Model Unspecific Search for New Physics with b-Hadrons in CMS

von

Erik Dietz-Laursonn

Diplomarbeit in Physik

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen

im Oktober 2010

angefertigt am

III. Physikalischen Institut A

bei

Prof. Dr. Thomas Hebbeker

Zweitgutachter

Prof. Dr. Christopher Wiebusch

Es ist keine Schande, nichts zu wissen, wohl aber, nichts lernen zu wollen.
Sokrates (470 - 399 v. Chr.), griechischer Philosoph

Wissenschaft und Kunst gehören der Welt an, und vor ihnen verschwinden die Schranken der Nationalität.
Johann Wolfgang von Goethe (1749 - 1832), deutscher Dichter

Abstract

Since the end of 2009, the LHC collider at CERN is accelerating protons and the large multi purpose detector CMS is taking collision data. The purpose of CMS is to perform precision measurements of Standard Model physics as well as searching for new physics, both at energy scales that could not be reached by collider experiments before. Many theories of how new physics at high energies might look like have been proposed. Together with possible models, that nobody has thought of yet, these theories of new physics result in an almost infinite number of possible experimental signatures, that should be investigated within the data. This cannot possibly be achieved by dedicated analyses, searching for specific signals only.

Model Unspecific Search in CMS (MUSiC) provides a complementary approach by generally scanning the data for deviations from the Standard Model, without restricting to a specific theory or signal. To achieve this, event classes are investigated, containing events with a similar object content. Taking into account statistical and systematic uncertainties, potential deviations are analysed and their significance is determined.

This work presents the MUSiC framework (which was developed at RWTH Aachen University) as well as an approach to implement the identification of b-hadrons into this framework. This approach has been performed, as b-hadrons play an important role in several precision measurements of the Standard Model as well as in many theories of new physics at high energies. Test scenarios with b-hadrons are presented, in order to determine the benefit of b-hadron identification with respect to the Model Unspecific Search in CMS.

Zusammenfassung

Seit Ende 2009 werden im LHC Teilchenbeschleuniger am CERN Protonen zur Kollision gebracht und deren Daten mit dem Vielzweck-Detektor CMS aufgezeichnet. Das Ziel von CMS ist die Durchführung von Präzisionsmessungen innerhalb des Standard Modells sowie die Suche nach neuer Physik, beides in Energiebereichen, die noch von keinem Beschleunigerexperiment zuvor erreicht wurden. Es wurden viele Theorien darüber vorgeschlagen, wie sich solche neue Physik äußert. Zusammen mit möglichen Modellen, welche bisher nicht erdacht wurden, führen diese Theorien zu einer nahezu unendlichen Anzahl an möglichen experimentellen Signaturen, welche in den Daten untersucht werden sollten. Dies ist nur mit dedizierten Analysen, welche ausschließlich nach speziellen Signalen suchen, nicht zu bewerkstelligen.

Die modellunabhängige Suche in CMS (MUSiC) bietet dadurch einen komplementären Ansatz, dass die Daten nach Abweichungen vom Standard Modell durchsucht werden, ohne sich auf eine spezielle Theorie oder Signatur zu beschränken. Dazu werden so genannte Ereignisklassen untersucht, welche Ereignisse mit ähnlichem Objektinhalt umfassen. Potentielle Abweichungen werden analysiert und ihre Signifikanz wird unter Berücksichtigung statistischer und systematischer Unsicherheiten bestimmt.

Diese Arbeit stellt sowohl das MUSiC Programmpaket vor (welches an der RWTH Aachen entwickelt wurde) als auch den Ansatz, die Identifikation von b-Hadronen in dieses Programmpaket zu implementieren. Dieser Ansatz wurde verfolgt, da b-Hadronen sowohl in vielen Präzisionsmessungen des Standard Modells als auch in vielen Theorien neuer Physik bei hohen Energien eine wichtige Rolle spielen. Zur Bestimmung des Nutzens der Identifikation von b-Hadronen für die modellunabhängige Suche in CMS werden Testszenarios mit b-Hadronen präsentiert.

Contents

1. Introduction

1.1. Useless Fundamental Research?

High energy particle physics is fundamental research. Its aim is to find out "what holds the world together in its inmost folds"[1], i.e. which elementary particles matter consists of and which elementary forces hold them together. According to today's knowledge of particle physics, matter can be described very well by the *Standard Model* of particle physics (see chapter 2.1), which contains *quarks* and *leptons* as the (currently) indivisible particles and describes their interactions by four fundamental forces (*electromagnetic*, *weak*, *strong*, *gravitational*). But it also has its weaknesses: First of all, the gravitational interaction is not yet properly described by the Standard Model. Moreover, an additional particle (the Higgs particle) has to be postulated to explain the existence of the particles' mass, but the existence of the Higgs particle could not yet be proven. Additionally, many hints have been found, that there has to be some physics beyond the Standard Model, e.g. it is clear today that we only know about 17 %[2] of the matter in the visible universe. That means that 83 % of the matter of the universe is unknown and this *Dark Matter* cannot be described by the Standard Model. To solve such problems huge particle colliders like the LHC[3] at CERN[4] are constructed (see chapter 3).

But what is the general use of fundamental research in particle physics? Why should people care about learning more precisely what matter consists of, if it does not seem to effect them anyway? In these times, when the economy all over the world just begins to recover from a global crisis, we (the scientist working for CERN) have to answer to the question, why to invest billions of Euros for a scientific project which has no practical use for the industry or the people. At the same time the German "science" journal 'Bild' (supported by an "expert", biochemist with a medical degree), is warning the people of black holes which would be created by the LHC and would destroy earth. This warning even culminates in a lawsuit at the Federal Constitutional Court of Germany against the LHC-experiment [2].

In this context, considered superficially, it seams irresponsible to run an experiment like the LHC. But if you dig deeper, you discover that it is not like that. First of all: Yes, under certain preconditions, the creation of so-called "Mini Black Holes" is theoretically possible at the LHC. But under the same preconditions, they would nearly instantly decay into

[1] "Faust", Johann Wolfgang von Goethe

[2] $\Omega_\mathrm{m} \approx 0.24$, $\Omega_\mathrm{b} \approx 0.0425$, $\Omega_\mathrm{dm} \approx 0.20$ [1], Ω_m is the overall matter density of the Universe, Ω_b is the matter density of the known, baryonic matter and Ω_dm is the matter density of the unknown dark matter.

[3] **L**arge **H**adron **C**ollider

[4] former "**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire", today "European Organisation for Nuclear Research"

elementary particles. Consequently, the same theory that postulates the creation of "Mini Black Holes" also postulates their harmlessness. Additionally, a simple fact makes clear that there is no imminent danger that the LHC-experiment will destroy the earth: Since billions of years, cosmic particles are colliding with particles of the terrestrial atmosphere. During these collisions, much more energy is released than during LHC collisions. And nevertheless we are still here.

The financial argument against fundamental research is not tenable either. In most cases there may be no direct practical use of the results of fundamental research for the industry or the people's everyday's life. But often the results become very important later: When Demokrit[5] or Rutherford[6] wondered about the structure of matter, no practical use of the results was in sight. But even if their theories had to be adapted to the results of newer experiments, e.g. quantum mechanical effects or the existence of *protons*, *neutrons* and later *quarks* as constituents of the nucleus, they still somehow form the base of today's modern physics and modern technical devices. And to give some concrete examples of direct or indirect use of particle physics research:

- Scintigraphy, Single Photon Emission Computed Tomography (SPECT) and especially Positron Emission Tomography (PET) are diagnostic methods in nuclear medicine, which use the knowledge of particle physics about radioactive decays and building of detector systems to diagnose for example metabolic or cancerous diseases.

- Important parts of the Internet as it is known today, were invented/developed by the British physicist Tim Berners-Lee at CERN in 1989, in order to ease communication and data transfer between people working for CERN throughout the whole world. And today, public life would nearly be unthinkable without it.

Now, with the beginning of the collisions at the LHC on November 23, 2009 [3] and the 7 TeV collisions since March 30, 2010 [4], an exciting time has begun. Within the next few years we will gather the data giving us the opportunity to gain a deeper knowledge of what matter is made of and how the universe formed. And with a little bit of luck, we will be able to answer fundamental questions of today's physics like "Why does matter have mass?", "What is the mysterious 'Dark Matter'?", "Are there more dimensions than the known three?" or "Are all fundamental forces different aspects of only one?" and questions that have always been asked by humans like "Where do we come from?".

The present diploma thesis is meant to do its small part to the goal of finding new physics and solving fundamental questions, by upgrading the MUSiC analysis tool which was developed at RWTH Aachen University. It describes the implementation of b-tagging into MUSiC and the testing of its functionality (chapter 6) and then give some examples of the thus gained improvement (chapter 7).

Again, the solution of these scientific questions will probably not be of use in everyday life. But who knows, which practical uses of the results or the infrastructure of the LHC-experiment humanity will benefit from in 20, 50 or 100 years, in addition to the scientific advance.

[5] Greek natural philosopher (460 - 371 B.C.) who coined the term "atom" as the smallest and indivisible parts of matter.

[6] New Zealand physicist (1871 - 1937) who created the first atomic model with *nucleus* and *electrons*.

1.2. Annotation

In this document the common *natural units* of particle physics (Heaviside-Lorentz units [5])

$$\hbar \equiv c \equiv \epsilon_0 \equiv \mu_0 \equiv 1 \tag{1.1}$$

are used. This leads to equations which tend to be clearer than they are when using SI units. For example:

$$E = \gamma m c^2 = \sqrt{m^2 c^4 + p^2 c^2} \quad \rightarrow \quad E = \gamma m = \sqrt{m^2 + p^2}. \tag{1.2}$$

Another consequence is that the important quantities have the same unit:

$$[length]^{-1} = [time]^{-1} = [mass] = [momentum] = [energy] = \text{eV}. \tag{1.3}$$

Therefore, "eV" is the common unit of particle physics, where 1 eV is the energy a particle with charge $Q = 1\,\text{e} = 1.602 \cdot 10^{-19}\,\text{C}$ gains when traversing a potential difference of 1 V.

To distinguish the "normal" three dimensional vectors from the relativistic, four dimensional Lorentz vectors the following convention is used: The symbol for a three dimensional vector is $\boldsymbol{p} = (p_x, p_y, p_z)$ with the absolute value $|\boldsymbol{p}| = p$, the symbol for a Lorentz vector is $\underline{\boldsymbol{p}} = (E, \boldsymbol{p})$ with the absolute value $|\underline{\boldsymbol{p}}|$.

Additionally the *Einstein notation*

$$a^i b_i \equiv \sum_i a^i b_i \tag{1.4}$$

is used in this document: If an index variable appears as an upper and as a lower index in one single term, this term is to be summed up over all values of this particular index. Thereby, Greek letters represent Lorentz indices.

Electrical charges are always given in units of the *elementary charge* $\text{e} = 1.602 \cdot 10^{-19}\,\text{C}$, i.e. the electrical charge of electrons is -1, meaning $-1\,\text{e} = -1.602 \cdot 10^{-19}\,\text{C}$.

If not specified differently, the name of the particles also includes their anti-particles, e.g. when talking about electrons also positrons are meant. Therefore, the charge is not mentioned either (e.g.: $\mu = \mu^{\pm}$).

In Feynman diagrams, the abscissa is the time-axis. Furthermore, Feynman's convention on the labelling of the particles is used: Fermions moving in positive direction of time (i.e. to the right) are particles and fermions moving in negative direction of time are anti-particles.

2. Theoretical Foundations

2.1. The Standard Model of Particle Physics

The question, what matter consists of, has been asked by humans for a long time [6, 7]. The idea that matter could consist of some sort of smallest, indivisible particles came up in ancient Greek for the first time. Around 400 B.C., the Greek philosopher Demokrit postulated that different combinations of diverse types of massive *atoms* form the different types of matter. Of course, at this time there was no possibility to prove such theories experimentally. Only about 2200 years later, at the beginning of the 19th century, John Dalton could show that the mass of molecules is a multiple of discrete masses, the nuclear masses of the atoms of the molecules. From that time on, better experimental devices were developed which allowed a deeper look into the structure of matter. Since the scattering experiment of Rutherford in 1911 [8], it is known that atoms are not as indivisible as they were thought of, but consist of a small *nucleus*, which carries nearly the whole mass, and *electrons* (which were discovered in 1897 by Thomson). Later, in 1919 and 1932, it was found that the nuclei are not indivisible either, but are composed of *protons* and *neutrons*. Finally in 1968 deep inelastic scattering experiments at the SLAC[7] proved that even protons and neutrons are not the smallest parts of the atomic nucleus, but are formed of *quarks* [9].

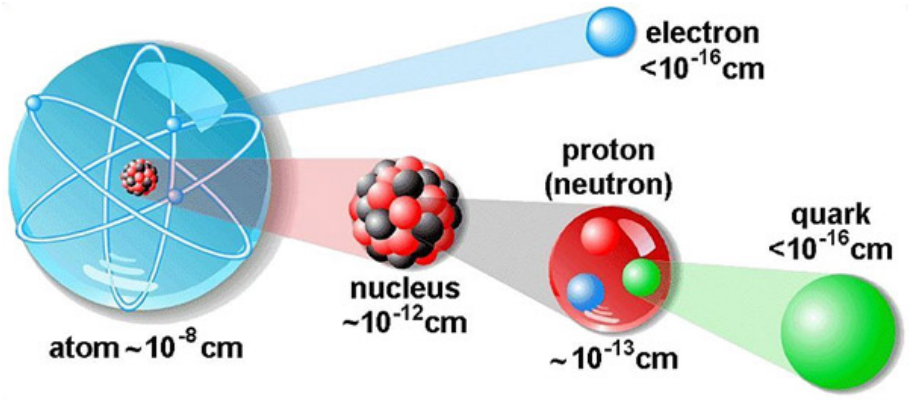


Figure 2.1.: Schematic structure of atoms [10].

Theoretical physics showed very early that the classical theories were unsuitable to describe physics at very small dimensions and distances and that new theories had to be developed. The result are *quantum mechanics* (Schrödinger et al., e.g. [11, 12]), *relativistic quantum mechanics* (Dirac et al.) and *quantum field theories* (e.g. [13]), which form the theoretical

[7] **S**tanford **L**inear **A**ccelerator **C**enter

foundation of the Standard Model of particle physics. As it provides the possibility to describe the composition of matter and three of the four known types of interactions between it very precisely, the Standard Model is one of the greatest achievements in physics in the last century and provides the basis on which to search for new physics.

2.1.1. Particle Zoo of the Standard Model

According to today's knowledge, all known matter of the visible universe consists of 12 point-like particles, *fermions*[8], which can be divided into two subclasses: *leptons* (electron e^-, electron neutrino ν_e, muon μ^-, muon neutrino ν_μ, tauon τ^-, tauon neutrino ν_τ), which can occur isolated, and *quarks* (up u, down d, charm c, strange s, top t, bottom b), which cannot be observed as isolated particles, due to the special properties of their *colour charge*: In nature exist only colourless isolated particles. This means that quarks can only be observed as particles consisting of more than one quark, which are called *hadrons*. One talks of the *Confinement* of quarks to hadrons.

The fermions interact via four basic forces: the *electromagnetic*, the *weak* and the *strong* force as well as *gravitation*. One should note that not every particle can take part in every interaction, e.g. an uncharged particle in not able to interact electromagnetically (cf. table 2.2). While gravitation cannot be described by the Standard Model (see chapter 2.1.7), it is possible to describe the interactions caused by the other three forces as an exchange of 12 *bosons*[9] (photon γ, W^+, W^-, Z^0, eight gluons with a different non-white combination of *colour*- and *anti-colour-charge* for each of them). The colour charge is a quantum number unique to the strong force, just like the electric charge to the electromagnetic force.

Gauge Bosons (spin = 1)				
	γ (photon)	W^+/W^-	Z^0	g (gluon)
Interaction	electromagnetic	weak	weak	strong
Interaction Partners	electric-charged particles	all fermions	all fermions	colour-charged particles
Electric Charge / e	0	+1 / −1	0	0
Colour Charge	-	-	-	8 combinations
Mass	0	$\approx 80.4\,\text{GeV}$	$\approx 91.2\,\text{GeV}$	0
Range	∞	$\sim 0.001\,\text{fm}$	$\sim 0.001\,\text{fm}$	$\sim 1\,\text{fm}$

Table 2.1.: The Gauge Bosons of the Standard Model an their properties [1, 14]. Since only massive particles can decay, the massless photon has an infinite range. In contrast the massless gluon has a very small effective range which is caused by the colourlessness of observable particles.

[8] Particles with half-integral spin, named after the Italian physicist Enrico Fermi who developed their quantum statistics.

[9] Particles with integral spin, named after the Indian physicist Satyendranath Bose who developed their quantum statistics.

The terms 'colour charge' and 'red', 'green', 'blue','anti-red', 'anti-green', 'anti-blue' for its possible values were chosen because of the analogy to the additive mixture of colours in chromatics: The combination of all three colour charges, all three anti-colour charges or a colour charge and its anti-colour charge result in an uncharged ("white") state. Therefore, the possible hadrons, which quarks can be combined to, are *mesons* (1 quark + 1 anti-quark, e.g. $\pi^+ = u\overline{d}$) and *baryons* (3 quarks or 3 anti-quarks, e.g. proton = uud, neutron = udd). Theoretically, also other combinations might be possible, e.g. *pentaquarks* (3 (anti-)quarks + 1 quark/anti-quark pair) or *glueballs* (a particle formed of self-interacting gluons), none of them yet discovered (e.g. cf. [15, 16]). Table 2.1 summarises the properties of the gauge bosons belonging to the three interactions described by the Standard Model.

Both, leptons and quarks, take part in weak interactions and can be grouped into three generations[10], due to their symmetry with respect to the weak force:

$$\underset{\text{I}}{\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}} \quad \underset{\text{II}}{\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}} \quad \underset{\text{III}}{\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}} \qquad \underset{\text{I}}{\begin{pmatrix}u\\ d\end{pmatrix}} \quad \underset{\text{II}}{\begin{pmatrix}c\\ s\end{pmatrix}} \quad \underset{\text{III}}{\begin{pmatrix}t\\ b\end{pmatrix}}. \tag{2.1}$$

While electromagnetic and strong interactions cannot change the type of a particle, the charged weak interaction can convert a "down-type" particle into an "up-type" particle and vice versa[11]. For instance a μ can convert into a ν_μ by emitting a W-boson. Due to this fact, the heavy charged leptons are able to decay into lighter ones as shown in figure 2.2. This results in a limited lifetime of muons and tauons, only the electron is stable because there is no possible decay into lighter particles. For the same reason all three neutrinos are stable.

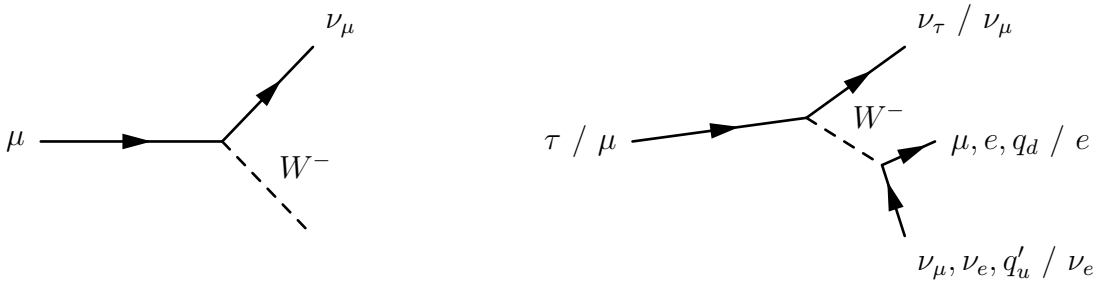


Figure 2.2.: Feynman diagrams of weak interactions: Conversion of a μ^- into a ν_μ by emitting a W-boson (left) and decay of a heavy charged lepton (right). In contrast to muons, tauons can also decay hadronically, since they have a much higher mass than those lightest hadron ($m_{\pi^0} \approx 135\,\text{MeV}$ [1]). q is only representing the quarks, which can be formed in this process for kinematic reasons.

Leptons can only convert into the lepton belonging to the same generation, e.g. the process $\mu \to \nu_\mu + W$ is possible whereas $\mu \to \nu_e + W$ is not possible. As far as quarks are concerned, things are different. When emitting a W-boson, an "up-type" quark can convert into a "down-type" quark of the same or another generation, i.e. the processes $t \to b+W$, $t \to s+W$ and $t \to d + W$ are all possible. This is the reason, why all four quarks of the higher

[10] More precisely, only fermions with left-handed helicity (projection of the spin $\vec{s}$ of a particle on the direction of its momentum: $\lambda = \vec{s}\frac{\vec{p}}{p}$; λ(left-handed fermions) $= -\frac{1}{2}$) form these doublets, since no right-handed neutrinos have experimentally been found (cf. chapter 2.1.5).

[11] respecting the laws of conservation, like conservation of lepton number or charge

generations (and in particular the b-quark, as shown in figure 2.3) are able to decay, unlike the leptonic neutrinos which are all stable.

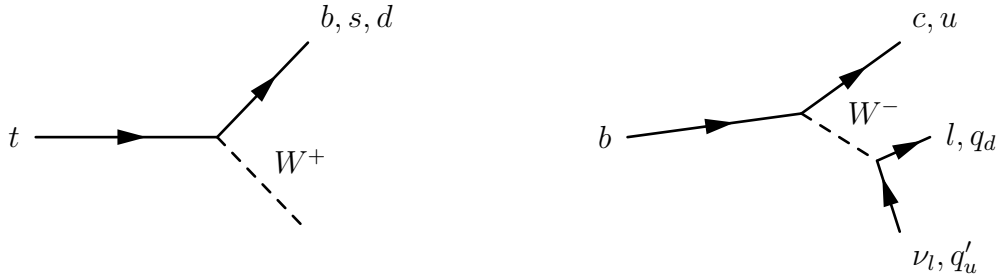


Figure 2.3.: Feynman diagrams of weak (*flavour-changing*[12]) interactions: Conversion of a t-quark into a b-, s-, or d-quark by emitting a W-boson (left) and decay of a b-quark (right). For kinematic reasons, l is standing for all charged leptons, while q is only representing the quarks of the first two generations.

This special behaviour of quarks in weak flavours-changing interactions can be summarised by the *CKM-matrix*[13] [1] or quark mixing matrix:

$$\begin{aligned} V_{CKM} &= \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \\ &= \begin{pmatrix} 0.97419 \pm 0.00022 & 0.2257 \pm 0.0010 & 0.00359 \pm 0.00016 \\ 0.2256 \pm 0.0010 & 0.97334 \pm 0.00023 & 0.0415^{+0.0010}_{-0.0011} \\ 0.00874^{+0.00026}_{-0.00037} & 0.0407 \pm 0.0010 & 0.999133^{+0.000044}_{-0.000043} \end{pmatrix}. \end{aligned} \tag{2.2}$$

The squares of its entries equal the conversion probabilities of the quarks, e.g. $|V_{ud}|^2$ is the probability of a d-quark converting into a u-quark (or vice versa) during a weak flavours-changing interaction. As one can see, the CKM-matrix is almost diagonal, therefore, conversion to other generations is highly suppressed, e.g. the t-quark converts nearly exclusively into a b-quark. For the same reason the b-quark (and thus hadrons contain b-quark) have a comparably long lifetime, since the lighter b-quark cannot covert into a heavier t-quark. Thus, only the conversion into a c-quark or a u-quark is possible, both of which are highly suppressed.

Table 2.2 summarises the properties of the fermions. As mentioned before, all fermions take part in weak interactions. As all quarks and all "down-type" leptons (e, μ, τ) carry an electric charge, they can also interact via the electromagnetic force. Only the quarks participate in strong interactions, due to the fact that they exclusively carry a colour charge.

As neutrinos carry neither an electric nor a colour charge, they can only take part in weak interactions. Due to the very small range of the weak interaction (cf. Table 2.1), they can fly very long distances in matter without interacting. For this reason, they are very difficult

[12] While each of the six types of quarks possesses a different flavour quantum number, for leptons only three different flavours exist (one for each generation). Thus, "flavour-changing" is normally referred to conversions between the generations and not within.

[13] **C**abibbo-**K**obayashi-**M**askawa matrix, named after its developers Nicola Cabibbo, Makoto Kobayashi and Toshihide Maskawa.

Fermions (spin = 1/2)					
			Charges		
	Name	Flavour	Electric / e	Colour	Mass
Leptons (interactions: weak for ν_l, electromagnetic and weak for l)	e^--neutrino	$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}$	0	-	$< 2\,\text{eV}$
	electron		-1	-	$\approx 0.511\,\text{MeV}$
	μ^--neutrino	$\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}$	0	-	$< 0.19\,\text{MeV}$
	muon		-1	-	$\approx 105.7\,\text{MeV}$
	τ^--neutrino	$\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}$	0	-	$< 18.2\,\text{MeV}$
	tauon		-1	-	$\approx 1777\,\text{MeV}$
Quarks (all interactions)	up	$\begin{pmatrix}u\\ d\end{pmatrix}$	$+2/3$	rgb	$2.55^{+0.75}_{-1.05}\,\text{MeV}$
	down		$-1/3$	rgb	$5.04^{+0.96}_{-1.54}\,\text{MeV}$
	charm	$\begin{pmatrix}c\\ s\end{pmatrix}$	$+2/3$	rgb	$1.27^{+0.07}_{-0.11}\,\text{GeV}$
	strange		$-1/3$	rgb	$105^{+25}_{-35}\,\text{MeV}$
	top	$\begin{pmatrix}t\\ b\end{pmatrix}$	$+2/3$	rgb	$171.3 \pm 2.3\,\text{GeV}$
	bottom		$-1/3$	rgb	$4.20^{+0.17}_{-0.07}\,\text{GeV}$

Table 2.2.: The fermions of the Standard Model and their properties [1, 14]. Since quarks cannot be observed as isolated particles but only in hadronised state, their masses have to be measured with different indirect methods. Detailed information about these methods can be found in [14]. For the same reason the smallest observable electric charge remains the elementary charge, which equals the charge of the charged leptons.

to detect and it needs huge detectors to measure neutrinos (e.g. like Super-Kamiokande[14]). Therefore, they only appear as missing energy in the detectors of colliders like the CMS detector (see Chapter 3.3). Since neutrinos are that hard to detect, it is also hard to determine their mass. In direct experiments (e.g. β-decay experiments) only upper limits can be found ($m_{\nu_e} < 2\,\text{eV}$ [1]). This is consistent with the theoretical description of the Standard Model, which does not allow massive neutrinos. But according to today's knowledge the masses of at least two of the neutrino flavours have to differ from zero. Otherwise, the effect of *neutrino oscillation*, that was for example measured at Super-Kamiokande [18], could not be explained, since its theory demands mass differences between the three types of neutrinos.

Additionally, an *anti-particle* exists for every fermion, which has the same nonadditive quantum numbers (like spin and mass) but the opposite additive quantum numbers (like lepton number, baryon number and charges). For example, the *positron* e^+ is the anti-particle of the electron e^- and the $\overline{u}$ is the anti-particle of the u-quark. They have the same mass and spin as the e^-/u but their charge is $+1$ / $-2/3$.

[14] A detector in Japan, with 50000 t of pure water and more than 13000 photo-multipliers, built especially for neutrino detection [17].

2.1.2. The Formalism of Quantum Field Theory

The mathematical background for the theoretical description of the Standard Model is Quantum Field Theory. It describes every particle (fermions and gauge bosons) as a quantised field and uses the following three concepts to achieve a model of matter and its interactions, which is consistent with the most experimental observations:

Lagrange Formalism

In classical mechanics, the equations of motion of the process to be described can be derived from its Lagrange function $L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)$[15], using *Hamilton's principle of stationary action*. This leads to the classical *Euler-Lagrange equation*:

$$\delta S = \delta \int_{t_1}^{t_2} L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)\ \mathrm{d}t \stackrel{!}{=} 0 \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)}{\partial \dot{\boldsymbol{q}}} - \frac{\partial L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)}{\partial \boldsymbol{q}} = 0\,. \tag{2.3}$$

In (quantum) field theory, the fields ϕ are described by their Lagrange density function $\mathcal{L}$ (Lagrangian) and the corresponding relativistic ansatz

$$\delta S = \delta \int_{t_1}^{t_2} \int \mathcal{L}\Big(\phi(\underline{\boldsymbol{x}}), \partial_\mu \phi(\underline{\boldsymbol{x}}), t\Big)\ \mathrm{d}^3x\ \mathrm{d}t \stackrel{!}{=} 0 \tag{2.4}$$

leads to the Euler-Lagrange equation for the Lagrange density function $\mathcal{L}$ of the described field ϕ:

$$\partial_\mu \left(\frac{\partial \mathcal{L}\Big(\phi(\underline{\boldsymbol{x}}), \partial_\mu \phi(\underline{\boldsymbol{x}})\Big)}{\partial\Big(\partial_\mu \phi(\underline{\boldsymbol{x}})\Big)} \right) - \frac{\partial \mathcal{L}\Big(\phi(\underline{\boldsymbol{x}}), \partial_\mu \phi(\underline{\boldsymbol{x}})\Big)}{\partial \phi(\underline{\boldsymbol{x}})} = 0\,. \tag{2.5}$$

Here ∂_μ represents for the relativistic equivalent of the nabla operator $\boldsymbol{\nabla}$ with

$$\partial_\mu = \left(\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right) = \left(\frac{\partial}{\partial t}, \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right) \quad \text{and} \quad \partial^\mu = \left(\frac{\partial}{\partial t}, -\boldsymbol{\nabla}\right). \tag{2.6}$$

Now a major theoretic problem is to find a Lagrangian that describes the properties and propagation of the particles correctly and at the same time satisfies the Euler-Lagrange equation. Such Lagrangians would then be able to describe the interactions between particles and the bosons transmitting these.

Gauge Principle

The basis of the gauge principle is the *Noether Theorem*[16]. It says, that every (global) mechanical symmetry of the Lagrangian corresponds to a conservation law[17]. When it comes to

[15] The Lagrange formalism of classical mechanics uses the Lagrange function $L = T - U$ to completely describe physical systems, using their kinetic energy T and generalised potential U.

[16] "If a system has a continuous symmetry property, then there is a corresponding conserved quantity", named after the German mathematician Emmy Noether who proved it.

[17] e.g.: symmetry under global translation of space-time $\underline{\boldsymbol{x}} \rightarrow$ conservation of relativistic momentum $\underline{\boldsymbol{p}}$; symmetry under global rotation $\rightarrow$ conservation of angular momentum $\boldsymbol{L}$.

gauge symmetries (i.e. symmetries under phase transformations), it is not that straightforward anymore: Now one can either state that an invariance of the Lagrangian under **global** phase transformations (with a generator χ and a real factor q)

$$\psi(\underline{\boldsymbol{x}}) \rightarrow \psi'(\underline{\boldsymbol{x}}) = e^{iq\chi}\psi(\underline{\boldsymbol{x}}) \tag{2.7}$$

corresponds to a conserved current and thus to the conversation of the corresponding charge. Or one can state the same but using **local** phase transformations (where the transformation depends on the position):

$$\psi(\underline{\boldsymbol{x}}) \rightarrow \psi'(\underline{\boldsymbol{x}}) = e^{iq\chi(\underline{\boldsymbol{x}})}\psi(\underline{\boldsymbol{x}}). \tag{2.8}$$

It can be shown, that both statements can be derived from the mathematical Noether Theorem [19]. For the gauge principle, they are used the other way round: If there is a known conservation quantity (like e.g. the electric current/charge), the Lagrangian has to fulfil at least the corresponding global symmetry (e.g. global phase invariance).

Demanding invariance under **local** phase transformations of the fields, one usually has to modify the Lagrangian and to introduce *gauge fields*, which can be identified with the interaction fields, as will be shown in the following sections. That means, demanding local phase invariance of the Lagrangian, one gets the description of the interactions "for free" (or at least relatively simple), in form of the gauge fields. Thus, the interaction of the fermion fields can be described in a very elegant way by demanding invariance under local phase transformations.

Quantum Fields

In rough analogy to non-relativistic Quantum Mechanics, fields can be quantised and be described by operators. As mentioned before, the Standard Model can be described by Quantum Field Theories. In contrast, the Lagrangians in this and the following sections are defined in terms of classical, non-quantised fields. "Real" Quantum Field Theories can be achieved via substituting the classical fields with quantum mechanical *field operators* $\hat{\psi}$ (Second Quantisation). They are defined in terms of *creation operators* and *annihilation operators*, which create or destroy the (anti-)particles corresponding to the field.

By introducing quantum fields and determining the expectation values of the temporal evolution of the quantum fields, it is possible to derive the *Feynman rules* from the Lagrangian, which describe the propagation or interaction of the fields [20]. They give the opportunity to calculate the *transition probability* $\mathcal{M}$ between an initial and a final state from the corresponding Feynman Diagrams. For experimental physics, this is an important quantity, as it gives the probability for the process to take place and thus the square of its absolute value is proportional to the cross section σ of the process ($\sigma \sim |\mathcal{M}|^2$), which is used in the experiments to determine how many events of the process should have occurred on average after a certain amount of collisions.

2.1.3. Quantum Electrodynamics (QED)

When looking at the electromagnetic interaction and using the Dirac Matrices γ^μ (with $\mu = 0, 1, 2, 3$), the Lagrangian of a free propagating particle in space-time is

$$\mathcal{L} = \overline{\psi}(\underline{\boldsymbol{x}})(i\gamma^\mu \partial_\mu - m_f)\psi(\underline{\boldsymbol{x}}), \tag{2.9}$$

and describes fermions with mass m_f. Here $\overline{\psi} = \psi^\dagger \gamma^0$ is the adjoint field of the fermion field ψ. Inserting this Lagrangian into the Euler-Lagrange equation (for $\overline{\psi}$), leads to the fermion's equation of motion, which is the well known *Dirac equation*:

$$(i\gamma^\mu \partial_\mu - m_f)\psi(\underline{\boldsymbol{x}}) = 0. \tag{2.10}$$

It was developed by Paul Dirac in 1928 and enabled him to predict the existence of "positively charged electrons", which were discovered in 1932.

$\mathcal{L}$ is obviously invariant under global phase transformations as χ is constant and does not depend on $\underline{\boldsymbol{x}}$, which ∂_μ acts on.

$$\psi(\underline{\boldsymbol{x}}) \to \psi'(\underline{\boldsymbol{x}}) = e^{iq\chi}\,\psi(\underline{\boldsymbol{x}}) \quad \Rightarrow \quad \overline{\psi}(\underline{\boldsymbol{x}}) \to \overline{\psi}'(\underline{\boldsymbol{x}}) = \overline{\psi}(\underline{\boldsymbol{x}})\,e^{-iq\chi} \tag{2.11}$$

In contrast, when applying a local phase transformation

$$\psi(\underline{\boldsymbol{x}}) \to \psi'(\underline{\boldsymbol{x}}) = e^{iq\chi(\underline{\boldsymbol{x}})}\,\psi(\underline{\boldsymbol{x}}) \tag{2.12}$$

the derivative of the field ψ becomes:

$$\partial_\mu \psi \to \partial_\mu \psi' = e^{iq\chi}\,\partial_\mu \psi + iq\,e^{iq\chi}\psi\,\partial_\mu \chi. \tag{2.13}$$

Therefore, $\mathcal{L}$ is not invariant under local gauge transformation:

$$\mathcal{L}' = \overline{\psi}(i\gamma^\mu \partial_\mu - m_f)\psi - q\,\overline{\psi}\gamma^\mu \psi\,\partial_\mu \chi = \mathcal{L} - q\,\overline{\psi}\gamma^\mu \psi\,\partial_\mu \chi \neq \mathcal{L}. \tag{2.14}$$

In order to restore the demanded invariance, the derivative has to be substituted by a *covariant derivative* of the form:

$$D_\mu = \partial_\mu - iqA_\mu(\underline{\boldsymbol{x}}). \tag{2.15}$$

Doing so, a new gauge field A_μ is implemented, which likewise has to be transformed if the field ψ is transformed. To adjust the term $-q\,\overline{\psi}\gamma^\mu \psi\,\partial_\mu \chi$ in equation (2.14) the gauge field has to transform like:

$$A_\mu(\underline{\boldsymbol{x}}) \to A'_\mu(\underline{\boldsymbol{x}}) = A_\mu(\underline{\boldsymbol{x}}) + \partial_\mu \chi(\underline{\boldsymbol{x}}). \tag{2.16}$$

This is exactly the same way that the electromagnetic vector potential A transforms in electrodynamics. Therefore the gauge field A_μ is identified with the photon field. In order to include its propagation, the kinetic term

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.17}$$

is added, which depends on the field strength tensor

$$F_{\mu\nu}(\underline{\boldsymbol{x}}) = \partial_\mu A_\nu(\underline{\boldsymbol{x}}) - \partial_\nu A_\mu(\underline{\boldsymbol{x}}) \tag{2.18}$$

and is well known from electrodynamics. Thus the final Lagrangian of QED has the form:

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(i\gamma^\mu D_\mu - m_f)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \overline{\psi}(i\gamma^\mu \partial_\mu - m_f)\psi + q\overline{\psi}\gamma^\mu A_\mu \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (2.19)$$

The first term describes the propagation of fermions, the last term describes the propagation of the gauge boson (the photon) and the central term describes the coupling between a photon and fermions with charge q. One should notice, that there is no mass term for the photon possible, which is due to the fact that a *boson mass term* of the form

$$\frac{m_\gamma^2}{2}A_\mu A^\mu \quad (2.20)$$

is not invariant under gauge transformations like shown in equation (2.16). And indeed, it has been shown experimentally, that photons are massless particles.

