dominant backgrounds of many new physics signals – $t\bar{t}$, W+jets and Z+jets – a constant k-factor has been applied consistently for all jet-bins and p_{T}-bins in order to reweigh the leading order cross section to the next-to-leading order prediction (obtained from MCFM [68]). For SUSY LM4 the Prospino 2 [69] next-to-leading order cross section is used utilizing the CTEQ6M PDF's [70]. A table summarizing all simulated samples used within this analysis and the corresponding cross sections can be found in Appendix B. For the future the utilization of tools like MC@NLO instead of assuming constant k-factors would be desirable.

II.2.2 Framework and workflow of MUSiC

A dedicated framework has been developed in the context of MUSiC in order to go through the whole chain from CMSSW-RECO files, containing reconstructed objects, up to the final results, the output of the MUSiC algorithm. Special care was taken to use standalone programs independent of any experiment-specific coding while still utilizing standard CMS objects as much as possible. In this way one can benefit from collaboration wide efforts concerning physics object identification and tools, but on the other hand minimize the constant changes needed in the background of a rapidly developing CMS framework.

These requirements have been met by interfacing a dedicated tool to the CMS software: The Physics eXtension Library (PXL [71], version 1.0.1) which is a class and template collection for four-vector analysis and hypothesis evolution in High Energy Physics analysis. In addition to this the common ROOT-package [72] is also used within several steps of the analysis. Taking a look at Figure II.2.1 one can easily go through the various steps of MUSiC:

The main input to the analysis chain are the standard CMSSW-RECO files containing all relevant physics objects, e.g. muon-tracks, vertices, jet-candidates, the HLT-trigger information or the MC information for generated events (3 MB per event in total). Using a standard ED-Analyzer within CMSSW these objects are then passed on to the PXL-package, reducing the information significantly to four-vectors plus additional information (5 kB per event), thus dropping information irrelevant for high-level analyses. An example would be the four-momentum of a global muon track (tracker plus muon system fit) and additional information such as isolation variables, number of hits of the track fit, χ^2 of the track fit etc. All these variables can be easily stored using PXL-particles and a catalog called UserRecord. Thus the interface to CMSSW has to be maintained only in this filling step in case of software changes within CMSSW. Another

advantage is the small size and fast access of the pxlio-format which allows to run the analysis interactively without the complex GRID-machinery.

These pxlio-files are then processed further, again using functions already provided by PXL: Within a single event loop two "Analysis-Processes" are performed in parallel. The ControlPlot-Factory is producing hundreds of plots needed to ensure and monitor the quality and correctness of the various selected objects. These distributions are saved in ROOT-format and also converted into a ps-file for an easy and direct examination. The EventClassFactory on the other hand sorts the events into event classes according to their topologies and computes some kinematic variables ($\sum p_T$, invariant Mass, Missing Transverse Energy). Here a C++ class (TEventClass) is persistently saved within a ROOT-file, allowing access to all relevant information such as contained physics samples, assumed luminosity or histograms of the kinematic variables. The selection cuts are performed within this step from pxlio- to ROOT-files (both for control-plots and event-classification). One should note that the interpretation of the event, be it generator level or using reconstructed objects, is distinguished using so-called PXL-EventViews. In this way both hypotheses can be processed in a modular way using the same code. Also one should mention that these steps have to be done for each simulated Monte Carlo sample separately, the individual TEventClass files can then be merged into a single ROOT-file containing all information needed as input to the search algorithm.

In the final step the search algorithm is executed (MISalgo) which reads in the files containing the TEventClasses and writes the results into a derived class, TECResult. Each event class (particle topology) is investigated and all kinematic distributions are scanned. The results (numbers characterizing them and plots illustrating them) are contained again in a ROOT-file, but for fast and easy examination some selected plots and tables of numbers are written to ps-files.

It should be noted that since no major software installation is needed for a large part of the analysis all steps but the first one (ePaxAnalyzer, see Figure II.2.1) can be performed in a fast and standalone way, requiring only a laptop with some disk space and ROOT.
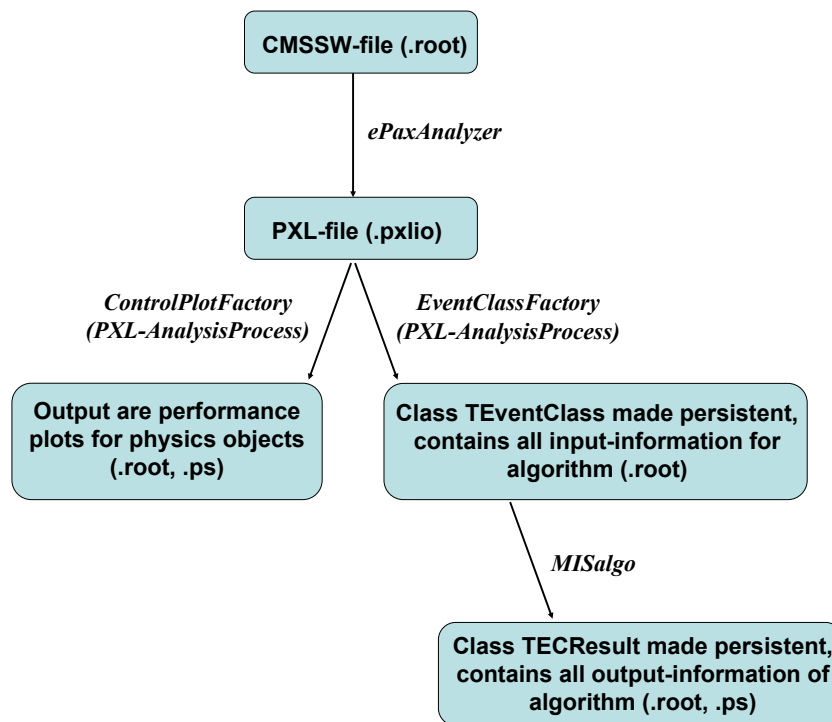
Figure II.2.1: *Framework and workflow of MUSiC, illustrating the path of the events starting from the CMS software up to the final plots of the search algorithm.*

Chapter II.3

Selection and Trigger Criteria

II.3.1 Selection criteria

The concept of a model independent search implies that selection cuts are not chosen or optimized assuming some specific signal beyond the Standard Model since this would introduce a strong bias. For MUSiC the aim of the selection cuts is to analyze standard physics objects which are well-studied and well-understood within the experiment, even if this implies some loss of statistics. As an example global muons with a relatively high p_T-threshold are used since these are expected to be well under control very early. Following this concept one does not distinguish between light quark jets and b quark jets in particular since b-tagging requires a detailed study outside the scope of this note. The strategy is to keep it simple and to focus on objects which are well-studied and understood by the numerous other analyses within CMS. In this way the estimation of efficiencies from data or selection cuts well-suited for rejecting misreconstructed objects in real data can be used within MUSiC but are taken as external inputs from dedicated studies. In this context MUSiC could serve as an additional cross check of these numbers or variables in a more general frame.

In the following the selection criteria for all physics objects are presented. The variable "reconstruction-efficiency" will appear several times. This is defined utilizing the MC-truth information of the simulated events. Therefore the generated objects (after the application of the selection and acceptance cuts like $|\eta|$, p_T or isolation) are matched to the reconstructed objects using a $\Delta R < 0.2$ criterion. The efficiency is then defined as:

$$\varepsilon_{reco} = \frac{N(\text{generated and matched to reco})}{N(\text{generated})} \qquad \text{(II.3.1)}$$

In the context of reconstruction-efficiency no trigger-requirement is used for the events. Trigger-efficiencies are studied seperately. Studies to derive efficiency estimates from data are currently under investigation within numerous other CMS studies, e.g. using the tag-and-probe method [73][74].

II.3.1.1   Muons

Global muons [22] are selected, i.e. a central track matched to the measurements in the muon system. In addition to this the standard muon isolation tool is used, more precisely the amount of energy deposited by tracks within a $R = 0.3$ cone around the muon. The following selection cuts are applied to ensure high reconstruction efficiency and low fake probability:

- $p_{\mathrm{T}}(\mu) > 30\,\mathrm{GeV}$

- $|\eta(\mu)| < 2.1$

- $R_{\mathrm{track\ isolation}} = \frac{\sum p_{\mathrm{T}}\ \mathrm{of\ tracks\ in\ 0.3\ cone\ excluding\ the\ muon\ itself}}{p_{\mathrm{T}}(\mu)} < 0.1$

- $N_{\mathrm{tracker\ hits}} > 8$

- $\chi^2/DoF/N_{\mathrm{valid\ hits}} < 1$

The chosen η-range is induced by the coverage of the muon detectors which provide the input to the L1 single muon trigger. The p_{T}-requirement of at least $30\,\mathrm{GeV}$ ensures that the muons are well above the trigger threshold (the HLT requirement for a single muon is approximately $15\,\mathrm{GeV}$). Also these muons should easily cross the iron of the return yoke, leading to track-segments within the muon system which can be combined with a track from the inner tracker. The cuts on number of hits and on the χ^2 of the muon track are designed to suppress mismeasured muon candidates which tend to have unphysically high p_T. The isolation variable helps to suppress muons originating from QCD multi-jet events. These non-prompt muons tend to be within or close to hadronic jets and are therefore more difficult to reconstruct, given the higher silicon track multiplicities. In addition high energetic particles might leak out of the calorimeter into the muon system (so called "punch-through") and cause higher muon segment multiplicities within the first muon stations.

Still it should be stressed that since the isolation criterion is quite loose one does not want to focus only on prompt muons. There is no stringent cut on the distance between the muon and the primary vertex, thus isolated muons from secondary vertices are not discarded a priori.

Figure II.3.1 shows some typical performance plots for the selected muons. The left plot displays the p_{T}-resolution (derived from simulated events) of the selected global muons as a function of the generated transverse momentum, using $t\bar{t}$ +jets, Z+jets and W+jets events from Chowder soup. The resolution is determined by fitting a Gaussian function to the variable $\frac{1/p_{\mathrm{T}}(gen) - 1/p_{\mathrm{T}}(rec)}{1/p_{\mathrm{T}}(gen)}$. This difference is proportional to the spatial resolution (sagitta measurement) of the detector which is expected to be Gaussian. An expected resolution of around $1\% - 2\%$ for global muons with $|\eta| < 2.1$ and $p_T < 200\,\mathrm{GeV}$ is obtained, comparable to the PTDR I studies [22]. The right plot presents the muon reconstruction efficiency after the application of all selection criteria. The turn-on curve has been created using a SUSY LM4 sample, matching generated muons to reconstructed global muons. An overall efficiency of more than 95% is found, almost flat in p_{T}.

II.3.1.2   Electrons

Gaussian-Sum-Filter-Pixel-Matched electrons [75] are selected. Here a supercluster (SC) measured by the electromagnetic calorimeter (ECAL) is used to find track seeds in the pixel detector.
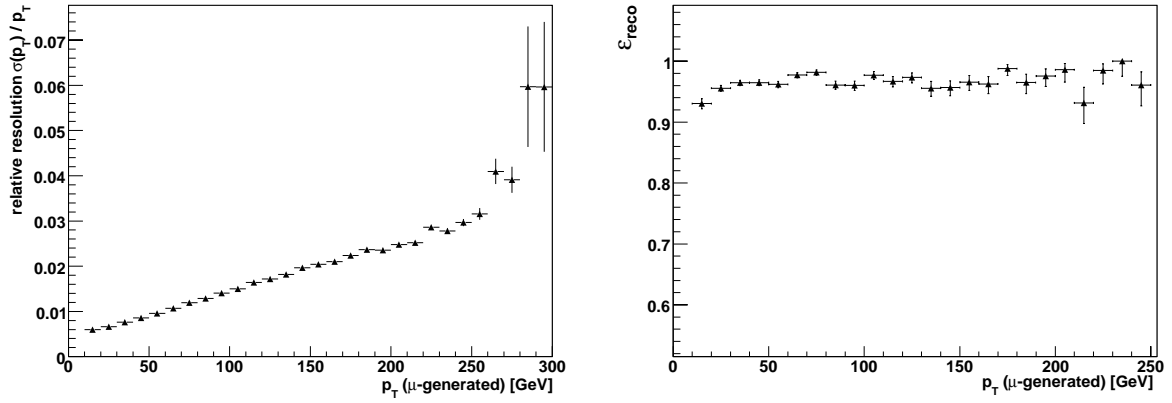
Figure II.3.1: *Left: p_T-resolution of global muons as a function of the muon momentum p_T at generator level using $t\bar{t}$ +jets, Z+jets and W+jets events from the Chowder soup. Right: Muon reconstruction efficiency as a function of the momentum p_T at generator level using the SUSY LM4 sample.*

These seeds are then used to propagate the electron through the silicon tracker. Since electrons are subjected to considerable bremsstrahlung in the detector material a Gaussian-Sum-Filter [76] is used which is a non-linear generalization of the Kalman-Filter. Here a mixture of Gaussians instead of a single Gaussian is able to account for the non-linear properties of energy loss by bremsstrahlung (Bethe-Heitler model). Finally the track parameters can be combined with the ECAL measurment. In addition to this the standard electron-gamma isolation tool and the standard "cut based" electron identification criteria are used. The following selection cuts are applied in order to ensure high reconstruction efficiency and low misidentification rates:

- $p_T(e) > 30$ GeV

- $|\eta(e)| < 2.5$

- Electron identification classification: *tight*

- $R_{\text{track isolation}} = \frac{\sum p_T \text{ of tracks in 0.3 cone}}{p_T(e)} < 0.1$

Electrons within $|\eta| < 2.5$ are expected to have the complete electromagnetic shower contained within the ECAL and the preshower detector can be used to suppress neutral pions. In addition to this electron candidates are required to have a matching inner track, thus limiting the selection to the tracker acceptance ($|\eta| \approx 2.5$). The $p_T > 30$ GeV matches to the muon requirement and also ensures that the electrons are well above the trigger threshold (HLT requirement for a single electron is approximately 15 GeV). The isolation criterion, like in the muon case, is based on tracker tracks within a cone around the electron. The electron momentum is subtracted by excluding an inner cone of $\Delta R < 0.015$ from the sum. This isolation ensures a clean electron measurement and rejects contamination from multi-jet events, e.g. jets with many $\pi^0 \to \gamma\gamma$ decays. Still, like in the muon selection, non-prompt electrons are not removed a priori. The cut-based electron identification tool uses the variables E/p, $E_{\text{had}}/E_{\text{elm}}$, the cluster shape $\sigma_{\eta\eta}$ and the matching between the track and the supercluster in η and ϕ. For the *tight* classification different cuts are applied to different electron classes, also distinguishing between electrons measured in the endcap and the barrel sub-detectors (see Table II.3.1).

Variable	Detector	Cut Value			Description
		Type 1	Type 2	Type 3	
$E_{\mathrm{had}}/E_{\mathrm{elm}}$	Barrel	< 0.042	< 0.05	< 0.045	energy ratio of the deposits within
	Endcap	< 0.037	< 0.055	< 0.050	HCAL and ECAL
$\sigma_{\eta\eta}$	Barrel	< 0.011	< 0.0125	< 0.01	shower shape variable representing
	Endcap	< 0.0252	< 0.0265	< 0.026	the lateral width in η
$|\Delta\phi_{\mathrm{in}}|$	Barrel	< 0.016	< 0.032	< 0.0525	difference between the weighted SC-ϕ
	Endcap	< 0.035	< 0.025	< 0.065	& the track-ϕ extrapolated to ECAL
$|\Delta\eta_{\mathrm{in}}|$	Barrel	< 0.0030	< 0.0055	< 0.0065	difference between the weighted SC-η
	Endcap	< 0.0055	< 0.0060	< 0.0075	& the track-η extrapolated to ECAL
$E_{\mathrm{seed}}/P_{\mathrm{in}}$	Barrel	> 0.94	> 0.24	> 0.11	ratio of the energy in the seed cluster
	Endcap	> 0.83	> 0.32	> 0.0	and the momentum of the track

Table II.3.1: *Variables and cuts used for the selection of "tight" electrons. Type 1 electrons have a f_{brem} (relative momentum change of the track between vertex and calorimeter entrance, which is proportional to the amount of bremsstrahlung emitted by the electron) of less than 6% / 10% in the barrel/endcap. If electrons exceed the previous cut but have an E/p between 0.8 and 1.2 they are classified as Type 2, otherwise as Type 3. Electrons with $f_{\mathrm{brem}} < 0.2$ and $E/p < 0.8$ are discarded as well as electrons which fulfill $E/p < 0.9 \cdot (1 - f_{\mathrm{brem}})$.*

Figure II.3.2 shows some representative performance plots of selected Gaussian-Sum-Filter-Pixel-Matched electrons. The left plot contains the p_{T}-resolution (derived from simulated events) of the selected barrel electrons ($|\eta| < 1$) as a function of the generated transverse momentum, using $t\bar{t}$ +jets, Z+jets and W+jets events from Chowder soup. The resolution is determined by fitting a Gaussian function to the variable $\frac{p_T(gen)-p_T(rec)}{p_T(gen)}$. An expected resolution of around 1.5% for electrons with $|\eta| < 1$ and $30\,\mathrm{GeV} <p_{\mathrm{T}} <100\,\mathrm{GeV}$ is obtained, comparable to the PTDR I studies [22]. The right plot presents the electron reconstruction efficiency after the application of all selection criteria. The distribution as a function of η has been created using a SUSY LM4 sample, matching generated electrons to reconstructed electrons. An overall efficiency of almost 90% is found. The transition region between the barrel and the endcap ECAL can be clearly seen as a dip at $|\eta| \approx 1.5$, the tracker acceptance is reflected in the drop of efficiency at $|\eta| > 2.4$.

II.3.1.3   Photons

Before discussing the photon selection it should be stressed that by the time this analysis was performed a standard photon ID was not yet developed. Thus an effort has been made to get the best of what was available, serving as a temporary solution.

So-called "corrected photons" [77] are used within this analysis which are reconstructed from a supercluster measured in the ECAL specifically corrected for photon objects. The standard electron-gamma isolation tool and a tool to identify photon conversions are used in order to determine if the photons are isolated or converted. Since no standardized photon identification tool existed at the time the analysis was designed, a likelihood algorithm is developed to identify photons. This algorithm is based on some old electron-ID likelihood code from a former CMS software package (ORCA). Separately for the endcap and the barrel regions, this algorithm uses various energy ratios (r_9, r_{19}, $e_{3\times3}/e_{5\times5}$ and $E_{\mathrm{had}}/E_{\mathrm{elm}}$) and shape variables ($\sigma_{\eta\eta}$, $\sigma_{\phi\phi}$ and $\sigma_{\eta\phi}$). In this context $e_{3\times3}$ is the cluster energy combining 3×3 crystals, $r_9 = e_{3\times3}/E_{Supercluster}$ and $r_{19} = E_{max}/e_{3\times3}$ where E_{max} is the crystal with the biggest energy. The decision if a photon is
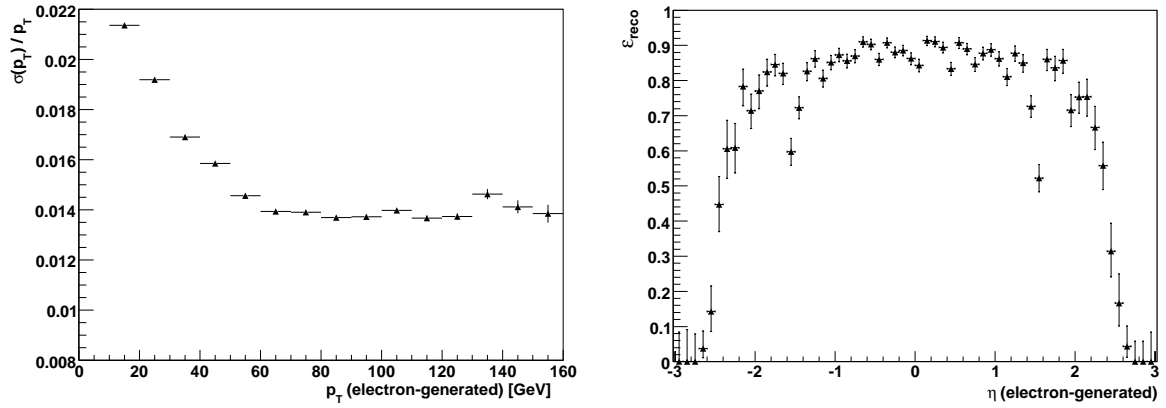
Figure II.3.2: *Left: Electron momentum resolution as a function of the momentum p_T at generator level using $t\bar{t}$ +jets, Z+jets and W+jets events from the Chowder soup, for barrel electrons with $|\eta| < 1$. Right: Electron reconstruction efficiency as a function of η at generator level using the SUSY LM4 sample.*

"good" or not (needed in order to determine the cut on the likelihood output) has been determined by matching the generated photons to the reconstructed ones. The following selection criteria are applied to ensure high reconstruction efficiencies and low misidentification rates:

- $p_T(\gamma) > 30\,\text{GeV}$

- $|\eta(\gamma)| < 2.5$

- veto on a matched pixel seed

- veto on converted photons

- $E_{\text{had}}/E_{\text{elm}} < 0.2$

- $R_{\text{track isolation}} = \frac{\sum p_T \text{ of tracks in 0.3 cone}}{p_T(\gamma)} < 0.1$

- likelihood output > 0.1

The acceptance cuts on $|\eta|$ and p_T follow the electron case. Measurements within the tracker are used to apply a track veto. The cut on the hadronic energy fraction, the veto on a pixel track seed and the veto on converted photons are chosen to ensure that only electromagnetic objects in the calorimeter without any tracks matched to them are considered, to guarantee that the selection is orthogonal to the electrons. The tool to identify photons which converted to a e^+e^- pair within the tracker is based on a dedicated tracking algorithm, identifying tracks pointing to a supercluster and reconstructing a potential conversion vertex. Since the tracker material budget is quite considerable (0.3-1.4 radiation lengths X_0 [78]) it might be desirable to include also converted photon candidates into the selection in the future. Still since these are more difficult to handle and detailed studies have only started recently the focus is on the easier scenario with unconverted photons first. Isolation is required in order to reject contamination from π^0 decays within hadronic jets. In this way photons with a considerable p_T originating from initial or final state radiation within the hard interaction or isolated photons coming from the decay of new particles (e.g. excited leptons) are selected.
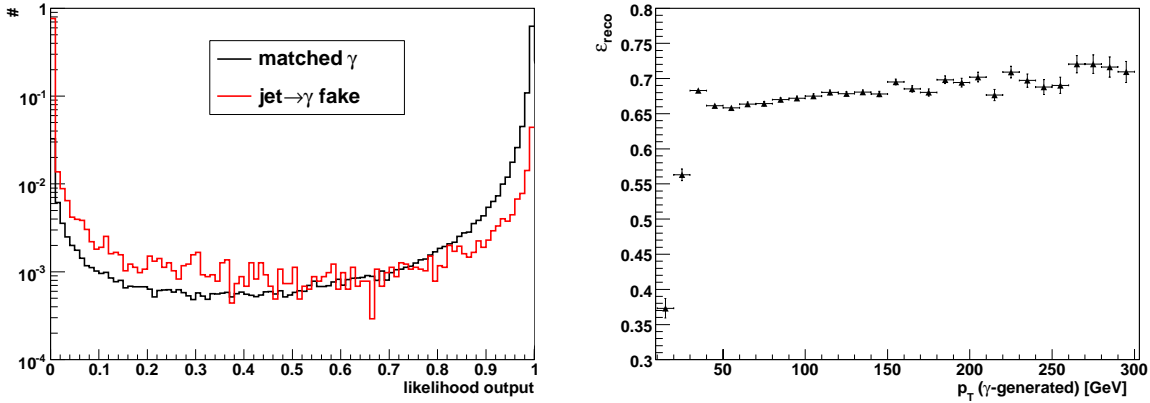
Figure II.3.3: *Left: Output of the likelihood discriminant for barrel photons ($|\eta| < 1.48$), using multi-jets and photon+jets events. Both curves are normalized to unity. Right: Photon reconstruction efficiency as a function of the momentum p_T at generator level using a photon+jets sample.*

Figure II.3.3 shows some representative performance plots of selected photon candidates. The left plot contains the output of the likelihood discriminant for barrel photons ($|\eta| < 1.48$), using multi-jets and photon+jets events (Gumbo soup). Both curves are normalized to unity. Reconstructed photons which cannot be matched with the $\Delta R < 0.2$ criterion to a generated isolated photon are labeled as "fakes". Since the sample analyzed contains QCD jet events these fakes mainly arise from jets, i.e. a hadronic jet with a high electromagnetic energy fraction. Especially jets containing a single or several π^0's can effectively fake a photon. One can see that the majority of the fake objects has a likelihood close to zero, thus they are discarded. Using the *likelihood* > 0.1 cut in total an algorithmic impurity (all events weighted with unity) of 0.99% for barrel photons and 1.3% for the endcap region is achieved. Note that since a π^0 can also decay within PYTHIA an irreducible background of matched reconstructed photons remains which originate from a π^0, requesting isolation both on generator and on reconstruction level. Only around 0.2% of the matched reconstructed photons come from a generated π^0, thus much less than the fake photons, i.e. unmatched, which enter the selection. These impurity numbers of course depend on the composition of the sample used, in this case multi-jets and photon+jets events.

