

**Search for new physics in pp collision events
with one electron and missing transverse energy
using CMS data**

von
Philipp Millet

Masterarbeit in Physik

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

Dezember 2013

erstellt im
III. Physikalischen Institut A
Prof. Dr. Thomas Hebbeker

Zweitgutachter
Prof. Dr. Christopher Wiebusch

Zusammenfassung

Das Thema dieser Arbeit ist die Suche nach neuer Physik mit einem Elektron und fehlender transversaler Energie im Endzustand. Die Suche basiert auf Daten von Proton-Proton Kollisionen bei einer Schwerpunktsenergie von 8 TeV, die mit dem CMS Detektor am LHC gemessen wurden. Insgesamt werden 20 fb^{-1} an Daten genutzt. Dies entspricht dem vollen Proton-Proton Datensatz aus dem Jahr 2012.

Viele Erweiterungen des Standardmodells der Teilchenphysik postulieren die Existenz von weiteren elektrisch geladenen Bosonen. Letztere werden oft W' genannt. In dieser Arbeit wird ein vereinfachtes W' -Modell benutzt, welches in seinen Eigenschaften dem Standardmodell gleicht. Es wird zwischen drei Szenarien unterschieden. Zwei der drei Szenarien enthalten Interferenzeffekte zwischen dem Standardmodell W und dem W' . Darüber hinaus wird ein Modell mit einer vier-Fermion Kontakwechselwirkung untersucht.

Die Ergebnisse der Suche zeigen keine signifikante Abweichung von der Standardmodellvorhersage. Die für die Bestimmung der Ausschlussgrenzen verwendete multi-bin Methode stellt eine Verallgemeinerung und Verbesserung der in bisherigen Suchen derselben Art verwendeten Methode dar. Sie erlaubt die Einbeziehung von Signalregionen mit destruktiver Interferenz in die Berechnung der Ausschlussgrenzen. Das W' Szenario ohne Interferenz kann für W' Massen bis zu 3.24 TeV ausgeschlossen werden (95% CL). Für die Szenarien mit Interferenz erhöhen sich die Ausschlussgrenzen auf 3.40 TeV beziehungsweise 3.55 TeV. Das Kontaktwechselwirkungsmodell kann bis zu einer Energieskala von 12.0 TeV ausgeschlossen werden. Die Ergebnisse dieser Arbeit werden zusammen mit den Ergebnissen einer ähnlichen Suche im Myon-Kanal veröffentlicht [1].

Abstract

In this thesis, a search for new physics in the electron + missing transverse energy channel is presented based on proton-proton collisions measured with the CMS detector at the LHC. The whole 2012 CMS dataset is used, which amounts to 20 fb^{-1} of collision data recorded at a center-of-mass energy of 8 TeV.

In many extensions of the Standard Model of particle physics, the existence of additional charged bosons called W' is predicted. For this search, a simple W' model with Standard Model-like properties is used as benchmark model with three different scenarios. Two of the three scenarios include interference effects between the W' and the Standard Model W . An additional interpretation is done in terms of an effective four fermion contact interaction theory.

No significant deviation from the Standard Model prediction is found. For the determination of exclusion limits on the parameters of the new theories, a multi-bin method was used which is a generalization of the methods used in previous searches and publications of the same kind. This allows the inclusion of signal regions with destructive interference into the limit calculation. The W' scenario without interference can be excluded at 95% CL for W' masses up to 3.24 TeV. In the interference scenarios, the exclusion limits extend to 3.40 TeV and 3.55 TeV respectively. The four fermion contact interaction can be excluded for energy scales up to 12.0 TeV. A similar analysis was performed in the muon channel and will be published together with the results from this thesis [1].

Contents

1. Introduction	1
I. Theoretical background	3
2. Standard Model of particle physics	3
3. Beyond Standard Model physics	9
3.1. Heavy charged gauge bosons	9
3.2. Helicity Non Conserving Contact Interaction	12
II. Experimental setup	13
4. Large Hadron Collider	13
5. Compact Muon Solenoid	15
6. Object reconstruction	21
6.1. Electron reconstruction	21
6.1.1. Overview	21
6.1.2. Electron clustering	23
6.1.3. Electron tracking	25
6.1.4. ECAL energy corrections	26
6.2. Missing Transverse Energy reconstruction	27
6.2.1. Particle Flow Missing Transverse Energy	28
6.2.2. Missing Transverse Energy corrections	29
III. Analysis	31
7. General signal signature	31
8. Dataset and Monte Carlo simulation of the Standard Model prediction	33
8.1. Dataset	33
8.2. Monte Carlo simulation of the Standard Model prediction	34
8.2.1. Standard Model processes	34
8.2.2. Pileup reweighting	37
9. Event selection and efficiencies	39
9.1. Event selection	39
9.1.1. Preselection	39
9.1.2. Electron ID	40

Contents

9.1.3. Signal specific selection	43
9.2. Selection efficiencies	43
9.2.1. Data - Monte Carlo scale factor	43
9.2.2. Background rejection efficiency	45
10. Estimation of the multi-jet contribution derived from data	47
10.1. Methodology	47
10.2. Measurement of the $t\bar{t}$ -ratio	48
10.3. Selection of the QCD template	52
10.4. Results	53
11. Parametrization of the $e + E_T^{\text{miss}}$ spectrum	55
12. Systematic uncertainties	59
13. Final M_T distribution	63
13.1. M_T resolution	63
13.2. Results	64
14. Signals	67
14.1. SSM W'	67
14.1.1. Scenario without interference	67
14.1.2. Scenarios with interference	69
14.2. HNC CI	71
IV. Statistical analysis	73
15. Introduction	73
15.1. Frequentist approach	73
15.2. Bayesian approach	75
16. SSM W' without interference	77
16.1. Single-bin method	77
16.2. Multi-bin method	79
16.3. Conclusion	81
17. SSM W' with interference	83
17.1. Bayesian multi-bin method	83
17.2. Hypothesis testing with CLs	84
17.3. Conclusion	85
18. HNC CI	87
19. Summary and conclusions	89
A. Appendix	91

1 Introduction

Elementary particle physics deals with the fundamental constituents of matter and their interactions. During the last century, a remarkable progress has been made in this field which lead to the formulation of the Standard Model of particle physics. The Standard Model is based on a small set of particles and forces and is able to describe a vast range of phenomena. It was tested over many years and is considered to be extremely successful. Despite its success, many open questions remain, which lead to the postulation of new theories trying to extend or replace the Standard Model.

Many of these new theories can be tested at the Large Hadron Collider, which is the current flagship experiment in the field of collider physics and allows to access previously unexplored energy regimes. In this thesis, the proton-proton (pp) collision data measured with the Compact Muon Solenoid (CMS) detector in 2012 is analyzed. CMS is one of the four main detectors installed at the LHC. The search channel of this analysis is the electron + missing transverse energy (E_T^{miss}) channel. E_T^{miss} can arise from particles which escape the detector without leaving a detectable signature (e.g. neutrinos).

The content of this thesis is divided into four parts. In the first part, an introduction to the theoretical background is given. This includes a summary of the Standard Model of particle physics as well as a presentation of the new physics models which are the baseline for this search. The second part explains the experimental setup and consists of a description of the CMS detector and the reconstruction processes involved during the reconstruction of a collision event. In the third part, the analysis is presented. As will be explained later, the assumptions on the properties of the new physics signal are rather general. Therefore, the analysis results can be interpreted in terms of many different new physics models. The last part is about the interpretation of the analysis results with statistical methods in terms of different new physics scenarios. In the end, a summary of the achieved results is given.

System of units

In particle physics, it is common practice to use a 'natural' system of units in which $\hbar = c = 1$. This leads to the same units for energy, mass and momentum (see equation 1.1).

$$[\text{energy}] = [\text{mass}] = [\text{momentum}] = 1\text{eV} \quad (1.1)$$

Energy, mass and momentum are usually expressed in units of electron Volt (eV). One eV is the energy a particle carrying the electrical charge of 1 e (elementary charge) gains, if it is accelerated with a Voltage of 1 V, and it corresponds to $1.602 \cdot 10^{-19}\text{J}$.

Coordinate system

The CMS detector is located in the north of the LHC collider ring. The origin of the CMS coordinate system is in the center of the detector at the nominal collision point. From this point, the x-axis points south towards the LHC center and the y-axis points

1. Introduction

vertically upwards. The z-axis is arranged along the beam axis and points into the west. These three axes form a Cartesian coordinate system. The x-y plane is usually called 'transverse plane'. By projecting the momentum of a particle onto the transverse plane, the transverse momentum can be obtained (similar for other variables). The azimuthal angle ϕ is measured in the transverse plane and is 0 for a vector pointing along the x-axis. The polar angle θ is defined with respect to the z axis. θ equals 0 for a vector pointing in the direction of the z-axis and π for a vector pointing in the opposite direction. Instead of θ , the pseudorapidity η as defined in equation 1.2 can be used.

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \quad (1.2)$$

The distance between two trajectories can be expressed with help of the variable ΔR which is defined in equation 1.3.

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (1.3)$$

Part I.

Theoretical background

2 Standard Model of particle physics

The Standard Model of particle physics is a theory describing the fundamental constituents of matter and the interactions between them. Being able to describe and predict a vast range of different phenomena, it is an extremely successful theory and has been tested over many decades. There are four known forces in the universe today: the electromagnetic, the strong, the weak and the gravitational force. From these four, the first three are included in the Standard Model. The electromagnetic and the weak force could be combined into the single framework of the electroweak force. Despite many efforts, no convincing way has been found to include the gravitational force or to combine strong and electroweak force. For the distances and energies relevant in particle physics, gravitation can be neglected since it is much weaker than the other three forces.

This chapter should serve as a general introduction into the concepts and phenomenology of the Standard Model. It includes a presentation of the fundamental constituents and forces, the concept of quantum field theories as well as a short summary of the cross-section calculation at a hadron collider. The chapter concludes with a list of flaws and remaining questions of the Standard Model. A detailed overview of the Standard Model and its history can be found in [2]. More information about the concept of quantum field theories and their application to the Standard Model is given in [3]. Unless otherwise stated, the content of this chapter is taken from those two sources.

Fundamental particles

The Standard Model is based on a few elementary particles which are assumed to be point-like. Beside of its mass, each particle can be characterized with a set of quantum numbers. A very important quantum number is the spin of the particle. Based on it, the particles are divided into fermions on the one side and bosons on the other side. The other quantum numbers are related to the interaction of the particle with the fundamental forces (e.g. electric charge \leftrightarrow electromagnetic force).

Fermions have a half-integer spin and can be further divided into quarks and leptons. Quarks participate in all three fundamental interactions included in the Standard Model. They have an electric charge of $+2/3$ or $-1/3$ (expressed in units of the elementary charge e). Leptons do not participate in the strong interaction. Here, the distinction between charged leptons with an electric charge of -1 participating in electromagnetic and weak interactions and uncharged leptons (electric charge of 0) can be made. The latter ones are called neutrinos and do only take part in weak interactions. Over the years, six quarks and six leptons have been discovered, which can be organized in three 'families' or 'generations'. The corresponding particles of each generation have the same quantum numbers but differ in their mass. Each generation consists of two quarks, a charged lepton (e, μ, τ) and a corresponding neutrino (ν_e, ν_μ, ν_τ). One of the two quarks is an 'up'-type quark (u, c, t) with an electric charge of $+2/3$ and the other one a 'down'-type quark (d, s, b) having an electric charge of $-1/3$. The Standard Model provides no explanation or prediction for the number of generations. With the exception of the neutrinos, all particles of the higher generations are unstable and decay into particles

2. Standard Model of particle physics

			bosons		
quarks	m = 2.3 MeV	m = 1.28 GeV	m = 174 GeV	m = 0 eV	m ≈ 125 GeV (?)
	u up-quark 1/2 +2/3	c charm-quark 1/2 +2/3	t top-quark 1/2 +2/3	g gluon 1 0	H Higgs boson 0 0
	m = 4.8 MeV	m = 95 MeV	m = 4.18 GeV	m = 0 eV	
	d down-quark 1/2 -1/3	s strange-quark 1/2 -1/3	b bottom-quark 1/2 -1/3	γ photon 1 0	
	m = 0 eV	m = 0 eV	m = 0 eV	m = 80.4 GeV	
	ν_e electron neutrino 1/2 0	ν_μ muon neutrino 1/2 0	ν_τ tau neutrino 1/2 0	W W boson 1 ±1	
leptons	m = 511 keV	m = 106 MeV	m = 1.78 GeV	m = 91.2 GeV	
	e electron 1/2 -1	μ muon 1/2 -1	τ tau 1/2 -1	Z Z boson 1 0	

Legend:

mass	spin
symbol	charge
name	

Figure 2.1. Fundamental particles of the Standard Model. Neutrino masses are approximated to zero. All values are taken from [7].

of the first generation via weak interaction. Thus, all observable matter is mainly built from particles of the first generation. In the following, the neutrinos are approximated to be stable and massless even though experimental data indicates the presence of an oscillation mechanism between the three neutrino generations which leads to non-zero masses (see [4] for details).

Bosons are particles with an integer spin. In the Standard Model, the fundamental forces are mediated through the exchange of 'force-carrier' bosons with a spin of 1. The force-carrier or the electromagnetic force is the photon. For the weak interaction, the exchange bosons are the two W^\pm bosons as well as the Z boson. The strong force is mediated via eight different gluons. While the photon and the gluons are massless, the W^\pm and Z bosons are found to be massive. The mechanism to introduce masses to the Standard Model particles is called Higgs mechanism [5, 6] and predicts the existence of a scalar boson called the Higgs boson. A summary of all Standard Model particles is given in figure 2.1. For each particle, an anti-particle which has the same mass but opposite charges (e.g the positron has an electric charge of +1) exists. In some cases, particle and anti-particle are identical (e.g. photon).

Theory of strong interactions

Quantum chromodynamics (QCD) [8, 9] is the theory of strong interactions based on the symmetry group $SU(3)$. As stated before, only quarks and gluons are involved in the strong interaction. Each quark exists in three different colors (usually dubbed red, green and blue) forming a color triplet. The color of a quark corresponds to the charge of the strong interaction. Anti-quarks thus carry opposite charges called anti-colors (anti-red, anti-green and anti-blue). The strength of the color charge is the same for all three colors.

In contrast to the electrically neutral photon, the gluons carry a combination of color and anti-color charge. Thus, a self-interaction between gluons is possible. The QCD has two special features called 'confinement' and 'asymptotic freedom'. Confinement means, that it is impossible to separate color charges. During the attempt to separate a quark anti-quark pair, the interaction energy will increase. At a certain distance a new quark anti-quark pair is generated such that two colorless quark anti-quark pairs are formed. Therefore, no free quarks or gluons exist. Instead, quarks appear in uncolored bound states as mesons (e.g pion) or baryons (e.g. proton or neutron). Mesons are combinations of quark and anti-quark pairs. Baryons are formed by combining three quarks or anti-quarks. Here, it is important to note that the combination of all three color charges can also form a colorless state. Similar to the electromagnetic and weak interaction, the coupling strength α_s of the QCD is not a constant, but a function of the momentum transfer Q^2 of the interaction. For large Q^2 , α_s will decrease and vanish asymptotically (\rightarrow asymptotic freedom). Thus, for large momentum transfers and small distances, quarks behave similar to free particles.

Theory of electro-weak interactions

The theory which describes electromagnetic interactions is called quantum electrodynamics (QED) [10] and is based on the symmetry group $U(1)_{\text{QED}}$. The electromagnetic force is mediated through the exchange of photons, which are massless, stable and uncharged bosons. Thus the force has an infinite range. Photons couple to the electric charge of particles. In the electromagnetic interactions, all quantum numbers are conserved.

The force carrier of weak interactions are the electrically charged W^\pm and the electrically neutral Z bosons. Due to the large mass of W^\pm and Z , the range of the weak interaction is very small. In weak interactions some unique features occur which are not present in QED and QCD. First of all, the charged current (W^\pm) can change the 'flavor' of particles (e.g. up-type quark \rightarrow down-type quark or electron \rightarrow electron neutrino). For leptons, the Standard Model in its simplest form (without the inclusion of neutrino oscillations (see [4] for details)) restricts these transitions to be within the same family (lepton number conservation). Due to an effect called quark-mixing, this is not true for quarks. The mixing of quarks is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix which is not predicted by the Standard Model and has to be determined via experiments (see [7] for details). Charged weak currents do only couple to left handed particles or right-handed anti-particles and do thus lead to maximal parity violation. Weak neutral currents are not flavor-changing but do also lead to a (smaller) violation of parity. The 'charge' of the weak interaction is the weak isospin T and its third component T^3 . Left-handed fermions are arranged in isospin doublets with $T^3 = \pm 1/2$, while right handed fermions are isospin singlets with $T^3 = 0$.

The combination of electromagnetic and weak interactions into a unified model is done in terms of the Glashow-Salam-Weinberg (GSW) model [11–13]. It is based on the symmetry group $SU(2)_L \otimes U(1)_Y$. The observable bosons (γ, Z, W^\pm) are linear combinations of the 4 generators ($W^{1,2,3}, B$) as shown in equation 2.1.

$$\begin{aligned}
 W^+ &= \frac{1}{\sqrt{2}} (W^1 + iW^2) & W^- &= \frac{1}{\sqrt{2}} (W^1 - iW^2) \\
 Z &= -\sin(\theta_w) B + \cos(\theta_w) W^3 & \gamma &= \cos(\theta_w) B + \sin(\theta_w) W^3
 \end{aligned}
 \tag{2.1}$$

2. Standard Model of particle physics

Here, θ_w is the electro-weak mixing angle which has to be measured. In the electro-weak theory the hypercharge Y is introduced which is the charge for the B coupling. It is related to the electric charge and the weak isospin via $Y/2 = Q - T_3$.

As stated before, the W^\pm and Z bosons are not massless. The Standard Model as presented so far, has no mechanism to introduce particle masses into the theory. On top of that, the 'manual' addition of mass terms to the theory breaks the local gauge invariance, which is one of the basic principles in quantum field theories. To include massive particles in the Standard Model, the Higgs mechanism is used. It is based on the principle of 'spontaneous symmetry breaking' and predicts the existence of a scalar particle called the Higgs boson. In the pre-LHC era, the Higgs boson was the only missing part in the set of fundamental particles of the Standard Model. At the LHC, a suitable candidate has been found [14, 15], which seems to have the properties of a Standard Model Higgs boson.

Formalism, principles and perturbation theory

The Standard Model is a renormalizable quantum field theory which is based on a local symmetry and it is expressed in terms of the Lagrange formalism. Here, local symmetry means that the Lagrangian of the theory is invariant under local (depending on space-time coordinates) transformations. This local gauge invariance is an essential principle in formulating the theory, since it allows to derive the dynamics of the system when starting from a free theory without interactions. Renormalization is a mathematical procedure to deal with infinite quantities occurring in the theory.

When calculating the transition probability for a given initial and final state (e.g. $q\bar{q} \rightarrow e^+e^-$), an infinite number of possible processes for this transition exists which have to be considered. Two possibilities for the chosen example process are visualized as Feynman diagrams in figure 2.2 (for an introduction of Feynman diagrams see [2]). In the calculation, each of the vertices (visualized as dots) leads to an additional factor in the transition amplitude. In the QED processes from figure 2.2, the lower diagram would have an additional factor of α (fine structure constant). Since α is small ($\approx 1/137$), the impact of the lower process to the total transition amplitude is suppressed compared to the impact of the upper one. Other diagrams with more vertices suffer from an even stronger suppression. Thus, the calculation of all the diagrams is not necessary in order to get a useful result for the transition probability. Instead, calculations in the Standard Model are done with help of perturbation series ordered in powers of the coupling constant. Processes from the lowest order are called Leading Order (LO) processes. The diagrams from the next higher order of the perturbation series are called Next to Leading Order (NLO) diagrams (and the next higher order \rightarrow NNLO etc.).

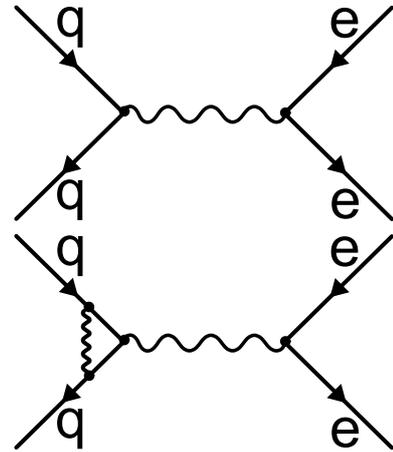


Figure 2.2. Two Feynman graphs for the process $q\bar{q} \rightarrow e^+e^-$.

Cross-sections at hadron-colliders

The concept of cross-sections is a fundamental ingredient in particle physics. Given a fixed target and an incoming particle beam, the cross-section can be seen as the transition

MSTW 2008 NLO PDFs (68% C.L.)

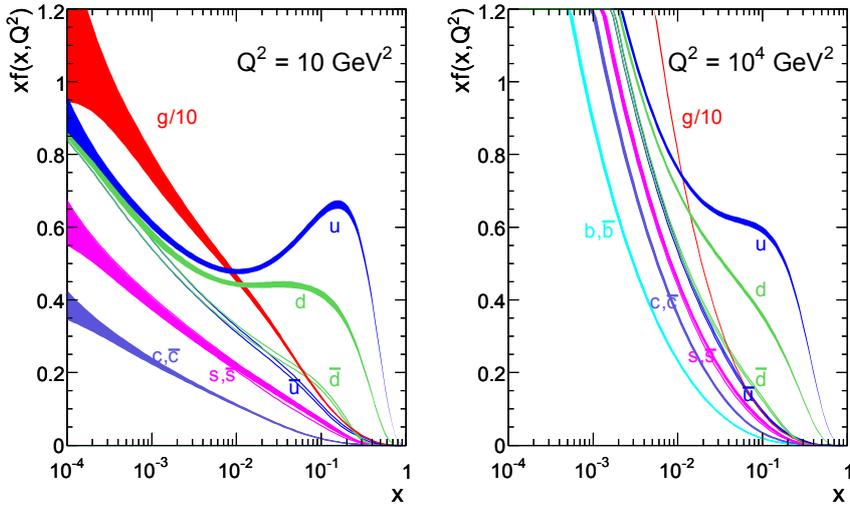


Figure 2.3. PDF functions from the MSTW2008 PDF set in NLO for two different values of Q^2 . Figure from [16].

rate per unit flux for a single target particle. The initial and final state of the transition have to be specified for the calculation. Cross-sections have the dimension $(\text{length})^2$ and are usually expressed in unit of 'barns' ($1 \text{ b} = 10^{-24} \text{ cm}^2$).

At the LHC, proton-proton collisions are measured. The main constituents of the proton are three quarks (two up-quarks, one down-quark), which are called valence-quarks. Beside the valence-quarks, gluons mediating the strong force between the quarks are present. Temporary conversion from gluons to quark anti-quark pairs lead to the presence of additional quarks and anti-quarks in the proton. Each of the constituents (or 'partons') will carry a certain fraction x of the protons momentum, and each constituent could take part in a hard scattering process of a proton-proton collision.

In order to calculate the cross-section for a specific process at a hadron-collider, the process is split into two parts. Everything which is related to the distribution of the proton momentum onto the different constituents is summarized in 'Parton Density Functions' (PDF). The PDF $f_j(x, Q^2)$ gives the probability to find a parton j with a momentum fraction x in a proton when probed at an energy scale of Q^2 . These PDF functions can not be calculated, but have to be extracted from data. More information about PDFs is given in chapter 12. Examples of PDFs can be seen in figure 2.3 for two different values of Q^2 . Once two partons (j,k) and their momentum fractions (x_1, x_2) are given, the cross-section for this specific process ($\hat{\sigma}_{j,k}$) can be calculated using quantum field theory. In equation 2.2, the general formula for the total cross-section σ of a process at a hadron collider is given. All possible parton pairs (j,k) have to be considered in the sum.

$$\sigma = \sum_{j,k} \int_0^1 \int_0^1 dx_1 dx_2 f_j(x_1, Q^2) f_k(x_2, Q^2) \hat{\sigma}_{j,k}(x_1, x_2, Q^2) \quad (2.2)$$

2. Standard Model of particle physics

Limitations

Even though the Standard Model is very successful, open questions remain. Some of the flaws and open questions are listed below.

- **Fine-tuning problem:** Loop corrections to the Higgs mass depend on a cut-off parameter up to which the Standard Model is believed to be valid. If this parameter is set to the Planck scale, the bare Higgs mass and the loop corrections have to be of the order of $\mathcal{O}(10^{19} \text{ GeV})$, match in the first 17 digits and nearly cancel each other in order to obtain an observed mass of $\mathcal{O}(100 \text{ GeV})$. This is considered to be unnatural.
- **Dark matter:** Several astronomical observations indicate the presence of dark matter, which constitutes a large part of the matter in the universe. The Standard Model provides no candidate which could explain the observations.
- **Matter anti-matter asymmetry:** The present universe consists almost exclusively of matter, while the big bang is thought to have created an equal amount of matter and anti-matter. A possible explanation of this asymmetry is CP violation. However, the CP violation in the Standard Model is not strong enough to explain the observed asymmetry.
- **Parity violation:** There is no explanation for the origin of parity violation in weak interactions. Its inclusion into the theory is done to match the observations.
- **Unification:** To unify the three fundamental forces of the Standard Model in one Grand Unified Theory (GUT), their coupling strengths have to become equal at a certain energy scale. This is not the case in the Standard Model.

3 Beyond Standard Model physics

A large range of Beyond the Standard Model (BSM) physics models exists, which try to address the open questions and issues of the Standard Model. In this analysis, a search for new physics with $e + E_{\text{T}}^{\text{miss}}$ final states is performed. A prominent example which would lead to this final state is the existence of additional heavy charged bosons. These bosons are usually called W' and are predicted in many models. In this chapter, two different models predicting the existence of a W' are briefly introduced. Then a benchmark model for the W' search is presented, which allows to focus on the general properties of a W' without having to rely on the details of a specific model. The chapter concludes with the presentation of an effective four fermion contact interaction theory, which could be a sign for fermion compositeness. Additional models can be found in [1].

3.1. Heavy charged gauge bosons

Left-Right symmetric models

In Left-Right (LR) symmetric models, the symmetry group of the Standard Model is extended, leading to a Lagrangian which is invariant under the exchange of left and right-handed particles. The symmetry has to be broken at low energies in order to reproduce the observed parity violation of charged currents. A simple way to obtain a LR-symmetric model is the addition of a $SU(2)_R$ group coupling only to right handed fermions. Similar as for the $SU(2)_L$ group of the Standard Model, an additional weak isospin T_R is introduced which is the charge related to the $SU(2)_R$. The total symmetry group for the electroweak sector is then $SU(2)_L \otimes SU(2)_R \otimes U(1)_{\tilde{Y}}$, where \tilde{Y} is a modified hypercharge. Hypercharge and weak isospins are related to the electrical charge via $Q = T_L^3 + T_R^3 + \tilde{Y}/2$. This model contains four electrically charged bosons W_L^\pm and W_R^\pm . To obtain the observed electro-weak Standard Model theory in the low energy regime, a modified version of the Higgs mechanism is needed. Depending on the details of this mechanism, this could lead to the postulation of additional Higgs bosons. The observable charged bosons are then obtained by a mixing of W_L and W_R as shown in equation 3.1.

$$\begin{aligned} W &= \cos(\zeta)W_L + \sin(\zeta)W_R \\ W' &= -\sin(\zeta)W_L + \cos(\zeta)W_R \end{aligned} \tag{3.1}$$

Here, ζ is a parameter which describes the mixing between the 'left' and 'right' gauge groups. Since the W corresponds to the Standard Model W boson, ζ has to be small in order to reproduce the observations (W only couples to left-handed fermions). The parameters of the model can be tuned such that the mass of the W' is large enough to be outside of the region excluded by experimental searches. In LR-symmetric models, right-handed neutrinos can naturally be included since they can couple to the W_R^\pm and are not 'sterile'. The presence of right-handed neutrinos allows the introduction of neutrino masses. With help of a 'see-saw' mechanism (see [17] for details), the masses of left

3. Beyond Standard Model physics

handed neutrinos can be kept very small to agree with the experimental observations. The information given in this section are based on [18, 19], where additional information can be found.

Universal Extra Dimensions

W' bosons are also predicted in various models postulating the existence of extra dimensions. Based on [20], an Universal Extra Dimension (UED) model will be briefly presented here. In UED models, the existence of one additional spatial dimension is predicted. In contrast to the other four dimensions, the fifth dimension is compact. All Standard Model particles are seen as the lowest excitation mode of Kaluza-Klein (KK) states (e.g. q_n , ℓ_n , W_n , etc. with $n=0$ for Standard Model particles). Thus the W' can be interpreted as higher KK mode of the W . Since only even KK states ($n=2,4,..$) are allowed to couple to the Standard Model particles and the W_n mass increases with n as shown in equation 3.2, the W_2 mode is the W' candidate of interest. Excitation modes with $n>2$ are not expected to be visible at present LHC conditions.

$$m_{W_n}^2 = m_W^2 + \left(\frac{n}{R}\right)^2 \quad (3.2)$$

The model has two parameters, which are called R and μ . R is proportional to the size of the extra dimension and μ controls the coupling strength to Standard Model particles.

Benchmark models

A priori, it is impossible to predict which of the many models containing new gauge bosons could be realized in nature. Instead of specializing on one specific model, it is utile and reasonable to choose a more generic approach in which the general properties of the different models are summarized without relying on their details. This can be done with help of a benchmark model. For additional gauge bosons, such a benchmark model is proposed in [21] and called the 'reference model'. Here, the existence of new gauge bosons W' and Z' is postulated without further specifying their origin. The new W' is supposed to have the same couplings as the Standard Model W meaning that it can decay into $q\bar{q}'$, $\ell\nu$ and also into WZ . Without further modification, the decay into WZ is the dominant decay channel since its decay width is proportional to $m_{W'}^5$. This would also lead to a large increase of the W' decay width. In many W' models (e.g. extended gauge model like the LR-symmetric model), the coupling to WZ is suppressed. Motivated from the expected suppression in extended gauge models a suppression factor of $\xi = m_W^2/m_{W'}^2$ can be introduced. The reference models allows to calculate W' cross-sections, decay widths and branching fractions.

Sequential Standard Model

The benchmark model used for W' searches in the $W' \rightarrow \ell\nu$ channels is called Sequential Standard Model (SSM) and is based on the reference model. Similar to the reference model, the W' is a heavy analogue of the Standard Model W with similar decay channels. If the W' mass is large enough, $W' \rightarrow t\bar{b}$ is an additional possible decay. The decay into WZ is assumed to be heavily suppressed. In the case where $m_{W'} \gg m_t$ and neglecting all fermion masses, the decay width of the W' to fermions is given by equation 3.3.¹

¹ $\Gamma_{W^{(\prime)} \rightarrow f\bar{f}'}$ is defined as the sum of the individual decays widths to all possible fermion pairs.

$$\Gamma_{W' \rightarrow f\bar{f}'} = \frac{4}{3} \frac{m_{W'}}{m_W} \Gamma_{W \rightarrow f\bar{f}'} \quad (3.3)$$

The factor $4/3$ accounts for the additional decay channel into $t\bar{b}$. This leads to a branching fraction around 8% for the $e\nu$ channel, which is the interesting channel for this analysis. The reference model does not specify whether the W' couples to right or left-handed fermions. In general both scenarios are possible. A general formulation of the Lagrangian for a W -like coupling to fermions is given in equation 3.4.

$$\mathcal{L}_{W_i f\bar{f}'} = \frac{1}{\sqrt{2}} V_{ff'} C_i^{\ell,q} \bar{f} \gamma_\mu \frac{1}{2} (1 - h_i \gamma_5) f' W_i^\mu + h.c. \quad (3.4)$$

Here, $V_{ff'}$ is the mixing matrix (in case of the Standard Model W : CKM matrix for quarks, unit matrix for leptons), $C_i^{\ell,q}$ is the coupling strength of the boson W_i to leptons (ℓ) or quarks (q) and $1/2 (1 - h_i \gamma_5)$ the projection operator for the left and right-handed component (with $P_L \rightarrow h_i = 1$ and $P_R \rightarrow h_i = -1$). The case $W_i = W$ with $h_i = 1$ and $C_i^\ell = C_i^q = g$ (g is the weak coupling constant) reproduces the Lagrangian of the Standard Model W coupling to fermions.