The QED was the first Quantum Field Theory which was developed to describe one of the fundamental interactions. It describes the experimental findings with a very high precision.

As its symmetry transformation $U = e^{iq\chi(\underline{\boldsymbol{x}})}$ is unitary ($U^\dagger U = 1$), the underlying symmetry of the QED is the $U(1)_{em}$ group, where U is a unitary 1×1 matrix. The other interactions need higher symmetry groups. These symmetry groups are the $SU(n)$[18] with unitary $n \times n$ matrices U and $n^2 - 1$ generators, i.e. gauge fields.

2.1.4. Quantum Chromodynamics (QCD)

QCD is the Quantum Field Theory of the strong interaction. Analogous to QED, it is based on symmetry under a local phase transformation in the colour space, i.e. the transformation for the colour charge. This charge is an additional quantum number of the quarks, which was established after the discovery of the Δ^{++} baryon. The Δ^{++} consists of three u-quarks and has spin $^3/_2$. As u-quarks possess spin $^1/_2$, the three u-quarks forming the Δ^{++} would occupy the same quantum state (same place, flavour, spin), which would break the Pauli Principle. Because the Δ^{++} has been discovered nevertheless, the quarks need to possess another quantum number (with at least 3 different values), which is the colour charge.

Starting from the Lagrangian of a free quark with colour a

$$\mathcal{L} = \overline{q}_a(\underline{\boldsymbol{x}})(i\gamma^\mu \partial_\mu - m_q)q_a(\underline{\boldsymbol{x}}) \qquad \text{with } a = 1, 2, 3 \quad (2.21)$$

and combining the quark's Dirac spinors of different colour to

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \quad (2.22)$$

the initial Lagrangian can be created:

$$\mathcal{L} = \overline{q}(\underline{\boldsymbol{x}})(i\gamma^\mu \partial_\mu - m_q)q(\underline{\boldsymbol{x}}). \quad (2.23)$$

[18] "S" standing for "Special" and indicating $\det(U) = 1$.

Demanding the invariance under local $SU(3)_C$ (C referring to "colour") transformation

$$q(\underline{\boldsymbol{x}}) \to q'(\underline{\boldsymbol{x}}) = e^{i\frac{g_s}{2}\theta^\alpha(\underline{\boldsymbol{x}})\lambda_\alpha}\, q(\underline{\boldsymbol{x}}) \qquad \text{with } \alpha = 1, ..., 8 \tag{2.24}$$

(where λ_α are the Gell-Mann matrices, g_s is the strong coupling constant and θ^α are the transformation angles), the covariant derivative becomes:

$$D_\mu = \partial_\mu - i\frac{g_s}{2}\lambda_\alpha A^\alpha_\mu(\underline{\boldsymbol{x}}). \tag{2.25}$$

With the structure constants $f_{\alpha\beta\gamma}$ the gluon fields A^α_μ transform like:

$$A^\alpha_\mu(\underline{\boldsymbol{x}}) \to A^\alpha_\mu(\underline{\boldsymbol{x}}) - \frac{1}{g_s}\partial_\mu\theta^\alpha(\underline{\boldsymbol{x}}) + f_{\alpha\beta\gamma}\theta^\beta(\underline{\boldsymbol{x}})A^\gamma_\mu(\underline{\boldsymbol{x}}). \tag{2.26}$$

Adding the kinetic term of the gluon field leads to the Lagrangian of the QCD:

$$\mathcal{L}_{\text{QCD}} = \sum_{a=1}^{3} \overline{q}_a(i\gamma^\mu D_\mu - m_q)q_a - \frac{1}{4}\sum_{\alpha=1}^{8} F_{\alpha,\,\mu\nu}F^{\mu\nu}_\alpha\,. \tag{2.27}$$

Like in QED, no gluon mass is allowed in the Lagrangian. The first term contains the interaction between quarks and gluons, described by

$$\frac{g_s}{2}\,\overline{q}_a\lambda_\alpha\gamma^\mu A^\alpha_\mu q_a. \tag{2.28}$$

But as the gluon field strength tensor is of the form

$$F^\alpha_{\mu\nu}(\underline{\boldsymbol{x}}) = \partial_\mu A^\alpha_\nu(\underline{\boldsymbol{x}}) - \partial_\nu A^\alpha_\mu(\underline{\boldsymbol{x}}) + g_s f_{\alpha\beta\gamma}A^\beta_\mu(\underline{\boldsymbol{x}})A^\gamma_\mu(\underline{\boldsymbol{x}}) \tag{2.29}$$

the kinetic term inserts three gluon and four gluon terms into the Lagrangian. Therefore, in contrast to the photons, the gluons are able to self-interact (which is due to the fact that they carry colour charges themselves).

With respect to the distance dependency of the strength, the strong force shows a special behaviour: While the electromagnetic and the weak interaction get weaker with higher distances, the strong interaction gets stronger. Thus, if two quarks depart far enough from each other, the gluon field between them contains enough energy to create new quarks. Therefore, the quarks *hadronise*[19] and are confined in hadrons. On the other hand, the strong interaction becomes weak for short distances (*asymptotic freedom*), i.e. nearly free quarks can be observed in high energy scattering experiments.

2.1.5. Electroweak Unification: Quantum Flavour Dynamics (QFD)

The description of the weak force is a bit more challenging than of the other forces. As mentioned before, no right-handed neutrinos seem to exist, or at least they do not interact with other particles (maybe apart from gravitational interactions). Furthermore, the weak interaction does not distinguish between the two left-handed particles of one generation.

[19] Except t-quarks, which decay almost instantaneously so that hadronisation cannot take place.

This behaviour can be described in an elegant way, using the *isospin* quantum number I (where I_3 is the component of the isospin in direction of the quantisation axis): Left-handed fermions can be grouped in *isospin doublets* (with $I = 1/2$)

$$\begin{pmatrix}\nu_e\\e^-\end{pmatrix}_L \quad \begin{pmatrix}\nu_\mu\\\mu^-\end{pmatrix}_L \quad \begin{pmatrix}\nu_\tau\\\tau^-\end{pmatrix}_L \quad \begin{pmatrix}u\\d\end{pmatrix}_L \quad \begin{pmatrix}c\\s\end{pmatrix}_L \quad \begin{pmatrix}t\\b\end{pmatrix}_L \qquad \begin{matrix}I_3 = 1/2\\I_3 = -1/2,\end{matrix} \tag{2.30}$$

while the right-handed fermions (except neutrinos) form isospin singlets (with $I = 0$):

$$e_R^- \quad \mu_R^- \quad \tau_R^- \quad u_R \quad d_R \quad c_R \quad s_R \quad t_R \quad b_R \qquad I_3 = 0. \tag{2.31}$$

The attempt to describe exclusively the weak interaction by demanding symmetry under local isospin transformation $SU(2)_L$ (i.e. transforming $I_3 = 1/2$ particles to $I_3 = -1/2$ particles) failed: Two of the three gauge bosons can indeed be combined to the observed $W^\pm$, but it is not possible to identify the third gauge boson (which does not transform the fermions) with the observed Z^0.

Therefore, the approach was made to combine the theoretical description of the weak and the electromagnetic interaction. For this purpose, the $SU(2)_L \times U(1)_Y$ group was formed by combining the (weak) $SU(2)_L$ group with the $U(1)_Y$ group, which acts on the *weak hypercharge* Y of the particles and is NOT the same as the $U(1)_{em}$ group. The weak hypercharge is calculated from the electric charge Q and I_3:

$$Y = 2\,(Q - I_3). \tag{2.32}$$

This equation results in the following hypercharges for the fermions:

	ν_L	l_L	ν_R	l_R	q_L^u	q_L^d	q_R^u	q_R^d
Q	0	-1	0	-1	$2/3$	$-1/3$	$2/3$	$-1/3$
I	$1/2$	$1/2$	0	0	$1/2$	$1/2$	0	0
I_3	$1/2$	$-1/2$	0	0	$1/2$	$-1/2$	0	0
Y	-1	-1	0	-2	$1/3$	$1/3$	$4/3$	$-2/3$

Table 2.3.: Electric charge, isospin and hypercharge of the left- and right-handed fermions. For the quarks, u and d refer to up-type and down-type quarks. Notice, that the hypercharge is equal for all left-handed leptons/quarks, as seen in experiments, and the right-handed neutrinos do not take part in the electroweak interaction, due to $I = Y = 0$.

Analogous to QED and QCD, the Lagrangian of the QFD can now be derived from an initial Lagrangian. Here one has to distinguish between the left- and right-handed fermion fields ψ_L and ψ_R, where the total fermion fields ψ can be written as the sum of the left- and right-handed fields:

$$\psi = \psi_L + \psi_R. \tag{2.33}$$

This distinction between left- and right-handed fermion fields would be possible in the QED or QCD Lagrangian, too, but would have no effect, since QED and QCD act on left- and the

right-handed particles in the same way. ψ_L and ψ_R can be derived from ψ, using the parity operators P_L and P_R:

$$\psi_L = P_L\,\psi = \frac{1-\gamma^5}{2}\psi \qquad \psi_R = P_R\,\psi = \frac{1+\gamma^5}{2}\psi \qquad \text{with } \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{2.34}$$

Due to $\gamma^5\gamma^\mu = -\gamma^\mu\gamma^5$ and $\gamma^{5\,\dagger} = \gamma^5$, fermion mass terms of the form

$$m_f\overline{\psi}\psi = m_f\big(\underbrace{\overline{\psi}_L\psi_L}_{0} + \underbrace{\overline{\psi}_R\psi_R}_{0} + \overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L\big) \tag{2.35}$$

can be formed, which are not invariant under a local $SU(2)_L \times U(1)_Y$ transformation

$$\begin{aligned} U(1)_Y: &\qquad \psi(\underline{\boldsymbol{x}}) \to e^{i\frac{g'}{2}Y\alpha(\underline{\boldsymbol{x}})}\,\psi(\underline{\boldsymbol{x}}) \\ SU(2)_L: &\qquad \psi(\underline{\boldsymbol{x}}) \to e^{ig\,I\,\sigma_\nu\beta^\nu(\underline{\boldsymbol{x}})}\,\psi(\underline{\boldsymbol{x}}) \end{aligned} \tag{2.36}$$

(with the four transformation angles α and β^ν with $\nu = 1,2,3$, the coupling constants g of $SU(2)_L$ and g' of $U(1)_Y$ and the three Pauli matrices σ_ν), as ψ_L and ψ_R transform differently. Therefore, no fermion masses are allowed in the QFD Lagrangian.

Demanding invariance under local $SU(2)_L \times U(1)_Y$ transformation, the covarinat derivative becomes

$$D_\mu = \partial_\mu - ig\,I\,\sigma_\nu W_\mu^\nu(\underline{\boldsymbol{x}}) - i\frac{g'}{2}YB_\mu(\underline{\boldsymbol{x}}) \tag{2.37}$$

and introduces four gauge fields, which transform like:

$$\begin{aligned} U(1)_Y: &\qquad B_\mu(\underline{\boldsymbol{x}}) \to B_\mu(\underline{\boldsymbol{x}}) - \frac{1}{g'}\partial_\mu\alpha(\underline{\boldsymbol{x}}) \\ SU(2)_L: &\qquad W_\mu^\nu(\underline{\boldsymbol{x}}) \to W_\mu^\nu(\underline{\boldsymbol{x}}) - \frac{1}{g}\partial_\mu\beta^\nu(\underline{\boldsymbol{x}}) + \epsilon_{\nu\delta\theta}\,\beta^\delta(\underline{\boldsymbol{x}})\,W_\mu^\theta(\underline{\boldsymbol{x}}). \end{aligned} \tag{2.38}$$

The Pauli matrices in equation (2.37) ensure that only those particles couple to the physical $W^\pm$ and Z^0 bosons, which have been observed doing so. Together with the kinetic terms of the gauge fields, the field strength tensors of the gauge fields

$$\begin{aligned} B_{\mu\nu}(\underline{\boldsymbol{x}}) &= \partial_\mu B_\nu(\underline{\boldsymbol{x}}) - \partial_\nu B_\mu(\underline{\boldsymbol{x}}) \\ W_{\mu\nu}^\theta(\underline{\boldsymbol{x}}) &= \partial_\mu W_\nu^\theta(\underline{\boldsymbol{x}}) - \partial_\nu W_\mu^\theta(\underline{\boldsymbol{x}}) + g\epsilon_{\nu\delta\theta}\,W_\mu^\delta(\underline{\boldsymbol{x}})W_\nu^\theta(\underline{\boldsymbol{x}}) \end{aligned} \tag{2.39}$$

and the two Lagrangians $\mathcal{L}_{\mathrm{SB}}$ and $\mathcal{L}_{\mathrm{YW}}$, the Lagrangian of the QFD becomes:

$$\mathcal{L}_{\mathrm{QFD}} = \sum_{\text{all fermions}} \overline{\psi}i\gamma^\mu D_\mu\psi - \frac{1}{4}\sum_{\theta=1}^{3} W_{\theta,\,\mu\nu}W_\theta^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \mathcal{L}_{SB} + \mathcal{L}_{YW}. \tag{2.40}$$

The Symmetry Breaking Lagrangian $\mathcal{L}_{\mathrm{SB}}$ and the Yukawa Lagrangian $\mathcal{L}_{\mathrm{YW}}$ provide the boson and fermion masses via the Higgs mechanism, which will be discussed in section 2.1.6. $\mathcal{L}_{\mathrm{QFD}}$ contains the terms describing the interaction between fermions and gauge bosons[20] as well as self-interaction terms of the W_μ^ν bosons, which lead to *self-interaction* of the $W^\pm$ bosons and the coupling of Z^0 and γ to $W^\pm$.

[20] Due to $\gamma^5\gamma^\mu\gamma^\nu = \gamma^\mu\gamma^\nu\gamma^5$, no coupling between left- and right-handed fields (e.g. $\overline{\psi}_L i\gamma^\mu D_\mu\psi_R$) remains, which would break the $SU(2)_L \times U(1)_Y$ symmetry.

In contrast to QED and QCD, it is not possible to directly identify the gauge bosons W_μ^ν and B_μ, which are the eigenstates of the electroweak interaction, with the observed $W^\pm$, Z^0 and γ. The latter can be obtained from the former by:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp i W_\mu^2\right) \\ Z_\mu &= \cos\theta_w W_\mu^3 - \sin\theta_w B_\mu \\ A_\mu &= \sin\theta_w W_\mu^3 + \cos\theta_w B_\mu. \end{aligned} \tag{2.41}$$

As the theory does not give a prediction for the *Weinberg angle* or *weak-mixing angle* θ_w, it has to be determined experimentally ($\sin^2\theta_w \approx 0.231$ [1]). As can be seen by combining equations (2.37) and (2.41), $W^\pm$ can only couple to left-handed fermions (as there is no W_μ^ν for $I = 0$), while Z^0 and γ interact with left- and right-handed fermions (via the B_μ).

Inserting (2.41) into the QFD Lagrangian (2.40), one indeed gets a Lagrangian describing the self-interaction of the physical gauge bosons (which will not be shown here, for simplicity purposes) and their interactions with the fermions:

$$\begin{aligned} \mathcal{L}_{\text{QFD}} = \sum_{\text{all fermions}} \overline{\psi} \Bigg(& i\gamma^\mu\partial_\mu - ig\, I\, \sqrt{2}\, \left(\sigma_+ W_\mu^- + \sigma_- W_\mu^+\right) \\ & - \left(ig\, I\, \sigma_3 \cos\theta_w - i\frac{g'}{2} Y \sin\theta_w\right) Z_\mu - \left(ig\, I\, \sigma_3 \sin\theta_w + i\frac{g'}{2} Y \cos\theta_w\right) A_\mu \Bigg) \psi \\ & - \frac{1}{4}\sum_{\theta=1}^{3} W_{\theta,\,\mu\nu} W_\theta^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \mathcal{L}_{SB} + \mathcal{L}_{YW}. \end{aligned} \tag{2.42}$$

The matrices $\sigma_\pm = {}^1\!/_2(\sigma_1 \pm i\sigma_2)$ ensure only the "right" pairs of particles to couple to $W^\pm$, e.g. $\nu_e \to W^+\, e^-$ or $e^- \to W^-\, \nu_e$, but not $\nu_e \to W^-\, e^-$ or $e^- \to W^+\, \nu_e$. The third Pauli matrix σ_3 makes the coupling of (left-handed) neutrinos with photons vanish and ensures the coupling constant of the photon-fermion-coupling to be the leptons' electric charge. Therefore, as the term for photon-fermion-coupling has to be equivalent to the corresponding QED term, the electroweak coupling constants have to fulfil the requirement:

$$g \sin\theta_w = g' \cos\theta_w = e. \tag{2.43}$$

One should note, that this equation relates the coupling constants of the $SU(2)_L$ group and the $U(1)_Y$ group to the electromagnetic coupling constant e and thus unifies both of the forces to the electroweak force.

2.1.6. Higgs Mechanism

The QFD (and thus the Standard Model, too) does only describe massless fermions and bosons, since their mass terms would break the $SU(2)_L \times U(1)_Y$ symmetry. This does obviously not correspond to the experimental observations, as the fermions (besides the neutrinos) and the gauge bosons $W^\pm$ and Z are found to be massive. The missing mass

terms in the Lagrangian can be generated by the *Higgs mechanism*[21]. It introduces a new field, the Higgs field, which forms a isospin doublet (with $I = 1/2$ and $Y = 1$)

$$\boldsymbol{\phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_1 + i\phi_2 \end{pmatrix} \tag{2.44}$$

and is assumed to interact with every particle with a coupling strength proportional to the particle's mass. A symmetric potential $V(\boldsymbol{\phi}) = -\mu^2 \boldsymbol{\phi}^\dagger \boldsymbol{\phi} + \lambda \boldsymbol{\phi}^\dagger \boldsymbol{\phi}$ is attributed to the Higgs field (shown in figure 2.4), which possesses a non-vanishing ground states and thus leads to spontaneous *symmetry breaking*. Therefore, the interaction between particles and the Higgs field generates the mass of the particles, which are massless before this interaction.

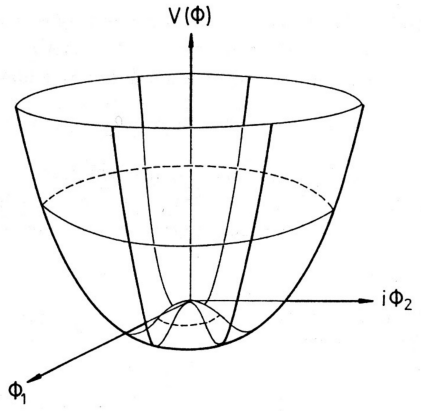


Figure 2.4.: The Higgs potential (taken from [13]).

More precisely, only the ground state of ϕ^0 is non-vanishing

$$\langle 0|\boldsymbol{\phi}|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \qquad \text{with } v = \sqrt{\frac{-\mu^2}{\lambda}}, \tag{2.45}$$

as a non-vanishing ground state of ϕ^+ would lead to a massive photon. "Choosing" one ground state breaks the symmetry and generates the masses.

The resulting Lagrangians of the Higgs field and its interactions are $\mathcal{L}_{\text{SB}}$ and $\mathcal{L}_{\text{YW}}$, which have been mentioned in chapter 2.1.5

$$\begin{aligned} \mathcal{L}_{\text{SB}} &= \mathcal{L}_{\text{Higgs}} = (D^\mu_{\text{QFD}} \boldsymbol{\phi})^\dagger (D_{\mu,QFD} \boldsymbol{\phi}) - V(\boldsymbol{\phi}) \\ \mathcal{L}_{\text{YW}} &= - \sum_{\text{all fermions}} \tilde{g}_f \left(\overline{\psi}_R (\boldsymbol{\phi}^\dagger \psi_L) + (\overline{\psi}_L \boldsymbol{\phi}) \psi_R \right) \end{aligned} \tag{2.46}$$

and for the masses one finds (with the Yukawa coupling constant $\tilde{g}_f$ for the interaction between Higgs field and fermion fields):

$$m_W = m_Z \cos\theta_w = \frac{gv}{2} \qquad\qquad m_f = \frac{\tilde{g}_f v}{\sqrt{2}}, \tag{2.47}$$

As hint for the realisation of the Higgs mechanism in nature, the relation $m_W = m_Z \cos\theta_w$ has been tested and proved experimentally.

[21] Named after one of its developers, the British physicist Peter Higgs.

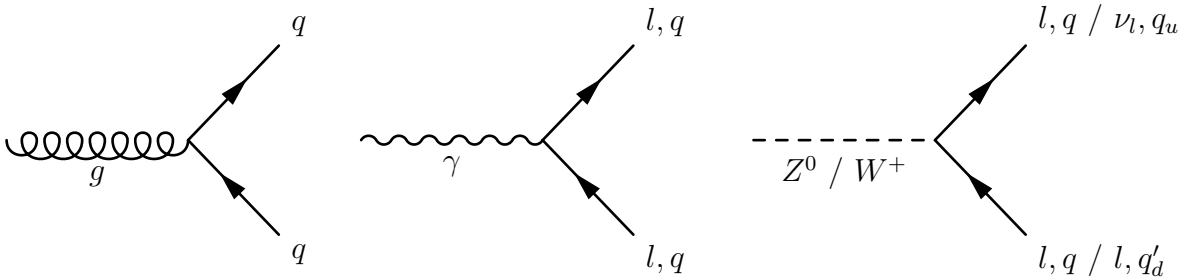
2.1.7. Unsolved Problems

The Standard Model is able to describe a wide range of particle physics with high precision and has been tested in many experiments. Nevertheless, it is not able to describe every experimental observation and there are still several open questions, for example:

- The Higgs mechanism is obligatory for the consistency of the Standard Model. Anyway, no experimental evidence of the Higgs boson has been found until now. Thus, the search for the Higgs boson is one of the main tasks of the LHC experiment.

- So far, there has been no success in developing a quantum mechanical description of gravity. The Standard Model is therefore not able to describe the gravitational force.

- As matter can be created from energy only together with antimatter, e.g. in processes like those shown below, there must have been the same amount of matter and antimatter shortly after the *big bang*. Though, a mechanism which enforced this is not known. Therefore it is one of the unsolved questions of particle physics, why we are living in a "matter"-world today, where no antimatter is left.



- The non-vanishing neutrino masses, which indications have been found for, cannot be described by the Standard Model, as they would lead to a connection of left- and right-handed states in the QFD. In this case, also right-handed neutrinos would have to be observed.

- The theoretical basis of the Standard Model are three different gauge groups with different coupling constants. With the QFD, it was already possible to unify two of them. The next step would be the unification of all three forces to one basic force in a Grand Unified Theory. However, this is not possible within the Standard Model, as not all the three coupling constants unify at one common energy scale. A possible solution can be provided by supersymmetric theories (cf. chapter 7.3).

3. Experimental Setup

MUSiC[22] is an analysis tool for collision data taken with the CMS detector. The CMS detector is part of the LHC experiment at CERN, which is the world's largest laboratory for particle physics and resides near Geneva in Switzerland.

3.1. Collider Physics

Particle colliders are built to investigate the interactions of particles. The aim of doing so is to perform **precision measurements of known physics** and to **search for new physics**. For both purposes, high *centre-of-mass energies*

$$\sqrt{s} = E_{\text{cm}} = 2E_{\text{beam}} \tag{3.1}$$

of the colliding beams are needed. There are two basic reason for the necessity of high $\sqrt{s}$: More energy is needed to create heavier particles, according to Einstein's formula

$$E = mc^2 \tag{3.2}$$

and high momenta p of the colliding particles result in small *de Broglie wavelengths*

$$\lambda = \frac{2\pi}{p} \tag{3.3}$$

and thus smaller structures can be displayed.

Basically, there are four options for designing a particle collider:

- Using a **linear** or a **circular collider**:
  In a linear collider, the particles are accelerated on a linear path, before they collide. Thus, every particle passes every accelerator device only once, which leads to a large material requirement. Furthermore, it is not possible to use particles, which did not collide, in a second collision. In a circular collider, the particles pass every part of the collider several times and therefore much less material (e.g. accelerator devices) is needed than with a linear collider. Additionally, after one collision, the remaining particles of the beams can be used for further collisions in a circular collider. On the

[22] Model Unspecific Search in CMS, a special algorithm developed at the RWTH Aachen, to realise model unspecific search for new physics for the data taken with CMS.

other hand, the reachable energy of circular colliders is limited by the energy losses due to *synchrotron radiation*

$$\Delta E_{\text{sync}} \sim \frac{E^4}{m^4 R}, \tag{3.4}$$

which depends on the energy E and the mass m of the accelerated particles, and the radius R of the collider, and the strength of the magnets guiding the particles.

- Using **leptons** or **hadrons**:
  As leptons are point-like particles without (known) substructure, they are much more adequate for performing precision measurements than hadrons. This is because their centre-of-mass energy in a collision equals the centre-of-mass energy of the beams $\sqrt{s}$ and thus is well known, whereas the interacting partons[23] carry an unknown fraction of the hadrons energy. Additionally, when colliding leptons, only the products of the interaction are detected, while using hadrons leads to many remains of the initial hadrons contaminating the interaction events. The reason for building hadron colliders is the higher reachable energy due to their heavily reduced synchrotron radiation, as a proton possesses a ≈ 2000 times larger mass than an electron (cf. equation 3.4). Accelerating and colliding heavier leptons (i.e. muons with $m_\mu \sim m_e$) thus would be the ideal enhancement in collider physics (more information e.g. on [21]). Unfortunately, this is technologically not yet possible, since muons have a too short live time ($\tau_\mu \approx 2.2 \cdot 10^{-6}\,\text{s}$ [1]).

Besides $\sqrt{s}$, the most important parameter of particle colliders is the instantaneous *luminosity* $\mathscr{L}$, as it is proportional to the event rate $\dot{N}_p$ of a certain process p with the *cross section* σ_p ($[\sigma] = 1\,\text{pb} = 10^{-36}\,\text{cm}^2$):

$$\dot{N}_p = \sigma_p \cdot \mathscr{L}. \tag{3.5}$$

In principle, the luminosity characterises the collision intensity of the two colliding beams. Larger luminosities result in a higher possibility of a hard interaction, i.e. a *pp*-collision during which new particles are produced. Assuming two beams colliding with a frequency f, each divided into N_B bunches with a Gaussian bunch profile (widths σ_x and σ_y perpendicular to the beam direction), of which each is filled with N particles, the instantaneous luminosity per interaction point is given by[24]:

$$\mathscr{L} = \frac{f \cdot N_B \cdot N^2}{4\pi\sigma_x\sigma_y}, \tag{3.6}$$

As the data are taken during a certain period of time, equation 3.5 has to be integrated. The total number of events seen by the experiment thus depends on the integrated luminosity $\mathscr{L}_{\text{int}}$ ($[\mathscr{L}_{\text{int}}] = 1\,\text{pb}^{-1} = 10^{36}\,\text{cm}^{-2}$):

$$N_p = \sigma_p \mathscr{L}_{\text{int}}. \tag{3.7}$$

[23] valence quarks (2 × up + 1 × down for a proton), sea quarks (quark/anti-quark pairs from temporal gluon splitting), gluons

[24] At LHC a geometric luminosity reduction factor F has to be multiplied to equation 3.6, due to the crossing angle at the interaction points [22].

3.2. The Large Hadron Collider (LHC)

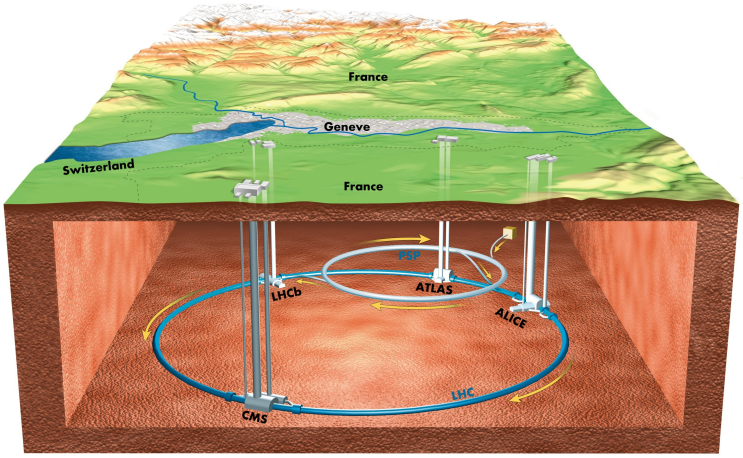


Figure 3.1.: Location of the LHC tunnel and the four main experiments (taken from [23] and modified).

The LHC [22] is a proton-proton collider[25,26] with a design centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$, built in the tunnel of the former e^+ e^--collider LEP[27]. As shown in figure 3.1, it is located under the border area of France and Switzerland. The high design luminosity of $\mathscr{L} = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ is the reason for using only protons (in contrast to $p\bar{p}$-colliders like e.g. Tevatron): It is much easier to provide enough protons than the same large number of anti-protons. In table 3.1, the design beam parameters of LHC are summarised.

Design beam parameter	Value
Energy of particles	7000 GeV
Luminosity $\mathscr{L}$	$10^{34}\,\text{cm}^{-2}\text{s}^{-1}$
Number N_B of bunches	2808
Mean beam size σ_x & σ_y at CMS interaction point	16.7 µm
Bunch spacing	25 ns
Number N of protons per bunch	$1.15 \cdot 10^{11}$
Cycling frequency	11245 Hz
Geometric luminosity reduction factor F	0.836
Beam lifetime	$\sim$ 15 h

Table 3.1.: Design beam parameters of the LHC particle collider [22].

LHC has a circumference of 26659 m and is located between 45 m and 170 m below the surface. Four main experiments are located at interaction points along the LHC tunnel:

[25] It is possible to use proton-proton collisions, because the proton energy is dominantly carried by gluons and sea quarks at the LHC energy scale. Otherwise, processes like e.g. the direct Z^0 production would be strongly suppressed.
[26] The plan is, to also accelerate lead ions at LHC, in order to perform heavy ion experiments at ALICE.
[27] **L**arge **E**lectron **P**ositron collider

- **A** **L**arge **I**on **C**ollider **E**xperiment (ALICE) - a detector specialised on measuring the fragments of heavy ion collisions.

- **A** **T**oroidal **L**HC **A**pparatu**S** (ATLAS) - one of the two general purpose detectors searching for physics beyond the Standard Model.

- **C**ompact **M**uon **S**olenoid (CMS) - the second general purpose detector.

- **L**arge **H**adron **C**ollider **b**eauty (LHCb) - a detector focusing on b-physics, the only detector with forward instead of symmetric geometry.

1232 *dipole magnets* with a length of 14.3 m and a peak dipole field of about 8.33 T are distributed along the tunnel to bend the proton beams on their way through the beam pipes. To be able to create such strong magnetic field, the dipole magnets need to consist of special superconducting materials and be cooled down to 1.9 K. Since the energy loss due to synchrotron radiation is negligible, the strength of the dipole magnets is the main limiting parameter to the beam energy.

As the two proton beams run in contrary directions, the dipole magnets need to have contrary polarities for each beam, in order to bend them on the same track. Therefore, two separated beam pipes are needed, only close to the interaction points, the beams may share one pipe. Because of the limited space available for the beam pipes and in order to reduce the efforts needed for cooling the magnets, a complex design has been chosen (cf. figure 3.2). It combines the two separated beam pipes with their own coils by a common cold mass and cryostat. The price of this solution is a magnetic coupling of both beam pipes, what complicates maintenance and reduces flexibility.

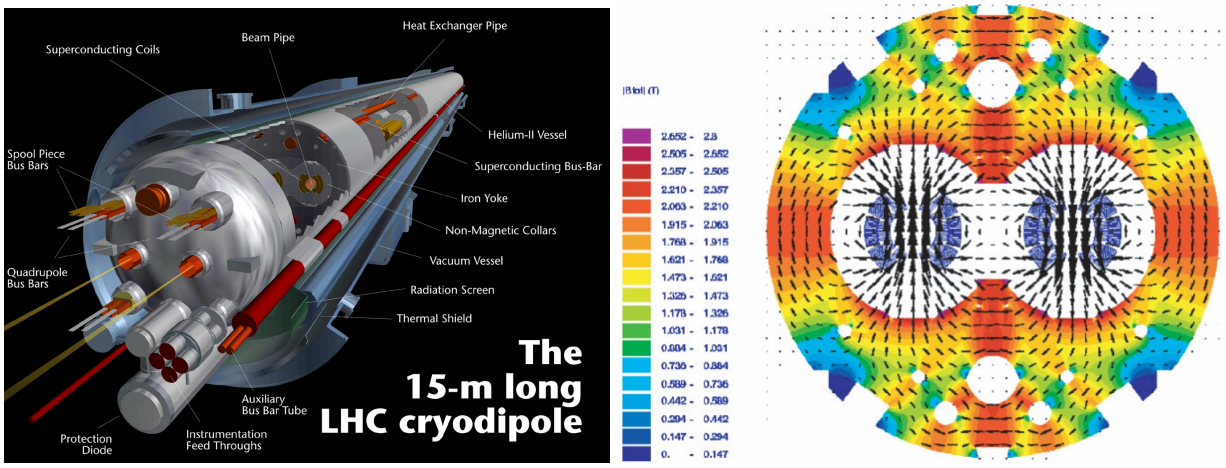


Figure 3.2.: Layout [24] (left) of and magnetic flux [25] (right) within the LHC cryodipoles.

The protons for the collisions are created by ionisation of hydrogen. Before entering the LHC, they have to pass through an *injector chain* of several acceleration cycles (Linac 2, Proton Synchrotron Booster (PSB), Proton Synchrotron (PS), Super Proton Synchrotron (SPS)), as the technology of LHC is not capable for handling too low energetic protons. When leaving

the injector chain, the protons have been grouped into bunches and accelerated to 450 GeV per beam. The injector chain with its accelerators is illustrated in figure 3.3.

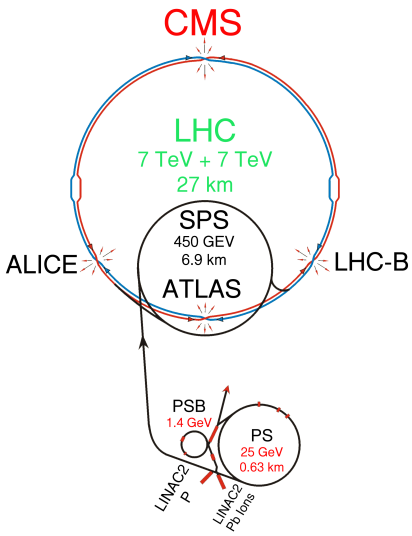


Figure 3.3.: Injector chain of the LHC [26].

3.3. The Compact Muon Solenoid (CMS)

The Compact Muon Solenoid [27] is a multi-purpose detector for precise measurement of the products from *pp*-collisions at LHC. An overview of its layout can be found in figure 3.4. Its parts were constructed on the surface and then assembled in a cavern about 100 m underground, near the French village Cessy. It consists of five wheel-like structures, which together form the barrel region, and two endcaps. All together, the detector has a length of 21 m, a height of 15 m, a width of 15 m and weighs 14000 t.

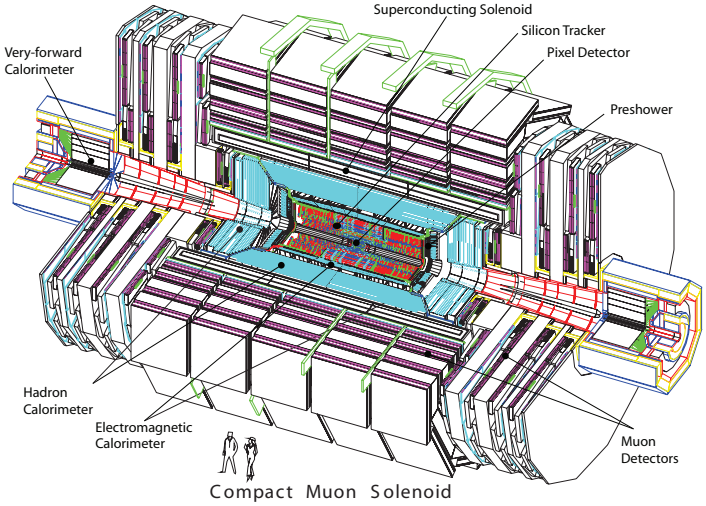


Figure 3.4.: Sketch of the complete CMS detector with its major components [28].