A more general number is the $jet \to \gamma$ fake probability. This can be estimated analyzing 9 million multi-jets events with on average two jets per event which could fake a photon. Using this sample one obtains 2400 fake photons. This leads to $p(jet \to \gamma) = 2400/(9000000 \cdot 2) = 0.01\%$.

The right plot presents the photon reconstruction efficiency after all selection criteria have been applied. In this case the topology-requirement of at least one electron or muon in the event has been disabled. The curve as a function of p_T has been created using photon+jets events (Gumbo soup), matching generated photons to reconstructed ones. An overall efficiency of almost 70% is found. This relatively low value can be explained by the fact that only unconverted photons are selected. Given the material budget of the silicon tracker the conversion probability is between $25\% - 60\%$ (depending on η).

II.3.1.4 Jets

The "iterative cone" jet algorithm [79] with a radius of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.5$ is used. The standard L2-L3 jet energy scale (JES) corrections [79] are applied in order to have a proper estimate of the jet energy at particle level. The following selection cuts aim for a high reconstruction efficiency and they ensure that only jets whose jet energy scale can be reasonably well-understood are picked up:

- $p_T(jet) > 60\,\mathrm{GeV}$

- $|\eta(jet)| < 2.5$

- $E_{\mathrm{had}}/E_{\mathrm{tot}} > 0.05$

The acceptance cut in $|\eta|$ ensures that the whole hadronic shower is contained within the barrel and endcap of the hadronic calorimeter (HCAL). The p_{T}-threshold ensures that the energy resolution of the hadronic jets is reasonable ($\approx 20\%$). Thus jet energy scale corrections should be well under control. A certain amount of hadronic energy is required in order to separate jets from electromagnetic objects in the calorimeter such as electrons or photons. For jets with a considerably large electromagnetic energy fraction standard jet energy scale corrections should not be applied, thus a restricttion to hadronic jets, which are easier to handle, is performed.

The left plot of Figure II.3.4 presents the jet reconstruction efficiency after the application of all selection criteria. In this case the topology-requirement of at least one electron or muon in the event has been dropped. The distribution as a function of the momentum p_{T} has been created using multi-jets and photon+jets events (Gumbo soup), matching iterative cone jets from all stable generator particles (except neutrinos/neutralinos) to hadronic jets at reconstruction level. A reconstruction efficiency close to 100% is found.
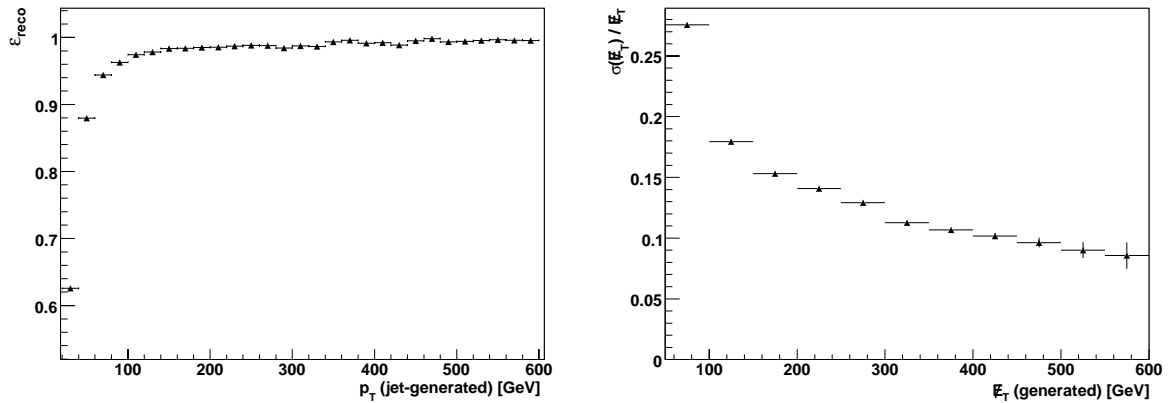

Figure II.3.4: *Left: Jet reconstruction efficiency for jets which have been reconstructed via the iterative cone algorithm with $\Delta R = 0.5$ as a function of $p_{\mathrm{T}}(gen)$ using a multi-jet and photon+jets sample. Right: $\not{E}_{\mathrm{T}}$-resolution as a function of the generated missing transverse energy, using ALPGEN W+jets events.*

II.3.1.5　Missing transverse energy

Since any error of the measurement of the physics objects discussed above propagates into the determination of $\not{E}_\mathrm{T}$, a relatively high threshold of

- $\not{E}_\mathrm{T} > 100\,\mathrm{GeV}$

is used. Standard corrections [80] are applied to the $\not{E}_\mathrm{T}$ object, accounting for jet energy scale corrections of the jets in the event, and subtracting muon momenta from the calorimeter based $\not{E}_\mathrm{T}$ estimate.

The right plot of Figure II.3.4 presents the relative $\not{E}_\mathrm{T}$-resolution (derived from simulated events) as a function of the generated missing transverse energy, using the ALPGEN $W+$jets events without requiring a selected lepton. The $\not{E}_\mathrm{T}$ at generator level is defined by adding up all stable particles within the calorimeter acceptance, excluding neutrinos and particles beyond the SM (e.g. neutralinos in case of SUSY). An expected resolution of just below 10% for $\not{E}_\mathrm{T}$ above 400 GeV is obtained, comparable to [80], which reflects the finite calorimeter resolution.

II.3.1.6　Suppression of instrumental backgrounds

While it is clear that with the arrival of first data various "cleaning" steps will be needed to select "good" runs without detector problems, also at the level of physics object reconstruction additional criteria are needed in order to minimize instrumental background from "fakes". This cleaning mainly refers to the removal of duplicate objects and the ambiguous interpretation of objects in the detector. For example a supercluster can be interpreted as an electron as well as a photon. The listed cleaning steps are carried out in the following sequence:

- Muon candidates which are closer than $\Delta R < 0.2$ to each other are cleaned, keeping only the one measured best (smaller χ^2/DoF). The cut is designed to remove ghost muons and other sources of duplicate muons.

- Electron candidates which are closer than $\Delta R < 0.2$ to each other, and which share either the inner track or the supercluster seed are cleaned, keeping only the more energetic one.

- Photon candidates which are closer than $\Delta R < 0.2$ to each other and which share the supercluster seed are cleaned, keeping only the more energetic one. Also photon candidates closer than $\Delta R < 0.2$ to an already selected electron are removed if the photon has the same supercluster seed as the electron. This should remove the ambiguity imposed by the fact that all superclusters can be interpreted as electrons as well as photons. Thus well-measured electrons receive a higher priority than photons.

- Jet candidates closer than $\Delta R < 0.2$ to an already selected electron or photon are removed to avoid an overlap of those collections.

So far no e/μ separation cut is included but could be added in the future.

II.3.2 High Level Trigger

The choice of the trigger menu used in this analysis is driven mainly by the requirement to combine triggers with a prescale factor of 1 (or at least using the same L1-definition) and HLT triggers which are expected to be "standard" at the LHC start-up and which are therefore commonly used and well-understood.

An "OR" of various single lepton and di-lepton HLT-triggers [30] is used, see Table II.3.2 for details on the various trigger requirements. These are combined into a muon-stream on one hand and an electron-stream on the other:

- single muon (with isolation) OR di-muon HLT (with and without isolation)

  [HLT1MuonIso || CandHLT2MuonIso || HLT2MuonNonIso]

- single electron OR di-electron HLT (both with and without isolation)

  [HLT1Electron || HLT1ElectronRelaxed || HLT2Electron || HLT2ElectronRelaxed || HLT1EMHighEt || HLT1EMVeryHighEt]

HLT Path	Technical Name	HLT Requirements	Rate (Hz)
Single Isolated μ	HLT1MuonIso	Input: L1 Muon with $p_\text{T} > 7\,\text{GeV}$ Threshold: $p_\text{T} > 11\,\text{GeV}$	18.3 ± 2.2
Double Isolated μ	CandHLT2MuonIso	Input: L1 Muons with $p_\text{T} > 3\,\text{GeV}$ Threshold: $p_\text{T} > 3\,\text{GeV}$ for both muons calorimeter and track isolation	—
Double Relaxed μ	HLT2MuonNonIso	Input: L1 Muons with $p_\text{T} > 3\,\text{GeV}$ Threshold: $p_\text{T} > 3\,\text{GeV}$ for both muons	12.3 ± 1.6
Single Isolated e	HLT1Electron	Input: L1 isolated e with $E_\text{T} > 12\,\text{GeV}$ Threshold: $E_\text{T} > 15\,\text{GeV}$ track isolation	17.1 ± 2.3
Single Relaxed e	HLT1ElectronRelaxed	Input: L1 e with $E_\text{T} > 15\,\text{GeV}$ Threshold: $E_\text{T} > 17\,\text{GeV}$	9.6 ± 1.3
Double Isolated e	HLT2Electron	Input: L1 isolated e's with $E_\text{T} > 8\,\text{GeV}$ Threshold: $E_\text{T} > 10\,\text{GeV}$ HCAL & Track isolation	0.2 ± 0.1
Double Relaxed e	HLT2ElectronRelaxed	Input: L1 e's with $E_\text{T} > 10\,\text{GeV}$ Threshold: $E_\text{T} > 12\,\text{GeV}$	0.8 ± 0.1
Single High E_T e/γ	HLT1EMHighEt	Input: L1 e's with $E_\text{T} > 15\,\text{GeV}$ Threshold: $E_\text{T} > 80\,\text{GeV}$ ECAL/HCAL/track isolation	0.5 ± 0.0
Single Very High E_T e/γ	HLT1EMVeryHighEt	Input: L1 e's with $E_\text{T} > 15\,\text{GeV}$ Threshold: $E_\text{T} > 200\,\text{GeV}$ No track and isolation requirement	0.1 ± 0.0

Table II.3.2: *Details on the High Level Triggers used within this analysis. The rates are estimated from the detector simulation within the "HLT exercise" [81] and assume $\mathcal{L} = 10^{32}\ \text{cm}^{-2}\text{s}^{-1}$.*
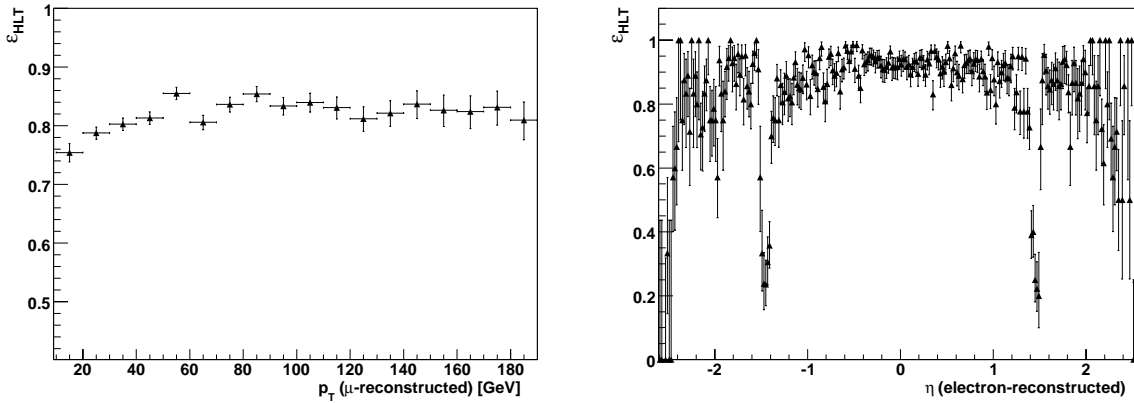
Figure II.3.5: *Left: HLT trigger efficiency using "OR" of all used muon triggers as a function of $p_T(rec)$ using the LM4 sample, with respect to selected events.   Right: HLT trigger efficiency using "OR" of all used electron triggers as a function of $\eta(rec)$ using the LM4 sample, with respect to selected events.*

After the application of all selection criteria, the muon and electron streams are merged into a single dataset, avoiding double counting in events where both electron and muon triggers have fired.

One representative efficiency plot for each data steam is chosen. The left plot of Figure II.3.5 shows the HLT trigger efficiency using an "OR" of all used muon trigger bits. Events from the SUSY LM4 sample have been used, and the trigger efficiency has been determined after all selection cuts have been applied, including the topology requirement of at least a single reconstructed muon. The efficiency curve as a function of p_T shows an overall level of well above 80%. Note that unfortunately the single muon trigger without isolation has not been included in the SUSY-HLT-skim used for the CSA07 samples. Including this trigger in future skims would increase the overall efficiency to about 90%. The right plot of Figure II.3.5 shows the HLT trigger efficiency using an "OR" of all used electron trigger bits, again using the SUSY LM4 sample and with respect to selected events. The efficiency curve as a function of η shows an overall level of almost 90%. The transition region between barrel and endcap ECAL can be seen clearly.

In addition to the specific turn-on curves also a single number for the global HLT trigger efficiency using an "OR" of all the triggers mentioned above is of interest. One obtains, again using the SUSY LM4 sample and with respect to selected events (statistical errors only):

- Muon Stream: $\varepsilon_{\mathrm{HLT}} = 82 \pm 0.4\%$

- Electron Stream: $\varepsilon_{\mathrm{HLT}} = 88 \pm 0.3\%.$

Chapter II.4

The Search Algorithm

At this point of the analysis, the events have been processed and physics objects satisfying the criteria mentioned above are identified. The composition of the event, i.e. the number of muons, jets etc. determines to which event class it is assigned. At the present time three distributions are investigated for each event class, thus limiting the number of distributions to be looked at and focusing on distributions which seem to be promising for spotting new physics:

- **Scalar sum of the transverse momentum** $\sum p_\mathrm{T}$ of all physics objects.
  For example for the class 1μ 2jet $\not{E}_\mathrm{T} + X$ one calculates
  $\sum p_\mathrm{T} = p_\mathrm{T}(\mu) + p_\mathrm{T}(\mathrm{jet}_1) + p_\mathrm{T}(\mathrm{jet}_2) + \not{E}_\mathrm{T}$.

- **Invariant mass** M_inv of all physics objects. For classes with $\not{E}_\mathrm{T}$ the transverse mass M_T is investigated.

- For classes with **missing transverse energy** the $\not{E}_\mathrm{T}$ variable is investigated separately.

The $\sum p_\mathrm{T}$ distribution is the most general quantity to be checked. The invariant mass has an obvious advantage for new particles produced as resonances like new heavy gauge bosons. Since many models beyond the SM aim at providing a candidate particle explaining dark matter in the universe, this particle would lead to a considerable amount of $\not{E}_\mathrm{T}$ in the event. An obvious example would be the lightest supersymmetric particle in SUSY extensions: here $\not{E}_\mathrm{T}$ is known to be an excellent variable for separating SUSY events from the SM background. Still this quantity will be hard to control and to understand at the beginning of data taking and thus the use of this quantity might be challenging. While in $\sum p_\mathrm{T}$ a model independent scan of all classes will be performed with the first data, this is not clear at the moment for M_inv and $\not{E}_\mathrm{T}$. Also in principle the implementation of even further variables can be done easily, if desired.

All distributions are input to the MUSiC algorithm (similar to the H1 analysis [61]) which scans them systematically for deviations, comparing the Standard Model expectation (Monte Carlo prediction) with the measured data. A bin width of $50\,\mathrm{GeV}$ for all distributions ($\sum p_\mathrm{T}$, M_inv and $\not{E}_\mathrm{T}$) is chosen in order to absorb detector resolution effects. This binning can be changed at any given time if desired, e.g. for classes with low statistics in the high-p_T tails. Still in principle by combining multiple bins to connected regions a wider binning is examined within the algorithm intrinsically. Having narrower bins has the advantage of being sensitive to narrower deviations, but also the disadvantage of a larger statistical penalty factor to be applied.

II.4.1   Steps of the algorithm and probability definition

First part

Each connected bin region is considered within the distributions, i.e. single bins (bin 10 or bin 200) as well as broad regions (bin $3 - 100$ or bin $150 - 155$). See also Figure II.4.1 for illustration. It is not considered meaningful to combine unconnected bins, e.g. combine bin 20 and bin 100 and bin 114 into one region.
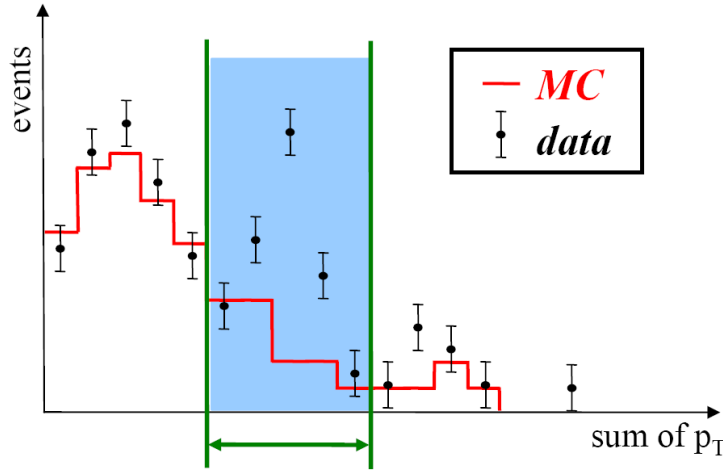

Figure II.4.1: *Illustration of a connected bin region within a kinematic distribution.*

For each connected region, a counting experiment is performed, adding up the various expected Monte Carlo contributions (N_{SM}) and comparing this sum to the amount of measured data (N_{data}). In addition to these two numbers also the uncertainty of the prediction $\delta(N_{\text{SM}})$ is used, i.e. the systematic and statistical uncertainties of simulated events contributing to this specific region. Then a Poisson probability is computed, determining how likely the prediction fluctuates up to or above the number of events seen in the data. The systematic uncertainties, taking correlations into account, are included using a convolution with a Gaussian:

$$
p = \begin{cases}
\displaystyle\sum_{i=N_{\text{data}}}^{\infty} A \cdot \int_{0}^{\infty} \mathrm{d}b \, \exp\left(\frac{-(b - N_{\text{SM}})^2}{2(\delta N_{\text{SM}})^2}\right) \cdot \frac{e^{-b}\, b^i}{i!} & \text{if} \quad N_{\text{data}} \geq N_{\text{SM}} \\[2ex]
\displaystyle\sum_{i=0}^{N_{\text{data}}} A \cdot \int_{0}^{\infty} \mathrm{d}b \, \exp\left(\frac{-(b - N_{\text{SM}})^2}{2(\delta N_{\text{SM}})^2}\right) \cdot \frac{e^{-b}\, b^i}{i!} & \text{if} \quad N_{\text{data}} < N_{\text{SM}}
\end{cases}
\quad , \qquad \text{(II.4.1)}
$$

where A ensures the normalization. From all the possible combinations of connected bins for one distribution, the region with the smallest p-value ($p_{\text{min}}^{\text{data}}$) is chosen. This is the place in the distribution where the biggest discrepancy between data and Monte Carlo prediction is found. It is called the **Region of Interest** for this distribution.

One should emphasize that this definition of p represents a Bayesian-frequentist hybrid since the true value of the background corresponds to one of the b's in the Gaussian integration. As

pointed out in [82] correct frequentist coverage cannot be guaranteed for all the parameter space and comparisons with other classes of algorithms are desirable for the future.

This simple but effective approach is sensitive to an excess of data as well as a deficit; it can detect large single bin fluctuations as well as possible signals spread over a large part of the distribution.