In general the cross-section of the process $q\bar{q}' \rightarrow W/W' \rightarrow \ell\nu$ will be proportional to the squared absolute value of the matrix element \mathcal{M} . $|\mathcal{M}|^2$ can be expanded as shown in equation 3.5.

$$|\mathcal{M}|^2 = |\mathcal{M}_W|^2 + |\mathcal{M}_{W'}|^2 + 2\mathcal{R}e(\mathcal{M}_W \mathcal{M}_{W'}^*) \quad (3.5)$$

A formula for the calculation of $\mathcal{M}_W \mathcal{M}_{W'}^*$ for one specific quark flavor combination (averaged over the initial spin configurations and accounting for the fact that the initial quark and anti-quark have to form a color neutral state) is given in equation 3.6. In this formula, the widths of the bosons have been neglected and z is the cosine of the angle between up-type quark and neutrino in the center-of-mass frame. It was further assumed, that the quark mixing matrix is the same for W and W' .

$$\mathcal{M}_W \mathcal{M}_{W'}^* = \frac{1}{192} |V_{qq'}|^2 \frac{C_W^\ell C_{W'}^{\ell*} C_W^q C_{W'}^{q*}}{(\hat{s} - m_W^2)(\hat{s} - m_{W'}^2)} \hat{s}^2 ((1 + h_W h_{W'})^2 (1 + z^2) + 2(h_W + h_{W'})^2 z) \quad (3.6)$$

As can be seen from equation 3.6, setting $h_{W'} = -1$ leads to a vanishing of the interference term (since $h_W = 1$). This corresponds to a W' which only couples to right-handed particles and left-handed anti-particles and is the most commonly used scenario for W' searches in the $\ell\nu$ channels. Here, the production of a Standard Model W and a W' are completely independent. This scenario is also used in this search and will be denoted SSM W' scenario in the following.

On top of the scenario without interference, two additional scenarios with interference are considered. In general the interference between W and W' can be constructive or destructive depending on the sign of $2\mathcal{R}e(\mathcal{M}_W \mathcal{M}_{W'}^*)$. Setting $h_{W'}$ to 1, leads to an interference term as given in equation 3.7.

$$2\mathcal{R}e(\mathcal{M}_W \mathcal{M}_{W'}^*) = 2\mathcal{R}e \left(\frac{g^2}{48} |V_{qq'}|^2 \frac{C_{W'}^\ell C_{W'}^q}{(\hat{s} - m_W^2)(\hat{s} - m_{W'}^2)} \hat{s}^2 ((1 + z^2) + 2z) \right) \quad (3.7)$$

In general, the coupling strengths of W and W' to leptons and quarks could be different. For this analysis, it is assumed that they have the same module ($|C_{W'}^{\ell,q}| = |C_W^{\ell,q}| = g$).

3. Beyond Standard Model physics

If $C_{W'}^\ell = C_{W'}^q = g$, the interference will be destructive when $\sqrt{\hat{s}} \in (m_W, m_{W'})$ and constructive for $\sqrt{\hat{s}}$ larger than $m_{W'}$ (see figure 14.5). This scenario will be called the same-sign SSM scenario (SSMS). In contrast to the SSMS, an opposite sign of the W' couplings ($C_{W'}^\ell \cdot C_{W'}^q = -g^2$) will lead to a constructive interference between $m_W - m_{W'}$ and a destructive interference beyond $m_{W'}$ (see figure 14.5). This scenario will be called the opposite-sign SSM scenario (SSMO). All formulas of this section are based on [22]. This reference also contains a full calculation of the differential $q\bar{q}' \rightarrow W/W' \rightarrow \ell\nu$ cross-section.

There are many exclusion limits for the existence of a W' from different experiments. Prior to the analysis of the 2012 LHC data, the most stringent exclusion limits for a SSM W' without interference effects were set by CMS ($m_{W'} > 2.5$ TeV [23]) and ATLAS ($m_{W'} > 2.55$ TeV [24]). Limits on a SSM W' with interference effects were set by CMS (between 2.43 TeV and 2.63 TeV depending on the scenario [23]). In the pre-LHC era the best limits (around 1.1 TeV) were obtained by CDF ([25]) and DO ([26, 27]). The strongest limits from indirect searches arise from measurements of the $K_L - K_S$ mass difference, where the existence of a right handed W' (as in the LR-symmetric model) can be excluded up to masses of 2.5 TeV. A list of W' limits from indirect searches can be found in [7]. In general, limits from indirect searches depend strongly on the assumed W' model.

3.2. Helicity Non Conserving Contact Interaction

The presence of mass hierarchies in the lepton and quark sector motivates the assumption, that quarks and leptons could be composite particles which are build from fundamental constituents. These constituents are called 'preons' in the literature [28]. There are many different models proposing a possible composite structure for leptons and quarks. A fundamental variable in such models is the energy scale Λ characterizing the strength of the preon bonding. If the center-of-mass energy is of the order of Λ , it should be possible to observe excited quarks and leptons. In the case $\sqrt{s} \ll \Lambda$, the compositeness could manifest itself as four fermion Contact Interaction (CI). Another possible origin for the appearance of a four fermion CI could be the exchange of a very heavy particle (c.f. Fermi's theory of weak interactions). Based on a search of the same kind in the $\mu + E_T^{\text{miss}}$ channel [29], the Helicity Non Conserving Contact Interaction (HNC CI) model [30] is chosen as benchmark CI model. The HNC CI is an effective theory which postulates a CI between two quarks and two leptons including processes like $q\bar{q}' \rightarrow \bar{\ell}\nu$. In this model, the interaction is only allowed if quark and anti-quark are right handed and both leptons left-handed or vice-versa. Therefore it is helicity non conserving. A direct consequence of this behavior is, that there is no interference with the Standard Model W , since the CI always includes a right-handed fermion or a left-handed anti-fermion. The differential cross-section of the HCN CI model is given in equation 3.8.

$$\frac{d\sigma}{d\cos(\theta)} = \frac{\pi\hat{s}}{12\Lambda^4} \quad (3.8)$$

A detailed overview of the HNC CI model can be found in [29]. The only existing limit for this specific model in the pre-LHC era was set by the CDF collaboration excluding a HNC CI in the electron channel for $\Lambda < 2.81$ TeV [31]. Prior to this analysis, CMS determined limits for the electron channel (10.5 TeV) and the muon channel (8.8 TeV) using a subset of the 2012 dataset [32].

Part II.

Experimental setup

4 Large Hadron Collider

The Large Hadron Collider (LHC) [33] is located at CERN near Geneva. It is installed in the 26.7 km long tunnel drilled for the LEP [34] accelerator and is designed to collide proton bunches at a maximal center-of-mass energy (\sqrt{s}) of 14 TeV as well as lead ion beams at an energy of 2.8 TeV per nucleon. From the final decision to build the LHC in 1994, it took 14 years of construction and installation until in summer 2008 proton beams circulated in the LHC for the first time. Due to a technical incident, the start of pp collisions was delayed until the end of 2009. After two years of pp collisions at $\sqrt{s} = 7$ TeV in 2010 and 2011, the center-of-mass energy was increased to 8 TeV in 2012.

One of the key parts of the LHC machine are the superconducting magnets which are used to guide the protons' trajectories. The magnets are operated at 1.9 K, and can produce magnetic fields up to 8.33 T. The cooling system needed to preserve the superconductivity is based on superfluid helium. For proton-proton collisions, the beams can not use the same beam pipe. The beam traveling clockwise through the LHC needs a magnetic field which points in the opposite direction than the magnetic field used to guide the beam running counter-clockwise. This, combined with the fact that the LHC tunnel is too small for the installation of two separate proton rings, led to the development of a special magnet design called "twin-bore" design. In this design the magnets can host two beampipes at the same time. Hadrons which enter the LHC are pre-accelerated with the help of different smaller accelerators located at CERN. An overview of the CERN accelerator complex is shown in figure 4.1.

The LHC hosts four main experiments which are called ATLAS (A Toroidal LHC Apparatus) [35], CMS (Compact Muon Solenoid) [36], ALICE (A Large Ion Collider Experiment) [37] and LHCb (Large Hadron Collider beauty experiment) [38]. ATLAS and CMS are multipurpose detectors which are constructed to investigate a broad spectrum of possible new physics. ALICE is optimized for lead-lead collisions. A main task of ALICE is to study the properties of the quark-gluon plasma. The LHCb experiment is specialized to study collision events involving b-quarks. One of its main physics tasks is to study the differences between matter and anti-matter in order to explain the asymmetry of matter and anti-matter in the present universe. The locations of the four main experiments at the LHC also shown in figure 4.1

The rate of events generated in pp beam collisions can be calculated according to equation 4.1.

$$\dot{N}_{ev} = L_{inst} \cdot \sigma \quad (4.1)$$

Here, L_{inst} is the instantaneous luminosity which depends on LHC machine parameters and σ is the total proton-proton cross-section at the given center-of-mass energy. The peak instantaneous luminosity achieved in 2012 was about $7.7 \times 10^{-33} \text{cm}^{-2}\text{s}^{-1}$ [40] for stable beams. A detailed description of the LHC design can be found in [33]. Unless stated otherwise, all information was taken from that reference.

4. Large Hadron Collider

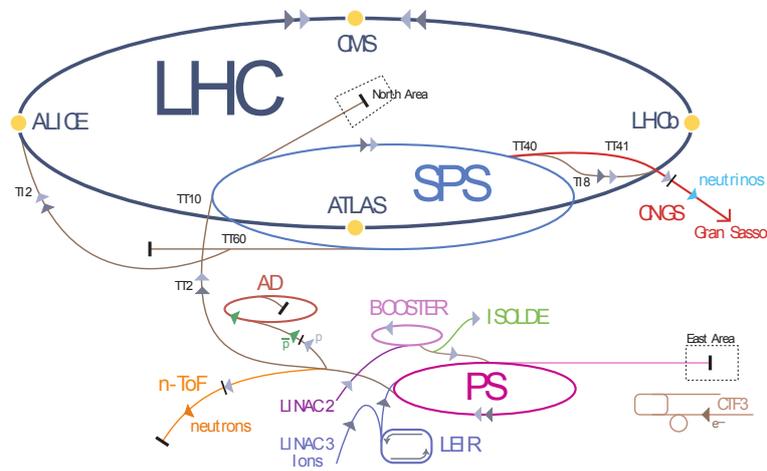


Figure 4.1. Overview of the accelerator complex at CERN. Shown is the LHC and its four main experiments ATLAS, CMS, ALICE and LHCb together with the pre-accelerator chain (LINAC2 \rightarrow Booster \rightarrow PS \rightarrow SPS). Also shown are some experiments and accelerators which are not part of the LHC. Figure taken from [39].

5 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a multipurpose detector operating at the LHC. One of the main physics tasks is the investigation of the electroweak symmetry breaking (e.g. the search for a Higgs boson). On top of that, the CMS detector offers the possibility to test a large set of alternatives to the Standard Model including new particles and forces at the TeV scale. A sketch of the CMS detector is shown in figure 5.1. It has a length of 21.6 m, a diameter of 14.6 m and weighs 12500 t. Starting from the innermost part, the structure of the CMS detector can be subdivided into the following components:

- **Tracker:** Measures the tracks and momentum of charged particles.
- **Electromagnetic Calorimeter (ECAL):** Measures the energy deposits from electrons and photons which are absorbed in the ECAL. Also measures energy deposits of hadrons on their way to the HCAL.
- **Hadronic Calorimeter (HCAL):** Measures the energy deposit of hadrons which are absorbed in the HCAL.
- **Superconducting Solenoid:** Creates a magnetic field which is needed for the momentum measurement of charged particles in the tracker.
- **Muon System:** Measures the momentum and tracks of muons which escape the detector.

In this chapter, each of the components will be briefly explained. This chapter is based on a detailed description of the CMS detector which can be found in [36].

Tracker

The tracker performs several discrete position measurements of charged particles emerging from proton-proton collisions. These measurements can be combined to reconstruct the trajectories of charged particles. The LHC imposes very challenging conditions to the design and operation of a tracking system. Given the design luminosity ($10^{34} \text{ cm}^{-2}\text{s}^{-1}$), the average number of particles per proton-proton bunch crossing is around 1000 [36]. This in combination with the small time interval between each bunch crossing (design value: 25 ns) requires a tracking system with a fast response and a high granularity. These are crucial in order to provide a good spatial and time resolution allowing to identify trajectories and assign them to the correct bunch crossing. However, a compromise has to be found between the performance of the tracking system and the amount of material which has to be installed inside the solenoid. The latter one should be as small as possible, since it will lead to multiple scattering, photon conversion, bremsstrahlung and nuclear interactions. The tracker is the detector system which has the smallest distance to the interaction point. Thus, it has to withstand severe radiation damage imposed by the high particle fluxes at the LHC.

5. Compact Muon Solenoid

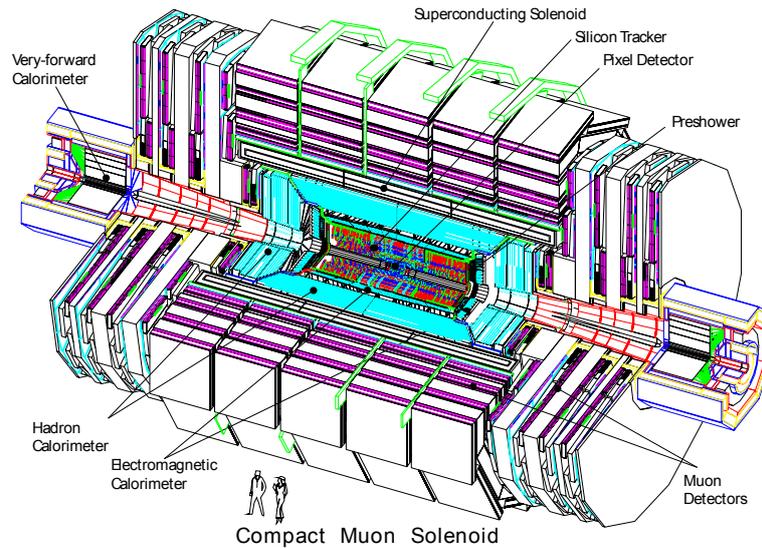


Figure 5.1. Overview of the CMS detector. Figure taken from [36].

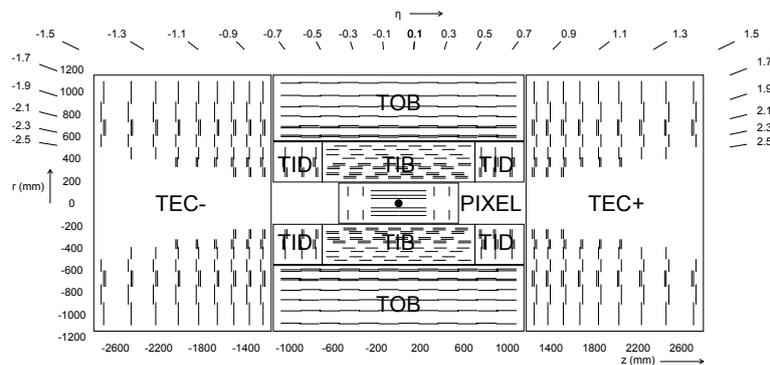


Figure 5.2. Scheme of the tracker design. Every line indicates a tracking module. Double lines represent elements with two tracking modules mounted back to back. Figure taken from [36].

The CMS tracker uses silicon based detector technology and is the largest tracker of this kind ever built. Constructing a silicon tracker of this size was only possible due to key innovations in the module assembly, sensor fabrication and read-out electronics leading to a large reduction of costs and an improved signal-to-noise performance. The tracker has a length of 5.8 m and a diameter of 2.5 m. It provides a coverage up to a pseudorapidity of $|\eta| < 2.5$. A sketch of the design is shown in figure 5.2.

The tracker is based on silicon pixel and strip detectors and can be divided into a barrel part and two tracker endcaps. The pixel detectors form the innermost part of the tracker and are arranged in 3 cylindrical layers between 4.4 cm and 10.2 cm from the beam-line. On each side of the barrel part, two additional disks of pixel detectors are installed. Each of the 66 million pixels has a size of $100 \times 150 \mu\text{m}^2$ in r - ϕ and z respectively. The fine granularity and the similar track resolution in r - ϕ and z direction is crucial for the reconstruction of primary and secondary vertices. The strip detectors in the barrel part are arranged in three different subsystems. The inner tracker part is divided into the Tracker Inner Barrel (TIB) and the 3 Tracker Inner Disks (TID) at

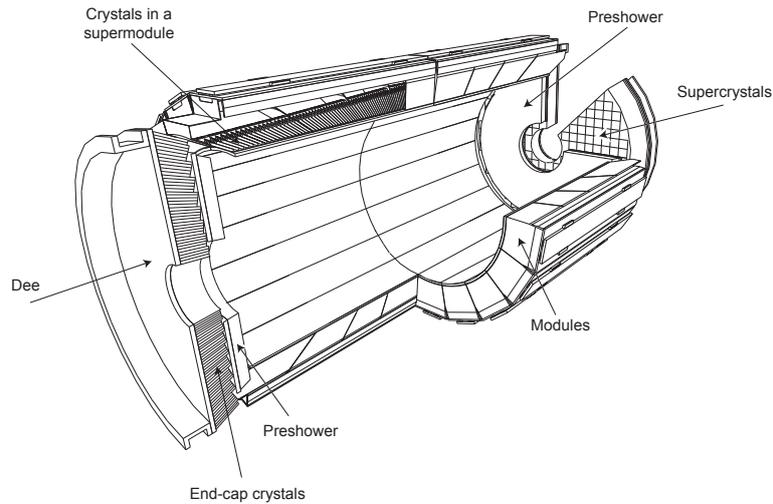


Figure 5.3. Scheme of the ECAL showing the arrangement of the crystals. Figure taken from [36].

each end. The thickness of the silicon micro-strips is $320\ \mu\text{m}$ while the surface varies between $80\ \mu\text{m}$ and $120\ \mu\text{m}$ in the TIB and $100\ \mu\text{m}$ and $141\ \mu\text{m}$ in the TID. Enclosing TIB and TID, the Tracker Outer Barrel (TOB) provides 6 additional layers of silicon-strip sensors. The barrel part of the tracker is complemented with two Tracker EndCaps (TEC). Each endcap is made out of 9 disks with up to 7 rings of micro-strip detectors. The size of the micro-strip sensors in TOB and TEC differ from the ones utilized in TIB and TID. More details can be found in [36].

ECAL

Similar to the tracker, the ECAL has to be fast, resistant to radiation damage and provide a fine granularity. It is divided into a barrel part and two endcaps which cover pseudorapidities of $|\eta| < 1.479$ and $1.479 < |\eta| < 3$ respectively. The ECAL consists of a large number of homogeneous lead tungstate (PbWO_4) crystals combined with photodetectors. Lead tungstate has a small Molière radius ($2.2\ \text{cm}$), a short radiation length ($0.89\ \text{cm}$) and a high density ($8.28\ \text{g}/\text{cm}^3$) which allows the construction of a very compact calorimeter and thus to install tracker, ECAL and HCAL within the solenoid. A sketch of the ECAL is shown in figure 5.3.

In the barrel part, 61200 crystals are installed. Each crystal is equipped with a pair of Avalanche Photo Diodes (APD) which are read out in parallel and are operated at a temperature of 18°C and a gain of 50. The temperature is maintained with help of a water cooling system. The barrel crystals have a truncated pyramid shape with a rectangular surface of $22 \times 22\ \text{mm}^2$ at the side closer to the beamspot and $26 \times 26\ \text{mm}^2$ at the other side. The length of the crystals is $23\ \text{cm}$ which corresponds to 25.8 radiation lengths. In order to minimize the alignment of particle trajectories and cracks between the crystals, the latter ones are installed in a 'quasi-projective' geometry. This means that the crystal axis is tilted (by 3°) with respect to the vector joining beamspot and crystal surface. The arrangement of the crystals into modules and supermodules is shown in figure 5.3.

In the endcaps, the crystals are arranged in units of 5×5 crystals. They have a surface

5. Compact Muon Solenoid

of $28.62 \times 28.62 \text{ mm}^2$ at the inner side, $30 \times 30 \text{ mm}^2$ at the outer side and a length of 22 cm ($\hat{=}$ 24.7 radiation lengths). Each endcap consists of 7324 crystals. Similar to the barrel crystals, the endcap crystals are slightly tilted (by $2^\circ - 8^\circ$) with respect to the beamspot. In the endcaps, Vacuum Photo Triodes (VPT) are used. They were specially designed for the use in the strong magnetic field. In front of each endcap, a preshower detector is installed. The preshower detectors are sampling calorimeters with 2 layers. Each layer consists of lead radiators and silicon strip sensors. The main purpose of the preshower detectors is to identify neutral pions.

The energy resolution of the ECAL was studied with test beams. The result is given in equation 5.1.

$$\left(\frac{\sigma}{E/\text{GeV}}\right)^2 = \left(\frac{2.8\%}{\sqrt{E/\text{GeV}}}\right)^2 + \left(\frac{0.12}{E/\text{GeV}}\right)^2 + (0.3\%)^2 \quad (5.1)$$

Here, the first term is called 'stochastic' term and accounts for fluctuations of the lateral shower containment and the energy deposited in the preshower detector. The second term is the 'noise' term and incorporates effects due to electronic and digitalization noise. The third term does not depend on E and is thus called 'constant' term. It accounts for intercalibration errors and non-uniformity of the light collection.

HCAL

The structure of the HCAL can be divided into a barrel part (HB), an outer barrel part (HO), two endcaps (HE) and two forward calorimeters (HF). The main parts are the HCAL Barrel (HB) and the HCAL Endcaps (HE). An overview of the HCAL design can be found in figure 5.4-left.

The HB is located between the ECAL barrel part and the solenoid, which limits the amount of material which can be used for the absorption of the shower. It is a sampling calorimeter in which absorber and plastic scintillator layers alternate. With the exception of the front and back layer which are made of steel the absorbing material is brass. The segmentation of the HB is 0.087×0.087 in $\eta \times \phi$ provided by 32 segments in η and 72 segments in ϕ . Since the scintillation light of all layers of an element at given η and ϕ is collected with wavelength shifting fibres and transported to a single photodiode, no segmentation in r is available (with the exception of the two outermost segments on each side). In order to catch the shower tails, an additional set of scintillators is placed outside of the solenoid (HO). Here, the solenoid is used as additional absorbing layer. HB and HO cover a pseudorapidity range of $|\eta| < 1.3$.

Similar to the HB the HE is designed as sampling calorimeter. It covers a pseudorapidity range of $1.3 < |\eta| < 3$ and has a segmentation of 0.087×0.087 for $|\eta| < 1.6$ and 0.17×0.17 for the remaining range. In contrast to the HB, the HE segments have a readout structure which allows for a coarse longitudinal segmentation.

At 11.2 m from the interaction point, the HFs are located extending to pseudorapidity coverage up to $|\eta| < 5.2$. For this high η range the rate of charged hadrons is extremely large. Therefore a key feature for the design of the HF is to resist the radiation damage. The HF is made of quartz fibers as active material which are inserted into grooved steel plates. It uses Cherenkov light which is emitted in the fibres by incident particles and measured with photomultiplier tubes. The large coverage in η is important for an accurate determination of E_T^{miss} (see section 6.2 for details).

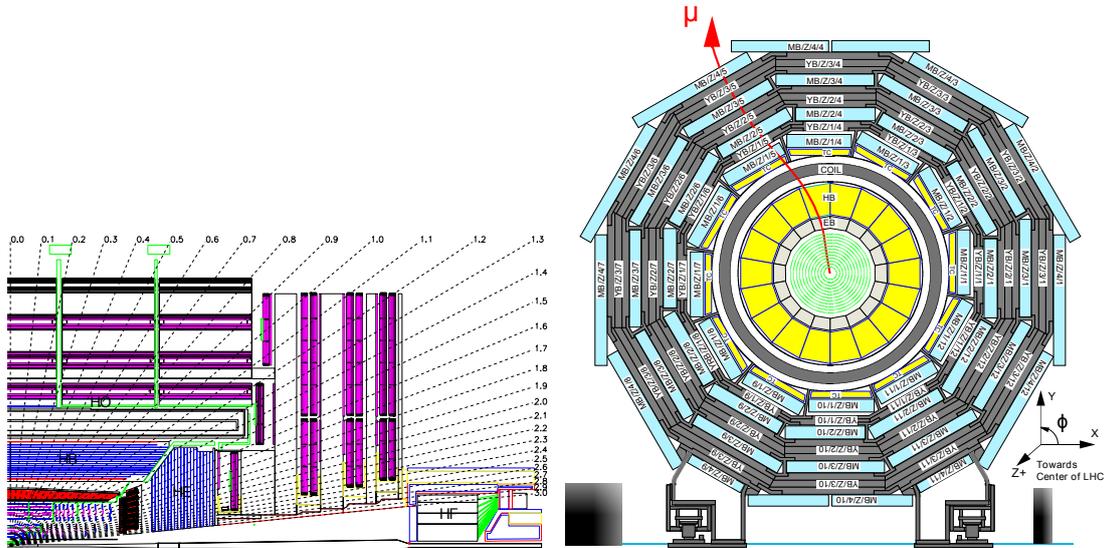


Figure 5.4. Left: Scheme of the HCAL design. Right: Layout of the CMS muon system in the barrel. Shown is a cross-section of one of the 5 wheels. MB labels the muon chambers, while YB labels elements of the iron yoke. Both figures are from [36].

Superconducting magnet

The momentum of charged particles can be determined by measuring the bending of the particles' trajectories in a magnetic field. In order to achieve precise measurements for high energetic particles, a strong magnetic field is needed. The magnetic field in the CMS detector is created by a superconducting solenoid which is located between HCAL and muon system. The magnet has a diameter of 6 m, a length of 12.5 m and a weight of 220 t. Its core part are the four layers of reinforced and stabilized NbTi windings. On the inside of the solenoid the homogeneous magnetic field reaches a strength of 3.8 T. Outside of the magnet, the magnetic flux is returned with help of an iron yoke whose total weight amounts to 10000 t.

Muon System

The muon system is located outside of the solenoid. Similar as for the other detector systems it can be divided into a barrel part and two endcaps. Three types of gaseous detectors are deployed. In the barrel part Drift Tube (DT) chambers are used. Along the z-axis, the barrel muon system is divided into 5 wheels. In each wheel 4 stations of muon detectors surround the solenoid and are installed around and in between the iron yoke. Each station contains 8 chambers measuring the r - Φ coordinate of muons. In the first 3 stations 4 additional chambers can measure the z coordinate of a muon. The barrel part of the muon system covers a pseudorapidity range of $|\eta| < 1.2$. To provide an independent and robust trigger system, Resistive Plate Chambers (RPC) are installed in the first two and in the last station. RPCs have a fast response and a good time resolution at the cost of a coarse spatial resolution. In the endcaps, the conditions are more challenging. The magnetic field is large and non-uniform and the muon rates as well as the background rates are high. This motivates to the use of Cathode Strip Chambers (CSC), which are resistant to radiation and provide a fast response as well as a fine segmentation. In each of the endcaps, 4 CSC stations are installed. With the

5. Compact Muon Solenoid

endcaps the pseudorapidity coverage is extended to $|\eta| < 2.4$.

Trigger and data acquisition

The enormous event rate at LHC conditions makes it impossible to read-out and store the data from all the events. Therefore a trigger system is needed to decide which of the events should be stored. An overview of the CMS Data Acquisition (DAQ) and trigger system is shown in figure 5.5. The CMS trigger system is divided into two levels called Level-1 (L1) trigger and High Level Trigger (HLT). Combining both trigger levels should reduce the rate of events by a factor of at least 10^6 . The L1 trigger system is based on programmable electronics and uses only coarsely binned data from the calorimeters and the muon system. This is crucial, since the L1 trigger system has to analyze every event and the data has to be temporally stored in the front-end electronics. Since the memory of the latter ones is limited, the L1 trigger has to be very fast. The L1 trigger system reduces the event rate down to a maximum of 100 kHz. The remaining event rate leads to very challenging conditions for the DAQ system. At the design luminosity, it has to treat a data-flow of ≈ 100 GB/s. If the L1 trigger fires, the event is passed to the HLT. The HLT is a software trigger which runs on a processor farm with approximately 13000 CPUs (see [41] for details). In contrast to the L1 trigger, the HLT can use the full event information from all the detector subsystems and perform similar operations as used for off-line analysis. The requirements for the trigger used in this analysis will be discussed in section 9.1.1. Combining L1 and HLT leads to a final trigger rate of a few 100 Hz. The remaining events are stored, processed and distributed with help of the LHC computing grid (see [42] for details).

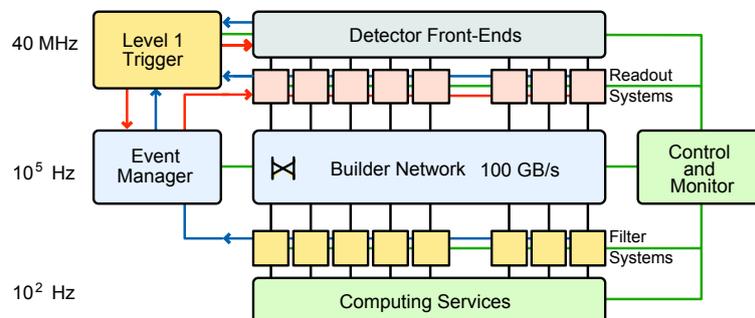


Figure 5.5. Overview of the CMS DAQ and trigger system. Figure taken from [36].

6 Object reconstruction

6.1. Electron reconstruction

In this section, a review of the electron reconstruction at CMS is presented. It starts with a general overview of the signature of an electron and the steps needed for the reconstruction. This introduction is followed by a more detailed description of some aspects of the electron reconstruction including the clustering algorithms, the reconstruction of electron tracks and the energy corrections of the ECAL measurements. The overview of the electron reconstruction is based on [43, 44]. In the following, the term electron designates electrons as well as positrons.

6.1.1. Overview

In an idealized scenario, the electron reconstruction is straight forward. Two major sources of information are available - the tracker and the ECAL. On its way to the ECAL, an electron will produce a measurable track, which is bent due to the magnetic field of the solenoid. When it reaches the ECAL it will deposit its energy in a small number of crystals (for 120 GeV barrel electrons 97% of the energy is deposited within a grid of 5×5 crystals [43]). The energy deposit in the ECAL should be very similar to the electron momentum as measured from the bending of the track. In reality the situation is much more complicated. The silicon tracker constitutes of a large amount of material, which the electrons have to pass. Due to the large material budget and the strong magnetic field, electrons passing the tracker will radiate a non negligible amount of bremsstrahlung before reaching the ECAL (at 50 GeV 35% (10 %) of the electrons radiate more than 70% (95%) of their initial energy [43]). To provide an accurate measurement of the electron energy with the ECAL, the energy deposits of bremsstrahlungs-photons in the ECAL have to be incorporated. The radiation of bremsstrahlung also complicates the track reconstruction, since it can lead to significant changes in the curvature of the electrons trajectory. On top of that, some of the bremsstrahlungs-photons will convert before reaching the ECAL producing secondary electrons, which can lead to ambiguous electron candidates. All these effects have to be considered in the reconstruction. In the following the general steps performed during the electron reconstruction are presented.

Electron seeding

The electron reconstruction starts with the reconstruction of the energy deposit in the ECAL. As said before, an electron or photon reaching the ECAL will deposit its energy in a small amount of adjacent crystals. During the reconstruction, the energy deposits in these crystals are grouped to form a cluster. The incorporation of bremsstrahlungs-photons is done by searching for the ECAL clusters of the latter ones and grouping their clusters together with the cluster of the primary electron to form a supercluster (a cluster of clusters). This is done with help of a clustering algorithm. Since barrel and endcap have a different crystal geometry and alignment, the clustering algorithms differ. For barrel electrons the hybrid algorithm is used, while in the endcap the multi5x5

6. Object reconstruction

algorithm is applied. These clustering algorithms are explained in 6.1.2. Both algorithms exploit the fact, that the magnetic field of the solenoid is aligned along the z-axis of the detector. Therefore, the bremsstrahlungs-photons will spread along the ϕ direction, while the spread in η will remain small. After the reconstruction of the supercluster, its position is used to search for an appropriate trajectory seed. Here, the fact that the energy weighted position of the supercluster (see 6.1.2 for details) corresponds on average to the impact point of an electron with the same energy which did not radiate bremsstrahlung is used. Based on the supercluster, this allows to define a search region for possible hits in the innermost tracker layers for each of the charge hypotheses. Pairs or triplets of hits in the pixel detector and TEC layers compatible with the supercluster form track seeds which are used to start the track reconstruction described in the following.

The procedure explained so far is called ECAL-driven seeding. It works well for isolated high energetic electrons ($p_T > 10$ GeV). For low p_T electrons and electrons inside of jets a complementary approach exists. In the other approach the reconstruction starts from information in the tracker and is therefore called tracker-driven seeding. Since this analysis is based on isolated high energetic electrons, the tracker-driven seeding is not of interest and thus not further explained here.