As the aim of CMS is to find new physics events hiding in collision data dominated by Standard Model events, especially good particle identification, energy resolution and momentum resolution are needed. Therefore, the detector consists of five main components (as shown in figure 3.5), namely (from the innermost to the outermost):

- The **tracking system**, allowing the reconstruction of the charged particles' trajectories.
- The **electromagnetic calorimeter**, which is used to measure the energy of electrons and photons.
- The **hadronic calorimeter**, measuring the energy of hadrons/jets.
- The **superconducting solenoid**, providing a strong magnetic field needed for determining the particles' momenta from the bending of their trajectories.
- The **muon system**, necessary for identifying muons and reconstructing their trajectories.

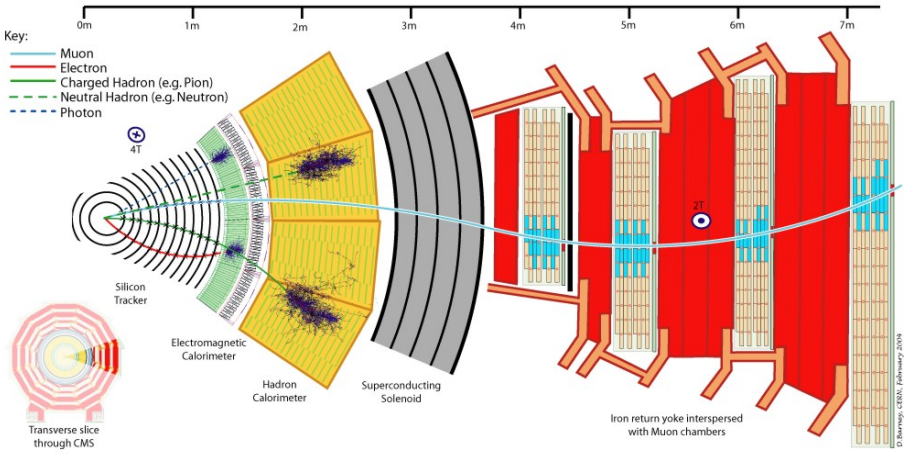


Figure 3.5.: Detector response to different types of particles [29]: Charged particles create hits (and thus bend tracks) in the tracker and the muon system (only muons should reach there), electrons and photons deploy their energy in the ECAL, hadrons in the HCAL.

Additionally, the detector almost entirely incloses the interaction point, which is a requirement for a good resolution of missing transverse momentum and missing transverse energy, which could indicate new physics.

3.3.1. CMS Coordinate System

In order to be able to uniquely describe reconstructed particles within the detector, a fixed, detector related coordinate system is used. The nominal interaction point is defined as the origin, the x-axis points to the centre of the LHC accelerator ring, the z-axis lies in beam

direction and the y-axis stands vertically to the surface. Due to symmetry, the particles' directions of flight are often described in spherical coordinates. ϕ is the polar angle, lying in the x-y-plane, measured from the x-axis in mathematical positive direction (i.e. the y-axis is at $\phi = 90\,°$). The azimuthal angle θ is measured from the z-axis in direction of the x-y-plane. It can be translated into the *pseudo-rapidity* η by

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_{||}}{p - p_{||}} \right) = - \ln \left(\tan \left(\frac{\theta}{2} \right) \right), \tag{3.8}$$

which is favoured due to the invariance of η-differences under Lorentz boost and its close relation to the (invariant) *rapidity* y in special relativity. Using these parameters, the distance between two particles can be defined as

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2}. \tag{3.9}$$

According to the Lorentz force (cf. equation 3.15, chapter 3.3.5), a magnetic field leads to magnetic deflection of charged particles moving through the magnetic field. This can be used to determine the particle's momentum (cf. chapters 3.3.2 and 3.3.6). As only the part of the momentum, which is transverse to the magnetic field, leads to magnetic deflection, only the *transverse momentum* p_T can be measured. In CMS, the magnetic field is in z-direction (cf. chapter 3.3.5), the transverse momentum is defined as

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2} = p \sin\theta. \tag{3.10}$$

Often, also the *transverse energy* E_T or *transverse missing energy* $\not{E}_\mathrm{T}$ are used. They are defined in the same way as the transverse momentum.

3.3.2. Tracking System

A crucial point of the reconstruction of an event is a reliable track measurement with a high spatial resolution. It is necessary for a good determination of the particles' momenta, the correct mapping of tracks and calorimeter entries as well as the reconstruction of vertices. Thus, without a good tracking system, b-tagging is impossible, since the highly efficient methods rely on the determination of secondary vertices from the decay of b-hadrons (cf. chapter 6.3).

In any case, the tracker [30] has to be the innermost detector layer, because particles are deflected and lose energy when travelling through matter. Additionally, to gain the best resolution possible and not to waste space within the detector, the tracker should be located as near to the interaction point as possible.

As the particles' interaction rate increases with the density of the material traversed, previous detectors often used **gas** filled drift chambers as or as part of the tracker system. Due to the long drift times of the ions, such techniques are relatively slow. Instead, CMS (with the expected time of 25 ns between two bunch crossings) relies on faster **silicon** based detectors.

They exploit the effect, that movable electrons and holes[28] are created in silicon by ionisation, if a charged particle traverses it. Applying high voltage, these movable charge carriers can be separated and measured as a current, which is illustrated in figure 3.6.

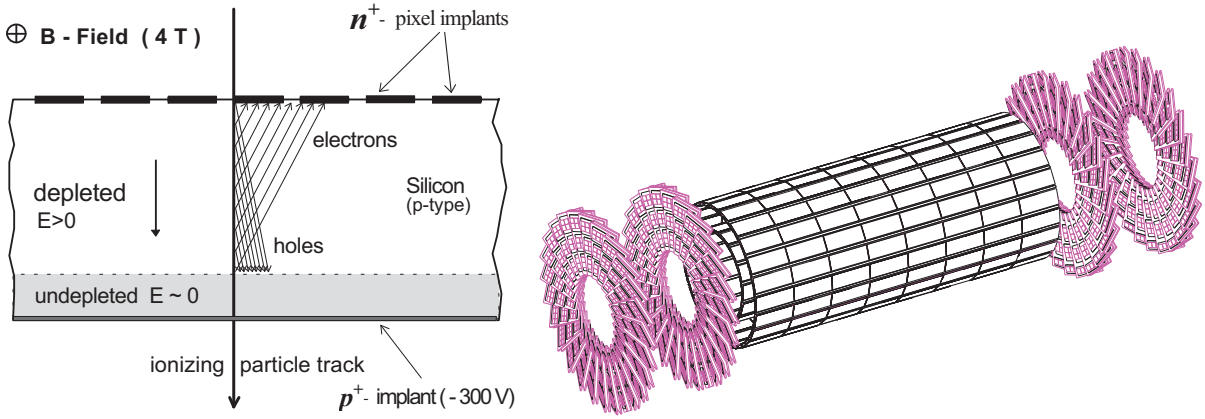


Figure 3.6.: Functionality of silicon based particle detection [30] (left) and layout of the pixel detector of the CMS tracking system [30] (right).

The CMS tracker consist of 1440 *pixel detector* modules and 15148 *strip detector* modules. A complete schematic illustration of the tracker and its coverage can be found in figure 3.7. They form the three layers of the tracker system, which covers a range of $|\eta| < 2.5$:

- **Pixel detector:**
  It forms the first layer of the tracker, located closest to the interaction point. It is composed of three cylindrical barrel layers and four disk-like endcaps. Its layout is shown in figure 3.6. The 66 million pixels possess an active area of about 1 m^2. Each pixel has a size of $100 \times 150\,\mu m^2$, allowing a good spatial resolution: Exploiting the effect of "charge sharing"[29] between the pixels, a single-point resolution in the order of 10 µm in x-y-direction and 20 µm in z-direction can be achieved [30].

  The major advantage of pixel detectors is the ability of determining all three coordinates of the particle transition simultaneously. Unfortunately, this has to be paid with an enormous number of readout channels needed and high construction costs. Therefore, the remaining parts of the tracker are composed of silicon micro-strip devices.

- **Inner silicon strip tracker** (Tracker Inner Barrel (TIB) & Tracker Inner Disks (TID)):
  It is composed of four barrel layers and three disks on each end, all consisting of silicon strip detectors. Thus, the number of needed readout channels as well as the production costs are reduced. Sequent layers have been mounted with a known angle between their strip direction, which allows the measurement of all three coordinates of the particle transition when combining the layers. A single hit resolution of 23 - 35 µm and a two-track separation of $< 200\,\mu m$ is achieved [30].

[28] Missing electrons, which can be treated as carriers of positive charge.
[29] The created charge carriers from one particle transition are distributed over a range of several pixels (cf. figure 3.6).

- **Outer silicon strip tracker** (Tracker Outer Barrel (TOB) & Tracker End Caps (TEC)):
  It has the same functionality as the inner silicon strip tracker. The TOB consists of six detection layers with higher distances between the strips, reducing the number of readout channels. Caused by the lower particle flow, a single hit resolution of 35 - 53 µm is reached anyway. The TEC consist of nine endcaps on each end of the TOB.

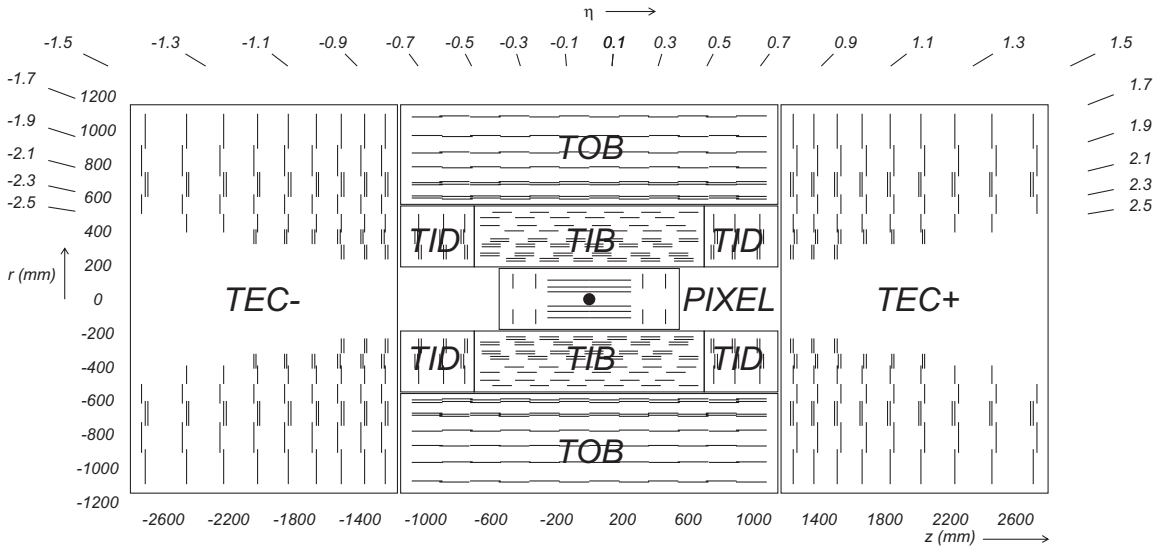


Figure 3.7.: Schematic illustration of the tracker system and its coverage [28].

Combining the 66 million pixels and 9.3 million strips of the whole tracker system leads to an expected transverse momentum resolution of

$$\left(\frac{\sigma(p_\mathrm{T})}{p_\mathrm{T}}\right)^2 = (a \cdot p_\mathrm{T}/\mathrm{GeV})^2 + (0.5\,\%)^2. \tag{3.11}$$

The linear term can be explained with the decrease of the bending power for higher p_T and the constant term results from multiple scattering. The value of a depends on η and varies between 15 % and 60 %.

3.3.3. Electromagnetic Calorimeter (ECAL)

The function of the electromagnetic calorimeter [31], which is sketched in figure 3.8, is to determine the energy of electrons and (high energy) photons. One way to do this is making them interact electromagnetically with a *scintillator* material until they are absorbed[30]. Then the particle's energy can be estimated by counting the number of optical photons created in the scintillating material, which is proportional to the energy deposit of the traversing particle. In order to achieve high efficiencies at low material thickness, the detector material needs to possess a high atomic number (resulting in a high number of electrons for electromagnetic interactions) and has to be transparent to its own scintillation light.

[30] In fact, also muons and hadrons will deploy some energy within the ECAL, but will not be absorbed.

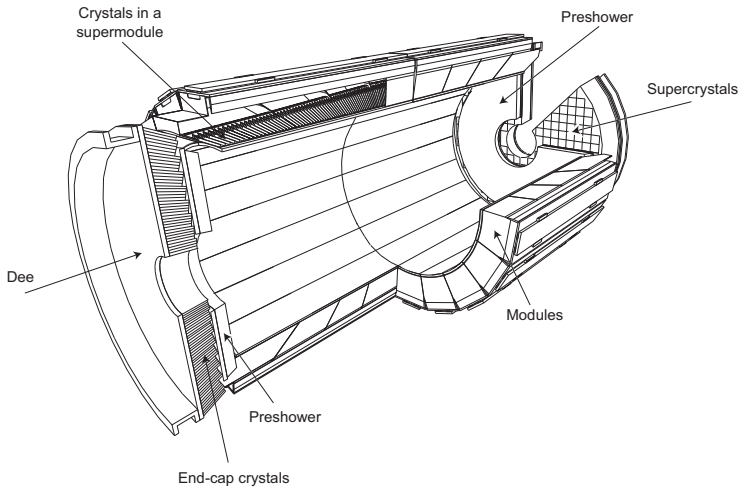


Figure 3.8.: Main components of the electromagnetic calorimeter [28].

For the CMS ECAL, lead tungstate ($PbWO_4$) has been chosen, which is a comparably fast inorganic scintillator (80 % of the scintillation light is emitted within 25 ns, allowing distinction between bunch crossings) with a high density and a good energy resolution. It is transparent to visible light and has a small *Molière radius*[31] ($R_M = 22\,\text{mm}$) and a short *radiation length*[32] ($X_0 = 8.9\,\text{mm}$), allowing a fine granularity and a compact ECAL, as the crystals have a length of $\approx 25 \cdot X_0$ and a width of $\approx 1 \cdot R_M$ [31].

The CMS ECAL consists of three components, which altogether cover a range of $|\eta| < 3$ (as shown in figure 3.9):

- **ECAL Barrel region** (EB):
  61200 $PbWO_4$ mono-crystals form the cylindrical barrel structure and cover a range up to $|\eta| < 1.479$.

- **ECAL Endcaps** (EE):
  The endcaps complete the ECAL at the margins of the barrel region. They consist of 3662 crystals each and cover the range $1.479 < |\eta| < 3$.

- **ECAL Preshower detector** (ES):
  This is an additional detector component, laying in front of the endcaps and covering $1.653 < |\eta| < 2.6$. It possesses a high granularity, achieving a two-dimensional resolution of 1.9 mm [31], and thus allows the discrimination of π^0 mesons decaying into two photons faking a high-energetic photon.

Being exposed to high radiation rates, the ECAL crystals will sustain radiation damages. As these damages but also recovery through thermal effects decreases with lower temperatures, a compromise had to be found. At 18 °C, which is the chosen operational temperature of

[31] Radius of a cylindrical volume around the original particle trajectory, containing $\approx 95\,\%$ of the total electromagnetic shower energy.

[32] Distance high-energetic electrons (photons) can travel through the material, before their energy (intensity) has decreased to $1/e$ ($(1/e)^{7/9}$) of its original value. An electromagnetic shower is about $20 \cdot X_0$ long.

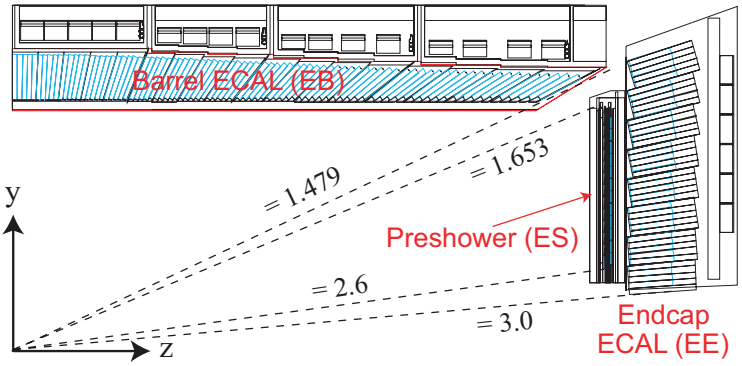


Figure 3.9.: Position and coverage of the ECAL components [32].

the ECAL, there is a equilibrium between damage and recovery.

The scintillators are read out via *photodetectors*, i.e. the scintillating light of the crystals is transformed into an electric signal. The photodetectors have to be radiation tolerant and operational within the high magnetic field of the solenoid. In the barrel region, Avalanche Photo Diodes [31] are used, which have a high quantum efficiency of 75 % for light at a wavelength of 430 nm and are insensitive to magnetic fields. In the endcaps the radiation rate is higher, therefore Vacuum Photo Triodes [31] have been chosen. The low quantum efficiency of 22 % is compensated by the larger surface coverage of the phototriodes.

Like every measurement, the energy determination is subject to limited precision of the measurement. The sources of uncertainties are calibration errors, intrinsic electronic noise, noise from *pile-up* events[33] and the probabilistic nature of photon absorption. During calibration measurements before final assembly, the expected resolution of the ECAL was determined to [28]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{12\,\%}{E/\mathrm{GeV}}\right)^2 + \left(\frac{2.8\,\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + (0.3\,\%)^2\,. \tag{3.12}$$

3.3.4. Hadron Calorimeter (HCAL)

Hadrons dominantly lose their energy via strong interaction and their *interaction length* λ_I[34] is much higher than the radiation length X_0 of electrons and photons. Therefore, high energetic hadrons pass the ECAL without losing much energy and a second calorimeter has to be installed right behind the ECAL. Since this hadronic calorimeter [33] has to provide enough nuclei for the hadrons to create hadronic showers, a very dense material has to be chosen. As the HCAL has to be limited to a reasonable extend, a material with a shorter λ_I than lead tungstate needs to be used. Therefore, the CMS HCAL is designed as a *sampling calorimeter*, alternating thick layers of additional "dead" absorber material (forcing

[33] events with more than one hard pp-interaction

[34] Distance in which all but a fraction 1/e of the incoming hadronic particles interacted. A hadronic shower is about $9 \cdot \lambda_\mathrm{I}$ long.

the showers) and thin layers of active material (detecting hadronic showers via scintillation light caused by charged particles). The HCAL consists of the following parts:

- **HCAL Barrel region** (HB):
  It is located between the EB and the solenoid and covers a range of $|\eta| < 1.3$. The HB contains brass absorber plates alternating with layers of plastic scintillator tiles, which have wavelength shifting fibres embedded for the signal readout. As this setup results in a material thickness of $10.6 \cdot \lambda_\mathrm{I}$ at $|\eta| = 1.3$ and only $5.82 \cdot \lambda_\mathrm{I}$ at $|\eta| = 0$, the additional HCAL Outer region is necessary. The segments of the HB have a tower-like readout, i.e. all the light collected by the scintillator tiles of one segment is directed to one Hybrid Photo Diode via fibres. Thus a two-dimensional resolution of $\Delta\phi = 5\,^\circ$ and $\Delta\eta = 0.087$ is obtained.

- **HCAL Outer region** (HO):
  In order to guarantee the containment of (lately developing and high energetic) hadronic showers within the calorimeter of the CMS barrel, an additional layer, the HO, is needed. It is also located in the barrel region, but outside the solenoid, and covers $|\eta| < 1.3$ as well. It consists of five wheels, placed in front of the iron return yoke. As the central region of the HB has the lowest material thickness with regard to the trajectory of the hadrons, for the central HO wheel two layers of scintillator tiles surround an absorber (iron) are used. The other four wheels are made of scintillator only, using the solenoid coil as absorber. The segmentation and readout of the HO reflects the HB's tower structure, in order to form combined HCAL towers. Considering all contributions from ECAL, HCAL, the solenoid, support structure and the first layer of the iron return yoke, a minimum material thickness of $11.8 \cdot \lambda_\mathrm{I}$ is achieved.

- **HCAL Endcaps** (HE):
  They cover a range of $1.3 < |\eta| < 3.0$ and basically work the same way as the HB. As in the HB, scintillator tiles are being read out collectively as HCAL towers. Their granularity decreases from $\Delta\phi = 5\,^\circ$ and $\Delta\eta = 0.087$ for $|\eta| < 1.6$ to $\Delta\phi = 10\,^\circ$ and $\Delta\eta = 0.17$ for $|\eta| > 1.6$. Together with the ECAL, the total material thickness is about $10 \cdot \lambda_\mathrm{I}$.

- **HCAL Forward region** (HF):
  It covers a range of $2.9 < |\eta| < 5.2$, which is not covered by any other detector part. To handle the very high particle fluxes[35] in the forward region of the detector, the design of the HF has to be quite different from the rest of the HCAL: Radiation tolerant quartz fibres are embedded in a $\approx 10 \cdot \lambda_\mathrm{I}$ long steel absorber. The charged shower particles generate Cherenkov light[36] within the fibres, which are bundled into towers of about $\Delta\phi = 10\,^\circ$ and $\Delta\eta = 0.175$.

Together, the components of the HCAL cover a range of $|\eta| < 5.2$, which is illustrated in figure 3.10, and only a small range of $< 0.7\,^\circ$ around the beam direction remains uncovered. This almost total enclosure of the interaction point is very important for the CMS detector's

[35] Per collision of two protons, $\approx 100\,\mathrm{GeV}$ are deployed in the large $|\eta| < 3$ detector region, in contrast to $\approx 760\,\mathrm{GeV}$ for $|\eta| > 3$ [33].

[36] When travelling through a material with a effective velocity of light lower than the particle's speed, a charged particle emits Cherenkov light.

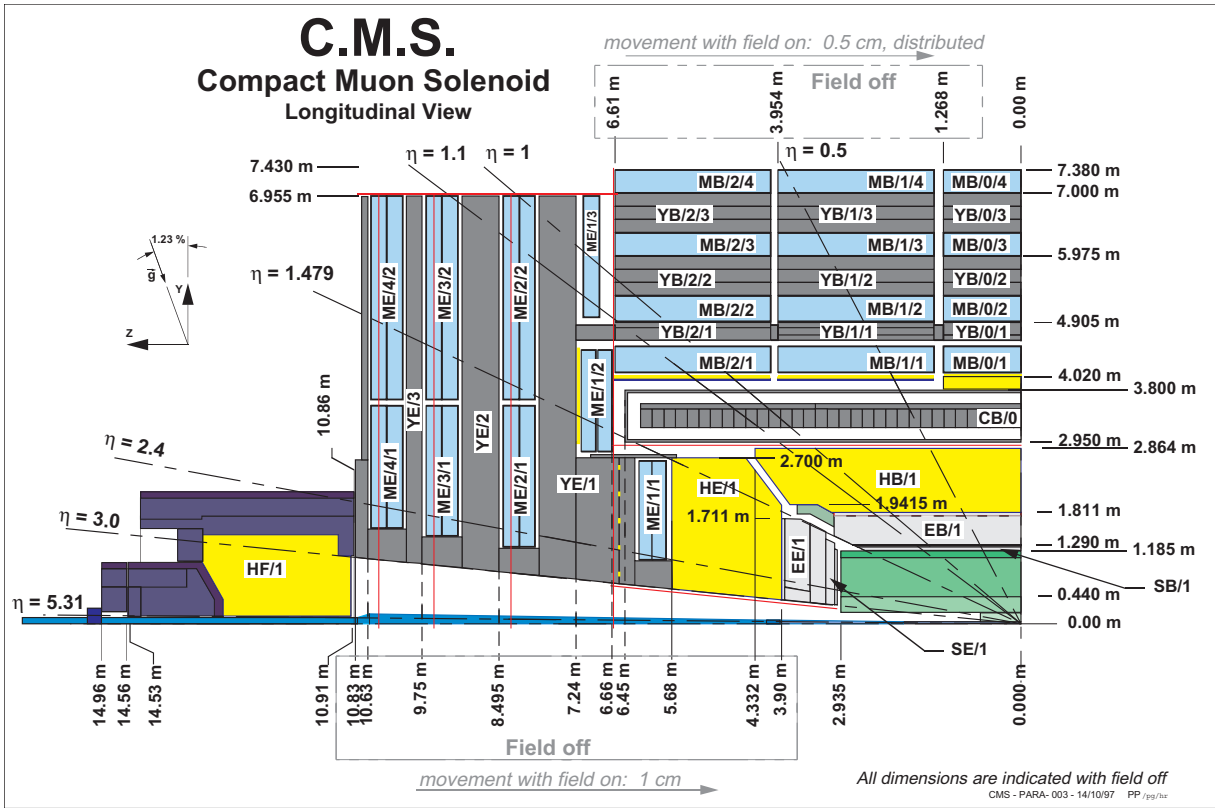


Figure 3.10.: Location and coverage of the HCAL and the muon system [34].

ability of determining possible missing transverse energy (cf. chapter 4.2.5), caused by neutrinos or new physics. The expected energy resolution of the full HCAL (including the HO) is [33]

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{120\,\%}{\sqrt{E/\mathrm{GeV}}}\right)^2 + (5\,\%)^2\,. \tag{3.13}$$

Combining the energy measurement of ECAL and HCAL, the CMS calorimeter is expected to achieve a resolution of

$$\left(\frac{\sigma_E}{E}\right)^2 \approx \left(\frac{100\,\%}{E/\mathrm{GeV}}\right)^2 + (4.5\,\%)^2 \tag{3.14}$$

for energies of $30\,\mathrm{GeV} < E < 1\,\mathrm{TeV}$.

3.3.5. Superconducting Solenoid

The solenoid [35] is the "heart" of the CMS detector. It is needed to provide the possibility of determination of the charged particles' momenta in the tracker and the muon system. According to the *Lorentz force*

$$\boldsymbol{F}_\mathrm{L} = q(\boldsymbol{v} \times \boldsymbol{B}), \tag{3.15}$$

a magnetic field leads to magnetic deflection of a particle with charge q, moving with velocity $\boldsymbol{v}$ through the magnetic field $\boldsymbol{B}$. In order to achieve a good p_T resolution even in very

high p_T ranges, the magnet is designed as a superconducting solenoid, possessing the ability of creating a magnetic field of 4 T[37] within its cylindrical coil (shown in figure 3.11). The coil is located between the Hadron Barrel calorimeter and the Hadron Outer calorimeter, which has the advantage that particles do not have to pass the solenoid's material before reaching the main parts of calorimeters. It has a diameter of 6 m and a length of 12.5 m and is cooled down to 4.6 K via liquid helium to operate in the superconducting state. When operating at full current, the solenoid stores 2.6 GJ of energy within its magnetic field.

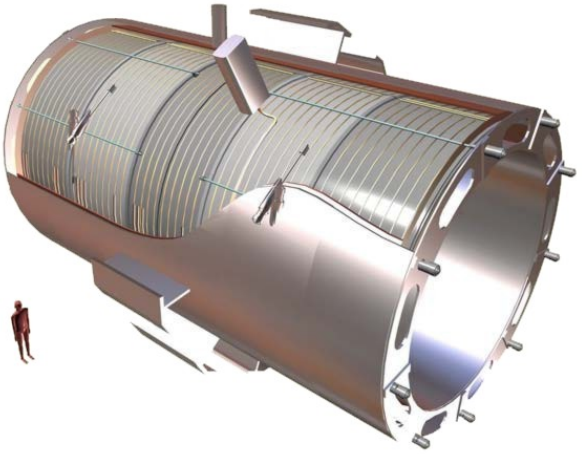


Figure 3.11.: Illustration of the CMS superconducting solenoid [28].

To return the magnetic flux, a 12000 t iron yoke in form of the five wheels and 2 endcaps is needed. As the iron yoke leads to a reverse magnetic field of about half the strength in the coil's outer region [36], the muon system is embedded in the iron yoke, in order to use the magnetic field for p_T determination.

3.3.6. Muon System

The tasks of the muon system [34] are identification of muons and determination of the p_T of muons. The muon identification is based on the fact, that every charged particle reaching the muon system **should** be a muon, as the muon system is located outside the HCAL. In reality also other charged particles may reach the muon system, e.g. due to "punch through"[38]. For the second task, the muon system has to be able to reconstruct the muons' trajectories like the tracker. Since large areas have to be covered, a silicon based setup (like for the tracking system) would be far too expensive. Therefore, gas detectors have been chosen. Additionally, the muon system has to contribute its part to the trigger system of CMS (cf. chapter 3.3.7). As gas detectors possess a too long response time, they are not suited for this purpose. Consequently, another component has been added: the Resistive Plate Chambers, which are described below.

[37] The CMS collaboration decided to run the detector with 3.8 T for the first years of operation.

[38] A "punch through" is a hadronic shower, which is not stopped in the HCAL and thus reaches the moun system.

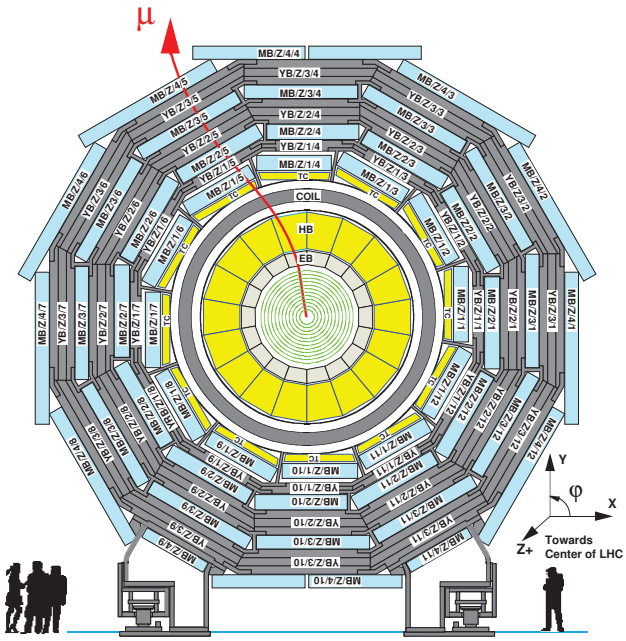


Figure 3.12.: Cross section of the barrel region of CMS [28], showing the positions of the Drift Tube chambers (labelled with "MB").

The muon system is located directly behind the calorimeter and is embedded into the iron return yoke. In total, it covers a range of $|\eta| < 2.4$ and is composed of three parts:

- **Drift Tube chambers** (DT):
  In the barrel region ($|\eta| < 1.2$) DTs with a spatial resolution of $\approx 100\,\mu\text{m}$ are used. They are mounted to the iron yoke wheels and form four cylindrical layers. This is illustrated in figure 3.12. Each DT consists of three Superlayers, except of the outermost layer, where only two Superlayers are used. One Superlayer is composed of four layers of drift tubes (cf. figure 3.13), which are shifted by half of a tube width each layer, in order to reduce the non-sensitive volume. To achieve a three-dimensional resolution, the second Superlayer of each DT is rotated by 90 ° with respect to the other Superlayers.

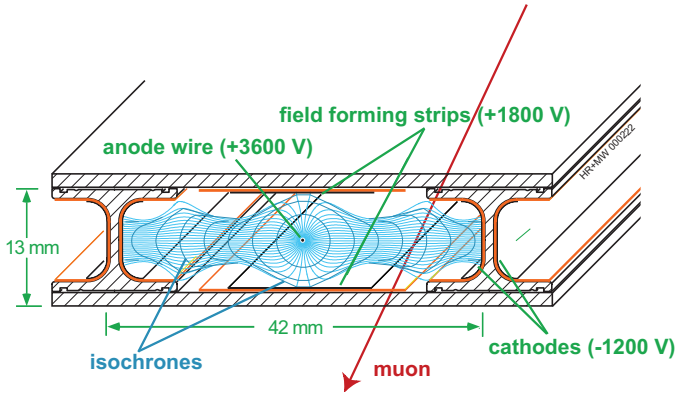


Figure 3.13.: Cross section of a CMS drift tube with the anode wire, which is spanned in the middle of the tube, and field lines of drift field [28].

The drift tubes are filled with a gas mixture of 85 % Ar and 15 % CO_2. Here Ar is the *active gas* while CO_2 is the *quenching gas*[39]. Charged particles traversing the gas volume ionise the Ar-atoms. Due to high voltage between the anode wire and the cathode strips, the resulting ions and electrons are separated and drift towards the cathode strips or anode wire, respectively. When reaching the anode wire, the electrons get extremely accelerated, because of the very high field density. This enables them to ionise further gas molecules, leading to an *avalanche effect* and resulting in a measurable electrical signal. The drift time of the drift tubes is limited to $\sim 380\,\text{ns}$ by the chosen composition of the gas mixture.

- **Cathode Strip Chambers** (CSC):
  In the endcap regions ($0.9 < |\eta| < 2.4$) CSCs with a spatial resolution of $\approx 50\,\mu\text{m}$ have been installed. In principal, they work the same way as the DTs, but due to the stronger and highly inhomogeneous magnetic field as well as the higher expected particle rates, a different detector design has been chosen (cf. figure 3.14).

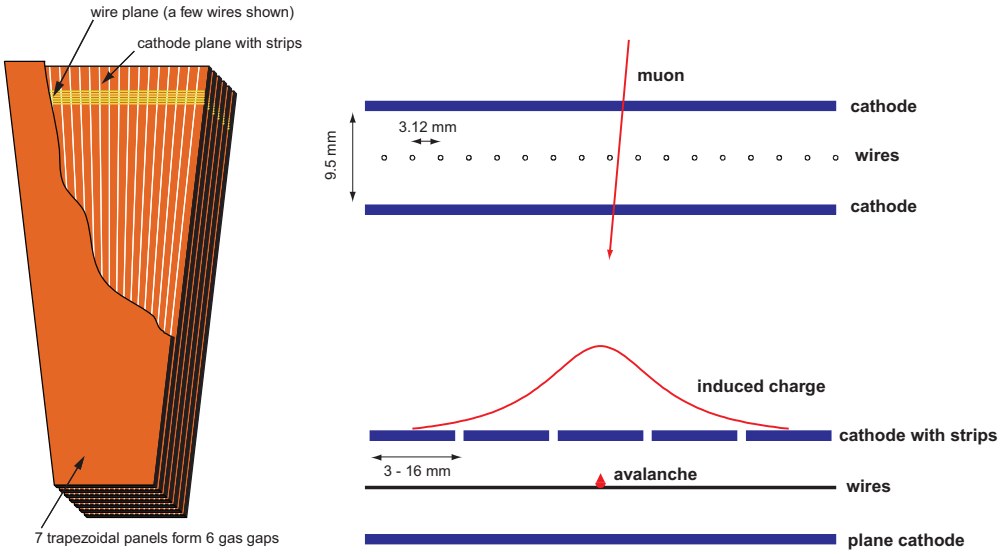


Figure 3.14.: Layout of the Cathode Strip Chambers with only a few wires shown (left) and schematic view of the functionality of a single gas gap (right) [32].

Multiwire proportional chambers with cathode strips are used, filled with a gas mixture of 40 % Ar, 50 % CO_2 and 10 % CF_4. While the cathode strips are radially orientated, the anode wires run orthogonally in azimuthal direction and have a spacing of 3.12 mm. All together, there are six layers of anode wires embedded by seven layers of cathode strip plates. High voltage is applied to the cathode plates, in order to induce electron drift and avalanche effect. The segmentation of the cathodes offers the possibility of measuring three-dimensionally with only one module, by interpolating the track position from the collected charge fractions of several cathode strips.

- **Resistive Plate Chambers** (RPC):
  In order to contribute to the L1 trigger stage and to map muon tracks to particular

[39] The task of the quenching gas is suppressing ionisation caused by photons from exited atoms.

bunch crossings, the muon system needs to have a fast component. Therefore, six layers of RPCs are embedded in the barrel and three layers in each endcap. Their layout is illustrated in figure 3.15. They consist of two thin gas-filled gaps between Bakelite plates. Two of these Bakelite chambers are combined via aluminium strips, which are used for the readout. High voltage is attached to the Bakelite plates, leading to an avalanche effect of free electrons created by incoming ionising radiation. Due to the small width of the gas gap, the avalanche reaches the Bakelite plates and is read out within $\sim 5\,\text{ns}$ (well below the critical 25 ns).

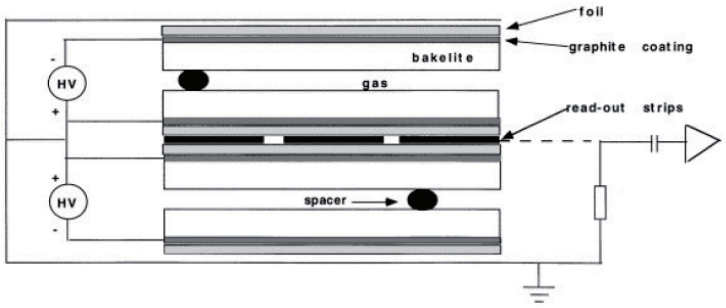


Figure 3.15.: Layout of the Resistive Plate Chambers [34].