One should stress the importance of including the error on the estimate of the Monte Carlo simulation into the probability definition. In this way the p-value gives the probability for the background to fluctuate up to the data and further, given the intrinsic uncertainties of the MC estimate. One can assign large errors to the value N_{SM} if the Monte Carlo events are expected to not describe the data well in a specific part of the phase space. If one expects 100% uncertainty in some exotic final state where one does not trust the Monte Carlo, some new physics signal might nevertheless lead to discrepancies far exceeding this large uncertainty and thus interesting physics might be revealed.

The different error contributions are assumed to be uncorrelated (e.g. luminosity uncertainty and jet energy scale uncertainty) and are thus added in quadrature. Correlations within a single error type, e.g. the luminosity uncertainty between simulated samples, are carefully included in the error estimate. The individual contributions will be discussed in detail in Section II.5.1. Ultimately the total systematic uncertainty is expressed as:

$$\delta N_{\mathrm{SM}} = \sqrt{\sigma_{\mathrm{stat}}^2 + \sum_i \sigma_{\mathrm{i,syst}}^2} \, , \tag{II.4.2}$$

where σ_{stat} represents the statistical error given the limited MC-statistics of the various samples used.

Second part

It is important to understand that the statistical estimator p alone is not sufficient to claim any evidence for a signal. A statistical penalty factor has to be applied to account for the large number of regions (connected bin combinations) investigated. This is done in the second step of the algorithm, determining the **event class significance** (per distribution) of the deviation found in the first step:

Toy Monte Carlo experiments are performed, assuming the background-only hypothesis. Therefore hypothetical data histograms (HDH) are created numerous times by varying the Monte Carlo prediction for each bin according to its statistical and systematic uncertainty. Again correlations within single error contributions have to be accounted for when creating the pseudo data. These hypothetical data are then fed again into the first step of the algorithm and compared to the Monte Carlo mean (results in $p_{\mathrm{min}}^{\mathrm{SM}}$). Again all possible connected regions are examined, not only the Region of Interest from the initial step 1. The event class significance of the deviation is defined as:

$$\widetilde{P} = \frac{\text{number of HDH with } p_{\mathrm{min}}^{\mathrm{SM}} \leq p_{\mathrm{min}}^{\mathrm{data}}}{\text{total number of HDH}} \, . \tag{II.4.3}$$

The value of $\widetilde{P}$ is the fraction of background-only toy experiments where a deviation even bigger than the one observed in the data is found, see also Figure II.4.2 for illustration.
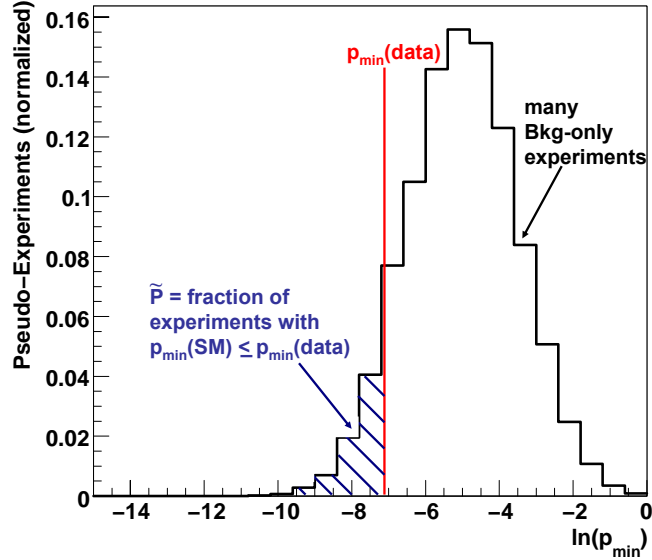

Figure II.4.2: *Definition of $\widetilde{P}$ when repeating the background-only hypothesis many times and comparing this to the observed deviation $p_{\mathrm{min}}^{\mathrm{data}}$.*

Performing these pseudo-experiments one jitters the Standard Model expectations and tests for signal-like fluctuations of the Standard Model. These fluctuation may appear in all regions considered within the algorithm, not only in the Region of Interest. The $\widetilde{P}$ can directly be translated into standard deviations (see Figure II.4.3) and is comparable to the widely used CL_{b} [83]. Since MUSiC is sensitive to an excess of data as well as a deficit a two-sided Gaussian is used for this translation.
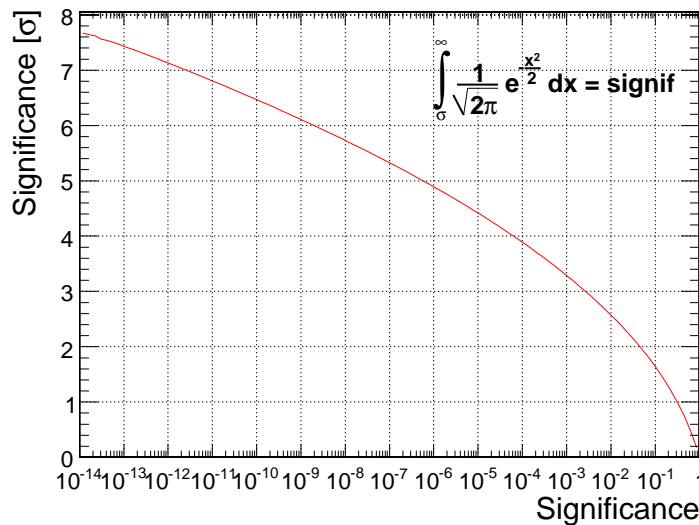

Figure II.4.3: *Translation of significance ($\widetilde{P}$) into number of standard deviations σ [84].*

II.4.2 Sensitivity study with simulated events

Since the LHC has not started yet it is clear that one has no N_{data} to compare with the Monte Carlo prediction. Still one can pick exemplary models beyond SM and test the sensitivity of MUSiC with them. Instead of only producing pseudo-data for the background-only hypothesis in step 2 one can also create toy data as input to step 1 assuming *signal + background*, i.e. add a signal distribution on top of the SM ones. In this way one can repeat several pseudo-CMS experiments and determine the expected event class significance of a possible signal present in the data. Figure II.4.4 illustrates this procedure, using the event class $1e\,5\text{jet} + X$ as an example: The green curve represents the pseudo-experiments where signal (SUSY LM4) plus background are assumed. With data this would correspond to a single line. The red curve on the other hand displays the multiple repetition of the SM expectation including its errors, thus this represents step 2 of the algorithm. The p and $\widetilde{P}$ values stated in the plots refer to the median of the left curve and then integrating the red curve beyond this median p_{min}. The interpretation of the two curves is clear: In the case that they are well-separated, $\widetilde{P}$ will be quite low and discovery is easy, as shown in the left plot where no systematic errors are assumed. But if one includes systematic uncertainties in the algorithm, the two hypotheses move close to each other and less than a 3σ deviation ($\approx 10^{-3}$) remains. This also underlines the importance of implementing systematic uncertainties into MUSiC which will be discussed below in Section II.5.1.
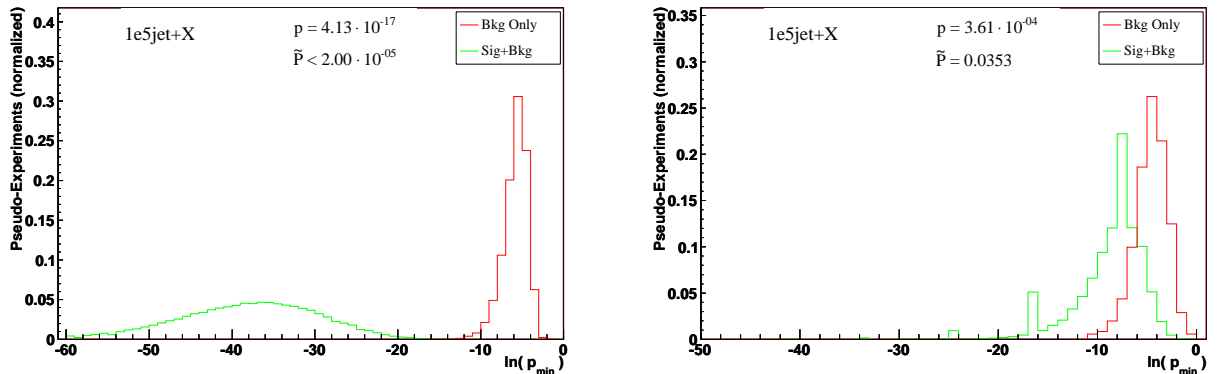

Figure II.4.4: *Signal plus background and background-only hypotheses for an LM4 event class, on the left without systematic uncertainties and on the right with all uncertainties included. The striking difference between both plots shows the importance of systematic uncertainties.*

II.4.3 Producing pseudo-data

Testing the significance of the deviation is done by dicing hypothetical data histograms, i.e. one changes the mean value N_{SM} slightly to reflect the inherent statistical and systematic errors. Of course, the assumption that these errors are both well-understood and realistic is crucial for this procedure. On the other hand huge deviations found in numerous event classes of the first LHC data could indicate that some uncertainties have been underestimated and/or additional errors have to be included.

In order to be able to separate deviations caused by new physics from background-only fluctuations the errors have to be included in a way similar to a real measurement of the CMS detector.

Thus correlations between bins and simulated samples are important. These will be discussed in detail in Section II.5.1, but some general comments on the implementation are done here as well.

The basis for the dicing is the probability density function in Equation (II.4.1), even though the actual dicing process is divided into several steps with respect to all error contributions. It is essential that contributions which are *statistically independent* can be *decoupled* and diced separately. There are three main dicing-contributions for each hypothetical data histogram:

- the **Poisson probability** to account for the statistical error of the actual measurement.

- the assumed **systematic errors** as part of the Gaussian convolution

- the **statistical errors** of the MC datasets as part of the Gaussian convolution

An example for a systematic uncertainty could be the error on the luminosity estimate. Assuming a 5% error all bins and all simulated samples are correlated with respect to this error. Thus for each set i of pseudo data a single number $\sigma_i(\text{lumi})$ is generated assuming a Gaussian with a mean of $\mu = 0$ and a width of $\sigma = 0.05$. This number could be -3% in one pseudo-experiment and $+10\%$ in another one, and all bins of all samples are scaled with an according factor. Thus magnitude and direction of the error is preserved for all bins and all MC-contributions. Similar considerations can also be made for the other systematic uncertainty contributions.

Figure II.4.5 serves as a control plot for the dicing procedure. Here the sum of the SM background is shown together with the total systematic uncertainty (shaded area). The data points correspond to the mean after many repetitions of the background-only hypothesis. The error bars correspond to the width of the variation for the many pseudo-data sets. One can nicely see that, as expected, the data points match the mean expectation of the Monte Carlo and the error bars reflect the total uncertainty estimate.
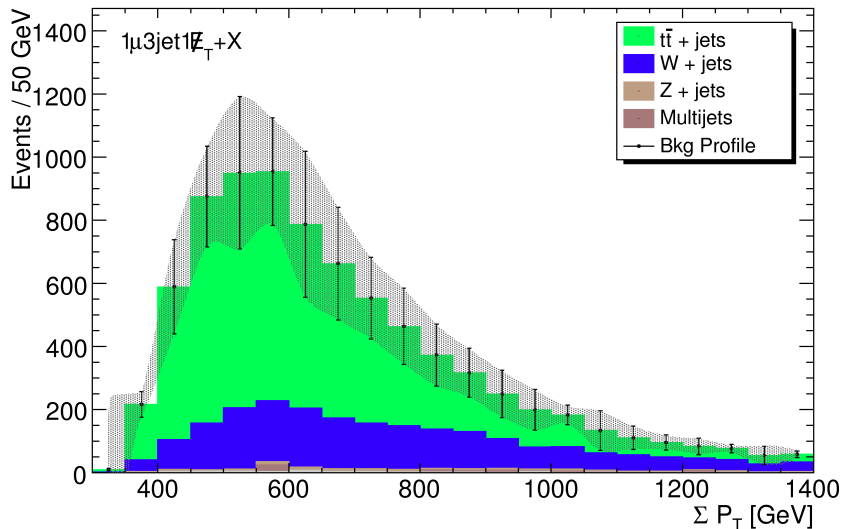

Figure II.4.5: *Sum of SM backgrounds and assumed systematic uncertainties (shaded area) in comparison with the distribution of the numerous pseudo-data sets. Data points corresponds to the mean of these sets, the error bars to the width of the variation.*

II.4.4   Discussion of methods for calculating the statistical significance

As already mentioned when introducing the definition of the statistical estimator p in Equation (II.4.1) the validity of the resulting value to some extend depends on the numbers and the problem in question. A detailed discussion of the properties of such a Bayesian-frequentist hybrid as well as comparisons to other methods can be found in [82]. While it is clear that including systematic uncertainties into hypothesis tests has been studied extensively by the professional statistics community, the reader should be aware of the possible shortcomings of the statistical estimator and that a satisfactory solution to all problems imaginable is almost impossible.

Still in the context of MUSiC the focus is not to give very precise significances but to act as an alarm system detecting interesting deviations. Since the number of data events, expected Monte Carlo events and its corresponding uncertainty is always known and stated, cross-checks using alternative statistical methods are possible and desired.

In [82] an estimator called Z_{Bi} is promoted since it is based purely on frequentist assumptions and shows good performance in many cases. Equation (II.4.1), called Z_N in this context, gives reasonable results for a mean background expectation and a Gaussian systematic uncertainty. Of course approximating the uncertainties by a Gaussian is a strong assumption which may not be true in all cases. Thus this should be understood as a pragmatic solution and the reader should be aware of possible deficits due to non-Gaussian tails. Figure II.4.6 shows a comparison of the p-values as determined by Z_N and by Z_{Bi}. Here a number of exclusive event classes have been used, assuming a SUSY LM4 signal present in the pseudo-data. One can clearly see a correlation between the results of both statistical methods, but also that a unique answer cannot be given in all cases.
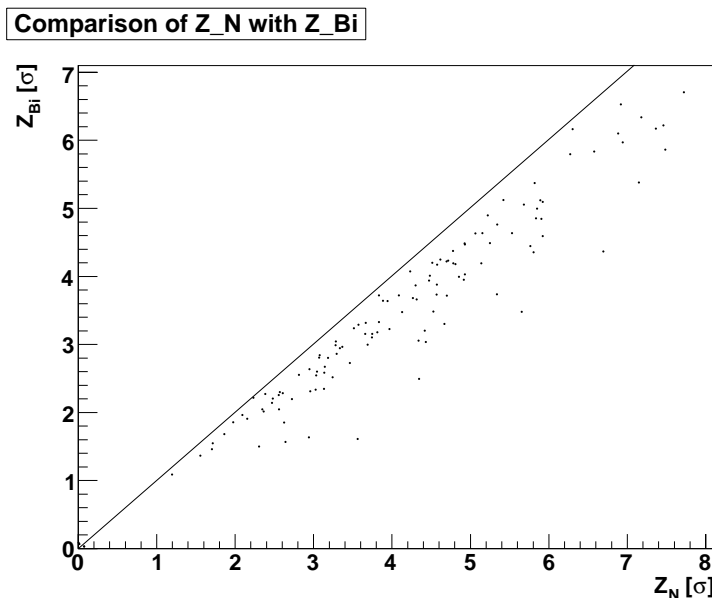

Figure II.4.6: *Comparison of p-values as computed by two distinct statistical methods [84].*

Chapter II.5

Systematic Uncertainties and Data-driven Methods

In this chapter the major contributions of systematic uncertainties which are included in the MUSiC search algorithm are discussed. Separately also the QCD multi-jet background which can be regarded as an additional source of uncertainty is studied. Since a reliable prediction using Monte Carlo simulation is almost impossible a first attempt to use a data-driven background estimate in the context of MUSiC is presented. In the future uncertainties of this estimate from data will of course also be included as a contribution to the systematic errors used in the algorithm.

II.5.1 Systematic uncertainties

As mentioned in the previous chapter, it is crucial to implement correct systematic uncertainty estimates in the algorithm in order to distinguish a true signal from a "fake" deviation caused by an unanticipated detector effect or an incorrect theoretical estimation of the Standard Model expectation. The following systematic uncertainties are assumed and included in MUSiC. Their magnitude is estimated in the context of 1 fb^{-1} of data, but the values can be adapted very easily:

- σ(integrated luminosity) = 5%

- σ(cross sections) = 10%

- σ(jet energy scale) = 5%, energy variation in jets propagated also into $\not{E}_{\mathrm{T}}$ estimate

- σ(efficiency correction factor) = 2% for e, μ, γ and 1% for jets

- σ(fake probability) = 100% for e, μ, γ

- statistical uncertainty of the Monte Carlo prediction, based on the amount of originally produced events per sample

It is important to stress that the algorithm also accounts for correlations within one error in the context of systematics. For global factors like cross sections all bins in a distribution and

the different sub-samples (jet multiplicity bins or p_T bins) are correlated. For the integrated luminosity even all physics processes are correlated. In addition to this, variations are not always just "up or down", the JES uncertainty actually redistributes the bins and is again correlated for all generated samples. These correlations have to be taken into account when computing p-values for a certain region and when generating pseudo-data for the whole distribution.

Luminosity

A luminosity error of 5% should be realistic assuming 1 fb^{-1} of data [17]. At this stage the LHC machine parameters should be well-known and the luminosity monitors should operate smoothly. In addition to this the W- and Z-peak can be used as "standard candles" to determine the luminosity assuming some theoretical predictions of the W and Z cross sections.

Cross section

The assumed cross section uncertainty of 10% is used for all Standard Model background processes, not distinguishing between different jet multiplicity bins or p_T bins of the generated samples. Studying the expected PDF uncertainties using the PDF reweighting method (see [85]) one can see that these typically range at the order of 2% to 8%, see [86] for details. The discussion of this and other theoretical uncertainties is still in flux, and the understanding of these numbers is likely to change in the future; therefore, it has been decided to use a single "conservative" number. Note that for some processes the uncertainties might be even higher and it is possible within MUSiC to specify the cross section uncertainty for each process individually. Thus the 10% only reflects the current understanding and might well be refined in the future.

Jet energy scale

The 5% uncertainty on the jet energy scale is taken from evaluations done by calorimeter and jet-reconstruction experts [79]. Within 1 fb^{-1} of data both MC based calibration techniques and data-driven methods can be used and compared to each other. In this way the simulation can be tuned to match the data, resulting in reliable jet energy scale corrections.

As mentioned above this error actually redistributes the bins and cannot be determined directly from the MC mean. All distributions analyzed with MUSiC are created also in an up-down variation, scaling all selected jets according to the assumed 5% error. The changes in the 4-vectors of the jets are summed up and the residual variation is vectorially subtracted from the $\slashed{E}_T$ estimate. In order to compute a p-value according to Equation II.4.1 the error has to be symmetrized. The direction of the error is taken from the up-variation and thus preserved for all bins, the value is symmetrized by averaging the up- and down-variation for each bin.

In principle also the muon energy scale and the electron energy scale have an uncertainty. Still these should be small compared to the JES, but the implementation within MUSiC is easy following the JES example.

Efficiency correction factor

For electrons, muons and photons efficiency correction factors are included into the MUSiC algorithm. These account for possible efficiency differences between data and Monte Carlo. Using data-driven methods (e.g. tag-and-probe technique on $Z \rightarrow \mu\mu$ events) reconstruction efficiencies can be measured and compared to the Monte Carlo estimate. This will result in correction factors to be applied to the simulated events. These correction factors are the result of careful and complex studies done by the various physics object groups (POG), see also recent studies for muons [73] and for electrons [74]. It should be stressed that MUSiC depends on the input of these numbers from the different groups. That is the reason why standard objects such as global muons or pixel-matched-electrons with standard identification cuts are used. In this way a duplication of work is avoided, synergy effects can be exploited and the scope of MUSiC remains feasible. On the other hand MUSiC can give feedback and spot possible limitations utilizing these numbers in a more broad context.

So far the correction factors are implemented as a function of p_{T} and η, with dummy values of unity until first data arrive. One gets for the original bin entry N_i

$$N_i' = N_i \cdot f_e^2 \cdot f_{\mathrm{jet}} \qquad \text{for the 2}e \text{ 1jet } (+X) \text{ event class.} \qquad \text{(II.5.1)}$$

Of course these Monte Carlo correction factors f_e and f_{jet} are only known up to a certain precision. For muons, electrons and photons a constant error of 2% is assumed for the correction factor. Since the jet reconstruction efficiency is close to 100% anyway and since QCD events will be available with almost unlimited statistics only a 1% error is assumed for jets. The error can be computed using simple error propagation on Equation II.5.1, respecting that all bins and all physics processes are correlated. For $\not{E}_{\mathrm{T}}$ no efficiency correction is planned. Here differences between data and MC are likely to be caused by resolution effects and thus an offline-smearing of the MC objects could be performed.

One should note that in the context of reconstruction efficiencies at first approximation MUSiC assumes them to be independent of the number of particles in the event. Of course for very complex particle topologies this might not be true, still it is hard to solve special issues and problems like this in a generic way for all event classes. This relates to the introductory remarks of the second part of this thesis: A deviation found by MUSiC has to be investigated in the following with more dedicated checks. MUSiC acts more like a warning system which cannot account for all details. Questions like the efficiencies within complex final states have to be addressed when investigating the deviation(s) found by MUSiC.

Fake probabilities

The estimation of fake probabilities for the reconstructed objects using first data will probably be a challenging task. Also huge differences with respect to the misreconstruction probabilities predicted by the detector simulation should not be surprising. In principle one could perform similar MC corrections as in the case of reconstruction efficiencies. Still, it is not clear which level of detector understanding and MC tuning is needed before this seems realistic. First studies using data-driven techniques have already started within the collaboration. In the scope of this thesis

and as a first attempt to implement this uncertainty it has been decided to rely on the MC-truth-knowledge for the moment. A reconstructed object not matching within a $\Delta R < 0.2$ criterion to a generated isolated particle is labeled as "fake". For jets and $\not{E}_{\mathrm{T}}$ the dominant uncertainty is already covered with the jet energy scale, thus fake errors are only assumed for muons, electrons and photons. As a conservative "guess" of the error on all fake probabilities 100% uncertainty is assumed. Thus for each event processed the number of fake objects is counted and an event weight for the "up"-variation is calculated:

$$weight_{\mathrm{fake}}^{\mathrm{up}} = 1 + \sqrt{(N_{\mathrm{f}}(e) \cdot \sigma_{\mathrm{f}}(e))^2 + (N_{\mathrm{f}}(\mu) \cdot \sigma_{\mathrm{f}}(\mu))^2 + (N_{\mathrm{f}}(\gamma) \cdot \sigma_{\mathrm{f}}(\gamma))^2} \quad , \qquad \text{(II.5.2)}$$

where $N_{\mathrm{f}}(e)$ denotes the number of fake electrons in this specific event and $\sigma_{\mathrm{f}}(e)$ the relative error of the fake probability. This results in an additional distribution where the fake probabilities are varied by one sigma. Again the differences between this distribution and the mean MC values can be computed and used for the algorithm. Since misreconstruction is an overall detector effect, all bins and all physics samples are correlated.