Electron tracking

The next step is to reconstruct the electron tracks starting with a trajectory seed. For this purpose a special track finding algorithm was developed, to cope with the energy losses due to bremsstrahlung. The tracks reconstructed with this algorithm are called Gaussian Sum Filter (GSF) tracks. More details are given in section 6.1.3.

Electron preselection

Electron candidates are formed by combining superclusters with appropriate tracks. Each candidate has to pass the following preselection:

- $E_{T,sc} > 4$ GeV
- $H/E < 0.15$
- $\Delta\eta_{in} = |\eta_{sc} - \eta_{in}^{extrap}| < 0.02$
- $\Delta\phi_{in} = |\phi_{sc} - \phi_{in}^{extrap}| < 0.15$

Here, $E_{T,sc}$ is the transverse energy of the supercluster and H/E is the ratio between the energy deposited in the HCAL in a cone of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.15$ around the supercluster position and the energy of the supercluster. These two requirements are already applied during the clustering step. $\Delta\phi_{in}$ ($\Delta\eta_{in}$) designates the difference between the ϕ_{sc} (η_{sc}) of the supercluster and the ϕ_{in}^{extrap} (η_{in}^{extrap}) extrapolated from the innermost track measurement to the ECAL.

Removal of ambiguous candidates

The conversion of bremsstrahlungs-photons can lead to duplicate electron candidates. This can happen, if a hard bremsstrahlungs-photon converts and the hits from the conversion legs are close to the expected position of the electron in the next tracker layer. These candidates will be associated to the same supercluster and will have GSF-tracks

close to each other. To resolve this ambiguity, an algorithm is used to choose the track of the primary electron based on the candidate properties. For candidates with ECAL-driven seeds, the candidate which has the innermost tracker hit is chosen. For candidate pairs which have the first hit on the same layer, the candidate with the better E/p ratio is selected.

Electron charge determination

The electron charge can be determined from the curvature of the GSF-track. To improve the charge measurement, two additional estimates are used. The final charge measurement of the candidate is determined by a majority method in which the charge predicted by at least two of the three methods is chosen. The other charge estimates are obtained from the curvature of the track reconstructed with the standard tracking algorithm (instead of the dedicated one for electrons) and from the charge of the supercluster. The latter one is obtained by calculating the sign of the difference in ϕ between two vectors. The first vector points from the beamspot to the supercluster while the second one connects beamspot and the first hit in the tracker.

Electron momentum determination

The momentum of the electron candidate can be obtained by two complementary methods. The first one is the energy of the supercluster after corrections (see 6.1.4). The second one is the p_T measurement from the tracker. In general, a weighted mean of these two estimates using the uncertainties of the individual measurements as weight is taken. For high p_T , the measurement will be dominated by the ECAL measurement. However, if the electron hits the ECAL in the region of the intermodule gaps, it can happen that even for high p_T the measurement is dominated by the tracker [45]. In these cases, there is a chance that a badly reconstructed track leads to a complete misreconstruction of the energy. Following the recommendations from [45], the ECAL measurement is used as electron energy in this analysis.

6.1.2. Electron clustering

Hybrid algorithm

The first step of the hybrid algorithm [46, 47] is to select the crystals with $E_T > 1$ GeV and list them in a descending order of E_T . These crystals can start the clustering. For each crystal in the list which is not already assigned to a cluster the algorithm proceeds as follows:

- construct a 5×1 domino in $\eta - \phi$ around the seed crystal
- repeat last step for crystals with the same η on a ϕ road of ± 17 crystals
- discard dominoes with $E_T < 0.1$ GeV
- group remaining dominoes which are connected to local clusters
- remove local cluster with a maximal domino energy < 0.35 GeV

A sketch of the algorithm is shown in figure 6.1-left. The local clusters in the $\eta - \phi$ road passing the energy threshold are combined to form a supercluster.

6. Object reconstruction

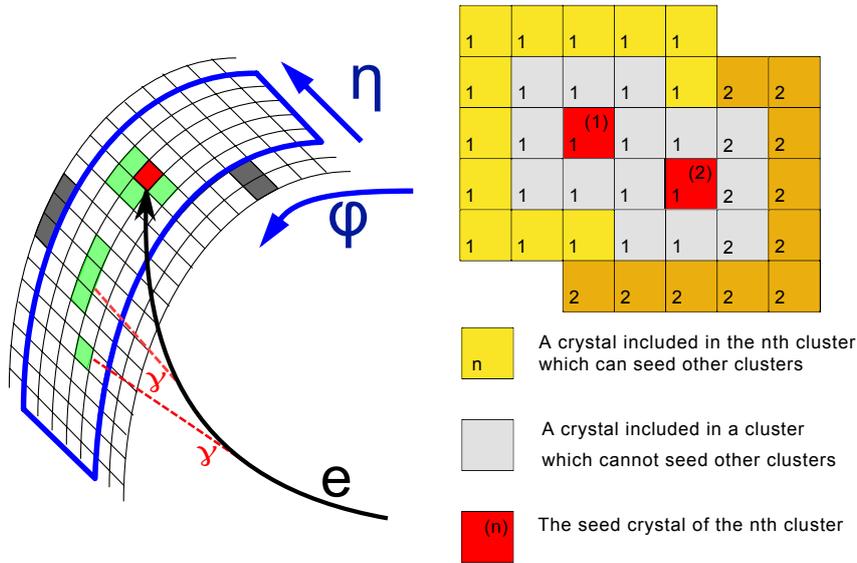


Figure 6.1. Left: Sketch of the hybrid algorithm. Shown is a part of the grid of barrel ECAL crystals. The colored squares represent crystals with an energy deposit. Green (grey) crystals mark energy deposits which are (not) selected by the algorithm. The red crystal is the seed crystal. The blue box represents the road of 5x1 dominoes which is created around the seed crystal in ϕ direction. Figure based on a similar figure from [46]. Right: Sketch of the multi5x5 algorithm. Shown is the clustering around a seed crystal (1) with the seeding of an additional 5×5 matrix around crystal (2). Figure taken from [47]

Multi5x5 algorithm

In contrary to the barrel, the endcap crystals do not follow a $\eta - \phi$ symmetry. Therefore, a different clustering algorithm is needed. The algorithm applied in the barrel is called 'multi5x5' algorithm [47]. Similar to the hybrid algorithm all crystals are sorted in descending order of E_T . Each crystal with $E_T > 0.18$ GeV can start the clustering. For each crystal which is not already assigned to a cluster and which is the local maximum of its four direct neighbours ('swiss-cross' pattern) the algorithm proceeds as follows:

- construct a 5×5 matrix around the seeding crystal
- group crystals of matrix to form a cluster (only crystals which are not part of other clusters)
- for each of the outer 16 crystals (if part of cluster): check seeding requirements
- for crystals passing seeding requirements:
 - repeat algorithm with new seed
 - add new crystals to first cluster

A sketch of this procedure is shown in figure 6.1-right. The clusters built as described above are again sorted in descending order of E_T . To combine the clusters into superclusters, rectangular windows in η and ϕ of $\Delta\eta = 0.14$ and $\Delta\phi = 0.6$ are created around clusters with an $E_T > 1$ GeV. Every energy deposit in this window is grouped into a supercluster. Each energy deposit can only be part of one supercluster.

Supercluster position

The position of a supercluster is calculated as a weighted mean of its crystal positions. The weights are calculated using the energy deposit in each crystal as shown in equation 6.1.

$$W_i = \begin{cases} W_0 + \log\left(\frac{E_i}{\sum_j E_j}\right) & \text{if } W_0 + \log\left(\frac{E_i}{\sum_j E_j}\right) > 0 \\ 0 & \text{if } W_0 + \log\left(\frac{E_i}{\sum_j E_j}\right) \leq 0 \end{cases} \quad (6.1)$$

Here, the sum loops over all the crystals which are part of the supercluster. The quantity W_0 limits the smallest energy deposit a crystal can have to take part in the position determination. Its value was optimized to be 4.2. This and more information can be found in [46].

6.1.3. Electron tracking

As mentioned above, the electron track reconstruction starts from tracking seeds in the innermost part of the tracker. The general tracking procedure applied at CMS is based on the Kalman filter and will be called Kalman filter tracking in the following. For electrons, the energy loss due to bremsstrahlung demands a modified algorithm which is called Gaussian Sum Filter (GSF). In the following, a brief overview of both algorithms is given. The information presented here are based on [48–51].

Kalman filter tracking

The Kalman filter provides a recursive estimator of a state vector describing a discrete dynamic system. Its application to track reconstruction allows the combination of the track finding procedure and the least square fit of the geometrical parameters describing the track. A track is characterized by a five dimensional state vector (position, momentum and curvature of the track expressed as function of e.g. the z-coordinate of the detector) and a covariance matrix of its parameters. The tracking starts from the track seed and proceeds with two alternating steps. First, the state vector and the covariance matrix is extrapolated to the next tracker layer, including material effects due to the interaction of the particle with the tracker material. In the case of minimum ionizing particles, these effects are mostly multiple coulomb scattering. For electrons, the energy loss due to bremsstrahlung has to be included, which is done by subtracting the mean energy loss from the momentum estimate and increasing the variance of the momentum estimate with the variance of the energy loss. This is an implicit approximation of the energy loss by a single Gaussian probability density function (pdf). The second step in the Kalman filter is to combine the predicted state vector with the actual measurement. In the end, a smoothing procedure is applied. For this purpose, the Kalman filter is applied a second time starting at the outermost hit of the track and extrapolating the track from the outside to the inside of the tracker. The smoothed track is the weighted mean of the two tracks.

GSF tracking

The Kalman filter is optimal, if the underlying pdfs are Gaussian. In the case of bremsstrahlung, this approximation does not hold. Following the model from Bethe

6. Object reconstruction

and Heitler, the pdf for the energy loss of an electron is given by equation 6.2.

$$f(z) = \frac{[-\ln(z)]^{\frac{t}{\ln(2)}-1}}{\Gamma(\frac{t}{\ln(2)})} \quad (6.2)$$

Here, z is the remaining fraction of energy after the passage of the material and t is the path length of the material expressed in units of radiation length.

In the GSF algorithm, the general structure of the Kalman filter with repeating extrapolation and update steps is maintained. The main innovation of the GSF tracking algorithm is to approximate the pdf from equation 6.2 by Gaussian mixtures of several Gaussian pdfs. Each component of this mixture describes a different hardness of the bremsstrahlung. It has been shown, that mixtures obtained by minimizing the quantity D_{CDF} provides a good approximation of $f(z)$. The definition of D_{CDF} is given in equation 6.3.

$$D_{CDF} = \int_{-\infty}^{\infty} |F(z) - G(z)| dz \quad (6.3)$$

Here $F(z)$ is the cumulative distribution function (cdf) of $f(z)$ and $G(z)$ is the cdf of the Gaussian mixture $g(z)$.

Another key feature of the GSF algorithm is, that not only the energy loss, but also the state vectors are described by Gaussian mixtures. Thus, the GSF algorithm can be seen as a multiple application of the Kalman filter in parallel. Since this process will lead to a combinatorial explosion of possible track candidates, only the best candidates are kept after each update step. During the update step, the components from the Gaussian mixture of the state vector are weighted according to their agreement with the measurement. In the end, the mode (component of the mixture with the highest weight) is taken to extract the track parameters.

6.1.4. ECAL energy corrections

The corrected supercluster energy in barrel (E_{SC}^{EB}) and endcap (E_{SC}^{EE}) is calculated according to equation 6.4.

$$\begin{aligned} E_{SC}^{EB} &= F_{e,\gamma} \cdot E_{SC}^{raw} \\ E_{SC}^{EE} &= F_{e,\gamma} \cdot (E_{SC}^{raw} + E_{ES}) \\ E_{SC}^{raw} &= \sum_j G \cdot C_j \cdot L_j(t) \cdot A_j \end{aligned} \quad (6.4)$$

In a first step, the calibrated energy of the supercluster (E_{SC}^{raw}) is calculated from the measured ADC counts (A_j) in each crystal j which is part of the supercluster. The calibration is factorized into three parts. $L_j(t)$ corrects for the transparency loss of the crystals during a run, which is time dependent and monitored with help of a laser system. C_j is the inter-calibration factor for crystal j and G the global energy scale calibration which is derived for barrel and endcap separately. The first calibration was done before the different subsystems of the detector were assembled, using a calibrated test beam. During the run periods, the calibration can be controlled and adjusted using well understood physics processes. Since the mass of the Z boson is known with high

precision, the process $Z \rightarrow ee$ provides an excellent control mechanism for the overall energy scale. Minimum bias events (randomly triggered events) should lead to a similar energy flow in ϕ and can be used for the inter-calibration. Another method is to monitor the E/p ratio of isolated electrons from $W \rightarrow e\nu$ events. In addition to this, π^0 and η mesons can be used for inter-calibration purposes. The second step on the way to the corrected energy is to apply the correction factor $F_{e,\gamma}$ to the calibrated supercluster energy.

In $F_{e,\gamma}$, corrections of energy containment effects are summarized. This includes corrections due to energy loss in the tracker material, variation of the amount of clustered energy as a function of the shower position and leakage of the shower out of crystal boundaries. $F_{e,\gamma}$ is derived separately for electrons and photons since their interaction with the tracker material differs. For superclusters in the endcap the energy deposit in the preshower detector (E_{ES}) has to be added to the supercluster energy.

The information given here is based on [52–54].

6.2. Missing Transverse Energy reconstruction

Neutrinos are the only known particles which leave the detector without producing a detectable signature. The partons involved in a hard scattering process within a proton-proton collision will usually carry an unequal fraction of the proton momentum. Therefore, the collision products will be boosted along the z -axis. The transverse momenta of the partons within the protons are negligible. Thus, the vectorial sum of the collision products' transverse momenta should vanish. The presence of a neutrino would spoil the balance of transverse momenta since its p_T is not measured. This can be exploited by defining a quantity called Missing Transverse Energy (\vec{E}_T), which is basically the negative vectorial sum of all measured transverse momenta in the event and quantifies the amount and direction of 'missing energy' in the transverse plane. Since a lot of BSM physics models predict particles which leave the detector without being detected, \vec{E}_T plays a crucial role in the search for new physics. The reconstruction of \vec{E}_T involves all measured particles in the event and is thus sensitive to various detector and reconstruction effects. Any misreconstruction or misidentification of a particle can have an impact on the reconstruction of \vec{E}_T . The \vec{E}_T reconstruction therefore is a very complex task and relies on a precise detector calibration. Several algorithms for the \vec{E}_T reconstruction were developed. The most common ones are Calo- \vec{E}_T , Track Corrected (TC) \vec{E}_T and Particle Flow (PF) \vec{E}_T . Except for the measurement of muons, the Calo- \vec{E}_T relies solely on the ECAL and HCAL measurements. The TC \vec{E}_T algorithm uses information from the tracker to complement the calorimeter measurements, while the PF \vec{E}_T is calculated from a complete list of particles reconstructed with the PF algorithm using a combination of all CMS sub-systems. A comparison of the performance of the three \vec{E}_T reconstruction algorithms is given in [55] showing that the PF algorithm leads to the best \vec{E}_T resolution. Over time, the PF \vec{E}_T has emerged as the most commonly used algorithm for the \vec{E}_T reconstruction. In this analysis, type-I corrected PF \vec{E}_T is used. This section is split in two parts. In the first part, the PF algorithm and the PF \vec{E}_T calculation are explained. The second part is about the corrections which are applied to the \vec{E}_T . The information given here is based on [55, 56]. More information about the \vec{E}_T reconstruction and performance can be found there. Here, the distinction between the Missing Transverse Energy vector (\vec{E}_T) and its absolute value ($E_T^{\text{miss}} = |\vec{E}_T|$) is made.

6. Object reconstruction

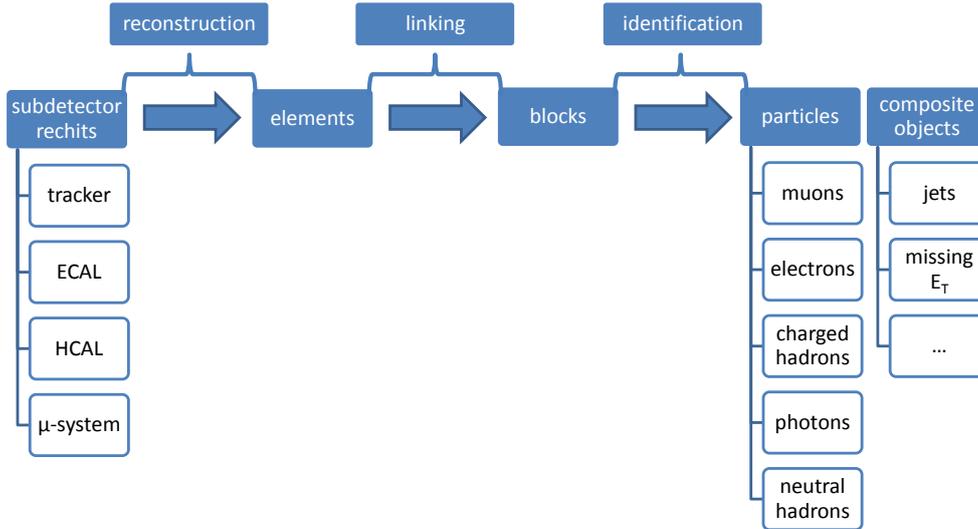


Figure 6.2. Overview of the PF algorithm.

6.2.1. Particle Flow Missing Transverse Energy

The strategy of the PF algorithm is to combine the information of all CMS sub-detectors in order to reconstruct and identify all stable particles (electrons, muons, photons, charged hadrons, neutral hadrons) in an event. Based on this list of particles, composite object like jets and \vec{E}_T are reconstructed. The PF algorithm can be divided in three steps which are briefly explained in the following. A sketch of the algorithm is shown in figure 6.2. The information given here are based on [57].

The first step of the PF algorithm is the reconstruction of tracks from charged particles in the tracker, clusters from energy deposits in the ECAL and HCAL as well as muon tracks from the muon system. These PF elements represent the reconstructed information from each sub-detector.

Each particle will, depending on its type, lead to the creation of several PF elements (e.g track and ECAL cluster in the case of an electron). For the reconstruction of the individual particles, the different PF elements have to be combined. This is done with help of a linking algorithm looping over pairs of PF elements and quantifying the matching between the two elements. For example the linking of an ECAL cluster and a charged particle track is done by extrapolating the track into the ECAL and check if the extrapolated track is within the boundaries of the cluster. In the latter case the track and the cluster are linked. The algorithm proceeds by grouping linked (directly or indirectly) elements into PF blocks. Since the CMS sub-detectors provide a high granularity the block usually consists only out of a few elements.

The key part of the PF algorithm is the identification of individual particles out of reconstructed PF blocks. This is done with an iterative approach starting with the identification of PF muons. Each time a particle is identified, the corresponding PF block is removed from the collection of PF blocks and the algorithm proceeds with the reduced collection. Blocks with a muon track element from the muon system and a charged particle track element are identified as global muon if a global fit of both tracks leads to a reasonable χ^2 and a momentum measurement which is compatible with the one from the charged particle track element. The next step is to identify PF electrons. A

pre-identification procedure is applied to charged particle track elements of appropriate blocks. For blocks passing the pre-identification, the track is refit using the GSF algorithm. The final identification is done based on a boosted decision tree using a large set of track and cluster properties. More Information about the PF electron identification can be found in [58]. After the application of additional quality criteria, each charged particle track linked to calorimeter clusters is treated as PF charged hadron. The particle momentum from the charged track is compared to the energy of the calorimetric clusters. In case that the calorimetric energy is significantly larger, PF photons and neutral hadrons are added to compensate the energy discrepancy. The last step of the particle identification is to identify all remaining ECAL (HCAL) clusters which are not linked to a charged particle track element as photons (neutral hadrons). The list of particles is then used to reconstruct composite objects like jets and $\vec{\cancel{E}}_T$. The PF $\vec{\cancel{E}}_T$ is defined as the transverse component of the negative vector sum of all PF particles as shown in equation 6.5.

$$\vec{\cancel{E}}_T^{\text{uncorr.}} = - \sum_i \vec{p}_{T,i} \quad (6.5)$$

6.2.2. Missing Transverse Energy corrections

Type-I correction

The type-I $\vec{\cancel{E}}_T$ corrections propagate the jet energy corrections (JEC) to the $\vec{\cancel{E}}_T$. CMS uses a factorized approach for the JEC which divides the correction into different levels. Each level corrects for a different effect and is applied to the jet energy obtained after application of the precedent correction levels. The commonly applied corrections are the L1, L2 and L3 correction. The L1 JEC corrects for the shift of the jet energy due to contribution of pile-up. The L2 JEC is called 'relative jet correction' and corrects the jet response to be flat in η . The L3 JEC is the 'absolute jet correction' and ensures a flat jet response in p_T . This and more information about JEC can be found in [59]. For the PF $\vec{\cancel{E}}_T$ type-I correction, the JEC are considered for all jets with $p_T > 10$ GeV and a fraction of energy deposited in the ECAL smaller than 90% of the jet energy.

The formula for the raw PF $\vec{\cancel{E}}_T$ given in equation 6.5 can be expanded by dividing the sum over all PF candidates into a sum of candidates within a jet and a second with the remaining candidates. The first sum can be replaced with a sum of the uncorrected jet p_T of all jets leading to the formula given in equation 6.6

$$\vec{\cancel{E}}_T^{\text{uncorr.}} = - \sum_{\text{jet}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T,i} \quad (6.6)$$

The sum of jet p_T can be further divided into jets with $\vec{p}_{T,\text{jet}}^{\text{L123}}$ larger than 10 GeV and jets with $\vec{p}_{T,\text{jet}}^{\text{L123}}$ smaller than 10 GeV. The resulting formula is given in equation 6.7. Here, $\vec{p}_{T,\text{jet}}^{\text{L123}}$ is the jet p_T after application of the L1, L2 and L3 JEC.

$$\vec{\cancel{E}}_T^{\text{uncorr.}} = - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} > 10\text{GeV}}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{\substack{\text{jet} \\ \vec{p}_{T,\text{jet}}^{\text{L123}} < 10\text{GeV}}} \vec{p}_{T,\text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T,i} \quad (6.7)$$

The final formula for the raw PF $\vec{\cancel{E}}_T$ is obtained by expanding the first term of equation 6.7 into a sum of the jets' p_T using the p_T after the L1 JEC ($\vec{p}_{T,\text{jet}}^{\text{L1}}$) and a sum of the

6. Object reconstruction

Table 6.1. Parameters for the \vec{E}_T - ϕ correction. [56]

	c_{x0} (GeV)	c_{xs} (GeV)	c_{y0} (GeV)	c_{ys} (GeV)
Data	0.27	0.32	-0.23	-0.17
Monte Carlo	0.12	0.02	0.28	-0.13

offset between raw jet p_T and $\vec{p}_{T, \text{jet}}^{\text{L1}}$. This is shown in equation 6.8.

$$\begin{aligned}
 \vec{E}_T^{\text{uncorr.}} = & - \sum_{\substack{\text{jet} \\ \vec{p}_{T, \text{jet}}^{\text{L123}} > 10\text{GeV}}} \vec{p}_{T, \text{jet}}^{\text{L1}} - \sum_{\substack{\text{jet} \\ \vec{p}_{T, \text{jet}}^{\text{L123}} > 10\text{GeV}}} (\vec{p}_{T, \text{jet}}^{\text{uncorr.}} - \vec{p}_{T, \text{jet}}^{\text{L1}}) \\
 & - \sum_{\substack{\text{jet} \\ \vec{p}_{T, \text{jet}}^{\text{L123}} < 10\text{GeV}}} \vec{p}_{T, \text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T, i}
 \end{aligned} \tag{6.8}$$

The type-I correction acts on the first term of the formula by replacing $\vec{p}_{T, \text{jet}}^{\text{L1}}$ correction with $\vec{p}_{T, \text{jet}}^{\text{L123}}$ which is the transverse momentum of the jets including L1, L2 and L3 JEC. The final formula for the PF type-I corrected \vec{E}_T is given in equation 6.9. All formulas are taken from [60].

$$\begin{aligned}
 \vec{E}_T^{\text{typeI}} = & - \sum_{\substack{\text{jet} \\ \vec{p}_{T, \text{jet}}^{\text{L123}} > 10\text{GeV}}} \vec{p}_{T, \text{jet}}^{\text{L123}} - \sum_{\substack{\text{jet} \\ \vec{p}_{T, \text{jet}}^{\text{L123}} > 10\text{GeV}}} (\vec{p}_{T, \text{jet}}^{\text{uncorr.}} - \vec{p}_{T, \text{jet}}^{\text{L1}}) \\
 & - \sum_{\substack{\text{jet} \\ \vec{p}_{T, \text{jet}}^{\text{L123}} < 10\text{GeV}}} \vec{p}_{T, \text{jet}}^{\text{uncorr.}} - \sum_{i \notin \text{jets}} \vec{p}_{T, i}
 \end{aligned} \tag{6.9}$$

ϕ correction

Studies in 2011 revealed the presence of an \vec{E}_T asymmetry in ϕ direction which is present in data and Monte Carlo. The origin of this effect is not yet completely understood. Possible explanations for this asymmetry are an imperfect alignment, a 4 mm shift of the beamline and detector center or a remaining ϕ dependence of the detector response after calibration and corrections. However, the asymmetry was found to be due to a shift in the E_x^{miss} and E_y^{miss} components and to be proportional to the number of reconstructed vertices (N_{vtx}). Based on $Z \rightarrow \mu\mu$ studies a correction procedure was developed (see [56] for details). The E_x^{miss} and E_y^{miss} components are corrected according to equation 6.10. The correction coefficients are given in table 6.1.

$$\begin{aligned}
 E_x^{\text{miss}, \phi} &= E_x^{\text{miss}} - (c_{x0} + c_{xs} \cdot N_{\text{vtx}}) \\
 E_y^{\text{miss}, \phi} &= E_y^{\text{miss}} - (c_{y0} + c_{ys} \cdot N_{\text{vtx}})
 \end{aligned} \tag{6.10}$$

Part III.

Analysis

7 General signal signature

This analysis searches for a deviation from the Standard Model expectation in events with E_T^{miss} and an isolated electron with high E_T ($E_T > 100$ GeV) where the E_T^{miss} originates from a particle which can not be detected with the CMS detector (e.g. a neutrino). As main discriminating variable the transverse mass M_T , which combines the available information about the electron and the E_T^{miss} in the transverse plane is used. A definition of M_T is given in equation 7.1.

$$M_T = \sqrt{2 \cdot E_{T,e} \cdot E_T^{\text{miss}} \cdot (1 - \cos(\Delta\Phi(e, \vec{E}_T^{\text{miss}})))} \quad (7.1)$$

The ratio E_T/E_T^{miss} is assumed to be around one and the angle between E_T and E_T^{miss} in the transverse plane $\Delta\Phi(e, \vec{E}_T^{\text{miss}})$ is assumed to be close to π for signal events. The existence of a signal would lead to a deviation in the M_T spectrum of the electron and E_T^{miss} (also 'e + E_T^{miss} spectrum' in the following) compared to the Standard Model prediction. The shape of the signal in the e + E_T^{miss} spectrum depends on the BSM physics model which is assumed. Since the assumptions made on the signal properties are rather general, an interpretation of the analysis results in terms of different BSM physics models is possible. A general sketch of the LO production process is shown in figure 7.1-left. The process in the bubble depends on the signal model.

The assumed signal properties can be motivated by looking at one of the possible BSM physics models in more detail. Since this search originates from a SSM W' search, the motivation is done in terms of this model. The leading order production process is shown in figure 7.1-right, where the bubble has been replaced with a W' boson. Since the W' has a very short lifetime, it decays almost instantly after its production. Therefore, it can only be reconstructed indirectly through its decay products. In the rest frame of the W' , the absolute value of the electron and neutrino momentum are equal and the angle between the two particles is π . As explained before, the partons producing the W' will usually carry unequal momentum fractions of their proton. Thus, the W' can be boosted along the z axis and the detector rest frame is not equal to the W' rest frame. For this reason the two given properties of the two body decay do not apply directly. However, in the transverse plane the absolute value of the electron and neutrino momentum should be approximately equal and the angle between the two particles around π . The neutrino

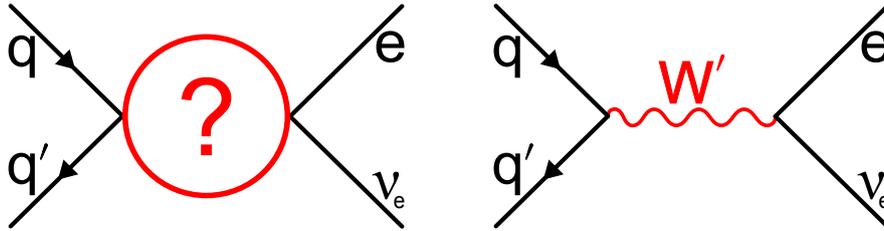


Figure 7.1. Left: General sketch of the signal signature. The process in the bubble depends on the physics model chosen for the signal. Right: Feynman graph of production and decay of a W' in LO.

7. General signal signature

leaves the detector without being detected. This spoils the balance of the measured momenta in the transverse plane and leads to E_T^{miss} . In a W' event like shown in figure 7.1-right the E_T^{miss} will be dominated by the neutrino. Since the invariant mass of the electron-neutrino system is inaccessible from the experimental side of view, the transverse mass M_T will be used as main discriminating variable instead. The transverse mass is the projection of the invariant mass on the transverse plane using E_T^{miss} as the neutrino momentum. In the definition given in equation 7.1, the transverse electron momentum is replaced with the transverse electron energy, which is considered to be identical to the transverse momentum for the energy regime of interest.

The signal properties explained here are taken from detailed studies of W' events (see [61–63]) and were verified to be applicable to the other BSM signal models used in this search (see [29, 64, 65]).

8 Dataset and Monte Carlo simulation of the Standard Model prediction

This chapter consists of two main parts. First an overview of the studied dataset is given. Then the modeling of the Standard Model contribution to the $e + E_T^{\text{miss}}$ spectrum is presented. This includes a summary of the different Standard Model processes which contribute to the $e + E_T^{\text{miss}}$ spectrum as well as the description of their simulation.

To perform an analysis using CMS data, the CMSSW software [66] is crucial. It includes various packages needed for simulation, reconstruction and calibration of CMS event data. CMS datasets and simulations are distributed over the computing centers located everywhere in the world and can be accessed via the LHC computing grid [42]. Since only a small subset of the information stored for each event is relevant for this analysis, it is utile to work with a reduced data format. In order to extract the relevant information and to store the output on the local computing center in Aachen, a dedicated software package [67] was developed and is maintained by various members of the III. Physikalischen Instituts A. The actual analysis is then performed on these reduced samples with help of the ROOT framework (version 5.34) [68]. With one exception (see chapter 10 for details), all contributions of Standard Model processes to the search region are evaluated using Monte Carlo simulations. This includes not only the simulation of the physics processes but also a simulation of the detector response and an emulation of the trigger chain. The simulation of the CMS detector was done with the GEANT4 software [69, 70] which is included in CMSSW.

8.1. Dataset

In 2012, CMS recorded 21.8 fb^{-1} of pp collision data at a center-of-mass energy of $\sqrt{s} = 8 \text{ TeV}$. From this total dataset, 19.7 fb^{-1} were certified by the Data Quality Monitoring (DQM) group. In general, only certified data is used for physics analysis, to ensure that all the subdetectors were working according to their specifications. The reconstruction of the collision events is done using the CMSSW software. At the end of the data recording period, the reconstruction is usually redone for the whole dataset to incorporate the improved knowledge about the detector and the running conditions gained during the data taking. This search uses the `SingleElectron` dataset from the 22Jan-ReReco reconstruction campaign done with CMSSW_5_3_X. The applied global tag, which encodes the detector and run conditions used for the reconstruction, is `FT_53_V21_AN5`. The dataset divided into the different run periods together with the corresponding luminosity and the chosen trigger for this analysis is shown in table 8.1. More information about the trigger is given in section 9.1.1.

8. Dataset and Monte Carlo simulation of the Standard Model prediction

Table 8.1. Overview of the 2012 pp collision dataset split into its different run periods. The full name of the dataset for each entry in the first column is /SingleElectron/Run2012[.]-22Jan2013-v1/AOD. Also shown is the integrated luminosity and the chosen trigger for each subset. The total luminosity adds up to 19.7 fb^{-1} .