The muon momentum resolution that can be achieved, varies from 8 - 60 % or from 20 - 100 %, depending on the η-region and the muon p_T. Hence, it can be heavily improved by combining the data of the muon system with the data of the tracking system, as shown in figure 3.16. Therefore, this is also picked up within the muon reconstruction (cf. chapter 4.2.1).

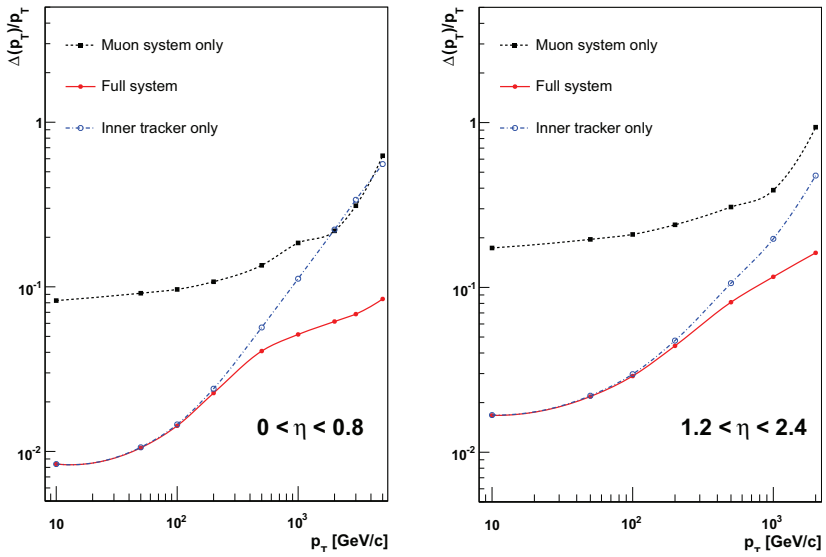


Figure 3.16.: Muon p_T-resolution versus muon p_T plotted for two η regions [28]: Obviously, the resolution can dramatically be improved by combining muon and tracking system.

3.3.7. Trigger

When running at its design luminosity, the LHC will deliver bunch crossings every 25 ns, each causing about 20 particle interactions. Most of these events are soft, i.e. no new particles are produced during the collision. Storing the data of all of these events is neither practicable with today's technology nor necessary. In order to select only interesting events and thus to reduce the event rate which has to be processed, a trigger system [37] has been developed for CMS. It consists of two stages:

- The Level 1 triggers (L1) are hardware based online triggers, meaning that they decide whether to save the events or not, directly after they have been recorded by the detector. In fact, the decision has to be made within 3 µs after each collision, because the data saved in the buffer are overwritten after this period. The L1 triggers leads to a reduction of the event rate from 40 MHz to 100 kHz, which is low enough to be saved and transferred to a computer farm.

- Events passing the L1 trigger are transferred to a computer farm, where they are processed by the second stage of the trigger system, the High Level Trigger (HLT). It is a software based offline trigger, that has more time for making decisions. Thus, it can use reconstruction algorithms to further reduce the event rates. Only permanently storing events that at least passed one HLT criteria, leads to a reduction of the event rate from 100 kHz to 100 Hz, which corresponds to the manageable rate 100 $^{\mathrm{MB}}/_{\mathrm{s}}$ to be stored on tape.

Different triggers exist for each of the trigger stages, which are specialised for finding special event types. There are e.g. muon triggers, deciding whether a muon is contained in the events.

4. Reconstruction Framework

As shown before, the CMS detector basically detects the positions of particle transition (in the tracker an the muon chambers) and the amount of energy, deployed in the different crystals of the electromagnetic calorimeter and the hadronic calorimeter. Since for every analysis the knowledge of the type of the involved particles, their momentum and trajectory is needed, the raw CMS data have to run through a reconstruction process. This chapter explains the particle reconstruction and the software needed, for brevity's and clarity's sake, only covering what is needed for the MUSiC framework (cf. chapter 5.3).

4.1. Software and Analysis Framework

The particle reconstruction is done within the **CMS S**oft**w**are (CMSSW) framework [38]. It is a framework that provides the software needed for the creation of Monte Carlo simulations and reconstruction of both simulated and detected data. It is built in a modular way, consisting of different code modules, written in `C++` and controlled via `Python` based configuration files. Therefore, physicists are able to use the specific software needed for their analysis, while benefiting from the same framework and core software. For this thesis, CMSSW 3.5.8 patch 3 has been used.

4.1.1. LHC Computing Grid

When reaching its design luminosity and centre of mass energy, the data production rate of the LHC experiment with its four detectors will be in the order of 10 PB[40] per year. A central storage of this enormous amount of data at a single location is not sensible with respect to availability for physicists around the world and data safety. Additionally, particle reconstruction and analyses require a high amount of computing power, which cannot possibly be provided by each single institution. Instead, the storage and processing of the data is spread around the world using the *LHC Computing Grid* [39]. It contains the CERN computing centre (where the primary backup will be stored) as well as storage and computing resources at hierarchical organised computer centres around the world. These computer centres are used to store the whole data taken, process them and make them available for physicists. Sending ones analysis program to the LHC Computing Grid rather than downloading the data and processing them locally, results in reduction of both: waiting time for the user and

[40] 1 PB = 1 PetaByte = 1024 TeraByte

required storage and computing capacity at the user's location. More information about the usage of the LHC Computing Grid, its advantages and challenges can be found in [40].

4.1.2. CRAB

In order to provide a user friendly framework for the submission and monitoring of CMSSW based "Grid jobs" a `Python` based tool was developed, the **C**MS **R**emote **A**nalysis **B**uilder (CRAB) [41, 42]. The task of CRAB is to automate the localisation of the desired data samples as well as the preparation and submission of Grid jobs. The only task remaining for the user, is to define jobs in form of CMSSW configuration files. Within the MUSiC framework, a further automation has been achieved by the "music_crab"- and the "watchdog"-script, which partly take over the creation of CMSSW configuration files and the monitoring processes, respectively.

In the course of preparing this thesis, CRAB and the LHC Computing Grid infrastructure were used to access the simulated Monte Carlo data, which were needed to verify the correct functionality and evaluate the benefit of the implementation of b-tagging into MUSiC (chapter 7). More precisely, they were used to run the MUSiC Skimmer on the Monte Carlo samples, which is described in chapter 5.3. The MUSiC Skimmer returns pxlio-files[41], which only contain the information needed as input for the MUSiC analysis and are sufficiently small for local storage.

4.1.3. PAT

For CMS, the data format of the reconstruction output is optimised for a maximum flexibility and high software performance. Additionally, the different reconstruction algorithms, which will be described later in this chapter, are as decoupled as possible, for the purpose of easier maintenance. As a consequence, the data format is not optimal for analyses, as some information is not easily accessible. The aim of the **P**hysics **A**nalysis **T**oolkit (PAT) [43, 44] is to bridge this gap by reorganising the information and thus providing easily accessible collections of analysis objects, like leptons, photon etc. Furthermore, when filling the collections, the objects can be cleaned and selected according to the special needs of every specific analysis, but using the same tools doing so. This approach can simplify the comparison of different analyses, which is a critical point in a collaboration as large as CMS.

4.1.4. PXL

The **P**hysics e**X**tension **L**ibrary (PXL) classes [45] provide `C++` based data containers which are optimised for high energy experiments in particle physics. During the reconstruction of a particle physics data event, all necessary event information can be stored in the PXL

[41] File format corresponding to the PXL data containers (cf. section 4.1.4).

classes, e.g. the particles and vertices of the events, or the 4-momenta of the single particles. Additionally, general information on a whole process can be stored in the pxl::EventView. This can easily be used for separate storage of data from generation and reconstruction of simulated events, which is obligatory for the determination of reconstruction efficiencies.

4.1.5. ROOT

ROOT [46] is a C++ based framework for scientific data analysis. It has been developed to satisfy the specific requirements of modern experimental particle physics with its very large quantities of data. It provides powerful mathematical tools for statistical evaluation of data as well as many libraries with tools for the graphical display of graphs and histograms. As ROOT is C++ based, including and using ROOT classes and libraries is easily possible in C++ based analyses like the MUSiC framework. For example, ROOT classes are used to create the control plots which monitor various aspect of the data during the event class generation (cf. chapter 5.3.2), especially the plots on b-tagging efficiencies and b-jet purity (cf. chapter 6.4.2), and for the kinematic distributions of the event classes within the search algorithm (cf. chapters 5.2.3 and 7).

4.2. Reconstruction and Identification of Particles

Although the present thesis focuses on jets (and the distinction of b-jets and other jets), all physics objects need to be well understood when intending to perform a model unspecific search. The simple reason is, that the event classes of MUSiC (cf. 5.2.2) will consist of several different physics object, combining their information.

Of course, the algorithms used in the particle reconstruction at CMS are not fixed, but may be improved any time. Especially now, as the first studies on the performance of the reconstruction algorithms on collision data get completed, there might be many developments considering these algorithms. Due to this fact, the following sections can only reflect the current status.

4.2.1. Muon Reconstruction

Despite particles from “punch through”, muons are the only particles that are supposed to produce hits in the muon system additionally to the hits in the tracker of the CMS detector. Thus, the standard muon reconstruction starts from the muon system [47]. Hits in the Drift Tubes or the Cathode Strip Chambers are individually connected to local track segments. Each of these track segments is used as a seed for the actual muon track, if it is geometrically matching with the beam spot, as the muons are supposed to be originated from there. Starting from the seeds, muon tracks are reconstructed by combining the hits of all three subdetectors (DT, CSC and RPC). To achieve this, an algorithm based on a

Kalman Filter [48] is used for finding possible tracks and fitting them to the hits. For each track seed, the final fit through the whole muon system results in a **"standalone muon"**, in case the χ^2 of the fit is acceptable.

To enhance the quality of the reconstructed muons, as a second step the hits of the muon system are combined with the hits of the tracker. Therefore, first of all, the process of track fitting is performed. I.e. tracks from the tracker are matched to the standalone muon tracks. Since all types of charged particles are detected by the tracker system, there are more tracks reconstructed in the tracker than in the muon system. So if it comes to more than one tracker track matching to one standalone muon track, the one with the best matching, i.e. the better χ^2 of the matching fit, is chosen. In the last step, one track is fitted to all hits belonging to the two matched tracks. Doing so, a muon is reconstructed by considering all information available, which is called **"global muon"**.

4.2.2. Electron Reconstruction

Electrons produce hits in the tracker and deploy their energy in the ECAL when forming an electromagnetic shower. As these showers spread over several ECAL crystals (typically $\sim 97\%$ of the shower energy will be contained in a 5×5 array of crystals [32]), single crystals have to be combined to clusters, in order to reconstruct the electron energy. Due to the fact, that electrons may lose a large fraction of their initial energy in form of bremsstrahlung before reaching the ECAL, the clustering process is more complicated than just choosing 5×5 arrays of crystals: To cover the deployed energy of the electron together with its photons from bremsstrahlung, superclusters are formed by combining nearby clusters fulfilling certain quality criteria [49].

These superclusters are used to identify seeds for the electron reconstruction within the pixel detector of the tracker. Using the energy estimate, given by the supercluster, the reconstruction algorithm searches for every combination of two hits in the pixel detector and/or inner silicon strip detector[42], which form a track geometrically matching with the beam spot and the supercluster (assuming an electrical charge of ± 1). Matching hits are used as seeds for finding and fitting possible electron trajectories. This is done using a Gaussian Sum Filter [51], which is a generalised, non-linear Kalman-Filter, and is able to take the electrons non-linear energy loss by bremsstrahlung into account. Finally, the electron energy is reconstructed by a weighted combination of the energy deposit of the combined crystals, which has to be corrected to the amount of bremsstrahlung emitted by the electron and the momentum measured by the tracker.

4.2.3. Photon Reconstruction

Just like electrons, photons form an electromagnetic shower in the ECAL deploying their energy there. With respect to the detector response, the only difference to electrons is

[42] A second, complementary approach has been developed in CMS [50]. As it is expected to be well suited for low-p_T and non-isolated electrons, this approach is not used for MUSiC.

the absences of hits in the tracker. Thus, the reconstruction of photons starts with the same clustering algorithms as the electron reconstruction [52, 50]. As a result, every photon candidate is also an electron candidate. The distinction between both types of particles, i.e. the identification of photons, is done e.g. by applying a track veto.

Like in the case of electron reconstruction, photon-specific energy scale corrections are applied when reconstructing the photon energy from the energy deposit of the superclusters. The photon's position when reaching the ECAL (and thus its trajectory) is reconstructed by determining the energy-weighted mean position of the crystals belonging to the supercluster.

4.2.4. Jet Reconstruction

Due to the confinement, quarks hadronise when they are created during an interaction. On their way through the detector, the heavier hadrons can decay into lighter ones as well as emit bosons. As the primary hadrons have a certain momentum, the decay products and the emitted bosons are boosted in their direction of flight. Thus, groups of particles heading into a similar direction are formed, which are called "jets". In figure 4.1 an event containing two jets, which has been observed with the CMS detector at $\sqrt{s} = 7\,\text{TeV}$, is shown in different

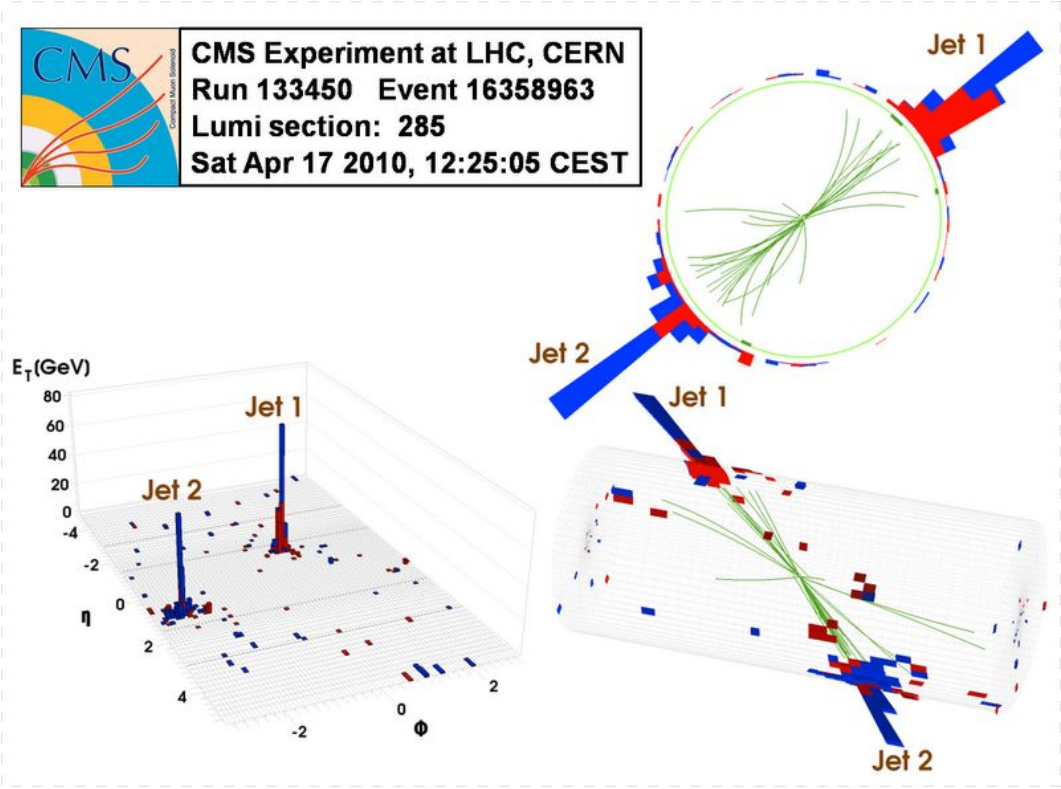


Figure 4.1.: Event display of an event with 2 jets [53], showing η-ϕ view (lower left), ρ-ϕ view (upper right) and 3D view (lower right). The transverse energy deposits in the calorimeter are shown for the ECAL (red) and HCAL (blue), reconstructed tracks are coloured green. The event was observed with the CMS detector at $\sqrt{s} = 7\,\text{TeV}$.

perspectives. In the ρ-ϕ view, one can easily recognise two groups of tracks (i.e. charged particles that moved in similar directions), the η-ϕ view clearly reveals the corresponding groups of calorimeter towers with energy deposition.

The aim of defining jets from the final-state particles is to gain knowledge on the underlying hard parton-level processes. As the LHC is a pp-collider, hadronic processes will play a dominant role in the interactions taking place at CMS. For this reason, a stable, efficient and fast jet reconstruction is necessary. Since jets are not well-defined particles, but groups of final-state particles, different approaches for reconstruction algorithms exist in CMS.

Because of the boost of the particles belonging to a jet, one can assume that jets can be enclosed by a cone, whose centre axis points to the moving direction of the initial parton. **Cone based** algorithms introduce this concept as a circle of fixed radius R in the η-ϕ-plane which should enclose the calorimeter entries belonging to the jet. Starting from a collection of seeding calorimeter towers, e.g. calorimeter towers with a certain minimum energy deposition, the standard cone algorithms search for stable cones. A cone is considered as stable, if the direction of the sum of the momenta of all particles within the cone equals the direction of the centre of the cone. Such a cone is identified as a jet. If the cone is not stable, the position of the cone is slightly changed and one iterates until the cone is found stable. This procedure is repeated until no further stable cones can be found. According to their treatment of overlapping stable cones, the cone algorithms can be divided in two subclasses:

- Splitting (merging) stable cones with an overlap of less (more) than a given fraction f, starting with the hardest pair of overlapping cones. One representative of this subclass is the "Midpoint Cone" algorithm, which is described in [54].

- Removing the calorimeter towers of a cone, which has been found to be stable, from the collection of all calorimeter towers before repeating the procedure. Algorithms using this approach always start with the hardest particle. The "Iterative Cone" algorithm [54] is a typical example.

To provide physically reasonable results, jet algorithms should be *collinear* and *infrared safe*. If an algorithm is collinear unsafe, it is sensitive to small substructures (e.g. if energy is distributed to two very close, collinear particles instead of one particle) and finds different stable cones, depending on the substructure. This is schematically shown in figure 4.2. Infrared unsafety means, adding a very small energy entry (e.g. from a soft gluon) leads to a different result of jet reconstruction (cf. figure 4.3).

Collinear or infrared unsafe algorithms obviously do not find all possible stable cones, which leads to an unwanted sensitivity to small changes in the event and thus to inconsistencies when comparing the results to theoretical predictions from perturbative calculations.

The cone algorithms explained above do not achieve both, collinear and infrared safety, as their seeded approach misses stable cones [54]. Therefore, several alternative algorithms have been developed, inherently fulfilling the criteria of collinear and infrared safety:

A cone based approach is the "Seedless Infrared Safe Cone" (SISCone) algorithm [56, 57].

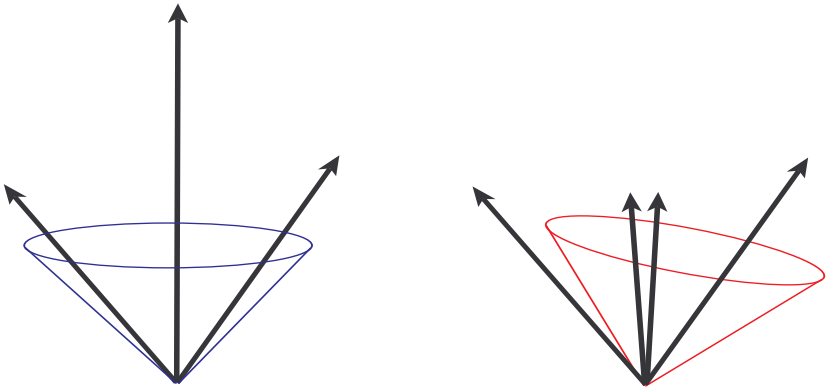


Figure 4.2.: Collinear unsafety of jet algorithms [55]: Using the Iterative Cone algorithm to find stable cones in this three particle event (left side), one stable cone containing all particles is found, as the central particle is the hardest one and thus the first seed. In this case, a 1-jet event will be reconstructed. If the hardest particle splits into two collinear particles (right side), the right particles becomes the first seed and a different jet configuration will be reconstructed. In this case, the algorithm will end up with a 2-jet event. One jet contains the particles within the plotted cone and one jet consists of the other particle.

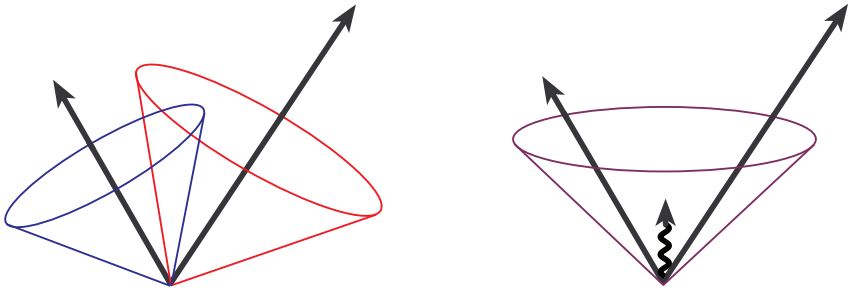


Figure 4.3.: Infrared unsafety of jet algorithms [55]: Using the Midpoint algorithm to find stable cones in this two particle event (left side), two stable cones are found, each containing one particle. Thus, a 2-jet event will be reconstructed. If a soft gluon is added (right side), a third stable cone is found and the two hard particles are reconstructed as a single jet.

This algorithm avoids to miss stable cones (and thus to be collinear or infrared unsafe) by constructing cones without using a collection of seeds. Instead, every cone with radius R in the y-ϕ-plane[43] is considered, which can be constructed with a pair of particles (i.e. calorimeter towers) on its edge. This ansatz is motivated by the simple fact, that the content of a cone can only change, if particles pass the edge of the cone. A cone is considered stable, if a new cone, pointing in the direction of the sum of the momenta of all particles within the original cone, contains the same particles as the original cone. The calorimeter towers of the stable cones are not removed, therefore, a split and merge procedure has to be applied, in order to prevent double counting of particles. This is illustrated in figure 4.4.

[43] Here $y = \frac{1}{2} \ln \left(\frac{E+p_{||}}{E-p_{||}} \right)$ is the rapidity.

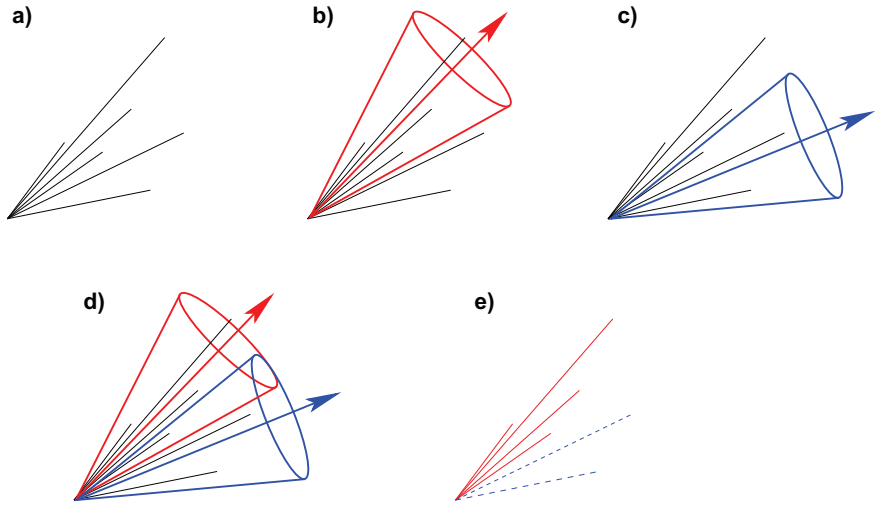


Figure 4.4.: Jet reconstruction with the SISCone algorithm [56]: In the collection of particles (a), two stable cones are identified (b,c) which are overlapping (d). These cones get split into two jets by distributing each particle, belonging to both cones, either to one or the other cone (e).

Another approach are **distance based** algorithms. Figure 4.5 illustrates the reconstruction of jets with distance based algorithms. They define a distance[44] d_{ij} between every pair of particles i and j and the distance d_i of every particle to the beam [58, 59]:

$$\begin{aligned} d_{ij} &= \min(k_{\mathrm{T}\,i}^{2p}, k_{\mathrm{T}\,j}^{2p}) \cdot \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \\ d_i &= k_{\mathrm{T}\,i}^{2p}. \end{aligned} \tag{4.1}$$

Here $k_{\mathrm{T}\,i}$, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. R is the radius parameter and p is a parameter defining the consideration of particles' momenta. The distances are calculated and the smallest one is chosen. If d_i is the smallest distance, the corresponding particle (or particle group) is considered as a jet and not considered anymore, when recalculating the distances. If d_{ij} is the smallest one, the particles i and j are combined. This procedure is repeated until only jets remain, i.e. all remaining particles are too far apart.

The radius parameter R in equation 4.1 ensures that only particles with a geometrical maximum distance $(y_i - y_j)^2 + (\phi_i - \phi_j)^2$ of R are combined and thus can be regarded as the maximum possible radius of jet cones reconstructed with the algorithm. The parameter p distinguishes the different distance based algorithms and defines the influence of the transverse momenta on the distances d_{ij} and d_i:

- The "k_{T}" algorithm [58] is defined by $p = 1$, which leads to a direct weighting of the distances with the transverse momenta. As a result, particles with smaller k_{T} are combined first by trend and only later added to particles with higher k_{T}. This may lead to rather strange and unintuitive jet shapes, which is shown in figure 4.6.

[44] Here the term "distance" is not to be taken literally, as the multiplicand is a relative distance and the multiplier may be a momentum, the integer 1 or an inverse momentum, depending on the value of p.

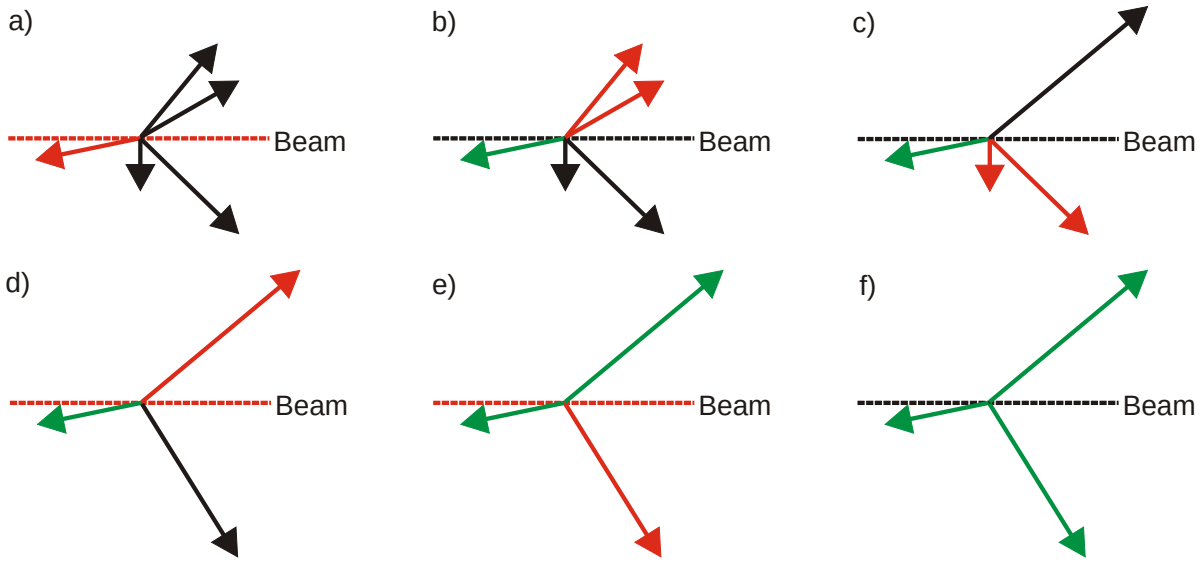


Figure 4.5.: Illustration (based on [55]) of the process of jet reconstruction with distance based algorithms (by the example of the k_T algorithm): The jets of a five-particle event are reconstructed. In each step, the objects with the smallest distance (red) and already identified jets (green) are highlighted. Jets are not considered for the distance calculation anymore. If two (combined) particles possess the smallest distance, they are combined (b $\rightarrow$ c & c $\rightarrow$ d). In contrast, if the distance of one (combined) particle and the beam is the smallest, the particle is considered as a jet (a $\rightarrow$ b & d $\rightarrow$ e & e $\rightarrow$ f). The algorithm stops when only jets remain and thus no distances can be calculated (f).

- The "Cambridge/Aachen" algorithm uses $p = 0$ and thus ignores the transverse momenta when calculating the distances.

- A relatively new algorithm is the "anti-k_T" algorithm [59]. The choice of $p = -1$ results in an inverse weighting of the distances with the transverse momenta. Thus, in contrast to the k_T algorithm, the anti-k_T algorithm tends to add low-k_T particles to the particles with the highest k_T first. Thus, the reconstructed jets have circular shapes (cf. figure 4.6), which matches the intuitive expectation.

With MUSiC, any of the explicated algorithms can be chosen for jet reconstruction. Of course, using a collinear or infrared unsafe algorithm is not advisable. For CMS, performing jet reconstruction with the SISCone and the anti-k_T algorithm is recommended instead. As anti-k_T is found to perform as good as SISCone but needs much less computing time, it is preferred and recommended as default algorithm by the CMS JetMET group [54]. Additionally, the SISCone algorithm shows bad a behaviour with respect to b-tagging [60]. Therefore, anti-k_T is used in the scope of this thesis, using $R = 0.5$[45] for the radius parameter, meaning that only particles with $\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2} \leq 0.5$ are combined.

To properly estimate energy of the reconstructed jets, the minimal required standard L1-L3 jet energy scale corrections [61] are applied. These corrections adjust different detector and

[45] i.e. half as big as used in figure 4.6

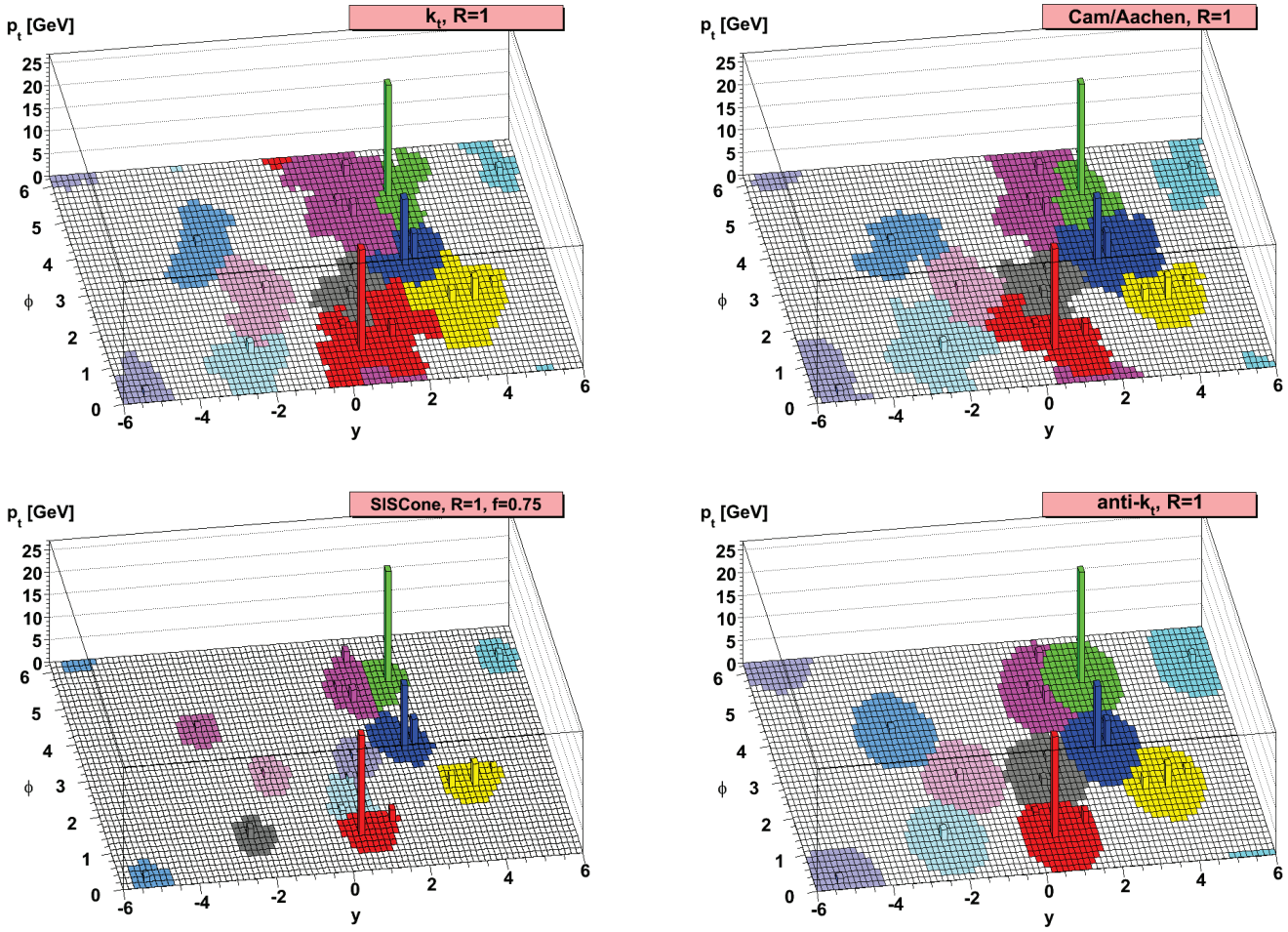


Figure 4.6.: Jet shapes of several jet algorithms [59]: A parton-level event together with $\sim$ 10000 randomly distributed soft particles ("ghosts", only needed to achieve visibility of the calorimeter towers) is used. The distance based algorithms with non-negative values of p (Cambridge/Aachen and k_T algorithm) result in complexly interlaced jet shapes which may totally dissent the intuitive expectation. For example, the purple jet (at $y \sim 2$, $\phi \sim 5$) consists of three low-k_T particles. However, in the upper plots, it gains many of the calorimeter towers, which one would intuitively attach to the green jet, consisting of one much harder particle. The actual shape of a k_T- or Cambridge/Aachen-jet depends on the specific distribution of soft ghosts and thus of soft particles or radiation in a jet. The anti-k_T algorithm and the cone based algorithms (in this case SISCone) result in jets with more intuitive circular shapes, which is not modified by soft particles. While the split and merge procedure of the cone based algorithms equally distributes the shared area of two overlapping jets, the anti-k_T algorithm inherently adds it to the jet with the harder particle.

physics effects:

- L1: An energy offset originated in pile-up and electronic noise is subtracted.
- L2: An angular dependent correction is applied, as the detector response varies with the pseudo-rapidity η
- L3: Like in L2, a p_{T} dependent correction is applied, due to detector response varying with the jet p_{T}.

Therefore, the corrected CaloJet energy can symbolically written as:

$$E_{\mathrm{corr}} = (E_{\mathrm{uncorr}} - E_{\mathrm{offset}}) \times \mathrm{C}(\eta) \times \mathrm{C}(p_{\mathrm{T}}). \tag{4.2}$$

Motivated by the concepts of robustness and generality, no further corrections (e.g. for the electromagnetic energy fraction or the flavour of the jet) are applied for jets in MUSiC.

As data taking at CMS has started, verification of the detector understanding, i.e. the agreement between Monte Carlo predictions and collision data, now becomes possible. First studies of the jet distributions in data at a centre-of-mass energy of 0.9 TeV, 2.36 TeV [62] and 7 TeV [63] shows a good agreement of data and Monte Carlo simulations.

4.2.5. Missing Transverse Energy

Since no detector can hermetically enclose the interaction point, decay products with large longitudinal momentum[46] cannot be detected. Nevertheless, all particles with a transverse momentum of

$$p_T > \frac{E}{\cosh 5} \approx 0.0135 \cdot E \tag{4.3}$$

can be detected [64], due to the large coverage range of the CMS detector ($|\eta| \leq 5$, cf. chapter 3.3.4). Thus, a rather precise survey of the momentum conservation in the transverse plane is possible[47]. Any measured significant imbalance, e.g. caused by neutrinos, leads to missing transverse energy ($\not\!\!E_{\mathrm{T}}$ or MET).

Obviously, $\not\!\!E_{\mathrm{T}}$ depends on the total transverse energy deployment in both calorimeters as well as the transverse momenta of muons, which practically do not deploy energy in the calorimeters. Therefore, $\not\!\!E_{\mathrm{T}}$ reconstruction consists of several steps and has to undergo several corrections:
The $\not\!\!E_{\mathrm{T}}$ reconstruction starts with summing up the deployed transverse energy of all calorimeter towers, then corrections are applied, which consider the jet energy scale corrections of the jets in the event. After that, the energy of muons leaving the detector has to be regarded, their small calorimetric energy deposition is vectorial added to $\not\!\!\boldsymbol{E}_{\mathrm{T}}$ and their momenta are vectorial subtracted. Additionally, corrections are applied, compensating effects of jets from

[46] i.e. momentum in beam direction
[47] Even with a hermetically inclosing detector, the conservation of the longitudinal momenta could not be surveyed, as the momenta of the interaction partons are unknown

decaying τ leptons and adding track based corrections as well as a dedicated treatment of unclustered calorimeter entries and out of cone effects [65].