Smearing corrections

Once first data have arrived the tuning of the detector simulation will start. In this context resolution differences between data and MC are likely to appear which would demand to further smear reconstructed objects in the simulation. Also the widths of these smearing functions are known only up to a certain precision. Thus one might consider varying this width by one sigma and further smear the MC. This would give an error estimate on the effect of these smearing steps performed in the simulation.

The implementation of these errors is very similar to jet energy scale variations and parts of the infrastructure could be re-used. Since without data these smearing corrections are not needed this systematic uncertainty is not included for the time being. Still it should be straightforward to include them in MUSiC in future versions.

Non-collision backgrounds

Since MUSiC is analyzing the event content of pp-collisions the contributions of other sources of particles can be regarded as another systematic uncertainty. While the effect of pile-up is expected to be small at initial luminosities, contributions from beam halo and cosmic muons are always existent. Both sources could cause deviations between data and SM simulation, especially in "exotic" channels with very high particle multiplicities. Still both sources are also expected to be relatively rare: The cosmic muons as well as the beam halo particles are asynchronous with respect to the hard interaction. Also, their passage through CMS is quite different from the particles originating from the vertex. Additionally, CMS is 90 m underground, such that the rate of cosmic muons arriving at CMS is only O(100 Hz). Thus, these events are unlikely to fire a trigger. When overlaid to a pp-triggered event the differences in timing and direction can hardly lead to high-quality reconstructed objects with central η and high-p_{T}.

Still they are an irreducible background which can affect data-MC comparisons. Luckily for both sources of particles dedicated Monte Carlo generators exist (see [28, 32] and part I of this thesis).

For future MC productions events from both beam halo and cosmics are planned to be mixed under the hard collision, just like for pile-up events. Ultimately of course real cosmic/beam halo/pile-up background events could also be overlaid. In addition to this, loose cuts on the extrapolations of the tracks to the vertex (Δz) can help to further reduce these backgrounds.

II.5.2 QCD background estimation from data

While for Standard Model processes like W+jets or $t\bar{t}$ +jets there are Monte Carlo generator tools which can produce fairly reliable predictions of shapes for the various distributions with high statistics, it is clear that for QCD multi-jet production the enormous cross sections exceed the computational resources available. In addition the theoretical uncertainties for multi-jet events are orders of magnitude larger than in the case of electro-weak processes. This analysis investigates events with at least a single isolated lepton. These are produced in QCD events only via non-prompt mechanisms or via misidentification, e.g. muons from b-jets or electrons from misidentified jets with a large pion fraction. Compared to the inclusive di-jet cross section these "fake" leptons are very rare, and thus difficult to model using inclusive QCD Monte Carlo samples.

For the MUSiC approach the aim is to estimate the QCD contribution from the data and to not rely on the simulated prediction. Since in a generic search one is looking at many different distributions and a diversity of final states, it seems not practical to define control regions for each specific event class. One has to use a more general estimate of the QCD background applicable to all classes. The uncertainties of such cross-class extrapolations have to be absorbed by an appropriate global uncertainty of the QCD estimate, which can be easily incorporated into the search algorithm.

The strategy used to estimate the QCD from data is similar to the methods commonly applied at the Tevatron [87]. A single selection cut, which is effective for distinguishing "fake" leptons from well-measured isolated ones, is inverted or relaxed. The sample with the inverted/relaxed cut is then used to model the shape of the QCD background, and a control region is defined where the sample is scaled to fill up the gap between the remaining SM Monte Carlo samples and the data.

This method is exercised using final states with muons; additional studies show that it works similarly in the electron case. The isolation cut is relaxed

$$0.1 < R_\text{track-isolation} = \frac{\sum p_T \text{ of tracks in 0.3 cone}}{p_T(\mu)} < 0.5. \qquad \text{(II.5.3)}$$

In this way one gets a sample with similar kinematics compared to the QCD events entering the final selection. Relaxing the cut even further one would risk to introduce larger differences in the distributions due to muons well within hadronic jets. Two control regions are defined,

- $110 - 150 \, \text{GeV}$ in the $\sum p_\text{T}$ distribution of the class 1μ 1jet+X

- $130 - 180 \, \text{GeV}$ in the $\sum p_\text{T}$ distribution of the class 1μ $\not{E}_\text{T}$+X.

These two inclusive classes represent quite different corners of the phase space analyzed with MUSiC, once requesting a lepton and a jet and once the combination of a lepton and missing

transverse energy. In this way one gets two independent estimates of the scale factor to be used. Furthermore the two regions are both located at the very low p_T-edge of the distributions, where signal contamination is expected to be small and QCD plus other SM processes dominate. From these control regions ($f_\text{QCD} = 0.3$ for class 1μ 1jet$+X$ and $f_\text{QCD} = 0.1$ for class 1μ $E\!\!\!/_\text{T}+X$) one obtains the following scale factor with its uncertainty:

$$f_\text{QCD} = \frac{\text{``data''} - \text{SM MC without QCD}}{\text{relaxed ``data''} - \text{relaxed SM MC without QCD}} = 0.2 \pm 0.1 \qquad \text{(II.5.4)}$$

The relative error of 50% indicates that the estimation of QCD from "data" for all event classes is not very precise. Nevertheless, since QCD contamination in the signal region is not very large, in most cases even such a large error should have a minor impact on the search sensitivity. It is more vital to get a proper shape of QCD in all classes without the enormous single bin fluctuations of a QCD Monte Carlo sample caused by the lack of MC statistics. Note that since this method is exercised only using a QCD Monte Carlo sample (as "data") the subtraction of the other SM samples is not needed. In any case the contribution in the denominator from relaxed SM MC without QCD is small since these mostly fulfill $R_\text{track-isolation} < 0.1$.

Figure II.5.1 illustrates this: Here the two event classes used for the normalization are shown, comparing the amount of QCD events which pass the cut relaxation to the rest of the SM processes (the error bars correspond to the uncertainty of the scaling factor). One can see that there is at least an order of magnitude between the QCD with relaxed cuts and the relaxed SM samples without QCD. Thus a possible uncertainty on the subtraction of the relaxed SM samples without QCD, see denominator of Equation II.5.2, is well absorbed by the overall 50% uncertainty of the QCD estimate.
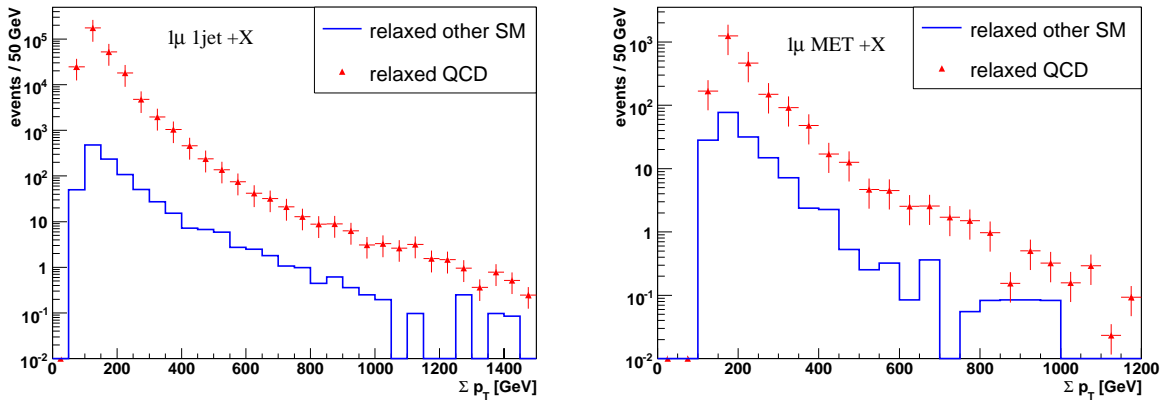

Figure II.5.1: *QCD Monte Carlo and other SM processes with inverted cuts in comparison, for the two event classes used for normalization.*

Figure II.5.2 shows the comparison of the QCD estimate from "data" with respect to the QCD Monte Carlo samples used to perform the cut inversion. The errors correspond to the uncertainty of the scaling factor. The sample with inverted cuts and the one fulfilling all final selection cuts agree well in terms of shape. Note that the event classes shown here do not contain the control regions, thus the agreement within the assumed errors serves as a good indication that the extrapolation from one final state topology to another works reasonably well.
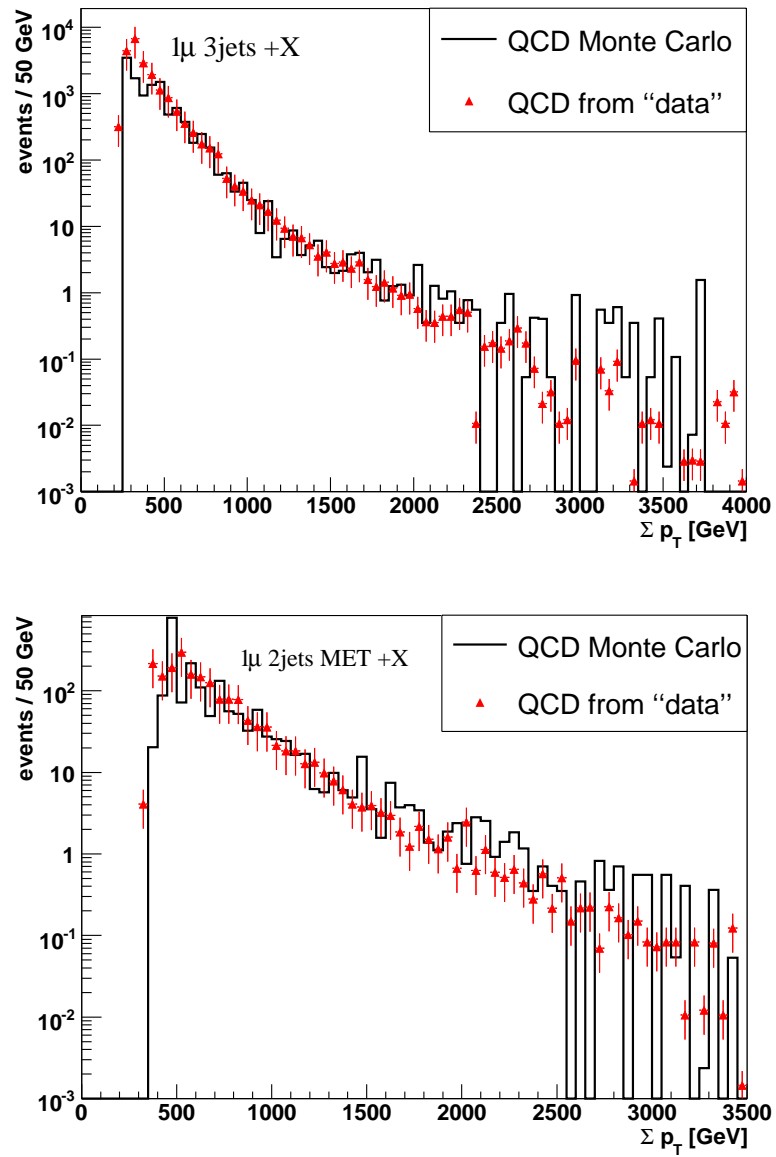
Figure II.5.2: *QCD Monte Carlo and estimate using cut inversion in comparison, for two representative event classes.*

Chapter II.6

Results and Summary

II.6.1 Results

After presenting the technical setup of the analysis, the selection cuts and the generic algorithm, now some representative results of this search approach will be discussed. For a model independent search it is difficult to summarize its performance without data to work with. One can only pick archetypical use cases and toy signals which surely can represent only a small part of the phase space covered by such a generic Ansatz. Still three scenarios which fit well to the concept of a model independent search will be discussed.

In order to set a reasonable threshold for a significant deviation event classes with $\widetilde{P} < 1 \cdot 10^{-3}$ are labaled as "interesting". This corresponds roughly to a $> 3.3\sigma$ deviation. It is clear that this value cannot justify any claim for discovery. Still the main focus of MUSiC is to spot discrepancies in the data worth looking at in detail, one does not aim to give a definitive answer to the question in which channels CMS in general could discover new physics. In this context it should also be noted that the precise value of $\widetilde{P}$ in the statistical sense is not of major importance. It is rather just an estimator defining which deviations are worth to be investigated further.

II.6.1.1 MUSiC timeline

One should emphasize that the succession of the following results also reflect the possible timeline of MUSiC in the context of first data:

With the first pb^{-1} to arrive the focus will not be on discovering new physics but on re-establishing the Standard Model, understanding the detector and validating the Monte Carlo predictions. One would concentrate on the high statistics parts of the distributions where the "SM candles" dominate. In this way it is also possible to ensure that initially one is not overwhelmed by deviations found by the algorithm, thus reducing the amount of distributions to be studied in detail. In this phase of data analysis MUSiC can contribute to the understanding of the detector and the tuning of the generators, see examples below.

In a next phase the focus will begin to shift also to the tails of the distributions where higher order effects like jet-multiplicities become important. Here the validation of the MC predictions will be crucial and comparisons of different event generators, e.g. PYTHIA vs ALPGEN, will be important. Also here MUSiC can help, comparing data and MC prediction in a large part of the

phase space. Thus while one generator might describe one part of the data properly, it might fail in another part. Each time new generator parameter tunes are available MUSiC can compare them to data in a general way and thus contribute to the overall generator validation.

After all initial problems have been solved and confidence in the understanding of the detector and the MC prediction is gained, the full dataset available can be analyzed and one can start looking for deviations from the Standard Model, see examples below.

II.6.1.2 Physics commissioning

Especially during early data taking ($\ll 1\,\text{fb}^{-1}$) the physics focus will not be to discover some signal beyond the Standard Model, but rather to re-establish the SM with the CMS detector. In order to measure the various SM candles properly a lot of work will be needed to understand the CMS detector properly. After years of construction and simulation studies, for the first time data will be recorded and can then be compared to the "ideal" Monte Carlo world. Differences between data and simulated events can arise from Monte Carlo generators not properly describing nature at $14\,\text{TeV}$ center of mass energy, or from a detector not working exactly as predicted by the detector simulation. Both aspects can be addressed using this generic search approach since measurements differing from the expectation can be revealed by the algorithm. Since a large part of the data is divided into event classes MUSiC can perform a general scan of the detector properties, possibly revealing unexpected discrepancies. Of course many detailed studies will examine efficiencies, resolutions and other detector properties to a great extent. Still, MUSiC can serve as a cross check. Thus it is interesting to see how well the data agree in general in the various event classes without extensive tuning and optimizations. On the other hand MUSiC can assist in the process of Monte Carlo tuning, monitoring the improvements of the SM Monte Carlos and giving feedback where additional changes might be needed.

In the context of this demonstration of physics commissioning the multi-jet background is not included since its large statistical fluctuations distorts the results and since it is irrelevant for the message here. With first data of course multi-jets will be considered though. Also for the first two examples the assumed cross section uncertainty is reduced from 10% to 3%. The conservative estimate used for the search for new physics is needed for the tails of the distributions where the variation of the PDFs yield cross sections not known precisely. With the first data the focus will rather be on the SM candles, and here uncertainties of 3% or below are more realistic [86]. Assuming a 10% error on all background processes would absorb much of the deviations caused by any detector inefficiency or mismeasurement.

Unexpected detector effects

As a first toy example a detector effect is introduced which is totally unexpected, thus not included within the systematic uncertainties. In order to do this the jet energy scale uncertainty of 5% is disabled and a JES of 10%-up in the pseudo-data is assumed. Figure II.6.1 shows the $\sum p_\text{T}$ distribution of the inclusive class with 1 muon and 4 jets for a single CMS pseudo-experiment. One can see that the variation in the jet energy scale leads to a considerable excess of data in the complete distribution. This is caused by the fact that the jet-p_T spectrum has a steep slope. By increasing the JES more jets are selected, resulting in an overall excess in the 4-jet channel. The region of interest selected by the algorithm is very wide and the p-

and $\widetilde{P}$-values are very small. Given the limited amount of background-only repetitions (step 2 of the MUSiC algorithm) only a lower limit for $\widetilde{P}$ can be given, resulting in a significant deviation of at least 4.4σ. Repeating the S+B hypothesis several times (S+B corresponds here to Standard Model with JES 10%-up) one can also determine the expected significance, which is $\widetilde{P}$ (expected) $> 4.4\sigma$. In general many event classes with jets show significant deviations thus revealing a possible problem.

If now the assumed JES uncertainty of 5% is re-introduced things change: Only an expected $\widetilde{P}$ value of 12% $\,\widehat{=}\,1.6\sigma$ is left, thus no significant deviation is found. This is caused by the fact that the assumed 5% JES uncertainty absorbs the 10% variation in the pseudo-data, measurement and expectation would agree within errors.
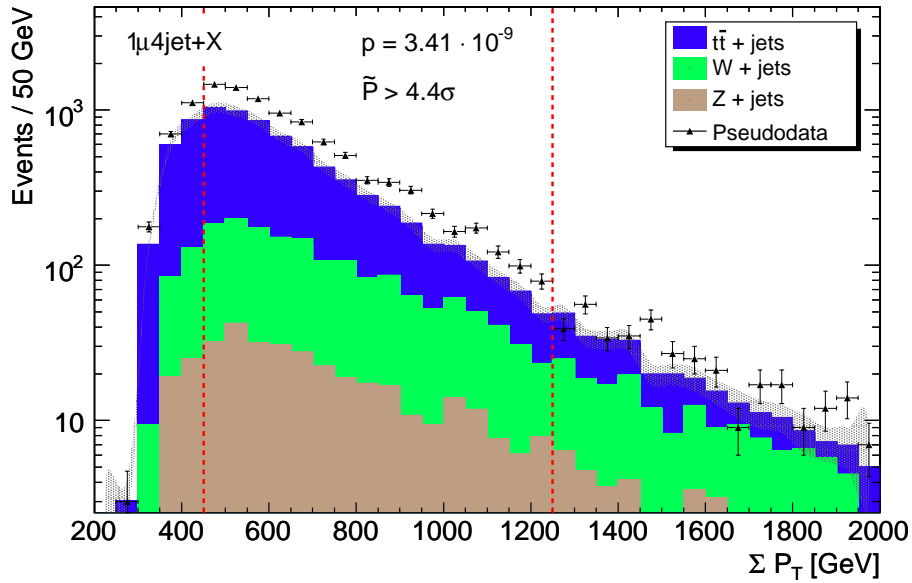

Figure II.6.1: *Event class $1\mu\ 4jet + X$ with JES 10% up in pseudo-data. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest (dotted lines): $N_{\text{data}} = 8831$ and $N_{\text{MC}} = 6202 \pm 448$*

Detector simulation tuning

In a second example the muon reconstruction efficiency is decreased by 15% in the pseudo-data, thus the drop in efficiency exceeds the assumed muon-efficiency uncertainty of 2% (see Section II.5.1). Effects like this could be caused by problems in the HLT-muon trigger and/or by unexpected inefficiencies in the muon reconstruction which are not included in the detector simulation yet. Figure II.6.2 shows the $\sum p_{\text{T}}$ distribution of the inclusive class with two muons for a single CMS pseudo-experiment (shaded area in last bin due to unphysical plotting artifact). One can see an obvious deficit of data compared to the SM expectation, reflecting the inefficiency in the data. The algorithm picks the tail of the Z-peak as the region of biggest discrepancy and the deviation well exceeds 3σ ($\widetilde{P} = 2.6 \cdot 10^{-4} \,\widehat{=}\, 3.7\sigma$).

Repeating many pseudo-experiments with the low muon-efficiency one can also determine an expected $\widetilde{P}$ which results in 3.2σ. On the other hand, without artificially lowering the efficiency by 15% one gets $\widetilde{P}$ (expected) $= 0.5$, thus all is well in agreement as expected. The expected significances indicate that the muon inefficiency can be observed, but does not exceed 5σ. The

assumed cross section and luminosity uncertainties are too large and absorb part of the muon inefficiency. Still, also other classes with muons show indications of a data deficit, thus one would get a consistent picture of a few channels with muons showing deviations at a few σ.

Both examples underline the importance of including all major systematic uncertainties at the correct order of magnitude into the algorithm. In this way the discrimination between a true signal of new physics and a simple detector effect seems possible.
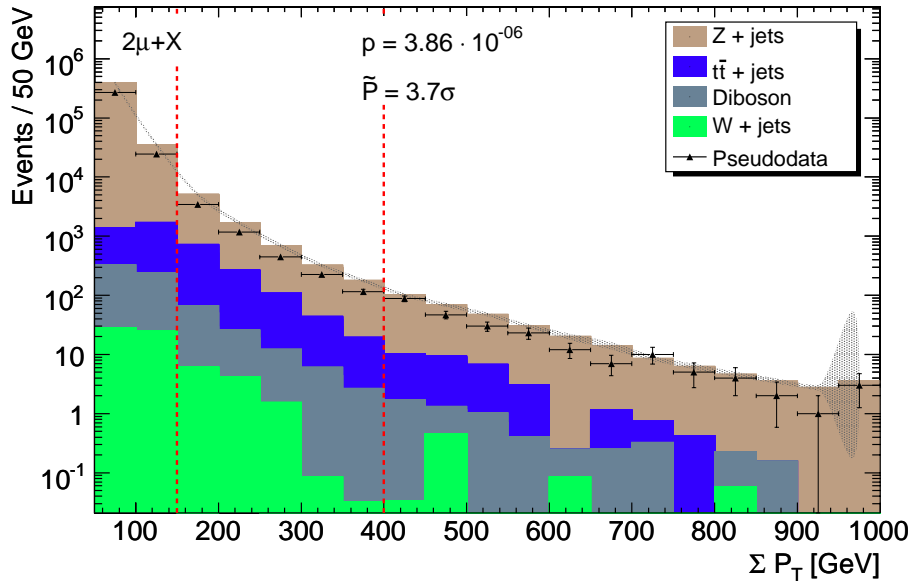

Figure II.6.2: *Event class $2\mu + X$ in case of a 15% drop in muon-efficiency in the pseudo-data. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest (dotted lines): $N_{\mathrm{data}} = 5416$ and $N_{\mathrm{MC}} = 7942 \pm 560$*

MC generator tuning

In the last toy example the possibility of a Monte Carlo prediction not describing the data properly is exercised. In order to do so one compares pseudo-data which are produced using ALPGEN $W + N_{\mathrm{jets}}$ samples to a SM expectation which relies on a PYTHIA inclusive $W \rightarrow e\nu$ sample (additional jets only from parton showering). Thus one would expect pseudo-data with higher jet-multiplicities and harder jet momenta. Since these higher order effects become important especially in the tails of the distributions the 10% cross section uncertainty are re-introduced.