Dataset Name	Run Range	Trigger	L (fb^{-1})
Run2012A	190456-193621	HLT_Ele80_CaloIdVT_TrkIdT	0.9
Run2012B	193833-196531	HLT_Ele80_CaloIdVT_GsfTrkIdT	4.4
Run2012C	198022-203742	HLT_Ele80_CaloIdVT_GsfTrkIdT	7.0
Run2012D	203777-208686	HLT_Ele80_CaloIdVT_GsfTrkIdT	7.4

8.2. Monte Carlo simulation of the Standard Model prediction

8.2.1. Standard Model processes

Several Standard Model processes contribute to the $e + E_T^{\text{miss}}$ spectrum at high M_T . The by far most important process is the off-shell production of W bosons. Other contributions come from $t\bar{t}$, single top, Drell-Yan and Diboson events. The misidentification from jets or photons as electrons leads to additional Standard Model events in the $e + E_T^{\text{miss}}$ spectrum. In the following, the different background contributions and their Monte Carlo simulation are explained in more detail. The production cross sections and numbers of generated events are taken from PREP [71], which is the CMS management system for Monte Carlo production. If possible, the cross sections are corrected with a k-factor ($k = \sigma_{\text{NjLO}}/\sigma_{\text{LO}}$) to include higher order corrections. If not stated otherwise, the higher order cross sections are taken from [72]. To compare the Monte Carlo simulations with the measured data, each sample is scaled with a factor given in equation 8.1 to match the integrated luminosity of the data.

$$w = \frac{\sigma \cdot L}{N_{MC}} \quad (8.1)$$

Here σ is the cross section of the process, N_{MC} the number of generated events and L the integrated luminosity of the data. A list of all the Standard Model Monte Carlo samples which are used in this analysis together with the corresponding cross sections and numbers of generated events is given in table A.1.

$W \rightarrow \ell\nu$

As explained in chapter 7, the signature of signal events can be summarized as W-like decay to an electron and a neutrino at high M_T leading to $E_T/E_T^{\text{miss}} \approx 1$ and $\Delta\Phi(e, \vec{E}_T) \approx \pi$. Therefore, the off-shell production of W bosons at high masses decaying into an electron and a neutrino is indistinguishable from possible signal events. The same holds for $W \rightarrow \tau\nu$ events, if the τ lepton decays into an electron and a neutrino. These events will be shifted to lower M_T compared to events, where the W directly decays into an electron and a neutrino. An explanation for this behavior can be found in [64]. Being the main SM contribution to the $e + E_T^{\text{miss}}$ spectrum, a precise knowledge of the $W \rightarrow \ell\nu$ process at high M_T is crucial.

8.2. Monte Carlo simulation of the Standard Model prediction

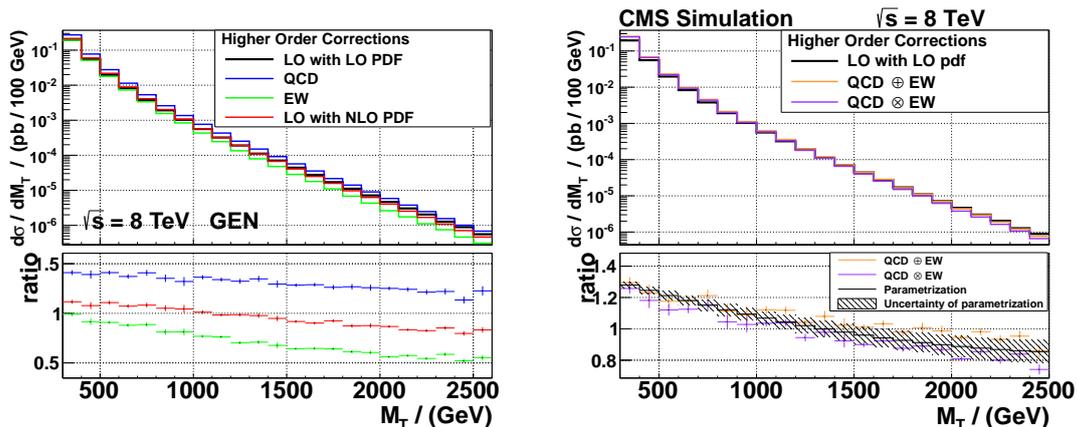


Figure 8.1. Left: Impact of QCD and EW NLO corrections on the differential cross section of the $W \rightarrow \ell\nu$ process as function of M_T . Both corrections are calculated separately. Right: Two different approaches (QCD \oplus EW additive approach, QCD \otimes EW multiplicative approach) for the combination of the individual QCD and EW correction. The ratio between the average of the two approaches and the LO prediction is used to correct the LO prediction. Before averaging the two approaches, they are parametrized with second order polynomials to obtain a smooth description. The result can be seen in the lower part of the plot. The difference between the two approaches is used as systematic uncertainty. Both plots by M. Olschewski [74].

The simulation of the $W \rightarrow \ell\nu$ process is split in three subsamples per lepton (e, τ) produced with PYTHIA [73] in LO. Besides a main sample produced without dedicated generator cuts, two high p_T samples binned in \hat{p}_T (first sample: $100 \text{ GeV} < \hat{p}_T < 500 \text{ GeV}$, second sample: $\hat{p}_T > 500 \text{ GeV}$) are used. These samples cover most of the M_T region which is analysed in this search and ensure a description of the $W \rightarrow \ell\nu$ process which is not limited by statistics, even at high M_T .

Higher order QCD and electroweak (EW) corrections have a sizeable impact on the production rate of the $W \rightarrow \ell\nu$ process. Since the strength of the individual corrections is not flat in M_T , it is useful to derive a M_T dependent correction factor. These calculations have been done by M. Olschewski [74] in NLO using MC@NLO [75] for the QCD and HORACE [76] for the EW part. The differential cross section with respect to M_T for the individual corrections are shown in figure 8.1-left. Also shown is the LO prediction and the LO prediction using a NLO PDF set. The latter one is needed for the combination of the QCD and EW corrections, since HORACE and MC@NLO both use a NLO PDF set. The PDF set for all three predictions using a NLO PDF set is CT10 [77]. The LO PDF set used for the Monte Carlo samples is CTEQ6L1 [78].

The incorporation of QCD NLO corrections lead to an enhancement of the $W \rightarrow \ell\nu$ production of around 40% at M_T values around 300 GeV and around 20% at $M_T \approx 2500$ GeV. The EW corrections have a stronger M_T dependence and lead to a decrease of the differential cross section. While at $M_T \approx 300$ GeV the impact is small, the differential cross section at $M_T \approx 2500$ GeV is reduced by a factor of approximately 0.5.

Due to mixed QCD and EW contributions, the combination of the two corrections from the individual calculations is not a trivial task. The combination is calculated with two different mathematical approaches. The first approach is given in equation 8.2, where $[d\sigma/d\mathcal{O}]_X$ is the differential cross section with respect to the observable \mathcal{O} and X refers to the order of the calculation (e.g. $[d\sigma/d\mathcal{O}]_{\text{QCD}}$ means NLO in QCD). This

8. Dataset and Monte Carlo simulation of the Standard Model prediction

additive approach should provide a good description if the correlation between the QCD and the EW correction is small.

$$\left[\frac{d\sigma}{d\mathcal{O}}\right]_{QCD\oplus EW} = \left[\frac{d\sigma}{d\mathcal{O}}\right]_{QCD} + \left[\frac{d\sigma}{d\mathcal{O}}\right]_{EW} - \left[\frac{d\sigma}{d\mathcal{O}}\right]_{LO} \quad (8.2)$$

The second one combines QCD and EW corrections in a multiplicative way, which is expected to be a good measure for strong correlations. The formula is given in equation 8.3.

$$\left[\frac{d\sigma}{d\mathcal{O}}\right]_{QCD\otimes EW} = \left(\frac{\left[\frac{d\sigma}{d\mathcal{O}}\right]_{QCD}}{\left[\frac{d\sigma}{d\mathcal{O}}\right]_{LO}}\right) \cdot \left[\frac{d\sigma}{d\mathcal{O}}\right]_{EW} \quad (8.3)$$

In both formulas the LO term refers to the prediction using the NLO PDF set. The results can be seen in figure 8.1-right. The difference between the two methods is around 10%. For this analysis, the average between the two methods is used and the difference is taken as a systematic uncertainty on the correction. More information about the combination of the two approaches can be found in [79]. The LO $W \rightarrow \ell\nu$ events from the Monte Carlo samples are corrected depending on their M_T by using a k-factor defined in equation 8.4 as event weight.

$$k(M_T) = \frac{\Delta\sigma(NLO)/\Delta M_T}{\Delta\sigma(LO)/\Delta M_T} \quad (8.4)$$

To obtain a smooth description of the k-factor, a parametrization of the individual corrections with polynomials of second order is done before combining them. The combined result can be seen in the ratio plot of figure 8.1-right together with the systematic uncertainty derived from the difference between the two combination methods. For M_T values around 300 GeV the resulting k-factor is around 1.3 and decreases with M_T . At $M_T \approx 2500$ the k-factor is around 0.9. This shows that the NLO corrections are very important for an accurate description of the $W \rightarrow \ell\nu$ process at high M_T .

According to [79, 80] the contribution from NNLO corrections are small and thus neglected here.

Multijet

Multijet events generated via strong interaction occur at a high rate at the LHC. If a jet is misidentified as electron and a misreconstruction of the jet energy leads to a non negligible amount of E_T^{miss} , multijet events can contribute to the $e + E_T^{\text{miss}}$ spectrum. The simulation is done with PYTHIA in multiple samples binned in the scale of the hard interaction \hat{p}_T . A filter is used to simulate only events which pass a minimal set of isolation criteria. This is done to maximize the amount of generated events, having a realistic chance to be misidentified as an electron. Jets which do not resemble an electron at all, are very unlikely to be misidentified and thus not of interest for an analysis as presented here. The contributions from multijet events will be titled 'QCD'. The QCD Monte Carlo is only used for cross-check purposes. For the analysis a data-driven method is used (see chapter 10).

$t\bar{t}$ and single top

The top quark decays via weak interaction into a b quark and a W boson. Events where the W subsequently decays into an electron and a neutrino will contribute to the $e + E_T^{\text{miss}}$ spectrum. Additional contributions arise from events, where the W decays into a τ and a neutrino and the τ further decays into an electron and a neutrino. Analog to the multijet background, jets emerging from the top decay could be misidentified as electrons. The $t\bar{t}$ process is simulated with MC@NLO including NLO QCD corrections. For the single top production, the processes are divided into the different production processes (s-channel, t-channel, associated (tW) production) and the charge of the top quark. Each of the 6 processes is simulated using the NLO generator POWHEG [81, 82].

$Z/\gamma^* \rightarrow \ell\ell$

The $Z/\gamma^* \rightarrow ee$ process is an important source of isolated electrons. A well reconstructed $Z/\gamma^* \rightarrow ee$ event will contain two isolated high energetic electrons and no substantial amount of E_T^{miss} . However, if one of the electrons is misreconstructed or outside the acceptance of the detector, enough E_T^{miss} can occur to contribute to the $e + E_T^{\text{miss}}$ spectrum. An additional contribution arises from $Z/\gamma^* \rightarrow \tau\tau$ events, where a τ further decays to an electron and a neutrino. The $Z/\gamma^* \rightarrow ee$ process is simulated with POWHEG in NLO, while the $Z/\gamma^* \rightarrow \tau\tau$ events are generated with PYTHIA. For both processes the samples are divided in subsamples binned in m_Z to provide enough statistics at high energies. The $Z/\gamma^* \rightarrow \ell\ell$ process will be called 'Drell-Yan' in the following.

Hard photon production

If a track is accidentally mismatched to the cluster of a photon in the ECAL, this photon can be misidentified as an electron. Due to initial or final state radiation, hard photons can be produced in QCD events. The simulation of these events is done with PYTHIA in several samples binned in the scale of the hard interaction. This process will be referred to as ' $\gamma + \text{Jets}$ '. Another possibility to obtain hard photons in QCD events is during the hadronization process. This is accounted for by PYTHIA during the simulation of parton showering.

WW, WZ, ZZ

The production of two weak bosons (WW, WZ, ZZ) is another Standard Model process to be considered. In the complicated topology of these events, there are lots of possibilities to produce an isolated high energetic lepton and E_T^{miss} . For each of the 3 processes one main sample and one high- p_T sample have been produced with PYTHIA. The 3 processes will be dubbed as 'Diboson' in the following.

8.2.2. Pileup reweighting

Due to the very high instantaneous luminosity at the LHC, there usually is more than one interaction per bunch crossing. In fact, the average number of interactions per bunch crossing is around 21 for 2012 LHC conditions [40]. The majority of these interactions are soft QCD multijet events with low momentum transfer. Every hard interaction which causes a trigger to fire and thus the event to be recorded, is accompanied by several additional interactions. The latter ones are called pileup. The number of additional

8. Dataset and Monte Carlo simulation of the Standard Model prediction

interactions can be estimated by multiplying the total inelastic proton proton cross section with the measured instantaneous luminosity for the current bunch crossing. Since every interaction should lead to one primary vertex, the number of vertices per event is closely related to the number of interactions per bunch crossing. Therefore counting the number of reconstructed vertices and dividing it by the vertex reconstruction efficiency is another method to estimate the amount of pileup. The official method at CMS is the first one. The second one is used as crosscheck.

Pileup is a very important effect at the LHC which has to be taken into account for the Monte Carlo generation. The amount of pileup depends very much on the present run conditions at the LHC, which change throughout the year. Since the Monte Carlo generation is usually done before or while the data is recorded, the simulation is done using an expected pileup scenario. The simulations used in this analysis are simulated with the pileup scenario S10 [83]. The relative frequencies of the number of interactions per bunch crossing in this scenario are shown in figure 8.2. Also shown is the corresponding distribution from data, where the number of interactions per bunch crossing is calculated via the first of the two methods described above.

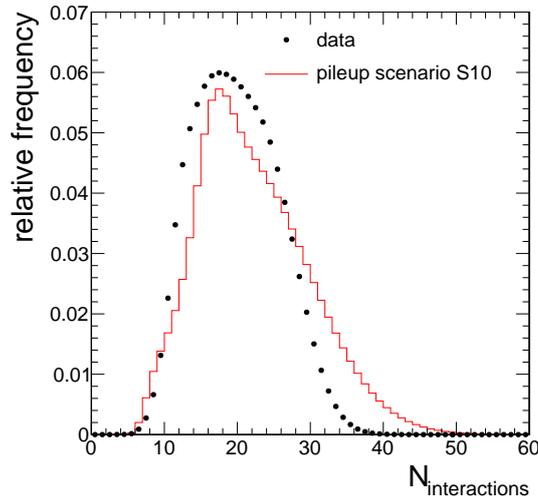


Figure 8.2. Relative frequency of the number of interactions per bunch crossing as defined in the pileup scenario S10 [83] used for Monte Carlo generation compared to the actual distribution measured in data.

To improve the agreement of the pileup simulation and the actual pileup occurring in data, a procedure called pileup reweighting is applied. For each number of interactions per bunch crossing shown in figure 8.2 a scalefactor is derived by dividing the value from the data distribution by the value from the Monte Carlo distribution. Each Monte Carlo event is then rescaled with the appropriate scalefactor. The information about pileup and pileup reweighting given in this section is based on [84, 85].

9 Event selection and efficiencies

The purpose of the event selection is to select events with signal-like properties, reject events from Standard Model processes if they are distinguishable from a possible signal and to ensure that the reconstructed objects of interest (e.g. the electron or E_T^{miss}) are well reconstructed. It can be divided into three parts. The first part, which will be called 'preselection' here, consists of the trigger and a set of general event quality criteria. The trigger ensures the existence of at least one electron candidate in the event. The second part is the electron ID. It is applied to all the electron candidates in events which pass the preselection. Its purpose is to select well reconstructed electrons and reject electron candidates which are misidentified photons or jets. The last part of the selection exploits the signal properties explained in chapter 7 to further separate signal from Standard Model events.

The efficiency of the reconstruction, the trigger and the electron ID is measured in data and Monte Carlo simulations. If the efficiencies differ between data and simulation, the simulation can be corrected by applying a data/MC scale factor.

In this chapter the selection and the corresponding efficiency measurement are presented.

9.1. Event selection

9.1.1. Preselection

Trigger

Since the signal signature predicts one highly energetic electron, the use of a single electron trigger for the event selection is convenient. The main trigger chosen for this analysis is the `HLT_Ele80_CaloIdVT_GsfTrkIdT` trigger which was unprescaled (every time the trigger fired the event was recorded) during the whole run range. For the first part of data taking (see RunA in table 8.1) this trigger was not implemented yet. Instead, the `HLT_Ele80_CaloIdVT_TrkIdT` trigger was used. The only difference between the two triggers is a use of a different tracking algorithm at HLT level. The conditions the triggers impose are summarized in table 9.1 [86]. For a definition of the variables see section 9.1.2.

These trigger constraints are tight enough to allow the trigger to run without a prescale while still being loose enough to be very efficient at high E_T . In general, a trigger will

Table 9.1. Summary of the conditions imposed by the single electron triggers `HLT_Ele80_CaloIdVT_GsfTrkIdT` and `HLT_Ele80_CaloIdVT_TrkIdT` [86].

Tag	Conditions Barrel		Conditions Endcap	
Ele80	$E_T > 80\text{GeV}$		$E_T > 80\text{GeV}$	
CaloIdVT	$H/E < 0.05$	$\sigma_{i\eta j\eta} < 0.011$	$H/E < 0.05$	$\sigma_{i\eta j\eta} < 0.031$
(Gsf)TrkIdT	$ \Delta\eta_{\text{in}} < 0.008$	$ \Delta\phi_{\text{in}} < 0.07$	$ \Delta\eta_{\text{in}} < 0.008$	$ \Delta\phi_{\text{in}} < 0.05$

9. Event selection and efficiencies

not be perfectly efficient with an instant turn-on and thus select no events below and every event above the energy threshold. Instead, the trigger efficiency as function of E_T will have a turn-on region around the energy threshold until it reaches an efficiency plateau. This assumption has to be verified by measuring the trigger efficiency as a function of E_T . This is shown in section 9.2.1. The exact details of this turn-on are very hard to model in Monte Carlo simulations. Therefore, it is suitable to use an offline E_T cut which is higher than the trigger threshold to avoid the turn-on regime. For this analysis the offline E_T cut is set to 100 GeV.

General event requirements

To ensure a good quality of the reconstructed events, some general constraints are applied. Each event is required to contain at least one well reconstructed vertex, which is less than 24 cm away from the center of the detector. The JetMET Physics Object Group (JetMET POG) developed a set of filters to reject events, where the E_T^{miss} is caused by a detector malfunction or noise [87]. They are officially recommended for every analysis using E_T^{miss} , and are applied for this analysis. A summary of these filters is given in table 9.2.

Table 9.2. Summary of the event filter recommended by the JetMET POG [87].

Filtername	Description
CSC tight beam halo	reject events with muons from beam halo
HBHE noise filter	reject events with anomalous HCAL noise due to instrumental issues
HCAL laser filter	reject events where the HCAL calibration laser fired during bunch crossing
ECAL dead cell (TP) filter	reject events where malfunctioning/dead ECAL crystals (about 1% of the crystals) lead to large energy loss
Tracking failure filter	reject events with a serverly displaced primary vertex or a too large number of clusters
Bad EE supercrystal filter	reject events where two regions of crystals in the endcap give anomalously high energies
ECAL laser filter	reject events where ECAL crystals with an unphysically large laser correction lead to large E_T^{miss}
Tracking odd events filter	reject events with aborted track reconstruction or affected by strip tracker coherent noise

9.1.2. Electron ID

To identify well reconstructed electrons, a set of requirements called electron ID is claimed. Here, the HEEP ID [45] is chosen. HEEP is the abbreviation of High Energy Electron Pairs, was originally designed for the Z' analysis [88] and is optimized for high- p_T electrons. A summary of the requirements of the HEEP-ID can be found in table 9.3. The lower E_T bound is replaced with the E_T cut of 100 GeV imposed by the trigger. In the following, the individual criteria will be explained.

Table 9.3. Overview of the official HEEP selection in version 4.1 [45]. The electrons are divided in barrel and endcap electrons using η_{sc} , which is the η of the electrons supercluster. In this analysis, the lower E_T limit is replaced with 100 GeV.

Official HEEP Selection v4.1		
Variable	Barrel ($\eta_{sc} < 1.422$)	Endcap ($1.56 < \eta_{sc} < 2.5$)
E_T	$> 35\text{GeV}$	$> 35\text{GeV}$
isEcalDriven	1	1
$ \Delta\eta_{in} $	< 0.005	< 0.007
$ \Delta\phi_{in} $	< 0.06	0.06
H/E	< 0.05	< 0.05
$\sigma_{i\eta i\eta}$	-	< 0.03
$E^{2x5}/E^{5x5} \parallel E^{1x5}/E^{5x5}$	$> 0.94 \parallel > 0.83$	-
Calorimetric isolation	$< 2 + 0.03 \cdot E_T + 0.28\rho$	$< 2.5 + 0.28\rho$ ($E_T < 50 \text{ GeV}$) $< 2.5 + 0.03 \cdot (E_T - 50) + 0.28\rho$ ($E_T > 50 \text{ GeV}$)
Track isolation	$< 5 \text{ GeV}$	$< 5 \text{ GeV}$
Inner layer lost hits	≤ 1	≤ 1
d_{xy}	< 0.02	< 0.05

isEcalDriven

As explained in section 6.1, the reconstruction of an electron can be seeded by the tracker or the ECAL. For high- p_T electrons the performance of the tracker is less good. Therefore, only electrons from ECAL seeds are used.

$|\Delta\eta_{in}|$ and $|\Delta\phi_{in}|$

$\Delta\eta_{in}$ ($\Delta\phi_{in}$) is defined as explained in section 6.1. This requirement checks for an acceptable matching of the electron track and the position of the supercluster. The constraint on $\Delta\eta_{in}$ is much stronger than the constraint on $\Delta\phi_{in}$, because of the larger spread of the electron's ECAL energy deposit in ϕ .

H/E

This variable is the ratio between the energy deposits in the HCAL and the ECAL. A definition is given in section 6.1. In general, electrons deposit most of their energy in the ECAL. Therefore H/E should be small for electrons.

$\sigma_{i\eta i\eta}$

The quantity $\sigma_{i\eta i\eta}$ is a measure for the spread of the ECAL energy deposit in η . The definition can be found in equation 9.1.

$$\sigma_{i\eta i\eta}^2 = \frac{\sum_i W_i \cdot \Delta\eta^2}{\sum_i W_i} \quad (9.1)$$

9. Event selection and efficiencies

Here, the index i loops over all crystals in a 5×5 grid around the seeding crystal. The quantity $\Delta\eta$ is the difference in η between crystal i and the crystal obtained by taking the average position of all crystals in the 5×5 grid, weighted according to the energy deposit in each crystal. It is calculated by multiplying the average crystal size in η with the difference between the two crystals expressed in units of number of crystals. The weights W_i used in equation 9.1 and for the calculation of the energy-weighted average position are defined as given in equation 6.1 using $W_0 = 4.7$. Since the electron hits in the ECAL have a small spread in η , $\sigma_{i\eta i\eta}$ is expected to be small for well reconstructed electrons. More details can be found in [89].

$E^{1 \times 5}/E^{5 \times 5}$ and $E^{2 \times 5}/E^{5 \times 5}$

$E^{i \times 5}/E^{5 \times 5}$ is another measure for the energy spread in η . $E^{5 \times 5}$ is the energy sum of all crystals in a 5×5 grid around the seed crystal. $E^{1 \times 5}$ ($E^{2 \times 5}$) is the sum of all crystal in a strip around the seed crystal containing 5 crystals in ϕ and 1 (2) crystals in η . For $E^{2 \times 5}$ two possibilities exists, from which the one with the higher value is chosen. Same as for $\sigma_{i\eta i\eta}$, $E^{i \times 5}/E^{5 \times 5}$ should be small for well reconstructed electrons. The electron is required to pass at least one of the two criteria.

Combined calorimetric isolation

The constraints on the isolation used for the HEEP ID are a mixture of ECAL and HCAL as well as relative and absolute isolations. The ECAL isolation is a sum of all the recorded hits in a cone of radius $\Delta R = 0.3$ around the position of the electron, excluding those hits which are in an inner cone of 3 crystals. Only hits with an $E_T > 0.08$ GeV in the barrel and $E_T > 0.1$ GeV in the endcap are considered. For the HCAL isolation the energy of all HCAL towers in a cone of radius $\Delta R = 0.3$ around the electron excluding an inner cone of radius $\Delta R = 0.15$ are summed up. The sum of ECAL and HCAL isolation is used for the selection. The ρ mentioned in table 9.3 is the mean energy density per effective area originating from pileup. Contributions from pileup will increase the value of the ECAL and HCAL isolation variables. This is accounted for by adding an estimate of the contribution from pileup to the maximum value allowed to pass the isolation.

Track isolation

The track isolation is defined as the sum of all tracks in a cone with radius $\Delta R = 0.3$ excluding an inner cone of radius $\Delta R = 0.04$. This quantity should be small for isolated electrons and helps to reject electrons which are part of a jet.

Inner layer lost hits

By extrapolating the track of the electron candidate back to the beam-line and counting the number of times the extrapolated track crosses active tracker material without a valid hit, the number of lost hits is obtained. The track of a prompt electron starts at the beam-line and is expected to produce hits in the inner tracker layers. The conversion of a prompt photon will most likely not occur directly at the primary vertex but somewhere inside the tracker. The track of an electron from such a conversion may not have any hits in the innermost tracker layers. Therefore, the number of lost hits can be used to reject electrons from photon conversion.

d_{xy}

The transverse impact parameter d_{xy} is defined as the minimal transverse distance of the electron track to the primary vertex. It can be used to reject electrons from photon conversion or electrons from the hadronization process of a jet. In contrast to prompt electrons, which are expected to have a small d_{xy} , the non-prompt electrons will on average have a greater distance to the primary vertex.

9.1.3. Signal specific selection

The following selection criteria exploit the characteristics of the signal explained in chapter 7. Since the signal signature contains exactly one high energetic and isolated electron, a veto is applied if a second HEEP electron with $E_T > 35$ GeV is found. This mainly helps to reduce the Drell-Yan background. The following two requirements make use of the two body decay kinematics explained in chapter 7. They have been studied, optimized and used in previous searches of the same kind [29, 61–64] and are used here without further investigation. The requirements are:

$$0.4 < \frac{E_T}{E_T^{\text{miss}}} < 1.5 \quad \Delta\Phi(e, \vec{E}_T) > 2.5$$

and will be called 'kinematic selection' in the following.

9.2. Selection efficiencies**9.2.1. Data - Monte Carlo scale factor**

The total selection efficiency is measured in data and simulation. To correct for possible differences between these two, a scale factor is applied to the simulated events. It is calculated by dividing the selection efficiency in data by the efficiency in Monte Carlo. The total selection efficiency can be factorized as shown in equation 9.2. Efficiency measurements are performed separately for barrel and endcap.

$$\epsilon = \epsilon_{GSF} \cdot \epsilon_{HEEP} \cdot \epsilon_{trigger} \quad (9.2)$$

Here ϵ_{GSF} is the efficiency to reconstruct a GSF electron from its supercluster in the ECAL. GSF electrons are electrons reconstructed as explained in section 6.1 using the GSF tracking algorithm. Since this efficiency is of general interest for all analysis using GSF electrons, ϵ_{GSF} is measured by the Egamma POG [90]. The results are given in table 9.4.

The second part ϵ_{heep} represents the efficiency for a GSF electron to pass the HEEP ID and is measured by the HEEP group [91]. The results for electrons with $E_T > 100$ GeV are shown in table 9.4. Since the evolution of the efficiencies at very high E_T is only known with low precision, an additional systematic uncertainty of 2% (4%) on the scale factor was added to the scale factor uncertainty following the recommendation of the HEEP group.

The last part $\epsilon_{trigger}$ refers to the efficiency of an HEEP electron to fire the single electron trigger used for this analysis. This was measured with 'tag and probe' techniques. An introduction to the tag and probe methodology can be found in [94]. For this efficiency measurement, tag and probe electrons are required to pass the HEEP

9. Event selection and efficiencies

Table 9.4. Summary of the efficiency measurements in data and Monte Carlo. The simulation is corrected by applying a data/Monte Carlo scale factor. Since the HLT_Ele80_CaloIdVT_TrkIdT is not simulated in Monte Carlo the scale factor is derived with respect to the HLT_Ele80_CaloIdVT_GsfTrkIdT trigger. (GSF reconstruction efficiencies from [90], HEEP efficiencies from [92], trigger efficiencies from [93])

Efficiency	Data	Monte Carlo	Data/MC
Barrel			
ϵ_{GSF}	0.985	0.989	0.995 ± 0.012
ϵ_{HEEP}	0.887	0.902	0.983 ± 0.027
$\epsilon_{\text{trigger}}$ (HLT_Ele80_CaloIdVT_TrkIdT)	0.974	not simulated	0.981 ± 0.006
$\epsilon_{\text{trigger}}$ (HLT_Ele80_CaloIdVT_GsfTrkIdT)	0.991	0.992	0.998 ± 0.002
combined data/MC scale factor			0.975 ± 0.023
Endcap			
ϵ_{GSF}	0.950	0.956	0.994 ± 0.015
ϵ_{HEEP}	0.907	0.922	0.984 ± 0.043
$\epsilon_{\text{trigger}}$ (HLT_Ele80_CaloIdVT_TrkIdT)	0.958	not simulated	0.975 ± 0.014
$\epsilon_{\text{trigger}}$ (HLT_Ele80_CaloIdVT_GsfTrkIdT)	0.976	0.983	0.993 ± 0.005
combined data/MC scale factor			0.970 ± 0.042

ID. If a probe electron fires the single electron trigger of interest, it is considered to be a passing probe. The very tight selection on tag and probe should ensure that the background contamination from non-Z events is negligible. This was cross-checked by evaluating the trigger efficiency with two methods. The first method is to simply count the events which passing or failing probes in the different invariant mass windows. Here, any background contamination can not be distinguished from $Z/\gamma^* \rightarrow ee$ events and will contribute to the efficiency measurement. For the second method, the invariant mass distribution of the tag and probe pairs is fitted with a combination of a Gaussian and an exponential function. The exponential function describes the background contribution, while the Gaussian part contains the $Z/\gamma^* \rightarrow ee$ events. This allows separating Z events from background events. The results of both methods are consistent within their uncertainties. For the high- E_T region, the number of available events is low. Since the fit done in the second method only works well if there is a sufficient amount of events, the first method is chosen here. The difference between them on the resulting data Monte Carlo scale factor is used as uncertainty. Additional information about the determination of the trigger efficiency can be found in [93].

The trigger efficiency curves versus the transverse energy of the electron can be seen in figure 9.1. The trigger efficiency curves have a sharp turn on at $E_T = 80$ GeV and reach a plateau region at around $E_T = 100$ GeV, where the efficiency stays flat versus E_T . In the plateau region, the trigger efficiencies in data and Monte Carlo agree to the level of 1-2%. In the turn-on region however, the differences are larger. Since this search is interested in the high E_T regime, it is suitable to apply an offline E_T cut of 100 GeV in order to avoid the turn-on region. For each efficiency curve, the region above 100 GeV

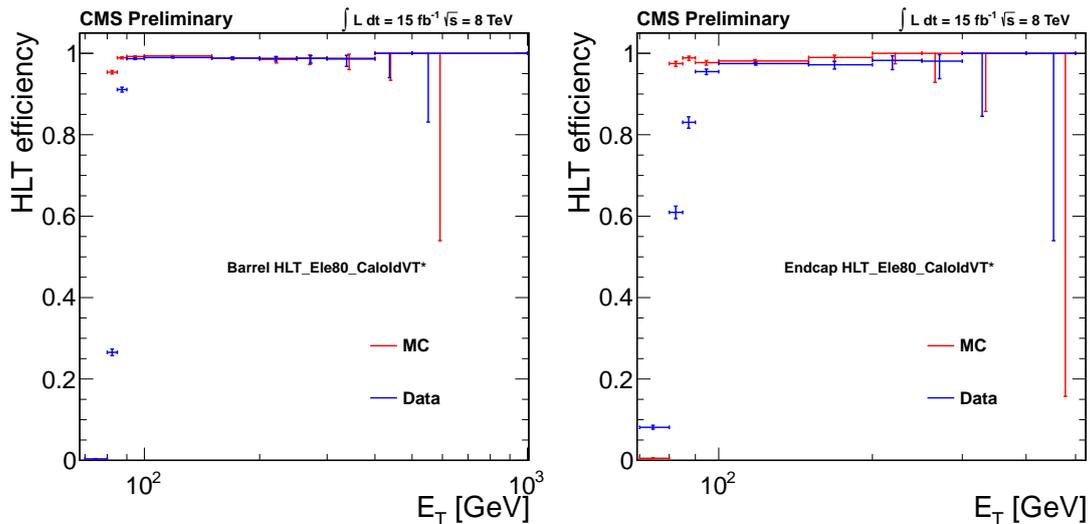


Figure 9.1. Trigger efficiency curves as function of E_T measured in data and Monte Carlo for barrel (left) and endcap (right) electrons. The efficiencies stay flat versus E_T for $E_T > 100$ GeV. Both plots by K. Padeken [93].

is fitted with a constant term. All results are summarized in table 9.4.