$\not{E}_{\mathrm{T}}$ relies on all detector components and is therefore extremely sensitive to not understood detector effects and detector malfunctions. Thus, its detailed understanding is an absolute requirement for the discovery of $\not{E}_{\mathrm{T}}$-based signatures of new physics. Gladly, a first study of the $\not{E}_{\mathrm{T}}$ distributions in collision data at a centre-of-mass energy of 7 TeV [66] came to the conclusion, that the measured $\not{E}_{\mathrm{T}}$ distribution is "in general agreement with expectations from the Monte Carlo simulation" after removing anomalous signals in the calorimeters.

5. Model Unspecific Search in CMS

The goal of a model unspecific search is to find deviations from the Standard Model in the data of a particle physical experiment. In order to achieve this, the selection cuts (selecting the events used for the analysis) are not optimised with respect to any signal, expected from new physics, but only serve the purpose to select events exclusively containing well understood objects. Doing so, a large fraction of the available data can be analysed, using a minimal number of assumptions. This is the precondition for finding unexpected deviations from the Standard Model, which is represented in Monte Carlo simulations. Similar approaches have already been applied successfully to other accelerator experiments before (e.g. [67, 68, 69, 70, 71]).

By now, there have been several publications by the developers of MUSiC and the members of the RWTH MUSiC group describing the motivation, concept, realisation and testing of MUSiC in a very detailed way (e.g. [72, 40, 73]). Therefore, I will restrict my description of the MUSiC framework to a short introduction to the concept of Model Unspecific Search and its realisation in CMS and explicitly refer to these previous publications.

5.1. Motivation

At the LHC, collision data is and will be taken at an energy scale never reached by any accelerator experiment before. Clearly, the hope is to find hints and evidences for new physics in the data. As shortly introduced in chapter 2.1.7, there are many reasons to expect new physics at these high energies. Unlike in previous experiments, where there was a manageable number of predictions[48], this leads to an almost infinite number of new theories of how this new physics will look like. The theorist Hitoshi Murayama very nicely summarised the situation in the picture shown in figure 5.1: It shows an extract of the schematic "landscape" of new physics, which could theoretically be realised in nature at high energies. As can be understood easily, even this extract could lead to an enormous number of topologies, which have to be studied. To achieve this, a very big number of dedicated analyses is needed. In case of Supersymmetry with its more than 100 free parameters (cf. chapter 7.3), an extremely large number of dedicated analyses would have to be performed. But the even greater challenge is represented by the areas marked as "not yet thought of". Of course, it is quite certain, that NOT every single possible theory has been thought of yet. Now, if a theory should be realised in nature, which has NOT been thought of, there will be no dedicated analysis searching for its resulting new physics and its (unexpected) signal may be missed.

[48] e.g. prediction of the existence of the t-quark

Figure 5.1.: Schematic "landscape" of possible new physics, realised in nature at high energies. Created by Hitoshi Murayama [74].

A solution of this problem can be provided by the model unspecific search. Following the saying 'expect the unexpected', it should systematically analyse the data with as little bias as possible[49], in order not to miss any possible unexpected signal. Thus, MUSiC should be sensitive to all kinds of deviations from the Standard Model, may they be caused by expected or unexpected signals of new physics. Especially in the first months after the beginning of taking data, MUSiC can be useful to improve the Monte Carlo Standard Model background simulations and the understanding of the detector, as at the start-up, MUSiC can quickly find discrepancies caused by detector effects[50] or suboptimal consideration of certain effects within the Monte Carlo simulations. On the other hand, this approach also has its drawbacks: For signals, where also a dedicated analysis exists, MUSiC is likely to be less sensitive. And it depends much more on a correct description of the background predictions by Monte Carlo generators, since a variety of final states is investigated.

One thing has to be made very clear: If a significant deviation is found with MUSiC, the conclusion, whether it is a discovery or not, can most likely only be drawn by more dedicated studies. Therefore, MUSiC has to be regarded as a global physics monitor, giving hints for interesting deviations from the Standard Model and being a complementary addition to the CMS physics program.

For the sake of completeness one has to mention, that there is also the possibility, that MUSiC might miss single models of new physics, too. E.g. MUSiC is probably not able to find the Standard Model Higgs (or at least not with the comparatively low statistics, which are sufficient for dedicated searches) [72].

[49] i.e. especially without selection cuts optimised on finding any specific topology

[50] This has been proved to be right, when the ECAL spikes (single noisy ECAL cells, faking high p_T photons) were found within the first data (e.g. [75]).

5.2. Concept of MUSiC

The concept of model unspecific searches fundamentally differs from the concept of the traditional, signal-driven searches for new physics. Searching for one certain theoretical model, the basic concept of traditional analyses is to find the separation cuts and final variables, leading to an optimal signal versus background ratio. To serve this goal, separation cuts are highly tuned to the theoretical model, searched in the analysis. In contrast, there is no optimisation with respect to a certain signal within MUSiC but the following guidelines have to be fulfilled:

- **Model independence:** No optimisation of selection cuts with respect to a certain expected signal.
- **Robustness:** Focus on well-understood physics objects, i.e. high p_{T}, central $|\eta|$ and solid object identification.
- **Simplicity:** The steps of the algorithm should be easy to follow, well tested statistical estimators and methods should be preferred.
- **Completeness:** Include any possible (systematic) uncertainties on the Standard Model prediction in the search algorithm.
- **Extended search:** Expect new physics contributed predominately to a single channel (resonance likes W') as well as new physics causing deviations in numerous final states (SUSY).

While the first item leads to separation cuts exclusively securing a solid object identification, the last item results in one of the basic concepts of MUSiC, the **classification of events**, which will be explained in chapter 5.2.2.

5.2.1. Lepton-Gamma-Triggering

In order to have a well defined trigger stream and in order to reduce the QCD multi-jet background, the events to be used within the analysis with MUSiC are restricted to events containing at least one charged lepton (electron or muon) or a photon. This LG-triggering[51] is also in accordance with the guideline of robustness mentioned above.

5.2.2. Event Classification

All events passing the selection cuts and the LG-triggering are sorted into classes, as shown in figure 5.2. Each *event class* is defined by the content (i.e. the amount of different physics

[51] "L" representing "lepton", "G" standing for "gamma"

objects, e.g. 1μ 3jet) of the events forming it, therefore, the event classes group the selected events according to their final state topology. Both, *exclusive* and *inclusive* event classes are filled. While an exclusive event class only consists of events, containing **exactly** the required number of physics object (e.g. 1 μ, no e, no γ, 3 jets, no $\not{E}_{\mathrm{T}}$ for the 1μ 3jet class), an inclusive event class is filled with all events, containing **at least** the required number of physics object and possibly something else (e.g. 2 μ, 1 e, 1 γ, 5 jets, $\not{E}_{\mathrm{T}}$ in the 1μ 3jet + X class). The latter are denoted with the suffix "+ X", which represents all physics objects exceeding requirements of the class[52]. While each selected event is present in exactly one exclusive event class, it will populate several inclusive event classes in most cases. The example event (1 μ, no e, no γ, 3 jets, no $\not{E}_{\mathrm{T}}$) is only filled into the exclusive 1μ 3jet class, but is present in the following inclusive classes: 1μ 3jet + X, 1μ 2jet + X, 1μ 1jet + X and 1μ + X.

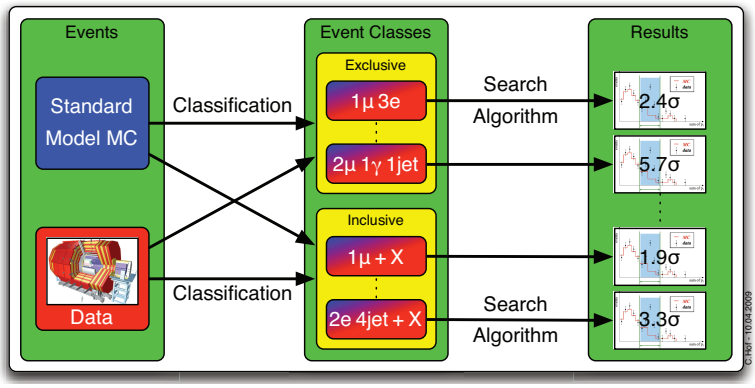


Figure 5.2.: Realisation of the concept of event classification in MUSiC [40]: Data events from the CMS detector and simulated events from Monte Carlo generators (MC) are classified by their topology and filled into the corresponding event classes. Within each class, distributions of certain sensitive variables are passed to a search algorithm, searching for deviations between data and simulation (cf. chapter 5.3).

Due to the overlaps, the statistical treatment of inclusive classes is more complicated, but their ability to combine many exclusive event classes might be useful when searching for new physics with complex decay chains (e.g. SUSY). Another example of the desirability of the event classes combination are events with a high jet multiplicity. Since Monte Carlo generators are not expected to correctly model the kinematics of a 8th or 9th jet, similar events with up to a certain jet multiplicity (e.g. 5 jets) might be investigated in exclusive classes (e.g. 1μ 1jet, 1μ 2jet, 1μ 3jet, 1μ 4jet and 1μ 5jet), while similar events with more jets are grouped into a single inclusive event class (e.g. 1μ 5jet + X).

[52] When filling the inclusive event classes, alike physics objects are ordered by their transverse momentum p_{T} and considered beginning with the one, possessing the largest p_{T}. This implies, that only the muon and the jet with the largest p_{T} enter the 1μ 1jet + X class, although the events entering this class might contain several muons, jets and other physics objects.

In order to perform such a classification, a clear definition of the physics objects is required. Currently the MUSiC analysis considers the following objects measured by the CMS detector:

- muons (μ)
- electrons (e)
- photons (γ)
- hadronic jets (jet)
- hadronic b-jets (b; optional physics object)
- missing transverse energy ($\not{E}_{\mathrm{T}}$)
- sum of the lepton charges ($Q = \sum Q_e + \sum Q_\mu$; optional physics object)

By now, two optional physics objects are available, in addition to the originally implemented physics objects. The implementation of hadronic b-jets is part of this thesis and will be further described in chapter 6. One could think of implementing even more physics objects (like the τ) in the future, but due to the guidelines of MUSiC, this would only be reasonable if their identification is well-studied and well-controlled in CMS.

When aiming to implement new physics objects, one should also be aware of the rapid increase of the number of event classes with an increasing number of physics objects: Considering only the basic physics objects (μ, e, γ, jet and $\not{E}_{\mathrm{T}}$) leads to a couple of hundreds event classes (exclusive and inclusive each), additional consideration of b-jets might increase this number by a considerable factor. As all event classes have to be investigated, an increasing number causes the computational time to increase as well as the penalty factor caused by the "look-elsewhere-effect" (cf. section 5.3.5).

5.2.3. Searching for Deviations

The essence of MUSiC is the search for deviations between the Standard Model prediction (represented by Monte Carlo simulations) and the data obtained with the CMS detector. These deviations can be quantified by comparing statistical distributions of every single event class as it would be expected (filled only with simulated events) and as it is seen in data (cf. also figure 5.2). In order to find deviations, the following variables are systematically analysed:

- the **total cross section**, i.e. number of events per class.
- **kinematic distributions** of an event class:
  - The **scalar sum of the transverse momentum** $\sum p_{\mathrm{T}}$ of all physics objects required by the event class. E.g. $1\mu\ 1e\ \not{E}_{\mathrm{T}}$ + X class, $\sum p_{\mathrm{T}} = p_{\mathrm{T}}(\mu) + p_{\mathrm{T}}(e) + \not{E}_{\mathrm{T}}$ is calculated. The "X" is not regarded.

– The **invariant mass** $M_{\text{inv}} = \sqrt{(\sum_i E^i)^2 - (\sum_i \boldsymbol{p}^i)^2}$ of all physics objects required by the event class. The invariant mass of the decay products of an on-shell particle[53] corresponds to the particles mass. Thus, if new physics should express via new heavy particles, it will lead to a peak in the M_{T} distribution. As the z-component of missing energy is unknown, for classes containing $\not{E}_{\text{T}}$, the **transverse mass** $M_{\text{T}} = \sqrt{(\sum_i E^i_{\text{T}})^2 - (\sum_i \boldsymbol{p}^i_{\text{T}})^2}$ is used. Here $\boldsymbol{p}_{\text{T}} = (p_x, p_y, 0)$ is the vector of the transverse momentum.

– The **missing transverse energy** distribution for event classes which contain $\not{E}_{\text{T}}$. Particles, which leave the detector undetected, lead to the presence of $\not{E}_{\text{T}}$. Therefore, new physics containing particles not interacting with the detector, e.g. gravitons or neutralinos, will express through the presence of much $\not{E}_{\text{T}}$.

These variables are expected to be sensitive to new physics as well as deviations caused by a limited understanding of the detector or imperfect tuning of the event simulation and generation.

5.3. MUSiC Framework

In order to realise the approach of a model unspecific search, a framework has been created, which is used to process the necessary steps. The different parts of this framework are described in the following sections. A summary of the work flow can be found in figure 5.3.

5.3.1. MUSiC Skimmer

The main input to the analysis with the MUSiC framework are of course the data and the Monte Carlo simulations. Those are available in form of CMSSW-RECO files, which contain all relevant information from the reconstruction process, e.g. particle candidates, their tracks, vertices, trigger information, b-tagging information, or Monte Carlo information (only for generated events). All in all, the amount of about 1 - 10 MB per event is stored in these files. The MUSiC Skimmer extracts the information which is relevant for the analysis and stores them in the pxlio-format provided by the PXl package (cf. chapter 4.1.4). Thus, the amount of stored data is reduced to about 5 kB per event, which leads to file sizes small enough for local storage (and access).

5.3.2. Classification and Quality Control

In the next part of the MUSiC framework, selection cuts are applied to the stored physics objects in order to assure that only well understood physics objects are used for the analysis,

[53] On-shell particle fulfil the equation $m^2 = E^2 - \boldsymbol{p}^2$.

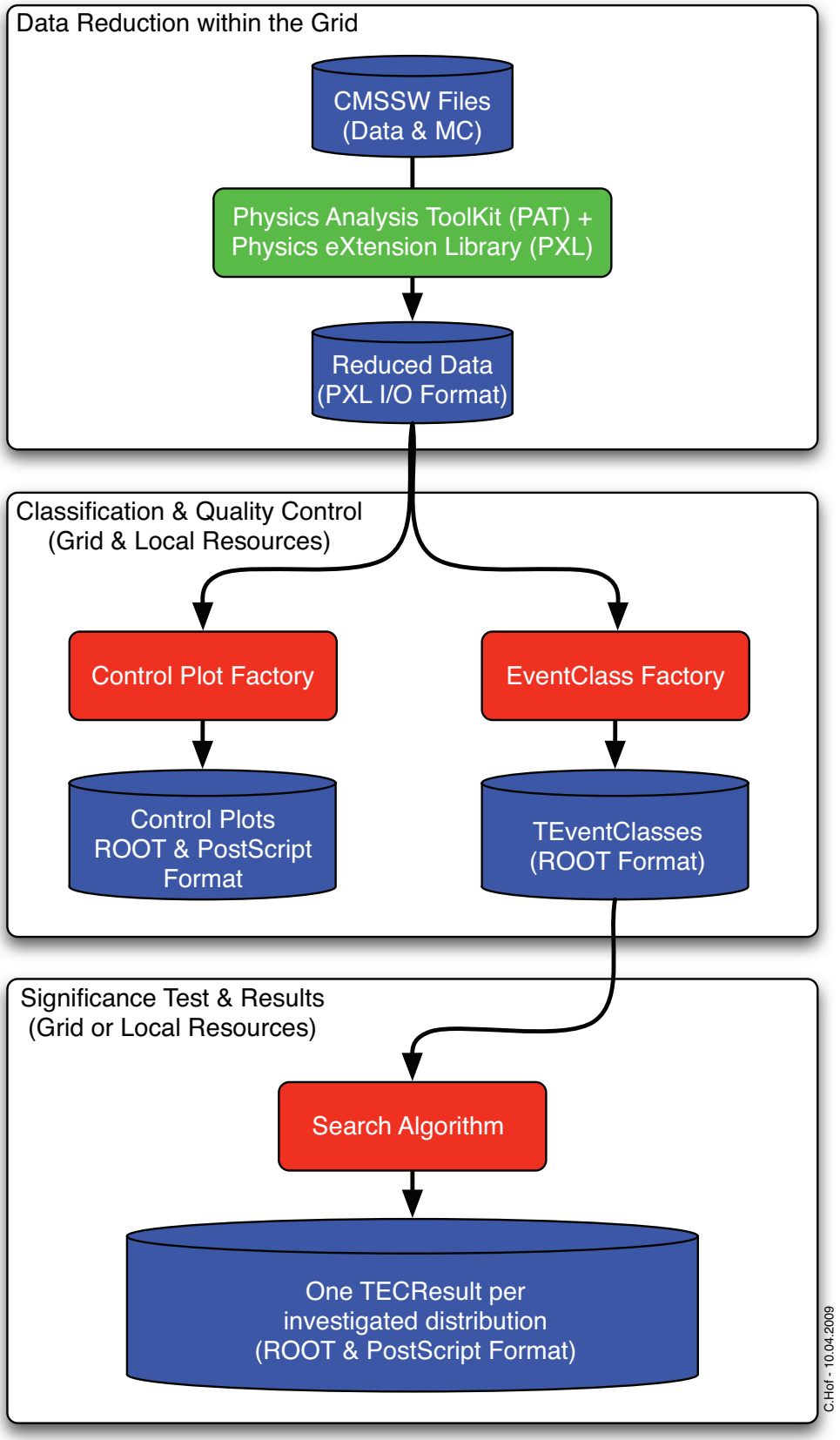


Figure 5.3.: Summary of the work flow within the MUSiC framework [40].

as demanded by the concept of robustness (cf. chapter 5.2). The selection criteria used in the scope of this thesis are presented in the next section. Then, the particles and events that passed the selection cuts are processed twice in parallel: Several hundred control plots are produced, allowing the monitoring of quality and correctness of the selected objects. These control plots for example show the p_T-, ϕ-, η- and reconstruction efficiency distributions for the different types of physics objects[54]. They are stored in a `ROOT`-file as well as in a ps-file for further analysis. At the same time, the events are distributed to the event classes. For each event class, kinematic distributions described in section 5.2.3, which form the base of the later search for deviations, are filled. They are stored to a `ROOT`-file, together with additional information, such as the contained process samples or the assumed luminosity.

5.3.3. Selection Criteria

Starting with a list of physics object candidates, one has to assure that there is not more than one object reconstructed of the same particle. Therefore, the first selection criteria applied on the candidates tries to exclude double-reconstructions by demanding a minimal distance between the physics objects. The applied criteria are summarised in the table below, where not selected objects are removed from the particle list:

criterion	**selected object**
muon candidates with $\Delta R < 0.2$	candidate with the track with lowest value of χ^2
electrons with $\Delta R < 0.2$, either sharing track or supercluster	candidate with the highest estimated energy
photons with $\Delta R < 0.2$ sharing the supercluster	candidate with the highest estimated energy

The next step of the particle selection is requiring at least one electron, muon or photon in each event, as described in the concept of LG-Triggering (cf. chapter 5.2.1). All events not passing this criteria are removed from the event list. Additionally, a list of triggers can be defined. By requesting each event to pass at least one trigger, Monte Carlo events, which would not be stored permanently when appearing in collision data (cf. chapter 3.3.7), can be removed. The full list of triggers used for MUSiC in the scope of this thesis can be found in the appendix (table A.1).

Subsequently, further selection criteria are applied, unique to each type of physics object (equal motivations are not repeated):

- **Electrons:**
  - $|\eta| < 2.5$, corresponding to the $|\eta|$-range covered by both, calorimetry and tracker.
  - $p_T > 30\,\text{GeV}$ for selection of well measured high energy electrons, in order to avoid the higher fake contributions expected for low momenta.

[54] Some control plots, which are produced for the purpose of monitoring b-tagging, can be found in chapter 6.4.

  - $\sum p_{\mathrm{T}}(\text{all tracks})/p_{\mathrm{T}}(\text{electron}) < 0.1$ within $\Delta R = 0.3$. This demanded track isolation rejects non-isolated electron, e.g. from hadronic jets, which we are not interested in.

  - $d_0 < 2\,\mathrm{mm}$ for the distance between the reconstructed primary vertex and the track of the electron candidate as discrimination against objects from pile-up interactions.

  - A "tight" electron identification [76] implementing a set of cuts with a high background rejection and a robust performance with respect to misalignment. The following variables are used:

    * H/E, which is the ratio between energy deposited in the HCAL and in the ECAL. Jets faking electrons very likely possess a higher value of H/E.

    * $\sigma_{\eta\eta}$, which is the shower shape variable, considering the η-extent of the supercluster. Electrons have a characteristic shower shape in η-direction.

    * $\Delta\eta_{\mathrm{in}}$, which is the difference between the η-positions of the supercluster and the associated track at the estimated vertex. $\Delta\eta_{\mathrm{in}}$ should be very small, as the tracks are not bent in η-direction.

    * $\Delta\phi_{\mathrm{in}}$, which is the difference between the ϕ-positions of the supercluster and the associated track at the estimated vertex. If bremsstrahlung has been correctly considered, they should be close to each other.

    * The ration $E_{\mathrm{seed}}/p_{\mathrm{in}}$ of the energy of the seeding supercluster and the estimated initial track momentum at the vertex.

    Different cuts on these variables are used for electrons with different characteristics (cf. table A.2) as well as for the barrel and the endcap regions of the ECAL. A list of the cuts can be found in table A.3.

- **Muons:**

  - $|\eta| < 2.1$

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

  - Track isolation of $\sum p_{\mathrm{T}}(\text{all tracks})/p_{\mathrm{T}}(\text{muon}) < 0.1$ within $\Delta R = 0.3$.

  - $N_{\mathrm{hits}} > 11$ & $\chi^2/\mathrm{dof} < 10$. The characteristics of the global track fit are used to select well reconstructed muon candidates and suppress mismeasured ones.

  - $d_0 < 2\,\mathrm{mm}$ as discrimination against objects from pile-up interactions and cosmic muons.

  - $C_{\mathrm{comp}} = 0.8 \cdot C_{\mathrm{calo}} + 1.2 \cdot C_{\mathrm{segment}} > 1$ is a quality criterion for muons, exploiting

the properties of muons as minimum ionising particles. C_{calo} is the probability of the muon candidate to be a muon, calculated from the energy deposits in the calorimeters. C_{segment} is the probability of the hits in the moun system to be caused by a muon. If for example a muon chamber registers no hit although is should have been passed by a muon, C_{segment} obtains a small value.

- **Photons:**
  - $|\eta| < 2.5$
  - $p_{\text{T}} > 30\,\text{GeV}$
  - Track isolation of $\sum p_{\text{T}}(\text{all tracks})/p_{\text{T}}(\text{photon}) < 0.1$ within $\Delta R = 0.3$.
  - HoEm $= {}^{E_{\text{HCAL}}}/_{E_{\text{ECAL}}} < 0.2$ to reject fake objects.
  - Pixel seed veto = true, to reject electrons.
  - Shower shape variable $R9 = {}^{E_{3\times3}}/_{E_{\text{SC}}} > 0.9$, where $E_{3\times3}$ is the energy deposit in the central 3×3 cluster and E_{SC} the energy deposit in the entire supercluster. Requiring a large $R9$ value helps to separate photons from hadronic showers, which tend to have broader showers.
- **Jets:**
  - $|\eta| < 2.5$
  - $p_{\text{T}} > 60\,\text{GeV}$
  - Hadronic energy fraction of at least 5 % to separate jets from photons and electrons.
- **Missing transverse energy:**
  - $\not{E}_{\text{T}} > 100\,\text{GeV}$, in order to ensure a robust identification.

5.3.4. MUSiC Search Algorithm

After the event classes have been filled, the `ROOT`-file they have been stored in can be processed by the MUSiC Search algorithm. It scans the kinematic distribution of the event classes for deviations from the background, i.e. Standard Model prediction, which of course has to be scaled according to luminosity and cross section. Within each distribution, the algorithm searches for that region of connected bins, which possesses the most significant deviation between data and background entries. This is illustrated in figure 5.4. Therefore, the entries of all bins, belonging to a possible region, are summed up. This is done for data and background individually, resulting in the total number of entries for data (N_{data}) and

for the Standard Model prediction (N_{SM}). To determine the deviation's significance, the uncertainties of the background prediction ($\delta(N_{\mathrm{SM}})$) have to be taken into account. Therefore, the probability p is calculated, of how likely the prediction fluctuates up to or above the number of events seen in the data:

$$p = \begin{cases} \sum_{i=N_{\mathrm{data}}}^{\infty} A \int_0^{\infty} \exp\left(\frac{-(b-N_{\mathrm{SM}})^2}{2\left(\delta(N_{\mathrm{SM}})\right)^2}\right) \cdot \frac{e^{-b}b^i}{i!}\,\mathrm{d}b & \text{for} N_{\mathrm{data}} \geq N_{\mathrm{SM}} \\ \sum_{i=0}^{N_{\mathrm{data}}} A \int_0^{\infty} \exp\left(\frac{-(b-N_{\mathrm{SM}})^2}{2\left(\delta(N_{\mathrm{SM}})\right)^2}\right) \cdot \frac{e^{-b}b^i}{i!}\mathrm{d}b & \text{for} N_{\mathrm{data}} < N_{\mathrm{SM}}\,, \end{cases} \tag{5.1}$$

where A is a normalisation factor. This is done by convoluting the Poisson distribution of the background's statistical fluctuation around its true value b with the Gaussian distribution of the fluctuation of b around N_{SM}. This fluctuation of b is due to the fact that, the true value of the background prediction is unknown and depends on the Monte Carlo prediction N_{SM} and its uncertainty $\delta(N_{\mathrm{SM}})$.

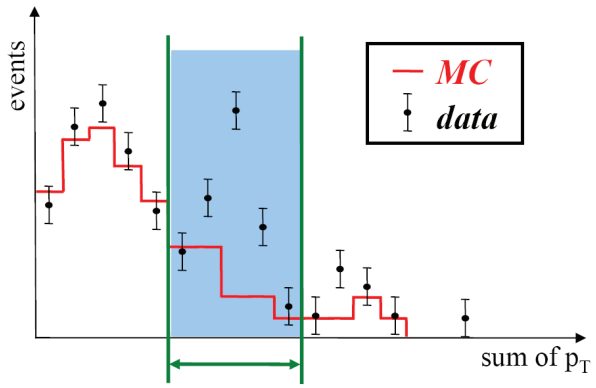


Figure 5.4.: Illustration of the search for the Region of Interest, meaning the region of combined bins with the most significant deviation between data and background [72].

From all possible regions of one distribution, the one with the smallest p-value (p_{min}) is the one in which the most significant deviation is found. It is called the "*Region of Interest*" of this distribution.

As the background description relies on Monte Carlo samples, which are scaled to the integrated Luminosity of the collision data, low statistics within the background distributions leads to a problem when determining the p-value: A bin with data entries but no background (and thus no background uncertainty) would possess a vanishing p-value. This equals to an infinite significance of the deviation, which is an overestimation, caused by the insufficient Standard Model prediction[55] of the background sample ($N_{\mathrm{SM}} = 0$). In order to avoid this and to estimate the significance conservatively, the background of such bins is filled up. This procedure tends to result in high p-values for the specific bins and a conservative estimation of the overall p-value of the considered region. However, filling up empty bins can also lead to

[55] i.e. prediction with insufficient statistics

unintuitive behaviour of the MUSiC search algorithm: When performing a dedicated analysis, one chooses a final state with as little background as possible. Then one tries to remove as much background of the investigated final state as possible (the whole background in an ideal case). The aim of this procedure is to achieve a maximal fraction of signal to background events within the data. For MUSiC, this procedure may not work due to the fill-up. Investigating an event class (= final state) with much data and almost no background may result in a small, very underestimated significance, instead of the high significance such an event class should possess. The reason is, that the empty background bins may be filled up with as many background entries as there are data entries (in an extreme case). Therefore, one should think about if there might be another solution to the problem of bins without background than the fill-up procedure.

5.3.5. Penalty Factors for a Diversified Search

As explained before, the Regions of Interest are determined by choosing the region with the smallest p-value from a set of many possible regions. Unfortunately, the p-value is not sufficient to claim any evidence for a signal. Due to the large number of regions investigated, a statistical penalty factor needs to be applied. This is necessary in order to account for the fact, that the more distributions are investigated, the more likely is a statistical fluctuation with an apparent high significance. This is called the "look-elsewhere-effect" [77]. Hence, it is important to determine the probability $\widetilde{P}$ of finding a smaller p-value than $p_{\text{min}}^{\text{data}}$ for any region within the distribution. $\widetilde{P}$ is the searched event class significance (after taking the "look-elsewhere-effect" into account). As, in general, there are correlations between the possible regions, due to overlaps between different regions or correlated errors, finding a general analytical formula for the determination of $\widetilde{P}$ is extremely challenging, if not impossible. A possible solution of this problem is approximating the result by dicing a large number of background-only “pseudo experiments”. With every pseudo experiment which is

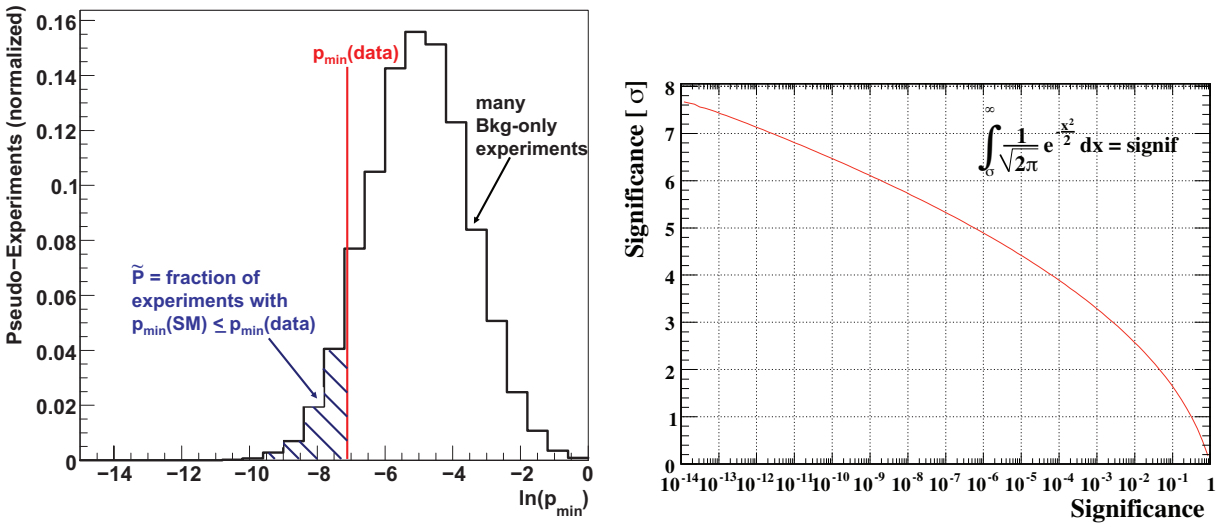


Figure 5.5.: Illustration for the definition of $\widetilde{P}$ [72] (left) and translation of $\widetilde{P}$ into number of standard deviations σ [40] (right).

diced, the entries of all background bins are fluctuated, considering statistical and systematic uncertainties and taking correlations into account. The systemic uncertainties, which are included in the MUSiC framework, are discussed in section 5.3.6. $\widetilde{P}$ is then defined as the fraction of background-only pseudo experiments possessing a Region of Interest with a deviation (caused by the expected statistical and systematic uncertainties of the Standard Model predictions) at least as significant as observed in the data:

$$\widetilde{P} = \frac{\text{number of SM-only pseudo experiments with } p_{\text{min}}^{\text{SM}} \leq p_{\text{min}}^{\text{data}}}{\text{total number of SM-only pseudo experiments}} . \tag{5.2}$$

This is illustrated in figure 5.5. This figure also shows, that $\widetilde{P}$ can directly be translated into standard deviations.

If a global search is performed, not investigating the distributions of one but many event classes, an additional trial factor has to be applied to account for the large number of final states looked at. Conservatively neglecting correlations between the event classes (which is of course not true for at least the inclusive event classes), the global experiment significance of one event class can be estimated as [72]

$$\widetilde{P}_{\text{CMS}} = 1 - (1 - \widetilde{P})^n, \tag{5.3}$$

when considering n distributions in total. In figure 5.6 the impact of this trial factor is displayed for several numbers of investigated distributions.

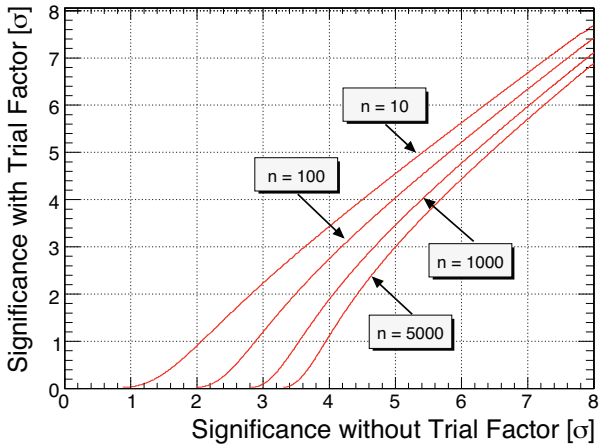


Figure 5.6.: Translation of the significance per event class $\widetilde{P}$ into a global-experiment significance by accounting for trial factors due to the total number n of investigated distributions [40].

Since the interpretation of $\widetilde{P}$ as significance of the deviations should not lead to an overestimation of the true significance (what is called undercoverage), the coverage properties of $\widetilde{P}$ have been studied [73]. As a result one can state, it has good coverage properties, but may suffer from undercoverage in some cases. Therefore, in order to provide the possibility of performing crosschecks of the results of $\widetilde{P}$, other possible functions have been implemented into the MUSiC framework [73] and can be used for describing the background fluctuation in equation 5.1.

5.3.6. Systematic Uncertainties

As explained before, the correct systematic uncertainty estimates have to be considered within the MUSiC search algorithm, as they contribute to the uncertainties of the Standard Model prediction. Thus, only by implementing the systematic uncertainties, deviations caused by detector effects or an incorrect theoretical prediction can be distinguished from deviations caused by a signal.

The following systematic uncertainties are included in the MUSiC framework[56]. They have been estimated in the context of $100\,\mathrm{pb}^{-1}$ to $10\,\mathrm{fb}^{-1}$ of data, which have been assumed for some of the benchmark scenarios used for testing MUSiC with b-tagging (cf. chapter 7).

- σ(integrated luminosity) = 5%
  Assuming $100\,\mathrm{pb}^{-1}$ to $10\,\mathrm{fb}^{-1}$ of data, a luminosity error of 5 % should be realistic value [78], as the aspired 1 - 2 % cannot be reached with this amount of data.

- σ(cross sections) = 10%
  The cross section uncertainty estimated by studying the expected PDF uncertainties typically possesses values of 2 - 8 % [79]. It is applied to all Standard Model background processes, without e.g. distinguishing between different p_T ranges. As the understanding of these uncertainties may change in the future, a conservative value has been chosen for MUSiC.

- σ(jet energy scale) = 5%
  The jet energy scale uncertainty describes how precise the energy of a jet can be measured within the CMS detector, which is also applied to the determination of $\not\!\!E_\mathrm{T}$. It is expected to be in the order of 5 % [61]. Also muons and electrons should possess an energy scale uncertainty, but they are expected to be negligible.

- σ(efficiency correction factor) = 2% for e, μ, γ and 1% for jets
  The efficiency correction factor considers possible efficiency deviations between data and Monte Carlo prediction. They result from complex studies, which are performed by the various physics object groups (POG). They are determined by data-driven methods, e.g. using the tag-and-probe technique on $Z \to \mu\mu$ events.

- σ(fake probability) = 100% for e, μ, γ
  Fake "rates" give the probability of reconstructing one object as another, e.g. reconstructing a jet as an electron. They are determined using Monte Carlo generator information. A reconstructed particle is a "fake" if no matching to a generated particle within $\Delta R < 0.2$ is possible. As a conservative "guess", 100 % uncertainty of the fake probability is assumed for electrons, muons and photons. For jets and $\not\!\!E_\mathrm{T}$ it is assumed, that the dominant uncertainties are covered by the jet energy scale error. Thus, no σ(fake probability) is applied to those objects.