Given the fact that the PYTHIA sample has only a rather small amount of statistics it is clear that in the very high-p_{T} tails and for high jet-multiplicities the PYTHIA sample does not contribute at all, leading to huge deviations with respect to the ALPGEN-pseudo-data. Thus the results presented here should not be interpreted as a detailed generator comparison but only as an illustration of the MUSiC performance.

Due to this lack of statistics the inclusive $1e\ 1jet\ \not{E}_{\mathrm{T}} + X$ event class is chosen where PYTHIA still contributes. Also an upper bound of 1350 GeV in $\sum p_{\mathrm{T}}$ is set for the search regions since above PYTHIA suffers from a lack of statistics. Figure II.6.3 shows the $\sum p_{\mathrm{T}}$ distribution of this inclusive event class, comparing ALPGEN-pseudo-data to a MC prediction using the PYTHIA

$W \to e\nu$ sample. One can clearly see an overall excess of pseudo-data and the algorithm picks the tail of the distribution where the excess well exceeds the systematic uncertainties. The overall excess can be explained by the fact that this event class is inclusive and that $\not{E}_{\mathrm{T}} > 100$ GeV is demanded. Since ALPGEN predicts considerably more events with many and/or hard jets these can all contribute to the event class. As the jets are harder and more numerous also the boosts of the decaying W-boson are more intense, leading to more events with large $\not{E}_{\mathrm{T}}$. The deviation in the Region of Interest is significant ($\widetilde{P} > 4.4\sigma$) and the deviation would even increase when allowing also regions > 1350 GeV.
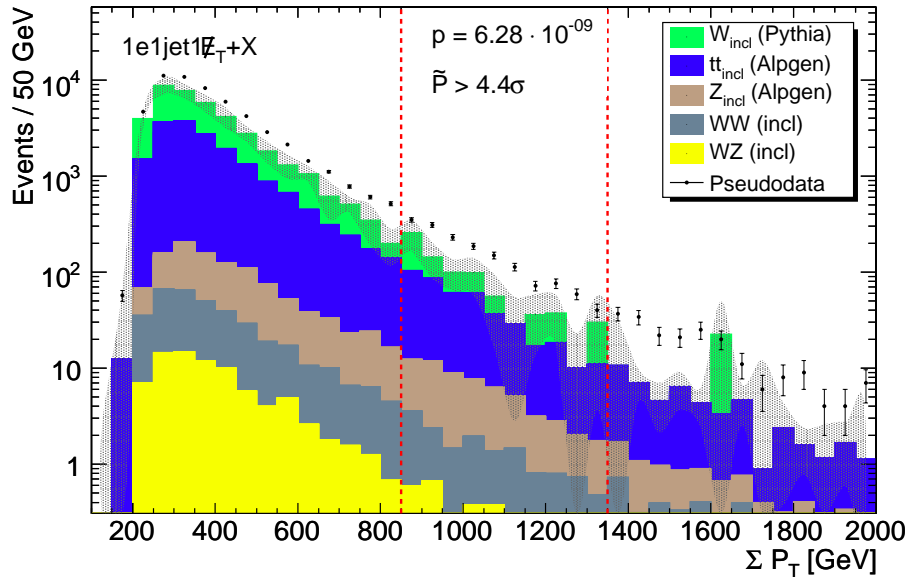

Figure II.6.3: *Event class* $1e\ 1jet\ \not{E}_{\mathrm{T}} + X$ *in case of pseudo-data following an* ALPGEN $W + N_{jets}$ *sample and MC expectation using* PYTHIA $W \to e\nu$ *instead. Upper border of Region of Interest constrained to* < 1350 GeV *(due to lack of* PYTHIA *statistics). The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the region of interest (dotted lines):* $N_{\mathrm{data}} = 1588$ *and* $N_{\mathrm{MC}} = 799 \pm 133$

II.6.1.3   Unexpected signal

Another important aspect of a model independent search would be the detection of a signal in corners of the phase space not analyzed within specific analyses so far. As mentioned in the motivation of the concept it is almost impossible to predict how new physics will look like at the LHC. Thus there is a possible danger in optimizing the selection with respect to a certain model since this might blind the analysis for a signal hidden in the data. Thus MUSiC can serve as an add-on to the physics program of CMS, scanning the data in a generic way and possibly spotting some unexpected deviation. Since no cut optimization can be performed within MUSiC of course such a "surprise" signal has to be rather prominent, thus large cross section at regions of the distribution where only little background is expected. Still there are several model candidates which fulfill these requirements, e.g. mini-black-holes, excited leptons or new heavy gauge bosons. Of course one might argue that for all these models dedicated analyses are already in place. Still, also in this context MUSiC is an independent cross check and is possibly sensitive to spectacular models not yet thought of.

New heavy gauge bosons

As a toy example the CSA07 sample cocktail incorporates a feature that is a nice dress rehearsal for MUSiC: A Z' signal with a mass of 1 TeV and a cross section of $\sigma = 365$ fb has been hidden within the "data" samples. As it is clear that a dedicated analysis (see e.g. [88]) or a search optimized for identifying mass resonances is superior to this approach, the results should be interpreted as a proof of principle that also the MUSiC algorithm is capable of finding a narrow excess.

Scanning all event classes the biggest discrepancy between pseudo-data and SM expectation is found in the M_{inv} distribution of the class $2e + X$, as one would expect. Figure II.6.4 shows this distribution for a single CMS experiment. The region of interest nicely selects the Z'-peak at 1 TeV and the p-value of 10^{-36} indicates that a discovery is possible. The $\widetilde{P}$ gives a lower limit of $> 4.4\sigma$, corresponding to a generation of 100 000 background-only experiments in the second step of the algorithm ($\widetilde{P} < 1/100\ 000 \,\widehat{=}\, \widetilde{P} > 4.4\sigma$, see Figure II.4.3). With a larger generation of pseudo-experiments also a $\gg 5\sigma$ deviation can be expected. Note that the signal contribution also spreads to lower M_{inv} masses since in the Z' generation interferences with Drell-Yan are also simulated.

Figure II.6.5 displays the distribution of the p-values repeating the S+B hypothesis and the background-only hypothesis many times. One can see that both curves are well-separated and a discovery is possible.
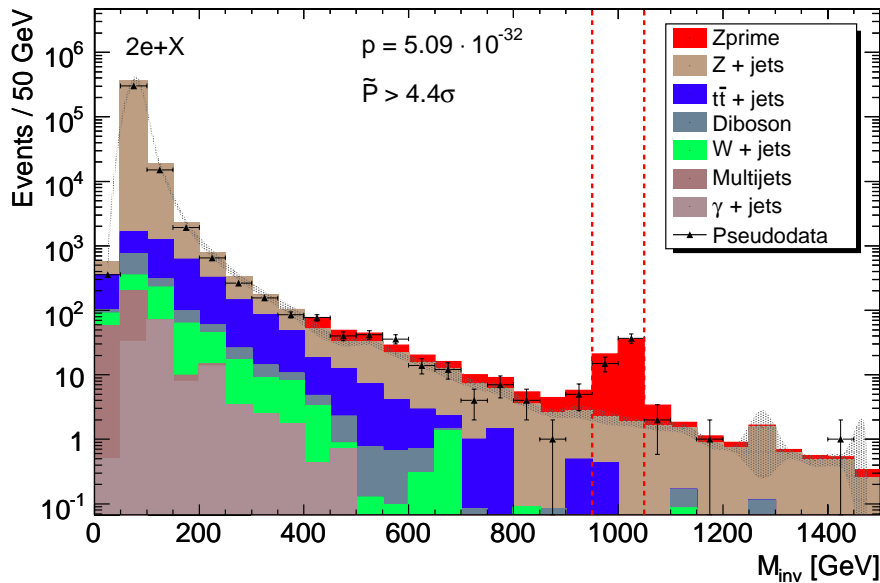

Figure II.6.4: *Single CMS experiment assuming 1 fb^{-1}, pseudo data are mixed with a 1 TeV Z'-signal. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest (dotted lines): $N_{data} = 52$ and $N_{MC} = 4 \pm 0.7$*

Note that also in the $\sum p_T$ distribution of the exclusive two electron class a significant deviation is found. No final states with muons show indications of the hidden Z'. This perfectly matches to the fact that only the decay channel into electron has been enabled during the signal sample generation.
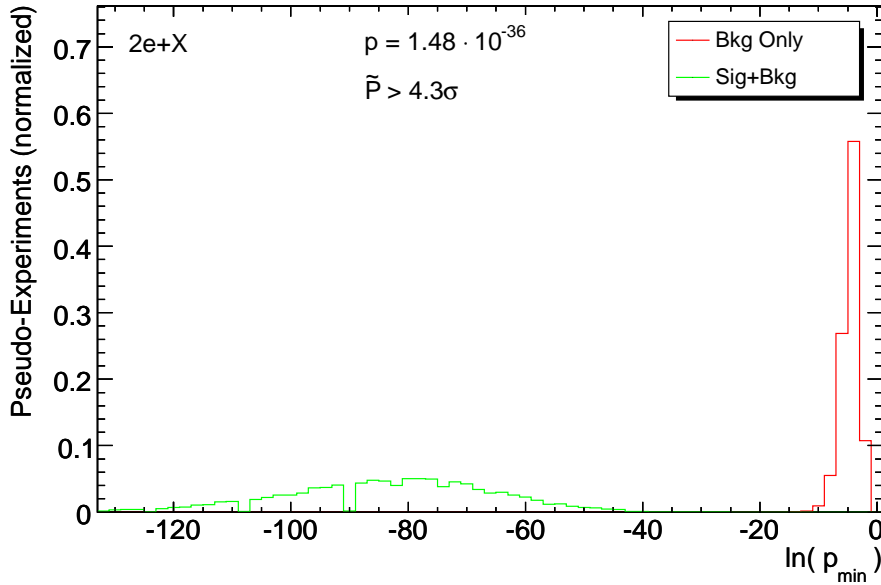
Figure II.6.5: *Many repeated CMS experiments assuming* $1\,\mathrm{fb}^{-1}$ *of data. The two curves represent the signal* (Z') *plus background (SM) hypothesis and the background-only hypothesis.*

Leptoquarks

As another exotic signal of new physics leptoquarks can be studied which arise in some Grand Unified Theories. These hypothetical massive particles can mediate transformations from leptons to quarks, thus establishing a symmetry between both fermionic species. In the context of LHC they can be produced as pairs, and a first generation leptoquark with a mass of $400\,\mathrm{GeV}$ and a cross section of $\sigma = 1200$ fb is considered. In particular $\beta = 1$ is used and thus only the decay $LQ \to e + u$ is allowed, while the decay into neutrinos is forbidden. Since leptoquarks are produced as pairs one can expect contributions especially in the $2e$ 2jet topology.

Figure II.6.6 shows the $\sum p_\mathrm{T}$ distribution of the $2e$ 2jet+X event class. One can clearly see the excess caused by the leptoquarks signal which sets in at twice its mass $(800\,\mathrm{GeV})$ and spreads up to the high-p_T tail. This is reflected by the region of interest of the algorithm and the small p-value $(2 \cdot 10^{-16})$ which results in a lower limit of $\widetilde{P} > 4.4\sigma$. Also when repeating the S+B hypothesis several times significant deviations can be found (p (expected) $= 5 \cdot 10^{-24}$ and $\widetilde{P}$ (expected) $> 4.4\sigma$), thus the leptoquark signal can be discriminated from the background-only hypothesis.

Note that a resonance peak reflecting the leptoquark mass cannot be observed since in this event class both electrons and jets are combined. In the $1e$ 1jet topology on the other hand the background dominates, thus the signal is not visible. Ideally in a dedicated search one would select events from the $2e$ 2jet topology and then construct the invariant masses of the electron-jet pairs. A similar thing cannot be done within MUSiC easily in a generic way since the combinatorics for complex topologies would be large. This is an example of the interplay between MUSiC and more dedicated analyses. In order to understand the excess found in the $2e$ 2jet+X event class and in order to gain information on the underlying theory additional studies have to be performed, e.g. looking at the invariant masses of the pairs of this topology.
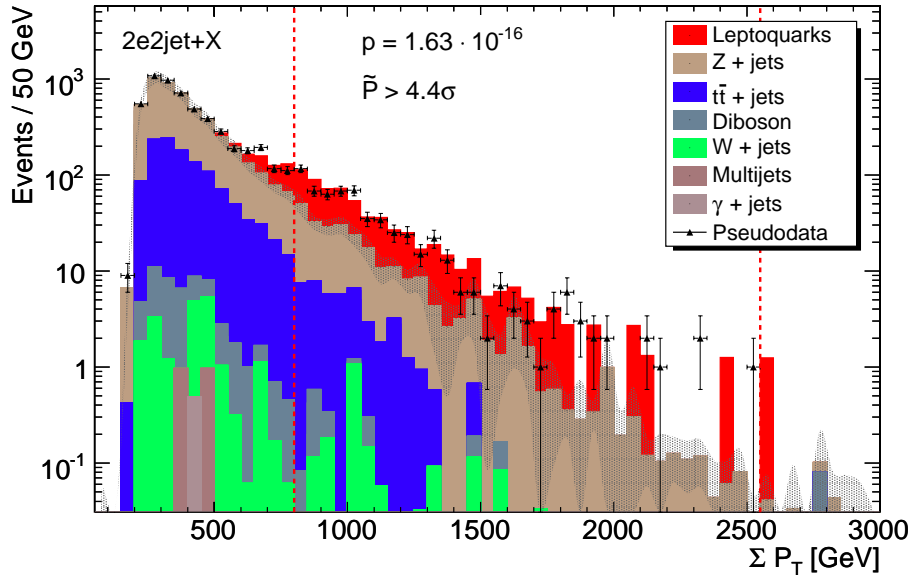
Figure II.6.6: *Single CMS experiment assuming* $1\,\mathrm{fb}^{-1}$*, pseudo data have* $400\,\mathrm{GeV}$ *1st generation lep-toquarks mixed under. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest (dotted lines):* $N_{\mathrm{data}} = 604$ *and* $N_{\mathrm{MC}} = 249 \pm 37$

Negative example: Higgs

In order to also give a "negative" search example there is another hidden signal within the CSA07 samples: A Standard Model Higgs with a mass of $150\,\mathrm{GeV}$ and a cross section of $\sigma = 325$ fb. Performing a global scan for this signal MUSiC unfortunately is not able to spot any deviation. Still also this is a positive/consistent result since the Higgs signal in the distributions analyzed is flooded by SM background. Thus there is no chance to discover anything without optimizing the selection cuts with respect to this specific Higgs signal.

Figure II.6.7 shows a representative distribution, using the $1e1\mu + X$ event class. For better visibility in this case the Higgs signal is not added to the SM background. One can see that in the kinematic region of the distribution where the Higgs contributes the SM is orders of magnitudes above the signal. The region of interest picked by the algorithm is not close to any signal, thus this is only a fluctuation of the background. This is reflected by the $\widetilde{P}$ value of 0.8σ, the deviation found in the pseudo-data agrees with the SM expectation well within the assumed uncertainties.

II.6.1.4   SUSY sensitivity in mSUGRA models

There are several evident reasons why such a generic search strategy as presented here might be a good supplement to more conventional analyses hunting for SUSY. First of all the large number of unconstrained parameters in most supersymmetric extensions of the SM also leads to an almost unlimited parameter space where nature can pick the SUSY scenario realized. Simplified models like MSSM try to shrink the number of free parameters using well or not so well-founded physics assumptions. Still the parameter space remains huge, with a multitude of possible "benchmark"
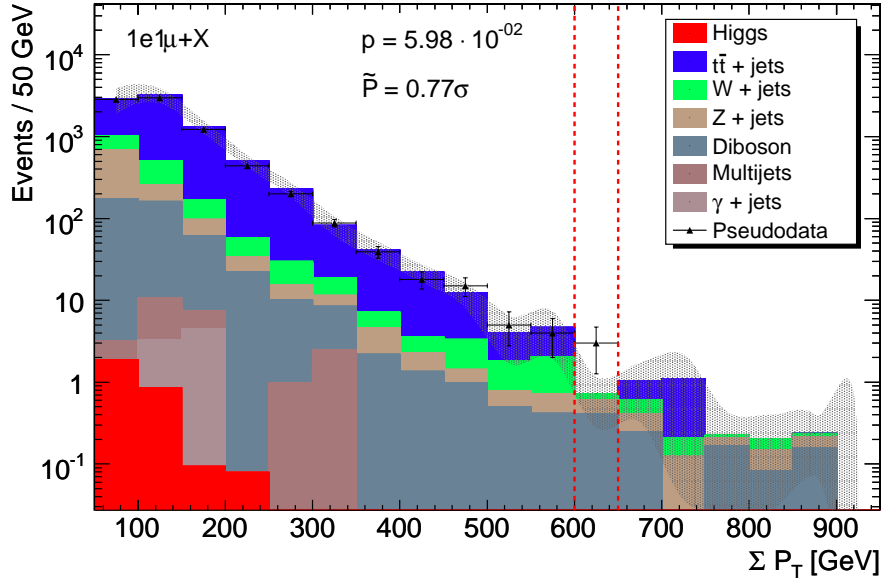
Figure II.6.7: *Event class $1e1\mu + X$ for a single experiment. The tiny Higgs signal is drawn in front of the SM background for better visibility. The shaded area corresponds to the syst. uncertainty of the SM expectation. $\widetilde{P}$ indicates a good agreement between pseudo-data and SM expectation and the algorithm is not sensitive to the Higgs signal in this distribution.*

points leading to completely different decay channels and signal topologies. Thus it might be dangerous to rely solely on analyses optimized on specific SUSY points.

Through their decay chains, SUSY particles often lead to spectacular cascades in the final state with high multiplicities of leptons and jets and a large amount of $\not{E}_T$ due to the LSP (lightest SUSY particle). So unlike single resonance production as for example $Z \to \mu\mu$, SUSY does not predominantly favor a single topology, but does contribute to a multitude of event classes within MUSiC. Thus this generic search aims to give a consistent picture of SUSY particles appearing on top of the Standard Model prediction. The information of significant deviations found in several classes could be combined to provide additional evidence for SUSY.

Since it would contradict the basic philosophy of MUSiC to really perform a large SUSY parameter scan the search results are just highlighted using a typical mSUGRA point:

- **SUSY LM4:** $m_0 = 210$; $m_{1/2} = 285$; $\tan\beta = 10$; $\mathrm{s}gn(\mu) = +$; $A_0 = 0$;
  σ (LO) = 20.5pb ; σ (NLO)= 27.7pb

For point LM4, the decay of the $\tilde{\chi}_2^0$ into on-shell Z's and $\tilde{\chi}_1^0$ is characteristic. $\tilde{q}\tilde{g}$ production is dominant, contributing about half of the total cross section. The squark and gluino masses are below $\sim 700\,\mathrm{GeV}$, leading to relatively large total cross sections at $14\,\mathrm{TeV}$ center of mass energy. Still, LM4 is not as close to the Tevatron exclusion limits [14] as for example LM1.

It will be discussed how well MUSiC performs in the various event classes and using the studied distributions. A more comprehensive analysis of different LM-points as well as a comparison to more model-specific analyses can be found in [89]; to summarize the results follow the expectations: With a dedicated search optimized for the specific point the expected significances are higher – or correspondingly, the required discovery luminosities are lower. Still MUSiC performs

best when several channels are combined into a comprehensive review in accordance with its strategy.

Global scan

As a first test a global scan of all event classes is performed, assuming 1 fb^{-1} of data. In order to not depend on fluctuations of a single CMS pseudo-experiment, but predict expected significances the S+B hypothesis is repeated several times. The median p-value of these experiments assuming SUSY plus SM can then be used to compute an expected $\widetilde{P}$, as explained earlier.

In total S+B (either signal or background MC) contribute to 375 inclusive classes and 315 exclusive classes. From the total number of 375 inclusive classes (315 exclusive classes) ≈ 110 (≈ 120) have so little signal or background MC entries that in most pseudo-experiments not a single pseudo-data point is diced.

LM4 has entries in 260 inclusive and 163 exclusive channels which show at least a minimal degree of discrepancy (single region with $N_{\mathrm{data}} - N_{\mathrm{MC}} > 3 \cdot \delta N_{\mathrm{MC}}$). Already these numbers indicate that SUSY is present in a large part of the data, thus many different topologies could give rise to a SUSY signal. One can summarize the results in the following way:

- LM4 contributes to 160 *exclusive* event classes (94 classes with $\not{E}_{\mathrm{T}}$):
  15% show significant deviations with $\widetilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in $\sum p_{\mathrm{T}}$
  19% show significant deviations with $\widetilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in M_{inv}
  38% show significant deviations with $\widetilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in $\not{E}_{\mathrm{T}}$

- LM4 contributes to 260 *inclusive* event classes (170 classes with $\not{E}_{\mathrm{T}}$):
  36% show significant deviations with $\widetilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in $\sum p_{\mathrm{T}}$
  33% show significant deviations with $\widetilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in M_{inv}
  59% show significant deviations with $\widetilde{P}$ (expected) $< 1 \cdot 10^{-3}$ in $\not{E}_{\mathrm{T}}$

Of course in the case of inclusive classes deviations found are "duplicated" in some way since 1μ 5jet events contribute to 1μ 2jet $+ X$, 1μ 3jet $+ X$ and so on. Nevertheless, when comparing similar final state topologies, the inclusive classes tend to have smaller p (expected) values and larger event class significances than the exclusive ones. This result that the inclusive channels look more promising is consistent with the expectation one would have for SUSY events: The complex decay chains produce a certain number of jets only with a certain probability, there is no single (exclusive) topology enhanced over others. Thus by summing up events of a certain topology inclusively more SUSY contributions can be integrated, and the excess observed in the data is more striking.