The data/Monte Carlo scale factors are also shown in table 9.4. For the total efficiency, the two triggers are combined by weighting them with the integrated luminosity of the run period in which they were active. The combined data/Monte Carlo scale factor for all three components in equation 9.2 is 0.975 ± 0.023 (0.970 ± 0.045) in the barrel (endcap).

9.2.2. Background rejection efficiency

The selection efficiencies for different Standard Model processes can be found in table 9.5. Starting point for the calculation of the efficiencies is the number of events passing the preselection. For each process and selection step, the first number specifies the efficiency with respect to the previous selection step while the number in brackets is the efficiency relative to number of events after the preselection.

Events from $W \rightarrow e\nu$, $Z/\gamma^* \rightarrow ee$ and diboson events contain prompt, isolated electrons which are assumed to pass the HEEP selection efficiently. The efficiencies for these processes are above 90% and agree well with this assumption. Around 84% of the QCD events passing the preselection are rejected by the HEEP requirements. For the other processes the HEEP selection efficiency is between 69% and 88%. The veto for events with a second HEEP electron should help to reject events from processes with more than one HEEP electron. This affects mainly $Z/\gamma^* \rightarrow ee$, where around 25% of the events passing the HEEP ID are rejected, but also $Z/\gamma^* \rightarrow \tau\tau$, diboson as well as $t\bar{t}$ and single top events.

The efficiency for $W \rightarrow e\nu$ events to pass the E_T/E_T^{miss} ($\Delta\Phi(e, \vec{E}_T)$) selection is around 12% (41%). The amount of W events passing the preselection does not only contain off-shell W events but also events from the W mass peak. On-shell W events are produced with a much higher probability than events from the off-shell mass tail. Most of these events are rejected by the electron E_T threshold of 100 GeV. However, if an on-shell W recoils against a high- p_T jet, the boost in transverse direction could lead to electrons

9. Event selection and efficiencies

Table 9.5. Event selection efficiency for different Standard Model processes. The first number for each entry is the efficiency calculated with respect to the previous selection step. The efficiency in brackets is calculated relative to the number of events which pass the preselection.

Process	HEEP ID	2nd e veto	E_T/E_T^{miss}	$\Delta\Phi(e, \vec{E}_T)$
$W \rightarrow e\nu$	0.923 (0.923)	> 0.999 (0.922)	0.118 (0.109)	0.407 (0.045)
$W \rightarrow \tau\nu$	0.706 (0.706)	0.995 (0.703)	0.360 (0.253)	0.174 (0.044)
QCD	0.163 (0.163)	> 0.999 (0.162)	0.003 (0.001)	0.247 (< 0.001)
$t\bar{t}$ + single top	0.877 (0.877)	0.950 (0.833)	0.257 (0.214)	0.204 (0.044)
$DY \rightarrow ee$	0.928 (0.928)	0.748 (0.694)	0.004 (0.003)	0.611 (0.002)
$DY \rightarrow \tau\tau$	0.711 (0.711)	0.974 (0.693)	0.321 (0.222)	0.109 (0.024)
γ + jets	0.691 (0.691)	> 0.999 (0.691)	0.002 (0.002)	0.403 (0.001)
Diboson	0.911 (0.911)	0.877 (0.799)	0.216 (0.172)	0.498 (0.086)

with $E_T > 100$ GeV. In these events the two-body decay kinematics do not apply and the E_T/E_T^{miss} ratio as well as the $\Delta\Phi(e, \vec{E}_T)$ constraint will reject these events. Using only the high- p_T tails with events fulfilling $m_W > 200$ GeV the corresponding efficiency to pass the E_T/E_T^{miss} ($\Delta\Phi(e, \vec{E}_T)$) selection is around 93% (96%).

Since the E_T/E_T^{miss} selection is an implicit E_T^{miss} cut ($E_T^{\text{miss}} \cdot 1.5 > E_T$), it will efficiently reject events with no physical source of E_T^{miss} (e.g. a neutrino) besides misreconstruction. This applies to QCD, $Z/\gamma^* \rightarrow ee$ and γ + Jets events, where more than 99% of the remaining events are rejected.

In figure 9.2 the number of expected events after the full selection for two different M_T thresholds are shown. As expected, the $W \rightarrow \ell\nu$ process is by far the most important process contributing around 81% of the Standard Model events in the $e + E_T^{\text{miss}}$ spectrum above $M_T = 220$ GeV. For an M_T threshold of 1500 GeV the process is even more important and contributes 97% of the Standard Model events.

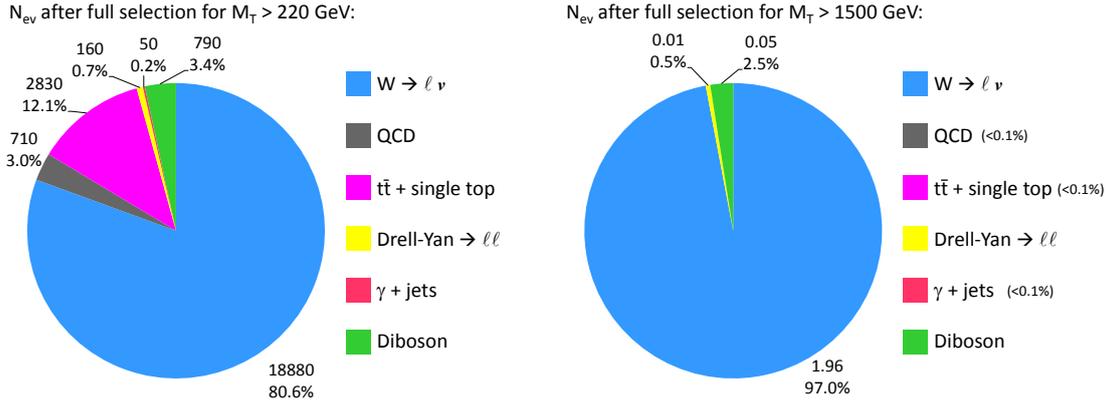


Figure 9.2. Number of expected Standard Model events after full selection for two different M_T thresholds and scaled to 19.7 fb^{-1} . Also given is the relative fraction of events for each process with respect to the total number of events. The expected number of QCD events is taken from the data-driven estimate explained in chapter 10.

10 Estimation of the multi-jet contribution derived from data

The description of QCD multi-jet processes via Monte Carlo simulations suffers from several limitations. The term 'QCD' refers not to one specific process, but to a large amount of different processes involving the strong interaction. This complicates the simulation of the QCD background. On top of that, the simulation is only available at LO accuracy. Due to the very high cross section, it is hardly possible to produce a sufficient amount of Monte Carlo events which is needed for an accurate description of the roughly 20 fb^{-1} of data (e.g. the $35 \cdot 10^6$ events generated for QCD processes with $80 \text{ GeV} < \hat{p}_T < 170 \text{ GeV}$ correspond to an integrated luminosity of only 0.19 fb^{-1}).

As explained in section 6.1.2, electrons are reconstructed by searching for energy deposits in the ECAL which resemble the pattern of an electron and looking for an appropriate track matching the ECAL hit. Jets lead to a large number of tracks and will deposit energy in the ECAL on their way to the HCAL. Therefore, it is plausible that a jet can fake the electron signature. Even though the fake probability is small, the multi-jet contribution to the $e + E_T^{\text{miss}}$ spectrum is of interest for this analysis since the number of multi-jet events is very large.

To improve the prediction of the multi-jet contribution, a data-driven approach can be used. Data-driven methods try to estimate the QCD contribution in the region of interest (here the $e + E_T^{\text{miss}}$ spectrum after full selection) by using data events which are not part of this region. Using these events, the multi-jet contribution can be extrapolated to the region of interest. In principle, two ingredients are needed. First, the shape of the QCD contribution in the region of interest has to be determined. This leads to the task of selecting an appropriate template of data events. Second, the normalization for this template is needed. In this chapter, the chosen method is explained followed by a presentation of the results.

10.1. Methodology

The principle of the method which is applied here is called 'ABCD method'. This kind of method is well established and commonly used. An example is the data-driven QCD method for the $W' \rightarrow \tau\nu$ analysis [64]. Events in the QCD template should be as similar as possible to QCD events in the region of interest but must not be part of it. The first part is achieved by requiring template events to pass as many selection steps as possible (see 10.3 for details). To distinguish between region of interest and template region the quality of the electron candidate is used. The electron candidate of template events is asked to fail the HEEP ID which ensures the distinction from the region of interest. Most of the electron candidates from QCD events are rejected by the HEEP ID (84% for events passing the preselection (see 9.2.2)). Therefore, the template will provide a large amount of QCD events which can be used to model the contribution in the region of interest. Since the amount of QCD events in the template will be much larger than the amount of QCD events in the region of interest, it can not directly be used but has

10. Estimation of the multi-jet contribution derived from data

to be scaled with a normalization factor. This factor can be determined with help of a tight-to-loose (ttl) ratio measurement and will be explained in section 10.3. The ttl-ratio measurement has to be performed with a dataset which is independent of the region of interest and the template region. To obtain an independent dataset, the E_T/E_T^{miss} ratio is used. Template events and events in the region of interest pass the E_T/E_T^{miss} cut of the event selection ($0.4 < E_T/E_T^{\text{miss}} < 1.5$). The ttl-ratio is measured with events where the E_T/E_T^{miss} ratio fulfills $1.5 < E_T/E_T^{\text{miss}} < 10$.

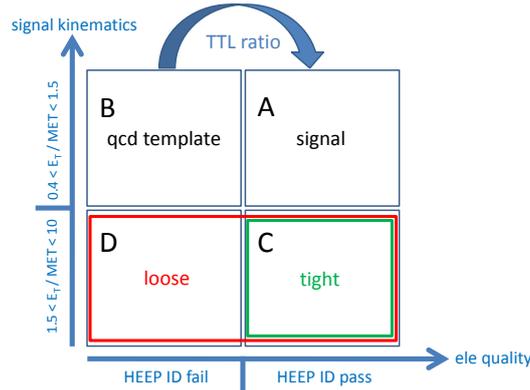


Figure 10.1. Sketch of the ABCD method.

A sketch of the procedure can be found in figure 10.1. Shown there is the classification of events into four regions (A,B,C,D) by means of the quality of the electron candidate (horizontal axis) and the E_T/E_T^{miss} ratio (vertical axis). Events in region A fulfill the E_T/E_T^{miss} selection and the electron candidate passes the HEEP ID. This region contains all the events which will contribute to the $e + E_T^{\text{miss}}$ spectrum after the full selection. It therefore is the region of interest (or 'signal region') for this analysis. The template is made from events in region B, which is disjunct from region A. Region B contains events which pass the E_T/E_T^{miss} cut, but where the electron candidate fails the HEEP ID. The E_T/E_T^{miss} criterion ensures that the template events contain a substantial amount of E_T^{miss} . This is important because the events in the QCD template should be as similar as possible to events from the region A. Well reconstructed dijet events with a negligible amount of E_T^{miss} for example will not pass the full event selection (especially the E_T/E_T^{miss} cut), even if one of the two jets would pass the HEEP selection. Thus events of this kind are not of interest for the QCD template. The ttl-ratio measurement is done with events from region C and D. Region C and D only contain events failing the E_T/E_T^{miss} selection and are thus independent of region A and B. Since QCD events generally do not have a physical source of E_T^{miss} but misreconstruction, region C and D will contain a large amount of QCD events. This is needed for an accurate measurement of the ttl-ratio.

10.2. Measurement of the ttl-ratio

A definition of the ttl-ratio r_{ttl} is given in equation 10.1.

$$r_{\text{ttl}} = \frac{t}{\bar{t}} = \frac{t}{t + \bar{t}} \quad (10.1)$$

Here, t is the number of events in region C ('tight' events), \bar{t} is the number of events

in region D ('not-tight' events) and l is the number of 'loose' events, which is the sum of tight and not-tight events. The ttl-ratio can be interpreted as the probability for a loose event to be tight and therefore as probability for an electron candidate from a QCD multi-jet event to pass the HEEP ID. To be considered as loose, an event has to fulfill the following criteria:

- pass preselection (see section 9.1.1)
 - electron candidate has $E_T > 100$ GeV
 - electron candidate fulfills trigger requirements (see table 9.1)
 - pass vertex requirements (see section 9.1.1)
 - pass eventfilter (see table 9.2)
- $1.5 < E_T/E_T^{\text{miss}} < 10$
- $\Delta\Phi(e, \vec{E}_T) > 2.5$
- $N_{\text{gsf,ele}}$ with $E_T > 10$ GeV = 1

In tight events, the electron candidate has - in addition to the loose criteria - to pass the HEEP ID and thus to be an isolated high quality candidate. The measurement of the ttl-ratio is done with the same dataset which is used for the analysis. The separation of region A/B and C/B by means of the E_T/E_T^{miss} criterion ensures that the subset of data used for the ttl-ratio measurement and the subsets in the signal and template region are independent. The lower E_T/E_T^{miss} bound for region C and D is given by the upper bound of the E_T/E_T^{miss} selection defining region A and B. The upper bound is chosen to be $E_T/E_T^{\text{miss}} < 10$ in order to have a sufficient amount of events for the ttl-ratio measurement, but to avoid events from the tail of the E_T/E_T^{miss} distribution. This helps to ensure that events from region C and D are as similar as possible to events from region A and B. Region C and D will not only contain QCD events, but also contaminations from other processes with real electrons and photons. The most important contamination arises from $W + \text{Jets}$ events, where the recoil against a jet can lead to events with a large E_T/E_T^{miss} value. In these events the electron will easily pass the HEEP ID which will lead to an artificial increase of the ttl-ratio since most of the $W + \text{Jets} \rightarrow e\nu$ events will be tight. Another important contamination is due to $\gamma + \text{Jets}$. Photons passing the preselection (and thus already misidentified as gsf electron) will have a non negligible chance ($\approx 69\%$ see section 9.2.2) to pass the HEEP selection. The contamination from $Z/\gamma^* \rightarrow ee$ is highly suppressed by requiring the number of gsf electron candidates with $E_T > 10$ GeV to be exactly 1. Additional small contaminations can arise from $t\bar{t}$, single top and diboson events. The contamination by all non-QCD processes is accounted for by direct Monte Carlo subtraction.

The results from the ttl-ratio measurement as function of E_T are summarized in figure 10.2-top. Shown is the ttl-ratio before and after Monte Carlo subtraction as function of E_T in three different η bins. The ttl-ratio decreases as function of E_T and becomes constant at around $E_T = 200 - 300$ GeV. For the constant part, it is around 7% in the barrel rising up to 25% for the outermost η regions of the ECAL endcap. For further usage and to smoothen statistical fluctuations, the ttl-ratio is parametrized with a second order polynomial for the low E_T region and a constant for the high E_T region. The form of the parametrization has no physical motivation besides describing the ttl-ratio well.

10. Estimation of the multi-jet contribution derived from data

To validate the $t\bar{t}$ -ratio measurement, three cross-checks have been applied. The first cross-check is to vary the upper E_T/E_T^{miss} boundary used to define region C and D between 8 and 12 and the lower bound between 1.5 and 2. Lowering (increasing) the upper E_T/E_T^{miss} bound will decrease (increase) the statistics available for the $t\bar{t}$ -ratio measurement, but also slightly ($\approx 1\%$) decrease (increase) the contamination from electroweak processes. The result of the $t\bar{t}$ -ratio measurement should nevertheless be the same. The variation of the $t\bar{t}$ -ratio due to the E_T/E_T^{miss} bound variation is shown in figure 10.2-bottom as grey shaded band. The agreement is within 5% for most of the E_T regime and within 30% for the whole E_T range. As a second cross-check, the cross-sections for the subtracted Monte Carlo samples are varied. The most important contributions arise from $W + \text{Jets}$ and $\gamma + \text{Jets}$. Following the $H \rightarrow \gamma\gamma$ analysis [95] a k -factor of 1.3 derived from data is applied to the $\gamma + \text{Jets}$ samples. The cross-section for this process is varied within $\pm 30\%$ around the corrected value. For the $W \rightarrow \ell\nu$ process the cross-section is varied by 10%, which corresponds to the uncertainty assigned to the higher order corrections for this process at high M_T (see 8.2). The cross-section for the other processes are left unchanged, since their impact is small. The resulting $t\bar{t}$ -ratio measurements are reparametrized and the result can be seen in figure 10.2-bottom as dashed purple lines. In the barrel the result changes within 15% while in the endcap the result varies within 8% ($|\eta| < 2$) and 5% ($|\eta| > 2$) respectively. The change in the barrel is larger, due to the higher contamination from non-QCD processes. The last cross-check is to determine the $t\bar{t}$ -ratio using QCD Monte Carlo simulations. Despite the fact, that the QCD simulation is far from being perfect and the number of simulated events is insufficient, the $t\bar{t}$ -ratio from QCD Monte Carlo can still be used as a rough cross-check. Since the QCD Monte Carlo does not contain any contamination from other processes, it also serves also as cross-check for the Monte Carlo subtraction. The result is shown in figure 10.2-bottom. The $t\bar{t}$ -ratio determined with QCD Monte Carlo shows stronger fluctuations than the measurement. In some bins the agreement of the $t\bar{t}$ -ratio from data and QCD Monte Carlo it is only within 30% and in one bin only within 50%. However, considering the large statistical uncertainties from the QCD Monte Carlo the overall agreement is acceptable.

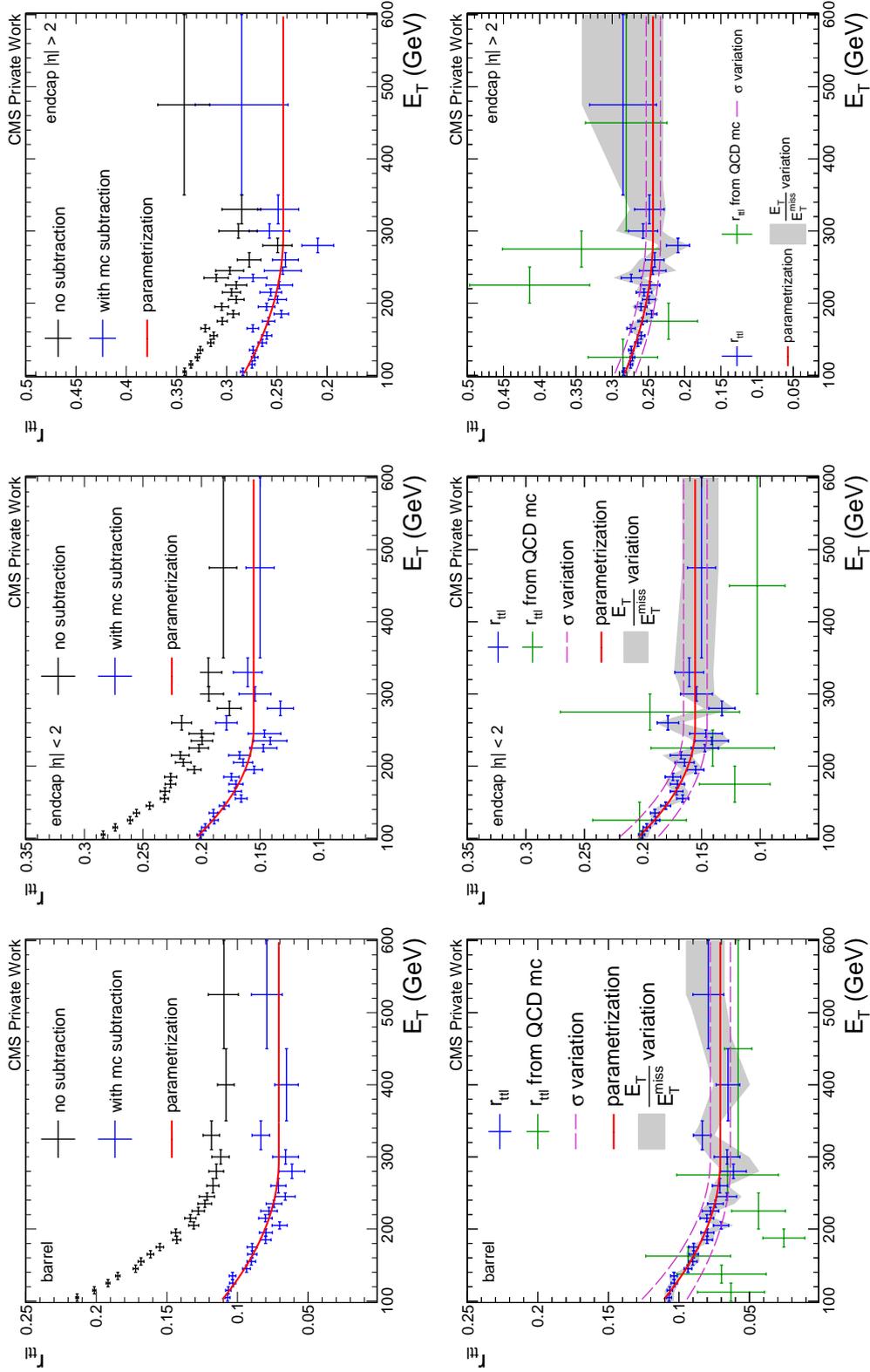


Figure 10.2. Top: Measurement of the $t\bar{t}l$ -ratio as function of E_T in three η bins before and after the subtraction of non QCD contamination. Bottom: Results from the $t\bar{t}l$ -ratio measurements together with performed cross-checks. The gray shaded band comes from the variation of the $t\bar{t}l$ -ratio when shifting the upper (lower) E_T/E_T^{miss} border of region C and D between 8 (1.5) and 12 (2). Shifting the cross-section for the $W \rightarrow \ell\nu$ and the $\gamma + \text{Jets}$ process by $\pm 10\%$ and $\pm 30\%$ respectively leads to the purple lines (after parametrization). The green points are the result of the $t\bar{t}l$ -ratio obtained using QCD Monte Carlo.

10.3. Selection of the QCD template

The QCD template consists of events from region B. Events in this region have to fulfill the following criteria:

- pass preselection (same as in section 10.2)
- electron candidate fails HEEP ID
- $0.4 > E_T/E_T^{\text{miss}} > 1.5$
- $\Delta\Phi(e, \vec{E}_T) > 2.5$
- $N_{\text{HEEP,ele}}$ with $E_T > 35 \text{ GeV} = 0$

To derive a scaling factor based on the results from 10.2, the electron candidate of the template event has to fulfill the same criteria as the electron candidate in loose events from region C or D. Following the definition of the ABCD plane, it is evident that region B events have to pass the E_T/E_T^{miss} selection and that the electron candidate has to fail the HEEP ID. The resulting template does not have the correct normalization to serve as a QCD estimate for region A. The normalization is achieved by weighting each event in the template with a scaling factor ϵ , which is defined in equation 10.2.

$$\epsilon(E_T, \eta) = \frac{r_{t\bar{t}}}{1 - r_{t\bar{t}}} \quad (10.2)$$

This formula can be obtained by solving equation 10.1 for t/\bar{t} , which is the ratio of tight-to-not-tight events. The scaling factor depends on E_T and η of the electron candidate. The next step is to proceed with the weighted events as if the electron candidate would have passed the HEEP ID. The selection contains a veto, if more than one HEEP electron is found in the event. Therefore, events containing one or more HEEP electrons should not be part of the template, even if they have an additional electron candidate fulfilling the region B criteria mentioned above. This is ensured by the requirement, that region B event do not contain any electron candidate passing the HEEP ID. In rare cases, events can contain more than one candidate fulfilling the requirements of region B. A priori each of the candidates has a certain probability to pass the HEEP ID. If more than one candidate passes the HEEP ID at the same time the event would be rejected by the second electron veto. Therefore, using the event once per candidate weighted with the scale-factor from equation 10.2 will lead to an overestimation of the contribution from this event. Using only one of the candidates scaled according to equation 10.2 will result in an underestimation and introduce the ambiguity of choosing one of the candidates. To resolve this, the event is used with each of the candidates but scaled with a reduced scale-factor. This factor is given in equation 10.3, where i labels the current candidate and j loops over the other candidates.

$$\epsilon_{\text{mod}} = \epsilon(E_{T,i}, \eta_i) \cdot \prod_{j \neq i} (1 - \epsilon(E_{T,j}, \eta_j)) \quad (10.3)$$

The impact of this reduced scale-factor on the total result is negligible since the fraction of events with more than one candidate is very small ($< 1\%$). All the scaled events in region B pass the signal specific selection (see section 9.1.3) by definition and form the data-driven estimate of the QCD background in the $e + E_T^{\text{miss}}$ spectrum after the

full selection. The template will contain contamination from non QCD processes. The same procedure described here for region B events, is applied to the Monte Carlo of contaminating processes. The resulting contamination is subtracted from the final QCD template after scaling. Similar to the $t\bar{t}$ -ratio measurement, the cross-sections of the $W \rightarrow \ell\nu$ and $\gamma + \text{Jets}$ process have been varied and the impact on the resulting template was found to be within 10%. As an additional cross-check the template is recreated using the $t\bar{t}$ -ratio without the parametrization. For M_T values smaller than 500 GeV, the change of the template is within 3%. For larger M_T the differences are larger but still within 10% for most of the M_T region. The larger difference for higher M_T is expected, since the fluctuations of the $t\bar{t}$ -ratio around the parametrization is larger for high E_T than for small E_T .

10.4. Results

The resulting QCD contribution to the signal region is shown in figure 10.3. The data-driven method provides a description of the QCD background up to high M_T . The fraction of QCD events compared to the total Standard Model prediction at low M_T ($M_T \approx 220$ GeV) is around 5% (see figure 13.2), which is compatible with the data-driven estimate obtained with a different method in a previous search of the same kind [62].

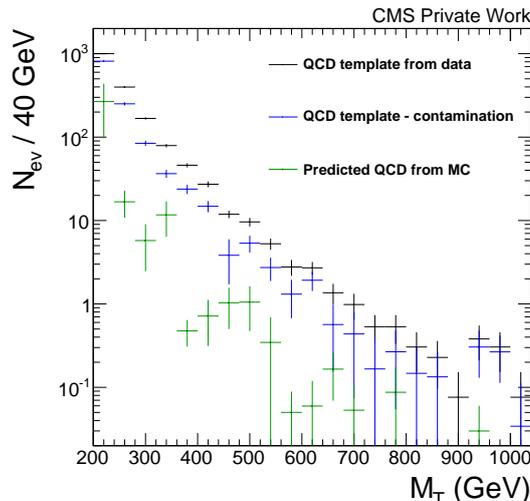


Figure 10.3. Data-driven estimation of the QCD background after full selection. Also shown is the QCD template before the subtraction of the contamination from non-QCD processes as well as the QCD contribution predicted from Monte Carlo .

In order to assign a total systematic uncertainty to the data-driven estimate, the various cross-checks are used. The variation of the cross-section for $t\bar{t}$ -ratio and template is fully correlated, while the E_T/E_T^{miss} variation and the uncertainty due to the parametrization are assumed to be uncorrelated. A total uncertainty of 40% is applied as a conservative estimate. Based on the cross-checks, this is the level to which the QCD estimate is understood.

11 Parametrization of the $e + E_T^{\text{miss}}$ spectrum

The predicted contribution of Standard Model processes to the $e + E_T^{\text{miss}}$ spectrum is shown in figure 11.1-top. It is a steeply falling curve as function of M_T . Even though a lot of effort was made to provide a sufficient amount of generated events, some statistical fluctuations still occur. Especially for very high M_T (> 3 TeV) the number of simulated events is low and the statistical fluctuations get more and more prominent. In order to obtain a smooth description of the Standard Model prediction over the whole M_T range of interest, the spectrum is parametrized with a function. A good description of the Standard Model contribution is essential for a comparison with the measured data using statistical methods (see chapter 15-18 for details). The chosen function is given in equation 11.1.

$$f(x) = e^{a+b \cdot x+c \cdot x^2} \cdot x^d \quad (11.1)$$

The non-parametrized prediction of the Standard Model contribution to the $e + E_T^{\text{miss}}$ spectrum will be called Monte Carlo prediction here, even if the description of multi-jet processes is based on a data-driven approach. Due to the E_T cut at 100 GeV the M_T spectrum flattens for M_T values around 200 GeV. The function given in equation 11.1 does not describe this behavior. Therefore, the beginning of the M_T spectrum is fitted with a Landau function which is found to describe the beginning of the spectrum well. The transition from the Landau function to the function f is done at the point where the difference between the functions is minimal. The point of minimal difference is at 293 GeV. In this region both functions describe the M_T spectrum well. Figure 11.2-left shows the beginning of the spectrum with the two functions. Figure 11.2-right shows the ratio plot for the combined function in this region. The transition from the first to the second function is smooth and does not produce a kink in the parametrization. The choice of the functions has no additional motivation besides providing a good description of the spectrum.

The combined result for the M_T range from 220 GeV to 4000 GeV can be seen in figure 11.1-top. Figure 11.1-mid shows the difference between the Monte Carlo and the fit divided by the value from the fit as a function of M_T . In figure 11.1-bottom a cumulative distribution is shown for fit and Monte Carlo. In this distribution the number of events integrated from a lower M_T threshold to up to 4000 GeV is plotted against the M_T threshold. With the exception of the region between 3000 GeV and 4000 GeV which is summarized in one bin, the binning for the ratio and the cumulative plot is the same as for the main plot showing the fit and the Monte Carlo. All three distributions indicate a very good agreement between the Monte Carlo and the fit. The resulting parameters from the fit are summarized in table 11.1.

11. Parametrization of the $e + E_T^{\text{miss}}$ spectrum

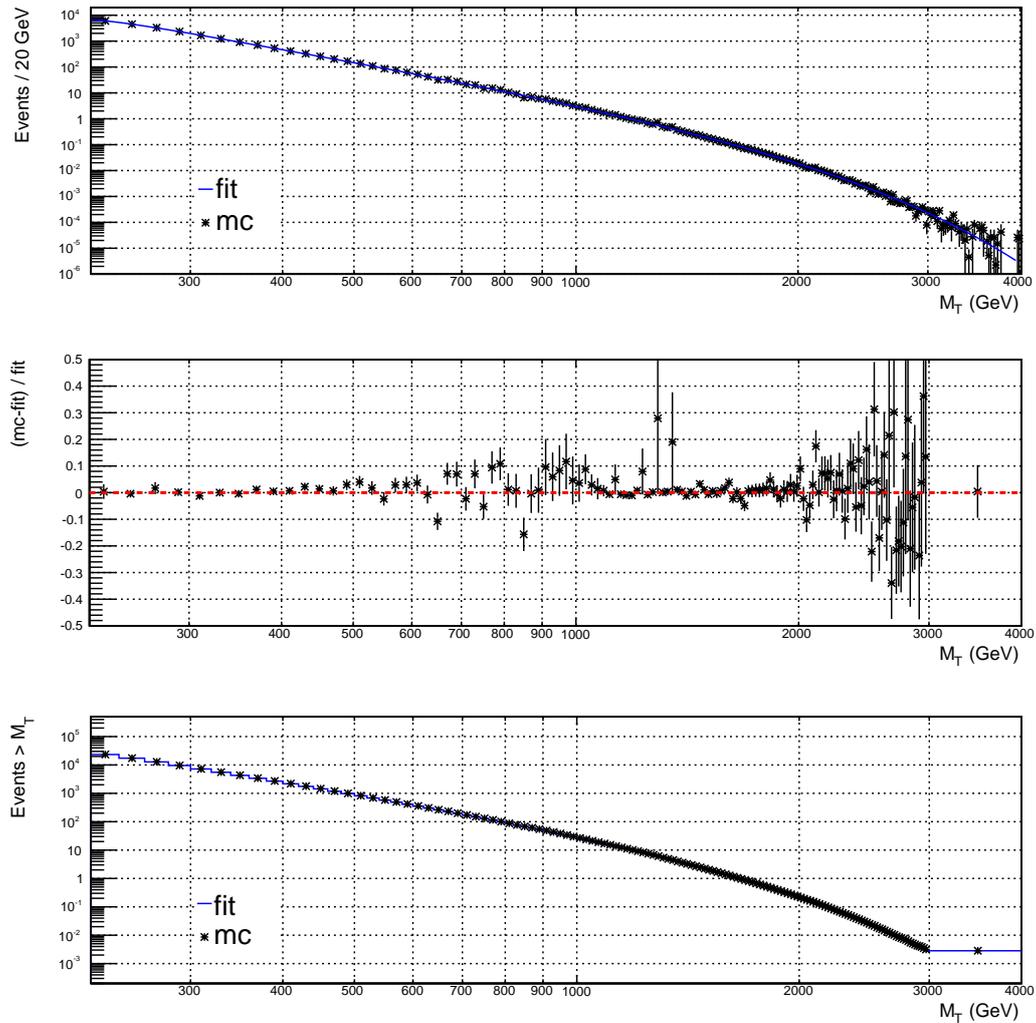


Figure 11.1. Top: Prediction of the Standard Model contribution to the $e + E_T^{\text{miss}}$ spectrum. Also shown is the combined parametrization of the Monte Carlo. Middle: Ratio of the difference between Monte Carlo and parametrization to the parametrization. Bottom: Cumulative distribution of Monte Carlo and parametrization, showing the number of events integrated from the lower M_T threshold given on the x-axis up to 4000 GeV.