[56] At the moment, all systematic uncertainties are included with constant values, i.e. their possible dependence of different physical processes/Monte Carlo samples is neglected

6. Implementation of b-Tagging into MUSiC

As mentioned before, this thesis concentrates on model unspecific searches (using MUSiC) for new physics containing b-hadrons and thus, b-jets. In order to do this, b-tagging had to be implemented into the MUSiC framework and its functionality had to be tested. This is described in the following chapter.

6.1. What is b-Tagging?

When a *b*-quark is produced at the interaction point (or *primary vertex*) of a particle collision[57], it hadronises like every quark, except the *t*-quark (cf. chapter 2.1.4), and departs from the primary vertex. If such a hadron itself decays, a *secondary vertex* is formed, which is the origin of all particles created in that decay. This is illustrated in figure 6.1. A jet (cf. chapter 4.2.4) is formed by the particles originating in the secondary vertex of the b-hadronic decay (and other particles from the hadronisation), which is called "b-jet".

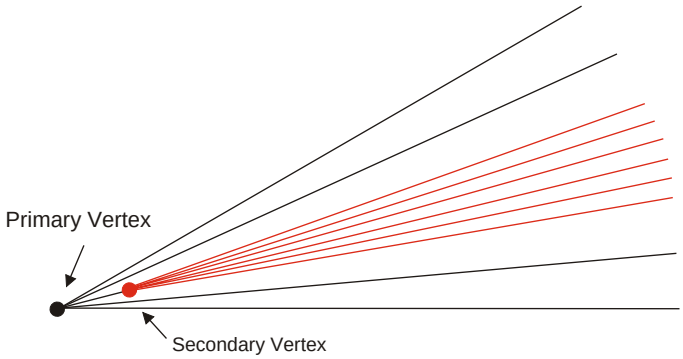


Figure 6.1.: Sketch showing the typical signature of b-jets: A b-hadron is created at the primary vertex and departs. It decays after a distance, which is long enough to be resolvable for the tracker. Thus, new trajectories start from the secondary vertex.

Most of the hadrons posses lifetimes that are either too long or too short for the secondary vertex to be detected: If the lifetime is too long, like e.g. $\tau_{\pi^\pm} \approx 2.6 \cdot 10^{-8}\,\mathrm{s}$ ($\Rightarrow c \cdot \tau_{\pi^\pm} \approx 7.8\,\mathrm{m}$) [1], the hadron will decay far beyond the tracking system[58], which covers a distance of about 3 m in its maximum extension (cf. figure 3.7). If, on the other hand, the lifetime is too short,

[57] i.e. proton collision in case of LHC
[58] supposed there was no calorimeter

like e.g. $\tau_{\Delta^{++}} \approx 5.6 \cdot 10^{-24}\,\mathrm{s}$ ($\Rightarrow c \cdot \tau_{\pi^0} \approx 1.7 \cdot 10^{-15}\,\mathrm{m}$)[59] [1], the distance between primary and secondary vertex is far below the spatial resolution of the tracking system ($\sim 10\,\mu\mathrm{m}$).

With respect to the detector characteristics, the b-hadrons happen to posses a medium lifetime of $\sim 10^{-12}\,\mathrm{s}$ [1], which is caused by the strong suppression of the only possible conversions to c- and u-quarks (cf. "CKM-matrix", chapter 2.1.1). This results in distances between the primary and secondary vertices of $c \cdot \tau_b \sim 500\,\mu\mathrm{m}$, which can be increased to $\sim 5\,\mathrm{mm}$ by the boost of the b-hadrons and which are perfectly measurable with the CMS tracking system. Therefore, jets evolving from b-quarks[60] created in a pp-collision at LHC can be distinguished from jets of lighter quarks and thus identified, i.e. b-tagged.

As b-quarks decay by emitting a W-boson, b-hadrons can decay hadronically (the W-boson converts to quarks) or semi-leptonically (the W-boson converts to leptons). This is something that can also be distinguished in CMS, since the leptonic fraction of the jet will show up in the ECAL, while the hadronic fraction will be detected in the HCAL. Figure 6.2 shows that this actually seems to be working with real collision data.

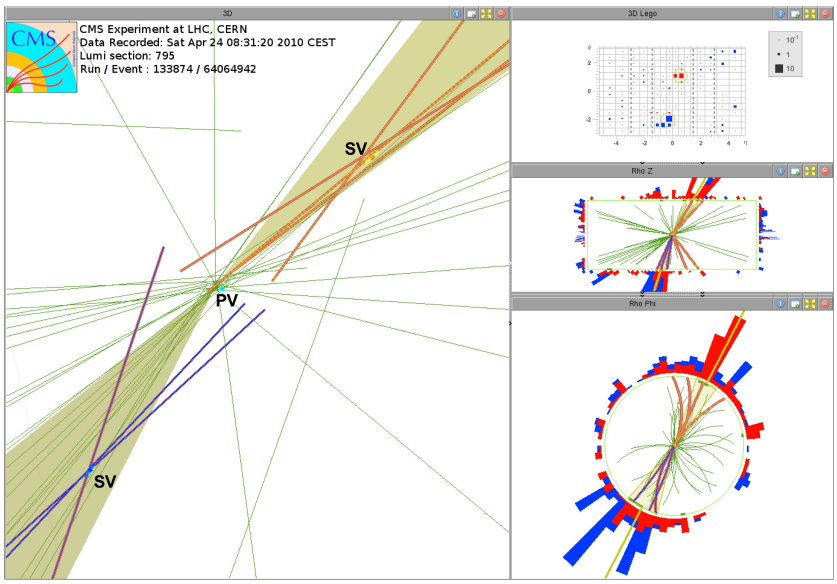


Figure 6.2.: Event display of a pp-collision event with $\sqrt{s} = 7\,\mathrm{TeV}$ at LHC, recorded with the CMS detector [80]: In the left picture, one can clearly recognise the primary vertex (PV) in the centre and two secondary vertices (SV) in opposite directions. The tracks of the particles created in the b-decays are extended in direction of the primary vertex. Due to the presence of two secondary vertices, this event is a double b-jet candidate. On the right side, the event is shown in different displays: the calorimeter entries in the η-ϕ-plane, calorimeter entries and tracks in the y-z-plane and in the x-y-plane (from top to bottom). It is clearly to be seen, that the upper b-jet has a high fraction of electromagnetic calorimeter entries (red), while the lower b-jet possesses a high fraction of hadronic calorimeter entries (blue). Thus, the upper b-jet is a candidate for a semi-leptonic b-decay and the lower one for a hadronic b-decay.

[59] $\tau = \frac{\hbar}{\Gamma} \approx \frac{6.582 \cdot 10^{-22}\,\mathrm{MeV \cdot s}}{118\,\mathrm{MeV}}$

[60] or t-quarks, as they almost exclusively convert to b-quarks (cf. "CKM-matrix", chapter 2.1.1)

A problem when trying to identify b-jets is the fact that also hadrons containing c-quarks possess a mean lifetime leading to a measurable distance between primary and secondary vertex. Therefore, it is harder to separate b-jets from c-jets than from lighter jets.

6.2. Motivation for Implementing b-Tagging into MUSiC

The general motivation of tagging b-jets is similar to the motivation of identifying all the other particles ($e^{\pm}$, γ, $\mu^{\pm}$, $\pi^{\pm}$,...) in a particle physics experiment: Identifying as many particles as possible helps understanding the Standard Model and reducing background events when searching for (new) physics. For example, it is much easier to find the Z peak in data, if the enormous QCD multi-jet background is reduced by demanding two leptons instead of two "particles" as final state. The same is true with b-tagging and b-jet dominated physics. This is why the identification of b-jets plays an important role in top physics (e.g. precision measurement of the top mass) and, of course, in b-physics (e.g. precision measurement of CP violation and CKM matrix).

In addition, what is more important for MUSiC, new physics at high energies like at LHC are expected to provide many possibilities for benefiting from the usage of b-tagging. Many existing theories contain interesting final states with b-jets, which could be found by MUSiC (i.e. they fulfil the requirements for events to be considered in MUSiC, cf. chapter 5.3.3). Some of them have been chosen to serve as benchmark scenarios, providing the possibility of testing the functionality and benefit of b-tagging in MUSiC (cf. chapter 7):

- The heavy quarks of a possible fourth generation would be able to convert to b- and t-quarks. Thus, they provide interesting finale states:

$$t'\,\overline{t'} \longrightarrow b\,\bar{b}\,W^+\,W^- \qquad\qquad b'\,\overline{b'} \longrightarrow t\,\bar{t}\,W^+\,W^- \longrightarrow b\,\bar{b}\,W^+\,W^-\,W^+\,W^-. \quad (6.1)$$

- Supersymmetric scenarios with a high value of $\tan\beta$ are expected to generate numerous b-jets in many different final states (due to long decay cascades). Therefore, they should be predestined for the combination of model unspecific search and b-tagging.

Additionally to the existing theories of new physics containing b-jets, of course there may be something that nobody has thought of yet, but which is realised in nature nevertheless. If this "something" dominantly shows in b-jets, then MUSiC with b-tagging should be much more sensitive in finding it than MUSiC without b-tagging.

For the sake of completeness, one should note at the end of this motivation chapter, that b-tagging is far from being a magic tool finding any (new) physics containing b-jets. If a so-called "golden channel" without b-jets exists, i.e. a final state with almost no or at least very low background, this final state will be more sensitive than final states containing b-jets, as there is always b-jet background (e.g. from the direct b or t production in the collisions). One example are the heavy gauge bosons Z' and W': These are heavy gauge bosons that are supposed to behave like their lighter partners Z^0 and $W^{\pm}$. However, due to their higher

mass, they are able to decay to t-quarks as well. This would possibly lead to interesting final states with respect to the usage of b-tagging:

$$Z' \longrightarrow t\,\bar{t} \longrightarrow b\,\bar{b}\,W^+\,W^- \qquad\qquad W'^{\pm} \longrightarrow t\,\bar{b}\,/\,\bar{t}\,b \longrightarrow b\,\bar{b}\,W^{\pm}. \tag{6.2}$$

Nevertheless, they possess final states with much higher significances of deviations from the Standard Model. Their golden channels are two high energetic leptons and one high energetic lepton + $\not{E}_{\mathrm{T}}$, respectively, which are orientated back-to-back. Therefore, b-tagging is not supposed to be of any advantage in these scenarios.

6.3. Realisation of b-Tagging in CMS

As outlined before, b-tagging is an important part of physics analysis at a high energy particle collider. Thus, several b-tagging algorithms have been developed and implemented in the CMSSW framework [81]. They use different variables, which possess the ability to separate b-jets from light jets (e.g. the impact parameter or the distance between primary and secondary vertex, cf. figure 6.3), and combine them to a discriminator variable each. These discriminator variables can then be used for b-tagging, since on average b-jets possess higher values than light jets. The basic functionality of the available algorithms is described below, especially considering the variables they use, their efficiency [82] and their robustness with respect to detector misalignment[61]. More specific information, e.g. on the specific methods used for track and vertex reconstruction (and particle identification in case of the lepton based algorithms) or the specific quality criteria for jet associated tracks, can be found in the cited sources.

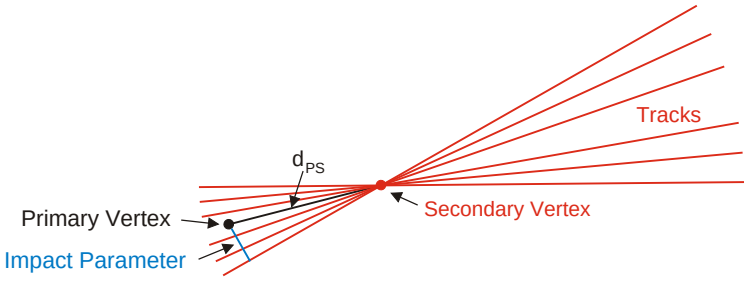


Figure 6.3.: Sketch illustrating the definition of the impact parameter, which is an important parameter for b-tagging algorithms: If a primary vertex has been reconstructed, the impact parameter of a track is the distance of its closest approach to the primary vertex. Obviously, the impact parameter of tracks coming from a secondary vertex tends to be bigger than of tracks having their origin in the primary vertex.

- **Lepton based algorithms:**
  In CMSSW, two lepton based algorithms [84] exist (**"Soft Electron"** and **"Soft Muon"**, respectively), which exploit the properties of soft leptons created in semi-

[61] The b-tagging performance depends on the track and vertex reconstruction and thus on the precise knowledge of the positions and orientations of the tracker modules, i.e. its alignment [83]

leptonic decays of b-jets. The following variables are combined to a discriminator variable, using a neural network:

- transverse momentum of the lepton relative to the jet axis
- distance ΔR between the lepton and the jet axis
- ratio of the lepton momentum to the calorimetric jet energy
- impact parameter significance (i.e. impact parameter divided by its uncertainty) of the lepton track
- calibrated calorimetric jet energy (for muons only)
- calorimetric jet pseudo-rapidity (for muons only)

The last two variables are additionally used with muons, in order to consider differences in detector response, depending on the jet energy and calorimeter region. As almost exclusively properties of the reconstructed soft lepton contribute to the discriminator variable, the soft lepton algorithms are very robust and do not suffer much from performance decrease due to misalignment. On the other hand, only $\sim 30\,\%$ of the b-hadrons decay semi-leptonically, thus the b-tagging efficiency of the soft lepton algorithms is heavily limited.

- **Impact parameter based algorithms:**
  Two different impact parameter based algorithms [85] are implemented in CMSSW: the **"Track Counting"** and the **"Jet Probability"** algorithm. The Track Counting algorithm is the more basic one. It sorts the tracks by descending impact parameter significance and uses the impact parameter significance of the n^{th} track as discriminator variable. n can be 2 or 3, dependent on choosing the "high efficiency" or the "high purity" algorithm version.

  The Jet Probability algorithm is more sophisticated. Based on the impact parameter, the probability for each track to come from the primary vertex is computed. The single probabilities of the tracks are then combined to a probability for the whole set of tracks having its origin in the primary vertex. If this probability is low, the corresponding jet is likely to be a b-jet. In comparison with the Track Counting algorithm, this method has the advantage that all tracks are used to determine the discriminator variable. An available variant of this algorithm, the "Jet B Probability" algorithm, gives more weight to the four most b-like tracks when combining the probabilities.

  The two impact parameter based algorithms achieve higher b-tagging efficiency than the lepton based algorithms as no b-jets have to be neglected by default. In return, a higher performance decrease due to misalignment has to be accepted, since the algorithms do not rely on well reconstructed leptons.

- **Secondary vertex based algorithm:**
  Instead of using the impact parameter of tracks, the **"Simple Secondary Vertex"** algorithm [86] tries to reconstruct secondary vertices from the tracks. If a secondary vertex has been found, the significance of the distance between primary and secondary vertex is directly used as the discriminator variable.

  As the Simple Secondary Vertex algorithm is dependent on reconstructing a secondary vertex (otherwise no discriminator variable can be determined), its b-tagging efficiency is restricted to the probability of finding an existing secondary vertex of a b-decay ($\sim$ 60 - 70 %). But this is very well acceptable for early data, since this algorithm has the lowest performance decrease due to misalignment of all b-tagging algorithms (except the lepton based algorithms).

- **Multi-variable based algorithm:**
  The most sophisticated approach of a b-tagging algorithm is the **"Combined Secondary Vertex"** algorithm [87]. It uses all known variables of the tracks and the secondary vertices possessing discriminating power between b-jets and light jets, and combines them to one discriminator variable. The choice of parameters used for the discriminator variable depends on the vertex category of the jet:

  1. Reco-Vertex: At least one reconstructed secondary vertex is found. If there is more than one secondary vertex, the tracks of all reconstructed secondary vertices are used for computing the discriminator variable.

  2. Pseudo-Vertex: If no reconstructed secondary vertex can be found but there are at least two tracks not compatible with the primary vertex (impact parameter significance > 2), a so-called pseudo-vertex is created from these tracks.

  3. No-Vertex: If neither 1. nor 2. is fulfilled.

  One variable is used for any vertex category (thus, every jet receives a discriminator variable):

  – impact parameter significances of the tracks (like used in both of the impact parameter based algorithms)

  In case if "No-Vertex", no additional variables are used, as they all require the existence of a secondary vertex. For jets of the "Reco-Vertex" or "Pseudo-Vertex" category, the following variables are added. Since the geometrical position of a pseudo-vertex does not have any physical meaning, the distance significance of secondary vertices is employed for jets of the "Reco-Vertex" category only.

  – distance significance (i.e. distance to the primary vertex, measured in the transverse plane, divided by its uncertainty) of the secondary vertex (for "Reco-Vertex" only)

  – invariant mass of charged particles (represented by tracks) associated to the secondary vertex (can be significantly higher for b-jets than e.g. for c-jets)

- multiplicity of tracks associated to the secondary vertex (is significantly higher for b-decays)

- energy fraction of tracks associated to the secondary vertex and all tracks of the jet

- rapidity $y = \frac{1}{2} \ln \left(\frac{E+p_{||}}{E-p_{||}} \right)$ of tracks associated to the secondary vertex with respect to the jet direction (smaller for b-jets than for c-jets, as b-jets (and thus their decay products) are less boosted, due to the higher b-mass)

- lowest impact parameter significances of the n tracks with the highest impact parameter significances, exceeding a charm mass threshold for the first time, starting from $n = 1$ (With 1.5 GeV, the charm mass threshold is lower than typical masses of c-hadrons, because neutral particles are missed and not all charged particles are reconstructed.)

These variables are combined via a Likelihood Ratio or, as a variant of the Combined Secondary Vertex algorithm, using a neural network in the "Reco-Vertex" case, which leads to a better b-tagging efficiency.

Due to the combination of all discriminating variables, that can be obtained from tracks and secondary vertices, the Combined Secondary Vertex algorithm achieves the best efficiencies of all b-tagging algorithms. However, with respect to misalignment, it is also the most sensitive one. Thus a well understood detector is needed, to be able to benefit from its advantages.

6.4. Integration of b-Tagging into MUSiC

Additionally to the standard physics objects (e, μ, γ, jet, $\not{E}_{\mathrm{T}}$), b-jets have been implemented into the MUSiC framework as an optional physics object. Now, when creating the event classes, one can decide whether or not to use them. In case of regarding b-jets as physics object, b-tagged jets are called "b-jet" (or simply "b") within MUSiC. More important is the fact that not b-tagged jets are referred to as "non-b-jet" or "jet". This means, the 1μ 3jet 2b event class contains a total of five jets (2 b-jets and 3 non-b-jet) rather than three jets.[62] Basically, b-tagging had to be implemented into several parts of MUSiC:

- For each jet, the resulting discriminator value of the available b-tagging algorithms has to be saved when skimming the CMS sample files.

- The event class factory had to be modified, so that it is able to use the discriminator values of the jets, in order to tag b-jets.

[62] With the current implementation of b-tagging in MUSiC, NO event class generally containing five jets (without the distinction of b- and non-b-jets) is created, like it is filled when using MUSiC without b-tagging. This could be changed in the future.

- Then, event classes containing b-jets have to be created as well as b-tagging specific control plots.

- Last but not least, the errors of the b-tag efficiency ϵ_b and of the mistag rates (which will be explained below) have to be considered within the MUSiC search algorithm, as they influence the uncertainties of the Standard Model prediction.

6.4.1. b-Discriminator

When intending to perform an analysis applying b-tagging, one important thing is to decide which algorithm/discriminator to use. This decision heavily depends on the goal of your analysis and the properties of the available discriminators (cf. chapter 6.3). If one e.g. wants to tag b-jets in the early data of CMS, one should choose a discriminator which

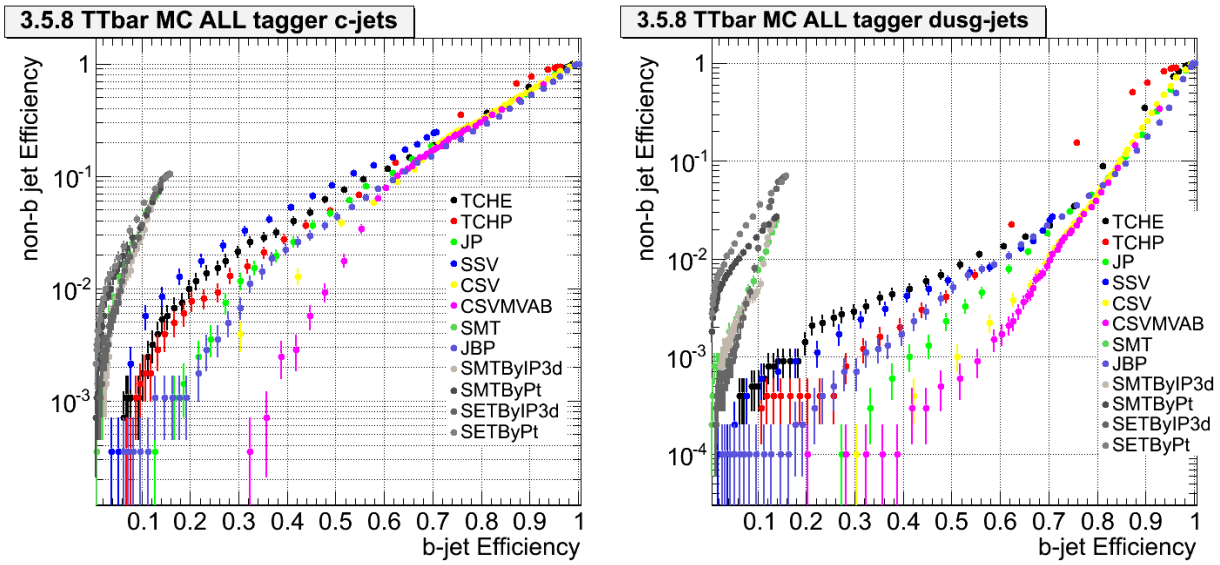


Figure 6.4.: Validation plots of the CMS b-tag group [82]. They show the mistag rate (non-b jet efficiency) of c-jets (left) and dusg[63]-jets (right), plotted versus the b-tagging efficiency. The lower mistag rates are achieved for a given b-tagging efficiency, the better the discrimination is. One can see, that CSVMVAB achieves (one of) the lowest mistag rates throughout the whole range of b-tagging efficiencies. The plots have been produced using CMSSW 3.5.8 and a $t\bar{t}$ sample. The labelling of the b-tagging algorithms is as follows: TCHE = Track Counting (high efficiency); TCHP = Track Counting (high purity); JP = Jet Probability; SSV = Simple Secondary Vertex; CSV = Combined Secondary Vertex; CSVMVAB = Combined Secondary Vertex (neural network variant); SMT = Soft Muon; JBP = Jet Probability (weighting variant); SMTByIP3d = Soft Muon ("only impact parameter" variant); SMTByPt = Soft Muon ("only transverse momentum" variant); SETByIP3d = Soft Electron ("only impact parameter" variant); SETByPt = Soft Electron ("only transverse momentum" variant).

[63] meaning jets from d-, u-, s-quarks or gluons

suffers as little as possible from performance decrease due to detector misalignment (as this is expected to be problematic in the early phase of data taking). Since the aim of MUSiC is to search for new physics within data of a well understood detector, i.e. not with early data, misalignment should not be a critical point. Thus, the choice has been made by determining the discriminator, which is expected to provide the best differentiation between b-jets and light jets, i.e. the best ratio of b-tag and mistag efficiency for a given value of ϵ_b. This is shown for the available discriminators in figure 6.4. It is obvious, that the neural network variant of the Combined Secondary Vertex algorithm ("CSVMVAB") reveals the best performance (or at least one of the best performances) over the full range of ϵ_b, especially for discriminating b-jets from c-jets for $\epsilon_b < 0.6$. Therefore, this discriminator has been chosen as default for MUSiC and has been used for the present thesis. The other algorithms have been implemented into the MUSiC framework anyway and can be chosen when creating event classes.

Of course, the expected performance of the algorithms is based on Monte Carlo studies (as the true flavour of a jet is unknown in data), but first studies on the distributions of b-tagging variables in collision data propose a good agreement with the Monte Carlo predictions [88]. Additionally, there are recent approaches aiming at a data-driven determination of the tagging efficiencies [89, 90].

6.4.2. Performance Plots of the b-Discriminator

Before the newly implemented b-tagging discriminator can be exploited to search for new physics, one should of course check the behaviour of variables related to its performance. If e.g. the discriminator distribution generated with MUSiC looks totally different than the expected one (as presented by the CMS b-tag group [82]), there might be a mistake in the implementation of the discriminator into MUSiC. Additionally, some performance plots (e.g. efficiency and purity distributions) are needed in order to define the discriminator value on which to cut. Therefore, several control plots are produced with MUSiC, using the default b-tagging algorithm (Combined Secondary Vertex, neural network variant). In order to gain higher statistics, the demand for at least one electron, muon or photon has not been used when producing the control plots. This does not effect the particles entering the event classes, as the selection criteria are not affected. But as more events are kept (those not containing leptons or photons) more jets can be examined when creating the control plots.

The following control plots have been generated using the Monte Carlo $t\bar{t}$ sample, assuming $\sqrt{s} = 7\,\text{TeV}$. As t-quarks almost exclusively convert into b-quarks (cf. chapter 2.1.1), the $t\bar{t}$ sample perfectly suits for the validation of b-tagging. Therefore, it is also used by the CMS b-tag group, what makes their validation plots comparable to some of the MUSiC control plots.

The **discriminator distribution** (cf. figure 6.5) shows the normalised number of jets versus the discriminator value they possess. The statistic errors on the bin entries are Poisson errors, calculated via $\sigma_N = \sqrt{N}$. As expected, the b-jets have higher discriminator values than the light jets, thus the discriminator is able to separate b-jets from other jets.

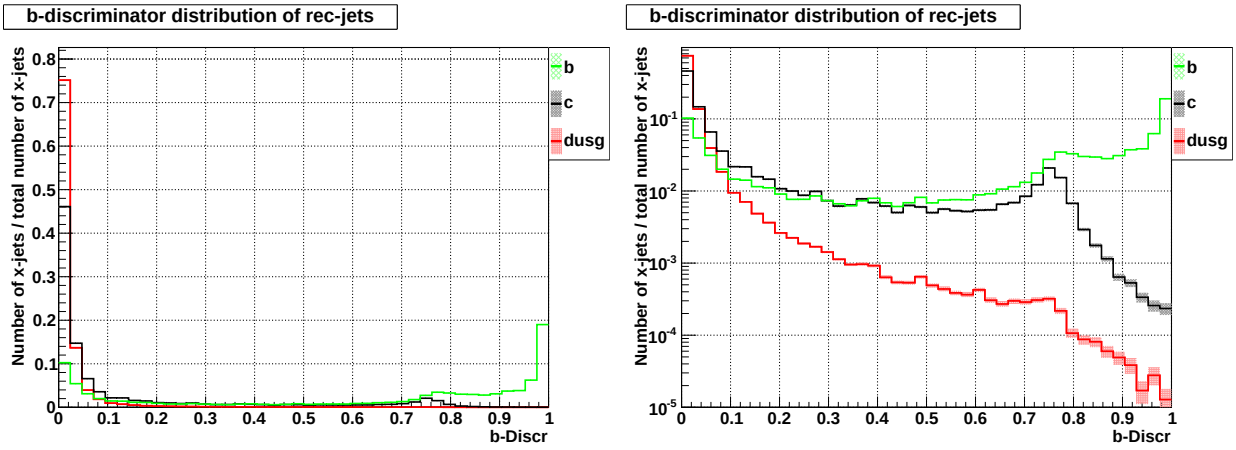


Figure 6.5.: Discriminator distribution of the b-tagging algorithm. The plots were created with MUSiC using the Monte Carlo $t\bar{t}$ sample ($\sqrt{s} = 7\,\text{TeV}$). The right plot shows the same distribution as the left plot, but with a logarithmic scale. The plotted errors are Poisson errors.

Figure 6.6 presents the *tagging efficiency distribution* and *distribution of "mistag rate" vs. b-tagging efficiency*. The tagging efficiency

$$\epsilon_q = \frac{N_q^{\text{b-tagged}}}{N_q^{\text{all}}} \qquad \text{with } q = \text{jet flavour} \tag{6.3}$$

is the fraction of jets, which are tagged as b-jets for a given discriminator cut. As expected, the b-tagging efficiencies are much higher than the mistag rates. Plotting the mistag rates versus the b-tagging efficiency, it turns out, that good b-tagging efficiencies can be achieved together with satisfactory low mistag rates (e.g. $\epsilon_c < 1\,\%$ for ϵ_b up to $47.5\,\%$).

For the estimation of the statistical errors, Clopper-Pearson confidence intervals [91] with $68.3\,\%$ confidence level have been chosen. In contrast to binomial errors $\sigma_\epsilon^{\text{bin}} = \sqrt{\frac{\epsilon(1-\epsilon)}{N^{\text{all}}}}$, which are often chosen to calculate the errors of efficiencies or efficiency-like parameters, they determine asymmetric errors, which are non-vanishing in case of $\epsilon = 0\,\%$ or $\epsilon = 100\,\%$. Additionally, the Clopper-Pearson intervals never suffer from undercoverage, but rather experience overcoverage [92]. This means, given a measured efficiency and a corresponding Clopper-Pearson interval with $68.3\,\%$ confidence level, one can be sure that this measured efficiency is not compatible with true efficiencies beyond the Clopper-Pearson interval with at least $68.3\,\%$ confidence level. Thus, using Clopper-Pearson intervals, errors are never underestimated but rather overestimated. This conservative error handling is important, as underestimated uncertainties of the Standard Model predictions would lead to overestimation of significances of the deviations in data.

In addition to the expected behaviour of the b-discriminator, a comparison between the plots shown above and the validation plots of the CMS b-tag group [82] shows a good agreement. Thus, it can be assumed, that b-tagging in MUSiC is working in a correct way.

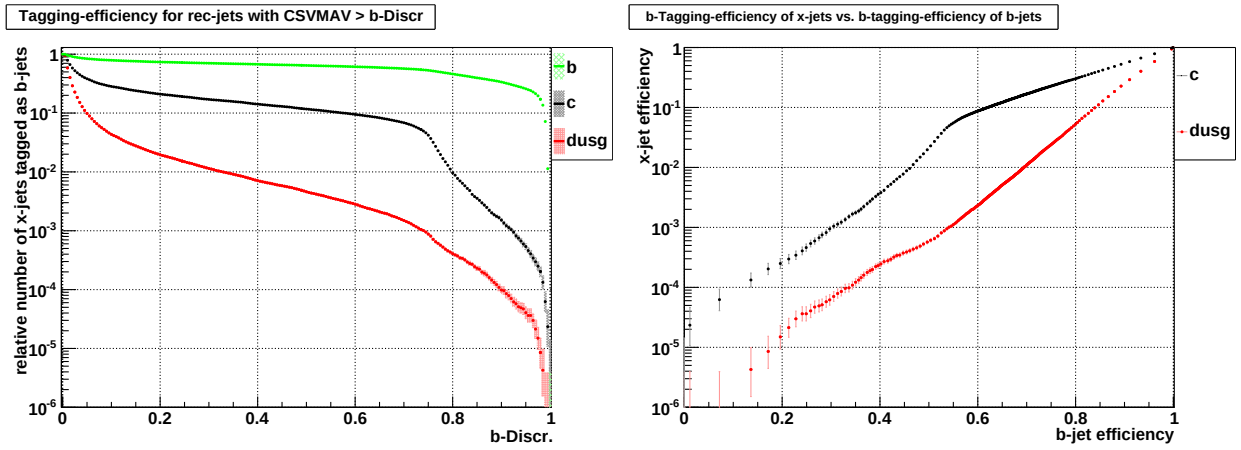


Figure 6.6.: Tagging efficiency distribution of the b-tagging algorithm (left) and distribution of mistag rate vs. b-tagging efficiency (right). The plots were created with MUSiC using the Monte Carlo $t\bar{t}$ sample ($\sqrt{s} = 7\,\text{TeV}$). The plotted errors are Clopper-Pearson confidence intervals with 68.3 % confidence level.

6.4.3. A Data-like Background Sample

In the last section, the $t\bar{t}$ sample has been used, in order to validate the functionality of the b-discriminator implemented into the MUSiC framework. The reason for choosing the $t\bar{t}$ sample for that purpose (the high fraction of b-jets contained) becomes a problem when intending to choose a b-discriminator cut value: As b-jets are dominating the sample from the beginning on, it is clear, that the b-tagged jets will possess a high *purity* (i.e. consist of b-jets to a large extent), regardless how high the mistag rates are. If, in contrast, a QCD multi-jet sample is used, which is heavily dominated by light jets, already a small mistag rate might result in a low purity of the b-tagged jets. Therefore, a "data-like" mixture of Monte Carlo samples is necessary, in order to determine the performance of the b-discriminator and to choose a b-discriminator cut value. This mixture has been modelled by combining Monte Carlo samples ($\sqrt{s} = 7\,\text{TeV}$) of the following processes, scaled due to their specific cross sections: QCD multi-jets, t-quark production (SingleTop, $t\bar{t}$), single boson + jets (WJets, ZJets, PhotonJets) and diboson (WW, WZ, ZZ, ZGamma). A complete list of the used samples and their cross sections can be found in the appendix (table A.4)

Using the sample mixture, the same control plots have been produced, which have been shown in section 6.4.2 for $t\bar{t}$ only. These plots can be found in figures 6.9 and 6.10 and have been used for choosing a cut value of the b-discriminator. This is described in section 6.4.4. Moreover, the higher statistics due to the combination of the Monte Carlo samples allows the production of additional control plots, namely the ϕ-, η- and p_{T}-dependency of the tagging efficiencies. They are shown in figure 6.8.

In order to produce the control plots of the sample mixture, the corresponding plots of all samples contained in the mixture have been scaled to the cross section and then combined. Due to the scaling, samples with high cross sections gain a large influence on the combined plots of the whole mixture. Thus, low statistics of such samples may result in large fluctuations and uncertainties in the mixture plots. The "QCD_Pt30" sample is a good example:

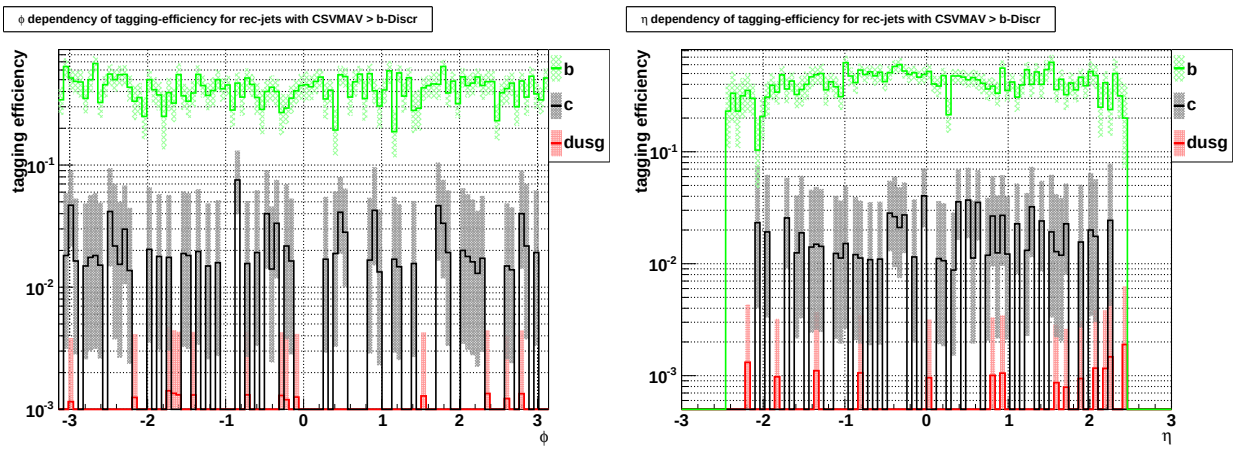


Figure 6.7.: ϕ- (left) and η-dependency (right) of the tagging efficiencies of the "QCD_Pt30" sample (assuming $\sqrt{s} = 7\,\text{TeV}$).