Two of the three kinematic distributions examined, $\sum p_{\mathrm{T}}$ and M_{inv} (M_{T}), lead to similar results. A systematic advantage of one of them cannot be observed in the presented LM4 analysis.

SUSY particles are produced as pairs. In the context of MUSiC all particles of the event class are combined to M_{inv} (M_{T}), thus not necessarily the correct particle combinations are found to produce resonance peaks. In addition to this the LSP distorts the picture such that only transverse masses can be constructed.

On the other hand one can observe a clear gain when using the $\not{E}_{\mathrm{T}}$ distribution. This is caused by the fact that the LSP in SUSY events leads to a considerable amount of missing transverse energy. Thus when analyzing event classes with leptons, jets and $\not{E}_{\mathrm{T}}$ the separation between SM

and SUSY is best when using the $\not{E}_T$ variable.

Unfortunately experiences from past accelerator experiments show that $\not{E}_T$ will be difficult to control and understand in the first data. Thus it might be desirable to also investigate classes without $\not{E}_T$ which show indications of SUSY.

As mentioned in Section II.6.1.1 it might be desirable in the beginning to discard deviations found by the algorithm which suffer from poor statistics. In this context a statistical failsafe is included and regions which have $N_{\mathrm{data}} < 3$ are discarded. For the $\sum p_T$ distribution this leads to 129 exclusive and 185 inclusive classes to be considered. 14% of these exclusive event classes have significant deviations and 32% of the inclusive ones. Thus also without the low statistics cases SUSY is prominent.

Investigating event classes

In the following a few representative event classes with significant discrepancies will be highlighted, stating their expected significances and showing some distributions of single CMS pseudo-experiments. Analyzing the many event classes which are above the 3-sigma threshold there are of course some topologies with spectacular particle multiplicities, e.g. 1μ 8jet $\not{E}_T$. Obviously one cannot expect the Monte Carlo prediction to perfectly match the measured data in these extreme kinematic regions. Then there are classes where SUSY contributes with little SM background or even no SM contribution at all, e.g. $2e$ 1μ 4jet $+ X$. It is clear that deviations found in classes with exotic particle combinations have to be taken with some care.

Still also these "exotic" classes are worth looking at since the spectacular SUSY decays will populate them while the SM (including its uncertainties) is almost negligible. If there are 100 events in the 6 muon channel where SM is close to zero something interesting is going on which can very likely not be explained by the SM alone or by MC not working properly. Note that for bins with pseudo-data entries where (given the limited statistics) not a single MC event is predicted, a conservative 68% upper Poisson limit of 1.15 events is used. This upper limit is applied to all samples contributing to this specific event class and then scaled according to the assumed luminosity.

In general there are several classes with a single lepton plus several jets and $\not{E}_T$ which show significant deviations. Using $\sum p_T$ the p-values and expected significances are not as low as using the $\not{E}_T$ variable. There is a considerable $t\bar{t}$ +jets and W+jets background left. The systematic errors of these backgrounds, especially the jet energy scale, lead to relative uncertainties around $20\% - 30\%$, thus signal and background cannot be separated perfectly.

The other type of class with significant deviations contains multi-leptons plus jets ($+\not{E}_T$). These channels look a bit more promising in the context of MUSiC since many combinations of 2 or 3 leptons show small p-values and $\widetilde{P}$. Without a dedicated cut optimization with respect to the SUSY signal the mere topology cut of the particle content suppresses the SM background considerably. For MUSiC the requirement of a spectacular final state already is a good discriminator between SUSY and SM.

Tables II.6.1 and II.6.2 show some typical event classes of different types where a significant deviation is found assuming 1 fb^{-1}. The table can be structured as classes with di-leptons plus jets and $\not{E}_T$, investigating the $\not{E}_T$ distributions and classes not utilizing the $\not{E}_T$-variable. The numbers and regions for a single CMS experiment are stated as well as $\widetilde{P}$ (expected) when re-

peating the S+B and B hypotheses multiple times:

		single CMS experiment				
Event Class	dis	ROI [GeV]	N_{data}	N_{MC}	p-value	$\widetilde{P}$
$1e\ 1\mu$ 3jet $\not{E}_{\text{T}} + X$	$\sum p_{\text{T}}$	$1000 - 2650$	188	61 ± 18	$2.6 \cdot 10^{-9}$	$> 4.4\sigma$
$2e$ 4jet $\not{E}_{\text{T}} + X$	$\sum p_{\text{T}}$	$1050 - 3350$	71	17 ± 4	$1.6 \cdot 10^{-12}$	$> 4.4\sigma$
2μ 3jet $\not{E}_{\text{T}}$	M_{T}	$1000 - 2400$	73	25 ± 5	$1.1 \cdot 10^{-8}$	$> 4.4\sigma$
1μ 3jet $\not{E}_{\text{T}} + X$	$\not{E}_{\text{T}}$	$400 - 950$	424	142 ± 40	$1.3 \cdot 10^{-10}$	$> 4.4\sigma$
$2e\ \not{E}_{\text{T}} + X$	$\not{E}_{\text{T}}$	$300 - 750$	123	22 ± 7	$9.0 \cdot 10^{-20}$	$> 4.4\sigma$
$1e\ 2\mu$ 3jet $+ X$	$\sum p_{\text{T}}$	$550 - 900$	12	0.9 ± 0.7	$1.2 \cdot 10^{-6}$	4.3σ
2μ 5jet $+ X$	$\sum p_{\text{T}}$	$700 - 2250$	112	49 ± 8	$2.7 \cdot 10^{-7}$	4.0σ

Table II.6.1: *Results for SUSY point LM4 using different distributions and assuming $1\ \text{fb}^{-1}$ of data. The numbers correspond to a single representative CMS experiment, the error on N_{MC} corresponds to the systematic uncertainty.*

		multiple experiments	
Event Class	dis	p (exp)	$\widetilde{P}$ (exp)
$1e\ 1\mu$ 3jet $\not{E}_{\text{T}} + X$	$\sum p_{\text{T}}$	$2.0 \cdot 10^{-7}$	4.0σ
$2e$ 4jet $\not{E}_{\text{T}} + X$	$\sum p_{\text{T}}$	$1.9 \cdot 10^{-14}$	$> 4.4\sigma$
2μ 3jet $\not{E}_{\text{T}}$	M_{T}	$1.5 \cdot 10^{-8}$	4.4σ
1μ 3jet $\not{E}_{\text{T}} + X$	$\not{E}_{\text{T}}$	$8.7 \cdot 10^{-12}$	$> 4.4\sigma$
$2e\ \not{E}_{\text{T}} + X$	$\not{E}_{\text{T}}$	$3.6 \cdot 10^{-17}$	$> 4.4\sigma$
$1e\ 2\mu$ 3jet $+ X$	$\sum p_{\text{T}}$	$6.8 \cdot 10^{-7}$	4.3σ
2μ 5jet $+ X$	$\sum p_{\text{T}}$	$4.8 \cdot 10^{-7}$	3.9σ

Table II.6.2: *Results for SUSY point LM4 using different distributions and assuming $1\ \text{fb}^{-1}$ of data. The numbers state expected p-value and significance when repeating S+B and B hypothesis multiple times.*

The following plots refer to the same event classes and same numbers as stated in Table II.6.1. Figure II.6.8 shows two distributions for a single CMS pseudo-experiment using event classes with di-leptons plus jets and $\not{E}_{\text{T}}$. The upper plot displays $\sum p_{\text{T}}$ for the inclusive class $1e\ 1\mu$ 3jet $\not{E}_{\text{T}} + X$, while the lower plot refers to the transverse mass of an exclusive class (2μ 3jet $\not{E}_{\text{T}}$). Both distributions show significant deviations well above 4σ. The expected significances $\widetilde{P}$ (expected) in Table II.6.2 indicate that these classes in general show indications of SUSY. In both cases $t\bar{t}$ is the dominant background, and in both distributions ($\sum p_{\text{T}}$ and M_{T}) $t\bar{t}$ peaks at lower energies while the signal populates also the tails of the distribution. This is reflected by the algorithm which in both cases picks a region of interest in the high-p_{T} tail where the SUSY contributions lead to a considerable excess.
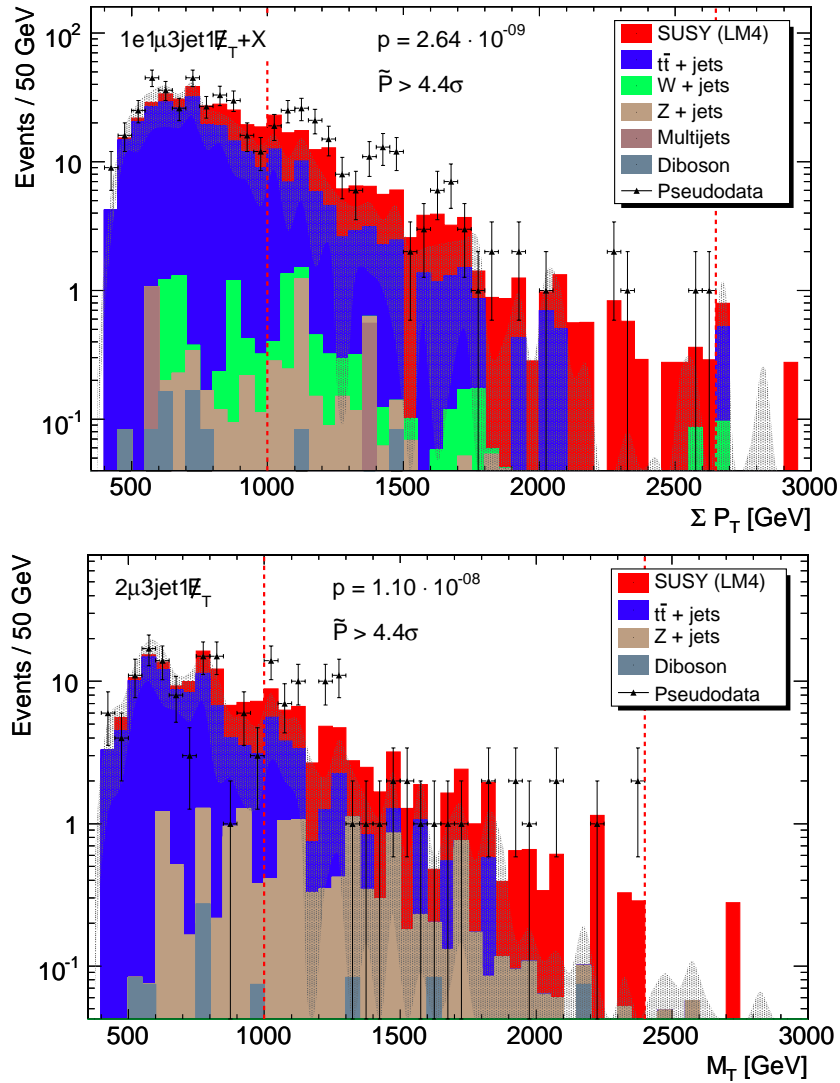
Figure II.6.8: *Results of a representative single pseudo experiment assuming* $1\,\text{fb}^{-1}$*, using event classes with di-lepton+jets+$\slashed{E}_\text{T}$. The dotted lines indicate the region of interest where the smallest p-value has been found, $\widetilde{P}$ corresponds to the significance assuming this single set of toy data. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest can be found in Table II.6.1.*

In general the biggest discrepancies between pseudo-data and MC expectation can be found in the $\slashed{E}_\text{T}$ distributions. Figure II.6.9 shows two typical $\slashed{E}_\text{T}$-distributions, on the top 1μ 3jet $\slashed{E}_\text{T}+X$ and on the bottom $2e$ $\slashed{E}_\text{T} + X$. In the single lepton as well as the di-lepton class the separation between signal and SM background is very good. While SM background events tend to have small $\slashed{E}_\text{T}$ the LSP in SUSY events leads to a considerable amount of missing transverse energy. Thus at high $\slashed{E}_\text{T}$ values MUSiC counts hundreds of the pseudo-data while only a fraction of this is expected by SM alone. These excesses are so large that relative uncertainties of the SM predictions of $\approx 30\%$ still lead to 4σ deviations.

On the other hand of course using $\slashed{E}_\text{T}$-distributions could also be problematic, especially with the early data in case this variable is not perfectly understood. In this context Figure II.6.10 shows
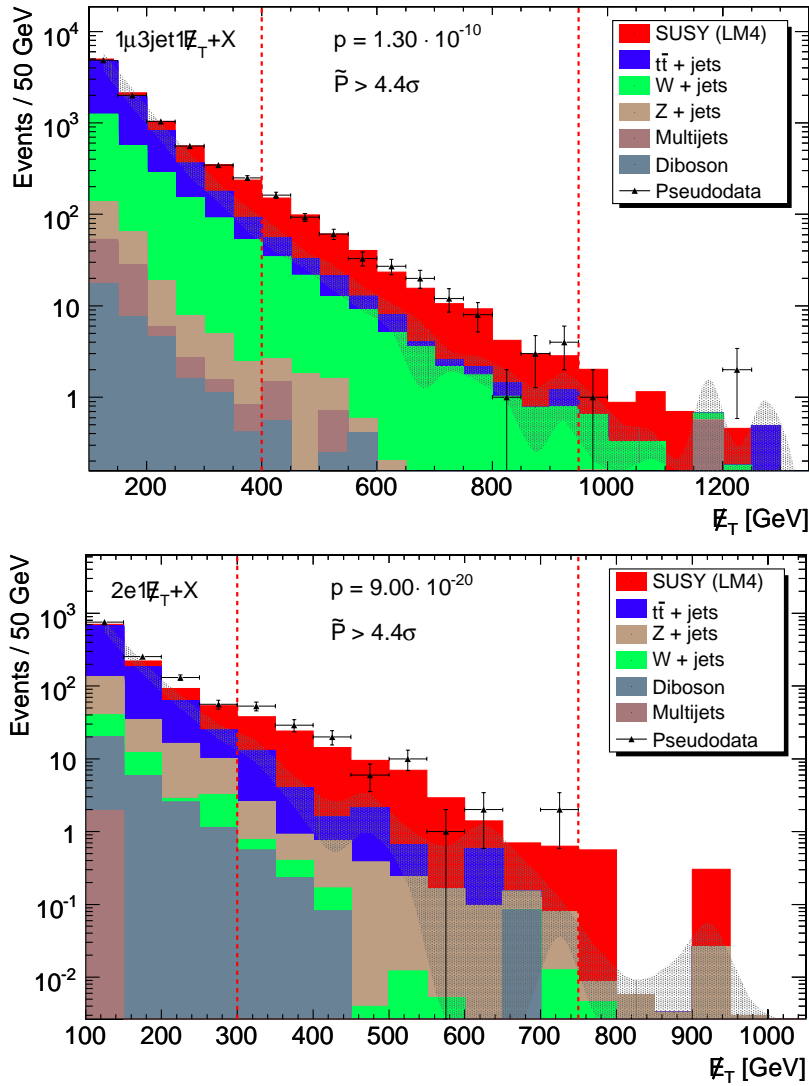
Figure II.6.9: *Results of a typical single pseudo experiment assuming 1 fb⁻¹, analyzing the $\not{E}_T$-distribution. The dotted lines indicate the region of interest where the smallest p-value has been found, $\widetilde{P}$ corresponds to the significance assuming this single set of toy data. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest can be found in Table II.6.1.*

two inclusive event classes which do not use $\not{E}_T$ at all. The upper plot refers to the 2μ $5\text{jet} + X$ class. Here almost all pseudo-data points are integrated within the Region of Interest, leading to a significant overall-excess. Due to the complex decay chains of the SUSY particles several highly energetic jets are produced. Thus the spectacular event class with leptons and 5 jets is well-populated.

The lower plot refers to the $1e$ 2μ $3\text{jet} + X$ event class. Here, using the MC statistics available, less than 1 event is predicted by the SM in the Region of Interest, while SUSY leads to a handful of events. By requiring three leptons and at least three jets the SM background is reduced so much that one does not need the additional separating variable $\not{E}_T$.
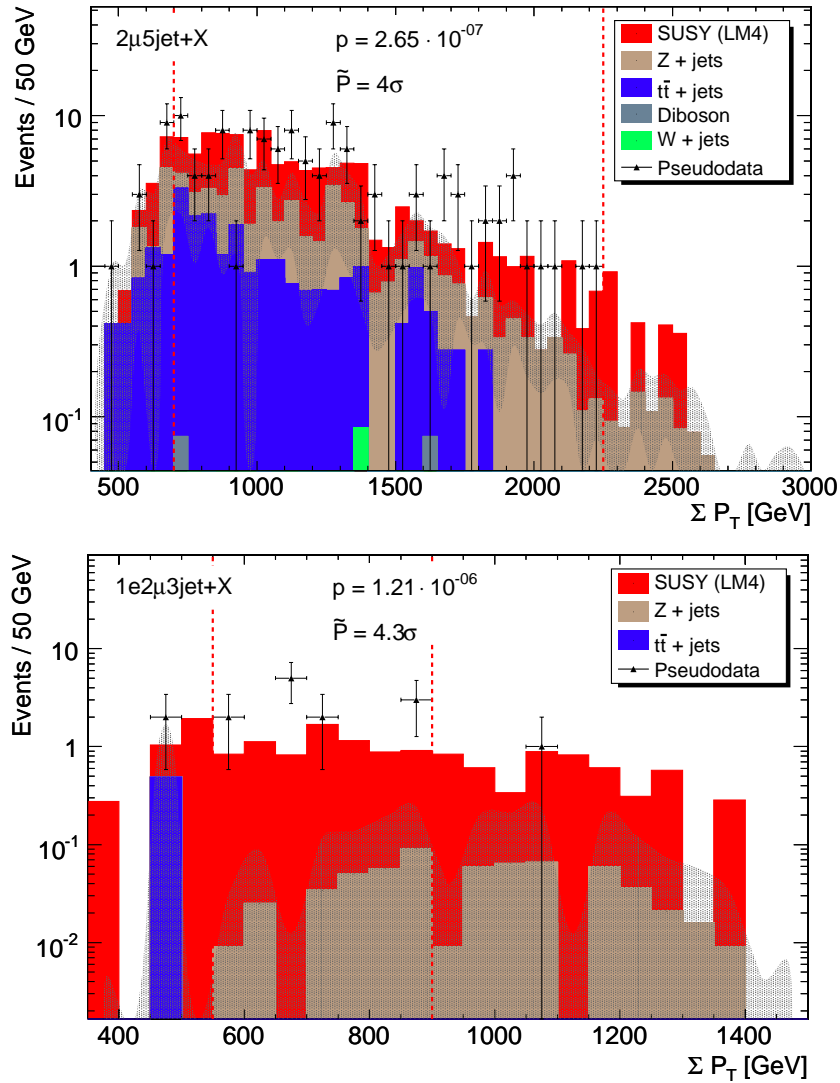
Figure II.6.10: *Results of a representative single pseudo experiment assuming $1\,\mathrm{fb}^{-1}$, using event classes without $\displaystyle{\not}E_{\mathrm{T}}$. The dotted lines indicate the region of interest where the smallest p-value has been found, $\widetilde{P}$ corresponds to the significance assuming this single set of toy data. The shaded area corresponds to the syst. uncertainty of the SM expectation. Numbers for the regions of interest can be found in Table II.6.1.*

II.6.1.5 Statistical interpretation of all event classes together

So far the individual event classes have been interpreted apart from the complete set of events. For each event class a significance $\widetilde{P}$ has been computed which easily can be translated into standard deviations, see Figure II.4.3.

When combining these numerous event classes a final trial factor can be estimated and applied to account for the multiple number of final state topologies looked at. A similar penalty factor must in principle also be used when considering the large number of independent analyses conducted by the whole CMS collaboration.

Conservatively neglecting correlations between the event classes (which is not true for the inclusive ones for sure), the final statistical estimator for the overall degree of agreement with the

Standard Model can be quantified using the formula

$$
\begin{aligned}
P_{\mathrm{CMS}} &= Pr(\widetilde{P} < \widetilde{P}_{\mathrm{single}} \text{ in any of n classes}) \\
&= 1 - Pr(\widetilde{P} > \widetilde{P}_{\mathrm{single}} \text{ in all n classes}) \\
&= 1 - (1 - \widetilde{P}_{\mathrm{single}})^n,
\end{aligned}
\tag{II.6.1}
$$

where $\widetilde{P}_{\mathrm{single}}$ is the significance of a certain event class and n refers to the total number of distributions analyzed. Figure II.6.11 displays this translation for various numbers of event classes considered. As an example, if 1000 classes are used a single *local* 5 σ deviation in a certain topology leads to roughly 3.5 σ considering the *global CMS* dataset.

Note that this global significance P_{CMS} would correspond to a *single* significant deviation found in the context of the many other classes analyzed. This corresponds to the question "if there is a single class with 5 σ, how probable does one get such a single 5 σ or better in any of the event classes when repeating the whole CMS experiment".

As seen in the SUSY example in general one expects several distributions to show a significant deviation. Therefore one could also compute a similar global significance using Binomial statistics for other cases, e.g. four classes with a 3 σ deviation or two classes with a 4 σ effect.
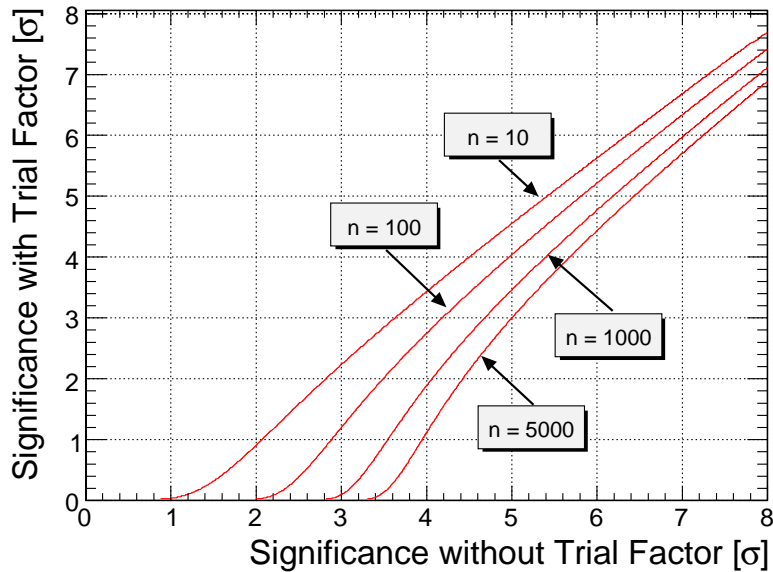

Figure II.6.11: *Translation of significance per event class* $(\widetilde{P})$ *into global-experiment significance after accounting for trial factors* (P_{CMS}) *[84].*

Another approach to quantify the global CMS accordance of data and Standard Model expectation is to plot the frequency distribution of the $\widetilde{P}$ values using all event classes analyzed. In a dataset where no signal beyond the SM is present the $\widetilde{P}$ values should peak at high values. If there is a signal leading to significant deviations in several event classes one would expect the tails of this global distribution to differ from the SM-only case. More entries than expected with small $\widetilde{P}$ should be observed, thus a discrepancy in the tails of this distribution between a SM-only CMS experiment and a CMS dataset including some signal should be seen.