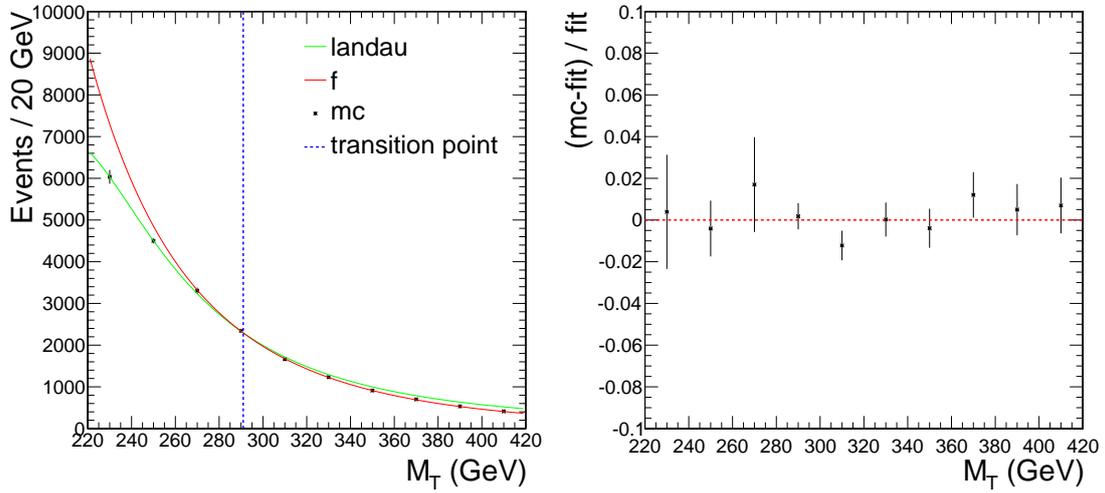


Figure 11.2. Left: Standard Model prediction for low M_T together with the fit of the Landau function and the function f to the Monte Carlo. The blue line marks the point of minimal difference, at which the transition from the Landau function to the function f is made. Right: Ratio of the difference between the Monte Carlo and the combined fit to the combined fit.

Table 11.1. Resulting parameters from the fit to the Monte Carlo for both functions. Also given is the χ^2/N_{DoF} of the fits.

f		Landau	
Parameter	Value	Parameter	Value
a	34.44	constant	3.83e+04
b	-8.84e-04	mpv	2.15e+02
c	-3.21e-07	sigma	20.77
d	-4.66		
χ^2/N_{DoF}	1.30	χ^2/N_{DoF}	0.68

12 Systematic uncertainties

Systematic uncertainties on the Standard Model contribution to the $e + E_T^{\text{miss}}$ spectrum arise from various sources. One part of these uncertainties is related to the theoretical input needed for the event generation. This includes uncertainties on PDFs, cross sections and higher order corrections. The other uncertainties originate from the experimental side. The precise knowledge of the calibration and performance of the detector is crucial for the simulation of collision events. For example, uncertainties on the luminosity measurement, the knowledge of the electron energy scale or reconstruction and selection efficiencies do have a direct impact on the prediction of the Standard Model contribution to the $e + E_T^{\text{miss}}$ spectrum. In the following, the sources of systematic uncertainties considered for this analysis are listed and explained. With the exception of the luminosity (see below for explanation), each systematic uncertainty can change the normalization as well as the shape of the M_T distribution. These uncertainties will be called shape uncertainties. For each shape uncertainty listed below, the quantity in question is varied according to its uncertainty and the M_T spectrum is reevaluated. The shifted M_T distributions will be fitted as explained in chapter 11 and compared to the unshifted fit.

Luminosity

The luminosity is measured with the pixel cluster counting method (see [96] for details). It is an essential ingredient for the calculation of the normalization of the Monte Carlo simulations. A larger (smaller) luminosity would lead to a constant upward (downward) shift of the Standard Model expectation but will not change the shape of the M_T distribution. The official uncertainty on the luminosity measurement with the pixel cluster counting method is 2.6% [96].

Electron energy measurement

For the electron energy measurement, two different sources of uncertainty are considered. These two sources are the electron energy resolution and the electron energy scale. The difference between the two can be explained with the following example. Consider a set of electrons with an energy of exactly 100 GeV. Each of the electrons is measured with the CMS detector. In general, the measured energy for the individual measurements will not be exactly 100 GeV but spread around 100. This spread is the result of the limited energy resolution of the measurement. The energy scale refers to the overall calibration of the energy measurement. In the example, an offset of +1 GeV on the energy scale would lead to the electron energy measurements spreading around 101 GeV as central value instead of 100 GeV. From fits to the Z peak in data and Monte Carlo, the electron energy resolution can be extracted. The resolution is found to be slightly worse in data compared to the resolution in Monte Carlo. The effect of this difference on the $e + E_T^{\text{miss}}$ spectrum can be quantified by adding an additional smearing to the electron energy resolution in the Monte Carlo. Based on [97], a value of 1.3% (2.8%) is used for barrel (endcap) electrons. The official uncertainty on the electron energy scale is 0.6% (1.5%)

12. Systematic uncertainties

in the barrel (endcaps) [98]. This is consistent with the shift of the energy scale obtained from fits to the Z peak in data and Monte Carlo [97].

E_T^{miss}

As explained in section 6.2, E_T^{miss} is a composite object. Therefore, it is not trivial to assign an uncertainty to the E_T^{miss} measurement. The official CMS recommendation is to decompose the E_T^{miss} into its components (jets, unclustered energy, τ , e and γ , μ), shift each component by its uncertainty and recalculate the E_T^{miss} with the shifted components [99]. This is done using the E_T^{miss} uncertainty tool provided by the JetMET POG [100]. The shift of the E_T^{miss} due to the electron energy scale is fully correlated with the uncertainty on the electron energy measurement due to the energy scale and has to be treated simultaneously. The same applies to the electron energy resolution.

Pileup reweighting

The uncertainty due to the pileup reweighting is incorporated following [101]. The method used to determine the number of collision events per bunch crossing needs the total inelastic proton proton cross-section as input (see section 8.2.2 for details). The cross section used for the reweighting is 69.4 mb, which is the extrapolation of the 7 TeV measurement to 8 TeV. $Z \rightarrow \mu\mu$ studies performed in 2012 show a better data - Monte Carlo agreement for the number of reconstructed vertices, if a cross section of 73.5 mb is used. The pileup reweighting is redone using the cross section from the $Z \rightarrow \mu\mu$ studies and the difference to the original M_T distribution is used as systematic uncertainty on the pileup reweighting.

QCD data-driven

As explained in section 10.3, a total uncertainty of 40% is assumed on the data-driven estimation of the multijet background.

Data-mc scale factor

The uncertainty on the scale factor presented in section 9.2.1 is 0.023 (0.042) in the barrel (endcap).

PDF

PDFs are an important ingredient for the Monte Carlo production. They have to be extracted from a large set of deep inelastic scattering data measured by various experiments. A PDF is obtained by performing a global fit to the data mentioned before and can then be extrapolated to the kinematic regions which are of interest at the LHC. There are multiple collaborations providing PDFs. The details of the method used to obtain the PDF differs between each group. Therefore, the PDFs are not identical but will disagree to some extent. The collaborations do not only provide a single PDF but PDF sets. A PDF set is made of a large number of different PDFs. Each set provides a central member representing the best fit to the data. The other members are obtained from variations of the fitted parameters and can be used to quantify the uncertainty on the central member due to uncertainties on the fit. Beside the fit uncertainties, additional uncertainties arise from external parameters of the PDF. These parameters are

fixed and not determined by the fit. For the PDF sets which are utilized here (see below), the most important one is the strong coupling constant α_s . To study impact of α_s uncertainties, the PDF sets are provided for different values of α_s .

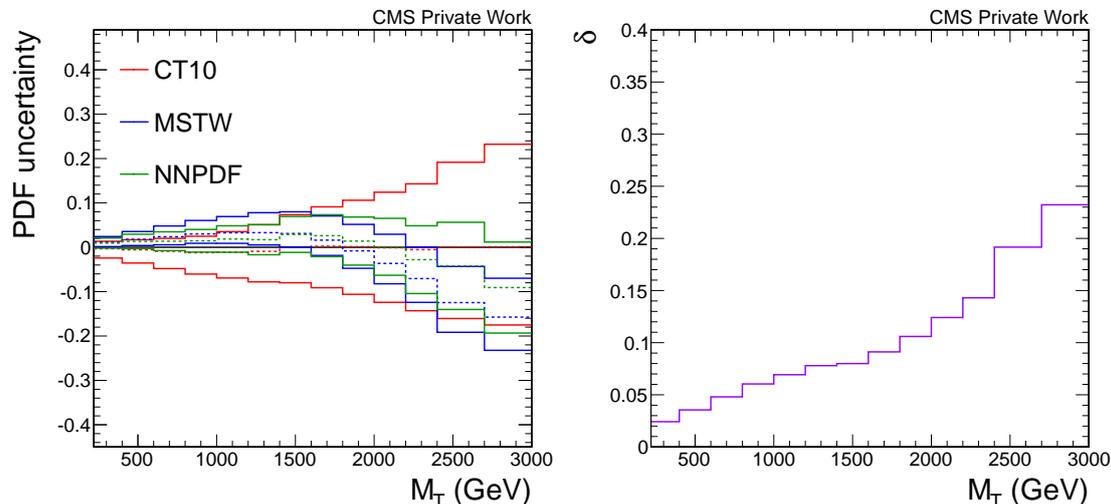


Figure 12.1. Left: Combined PDF+ α_s uncertainties derived with three different PDF sets (CT10, MSTW2008, NNPDF 2.3) following the PDF4LHC recipe as a function of M_T . The curves are normalized to the central value M . The dashed lines represent the mean values of the three PDF sets. Right: Percentual width of the envelope from the three PDF sets in left figure as a function of M_T .

In [102] two different methods are proposed to evaluate uncertainties due to PDFs. The first approach is to regenerate Monte Carlo samples for each PDF member and investigate the changes in the observable of interest for the different members using the additional Monte Carlo samples. Due to the large number of different PDF sets and the even larger number of members per PDF set this approach is not very practical. The idea of the second method is to apply a reweighting procedure to the existing Monte Carlo samples which were produced with one specific PDF. This allows to study the impact of PDF uncertainties on the observable of interest without having to produce additional Monte Carlo samples. For each event and each member of the PDF set under study, a weight is calculated based on the truth information about the x and Q^2 of the event. In the case of NNPDF, a small fraction of these weights will be negative. For high M_T , where the Monte Carlo statistics is low, this can lead to unphysical fluctuations in the reweighting procedure. Therefore, negative weights are set to zero. The weight is calculated according to equation 12.1, where n labels the events, S_0 labels the original PDF and S_i labels the new PDF.

$$W_n^i = \frac{f(x_1, Q^2; S_i) f(x_2, Q^2; S_i)}{f(x_1, Q^2; S_0) f(x_2, Q^2; S_0)} \quad (12.1)$$

The implementation of the PDF+ α_s uncertainty calculation is done according to the PDF4LHC [103, 104] recipe. To access the different PDF sets, the LHAPDF [105] interface is used. The PDF sets considered for the uncertainty calculation are CT10 [77], MSTW2008 [16, 106] and NNPDF 2.3 [107]. The PDF4LHC recipe gives appropriate prescriptions to calculate the combined PDF+ α_s uncertainties for each of the three PDF sets. These prescriptions are applied here (for formulas see [103, 104]). As total uncertainty the

12. Systematic uncertainties

envelope of the uncertainties around the central value is used. The central value M is calculated with help of the formulas given in equation 12.2, where i labels the three different PDF sets (CT10, MSTW2008, NNPDF 2.3) and \mathcal{O}_0^i labels the observable of interest evaluated with the central PDF member of PDF set i . The percentual width of the envelope is called δ .

$$\begin{aligned}
 U &= \max_i \{ \mathcal{O}_0^i + \sigma^{(i)}(\alpha_s + PDF, +) \} & L &= \max_i \{ \mathcal{O}_0^i - \sigma^{(i)}(\alpha_s + PDF, -) \} \\
 M &= \frac{U + L}{2} & \delta &= \frac{U - M}{M}
 \end{aligned}
 \tag{12.2}$$

In figure 12.1-left the resulting uncertainties for the three PDF sets are shown as a function of M_T normalized to the central value M . The resulting δ of the envelope is shown in figure 12.1-right. The total PDF+ α_s uncertainty rises with M_T and is mostly dominated by CT10.

k-factor

As explained in section 8.2, the difference between the two combination methods of the EW and QCD corrections is used as uncertainty.

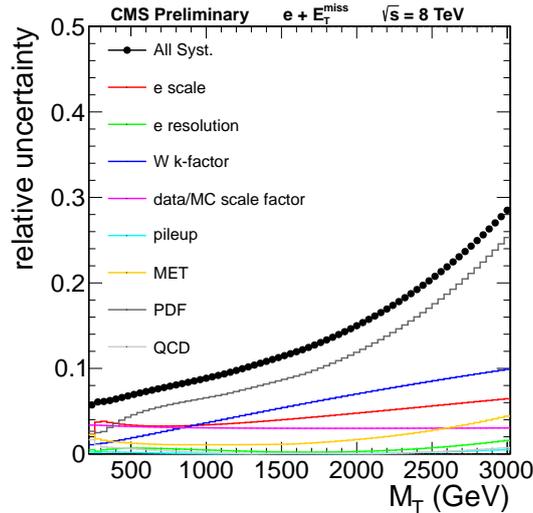


Figure 12.2. Relative systematic uncertainty on the $e + E_T^{\text{miss}}$ spectrum as function of M_T . The total uncertainty is shown together with the individual contributions.

The impact of the systematic uncertainties on the $e + E_T^{\text{miss}}$ spectrum is summarized in figure 12.2. Shown there is the resulting total relative systematic uncertainty as well as the contribution from the individual sources as function of M_T . The total uncertainty is around 6-10% for the low M_T region with $M_T < 1.2$ TeV. It increases as a function of M_T and amounts to 28% at 3 TeV. The total uncertainty is dominated by theoretical uncertainties. Especially for high M_T the PDF+ α_s uncertainty is by far the most important uncertainty with a value close to $\approx 25\%$ at 3 TeV. The most important experimental uncertainty is due to the electron scale. At 3 TeV it is around 6%. For the low M_T region around 300 GeV, the contribution from PDF+ α_s is comparable to the contribution from electron scale and scale factor uncertainty. Each of the three has a relative uncertainty of around 3-4%.

13 Final M_T distribution

In this chapter, the final analysis results are shown. First, the results of a Monte Carlo study of the M_T resolution are given, which are used to choose the binning for the M_T spectrum. Then, the final M_T spectrum is presented and a first comparison of data and Standard Model prediction is done.

13.1. M_T resolution

The binning of the $e + E_T^{\text{miss}}$ spectrum has to be chosen consistent to the M_T resolution. A binning with bin-widths much smaller than the resolution could lead to unphysical structures in the M_T spectrum and should be avoided. In general, the M_T resolution will not be constant but a function of M_T and will depend on the electron energy and the E_T^{miss} resolution. Since E_T^{miss} is a composite object, the resolution can depend on the actual composition of the event. For this analysis, the by far most important Standard Model contribution arises from the $W \rightarrow e\nu$ process. In these events, the E_T^{miss} resolution will be dominated by the electron energy resolution, especially for high M_T . To choose the binning for this analysis, a study was performed using $W \rightarrow e\nu$ Monte Carlo samples. The result can be found in figure 13.1 and shows the M_T resolution as a function of M_T .

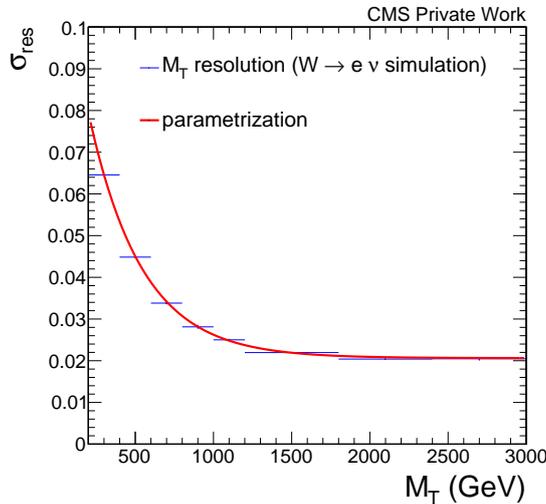


Figure 13.1. M_T resolution as function of M_T determined with $W \rightarrow e\nu$ Monte Carlo.

The difference between reconstructed M_T and truth M_T normalized to the reconstructed M_T is determined for each event. Here, the M_T range is divided into bins as shown in figure 13.1. In each bin the RMS of the normalized difference between reconstructed and truth M_T is calculated which corresponds to the M_T resolution σ_{res} in that bin. In addition to the resolution in each bin, a parametrization of the M_T resolution is shown in figure 13.1, which is used to determine the bin-widths for the $e + E_T^{\text{miss}}$ spectrum. Since the M_T resolution is not flat in M_T , a variable binning is used.

13. Final M_T distribution

13.2. Results

In figure 13.2-left, the M_T distribution from 20 fb^{-1} of CMS collision data after the full analysis is shown together with the predicted contribution from Standard Model processes. The data is shown as black points while the Standard Model prediction is shown as stack divided into the individual contributions from $W \rightarrow \ell\nu$, multi-jet, top, γ + jets, Drell-Yan and Diboson processes. The systematic uncertainty on the Standard Model prediction is shown as grey shaded band. Figure 13.2-right shows a cumulative distribution for data and Monte Carlo, where the number of events integrated from a lower M_T bound up to 4000 GeV is shown as a function of the lower M_T bound. In both plots, no significant deviation of the data from the Standard Model prediction is visible. The data event with the highest M_T has a M_T of 2.31 TeV.

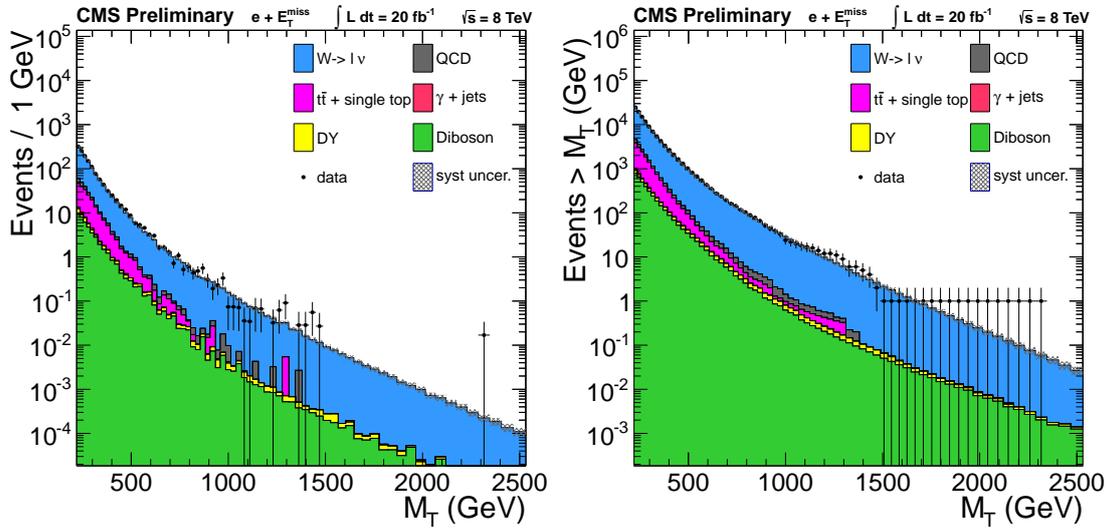


Figure 13.2. Left: M_T distribution of data (black points) and Standard Model prediction (divided in different processes shown as stack) after full selection. The total systematic uncertainty is shown as grey shaded band. Right: Cumulative distribution of data and Standard Model prediction as a function of M_T . Each bin represents the number of events integrated from the x-value of the bin up to 4000 GeV.

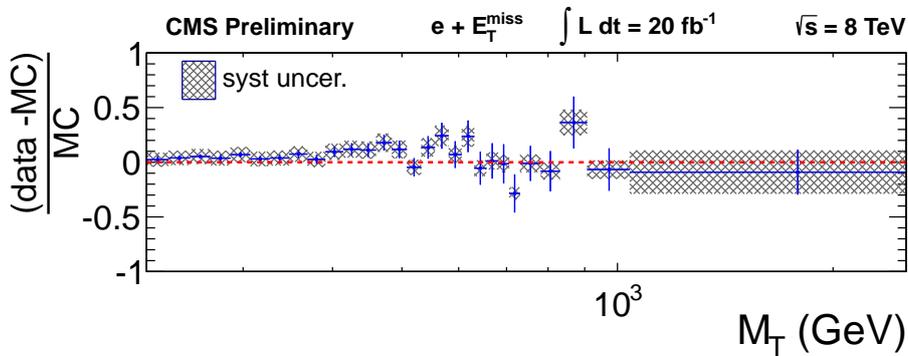


Figure 13.3. Difference between data and Standard Model prediction normalized to the Standard Model prediction. Each bin contains at least 20 events predicted from Standard Model processes.

Figure 13.3 shows the difference between data and Standard Model prediction normalized to the Standard Model prediction as a function of M_T . With the binning used in figure 13.2, the expected number of events per bin is very small for high M_T . Most of the bins with M_T larger than 1 TeV do not have any contribution from data. The few bins containing an event from data would lead to a very large value in this ratio plot. In order to keep all points of the ratio plots visible, the y-axis range would have to be very large. This would spoil the utility of the ratio plot for low M_T , where the ratio is expected to be close to 0. Therefore, starting from the binning used in figure 13.2, bins of the ratio plot are merged until the Standard Model prediction for each bin is at least 20 events. Coupled with the logarithmic x-axis, figure 13.3 allows a closer look at the data - Standard Model agreement. Up to M_T values of 390 GeV, the agreement between data and Standard Model prediction is very good and the relative deviations close to 0. Between 390 GeV and 630 GeV, some deviations occur which are not included in the 1σ band of the systematic uncertainties anymore. With the exception of two bins, the agreement beyond 630 GeV is again very good. If the data follows the Standard Model prediction, the expected value of the ratio shown in figure 13.3 is 0 for each bin. This expected value can be used to perform a χ^2 test with the 28 bins of the ratio plot. Taking only the systematic uncertainties into account, a χ^2/N_{Dof} of 2.3 is achieved. Allowing the total normalization of the Standard Model prediction to float by 2.6%, which corresponds to the systematic uncertainty of the luminosity measurement and is well within the total normalization uncertainty (see chapter 12), a χ^2/N_{Dof} of 1.3 is obtained. Overall, the measured data agrees very well with the Standard Model prediction.

In table 13.1 the number of events integrated from a lower M_T threshold up to 4000 GeV for data and Standard Model prediction is shown for 6 different thresholds between 500 GeV and 2500 GeV. For M_T thresholds up to 1 TeV the data and Standard Model prediction agree well within the total uncertainty of the Standard Model prediction. Beyond 1.5 TeV one data event is found while the integrated Standard Model prediction is around 2 events. The difference between data and Standard Model prediction is not covered by the uncertainty given in table 13.1. Choosing a lower M_T threshold of 2.3 TeV just before the remaining data event, leads to a Standard Model expectation of 0.056 events. Again, the difference to the number of measured data events is not covered by the uncertainty given in table 13.1. At first sight, this looks like a significant deviation of Standard Model and data. But so far two additional aspects have not been considered. The first one is that data events only occur in discrete numbers. When measuring pp collisions, there is no such thing as half an event. The other aspect is that for M_T larger than 1.5 TeV the expected number of events per bin is very small (especially smaller than 1). For the high M_T region, the fact that the data is only a limited sample of the physics realized in nature becomes evident. Collecting another 20 fb^{-1} of pp collision data at 8 TeV under the same conditions as 2012 would lead to a result, which would most probably look very similar for low M_T , while for high M_T the position of the few data events over the large amount of bins might look completely different. As said before the expected integrated number of events for a lower M_T threshold of 1.5 TeV is around 2 with an uncertainty of approximately 0.26. Assuming a Poisson distribution with a mean value of 2, the probability to observe exactly 2 events is 0.27 which is equal to the probability to observe exactly 1 event. Therefore, the deviation of data and Standard Model prediction is not significant, even if the difference is not covered by the uncertainty of the Standard Model prediction. The same reasoning applies for

13. Final M_T distribution

Table 13.1. Number of events integrated from a lower M_T threshold up to 4 TeV for data and Standard Model prediction. Also shown is the total uncertainty on the Standard Model prediction

M_T^{\min} (GeV)	500	1000	1500	2000	2300	2500
Data	880	24	1	1	1	0
SM prediction	814	26.0	2.02	0.207	0.056	0.0232
Uncertainty	± 62	± 2.6	± 0.26	± 0.038	± 0.012	± 0.0059

the threshold at 2.3 TeV. The probability to observe 1 or even more events when 0.056 events are expected (p-value) is around 0.054 which corresponds to a significance of 1.6σ .

A detailed analysis of the measured data in the context of different BSM physics models is performed in chapter 15-18.

14 Signals

In this chapter, the analysis of the different BSM physics models under investigation are presented. An introduction into the theory was already given in chapter 3. Therefore, this chapter focusses on the signal properties as well as the simulation and the analysis of the signal samples. For all models presented here, the general signal properties explained in chapter 7 do apply. The first BSM model which will be discussed, is the SSM W' model without interference with the Standard Model W , followed by W' models with interference effects (SSMS and SSMO). Then, the analysis of HNC CI signals is explained. The results from this analysis are interpreted in the context of various other models. This includes models with extra dimensions (split Universal Extra Dimensions model and TeV-1 model) as well as an effective field theory dark matter model with different coupling and interference scenarios. Information about the models and results can be found in [1]. A detailed description of the dark matter analysis can be found in [65].

14.1. SSM W'

In section 3.1, different theoretical models have been introduced which predict the existence of a heavy charged boson called W' . In order to cover a large set of different theories, it is useful not to rely on the details of specific models but to summarize the common properties in a generalized model, which can be used as a benchmark model. In this thesis the SSM model is used as W' model. It includes W' scenarios without interference of Standard Model W and W' as well as scenarios with constructive and destructive interference. In figure 14.1, the Feynman graph of the leading order production process at the LHC is shown. The event signature of W' events was already presented in chapter 7. In the following, the simulation of signal samples, the shape of the signal in the M_T spectrum, the signal selection efficiencies and the differences between the three W' scenarios are presented.

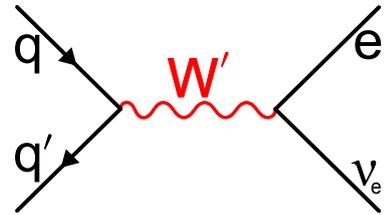


Figure 14.1. Feynman graph of the production and subsequent decay of a W' .

14.1.1. Scenario without interference

The W' has a similar experimental signature as the Standard Model W -boson, but at much higher masses. In the invariant mass spectrum of the electron neutrino system, a W' would produce a Breit-Wigner peak around the W' mass. As said before, the invariant mass is not accessible from the experimental side of view and the transverse mass (see equation 7.1) is used instead. Projecting the signal shape to the transverse plane leads to a Jacobian peak. The signal contribution increases with M_T up to $M_T = m_{W'}$. Beyond $M_T = m_{W'}$ a sharp decrease of the signal contribution occurs. In figure 14.2, the signal shape is shown for four different W' masses.

14. Signals

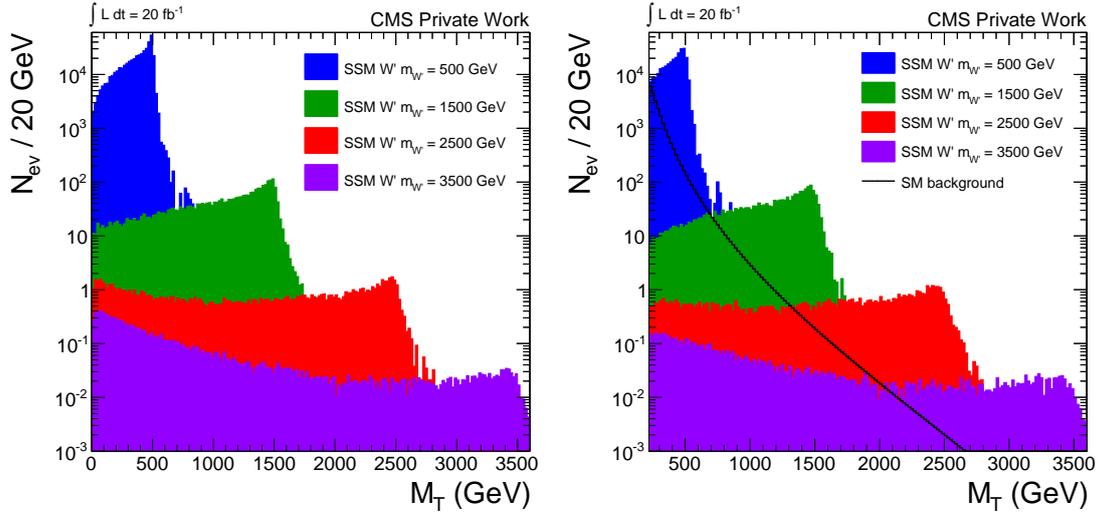


Figure 14.2. Left: M_T distribution of four different SSM W' samples on generator level. The samples are drawn in front of each other. Right: M_T distribution of four different SSM W' samples after full selection. The samples are drawn in front of each other. Also shown is the expected SM background.

For low $m_{W'}$ the Jacobian peak structure is clearly visible. The width of the peak increases with $m_{W'}$, due to the linear dependence of the W' decay width on $m_{W'}$ (see equation 3.3). For high $m_{W'}$ an additional effect is visible. The fraction of W' bosons produced off-shell with a lower mass than $m_{W'}$ increases with $m_{W'}$. This leads to an increase of the signal contribution at lower M_T . The reason for this increase is the limited center of mass energy of the LHC. For very high $m_{W'}$, it is less probable to produce an on-shell W' , since only the extreme tails of the PDFs contribute. An extensive study of the SSM W' signal shape on generator level can be found in [62].

The simulation of the SSM W' signal samples is done with Pythia in LO. In total 27 samples have been produced with masses between $m_{W'} = 300$ GeV and $m_{W'} = 4000$ GeV. For each mass point, 20000 events have been produced. All samples are from the official CMS Monte Carlo production and produced in the production campaign **Summer12**. The cross-section is corrected to NNLO using k-factors ranging from 1.14 to 1.36 (see [108] for details). A summary of all samples with corresponding W' mass, cross-section and k-factor can be found in table A.2. In figure 14.3-left the LO and NNLO cross-section is shown as a function of the W' mass. The NNLO cross-section is around 113.5 pb for a W' mass of 300 GeV and decreases down to 0.25 fb for $m_{W'} = 4000$ GeV. The cross-section uncertainty due to PDF uncertainties is calculated using a similar reweighting procedure explained chapter 12. In figure 14.3, the uncertainties for the 3 PDF sets CT10, MSTW2008 and NNPDF 2.3 are shown for each of the produced W' mass points. As total uncertainty the envelope is taken. For a W' mass of 300 GeV, the resulting uncertainty is around 2.2% rising up to 9.5% for a W' mass of 2600 GeV and decreasing again to 3.6% for a W' mass of 4000 GeV. The decrease of the uncertainty for higher W' masses is due to the increasing part of off-shell production at lower M_T , where the PDFs are known with a higher precision.