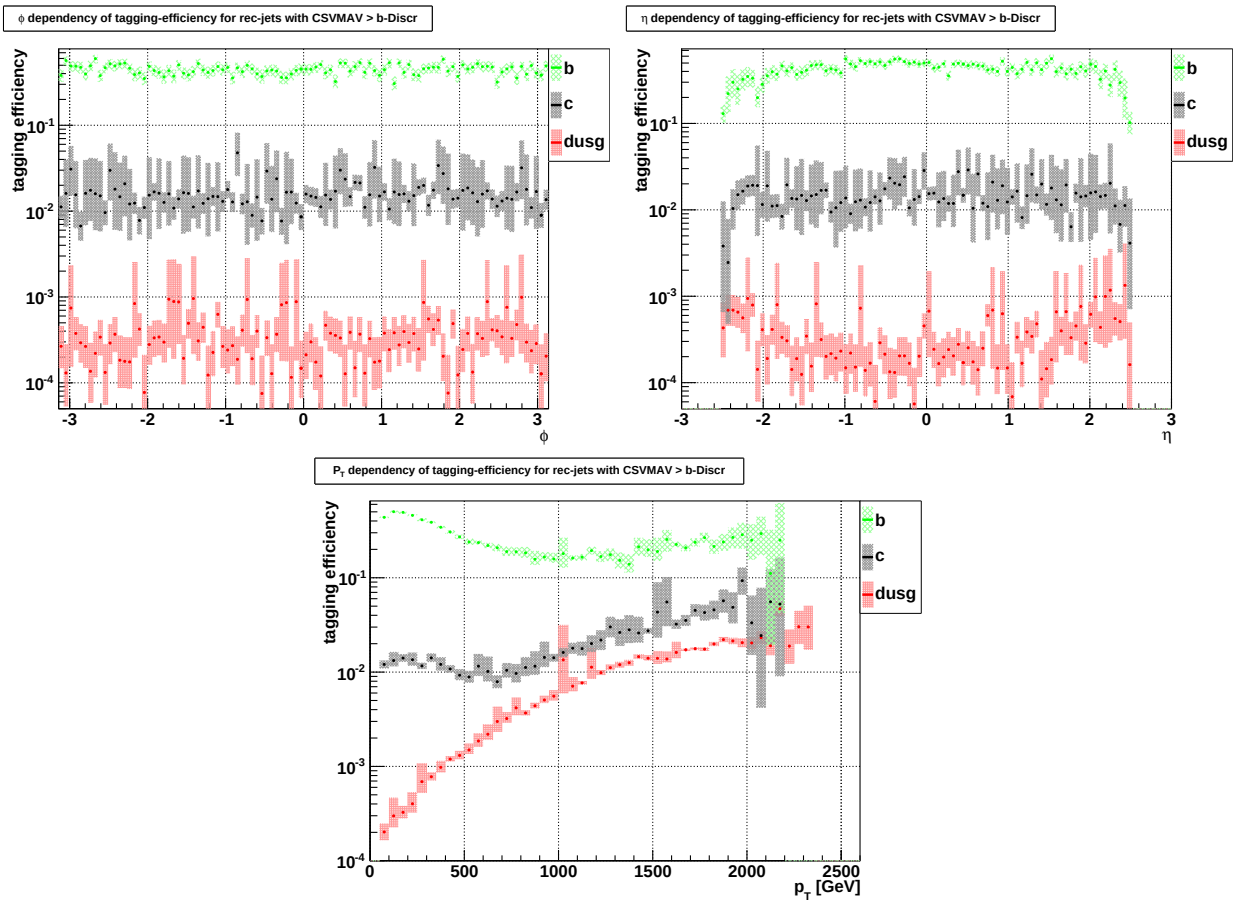


Figure 6.8.: ϕ- (top left), η- (top right) and p_T-dependency (bottom) of the tagging efficiencies of the sample mixture (assuming $\sqrt{s} = 7\,\text{TeV}$). The large fluctuations and uncertainties are mainly caused by the low statistics of the"QCD_Pt30" sample shown above.

It possesses the largest cross section of all used samples and suffers from low statistics (due to the high p_T selection cuts, cf. chapter 5.3.3). As a result, its bins with low statistics clearly coincide with the strongly fluctuating bins of the corresponding mixture plots, which also show the largest uncertainties (cf. figures 6.7 and 6.8).

The efficiency of the sample mixture does not show a ϕ-dependency, exactly as one would expect, since the CMS detector is symmetrically built in ϕ-direction. For high values of η, the tagging efficiencies become worse, due to the worse resolution in this area. As the particles tend to be more boosted for higher p_T values and thus the reconstruction of a secondary vertex as well as the determination of the impact parameter becomes more difficult, the mistag efficiencies extremely increase for higher p_T values.

Since the Monte Carlo samples contained in the data-like mixture are scaled to their cross sections, the uncertainties of the individual samples will have variable influence on the uncertainty of the combination. In order to consider this, the uncertainties of the combined plots have been estimated by adding (subtracting) the uncertainty to (from) the efficiencies of the separate samples before combining them. Thus, an upper (lower) limit on the uncertainty range could be estimated.

6.4.4. Efficiency and Purity based Choice of the Discriminator Cut

To process b-tagging with a b-discriminator, a cut value has to be defined. All jets possessing a discriminator value greater than this cut value are considered as b-jets, whereas all other jets are called non-b-jets. The cut value has to be chosen in a way, that a compromise between a high purity and a reasonable efficiency is found. Therefore, the discriminator distribution, efficiency distribution and purity distribution of the sample mixture are produced. They are shown in figures 6.9, 6.10 and 6.11. The shown uncertainties have been estimated as described in section 6.4.3. The distributions show a strong agreement with the plots of the $t\bar{t}$ sample (cf. section 6.4.2). Discriminator values of above 0.75 are dominantly populated by

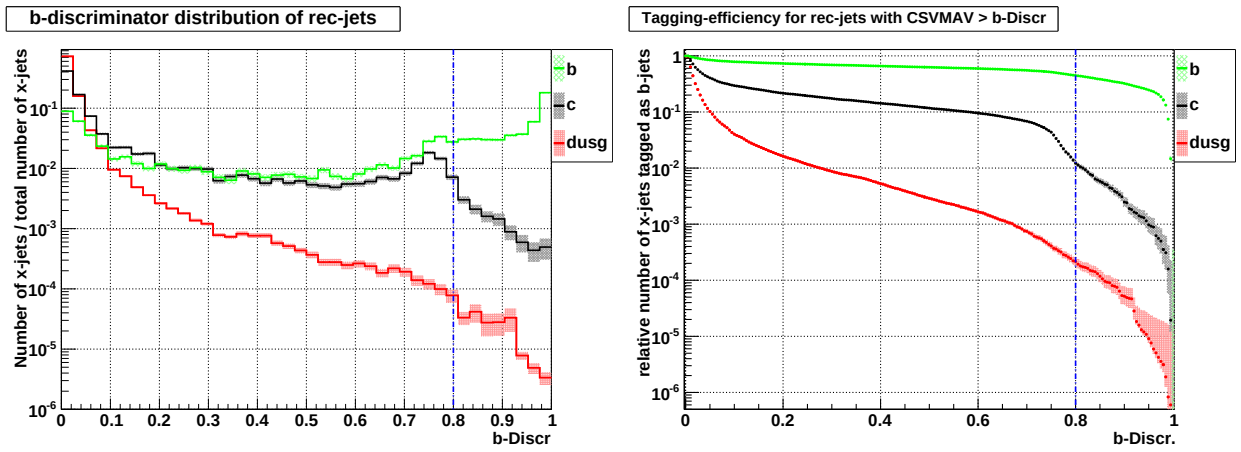


Figure 6.9.: Discriminator distribution (left) and tagging efficiency distribution (right) of the data-like sample mixture ($\sqrt{s} = 7\,\text{TeV}$). The blue lines indicate the cut value that has been chosen.

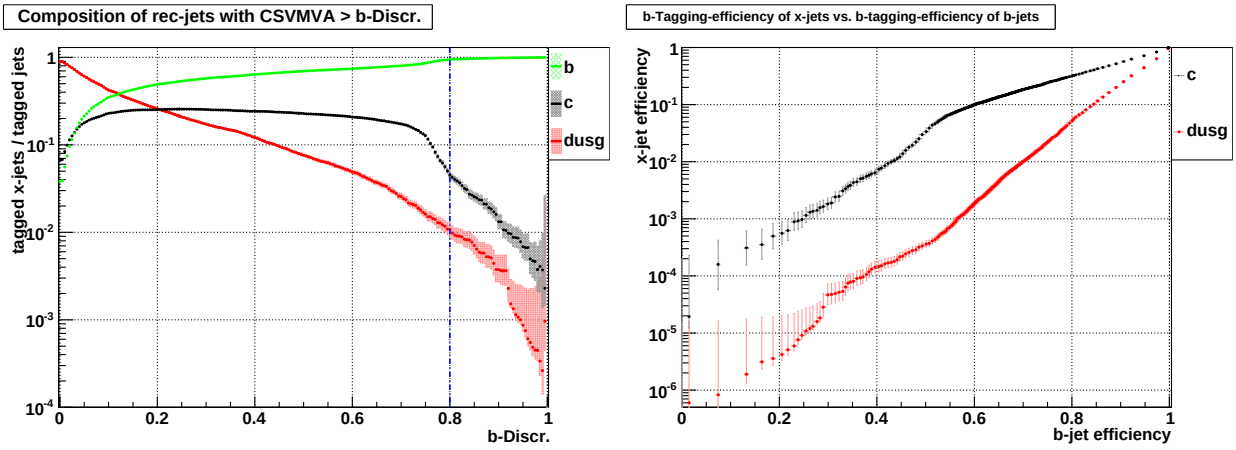


Figure 6.10.: Purity distribution (left) and distribution of mistag rate vs. b-tagging efficiency (right) of the data-like sample mixture ($\sqrt{s} = 7\,\text{TeV}$). The blue lines indicate the cut value that has been chosen.

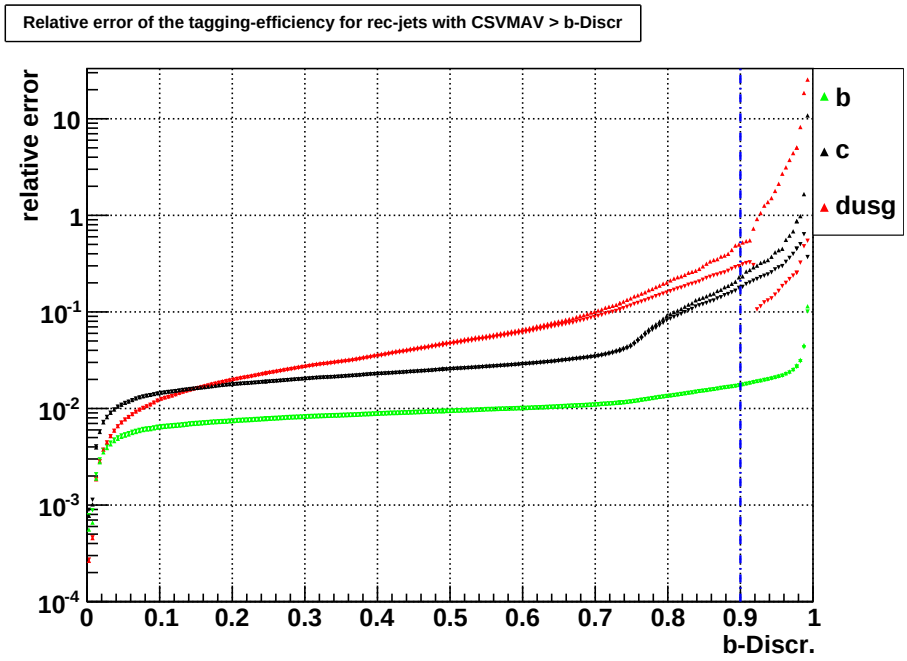


Figure 6.11.: Distribution of relative efficiency errors of the data-like sample mixture ($\sqrt{s} = 7\,\text{TeV}$). Triangles pointing upwards stand for the upper relative efficiency errors, while triangles pointing downwards represent the lower relative efficiency errors. The blue line indicates the discriminator value which has been used to estimate the maximum relative efficiency errors.

b-jets. For cutting values above 0.75 the mistag rates and the fraction of light jets tagged as b-jets rapidly decreases. As also the b-tagging efficiency decreases and the efficiency as well as the purity distribution contain steps (caused by low statistics) above discriminator values of about 0.9, a discriminator cut value of 0.8 has been chosen. Applying this cut value, the following performance parameters are achieved for the data-like sample mixture, assuming $\sqrt{s} = 7\,\text{TeV}$:

	b-jets	c-jets	dusg-jets
tagging efficiency	$44.7\,\%^{+0.7\,\%}_{-0.7\,\%}$	$1.22\,\%^{+0.12\,\%}_{-0.10\,\%}$	$0.022\,\%^{+0.005\,\%}_{-0.004\,\%}$
$^{\text{b-tagged x-jets}}/_{\text{all x-jets}}$	$94.4\,\%^{+0.4\,\%}_{-0.5\,\%}$	$4.5\,\%^{+0.5\,\%}_{-0.4\,\%}$	$1.1\,\%^{+0.3\,\%}_{-0.2\,\%}$

To determine the significances of deviations found with MUSiC, the uncertainties of the Standard Model prediction have to be considered (cf. chapter 5.3.4). One part of the uncertainties has its origin in the uncertainties of the tagging efficiencies. As the discriminator cut can be chosen freely and only values greater than 0.9 should be avoided (due to lack of statistics), the uncertainties of the b-tagging efficiency and the mistag rates are conservatively estimated by determining their maximum value on the discriminator interval of 0 - 0.9. Thus, the relative uncertainties are assumed with 2 % for the b-tagging efficiency, 30 % for the c-jet mistag rate and 60 % for the dusg-jet mistag rate (cf. figure 6.11).

6.4.5. Class Dependency of the Purity

When considering the purity of the b-tagged jets, the big strength of MUSiC, searching in many event classes instead of a few final states, leads to a problem: The event classes will be filled unbalanced by specific processes. For example QCD multi-jets will dominantly be present in classes with only one lepton/photon and a high multiplicity of jets. On the other hand, diboson events will enter event classes with two or more leptons and low multiplicity of jets. In the extreme case, single event classes may be filled exclusively by QCD multi-jet events while other event classes are only entered by TTbar events. This would obviously result in very different values of the purity of b-tagged jets in those classes. Therefore, the purity's event class dependency has to be investigated, using the data-like sample mixture.

In table 6.1 the results for some of the $\sim$ 1800 event classes are summarised. Therefore, typically behaving event classes have been chosen. One can see, that the b-jet purity is very

event class	purity
1e 1b 1met	$96.9\,\%^{+2.0\,\%}_{-2.7\,\%}$
1e 1jet 1b 1met	$98.6\,\%^{+0.9\,\%}_{-2.7\,\%}$
1e 1mu 2b 1met	$99.7\,\%^{+0.3\,\%}_{-5.8\,\%}$
1gam 5jet 1b 1met	$99.995\,\%^{+0.005\,\%}_{-83.177\,\%}$

Table 6.1.: Purities and their absolute errors of 4 typical of about 1800 event classes produced with the data-like Sample mixture with a discriminator cut at 0.8. As the event classes of the different samples contained in the data-like mixture are combined, the uncertainties are estimated in the same way like when combining their plots: An upper (lower) limit on the uncertainty range is estimated by adding (subtracting) the uncertainty to (from) the purities of the separate samples before combining them. Thus, very asymmetric and large error ranges may be the result of event classes with low statistics from samples possessing a large cross section (cf. section 6.4.3).

high throughout all event classes. Though, "ordinary" event classes (like 1e 1b $\not\!\!E_{\mathrm{T}}$) possess high statistics, leading to low purity uncertainties, while more "exotic" event classes (like 1gam 5jet 1b $\not\!\!E_{\mathrm{T}}$) suffer from low statistics and thus very high purity uncertainties. Varying the discriminator cut by ± 0.05 reveals no discrepancy of this general conclusion. Merely the absolute values change dependent on the discriminator cut, but the "ordinary" event classes always possess much smaller uncertainties than the "exotic" ones.

7. Test Scenarios with b-Jets

Tests have been processed to determine the benefit of using MUSiC with b-tagging when searching for new physics. Of course, in order to find out whether MUSiC would be able to find new physics, specific models have to be used. Therefore, Monte Carlo samples of different benchmark scenarios have been mixed to the Standard Model samples. The results are described in this chapter.

Several processes have been used as signal, assuming a centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$. Each of these signal processes has been searched for with MUSiC, using four different configurations:

- With b-tagging and the discriminator cut of 0.8, which has been chosen in chapter 6.4.4.
- Without b-tagging, in order to compare the significances of the deviations, which have been found with and without b-tagging, and thus determine, if b-tagging gives an advantage.
- With b-tagging and a discriminator cut of 0.75 and 0.85, respectively, to estimate the influence of the chosen cut value on the significances.

7.1. Top Pair Production

The first process that has been searched for does not belong to the category of new physics. It is the Standard Model $t\bar{t}$ production, which is the main source of background when searching for b-jets. Therefore, it is a reasonable choice to check, whether b-tagging is of advantage when searching for $t\bar{t}$ with MUSiC. If already in this case no advantage of using b-tagging could be found, searches for new physics with b-jets would be rather hopeless. Additionally, the rediscovery of the t-quark is an important step in verifying the Standard Model at the LHC energy scale.

7.1.1. Significant Event Classes

In order to search for $t\bar{t}$, all Monte Carlo samples of the Standard Model mixture but the $t\bar{t}$ sample have been used as background samples, while the $t\bar{t}$ sample has been used as signal. Assuming a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ and an integrated luminosity of $100\,\text{pb}^{-1}$,

those classes have been investigated, which are supposed to be the most significant ones. As explained in chapter 2.1.1, the t-quark nearly instantaneously decays into a b-quark and a W-boson almost exclusively. Then the b-quark hadronises and forms a b-jet, while the W decays leptonically ($\approx 30\,\%$) or hadronically ($\approx 70\,\%$), creating one charged lepton and a neutrino ($\not{E}_{\mathrm{T}}$) or two jets. Thus, the event classes $2e\ 2b\ \not{E}_{\mathrm{T}}$, $1e\ 1\mu\ 2b\ \not{E}_{\mathrm{T}}$ and $2\mu\ 2b\ \not{E}_{\mathrm{T}}$ should provide the most significant deviations. Unfortunately, these event classes suffer from low statistics due to the fact that both b-jets of a $t\bar{t}$ event need to be identified. With a b-tagging efficiency of $47.7 \pm 0.7\,\%$, not even every fourth event belonging to these event classes is contained in them. Therefore, no significant deviations have been found in these event classes (assuming an integrated luminosity of $100\,\mathrm{pb}^{-1}$).

Thus, a rough[64] global scan of all event classes has been made, in order to determine the event classes with the most significant deviations. The global scan showed large significances in many event classes. The most significances were found in the $1e$ 1jet $1b\ \not{E}_{\mathrm{T}} + \mathrm{X}$ and the $1\mu\ 2b + \mathrm{X}$ event classes, which can be found in figures 7.1 and 7.2. This can be explained by

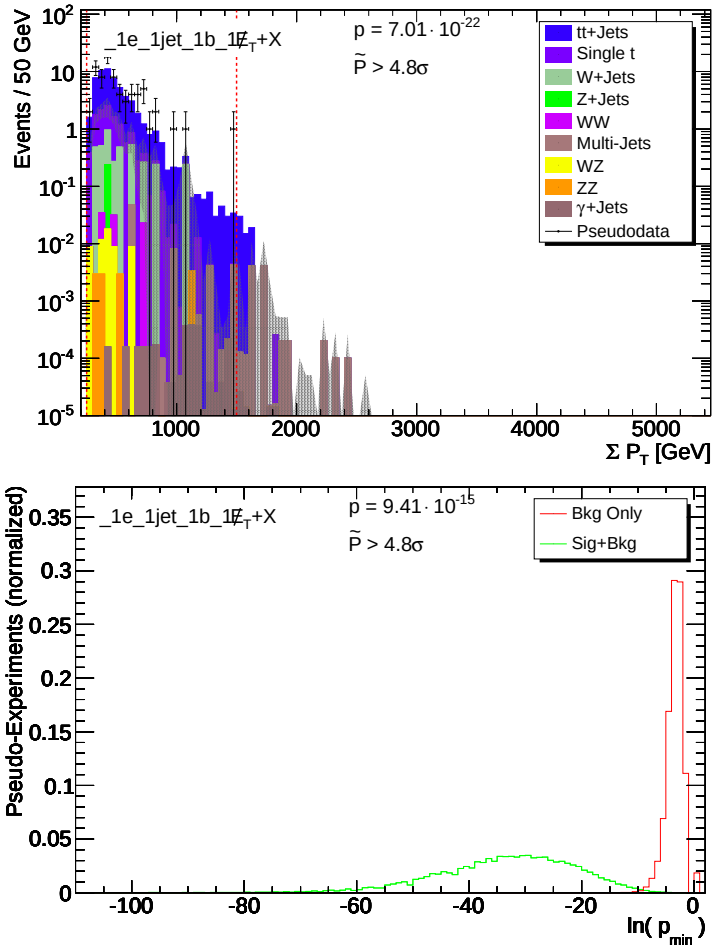


Figure 7.1.: Kinematic distribution (top) and p-value distribution of the diced pseudo experiments (bottom) of one of the most significant event classes when using b-tagging to search for $t\bar{t}$. A significant deviation as well as a good separation of the diced pseudo experiments for background + signal (the complete Standard Model) and background only (the Standard Model without $t\bar{t}$) can be found.

[64] Dicing the background (background + signal) only 5000 (1000) times, which is far too little to give an exact estimation of the significances, but considerably reduces the time needed for the analysis.

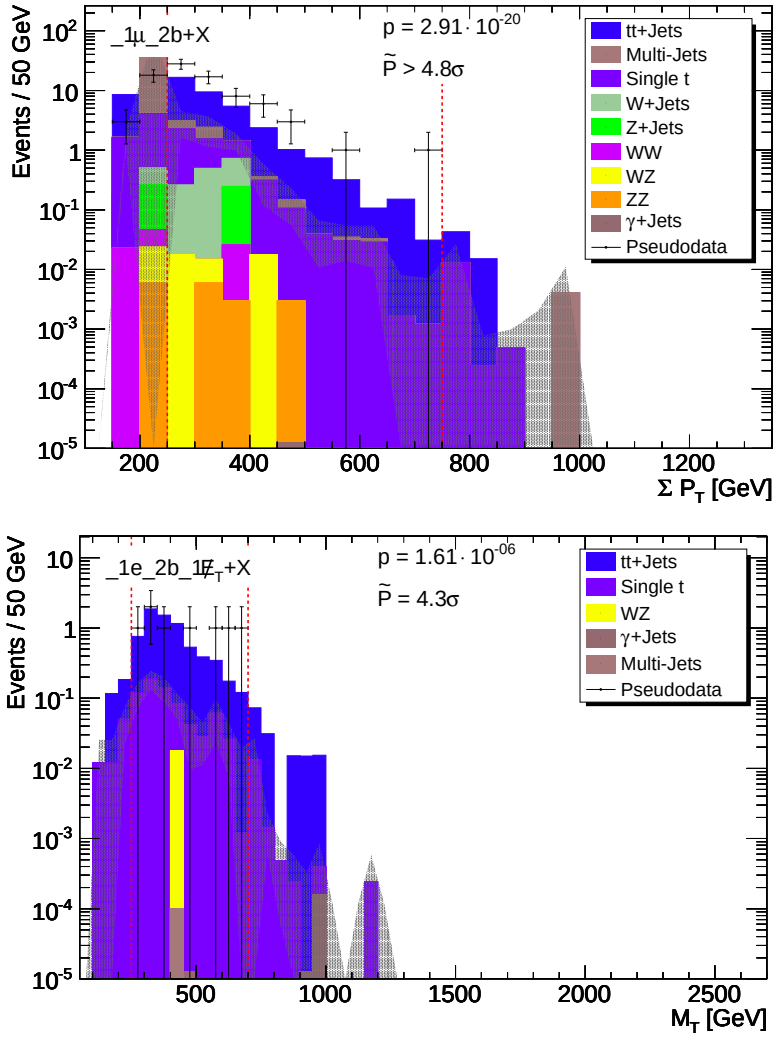


Figure 7.2.: Kinematic distribution of two more event classes when using b-tagging to search for $t\bar{t}$. One can see that the more specific event class (bottom) heavily suffers from lower statistics.

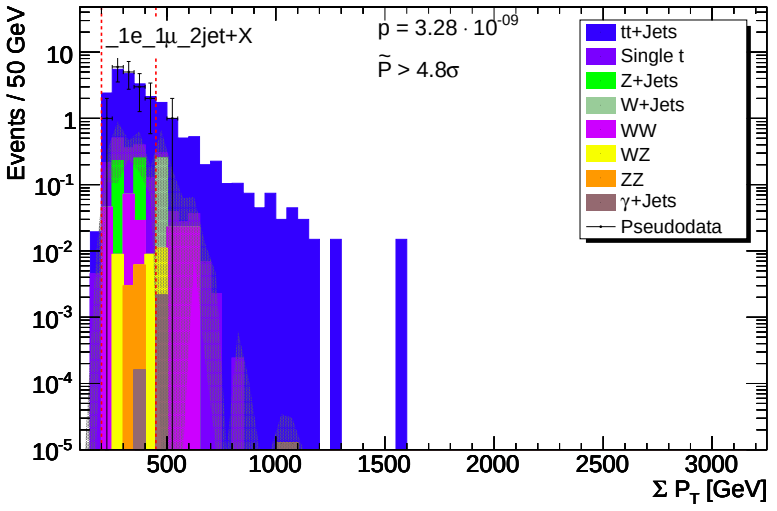


Figure 7.3.: Kinematic distribution of the most significant event class when searching for $t\bar{t}$ without using b-tagging.

the fact, that the b-production in the Standard Model is heavily dominated by the indirect production via $t\bar{t}$. Thus, every event class containing b-jets is dominated by $t\bar{t}$, too. The more specific event classes (e.g. $1e\ 2b\ \not{E}_{\mathrm{T}} + \mathrm{X}$, cf. figure 7.2) strongly suffer from lower statistics. In contrast, the more general event classes benefit from high statistics, since they combine the more specific ones.

The transparent error bands shown in the plots represent the uncertainties of the Standard Model prediction. They include the statistical uncertainties of the Monte Carlo Samples and all systematic general uncertainties (as described in chapter 5.3.6) as well as the uncertainties of the tagging efficiencies (as introduced in chapter 6.4.4).

The comparison of the efficiencies gained with and without using b-tagging shows a clear advantage of using b-tagging when searching for $t\bar{t}$. Without b-tagging, more specific classes (containing two leptons and two jets) have to be investigated. As can be seen in figure 7.3, these event classes are dominated by $t\bar{t}$, too, but achieve lower statistics than the corresponding event classes containing b-jets. Additionally, only a few significant event classes can be found without using b-tagging, whereas many significant event classes show up when using b-tagging.

Varying the b-tagging cut value slightly changes the absolute values of the significances, but the overall result remains the same.

7.1.2. Luminosity Scan

Using the most significant event class, a study of the discovery potential 7 TeV has been performed. Therefore, the assumed integrated luminosity has been varied from $1\,\mathrm{pb}^{-1}$ to $50\,\mathrm{pb}^{-1}$. It shows, that (when using b-tagging) deviations with significances of above $3\,\sigma$ can already be found with integrated luminosities of $10\,\mathrm{pb}^{-1}$. Significances of more than $4.8\,\sigma$ can be achieved with integrated luminosities of $20\,\mathrm{pb}^{-1}$. In Contrast, it would not be possible to discover $t\bar{t}$, using MUSiC without b-tagging, even with an integrated luminosity of up to $50\,\mathrm{pb}^{-1}$.

7.2. Fourth Generation Quarks

According to the Standard Model, matter consists of three generations of leptons and three generations of quarks (cf. chapter 2.1.1). However, nothing within the theory of the Standard Model limits the number of possible generation to that number. The Standard Model is rather built with only three generations due to the facts, that no other generations have been discovered (yet) and that the number of existing neutrino types (with $m_\nu < \frac{1}{2}m_Z$) is $N_\nu \approx 3$ according to the data taken during the LEP experiment [1]. Nevertheless, it can be shown, that the integration of a fourth generation of quarks into the Standard Model is possible (e.g. [93]). Since no fourth generation quark has been discovered yet, they need to possess a higher mass than the known quarks. Their mass also has to be high enough

so that the fourth generation quarks would not have been created in the former particle collider experiments. Due to the high $\sqrt{s}$ that is achieved at LHC, it will be possible to search for possible fourth generation quarks in the near future. One of the possible two fourth generation quarks is a heavy b-like quark, which is called b'. As it possesses a higher mass, the b' could possibly decay to a t-quark and a W-boson (as shown in figure 7.4), assuming the same behaviour like the quarks of the other three generations.

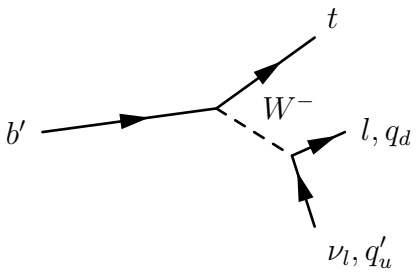


Figure 7.4.: Feynman diagram of the b'-quark decay.

A Monte Carlo sample of the $b'\,\overline{b'} \to t\,\bar{t}\;W^+\,W^- \to b\,\bar{b}\;W^+\,W^-\,W^+\,W^-$ decay channel with an assumed quark mass of $m_{b'} = 300\,\text{GeV}$ has been used to perform an analysis assuming a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ and an integrated luminosity of $2.5\,\text{fb}^{-1}$. Despite using this decay channel, no high significances can be expected in 1 lepton $1b$ $\not{E}_\text{T}$ event classes. These event classes are dominated by $t\bar{t}$, as this process possesses a cross section, which is about 20 times larger than the cross section of the used $b'\overline{b'}$ decay channel. Therefore, event classes with high lepton or jet multiplicities have to be investigated, since the four W-bosons of the decay $b'\overline{b'}$ can end up as four leptons or eight jets in an extreme case. Similar to the $t\bar{t}$ search, event classes containing both b-jets of the decay chain suffer from the b-tagging efficiency. Thus, event classes containing one b-jet + X (in addition to the leptons and jets) possess more significant deviations.

Investigating the benefit of using b-tagging for a search for b' gives an ambivalent result: On the one hand, event classes containing at least two leptons and one b-jet or one lepton, a high multiplicity of jets and one b-jet (e.g. figure 7.5) show high significances as these event classes are specific enough to allow a separation of b' from $t\,\bar{t}$ and at the same time general enough not to suffer from too low statistics. On the other hand, significant deviations can also be found without using b-tagging, when investigating more specific event classes, i.e. event classes with at least two leptons and a high jet multiplicity at the same time (e.g. figure 7.6).

One can conclude, that b-tagging is not such a big advantage in searches for signals which are very similar to $t\,\bar{t}$ (which is not surprising). This result does not depend on the choice of the b-tagging cut value, only the absolute values of the significances slightly change when varying the cut value.

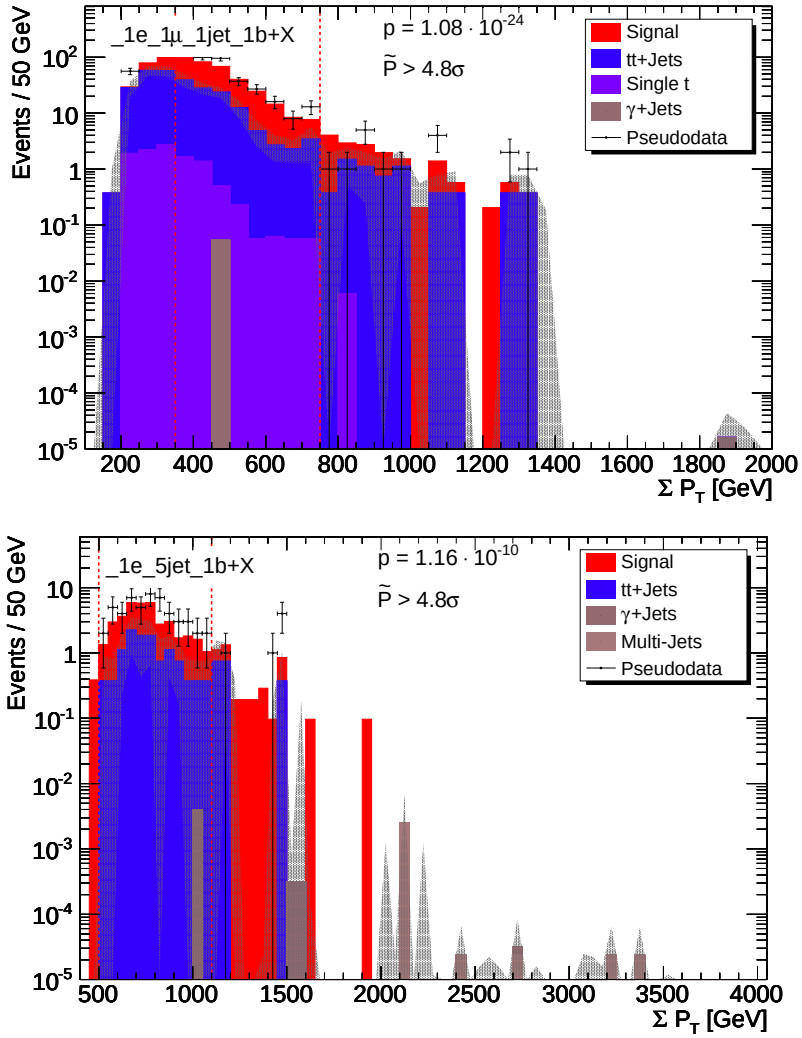


Figure 7.5.: Kinematic distributions of two of the most significant event classes when using b-tagging to search for b'.

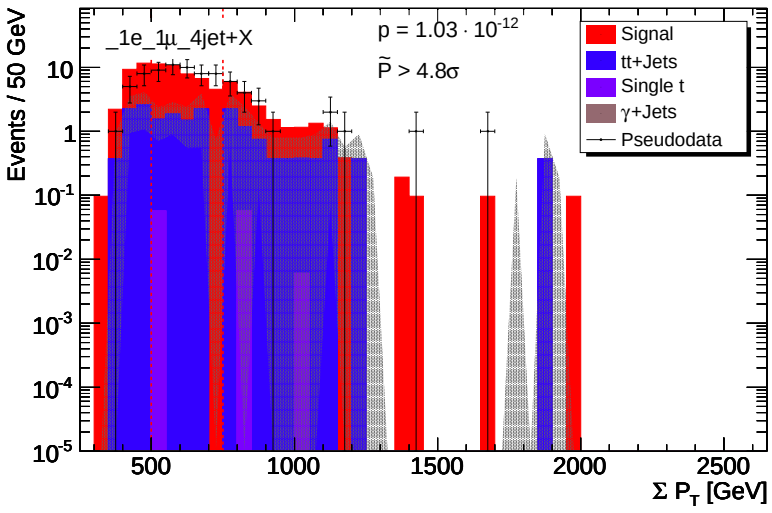


Figure 7.6.: Kinematic distribution of one of the most significant event classes when searching for b' without using b-tagging.

7.3. Supersymmetry

As mentioned in chapter 2.1.7, there are several questions that cannot be solved by the Standard Model alone, e.g. "What does dark matter consist of?", "Is there a unification of the three known forces?" and "How can the gravitational force be described?". To solve these questions, many extensions of the Standard Model have been developed. One of them is *Supersymmetry*. According to the Theory of Supersymmetry, a supersymmetric partner (sparticle) exists to each Standard Model particle, whose spin J differs by $1/2$ from the spin of the (Standard Model) particles, e.g. $J_{\text{lepton,quark}} = 1/2$, $J_{\text{slepton,squark}} = 0$. Without going into detail, it should be mentioned, that the theory of Supersymmetry can solve or help to solve the open questions mentioned above. The lightest sparticle is a promising candidate for dark matter, Supersymmetry leads to a unification of the three Standard Model forces and local Supersymmetry transformations could lead to the introduction of a gravitino and a graviton as the moderators of gravity.

As Supersymmetry leads to a vast number of free parameters, usually different scenarios are used, in order to reduce the number of parameters and thus be able to describe the phenomenology of Supersymmetry. One of these scenarios is the *Minimal Supersymmetric Standard Model* (MSSM), which represents the minimal extension of the Standard Model which is consistent with current experimental data and able to provide masses to all observed particles. Since the MSSM still contains more than 100 free parameters, their number has to be reduced further, to gain scenarios which can be used to make predictions. A prominent scenario is *minimal supergravity* (mSUGRA). Here the Supersymmetry breaking[65] takes places in a hidden sector (which cannot be observed) and is transferred to the visible world by the gravitational interaction. Requiring some additional assumptions (like e.g. CP-violation compatible with the measurements, i.e. a very small effect) one ends up with only five free parameters:

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

As these five parameters still lead to a vast parameter space, with the predictions strongly depending on their choice, physicists have agreed on several benchmark points. These points in the parameter space have been chosen as exemplary parameter sets, representing important phenomenological cases, which are not yet ruled out experimentally. In the following analysis, the benchmark point LM9 (Low Mass 9) has been used for testing purpose. Its parameters are summarised in table 7.1. It has been chosen, as a higher value of $\tan(\beta)$ is expected to lead to more b-jets being produced in the decay chains.

[65] The Supersymmetry breaking leads to differences between the particle and sparticle masses, which have to exist, since no sparticle has been discovered yet (leading to the assumption of high sparticle masses).

$m_{1/2}$ / GeV	m_0 / GeV	A_0	$\tan(\beta)$	$\mathrm{sgn}(\mu)$
175	1450	0	50	+

Table 7.1.: SUSY parameters of the LM9 benchmark point.