Figure II.6.12 gives two examples for such a distribution, using the $\sum p_{\mathrm{T}}$ distribution in the

exclusive case and in the inclusive case. Here the $\widetilde{P}$ values ($-log_{10}\widetilde{P}$, thus $3 \mathrel{\widehat{=}} 3.3\sigma$) of all event classes with pseudo-data entries are charted in a histogram. The black curve refers to the expectation of a SM-only dataset. Here the distributions of several single CMS experiments without any signal are averaged in order to give a reliable prediction. The points on the other hand correspond to a single CMS experiment assuming SUSY LM4 is realized in nature. One can clearly see that the SUSY contribution leads to significant deviations in numerous classes. Thus one gets many entries at small $\widetilde{P}$ which are not expected from the SM prediction. The discrepancy between SM-expectation and pseudo-data is even more prominent in the inclusive case. Note that classes where only an upper limit can be set ($\widetilde{P} < X$, indicated by the red arrow) all contribute to the very last bin.

Integrating the SM-only curve one can determine a similar estimator as the P_{CMS} discussed above. The tail corresponds to the global trial factor, but again only for a deviation in a single event class.

The discussion of global trial factors above indicate that it might be desirable to constrain the number of distributions looked at to a minimum. In the context of MUSiC it is clear that $\sum p_{\mathrm{T}}$ of all event classes will be scanned for deviations in a generic way minimizing any bias towards a certain model beyond the SM. Including transverse mass, $\not{E}_{\mathrm{T}}$ or possible additional distributions might look promising for certain models, e.g. M_{inv} for Z$'$ or $\not{E}_{\mathrm{T}}$ for SUSY.

An interesting approach to lower the penalty of the trial factor is to use a so called hypothesis ranking [90]. Its feasibility within the MUSiC algorithm is currently under investigation but beyond the scope of this sensitivity study. In general this technique could be a good solution to include additional promising kinematic distributions for certain event classes without blowing up the global trial factor.

II.6.2 Summary and conclusions

An analysis strategy novel in CMS has been introduced, MUSiC (Model Unspecific Search in CMS), which systematically scans various final state topologies for significant deviations from the Standard Model expectation. Only standard selection criteria are utilized, focusing on well understood high-p_{T} objects, without any optimization towards a certain signal. A dedicated algorithm selects the region of biggest discrepancy within a certain distribution and estimates the significance of this deviation, taking into account statistical as well as systematic uncertainties. Especially in the context of LHC start-up such a strategy will be an autonomous cross-check and a good supplement to the CMS physics program since reliable predictions of how physics beyond the Standard Model could appear are difficult. Several representative scenarios of possible deviations in the data have been discussed and it has been demonstrated that MUSiC is sensitive to detector effects and generator issues, gold plated signatures for new physics as well as complex models such as Supersymmetry. As a global physics monitor MUSiC will help to stay alert to all possibilities, be it SUSY or an unexpected signal, or the understanding of the detector and backgrounds.
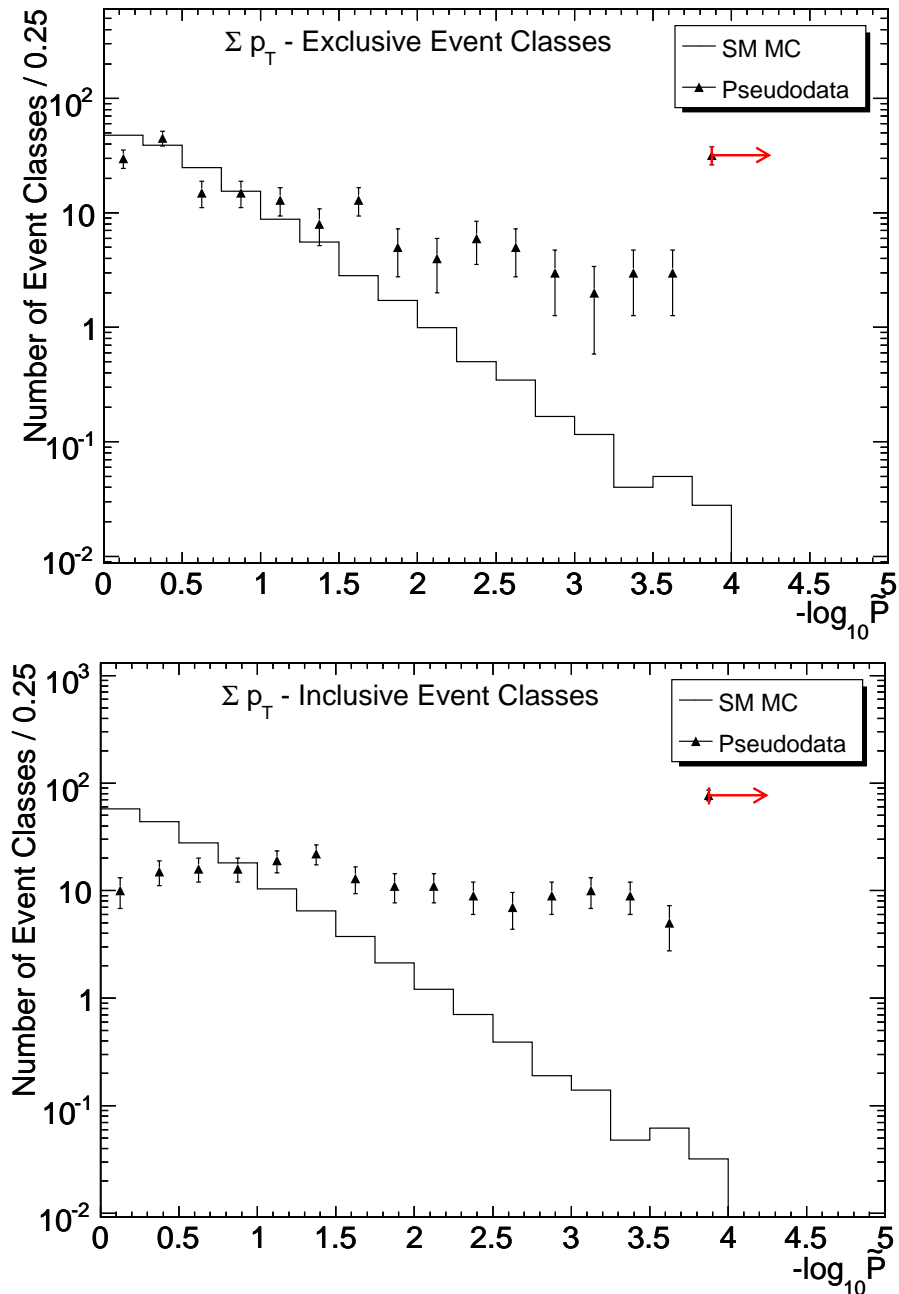
Figure II.6.12: *Frequency distribution of the $\widetilde{P}$ values using all event classes which have pseudo-data entries (exclusive on the top, inclusive on the bottom), using $\sum p_T$ distribution and assuming $1\,\text{fb}^{-1}$. The black curve refers to an averaged CMS experiment with SM-only, the points correspond to a single CMS dataset with SUSY LM4 present.*

Part III

Global Summary

From Construction of an Experiment to Physics Studies Awaiting First Data

The contents of this thesis reflect the evolution and progress of the CMS experiment during the past years. With the assembly of the various detector components at CERN large scale tests could be performed, culminating in the Magnet Test/Cosmic Challenge in 2006. During this cosmic ray data taking period, including the operation of the 4 T magnetic field, for the first time the whole chain from data acquisition to offline data analysis was run through. The first part of this thesis represents the last detail of this analysis chain, comparing the recorded data to expectations from simulation. In this context a dedicated cosmic muon generator, CMSCGEN, has been implemented and validated. Ranging from very basic quantities, e.g. the drift times of DT chambers, up to high level objects such as p_T-measurements of cosmic tracks, a good agreement can be found between simulation and actual measurement. This gives great confidence in our proper understanding of the detector and promises to contribute to a smooth operation during first p-p collisions. In addition to this, cosmic ray muons are an important calibration and alignment source also during LHC operation, thus in this context a cosmic generator will be very useful for various studies.

As already indicated, the commissioning efforts prior to the first collisions are essential to the success of the whole CMS experiment. Especially in searches for new physics leptons are a very clean and attractive signature, and tests of the tracker and muon system using cosmic rays clearly contribute to an efficient lepton identification. In the context of searches in the second part of this thesis a model independent analysis approach, novel in CMS, is introduced: MUSiC – Model Unspecific Search in CMS. In this Ansatz priority is not given to a specific extension of the Standard Model but the events are categorized according to particle content and are analyzed systematically for deviations from the Monte Carlo simulation. Since LHC is entering a new territory in the TeV range and since reliable theoretical predictions of the new physics to be discovered are problematic, this approach promises to be open to surprises and valuable with first data. It has been demonstrated in various examples that MUSiC can contribute to the initial understanding of the detector and the tuning of the simulation in a global way, and that it is sensitive to a wide array of physics beyond the Standard Model. Acting as a global physics monitor MUSiC can serve as a cross-check and add-on to the more dedicated analyses.

To conclude, with the help of cosmic muon ray commissioning CMS is well prepared and eager to take first collision data in 2009. MUSiC can then help to get a global picture of these anticipated data and hopefully contribute to the discovery of new phenomena at the TeV scale; new phenomena which could revolutionize our understanding of the basic laws of nature and the history of the universe.

Appendix A

Muon Flux Parameterization

The procedure outlined in [36] is followed closely, see also [34] for more information: The total muon flux, integrated over the full range of zenith angles (vertical to horizontal) and all azimuthal angles, at the flat CMS surface area (ignoring buildings and other structures) at a certain altitude is parametrized in the form

$$\frac{d\,\Phi}{d\,p} = C_{norm} \cdot \frac{1}{p^3} \cdot s(L) \quad . \tag{A.1}$$

Φ denotes the number of muons per time and area hitting a (horizontal) surface area, p is the muon momentum and $L = \log_{10}(p/\mathrm{GeV})$. The flux normalization constant C_{norm} will be discussed later. With this parametrization of the steeply falling muon momentum spectrum the function $s(L)$ (which one needs to determine) varies rather little with L. The definition (A.1) implies that $s(L)$ can be written as

$$s(L) \sim p^2 \cdot \frac{d\,\Phi}{d\,L} \quad . \tag{A.2}$$

This distribution can be obtained from the histogram of the logarithm L, by applying a weight of $(p/\mathrm{GeV})^2 = 10^{2L}$ to each muon generated by CORSIKA. Since $s(L)$ is a slowly varying function of L it is approximated by a polynomial:

$$s(L) = a_0 + a_1\,L + \ldots + a_6\,L^6 \tag{A.3}$$

It turns out that a polynomial of degree six is sufficient for the purposes in question.

The distribution of the azimuthal angle ϕ is assumed to be flat, the dependence on the momentum dependent zenith angle (here: $\theta = 180^0$ for a vertically down-going muon) is parametrized by the normalized function $z(c, L)$ with $c = \cos\theta$:

$$z(c, L) = b_0(L) + b_1(L)\,c + b_2(L)\,c^2 \tag{A.4}$$

Spherical coordinates with the z axis pointing vertically upwards are used; thus vertical muons correspond to $c = -1$, horizontal ones to $c = 0$; these numbers define the range of c values to be considered. Note that the three coefficients b_i are not independent from each other since the integral $\int z(c, L)$ is one by definition. Also the coefficients b_i depend on the muon momentum.

The global constant C_{norm} can be derived by comparing the predicted muon flux with the measured one, for vertical muons at the reference momentum of $100\,\mathrm{GeV}$.

Thus the full parametrization has the following simple form:

$$\boxed{\frac{d\,\Phi}{dp\,dc\,d\phi} = C_{norm} \cdot \frac{1}{p^3} \cdot s(L) \cdot z(c,L) \cdot \frac{1}{2\pi}} \qquad (A.5)$$

Appendix B

Cross Sections

Cross Sections

Dataset	Cross Section [fb]	Number of Events	Generator	Comments
SUSY LM4	27700	100k	PYTHIA	NLO (Prospino 2)
Hidden Higgs	325	45k	PYTHIA	m = 150 GeV, LO
Hidden Z$'$	365	12k	PYTHIA	m = 1 GeV, LO
Leptoquark	1200	1k	PYTHIA	m = 400 GeV, LO

Table B.1: *Datasets used as benchmark signals in this analysis. The SUSY total NLO cross sections have been obtained by Prospino 2 (using CTEQ6M PDF), while SoftSusy has been used to calculate the mass spectrum and SUSY-Hit to decay the particles. Finally* PYTHIA *is used for the simulation of the parton shower. The number of simulated events entering the analysis is given in the third column.*

Dataset	Cross Section $[fb]$	# of Events	Generator	Comments
W_0jets	$5.07 \cdot 10^7$	8.8M	ALPGEN	Chowder, NLO (k = 1.12)
W_1jets_0_100	$1.03 \cdot 10^7$	9.1M	ALPGEN	Chowder, NLO (k = 1.12)
W_1jets_100_300	$2.87 \cdot 10^5$	247k	ALPGEN	Chowder, NLO (k = 1.12)
W_1jets_300_800	3260	57k	ALPGEN	NLO (k = 1.12)
W_1jets_800_1600	17.8	60k	ALPGEN	NLO (k = 1.12)
W_2jets_0_100	$2.84 \cdot 10^6$	2.4M	ALPGEN	Chowder, NLO (k = 1.12)
W_2jets_100_300	$2.52 \cdot 10^5$	287k	ALPGEN	Chowder, NLO (k = 1.12)
W_2jets_300_800	4530	25k	ALPGEN	NLO (k = 1.12)
W_2jets_800_1600	35.3	54k	ALPGEN	NLO (k = 1.12)
W_3jets_0_100	$6.59 \cdot 10^5$	353k	ALPGEN	Chowder, NLO (k = 1.12)
W_3jets_100_300	$1.20 \cdot 10^5$	118k	ALPGEN	Chowder, NLO (k = 1.12)
W_3jets_300_800	3440	107k	ALPGEN	NLO (k = 1.12)
W_3jets_800_1600	33.7	53k	ALPGEN	NLO (k = 1.12)
W_4jets_0_100	$1.38 \cdot 10^5$	126k	ALPGEN	Chowder, NLO (k = 1.12)
W_4jets_100_300	42300	40k	ALPGEN	Chowder, NLO (k = 1.12)
W_4jets_300_800	1750	29k	ALPGEN	NLO (k = 1.12)
W_4jets_800_1600	21.1	55k	ALPGEN	NLO (k = 1.12)
W_5jets_0_100	84300	62k	ALPGEN	Chowder, NLO (k = 1.12)
W_5jets_100_300	44400	44k	ALPGEN	Chowder, NLO (k = 1.12)
W_5jets_300_800	3390	40k	ALPGEN	NLO (k = 1.12)
W_5jets_800_1600	66.3	17k	ALPGEN	NLO (k = 1.12)
W_enu	$1.92 \cdot 10^7$	1M	PYTHIA	NLO (k = 1.12)

Table B.2: *Standard Model background datasets used in the analysis. The cross section, the generator and the approximate number of events used in the analyses are stated. The comment column contains the k-factor and the "soup" from which the dataset has been extracted.*

Dataset	Cross Section [fb]	# of Events	Generator	Comments
Z_0jets	$5.04 \cdot 10^6$	3.3M	ALPGEN	Chowder, NLO (k = 1.12)
Z_1jets_0_100	$1.04 \cdot 10^6$	945k	ALPGEN	Chowder, NLO (k = 1.12)
Z_1jets_100_300	33600	36k	ALPGEN	Chowder, NLO (k = 1.12)
Z_1jets_300_800	403	30k	ALPGEN	NLO (k = 1.12)
Z_1jets_800_1600	2.24	13k	ALPGEN	NLO (k = 1.12)
Z_2jets_0_100	$3.02 \cdot 10^5$	289k	ALPGEN	Chowder, NLO (k = 1.12)
Z_2jets_100_300	31600	35k	ALPGEN	Chowder, NLO (k = 1.12)
Z_2jets_300_800	616	29k	ALPGEN	NLO (k = 1.12)
Z_2jets_800_1600	4.48	22k	ALPGEN	NLO (k = 1.12)
Z_3jets_0_100	77300	73k	ALPGEN	Chowder, NLO (k = 1.12)
Z_3jets_100_300	14400	24k	ALPGEN	Chowder, NLO (k = 1.12)
Z_3jets_300_800	448	28k	ALPGEN	NLO (k = 1.12)
Z_3jets_800_1600	4.26	16k	ALPGEN	NLO (k = 1.12)
Z_4jets_0_100	15500	33k	ALPGEN	Chowder, NLO (k = 1.12)
Z_4jets_100_300	4700	7k	ALPGEN	Chowder, NLO (k = 1.12)
Z_4jets_300_800	224	25k	ALPGEN	NLO (k = 1.12)
Z_4jets_800_1600	2.8	12k	ALPGEN	NLO (k = 1.12)
Z_5jets_0_100	9740	12k	ALPGEN	Chowder, NLO (k = 1.12)
Z_5jets_100_300	5710	6k	ALPGEN	Chowder, NLO (k = 1.12)
Z_5jets_300_800	504	25k	ALPGEN	NLO (k = 1.12)
Z_5jets_800_1600	8.3	37k	ALPGEN	NLO (k = 1.12)
DrellYan_ee_200	1660	42k	PYTHIA	LO
DrellYan_ee_500	86.2	65k	PYTHIA	LO
DrellYan_ee_1000	7.45	12k	PYTHIA	LO
DrellYan_ee_1500	1.22	4k	PYTHIA	LO
DrellYan_ee_2000	0.26	10k	PYTHIA	LO
DrellYan_mumu_200	1660	42k	PYTHIA	LO
DrellYan_mumu_500	86.2	42k	PYTHIA	LO
DrellYan_mumu_1000	7.45	13k	PYTHIA	LO
DrellYan_mumu_1500	1.22	14k	PYTHIA	LO
DrellYan_mumu_2000	0.26	13k	PYTHIA	LO
$t\bar{t}$_0jets	$6.19 \cdot 10^5$	1.5M	ALPGEN	Chowder, NLO (k = 1.85)
$t\bar{t}$_1jets	$1.76 \cdot 10^5$	362k	ALPGEN	Chowder, NLO (k = 1.85)
$t\bar{t}$_2jets	33700	81k	ALPGEN	Chowder, NLO (k = 1.85)
$t\bar{t}$_3jets	5920	14k	ALPGEN	Chowder, NLO (k = 1.85)
$t\bar{t}$_4jets	1480	5.3k	ALPGEN	Chowder, NLO (k = 1.85)
WW	70000	850k	PYTHIA	LO
WZ	26900	360k	PYTHIA	LO
ZZ	11200	140k	PYTHIA	LO

Table B.3: *More Standard Model datasets used in the analysis. The cross section, the generator and the approximate number of events used in the analyses are stated. The comment column contains the k-factor and the "soup" from which the dataset has been extracted.*

Dataset	Cross Section [fb]	# of Events	Generator	Comments
QCD_0_15	$5.30 \cdot 10^{13}$	14M	PYTHIA	Gumbo, LO
QCD_15_20	$1.46 \cdot 10^{12}$	1.7M	PYTHIA	Gumbo, LO
QCD_20_30	$6.30 \cdot 10^{11}$	2.7M	PYTHIA	Gumbo, LO
QCD_30_50	$1.63 \cdot 10^{11}$	2.5M	PYTHIA	Gumbo, LO
QCD_50_80	$2.16 \cdot 10^{10}$	2.5M	PYTHIA	Gumbo, LO
QCD_80_120	$3.08 \cdot 10^{9}$	1.2M	PYTHIA	Gumbo, LO
QCD_120_170	$4.94 \cdot 10^{8}$	1.3M	PYTHIA	Gumbo, LO
QCD_170_230	$1.01 \cdot 10^{8}$	1.2M	PYTHIA	Gumbo, LO
QCD_230_300	$2.45 \cdot 10^{7}$	1.2M	PYTHIA	Gumbo, LO
QCD_300_380	$6.24 \cdot 10^{6}$	1.2M	PYTHIA	Gumbo, LO
QCD_380_470	$1.78 \cdot 10^{6}$	1.2M	PYTHIA	Gumbo, LO
QCD_470_600	$6.83 \cdot 10^{5}$	1.2M	PYTHIA	Gumbo, LO
QCD_600_800	$2.04 \cdot 10^{5}$	500k	PYTHIA	Gumbo, LO
QCD_800_1000	35100	100k	PYTHIA	Gumbo, LO
QCD_1000_1400	10900	30k	PYTHIA	Gumbo, LO
QCD_1400_1800	1600	30k	PYTHIA	Gumbo, LO
QCD_1800_2200	145	20k	PYTHIA	Gumbo, LO
QCD_2200_2600	23.8	10k	PYTHIA	Gumbo, LO
QCD_2600_3000	4.29	10k	PYTHIA	Gumbo, LO
QCD_3000_3500	0.84	10k	PYTHIA	Gumbo, LO
QCD_3500_inf	0.11	10k	PYTHIA	Gumbo, LO
PhotonJets_0_15	$1.70 \cdot 10^{11}$	300k	PYTHIA	Gumbo, LO
PhotonJets_15_20	$2.57 \cdot 10^{8}$	520k	PYTHIA	Gumbo, LO
PhotonJets_20_30	$1.32 \cdot 10^{8}$	600k	PYTHIA	Gumbo, LO
PhotonJets_30_50	$4.10 \cdot 10^{7}$	510k	PYTHIA	Gumbo, LO
PhotonJets_50_80	$7.20 \cdot 10^{6}$	520k	PYTHIA	Gumbo, LO
PhotonJets_80_120	$1.30 \cdot 10^{6}$	530k	PYTHIA	Gumbo, LO
PhotonJets_120_170	$2.75 \cdot 10^{5}$	560k	PYTHIA	Gumbo, LO
PhotonJets_170_300	87000	200k	PYTHIA	Gumbo, LO
PhotonJets_300_500	8200	30k	PYTHIA	Gumbo, LO
PhotonJets_500_7000	870	30k	PYTHIA	Gumbo, LO
emQCD	$5.34 \cdot 10^{11}$	8.7M	PYTHIA	Stew, LO
emQCD_bb_5_50	$1.70 \cdot 10^{10}$	3.0M	PYTHIA	Stew, LO
emQCD_bb_50_170	$1.65 \cdot 10^{8}$	3.0M	PYTHIA	Stew, LO
emQCD_bb_170_inf	$2.54 \cdot 10^{6}$	2.6M	PYTHIA	Stew, LO
muQCD	$4.40 \cdot 10^{10}$	20M	PYTHIA	Stew, LO
B_JPsi	$4.27 \cdot 10^{7}$	500k	PYTHIA	Stew, LO
BottomOnia_0_20	$3.20 \cdot 10^{7}$	1M	PYTHIA	Stew, LO
BottomOnia_20_inf	$4.25 \cdot 10^{5}$	1M	PYTHIA	Stew, LO
CharmOnia_0_20	$3.27 \cdot 10^{8}$	1M	PYTHIA	Stew, LO
CharmOnia_20_inf	$2.47 \cdot 10^{6}$	1M	PYTHIA	Stew, LO

Table B.4: *More Standard Model datasets used in the analysis. The cross section, the generator and the approximate number of events used in the analyses are stated. The comment column contains the k-factor and the "soup" from which the dataset has been extracted.*

Bibliography

[1] M. Peskin, D. Schroeder. "An Introduction to Quantum Field Theory". (1995).