In figure 14.2-right, the reconstructed M_T spectra after the full event selection described in chapter 9 can be seen for four different W' masses. The M_T spectra are shown starting at $M_T = 220$ GeV together with the predicted Standard Model background.

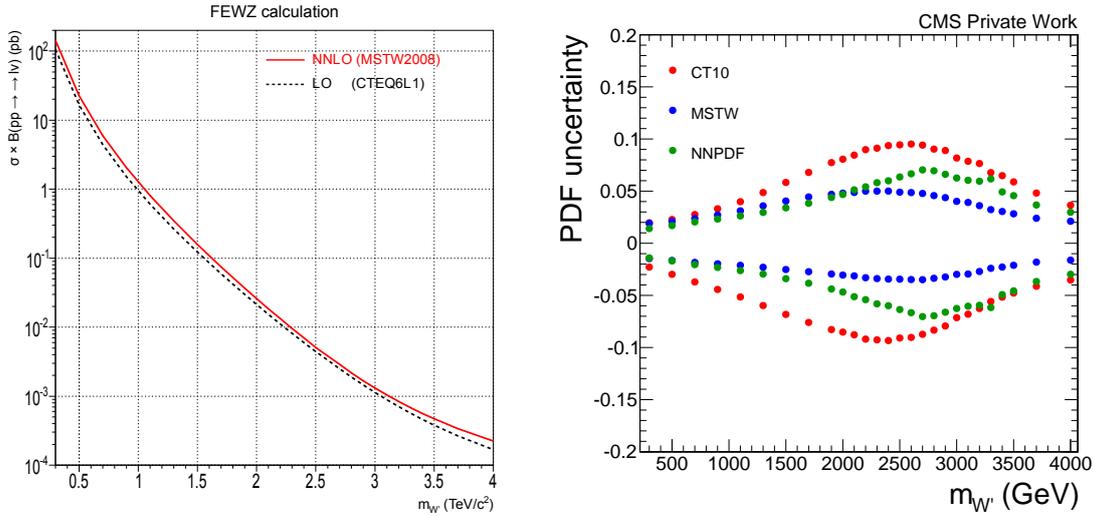


Figure 14.3. Left: SSM W' LO and NNLO cross-section as a function of $m_{W'}$. Figure from [108]. Right: Cross-section uncertainties resulting from PDF uncertainties. The uncertainties were derived using three different PDF sets (CT10, MSTW2008 and NNPDF 2.3).

The reconstructed shape differs very little from the reconstructed ones. The peak structure and the increasing fraction of off-shell production with rising $m_{W'}$ remains clearly visible. Due to limited detector acceptance as well as reconstruction and selection efficiencies smaller than 1, the amount of reconstructed signal events is obviously smaller than on generator level.

The overall acceptance times selection efficiency is around 70% for $m_{W'}$ between 1 TeV and 2.5 TeV, decreasing down to 50% for $m_{W'} = 4$ TeV and down to 40% for $m_{W'} = 0.3$ TeV. In figure 14.4, the acceptance times efficiency is shown divided in geometrical acceptance, preselection, HEEP ID and kinematic selection.

The geometrical acceptance is around 90%. The largest efficiency loss occurs during the preselection which contains the electron energy cut of 100 GeV. This cut is the main reason for the decrease of efficiency, since for small $m_{W'}$ a larger part of the signal is below that threshold and for high $m_{W'}$ the increasing fraction of off-shell production leads to a larger amount of event which do not pass the E_T cut.

14.1.2. Scenarios with interference

When interference effects between the Standard Model W and a potential SSM W' are taken into account, the signal shape changes. As explained in section 3.1, there is an interference between W and W' , if the latter one couples to left-handed fermions. The interference can be constructive or destructive depending on the sign of the W' coupling to quarks and leptons. If the coupling of the W' to leptons and quarks have the same sign, destructive interference in the mass region between m_W and $m_{W'}$ occurs. This scenario is called SSMS. In the other scenario called SSMO, the couplings have opposite signs leading to constructive interference in the mass region mentioned before. For masses larger than $m_{W'}$ the effect of the interference changes in both scenarios. The interference in the SSMO (SSMS) scenario is then destructive (constructive). An example of the different shapes is shown in figure 14.5.

14. Signals

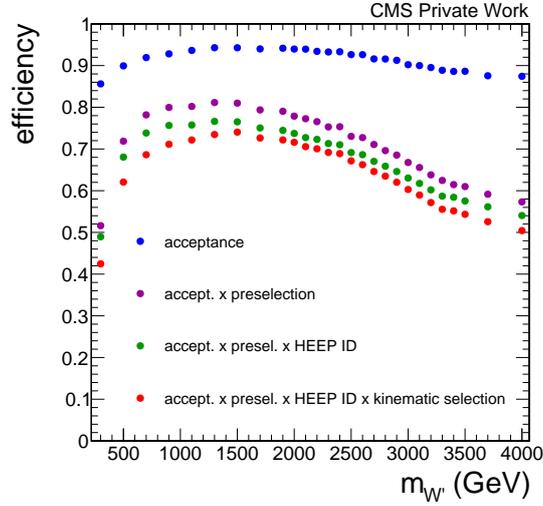


Figure 14.4. Overall acceptance times efficiency as a function of the SSM W' mass (red dots). Also shown is the geometrical acceptance (blue dots) and the selection efficiency for different steps of the event selection (purple and green dots).

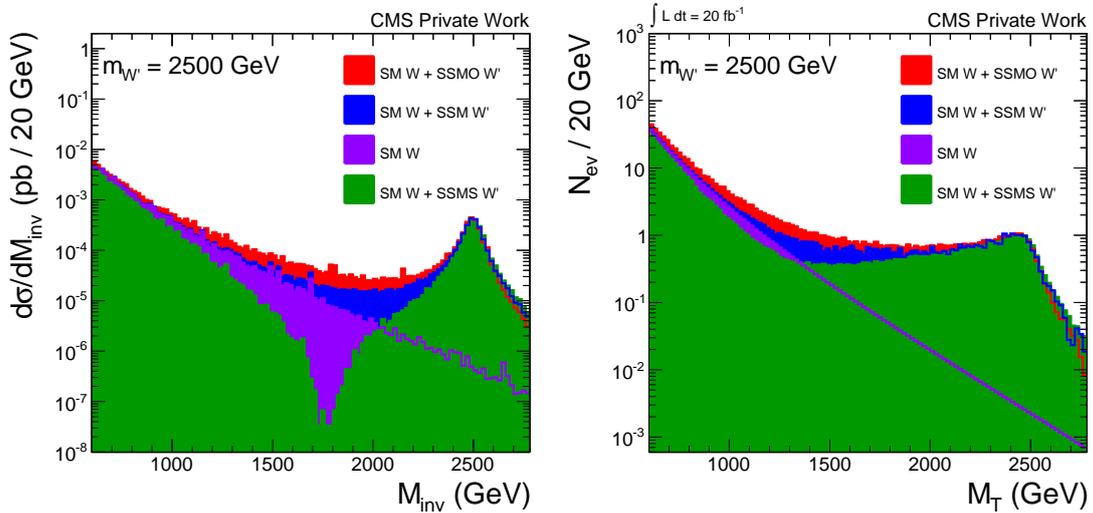


Figure 14.5. Left: Differential cross-section of $W + W'$ for three different SSM W' scenarios with constructive (red curve), destructive (green curve) and no interference (blue curve) as function of the invariant mass M_{inv} . Also shown is the distribution of the Standard Model W without W' contribution (purple curve). The distributions are drawn in front of each other. This figure was made based on LHE files provided by Mark Olschewski. Right: M_T distribution of $W + \text{SSMO } W'$, $W + \text{SSM } W'$, $W + \text{SSMS } W'$ and W only. The same color code as for the left figure is used. Again the distributions are drawn in front of each other.

In figure 14.5-left, the invariant mass distributions of the Standard Model W, W + SSM W', W + SSMO W' and W + SSMS W' are shown. At the W' mass peak, the W + W' event yield is equal in all three scenarios. For invariant masses smaller than $m_{W'}$ the effect of the W - W' interference becomes clearly visible. In the SSMO case, the constructive interference of W and W' leads to an enhancement of the W + W' production in the region between m_W and $m_{W'}$ (red curve). The resulting event yield is therefore higher than the sum of W and SSM W' (blue curve). In the case of destructive interference (green curve), the expected number of W + W' events is even lower than the expected events from the Standard Model W (purple curve). Here, the presence of the signal leads to a decrease of the expected events compared to the Standard Model prediction. The same behavior can be seen in the M_T spectrum, even though the effect is less significant after the projection onto the transverse plane.

The simulation of the signal samples was done with MADGRAPH. Due to the interference, the W' has to be simulated together with the Standard Model W. For each mass point, the simulation is divided into three subsamples divided by the p_T of the electron. Each subsample contains 100000 events. In total 18 mass points with W' masses between 300 GeV and 4000 GeV have been produced for each of the two scenarios. The coupling strength of the W' to fermions is chosen to be identical to the one of the Standard Model W. A summary of all produced mass points with the corresponding cross-sections can be found in table A.3.

14.2. HNC CI

In section 3.2, the HNC CI model was presented. It predicts the existence of a contact interaction leading to a process shown in figure 14.6. The theory has one adjustable parameter called Λ (see section 3.2 for details). The simulation of the HNC CI signal was done with PYTHIA in LO. Due to technical details of the implementation in PYTHIA, the simulation has to be done together with the Standard Model W, although no interference between the W and the CI occurs. Thus the W has to be subtracted from the combined sample to obtain the 'pure' signal.

A previous study [29] has shown, that the shape of the CI signal does not change with Λ but simply scales up and down with Λ^{-4} . Therefore, the use of a single signal sample is sufficient. The sample which is used for this analysis was produced for $\Lambda = 3$ TeV using a generator cut of $\hat{p}_T > 300$ GeV. On top of that, a W sample was produced with the same generator settings, which is used to subtract the W contribution of the combined sample. Additional information about the produced samples is summarized in table A.4. In figure 14.7-left, the combined CI + W distribution is shown for different values of Λ . The signal after subtraction of the W can be seen in figure 14.7-right. All distributions in figure 14.7 are based on the one produced sample. The CI part was scaled according to equation 3.8. As can be seen, the HNC CI leads to a non-resonant excess in the M_T distribution decreasing as function of M_T . The LO CI cross-section at $\Lambda = 3$ TeV is about 0.55 pb. From a similar study as performed for the SSM W', the PDF uncertainty on the cross-section is determined to be around 5.8%. The acceptance times efficiency is around 80% after the full selection.

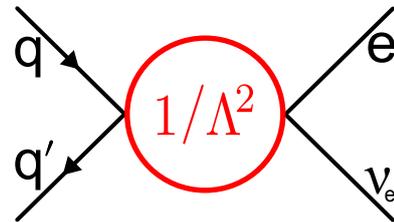


Figure 14.6. Feynman graph of the HNC CI model.

14. Signals

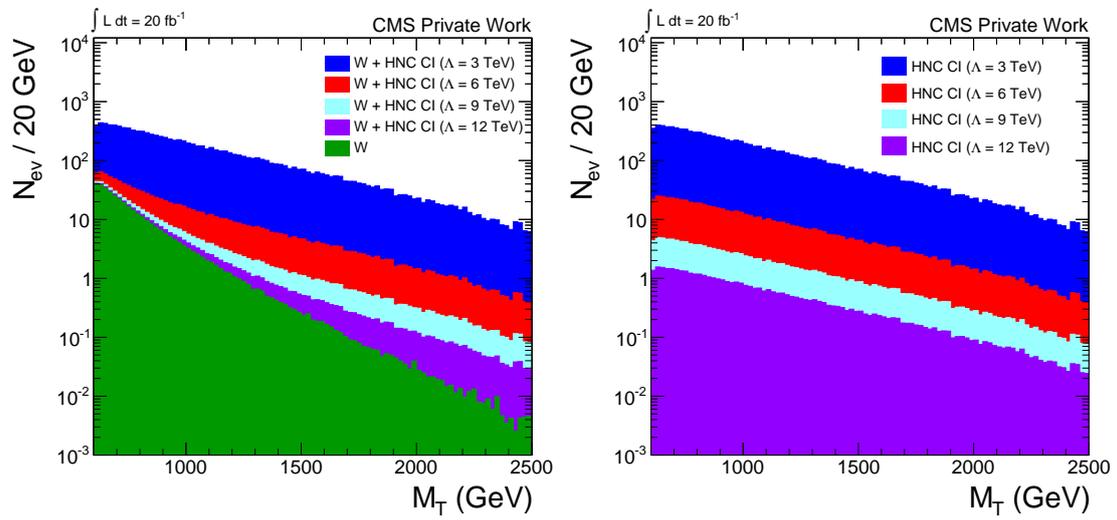


Figure 14.7. M_T distribution of the combined W + CI (left) and CI only (right) sample for different values of Λ .

Part IV.

Statistical analysis

15 Introduction

After an introduction into the theoretical and experimental background, the analysis and its result were presented. The remaining challenge is to interpret the analysis results in terms of a search for new physics using several different signal models. This implies to quantify and compare the agreement of the measurement with the Standard Model hypothesis (H_0 ; assuming the absence of new physics) and a signal hypothesis (H_1 ; assuming the presence of a signal according to the BSM physics model under investigation). Based on this result, the decision if one of the hypotheses can be rejected is taken.

In principle, two different approaches are possible. The 'optimistic' approach is the significance estimation, which tries to quantify the compatibility (or incompatibility) of the measurement with H_0 . If the incompatibility of H_0 and the measurement is significant enough (5σ), the measurement is assumed to indicate the presence of some alternative hypothesis and the H_0 hypothesis is rejected. The 'pessimistic approach' starts with H_1 and tries to exclude it. Here, this consists mainly in deriving an interval in which the parameter of interest (e.g. the signal cross-section) can be excluded at a certain confidence level, given the observed data. Since the simple 'goodness of fit' test performed in chapter 13 indicates a fair agreement between observation and Standard Model, this chapter will focus on limit setting. Nevertheless, a significant excess of the data from the Standard Model should manifest itself in a loss of exclusion power in the region of the excess compared to the expected exclusion limit assuming H_0 .

Another decision which has to be made is whether to use Bayesian or frequentist statistics. Both approaches provide a framework for limit setting. With the exception of the SSM W' model with interference, where both approaches are used, the Bayesian approach is chosen.

All limit calculations for this analysis are calculated with the 'Higgs combine tool' [109] provided by the CMS Higgs Physics Analysis Group. It is a software package based on RooStats [110] and provides a large set of different functions including Bayesian and frequentist limit setting, significance estimation and 'goodness of fit' tests. In the following, a short introduction into frequentist and Bayesian statistics is given. The content of this chapter is mainly taken from [7, 111].

15.1. Frequentist approach

In the frequentist approach the probability $P(A)$ is defined as the relative frequency of A in the limit of infinite repetitions of an experiment in which A is a possible outcome. A common task involving statistical methods is the parameter estimation. Let x be a vector of data, $f(x_i, \theta)$ the probability density function (pdf) describing each measurement x_i and θ an unknown parameter of f which is to be estimated based on the data. Following the method of maximum likelihood an estimator $\hat{\theta}$ can be found by maximising the likelihood function given in equation 15.1 with respect to θ .

15. Introduction

$$L(x|\theta) = \prod_{i=1}^N f(x_i, \theta) \quad (15.1)$$

If the underlying model $f(x, \theta)$ is not known, the frequentist hypothesis testing can be used to test one hypothesis for $f(x, \theta)$ against another one. The first part of formulating a hypothesis test consists in defining a test statistic. Based on the result of the test statistic using the observed data, a decision whether to reject a hypothesis or not is taken. The second part of constructing a hypothesis test is to define a region of test statistic results for which the hypothesis H_1 is to be rejected. This 'rejection region' is chosen in a way that the probability to reject H_1 if H_1 is true, is less than or equal to the significance level α of the test. In this context the type-I and type-II error can be defined. A type-I error is to reject H_1 if it is true. As said before this occurs with the probability smaller or equal to α . A type-II error is defined as not rejecting H_1 if an alternative hypothesis H_0 is true.

The Neyman-Pearson lemma (see [7] for details) states, that for simple hypotheses (without any undetermined parameter) the use of the likelihood ratio (see equation 15.2) as test statistic minimizes the type-II error rate at a given significance level α .

$$\lambda(x) = \frac{L(x|H_1)}{L(x|H_0)} \quad (15.2)$$

H_0 is rejected if $\lambda(x)$ is smaller than a constant value depending on α . Because of technical reasons, $-2\ln(\lambda(x))$ is usually chosen instead of $\lambda(x)$. In this analysis, the CLs method [112] is used, which is named after its test statistic and based on the likelihood ratio. From its first use at LEP to the use at TeVatron and LHC, the CLs method has evolved and exists now in different versions. A comparison of its different forms can be found in [113]. The explanations given here will be restricted to the version which is used for this analysis. The CLs test statistic can be written with the help of p-values. A p-value tries to quantify the 'degree of surprise' of an observation with respect to a hypothesis H . This is done with help of a test statistic. A p-value is defined as the probability to observe an equally or more extreme value of the test statistic (assuming H is true) than obtained from data. Small p-values indicate that the observation is unlikely assuming that H is true. The definition of CLs is given in equation 15.3, where p_μ (p_0) is the p-value of the signal (background) hypothesis H_1 (H_0).

$$CLs = \frac{p_\mu}{1 - p_0} \quad (15.3)$$

Here, the denominator serves as penalty factor to avoid the exclusion of a signal if the analysis is not sensitive to it and the p-values are calculated based on the likelihood ratio as test statistic. The pdfs in the likelihoods will depend on parameters which are known within some uncertainty (e.g the integrated luminosity) and which will be called nuisance parameters ν . To incorporate systematic uncertainties on the nuisance parameters, each likelihood is extended by multiplying it with prior functions (see next section) for the nuisance parameters. This corresponds to the Bayesian treatment of systematic uncertainties. Thus, the CLs method applied here is not a purely frequentist approach but a 'frequentist-Bayesian hybrid' method. Each likelihood is maximised with respect to its nuisance parameters ($\hat{\nu}_\mu, \hat{\nu}_0$). The resulting likelihood ratio is shown in equation 15.4.

$$q_\mu = -2\ln(Q) = -2\ln\left(\frac{L(x|\mu, \hat{\nu}_\mu)}{L(x|\mu=0, \hat{\nu}_0)}\right) \quad (15.4)$$

Here, μ is the parameter of interest characterizing the signal strength ($\mu = 0$ is the Standard Model hypothesis). The signal hypothesis is excluded at 95% CL for all values of μ where CLs < 0.05.

15.2. Bayesian approach

The definition of probability in the Bayesian approach differs from the frequentist one. The probability $P(A)$ is seen as the degree of belief that A is true and thus has a subjective element. Let x be a vector of data and θ the parameter of a model describing the data which is to be determined. In the Bayesian approach it is possible to formulate a 'posterior pdf' $p(\theta|x)$ for θ given the data x . Following Bayes theorem (see [7] for details) $p(\theta|x)$ can be calculated according to equation 15.5.

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int d\theta' L(x|\theta')\pi(\theta')} \quad (15.5)$$

Here $L(x|\theta)$ is the likelihood function and $\pi(\theta)$ is the 'prior pdf' for θ . In $\pi(\theta)$ the knowledge about θ before the incorporation of the data is summarized. For the formulation of the prior, no fundamental procedure exists. It can be driven by previous measurements or theoretical arguments. In many cases, the parameter of interest of a search for new physics is the signal cross-section. If no a priori information about the cross-section is available, this can be expressed in the prior by choosing a uniform distribution truncated at 0. This way, all positive values for the signal cross-section are assumed to be equally likely. The Bayesian approach provides a natural mechanism to treat systematic uncertainties. If the model mentioned above also depends on a set of parameters ν , the posterior pdf, the prior pdf and the likelihood will depend on ν as shown in equation 15.6.

$$p(\theta, \nu|x) = \frac{L(x|\theta, \nu)\pi(\theta, \nu)}{\int d\theta' d\nu' L(x|\theta', \nu')\pi(\theta', \nu')} \quad (15.6)$$

The parameters ν are the nuisance parameters and are assumed to be known within a given uncertainty. To obtain the posterior pdf as function of θ , an integration over all possible values for the nuisance parameters can be done. Assuming that the joint pdf $\pi(\theta, \nu)$ factorizes, $p(\theta|x)$ can be written as given in equation 15.7.

$$p(\theta|x) = \frac{\int d\nu' L(x|\theta, \nu')\pi(\theta)\pi(\nu')}{\int d\theta' d\nu' L(x|\theta', \nu')\pi(\theta')\pi(\nu')} \quad (15.7)$$

Since the nuisance parameters are known within an uncertainty, this knowledge is incorporated in the prior. For a parameter ν_i with an uncertainty σ_{ν_i} a Gaussian function centered at ν_i with σ_{ν_i} as standard deviation can be used. If ν_i is bound to positive, log-normal functions have proven to be a good choice. Integrating equation 15.7 over an interval $[\theta_1, \theta_2]$ yields the degree of believe that $\theta \in [\theta_1, \theta_2]$, which is used to define Bayesian credible intervals. These integrals can usually not be solved analytically. Therefore the use of numerical methods is indispensable. In the software framework which is used for this analysis, the Markov Chain Monte Carlo (MCMC) method is

15. Introduction

used. A short introduction of the MCMC method can be found in [7]. Upper limits on θ at credibility level $1-\alpha$ are derived by solving equation 15.8 for θ_{up} .

$$1 - \alpha = \int_{-\infty}^{\theta_{\text{up}}} d\theta p(\theta|x) \quad (15.8)$$

This equation is used to determine upper cross-section limits on the signal cross-section by setting $1-\alpha$ to 0.95 and choosing the signal cross-section as parameter of interest.

16 SSM W' without interference

In this chapter, the determination of upper cross-section limits on the production of a SSM W' without interference effects is presented. The expected Standard Model background and the measured data is shown again in figure 16.1 together with four SSM W' signals. The cross-section limits are calculated with two different methods. First a single-bin method is used which integrates the M_T spectrum above a certain threshold and performs a Poisson counting experiment. A similar method was used in previous master theses [29, 61, 62, 64] and publications [23, 32, 114] of W' searches. Then a generalization of the single-bin method to multiple bins is presented. This method will be called multi-bin method and has several advantages which will be explained later.

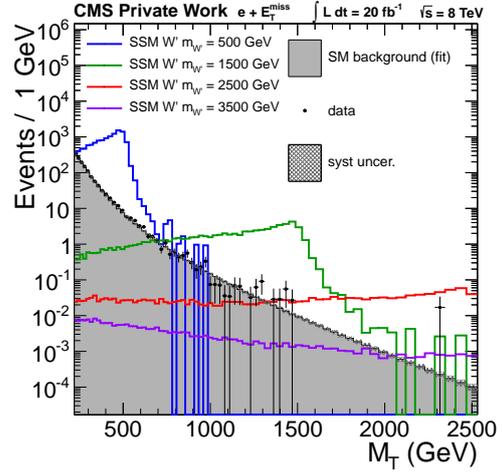


Figure 16.1. M_T distribution for expected Standard Model background (fit presented in chapter 11), data and four different SSM W' mass points. The curves are not stacked.

16.1. Single-bin method

A conceptually simple but yet powerful approach for setting upper cross-section limits is based on comparing the number of expected events with the number of observed events. Since the existence of a SSM W' would lead to an excess of events at high M_T above the exponentially falling Standard Model background, it is utile to set a lower M_T threshold and to use only the number of events above this threshold. This will improve the signal to background ratio and increase the exclusion power of the method. The optimal value for M_T^{\min} will in general depend on the signal shape and cross-section and its determination will be explained later. In this approach all bins above M_T^{\min} are merged into a single bin. Therefore, this method will be called single-bin method. The number of events for this bin (N_{exp}) can be seen as the result of a Poisson counting experiment. The formula for calculation of N_{exp} and the corresponding likelihood function is given in equation 16.1 and 16.2 respectively.

$$N_{\text{exp}} = \mathcal{A}(\mathcal{L}, \epsilon) \cdot \sigma_s + b \quad (16.1)$$

$$L = \frac{(N_{\text{exp}})^{N_{\text{obs}}}}{N_{\text{obs}}!} \cdot e^{-N_{\text{exp}}} = \frac{(\mathcal{A} \cdot \sigma_s + b)^{N_{\text{obs}}}}{N_{\text{obs}}!} \cdot e^{-(\mathcal{A} \cdot \sigma_s + b)} \quad (16.2)$$

Here \mathcal{A} is the signal acceptance, σ_s the signal cross-section, b the number of expected Standard Model events and N_{obs} the number of observed data events in this bin. The signal acceptance \mathcal{A} as used in this context is defined to be a product of the integrated luminosity \mathcal{L} and the efficiency times geometrical acceptance ϵ for the chosen M_T^{\min} . The

16. SSM W' without interference

signal cross-section σ_s is the parameter of interest on which a upper limit is to be set. \mathcal{L} , ϵ and b are treated as nuisance parameters. Systematic uncertainties on the nuisance parameters are incorporated by multiplying the likelihood with the prior functions of the nuisance parameters. In this model, log-normal functions are used as priors. The posterior pdf for the parameter of interest σ_s is shown in equation 16.3.

$$\Pi(\sigma_s|N_{obs}) = \int d\mathcal{L} d\epsilon db \pi(\sigma_s)\pi(\mathcal{L}) \cdot \frac{(\pi(\epsilon) \cdot \mathcal{A} \cdot \sigma_s + \pi(b) \cdot b)^{N_{obs}}}{N_{obs}!} \cdot e^{-(\pi(\epsilon) \cdot \mathcal{A} \cdot \sigma_s + \pi(b) \cdot b)} \quad (16.3)$$

Here, $\pi(\mathcal{L})$, $\pi(\epsilon)$ and $\pi(b)$ represent the priors for the nuisance parameters and $\pi(\sigma_s)$ the prior for parameter of interest. For σ_s a uniform prior is taken. With $\Pi(\sigma_s|N_{obs})$ the upper limits on σ_s can be calculated according to equation 15.8. To quantify the expected exclusion limit and possible variations due to statistical and systematic uncertainties, pseudo experiments are used. Assuming the absence of any signal ($\sigma_s = 0$) the outcome of a pseudo experiment N_{obs}^{pseudo} is determined by dicing the nuisance parameters according to their pdfs and drawing a N_{obs}^{pseudo} from a Poisson distribution using the Standard Model expectation with the diced nuisance parameters as mean. For each SSM mass point, 4200 limits based on pseudo experiments are calculated. The median of this set of limits is used as expected exclusion limit. To quantify the spread of the expected limit a 1σ (2σ) band is determined by taking the range centered around the median which contains 68.2% (95.4%) of all the limits from pseudo experiments.

The lower M_T threshold is chosen by optimizing it for the best expected limit. For each mass point the expected exclusion limit is calculated using thresholds between 250 GeV and 2500 GeV (in 50 GeV steps) with a reduced number of pseudo experiments (1250 per threshold and mass point). The M_T threshold leading to the best expected limit is then used for the final limit calculation. The result of the optimization process is shown in figure 16.2-left where the expected upper limit on cross-section times branching fraction is given as function of M_T^{min} for different W' mass points. The cross-section times branching fraction ($\sigma \times B$) corresponds to the signal cross-section in the electron channel σ_s which was used in the formulas before.

The main signal contribution of a SSM W' is accumulated in the Jacobian peak at $m_{W'}$ if the W' mass is not too high (cf. chapter 14.1 and figure 14.2). For very high masses the off-shell contribution increases and the peak is less pronounced. The expected upper cross-section limit for small and intermediate masses ($m_{W'} \in [300 \text{ GeV}, 2500 \text{ GeV}]$) improves (gets smaller) with rising M_T^{min} up to the point where the threshold starts to cut into the Jacobian peak leading to a drastic decrease of the exclusion power. An increase of M_T^{min} way below the Jacobian peak leads to a small loss of signal contribution compared to a large reduction of the (exponentially falling) Standard Model contribution. Close to the peak, an increase of M_T^{min} would lead to the opposite effect (large loss of signal contribution compared to a relatively small reduction of background). For high $m_{W'}$ ($m_{W'} > 2500 \text{ GeV}$) the contribution from the off-shell part starts to get relevant compared to the signal contribution in the peak. The optimization procedure leads to a broad minimum around 1750 GeV. For very high $m_{W'}$ (4000 GeV) the optimal M_T^{min} starts to decrease again, since the off-shell contribution dominates the contribution from the Jacobian peak.

The SSM W' upper cross-section limits obtained with the single-bin method are shown in figure 16.2-right. Shown there is the upper limit on cross-section times branching

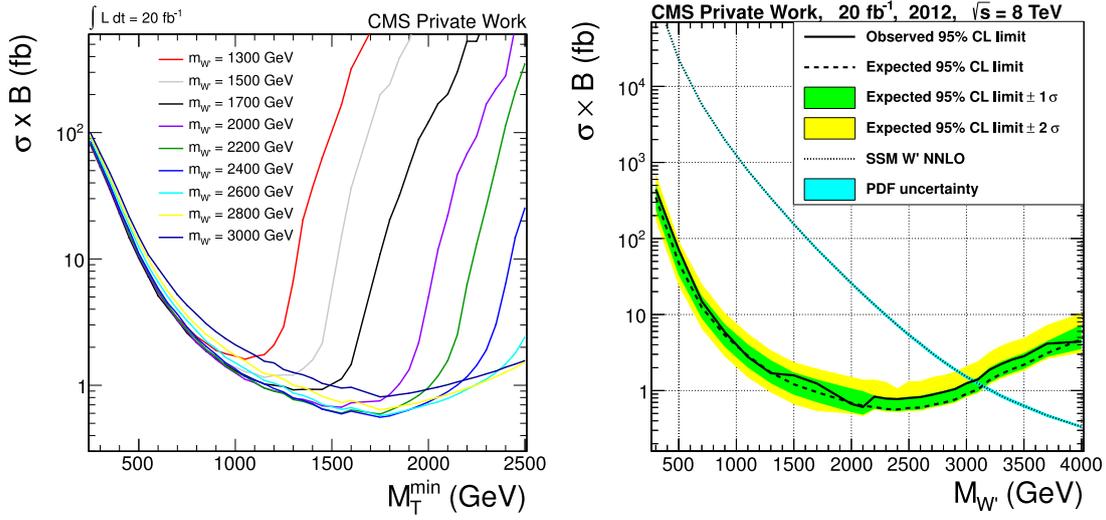


Figure 16.2. Left: Result of the M_T^{\min} optimization procedure. Shown is the expected upper cross-section limit as function of M_T^{\min} for various $m_{W'}$. Right: Upper limit on cross-section times branching fraction as function of $m_{W'}$. The black solid line is the observed limit. The expected limit is shown as black dashed line. The corresponding 1 and 2 sigma bands are shown in green and yellow. Every mass point where the observed line is below the predicted SSM W' cross-section (dotted black line) is excluded. All limits at 95% CL.

fraction as function of $m_{W'}$. The solid black line is the observed exclusion limit obtained with the measured data. The dashed black line and the green and yellow band represent the expected limit with 1 and 2 σ bands from the calculation of pseudo experiments. Over the whole $m_{W'}$ range the observed exclusion limit is within the 2 σ band around the expected limit. Thus no strong sign for the existence of a W' is found. The W' cross-section predicted in the SSM model is shown as black dotted line. For each W' mass, where the observed upper cross-section limit is below the predicted SSM cross-section the SSM W' model is excluded at 95% CL. With the single-bin method a SSM W' can be excluded at 95% for masses below $m_{W'} < 3.06$ TeV compared to an expected limit of $m_{W'} < 3.14$ TeV.

16.2. Multi-bin method

The single-bin approach uses only a small amount of the available information. During the optimization procedure, expected upper cross-section limits are calculated for a large set of different M_T thresholds. Since the bins above M_T^{\min} are merged, all information about the distribution of the events is lost. The single-bin approach is thus not really sensitive to the shape of the signal. Once the optimization is done, the information about the M_T spectrum below M_T^{\min} has no impact on the limit. Especially the information about the measured data below M_T^{\min} and the distribution of the data points above M_T^{\min} is not used at all, since the optimization is done based solely on simulations. The Standard Model expectation for N_{ev} with $M_T^{\min} = 1$ TeV is around 26 ± 2.6 compared to 24 events measured in data. For $M_T^{\min} = 1.5$ TeV the Standard Model expectation is only around 2 events. If a considerable amount of the 24 data events were located in a peak above $M_T = 1.5$ TeV the data distribution would show a significant positive excess in the peak region and a negative excess in the region before the peak (similar

16. SSM W' without interference

to a SSMS W' signature (cf. section 14.1.2)). A limit calculation using $M_T^{\min} = 1$ TeV would be blind to this discrepancy. Of course, such an excess should already be visible in the M_T spectrum and thus noticed anyway. Nevertheless, this example showcases one of the weaknesses of the single-bin method.