A global scan of all event classes has been performed, assuming a centre-of-mass energy of $\sqrt{s} = 7\,\mathrm{TeV}$ and an integrated luminosity of $10\,\mathrm{fb}^{-1}$. In total, the LM9 signal leads to 974 event classes (375 exclusive and 599 inclusive ones), when using a b-jet discriminator cut of 0.8. 167 (37 exclusive and 130 inclusive) of theses event classes show an expected significance of $\tilde{P} > 3\sigma$. This high number of significant event classes (considering the relatively low cross section of Supersymmetry[66], cf. table A.4) indicates, that SUSY manifests in many event classes, as expected due to the long decay chains. One can also see, that many of the significant event classes contain at least one b-jet (28 exclusive and 96 inclusive

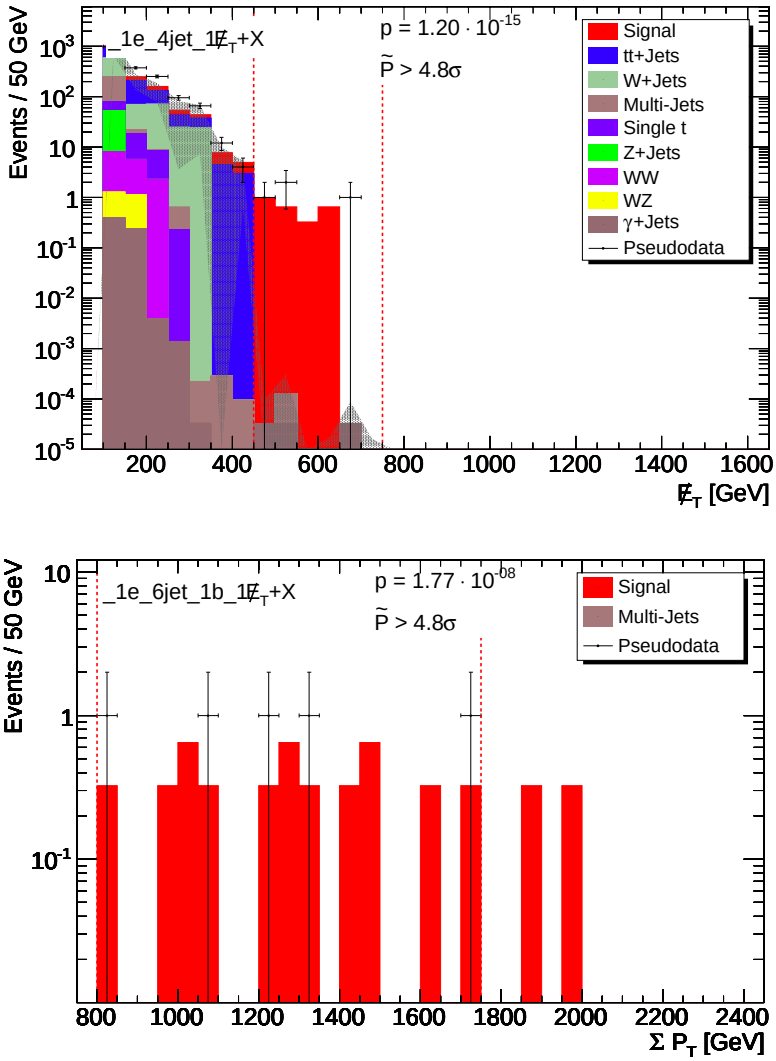


Figure 7.7.: Kinematic distributions of two of the most significant event classes when using b-tagging to search for the SUSY LM9 benchmark point. The majority of the event classes only contains signal entries and no expected background, like the distribution on the right.

[66] The other event classes are either dominated by $t\bar{t}$ or suffer from low statistics.

event classes), as expected. An expected significance of even $\widetilde{P} > 4.8\sigma$ can be found in 143 (30 exclusive and 113 inclusive) of the significant event classes. When applying the penalty factor for searching in many event classes (cf. chapter 5.3.5 and figure 5.6), this equals a significance of $> 3\sigma$, i.e. the LM9 signal leads to deviations which remain significant when performing a global search. Figure 7.7 shows two of the event classes possessing $\widetilde{P} > 4.8\sigma$.

Comparing the results of the b-tagging based search and the analysis without b-tagging, considerably more deviations with high significances can be found when using b-tagging. Without b-tagging, the LM9 signal leads to 441 event classes (176 exclusive and 265 inclusive ones). From these event classes 62 (18 exclusive and 44 inclusive) achieve significances of $\widetilde{P} > 3\sigma$ and only 49 event classes (13 exclusive and 36 inclusive ones) achieve significances of $\widetilde{P} > 4.8\sigma$. Thus, the LM9 signal can be found easier with MUSiC, if b-tagging is used. This result does not change when varying the b-tagging cut value by ± 0.05, which is due to the fact that the majority of the highly significant classes possesses no expected background. Thus every single signal entry leads to a high significance, i.e. the variation of the total number of entries caused by varying the cut value only has a minor effect on the large number of highly significant classes, which are found with MUSiC.

8. Conclusion

Several topics related to the extension of a Model Unspecific Search at CMS with respect to tagging b-jets have been presented.

The MUSiC Analysis framework has been extended successfully to provide the possibility of identifying b-jets within the events and process event classes containing b-jets. The functionality of the implementation of b-tagging into the MUSiC framework has successfully been verified and a reasonable discriminator cut value as well as the resulting b-tagging efficiency, mistag rates and the b-jet purity have been determined.

Simulated test scenarios have been investigated to determine the benefit from b-tagging when performing a search for new physics with MUSiC. It has been found, that a significantly lower integrated luminosity is sufficient to recover the top quark ($t\,\bar{t}$) in the early CMS collision data. Additionally, it has been shown for the example of the supersymmetric LM9 benchmark point, that searches for new physics leading to final states with many b-jets or complex final states containing at least one b-jet, benefit from using b-tagging. Unfortunately, as $t\,\bar{t}$ possesses a relatively high cross section, there is no benefit from using b-tagging in searches for new physics resulting in final states, which are similar to the $t\,\bar{t}$ final states, e.g. b'.

A. Appendix: Tables

Name	L1 prerequisite	requirement
Muon Trigger		
Mu9	≥ 1 μ candidate with $p_T > 7\,\text{GeV}$	≥ 1 μ candidate with $p_T > 9\,\text{GeV}$
DoubleMu0	≥ 2 μ candidates (no p_T requirement)	≥ 2 μ candidates (no p_T requirement)
Electron Trigger		
Ele20_SW_L1R	≥ 1 e/γ candidate with $p_T > 8\,\text{GeV}$	≥ 1 e candidate with $p_T > 20\,\text{GeV}$
DoubleEle10_SW_L1R	≥ 2 e/γ candidates with $p_T > 5\,\text{GeV}$	≥ 2 e candidates with $p_T > 10\,\text{GeV}$
Photon Trigger		
Photon25_L1R	≥ 1 e/γ candidate with $p_T > 8\,\text{GeV}$	≥ 1 γ candidate with $p_T > 25\,\text{GeV}$
DoublePhoton15_L1R	≥ 2 e/γ candidates with $p_T > 5\,\text{GeV}$	≥ 2 γ candidates with $p_T > 15\,\text{GeV}$
Cross-channel Trigger		
L1Mu14_L1SingleEG10	≥ 1 μ candidate with $p_T > 14\,\text{GeV}$ **AND** ≥ 1 e/γ candidate with $p_T > 10\,\text{GeV}$	
L2Mu5_Photon9_L1R	≥ 1 μ candidate with $p_T > 5\,\text{GeV}$ **AND** ≥ 1 e/γ candidate with $p_T > 5\,\text{GeV}$	

Table A.1.: HLT triggers currently used in the MUSiC analysis.

category 1	$0.8 < E/p_{\text{in}} < 1.2 \quad \wedge \quad f_{\text{brem}} < 0.06$ (barrel) or < 0.1(endcap)
category 2	$0.8 < E/p_{\text{in}} < 1.2 \quad \wedge \quad f_{\text{brem}} > 0.06$ (barrel) or > 0.1(endcap)
category 3	$E/p_{\text{in}} < 0.8 \quad \vee \quad E/p_{\text{in}} > 1.2$

Table A.2.: Definition of the electron categories used for electron identification cuts. p_{in} is the estimated momentum at the track vertex, $f_{\text{brem}} = (p_{\text{in}} - p_{\text{out}})/p_{\text{in}}$ is an estimation of the effects due to bremsstrahlung.

variable	region	category 1	category 2	category 3
H/E	barrel	< 0.042	< 0.050	< 0.045
	endcap	< 0.037	< 0.055	< 0.050
$\sigma_{\eta\eta}$	barrel	< 0.011	< 0.0125	< 0.01
	endcap	< 0.0252	< 0.0265	< 0.026
$\Delta\eta_{\text{in}}$	barrel	< 0.003	< 0.0055	< 0.0065
	endcap	< 0.0055	< 0.006	< 0.0075
$\Delta\phi_{\text{in}}$	barrel	< 0.016	< 0.032	< 0.0525
	endcap	< 0.035	< 0.025	< 0.065
$E_{\text{seed}}/p_{\text{in}}$	barrel	> 0.94	> 0.24	> 0.11
	endcap	> 0.83	> 0.32	> 0

Table A.3.: Set of cuts for the "tight" electron identification. For $E/p_{\text{in}} > 1.5$, the $\Delta\phi_{\text{in}}$ cuts are replaced by 0.09 in the barrel region and 0.092 in the endcaps

Process	Cross section / pb	# of events	Path in dbs
PYTHIA samples			
QCD	$6.041 \cdot 10^7$	5269664	/QCD_Pt30/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$9.238 \cdot 10^5$	3221800	/QCD_Pt80/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$2.547 \cdot 10^4$	3171950	/QCD_Pt170/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$1.256 \cdot 10^3$	3282665	/QCD_Pt300/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	87.98	2159497	/QCD_Pt470/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	2.186	2181700	/QCD_Pt800/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$1.122 \cdot 10^{-2}$	1185024	/QCD_Pt1400/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$t\bar{t}$	94.30	632010	/TTbar/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
γ + jets	$2.007 \cdot 10^4$	1026794	/PhotonJet_Pt30/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$5.565 \cdot 10^2$	1287711	/PhotonJet_Pt80/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	24.37	1139400	/PhotonJet_Pt170/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	1.636	1034266	/PhotonJet_Pt300/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$1.360 \cdot 10^{-1}$	1091179	/PhotonJet_Pt470/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$3.477 \cdot 10^{-3}$	1075640	/PhotonJet_Pt800/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	$1.286 \cdot 10^{-5}$	1331025	/PhotonJet_Pt1400/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$W \to e\nu$	$7.899 \cdot 10^3$	2087865	/Wenu/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$W \to \mu\nu$	$7.899 \cdot 10^3$	3116114	/Wmunu/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$W \to \tau\nu$	$7.899 \cdot 10^3$	2185655	/Wtaunu/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$Z \to ee$	$1.300 \cdot 10^3$	2461575	/Zee/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$Z \to \mu\mu$	$1.300 \cdot 10^3$	2872536	/Zmumu/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$Z \to \tau\tau$	$1.300 \cdot 10^3$	2195255	/Ztautau/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
WW	28	122980	/WW/Summer09-MC_31X_V3_7TeV-v1/GEN-SIM-RECO
WZ	10.5	118130	/WZ/Summer09-MC_31X_V3_7TeV-v1/GEN-SIM-RECO
ZZ	4.3	145368	/ZZ/Summer09-MC_31X_V3_7TeV-v1/GEN-SIM-RECO
$Z\gamma \to b\bar{b}\gamma$	8	109615	/ZGToBBbar_7TeV-pythia6/Spring10-START3X_V26-v1/GEN-SIM-RECO
SUSY LM9	7.134	220000	/LM9/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
MADGRAPH samples			
SingleTop	$9.9 \cdot 10^{-1}$	412055	/SingleTop_sChannel-madgraph/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	63	528593	/SingleTop_tChannel-madgraph/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
	10.56	466437	/SingleTop_tWChannel-madgraph/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
W + jets	$2.417 \cdot 10^4$	10068895	/WJets-madgraph/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
Z + jets	$2.4 \cdot 10^3$	1084921	/ZJets-madgraph/Spring10-START3X_V26_S09-v1/GEN-SIM-RECO
$b'\bar{b}'$	4.510	116783	/BprimeToTWTW_WInc_M-300_7TeV-madgraph/Spring10-START3X_V26-v1/GEN-SIM-RECO

Table A.4.: Used Monte Carlo samples, simulating 7 TeV collisions.

List of Figures

List of Tables

List of Equations

Bibliography

[1] Particle Data Group. *Particle Physics Booklet*. 2008th ed. URL: http://pdg.lbl.gov/2009/download/rpp-2008-booklet.pdf. (Cit. on pp. 1, 6–9, 17, 22, 65, 66, 84, 97, 99).

[2] Bundesverfassungsgericht. *Verfassungsbeschwerde gegen Versuchsreihen am "CERN" unzulässig*. Press Release. 2010. URL: http://www.bundesverfassungsgericht.de/pressemitteilungen/bvg10-014.html. (Cit. on p. 1).

[3] CERN. *Two circulating beams bring first collisions in the LHC*. Press Release. 2010. URL: http://public.web.cern.ch/press/pressreleases/Releases2009/PR17.09E.html. (Cit. on p. 2).

[4] CERN. *LHC research programme gets underway*. Press Release. 2010. URL: http://public.web.cern.ch/press/PressReleases/Releases2010/PR07.10E.html. (Cit. on p. 2).

[5] Prof. M.A. Thomson. *Particle Physics*. Lecture. Department of Physics, University of Cambridge. 2009. URL: http://www.hep.phy.cam.ac.uk/~thomson/lectures/partIIIparticles/Handout1_2009.pdf. (Cit. on p. 3).

[6] Welt der Physik. 2004. URL: http://www.weltderphysik.de/de/1035.php. (Cit. on p. 5).

[7] Wolfgang Demtröder. 'Zeittafel zur Kern- und Hochenergiephysik'. In: *Experimentalphysik 4 - Kern-, Teilchen- und Astrophysik*. 2nd ed. Spinger-Verlag, 2005, pp. 435, 436. ISBN: 3-540-21451-8. (Cit. on p. 5).

[8] E. Rutherford. *The Scattering of α and β Particles by Matter and the Structure of the Atom*. Philosophical Magazine. 1911. URL: http://www.ffn.ub.es/luisnavarro/nuevo_maletin/Rutherford%20(1911),%20Structure%20atom%20.pdf. (Cit. on p. 5).

[9] University of Oxford Department of Physics. *The Stanford Linear Accelerator Center (SLAC)*. URL: http://www.physics.ox.ac.uk/documents/PUS/dis/SLAC.htm. (Cit. on p. 5).

[10] ETH Zurich Institute for Particle Physics. URL: http://www.ipp.phys.ethz.ch/aboutus/_media/atom_zoom_b.jpg. (Cit. on pp. 5, 95).

[11] Wolfgang Nolting. *Grundkurs Theoretische Physik 5/1: Quantenmechanik - Grundlagen*. 6th ed. Springer-Verlag, 2004. ISBN: 3-540-40071-0. (Cit. on p. 5).

[12] Wolfgang Nolting. *Grundkurs Theoretische Physik 5/2: Quantenmechanik - Methoden und Anwendungen*. 5th ed. Springer-Verlag, 2004. ISBN: 3-540-40072-9. (Cit. on p. 5).

[13] Peter Schmüser. *Feynman-Graphen und Eichtheorien für Experimentalphysiker*. 2nd ed. Springer-Verlag, 1995. ISBN: 3-540-58486-2. (Cit. on pp. 5, 18, 95).

[14] Particle Data Group. *Particle Listings*. 2009th ed. URL: http://pdg.lbl.gov/2009/listings/contents_listings.html. (Cit. on pp. 6, 9, 97).

[15] Nelli Pukhaeva. *Search for Pentaquarks in the Hadronic Decays of the Z Boson with the DELPHI detector at LEP*. International Europhysics Conference on High Energy Physics. 2005. URL: http://cdsweb.cern.ch/record/1109383/files/084HEP2005_084.pdf. (Cit. on p. 7).

[16] Wolfgang Ochs. *Searching for the Scalar Glueball*. 2008. URL: http://arxiv.org/PS_cache/arxiv/pdf/0805/0805.0997v1.pdf. (Cit. on p. 7).

[17] Official homepage of the Super-Kamiokande experiment. URL: http://www-sk.icrr.u-tokyo.ac.jp/sk/index-e.html. (Cit. on p. 9).

[18] Super-Kamiokande Collaboration. *A Measurement of Atmospheric Neutrino Oscillation Parameters by Super-Kamiokande I*. 2005. URL: http://arxiv.org/PS_cache/hep-ex/pdf/0501/0501064v2.pdf. (Cit. on p. 9).

[19] Katherine Brading. *Which Symmetry? Noether, Weyl, and Conservation of Electric Charge*. 2002. URL: http://www.nd.edu/~kbrading/Research/WhichSymmetryStudiesJuly01.pdf. (Cit. on p. 11).

[20] Gerhard Soff. *Quantenfeldtheorie*. Lecture. Institute for Theoretical Physics, Technical University of Dresden. 2002. URL: http://www.physik.tu-dresden.de/itp/members/soffscripte/qf_2001.pdf. (Cit. on p. 11).

[21] URL: http://puhep1.princeton.edu/mumu/physics/. (Cit. on p. 22).

[22] Official homepage of the LHC experiment design report. URL: http://lhc.web.cern.ch/lhc/LHC-DesignReport.html. (Cit. on pp. 22, 23, 97).

[23] CERN. URL: http://cdsweb.cern.ch/record/987579. (Cit. on pp. 23, 95).

[24] CERN. URL: http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/images/cryodipole.jpg. (Cit. on pp. 24, 95).

[25] Lyndon Evans. *The Large Hadron Collider*. New Journal of Physics. 2007. URL: http://iopscience.iop.org/1367-2630/9/9/335/pdf/njp7_9_335.pdf. (Cit. on pp. 24, 95).

[26] Metin Ata. 'Studies of Models with Large Extra Dimensions in CMS'. Diploma thesis. RWTH Aachen, 2009. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/ata_diploma.pdf. (Cit. on pp. 25, 95).

[27] Official homepage of the CMS experiment. URL: http://cms.web.cern.ch/cms/index.html. (Cit. on p. 25).

[28] CMS Collaboration. *The CMS experiment at the CERN LHC*. 2008. URL: http://www.hephy.at/fileadmin/user_upload/Publikationen/jinst8_08_s08004.pdf. (Cit. on pp. 25, 29–31, 34, 35, 37, 95).

[29] Official CMS homepage with explanations to the detector. URL: http://cms.web.cern.ch/cms/Detector/Work/index.html. (Cit. on pp. 26, 95).

[30] CMS Collaboration. *CMS Tracker Technical Design Report*. 1998. URL: http://cmsdoc.cern.ch/cms/TDR/TRACKER/tracker_tdr.html. (Cit. on pp. 27, 28, 95).

[31] CMS Collaboration. *CMS ECAL Technical Design Report*. 1997. URL: http://cmsdoc.cern.ch/cms/TDR/ECAL/ecal.html. (Cit. on pp. 29–31).

[32] CMS Collaboration. *CMS Physics - Technical Design Report Vol I: Detector Performance and Software*. 2006. URL: http://cmsdoc.cern.ch/cms/cpt/tdr. (Cit. on pp. 31, 36, 42, 95).

[33] CMS Collaboration. *CMS HCAL Technical Design Report*. 1997. URL: http://cmsdoc.cern.ch/cms/TDR/HCAL/hcal.html. (Cit. on pp. 31–33).

[34] CMS Collaboration. *CMS Muon Technical Design Report*. 1997. URL: http://cmsdoc.cern.ch/cms/TDR/MUON/muon.html. (Cit. on pp. 33, 34, 37, 95).

[35] CMS Collaboration. *CMS: The Magnet Project - Technical Design Report*. 1997. URL: http://cmsdoc.cern.ch/cms/TDR/MAGNET/CMS_TDR_10_97/CMS_Pageshtml/toc.html. (Cit. on p. 33).

[36] CMS Collaboration. *Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays*. CMS Paper (CMS PAPER CFT-09-015). 2010. URL: http://arxiv.org/PS_cache/arxiv/pdf/0910/0910.5530v2.pdf. (Cit. on p. 34).

[37] CMS Collaboration. *CMS The TriDAS Project - Technical Design Report, Vol I: The Trigger Systems*. 2000. URL: http://cmsdoc.cern.ch/cms/TDR/TRIGGER-public/CMSTrigTDR.pdf. (Cit. on p. 38).

[38] CMS Collaboration. *Twiki documentation of data analysis at CMS*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/WorkBook. (Cit. on p. 39).

[39] CMS Collaboration. *CMS - The Computing Project: Technical Design Report*. 2005. URL: http://cdsweb.cern.ch/record/838359/files/lhcc-2005-023.pdf. (Cit. on p. 39).

[40] Carsten Hof. 'Implementation of a Model-Independent Search for New Physics with the CMS Detector Exploiting the World-Wide LHC Computing Grid'. PhD thesis. RWTH Aachen, 2009. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/hof_phd.pdf. (Cit. on pp. 40, 51, 54, 57, 62, 63, 95).

[41] CMS Collaboration. *Twiki documentation of data analysis at CMS, using the LHC Computing Grid*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookRunningGrid. (Cit. on p. 40).

[42] Stefano Lacaprara, Federica Fanzago, and Marco Corvo. *CRAB tutorial - Cms Remote Analysis Builder*. Talk. 2005. URL: http://www.pd.infn.it/~lacaprar/talks/CRAB_tutorial.pdf. (Cit. on p. 40).

[43] W. Adam, V. Adler, et al. *PAT: the CMS Physics Analysis Toolkit*. CMS Conference Report (CMS CR-2009/083). 2009. URL: http://cdsweb.cern.ch/record/1196152. (Cit. on p. 40).

[44] CMS Collaboration. *Twiki documentation of data analysis at CMS, using the Physics Analysis Toolkit*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookPAT. (Cit. on p. 40).

[45] Official Physics eXtension Library homepage. URL: http://sourceforge.net/apps/trac/pxl/. (Cit. on p. 40).

[46] Official ROOT homepage. URL: http://root.cern.ch/drupal/. (Cit. on p. 41).

[47] S. Baffioni, C. Charlot, et al. *Muon Reconstruction in the CMS Detector*. CMS Analysis Note (CMS AN-2008/097). 2008. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2008_097_v4.pdf. (Cit. on p. 41).

[48] R. Frühwirth. 'Application of Kalman filtering to track and vertex fitting'. In: *Nuclear Instruments and Methods in Physics Research Section A* 262 (1987), pp. 444–450. URL: http://cdsweb.cern.ch/record/178627. (Cit. on p. 42).

[49] G. Abbiendi, J. Alcaraz, et al. *Electron Reconstruction in CMS*. CMS Note (CMS NOTE 2006/40). 2006. URL: http://cdsweb.cern.ch/record/934070/files/NOTE2006_040.pdf. (Cit. on p. 42).

[50] Pascal Vanlaer. *Electron and photon reconstruction in CMS*. CMS Conference Report (CMS CR-2009/201). 2009. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?type=CR&year=2009&files=CR2009_201.pdf. (Cit. on pp. 42, 43).

[51] R. Frühwirth. 'Track Fitting with Non-Gaussian Noise'. In: *Computer Physics Communications* 100 (1997), pp. 1–16. URL: http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=1997CoPhC.100....1F&link_type=EJOURNAL. (Cit. on p. 42).

[52] J. Nysten. *Photon Reconstruction in CMS*. CMS Conference Report (CMS CR 2004/004). 2004. URL: http://cms-secr.web.cern.ch/cms-secr/documents/04/cr04_004.pdf. (Cit. on p. 43).

[53] Official CMS homepage with CMS-related pictures. URL: http://cms.web.cern.ch/cms/Media/Images/EventDisplays/7_0TeVCollisions/index.html. (Cit. on pp. 43, 95).

[54] CMS Collaboration. *Performance of Jet Algorithms in CMS*. CMS Physics Analysis Summary (CMS PAS JME-07-003). 2009. URL: http://cdsweb.cern.ch/record/1198227/files/JME-07-003-pas.pdf. (Cit. on pp. 44, 47).

[55] Gerald C. Blazey, Jay R. Dittmann, et al. *Run II Jet Physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop*. 2000. URL: http://arxiv.org/PS_cache/hep-ex/pdf/0005/0005012v2.pdf. (Cit. on pp. 45, 47, 95).

[56] A. Bhatti, F. Chlebana, et al. *Performance of the SISCone Jet Clustering Algorithm*. CMS Analysis Note (CMS AN-2008/002). 2008. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2008_002_v3.pdf. (Cit. on pp. 44, 46, 95).

[57] Gavin P. Salam and Gregory Soyez. *A practical Seedless Infrared-Safe Cone jet algorithm*. 2007. URL: http://arxiv.org/PS_cache/arxiv/pdf/0704/0704.0292v2.pdf. (Cit. on p. 44).

[58] P. Schieferdecker, M. Vazquez Acosta, et al. *Performance of the k_t jet algorithm in CMSSW*. CMS Internal Note (CMS IN 2007/062). 2007. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?type=IN&year=2007&files=IN2007_062.pdf. (Cit. on p. 46).

[59] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. *The anti-k_t jet clustering algorithm*. 2008. URL: http://arxiv.org/PS_cache/arxiv/pdf/0802/0802.1189v2.pdf. (Cit. on pp. 46–48, 95).

[60] Christophe Saout. *Using Anti-Kt 0.5 as Default Jet Algorithm for b-Tagging instead of Iter. Cone 0.5*. b-tag group meeting, 21.09.09. URL: http://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=68060. (Cit. on p. 47).

[61] CMS Collaboration. *Plans for Jet Energy Corrections at CMS*. CMS Physics Analysis Summary (CMS PAS JME-07-002). 2008. URL: http://cdsweb.cern.ch/record/1194485/files/JME-07-002-pas.pdf. (Cit. on pp. 47, 64).

[62] CMS Collaboration. *Jets in 0.9 and 2.36 TeV pp Collisions*. CMS Physics Analysis Summary (CMS PAS JME-10-001). 2010. URL: http://cdsweb.cern.ch/record/1248210/files/JME-10-001-pas.pdf. (Cit. on p. 49).

[63] CMS Collaboration. *Jet and MET Commissioning Results from 7 TeV Collision Data*. CMS Detector Performance Summary (CMS DPS -2010/014). 2010. URL: http://cms.cern.ch/iCMS/user/getnote?type=DP&year=2010&files=DP2010_014.pdf. (Cit. on p. 49).

[64] S. Esen, G. Landsberg, et al. *$\not\!\!E_T$ Performance in CMS*. CMS Analysis Note (CMS AN-2007/041). 2008. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2007_041_v6.pdf. (Cit. on p. 49).

[65] Greg Landsberg and Filip Moortgat. *MET Reconstruction, Performance, and Validation*. CMS Analysis Note (CMS AN-2008/089). 2009. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2008_089_v3.pdf. (Cit. on p. 50).

[66] CMS Collaboration. *Missing Transverse Energy Performance in Minimum-Bias and Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV*. CMS Physics Analysis Summary (CMS PAS JME-10-004). 2010. URL: http://cms-physics.web.cern.ch/cms-physics/public/JME-10-004-pas.pdf. (Cit. on p. 50).

[67] Thomas Hebbeker. *A Global Comparison between L3 Data and Standard Model Monte Carlo - a first attempt*. L3 note 2305. 1998. URL: http://web.physik.rwth-aachen.de/~hebbeker/l3note_2305.pdf. (Cit. on p. 51).

[68] D0 Collaboration. *Search for New Physics in e mu X Data at D0 Using Sleuth: A Quasi-Model-Independent Search Strategy for New Physics*. Physical Review D 62:092004. 2000. URL: http://arxiv.org/PS_cache/hep-ex/pdf/0006/0006011v2.pdf. (Cit. on p. 51).

[69] D0 Collaboration. *A Quasi-Model-Independent Search for New Physics at Large Transverse Momentum*. Physical Review D 64:012004. 2001. URL: http://cdsweb.cern.ch/record/477731/files/0011067.pdf. (Cit. on p. 51).

[70] CDF Collaboration. *Model-Independent and Quasi-Model-Independent Search for New Physics at CDF*. Physical Review D 78:012002. 2008. URL: http://www-cdf.fnal.gov/physics/preprints/cdf8952_prd1_v03.pdf. (Cit. on p. 51).

[71] H1 Collaboration. *A General Search for New Phenomena at HERA*. Physics Letters B674:257-268. 2009. URL: http://arxiv.org/PS_cache/arxiv/pdf/0901/0901.0507v1.pdf. (Cit. on p. 51).

[72] Philipp Biallass. 'Commissioning of the CMS Muon Detector and Development of Generic Search Strategies for New Physics'. PhD thesis. RWTH Aachen, 2009. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/biallass_phd.pdf. (Cit. on pp. 51, 52, 61–63).

[73] Stefan Schmitz. 'Model Unspecific Search for New Physics with High p_T Photons in CMS'. Diploma thesis. RWTH Aachen, 2009. URL: http://web.physik.rwth-aachen.de/~hebbeker/theses/schmitz_diploma.pdf. (Cit. on pp. 51, 63).

[74] Hitoshi Murayama. *Outlook: The Next Twenty Years*. Talk at Lepton Photon 2003 at Fermilab. URL: http://hitoshi.berkeley.edu/LP03/Outlook.pdf. (Cit. on pp. 52, 95).

[75] Holger Pieta. *MUSiC - Model Unspecific Search in CMS - First data*. 2010. URL: http://indico.cern.ch/getFile.py/access?contribId=3&resId=0&materialId=slides&confId=76791. (Cit. on p. 52).

[76] J. Branson, M. Gallinaro, P. Ribeiro, R. Salerno, and M. Sani. *A cut based method for electron identification in CMS*. CMS Analysis Note (CMS AN-2008/082). 2008. URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2008_082_v1.pdf. (Cit. on p. 59).

[77] Bruce Knuteson. *Systematic analysis of HEP collider data*. Nuclear Instruments and Methods in Physics Research. 2004. URL: http://cdsweb.cern.ch/record/687640. (Cit. on p. 62).

[78] CMS Collaboration. *CMS Physics - Technical Design Report Volume II: Physics Performance*. 2006. URL: http://cmsdoc.cern.ch/cms/cpt/tdr/ptdr1_final_colour.pdf. (Cit. on p. 64).

[79] URL: https://twiki.cern.ch/twiki/bin/view/CMS/AachenPdfUncertainties. (Cit. on p. 64).

[80] CMS Collaboration. *B-tagging observables and event display from first 7 TeV data*. CMS Detector Performance Summary (CMS DPS-2010/015). 2010. URL: http://cms.cern.ch/iCMS/user/getnote?type=DP&year=2010&files=DP2010_015.pdf. (Cit. on pp. 66, 95).

[81] CMS Collaboration. *Algorithms for b Jet Identification in CMS*. CMS Physics Analysis Summary (CMS PAS BTV-09-001). 2009. URL: http://cdsweb.cern.ch/record/1194494/files/BTV-09-001-pas.pdf. (Cit. on p. 68).

[82] Homepage with plots for b-tagging validation. URL: https://cms-btag-validation.web.cern.ch/cms-btag-validation/validation/. (Cit. on pp. 68, 72–74, 96).

[83] CMS Collaboration. *Impact of Tracker Misalignment on the CMS b-Tagging Performance*. CMS Physics Analysis Summary (CMS PAS BTV-07-003). 2009. URL: http://cdsweb.cern.ch/record/1198691/files/BTV-07-003-pas.pdf. (Cit. on p. 68).

[84] P. Demin, S. de Visscher, A. Bocci, and R. Ranieri. *Tagging b jets with electrons and muons at CMS*. CMS Note (CMS NOTE 2006/043). 2006. URL: http://cdsweb.cern.ch/record/951390/files/NOTE2006_043.pdf. (Cit. on p. 68).

[85] A. Rizzi, F. Palla, and G. Segneri. *Track impact parameter based b-tagging with CMS*. CMS Note (CMS NOTE 2006/019). 2006. URL: http://cdsweb.cern.ch/record/927385/files/NOTE2006_019.pdf. (Cit. on p. 69).

[86] Armin Scheurer. 'Algorithms for the Identification of b-Quark Jets with First Data at CMS'. PhD thesis. Universität Karlsruhe, 2008. URL: http://ekpwww.physik.uni-karlsruhe.de/pub/web/thesis/iekp-ka2008-19.pdf. (Cit. on p. 70).

[87] C. Weiser. *A Combined Secondary Vertex Based B-Tagging Algorithm in CMS*. CMS Note (CMS NOTE 2006/014). 2006. URL: http://cdsweb.cern.ch/record/927399/files/NOTE2006_014.pdf. (Cit. on p. 70).

[88] CMS Collaboration. *Commissioning of b-jet identification with pp collisions at* $\sqrt{s} = 7$ *TeV*. CMS Physics Analysis Summary (CMS PAS BTV-10-001). 2010. URL: http://cdsweb.cern.ch/record/1279144/files/BTV-10-001-pas.pdf. (Cit. on p. 73).

[89] S. Lowette, J. D'Hondt, J. Heyninck, and P. Vanlaer. *Offline Calibration of b-Jet Identification Efficiencies*. CMS Note (CMS NOTE 2006/013). 2006. URL: http://cdsweb.cern.ch/record/927388/files/NOTE2006_013.pdf. (Cit. on p. 73).

[90] CMS Collaboration. *Performance Measurement of b-tagging Algorithms Using Data containing Muons within Jets*. CMS Physics Analysis Summary (CMS PAS BTV-07-001). 2009. URL: http://cdsweb.cern.ch/record/1194483/files/BTV-07-001-pas.pdf. (Cit. on p. 73).

[91] C. J. Clopper and E. S. Pearson. *The Use of Confidence or Fiducial Limits illustrated in the Case of the Binomial*. 1934. URL: http://biomet.oxfordjournals.org/content/26/4/404.full.pdf. (Cit. on p. 74).

[92] Robert D. Cousins, Kathryn E. Hymes, and Jordan Tucker. *Frequentist Evaluation of Intervals Estimated for a Binomial Parameter and for the Ratio of Poisson Means*. 2009. URL: http://arxiv.org/PS_cache/arxiv/pdf/0905/0905.3831v2.pdf. (Cit. on p. 74).

[93] C. D. Froggatt, D. J. Smith, and H. B. Nielsen. *Could There Be a Fourth Generation of Quarks Without More Leptons?* 1995. URL: http://arxiv.org/PS_cache/hep-ph/pdf/9603/9603436v1.pdf. (Cit. on p. 84).

Danksagung

Schlussendlich neigt sich die Arbeit an meiner Diplomarbeit dem Ende entgegen. Daher ist es an der Zeit, den Personen zu danken, ohne welche es diese Diplomarbeit nicht gäbe.

Zu allererst möchte ich mich bei meinen Kollegen am III. Physikalischen Institut A für das sehr gute Arbeitsklima, das Korrekturlesen diverser Vorträge und Diplomarbeitsentwürfe sowie die vielen offenen Ohren im Bezug auf Fragen und Probleme bedanken. Besonders zu erwähnen sind hier: Metin Ata, Michael Bontenackels (für die schnellst mögliche Beseitigung diverser Computerprobleme), Adriana Del Piero (als unverzichtbare, immer fröhliche Ansprechpartnerin im Sekretariat), Deborah Duchardt, Andreas Güth, Prof. Dr. Thomas Hebbeker, Dr. Kerstin Hoepfner, Dr. Carsten Hof, Dr. Carsten Magaß, Dr. Markus Merschmeyer, Dr. Arnd Meyer, Mark Olschewski, Paul Papacz, Holger Pieta, Stefan Schmitz sowie Peter Vonhoegen.

Besonders bedanken möchte ich mich bei Prof. Dr. Thomas Hebbeker für die Möglichkeit meine Diplomarbeit bei der CMS Gruppe des III. Physikalischen Institut A zu verfassen, sowie für seine konstruktive Kritik an ihrer Entwurfversion. Ebenso gilt mein Dank Prof. Dr. Christopher Wiebusch vom III. Physikalischen Institut B, der sich bereit erklärte, meine Diplomarbeit als Zweitgutachter zu bewerten.

Außerhalb der Welt der Physik gebührt mein Dank sowohl Anna-Maria Panitz für ihren Einsatz bei der Ausmerzung jeglicher Fehler im Bezug auf die englische Sprache, als auch meiner Familie, die mir das Studium überhaupt erst ermöglicht und mich in jeder erdenklichen Hinsicht dabei unterstützt hat.

Danke!

Selbstständigkeitserklärung

Ich, Erik Dietz-Laursonn, erkläre hiermit, dass ich die Diplomarbeit ohne fremde Hilfe und ohne Benutzung anderer, als der angegebenen Quellen und Hilfsmittel angefertigt habe. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Aachen, den 22.10.2010

(Dietz-Laursonn)