[2] The Super-Kamiokande Collaboration. "Evidence for Oscillation of Atmospheric Neutrinos". Phys. Rev. Lett. 81 (1998) 1562.

[3] `http://sciencepark.etacude.com/particle/force_unification.gif`.

[4] G. Ross. "Grand Unified Theories". Addison-Wesley (1984).

[5] Lecture by P. Zerwas. "Supersymmetry: Base and LHC Projection".

[6] W-M Yao *et al.*. "Review of Particle Physics". J. Phys. G (2006): Nucl. Part. Phys. 33, Equation 27.12.

[7] S. Weinberg. "Implications of dynamical symmetry breaking: An addendum". Phys. Rev. D19 (1979) 1277.

[8] N. Arkani-Hamed *et al.*. "Electroweak symmetry breaking from dimensional deconstruction". Phys. Lett. B 513 (2001) 232.

[9] E.J. Eichten *et al.*. "New Tests for Quark and Lepton Substructure". Phys. Rev. Lett. 50 (1983) 811.

[10] J. Wudka. "Composite leptoquarks ". Phys. Lett. B 167 (1986) 337.

[11] N. Arkani-Hamed *et al.*. "The hierarchy problem and new dimensions at a millimeter". Phys. Lett. B 429 (1998) 263.

[12] L. Randall and R. Sundrum. "A Large Mass Hierarchy from a Small Extra Dimension". Phys. Rev. Lett. 83 (1999) 3370.

[13] Y.A. Golfand and E.P. Likhtman. JETP Lett. 13 (1971) 323;
D.V. Volkov and V.P. Akulov. JETP Lett. 16 (1972) 438;
D.V. Volkov and V.P. Akulov. Phys. Lett. 46 B (1973) 109;
J. Wess and B. Zumino. Nucl. Phys. B 70 (1974) 39.

[14] V.M. Abazov *et al.* (D0 Collaboration). "Search for squarks and gluinos in events with jets and missing transverse energy using 2.1 fb^{-1} of $p\overline{p}$ collision data at $\sqrt{s} = 1.96$ TeV". Phys. Lett. B 660 (2008) 449.

[15] E. Komatsu *et al.*. "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation". arXiv:0803.0547v2 [astro-ph].

[16] P. Fayet. "Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions". Phys. Lett. B 69 (1977) 489.

[17] The CMS Collaboration. CMS Physics TDR Volume II: "Physics Performance". J. Phys. G: Nucl. Part. Phys. 34 (2007) 995.

[18] shown at DIS2008 in London, March 2008. H1prelim-08-045.
`https://www-h1.desy.de/publications/H1results.html`.

[19] shown at GIF conference in Annecy, 2001.
`http://cmsinfo.cern.ch/outreach/CMSdocuments/JimGIF/GIFVirdee1_index.html`.

[20] L. Evans. "The Large Hadron Collider". New J. Phys. 9 (2007) 335.

[21] current LHC temperature. `http://hcc.web.cern.ch/hcc/field.php`.

[22] The CMS Collaboration. CMS Physics TDR Volume 1: "Detector Performance and Software". CERN/LHCC 2006-001 (Figure 9.3/9.4., Figure 1.3.).

[23] The CMS Collaboration. "The CMS experiment at the CERN LHC". JINST 3 (2008) S08004.

[24] CERN Document Server. `http://cdsweb.cern.ch/`.

[25] J. Knobloch *et al.*. "LHC Computing Grid – Technical Design Report". CERN-LHCC-2005-024 (2005).

[26] The TOTEM Collaboration. "The TOTEM Experiment at the CERN Large Hadron Collider". JINST 3 (2008) S08007.

[27] C. D. Jones *et al.*. "The new CMS data model and framework". CHEP06 Conference Proceedings (2007).

[28] P. Biallass *et al.*. "Simulation of Cosmic Muons and Comparison with Data from the Cosmic Challenge using Drift Tube Chambers". CERN CMS NOTE 2007/024.

[29] P. Biallass *et al.*. "First Measurements of Cosmic Muons with Magnetic Field in CMS". CERN CMS CR-2007/048. Prepared for EPS-HEP2007 conference, published in J. Phys.: Conf. Ser. 110 Volume 110 (2008) 122004.

[30] The CMS Collaboration. The TriDAS Project, Technical Design Report, Volume 1: "The Trigger Systems". CERN LHCC 2000-38 (2000);
The CMS Collaboration. The TriDAS Project, Technical Design Report, Volume 2: "Data Acquisition and High-Level Trigger". CERN LHCC 2002-36 (2002).

[31] V. Drollinger. "Simulation of Beam Halo and Cosmic Muons". CERN CMS NOTE 2005/012.

[32] E. Barberis, P. Biallass, V. Drollinger, K. Hoepfner, D.R. Wood. "Trigger and Reconstruction Studies with Beam Halo and Cosmic Muons". CERN CMS NOTE 2006/012.

[33] E. Barberis, P. Biallass, V. Drollinger, K. Hoepfner, D.R. Wood. "Trigger and Reconstruction Studies with Beam Halo and Cosmic Muons". CMS AN-2005/046.

[34] P. Biallass and T. Hebbeker. "Improved Parametrization of the Cosmic Muon Flux for the Generator CMSCGEN". CMS IN-2007/033.

[35] P. K. F. Grieder. "Cosmic Rays at Earth: Researcher's Reference Manual and Data Book". Amsterdam, Netherlands: Elsevier (2001).

[36] T. Hebbeker, A. Korn. "Simulation Programs for the L3+Cosmics Experiment". L3-C note (1998). `http://web.physik.rwth-aachen.de/~hebbeker/l3csim.pdf`.

[37] O. Adriani *et al.*. "The L3+C detector, a unique tool-set to study cosmic rays". Nucl. Instrum. Meth. A 488 (2002) 209.

[38] D. Heck, J. Knapp *et al.*. "CORSIKA: A Monte Carlo Code to simulate Extensive Air Showers". Forschungszentrum Karlsruhe, Report FZKA 6019 (1998).

[39] P. Achard *et al.* [L3 Collaboration]. "Measurement of the atmospheric muon spectrum from 20-GeV to 3000-GeV". Phys. Lett. B 598 (2004) 15.

[40] M. Unger. "Measurement of the muon momentum spectrum of atmospheric muons with the L3 Detector". PhD thesis, Humboldt University Berlin (2003).

[41] T. Hebbeker and C. Timmermans. "A Compilation of High Energy Atmospheric Muon Data at Sea Level". Astropart. Phys. 18 (2002) 107.

[42] J. Allison *et al.*. "Geant4 developments and applications". IEEE Trans. Nucl. Sci. 53 (2006) 270.

[43] D. Spiga *et al.*. "CRAB: the CMS distributed analysis tool development and design". Nucl. Phys. B – Proceeding Supplements (2008).
`http://cmsdoc.cern.ch/cms/ccs/wm/www/Crab/`.

[44] N. Amapane *et al.*. "Offline Calibration Procedure of the Drift Tube Detectors". CERN CMS NOTE 2007/034.

[45] "The SQLite Web Page". `http://www.sqlite.org/`.

[46] J. Puerta-Pelayo, M.C. Fouz, P. Garcia-Abia. "Parametrization of the Response of the Muon Barrel Drift Tubes". CERN CMS NOTE 2005/018.

[47] "The IGUANA Web Page". `http://iguana.web.cern.ch/iguana/`.

[48] M. Benettoni *et al.*. "CMS DT Chambers: Optimized Measurement of Cosmic Rays Crossing Time in absence of Magnetic Field". CERN CMS NOTE 2008/17.

[49] S. Bolognesi *et al.*. "Measurement of Drift Velocity in the CMS Barrel Muon Chambers at the CMS Magnet Test Cosmic Challenge". CERN CMS NOTE 2008/03.

[50] C. Liu, N. Neumeister. "Reconstruction of Cosmic and Beam-Halo Muons". CERN CMS NOTE 2008/01.

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

[52] R. Frühwirth. "Application Of Kalman Filtering To Track And Vertex Fitting". Nucl. Instrum. Meth. A 262 (1987) 444.

[53] M. Aldaya, P. Garcia-Abia. "Measurement of the charge ratio of cosmic muons using CMS data". CERN CMS NOTE 2008/016.

[54] P. Arce *et al.*. "Bunched beam test of the CMS drift tubes local muon trigger". Nucl. Instrum. Meth. A 534 (2004) 441.

[55] F.R. Cavallo, P. Zotto. "Private communication" (2006).

[56] The CMS Collaboration. "MUSiC – An Automated Scan for Deviations between Data and Monte Carlo Simulation". CMS PAS EXO-08-005.

[57] P. Biallass *et al.*. "MUSiC – An Automated Scan for Deviations between Data and Monte Carlo Simulation". CMS AN-2008/065.

[58] P. Biallass. "MUSiC – A General Search for Deviations from Monte Carlo Predictions in CMS". CERN CMS CR-2009/015. Prepared for DISCRETE '08 conference, to be published in *Journal of Physics: Conference Series*.

[59] C. Hof. PhD thesis in preparation.

[60] B. Abbott *et al.* (D0 Collaboration). "Search for New Physics in e mu X Data at D0 Using Sleuth: A Quasi-Model-Independent Search Strategy for New Physics". Phys. Rev. D 62 (2000) 092004;
V.M. Abazov *et al.* (D0 Collaboration). "A Quasi-Model-Independent Search for New Physics at Large Transverse Momentum". Phys. Rev. D 64 (2001) 012004.

[61] A. Aktas *et al.* (H1 Collaboration). "A general search for new phenomena in ep scattering at HERA". Phys. Lett. B 602 (2004) 14.

[62] T. Aaltonen *et al.* (CDF Collaboration). "Model-Independent and Quasi-Model-Independent Search for New Physics at CDF". Phys. Rev. D 78 (2008) 012002.

[63] T. Hebbeker. "A Global Comparison between L3 Data and Standard Model Monte Carlo – a first attempt". L3 note 2305 (1998).
`http://web.physik.rwth-aachen.de/~hebbeker/l3note_2305.pdf`.

[64] P. Biallass."Model Independent Search for Deviations from the Standard Model at the Tevatron". Diploma thesis, RWTH Aachen (2004).
`http://web.physik.rwth-aachen.de/~hebbeker/theses/biallass_diploma.pdf`.

[65] M.L. Mangano *et al.*. "ALPGEN, a generator for hard multiparton processes in hadronic collisions". JHEP 0307:001 (2003), hep-ph/0206293.

[66] T. Sjostrand, S. Mrenna and P. Skands. "PYTHIA 6.4 Physics and Manual". JHEP 0605:026 (2006), hep-ph/0603175.

[67] B.C. Allanach. "SOFTSUSY: a program for calculating supersymmetric spectra". Comput. Phys. Commun. 143 (2002) 305-331.

[68] J.M. Campbell and R.K. Ellis. "Next-to-leading order corrections to W+2 jet and Z+2 jet production at hadron colliders". Phys. Rev. D 65, 113007 (2002). `http://mcfm.fnal.gov/`.

[69] W. Beenakker, R. Hopker and M. Spira. "PROSPINO: A program for the PROduction of Supersymmetric Particles In Next-to-leading Order QCD". arXiv:hep-ph/9611232.
`http://www.ph.ed.ac.uk/~tplehn/prospino/`

[70] J. Pumplin *et al.*. "New generation of parton distributions with uncertainties from global QCD analysis". J. High Energy Phys. 0207 (2002) 012.

[71] M. Erdmann *et al.*. "Concepts, developments and advanced applications of the PAX toolkit". arXiv:physics/0605063. `http://pax.web.cern.ch/pax/pxl/`.

[72] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". Nucl. Instrum. Meth. A 389 (1997) 81. `http://root.cern.ch/`.

[73] The CMS Collaboration. "Towards a measurement of the inclusive $W \to \mu\nu$ and $Z \to \mu^+\mu^-$ cross sections in pp collisions at $\sqrt{s} = 14$ TeV". CMS PAS EWK-07-002.

[74] The CMS Collaboration. "Measuring Electron Efficiencies at CMS with Early Data". CMS PAS EGM-07-001.

[75] S. Baffioni *et al.*. "Electron Reconstruction in CMS". CERN CMS Note 2006/40.

[76] R. Frühwirth. "Track fitting with non-Gaussian noise". Comput. Phys. Commun. 100 (1997) p. 1-16.

[77] N. Marinelli. "Track finding and identification of converted photons". CMS NOTE 2006/005.

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

[79] The CMS Collaboration. "Performance of Jet Algorithms in CMS". CMS PAS JME-07-003; The CMS Collaboration. "Plans for Jet Energy Corrections at CMS". CMS PAS JME-07-002.

[80] The CMS Collaboration. "$\not{E}_{\mathrm{T}}$ Performance in CMS". CMS PAS JME-07-001.

[81] D. Acosta *et al.*. "CMS High Level Trigger". CERN/LHCC 2007-021 (2007).

[82] R.D. Cousins *et al.*. "Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process". Nucl. Instr. Meth. A 595 (2008) 480–501.

[83] A.L. Read. "Presentation of search results: The CL(s) technique". J. Phys. G 28 (2002) 2693.

[84] C. Hof. "Private communication" (2008).

[85] D. Bourilkov, R. C. Group and M. R. Whalley. "LHAPDF: PDF use from the Tevatron to the LHC". arXiv:hep-ph/0605240.

[86] `https://twiki.cern.ch/twiki/bin/view/CMS/AachenPdfUncertainties`

[87] The methods are discussed in numerous D0 and CDF publications. A recent example is: V. M. Abazov *et al.* (D0 Collaboration). "Evidence for Production of Single Top Quarks and First Direct Measurement of $|Vtb|$". Phys. Rev. Lett. 98 (2007) 181802.

[88] The CMS Collaboration. "Search for high mass resonance production decaying into an electron pair in the CMS experiment". CMS PAS EXO-08-001.

[89] P. Biallass *et al.*. "Search Strategies for mSUGRA in the Muons+Jets+MET Final State". CMS AN-2008/034.

[90] S. D. Biller. "Hypothesis ranking and the context of probabilities in an open-end search". Astropart. Phys. 4 (1996) 285.

Acknowledgments

Concerning the more "private" acknowledgments, first of all apologies for the maybe uninspired style, but after almost 4 years these final words are not about being creative but about being *final*.

Nevertheless, I would like to express deep gratitude to Prof. Thomas Hebbeker for not only supervising and supporting my thesis and my work, but also for giving true advice and guidance to this point. I would like to stress that I always felt very comfortable being entrusted in his care as a PhD student and appreciate very much that he respects my work and my personality as a valuable part of his research group.

In the same manner I would like to thank Prof. Christopher Wiebusch who provided very interesting discussions on the statistics of this thesis and agreed to be my second referee.

After almost 5 years at the institute IIIA there are of course countless people I met and which helped me during this time. Many thanks to my (ex-)office colleagues Lotte Wilke, Carsten Magass, Tanja Rommerskirchen, Jens Frangenheim and Clemens Zeidler for shaping a really nice working atmosphere. Special thanks to Kerstin Hoepfner and Arnd Meyer for always having time and interest to discuss problems and share their experiences. Many thanks to the members of our local CMS group – Stefan Schmitz, Walter Bender, Holger Pieta, Michael Sowa, Oleg Tsigenov, Hendrik Jansen, Metin Ata, Paul Papacz, Daniel Teyssier, Markus Merschmeyer, Lars Sonnenschein, Peter Kreuzer, Hans Reithler, Prof. Martin Erdmann, Andreas Hinzmann, Jan Steggemann and Tatsiana Klimkovich – for all the useful discussions on physics and the countless help in technical problems, ranging from CMSSW over the GRID to PXL. Also many thanks to our colleagues from IIIB for keeping the GRID infrastructure alive and for the nice time skiing in Saas Grund.

Let me repeat again special thanks to Carsten Hof who taught me a lot in computing and programming – thanks for always remaining patient! I really think we both complemented one another perfectly. Special thanks also to Michael Bontenackels for being the good spirit of the institute and for always providing help in any occasion; thanks also for designing the nice MUSiC logo.

As there also is a life next to work, I would like to express honest gratitude to my parents, to my brother Christof and his wife Delia as all of you always supported me, financially and emotionally, and believed in the decisions I had to make.

For more than just support, I would like to thank my friends which I have known now for more than half of my life: Kevin, Marcel, Tim, Mascha and Jennifer. Even though we are starting to be spread around the whole country I really hope we will continue to stay in contact as all the years before.

Then I want to thank my friends from Aachen for all the good times we spent together (almost half of my life): Philip, Sandra, Patrik, Maria, JarJar, Thomas, Nico, Sebastian, Benno, Vanessa, Henning and Verena. And many thanks to all of you again for not only spending the good times together but also for being there during hard times.

I often wonder if I'll ever finish all I've started, and the answer I have found is NO.

No, I will never finish all that I have started because life is about doing, the process

And not the result. Life is about doing whether you want to or not.

MY LIFE!: A constant work in progress and I wouldn't have it any other way.

– lyrics by Set Your Goals

Curriculum Vitae

Personal Data

Name	Philipp Alexander Biallass
Address	Matthiashofstr. 8 52064 Aachen
Phone	+49 (0)179 23 77 38 2 (cellular)
Email	biallass@cern.ch
Birth Date	1. August 1978
Birth Place	Duisburg
Citizenship	German

Education

03/2005 - 03/2009	Ph.D. in Physics, RWTH Aachen University (final grade: summa cum laude)
10/1999 - 02/2005	Diploma in Physics, RWTH Aachen University (final grade: summa cum laude)
08/1989 - 06/1998	Abitur, Steinbart-Gymnasium Duisburg (final grade: 1.3)

Research Interest

Particle physics, search for new phenomena at large collider experiments

Research Experience

Ph.D. Thesis   Participation in the CMS experiment at CERN:
- contributions to the commissioning of the barrel muon system and the Magnet Test/Cosmic Challenge
- 12 weeks stay at CERN during the Magnet Test/Cosmic Challenge
- implementation of cosmic muon simulation and comparisons to first CMS measurements
- detector operation shifts for DT muon system
- feasibility study of a generic physics data analysis, searching for deviations from the Standard Model expectation caused by new physics signals
- title of thesis: "Commissioning of the CMS Muon Detector and Development of Generic Search Strategies for New Physics"
- participation at the 2006 CERN-Fermilab Hadron Collider Physics Summer School
- poster presentation at EPS 2007 conference
- talk at DISCRETE'08 conference
- various talks at the German Physics Society conferences (DPG)

Diploma Thesis	Participation in the DØ experiment at Fermilab: model independent search for new physics based on TEVATRON Run II data (≈ 230 pb^{-1}), title of thesis: "Model Independent Search for Deviations from the Standard Model in Final States with Missing Transverse Energy at the Tevatron"
Summer Student	2 month internship at the BaBar experiment at SLAC in 2003: participation in the electromagnetic calorimeter group of BaBar, installation of a movable scintillator detector to measure the positron-background of the accelerator/detector
Student Project	Participation in the 5th Student Parabolic Flight Campaign (ESA) in 2002, including one year of preparation for a microgravity experiment
Student Job	2001-2003 as student research assistant at the Fraunhofer Institute for Laser Technology

Honors & Awards

3 years scholarship of the DFG Graduate College "Elementary Particle Physics at the TeV Scale" during Ph.D. thesis

Diploma thesis awarded with Friedrich-Wilhelm-Preis 2005 for "outstanding scientific work"

Physics diploma awarded with Springorum-Denkmünze and Schöneborn-Preis for final grade "summa cum laude"

Abitur (German general qualification for university entrance) awarded with Dr.-Hans-Peter-Wüst-Preis for graduating as best of the class

Publications (only excerpt)

P. Biallass *et al.*, "Trigger and reconstruction studies with beam halo and cosmic muons", CERN-CMS-NOTE-2006-012 (contribution to "Technical design report" CMS-TDR-008-1)

P. Biallass *et al.*, "Simulation of Cosmic Muons and Comparison with Data from the Cosmic Challenge using Drift Tube Chambers", CMS-NOTE-2007-024

P. Biallass *et al.*, "MUSiC – An Automated Scan for Deviations between Data and Monte Carlo Simulation", CMS PAS EXO-2008-005

Other Skills

Languages	English, French (spoken and written) Latin
Programming	Java, C, C++
Operating Systems	Windows, Linux
Miscellaneous	ROOT, Powerpoint, LaTeX, Emacs
Teaching	supervision of lab courses and diploma students

Aachen, March 29, 2009