To overcome the drawbacks mentioned above, a multi-bin method can be used which is a generalization of the single-bin method. Instead of using one large bin starting at M_T^{\min} , the whole M_T range is used divided into multiple bins. For each bin a Poisson counting experiment is performed similar to the single-bin method. The combination of all the bins is done by multiplying the single-bin likelihoods as shown in equation 16.4.

$$L = \prod_{k=1}^n \frac{(N_{exp}^{(k)})^{N_{obs}^{(k)}}}{N_{obs}^{(k)}!} \cdot e^{-N_{exp}^{(k)}} = \prod_{k=1}^n \frac{(\mathcal{A}^{(k)} \cdot \sigma_s + b^{(k)})^{N_{obs}^{(k)}}}{N_{obs}^{(k)}!} \cdot e^{-(\mathcal{A}^{(k)} \cdot \sigma_s + b^{(k)})} \quad (16.4)$$

The product loops over all bins which are used for the limit calculation. Similar to equation 16.3, $\mathcal{A}^{(k)}$ and $b^{(k)}$ are the signal acceptance and Standard Model contribution in bin k . For the treatment of systematic uncertainties the distinction between normalization and shape uncertainties is made. An example for a normalization uncertainty is the uncertainty on the integrated luminosity. As already explained in chapter 12, a different value for the luminosity would result in a simple rescaling of the whole M_T spectrum. For normalization uncertainties log normals priors are used. Shape uncertainties (e.g. on the electron energy scale) can alter the normalization and the shape of the M_T spectrum. They are included using a technique called 'vertical shape morphing'. An explanation of this method can be found in [115]. Since the multi-bin method uses the full M_T range, the optimization procedure becomes obsolete. On top of that, the method is sensitive to the distributions of the events as function of M_T and thus sensitive to the shape of the signal.

For the limit calculation, a similar binning as shown in figure 13.2 (motivated by the M_T resolution) is used. However, the multi-bin method was tested to be stable with respect to variations of the binwidth as can be seen in figure 16.3-left. Here, the expected limit is shown for eight different $m_{W'}$ using six different binwidths. The black line labeled 'variable binning' is the one which is used for the final calculation. The 1 and 2 σ bonds belong to the limit calculated with the variable binning. With the exception of the 420 GeV binning, the expected limits are independent on the chosen binwidth for the whole $m_{W'}$ range. The limit using the 420 GeV is significantly worse for $m_{W'}$ smaller than 1.5 TeV. For these masses, the signal contribution is mainly accumulated in the Jacobian peak. Thus, the signal will be contained in one of the bins, and the limit does not profit from the division of the $e + E_T^{\text{miss}}$ spectrum into multiple bins.

In figure 16.3-right, the result of the multi-bin method is shown. The structure of the plot is similar to the single-bin exclusion plot. As a comparison, the expected limit of the single-bin method is shown as red dashed line. With the multi-bin method, a SSM W' can be excluded for $m_{W'}$ smaller than 3.24 TeV compared with an expected limit of 3.18 TeV.

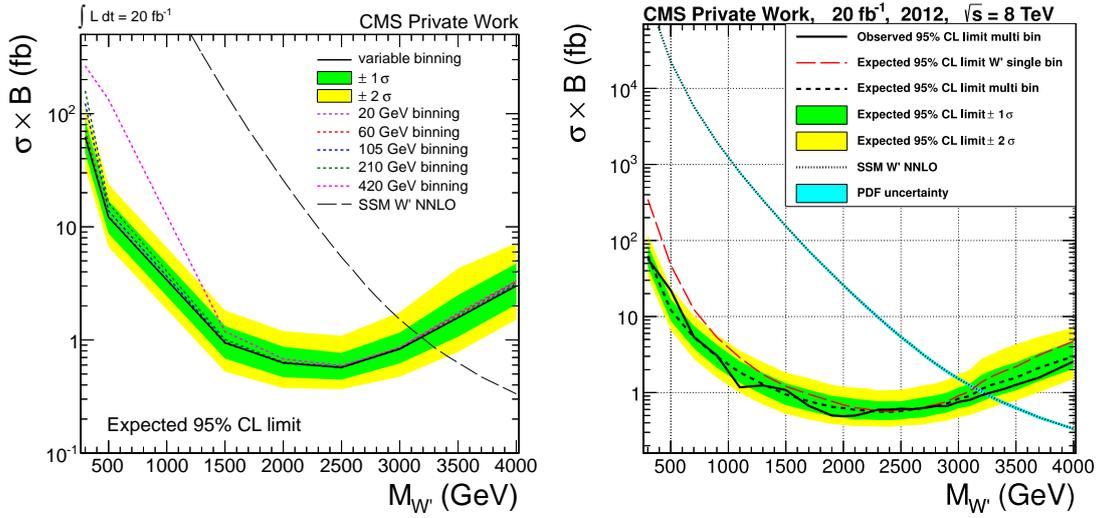


Figure 16.3. Left: Expected upper cross-section limit for 8 different $m_{W'}$ using 6 different M_T binnings. Also shown is the 1 and 2 σ bond of limit calculated with the variable binning. Right: Upper limit on cross-section times branching fraction as function of $m_{W'}$ calculated with the multi-bin method. Also shown is the expected limit from the single-bin method.

16.3. Conclusion

The exclusion limits obtained with the two methods are summarized in table 16.1. As can be seen in 16.3-right, both methods are comparable for $m_{W'}$ larger than 1.5 TeV with the multi-bin method being slightly better. In the low $m_{W'}$ regime, the multi-bin method yields significantly better exclusion limits. For a $m_{W'}$ of 300 GeV the upper cross-section limit of the multi-bin method is 6 times better (lower) than the single-bin limit. Here, the use of additional information about the distribution of the events - located in a narrow Jacobian peak - leads to a large increase of exclusion power.

Table 16.1. Summary of the SSM W' exclusion limits without interference obtained with the single-bin and the multi-bin method. All Limits are given at 95% CL.

Method	Observed limit (95% CL)	Expected limit (95% CL)
single-bin	$m_{W'} > 3.06$ TeV	$m_{W'} > 3.14$ TeV
multi-bin	$m_{W'} > 3.24$ TeV	$m_{W'} > 3.18$ TeV

17 SSM W' with interference

This chapter is about the limit determination in the SSMS and SSMO scenario, where the W' interferes with the Standard Model W . As explained in chapter 14, the interference alters the signal shape. The signal will be partly positive and partly negative (see figure 14.5) depending on the scenario, $m_{W'}$ and M_T . This leads to additional challenges when setting limits. A single-bin method as explained in chapter 16 is clearly suboptimal. Merging the signal into a single-bin bin would lead to a cancellation of positive and negative signal contributions. This issue does not occur, when using a method with multiple bins. For this reason and because of the improved sensitivity for low masses, the multi-bin approach is chosen here. In a first attempt, limits are calculated with the Bayesian multi-bin method presented in chapter 16. Then, a second method based on CLs is introduced. This method solves a major issue encountered with the first approach which will be explained later.

17.1. Bayesian multi-bin method

Just like previously, the combine tool needs the Standard Model background and the signal as separate input. While this is trivial in case of no interference where Standard Model W and W' are simulated separately, it demands an additional treatment of the signal samples in the SSMO and SSMS scenario. Due to the interference, W' and Standard Model W have to be simulated together. Thus the latter one has to be subtracted to obtain the 'pure' signal samples.

The subtraction of the Standard Model W contribution will lead to a negative signal in some part of the M_T range. This happens in M_T regions, where the interference is destructive and strong enough to lead to an event yield which is smaller than the event yield from the Standard Model W alone. In the SSMS scenario, this region is located between m_W and $m_{W'}$. Figure 14.5 shows the three W' scenarios together with the Standard Model W for a $m_{W'}$ of 2.5 TeV. For this mass point, the region $M_T < 1350$ GeV, where the SSMS $W' +$ Standard Model W (green curve) is below the Standard Model W (purple curve), will have a negative signal prediction. In the SSMO scenario, destructive interference occurs for masses larger than $m_{W'}$. Here, the remaining signal contribution is small and the effect is much less important for the limit calculation compared to the SSMS scenario. Nevertheless it will lead to a negative signal at some point. Examples for the signal before and after subtraction are shown in figure 17.1 for both scenarios. For the Standard Model W subtraction, the same samples as for the background evaluation are used. A similar fit as described in chapter 11 is performed to obtain a smooth parametrization of the W contribution.

In principle the obtained pure signal samples (c.f. figure 17.1) can be used as input for the multi-bin method described in section 16.2. Unfortunately the current implementation of the multi-bin method in the combine tool can not treat negative signals with shape uncertainties. Therefore, the limit calculation has to be restricted to the M_T range $[M_T^{\min}, M_T^{\max}]$ where the signal is positive. In the SSMS scenario, M_T^{\min} increases with $m_{W'}$ and ranges between 220 GeV ($m_{W'} = 300$ GeV) and 2730 GeV ($m_{W'} = 4000$ GeV).

17. SSM W' with interference

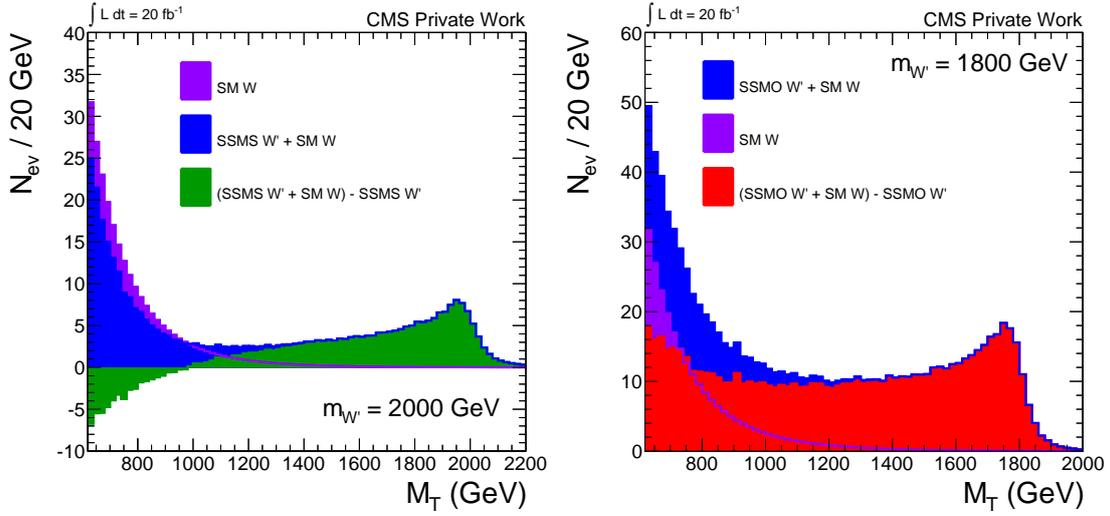


Figure 17.1. W' signal before and after the subtraction of the Standard Model W in the SSMS (left) and SSMO (right) scenario for two different $m_{W'}$.

For the SSMS mass point shown in figure 17.1, M_T^{\min} is 1020 GeV. The M_T^{\max} value for the SSMO scenario is 380 GeV for $m_{W'} = 300$ GeV and reaches 4000 GeV for the 4000 GeV W' sample. A list of all M_T^{\min} and M_T^{\max} values can be found in table A.5.

The style of the exclusion plot differs from the plots shown so far. Here, the ratio of the excluded cross-section and the cross-section predicted in the SSMS /SSMO model is shown on the y-axis. For every mass point, where the limit is below the red line marking $\sigma_{\text{excl.}}/\sigma_{\text{SSM } W'}$ values of 1, the model is excluded at 95% CL. This plotting style is chosen, since the W' signal cross-section without the W contribution is not a well defined quantity in scenarios with interference. In the SSMO scenario a W' can be excluded at 95% CL for $m_{W'}$ up to 3.5 TeV (both expected and observed). The exclusion limit in the SSMS scenario is 3.0 TeV compared to an expected limit of 3.1 TeV.

17.2. Hypothesis testing with CLs

Especially in the SSMS scenario, the restriction to the positive signal region is unsatisfying. Here, the qualitative difference to the other scenarios is the region where the signal leads to a decrease of expected events compared to the Standard Model. To use the whole signal range in both scenarios, a different approach is used. Following the recommendations of the statistics committee, the limit setting is done in form of a hypothesis test using the CLs method. Here, the Standard Model hypothesis (H_0) is tested against the Standard Model + SSM W' hypothesis (H_1).

The construction of H_0 is straight forward. It is based on the Standard Model expectation derived in chapter 11, which is also used for the Bayesian limits. For H_1 , the construction is a bit more complicated. Adding all non- W Standard Model backgrounds to the generated $W + W'$ samples is not sufficient, since the W in the signal samples is produced at LO. Instead, the LO W contribution is subtracted from the signal samples (see Bayesian approach), scaled with the M_T binned k-factor and added back again to the 'pure' signal samples. This treatment ensures, that the Standard Model predictions are the same for both hypotheses.

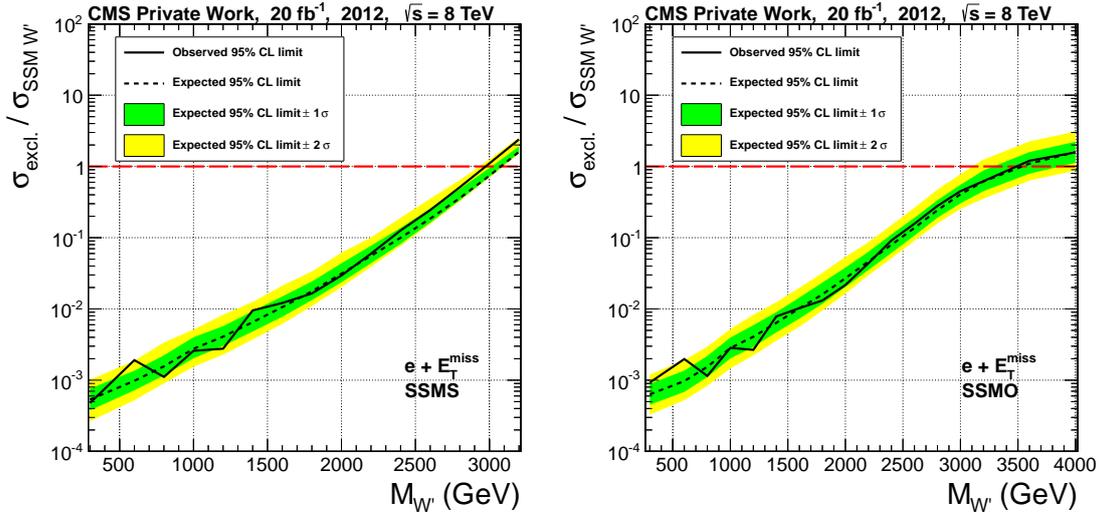


Figure 17.2. Upper limits on the ratio $\sigma_{\text{excl.}}/\sigma_{\text{SSM } W'}$ in the SSMS (left) and SSMO (right) scenario. All limits at 95% CL.

For both hypothesis a large number (between $5 \cdot 10^5$ and $1 \cdot 10^7$ per mass point and hypothesis) of pseudo experiments are drawn. With each of the pseudo experiments, the test statistic $-2\ln(Q)$ (see equation 15.2) is calculated. This leads to frequency distributions of the test statistic for H_0 and H_1 . An example is shown in figure 17.3 for a $m_{W'}$ of 4 TeV in the SSMO scenario. The black line in figure 17.3 is the test statistic result obtained using the measured data. Based on these distributions the p-values p_0 and p_μ can be calculated. The integral of the distributions' tails up to the data line (cf. green areas in figure 17.3) divided by the integral over the whole distribution corresponds to the probability to obtain a $-2\ln(Q)$ value which is equal or more extreme than the data value and thus to p_0 and p_μ . With p_0 and p_μ it is possible to calculate the quantity CLs according to equation 15.3. The calculation of the expected limit is done using the median of the H_0 distribution as 'data' line. For the 1σ and 2σ band, the test statistic values which form the 68.2% and 95.4% interval around the median of the H_0 distribution are taken as integration border.

The resulting limits are shown in figure 17.4-left for the SSMS and figure 17.4-right for the SSMO scenario. Here, the CLs value is shown as a function of the W' mass. Every mass point leading to a CLs value below the red line (CLs = 0.05) can be excluded at 95% CL. With the CLs method a SSMS (SSMO) W' can be excluded for masses below 3.4 TeV (3.55 TeV) compared to an expected limit of 3.5 TeV (3.6 TeV).

17.3. Conclusion

With the CLs method it was possible to include the negative signal contributions into the limit leading to a significantly better limit in the SSMS scenario. The SSMO limit is slightly better in the CLs approach. Possible reasons for this are the increased M_T range used for the limit, conceptual differences between the two methods and the fact that only two signal samples are available in the region between 3.2 TeV and 3.6 TeV and thus the mass range where the limit has to be interpolated is rather large.

17. SSM W' with interference

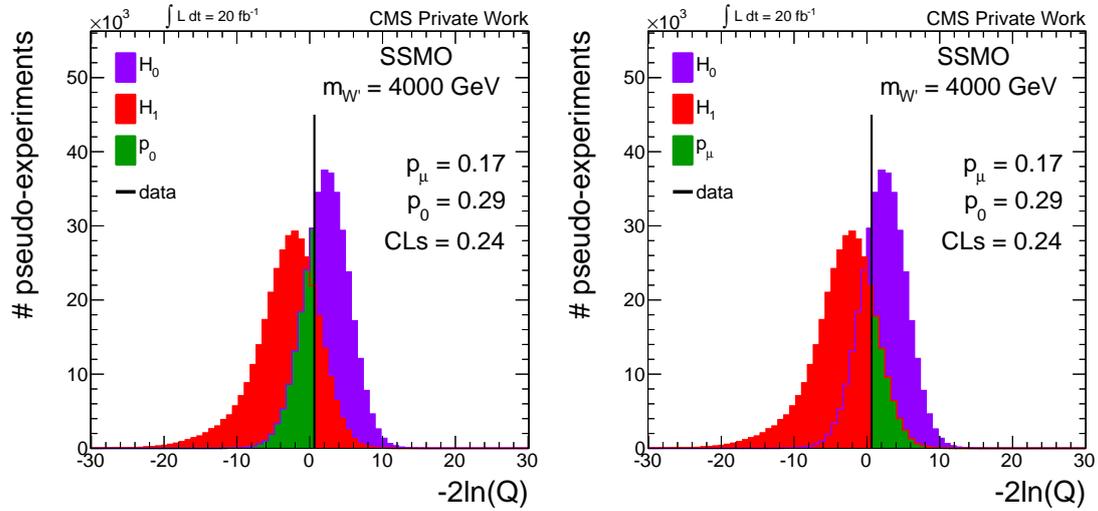


Figure 17.3. Test statistic distribution of the pseudo-experiments for the Standard Model hypothesis (H_0) and the Standard Model + SSM W' hypothesis (H_1). Also shown is the outcome of the test statistic obtained with the measured data. The integral of the green area in the left (right) plot divided by the integral over the whole H_0 (H_1) distribution, corresponds to the p value of the H_0 (H_1) hypothesis. These p values are used for the calculation of CLs.

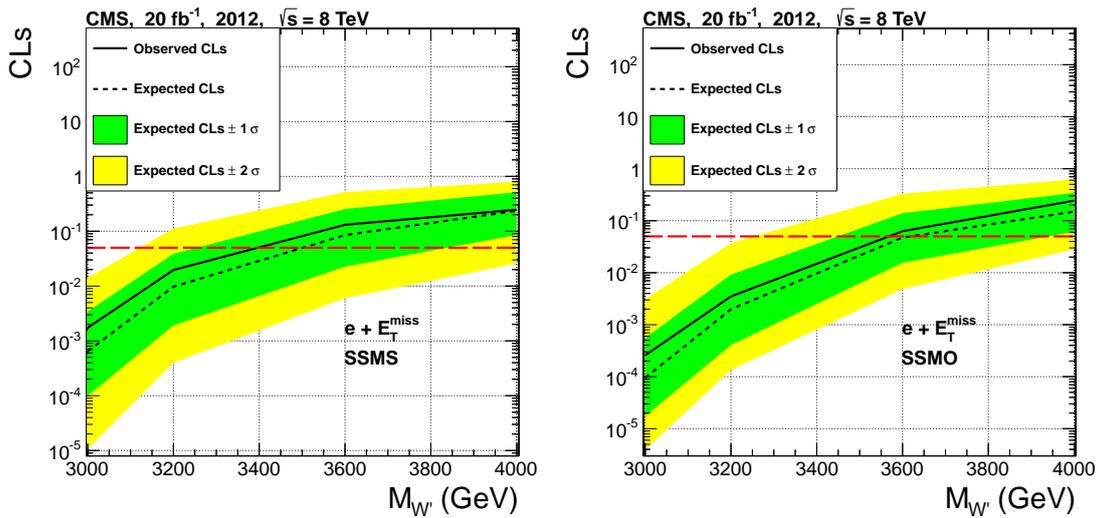


Figure 17.4. CLs limit in the SSMS (left) and SSMO (right) W' scenario. For every W' mass, where the limit is below the red line ($CLs=0.05$) can be excluded at 95% CL.

18 HNC CI

In this chapter, the upper cross-section limit for the HNC-CI model is presented. Figure 18.1-left shows the expected Standard Model background together with the measured data and the CI signal for four different values of Λ . As explained in chapter 14, only one simulated sample is used which can be rescaled for different Λ . The same Bayesian limit approach as presented in chapter 16 is applied. The M_T range which is used for the limit calculation is restricted to the range in which the signal was produced (see figure 18.1-left).

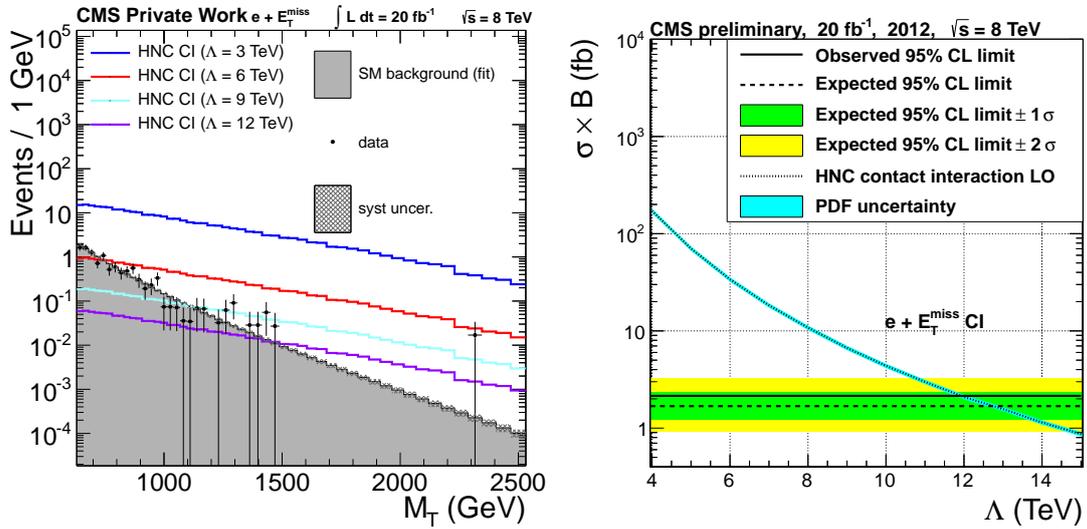


Figure 18.1. Left: M_T distribution for expected Standard Model background (fit presented in chapter 11), data and four different CI samples. The curves are drawn in front of each other. Right: Upper cross-section limit for the HNC-CI model.

Figure 18.1-right shows the resulting limit on cross-section times branching fraction. The limit is calculated with the one simulated sample and used for all values of Λ . This can be done, since the signal shape does not depend on Λ and simply scales as a function of Λ . Thus, the cross-section which can be excluded does not depend on Λ . A HNC-CI can be excluded at 95% CL for Λ smaller than 12.0 TeV compared to an expected limit of 12.7 TeV.

19 Summary and conclusions

In this thesis, a search for new physics in the $e + E_{\text{T}}^{\text{miss}}$ channel was presented. It was performed using the whole 2012 pp collision dataset recorded with the CMS detector. This dataset includes roughly 20 fb^{-1} of pp collision at $\sqrt{s} = 8 \text{ TeV}$.

The analysis is based on and extends previous theses about searches of a similar kind [29, 61–64]. A large set of different Monte Carlo simulations has been used to provide a good description of the Standard Model contribution to the $e + E_{\text{T}}^{\text{miss}}$ spectrum up to high M_{T} . Higher order QCD corrections to Standard Model processes have been included. For the main background - the Standard Model W - also electroweak higher order corrections were incorporated. To improve the Standard Model prediction, a data-driven approach was used for the estimation of multi-jet processes for which the Monte Carlo simulation is known to be suboptimal. The impact of systematic uncertainties on the $e + E_{\text{T}}^{\text{miss}}$ spectrum was studied up to high M_{T} , which is a crucial ingredient for the interpretation of the analysis results with statistical methods. No significant deviation of the data from the Standard Model prediction has been found.

The analysis results were interpreted in terms of different signal hypotheses including the SSM W' scenario without interference effects, two SSM W' scenarios with interference effects (SSMS/SSMO) and the HNC contact interaction scenario. Exclusion limits were set for all four scenarios. In previous theses and publications [23, 29, 32, 61, 62, 64, 114], a single-bin method was used for the limit-setting procedure. Here, a generalization of the single-bin method to a multi-bin method was realized. The new method is sensitive to the shape of the signal and was tested to be stable against variations of the bin-size. A comparison of the two methods in the SSM W' scenario without interference showed a significant increase of the sensitivity for low $m_{W'}$ (≈ 6 times better cross-section limit for $m_{W'} = 300 \text{ GeV}$) and a slightly better sensitivity for high $m_{W'}$ by using the new method. With help of the multi-bin method, it was possible to set exclusion limits in the SSMS and SSMO scenario including the M_{T} range where the interference is destructive and leads to negative signal contributions. In the SSMS scenario, the destructive interference is of special interest, since observing a deficit in the high M_{T} region could be a first hint for the existence of a very heavy W' . A summary of the obtained limits is shown in table 19.1.

The results of this analysis will be published [1] together with the results from a similar analysis in the muon channel [65]. Combining the electron and muon channel, a SSM W' without interference can be excluded for masses up to 3.35 TeV. This is the current world's best limit. In [1], the interpretation of the analysis results in terms of additional signal models can be found. The additional models include a Dark Matter model with different scenarios and models with extra dimensions. Based on the work in this thesis, a further improvement of the SSM W' limits with interference was developed by Mark Olschewski. This new approach is called 'generalized couplings' and allows the variation of the W' coupling strength to fermions, which was fixed to have the same absolute value

19. Summary and conclusions

Table 19.1. Summary of the exclusion limits. The W' limits are set on the W' mass $m_{W'}$. W' bosons with masses below the limit can be excluded. In the CI hypothesis, the limit is set on the energy scale Λ . Also shown is the SSM W' limit after combination with the muon channel (see [65] for details). All Limits are given at 95% CL.

Model	Observed limit (95% CL)	Expected limit (95% CL)
SSM W'	$m_{W'} > 3.24 \text{ TeV}$	$m_{W'} > 3.18 \text{ TeV}$
SSMS W'	$m_{W'} > 3.40 \text{ TeV}$	$m_{W'} > 3.50 \text{ TeV}$
SSMO W'	$m_{W'} > 3.55 \text{ TeV}$	$m_{W'} > 3.60 \text{ TeV}$
HNC CI	$\Lambda > 12.0 \text{ TeV}$	$\Lambda > 12.7 \text{ TeV}$
SSM W' e + μ [1]	$m_{W'} > 3.35 \text{ TeV}$	$m_{W'} > 3.30 \text{ TeV}$

as the Standard Model W boson in this analysis. The variation of the coupling strength for a fixed $m_{W'}$ leads to a variation of the signal shape. Results and details about the generalized couplings can be found in [1].

A Appendix

Table A.1. Overview of the Standard Model Monte Carlo samples. All samples are from the official CMS Monte Carlo production (prod. campaign **Summer12**, frontier conditions **START53.V7A**, see PREP [71] for full dataset names). Here, σ_{prod} is the production cross-section (LO for PYTHIA and MADGRAPH [116], NLO for POWHEG and MC@NLO). If higher order corrections are available, they are included by applying a k-factor to σ_{prod} . The corrected cross-sections are given in the column labelled σ .

Generator	Process	Kinematic cuts (GeV)	σ_{prod} (pb)	σ (pb)	N_{ev}
PYTHIA	$W \rightarrow e\nu$	-	9140	see 8.2.1	$\sim 5\text{M}$
PYTHIA	$W \rightarrow e\nu$	$100 < \hat{p}_T < 500$	1.457	see 8.2.1	$\sim 1\text{M}$
PYTHIA	$W \rightarrow e\nu$	$\hat{p}_T > 500$	0.001525	see 8.2.1	$\sim 1\text{M}$
PYTHIA	$W \rightarrow \tau\nu$	-	9170	see 8.2.1	$\sim 5\text{M}$
PYTHIA	$W \rightarrow \tau\nu$	$100 < \hat{p}_T < 500$	1.457	see 8.2.1	$\sim 1\text{M}$
PYTHIA	$W \rightarrow \tau\nu$	$\hat{p}_T > 500$	0.001525	see 8.2.1	$\sim 1\text{M}$
PYTHIA	QCD EM enriched	$30 < \hat{p}_T < 80$	4615893	-	$\sim 35\text{M}$
PYTHIA	QCD EM enriched	$80 < \hat{p}_T < 170$	183294	-	$\sim 35\text{M}$
PYTHIA	QCD EM enriched	$170 < \hat{p}_T < 250$	4587	-	$\sim 35\text{M}$
PYTHIA	QCD EM enriched	$250 < \hat{p}_T < 350$	556.75	-	$\sim 35\text{M}$
PYTHIA	QCD EM enriched	$\hat{p}_T > 350$	89.1	-	$\sim 35\text{M}$
MC@NLO	$t\bar{t}$	-	211.1	245.8 (NNLO)	$\sim 33\text{M}$
POWHEG	$t \rightarrow b\nu$ (s-Channel)	-	2.82	3.79 (approx NNLO)	$\sim 260\text{K}$
POWHEG	$t \rightarrow b\nu$ (t-Channel)	-	47.0	56.4 (approx NNLO)	$\sim 100\text{K}$
POWHEG	$t \rightarrow b\nu$ (tW-Channel)	-	10.7	11.1 (approx NNLO)	$\sim 500\text{K}$
POWHEG	$\bar{t} \rightarrow b\nu$ (s-Channel)	-	1.57	1.76 (approx NNLO)	$\sim 140\text{K}$
POWHEG	$\bar{t} \rightarrow b\nu$ (t-Channel)	-	25.0	30.7 (approx NNLO)	$\sim 2\text{M}$
POWHEG	$\bar{t} \rightarrow b\nu$ (tW-Channel)	-	10.7	11.1 (approx NNLO)	$\sim 500\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 20$	1871	1915 (NNLO)	$\sim 50\text{M}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 120$	11.89	12.17 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 200$	1.48	1.52 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 400$	0.1085	0.1112 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 500$	0.04409	0.04519 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 700$	0.01025	0.01049 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 800$	0.005491	0.005620 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 1000$	0.001796	0.001838 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 1500$	0.0001705	0.0001745 (NNLO)	$\sim 100\text{K}$
POWHEG	$Z \rightarrow ee$	$m_{ee} > 2000$	0.0000221	0.0000225 (NNLO)	$\sim 100\text{K}$
PYTHIA	$Z \rightarrow \tau\tau$	$m_{\tau\tau} > 20$	1510	1915 (NNLO)	$\sim 19\text{M}$
PYTHIA	$Z \rightarrow \tau\tau$	$100 < m_{\tau\tau} < 200$	34.92	44.29 (NNLO)	$\sim 2\text{M}$
PYTHIA	$Z \rightarrow \tau\tau$	$200 < m_{\tau\tau} < 400$	1.181	1.498 (NNLO)	$\sim 1\text{M}$
PYTHIA	$Z \rightarrow \tau\tau$	$400 < m_{\tau\tau} < 800$	0.087	0.110 (NNLO)	$\sim 1\text{M}$
PYTHIA	$Z \rightarrow \tau\tau$	$m_{\tau\tau} > 800$	0.0045	0.0057 (NNLO)	$\sim 1\text{M}$
PYTHIA	$\gamma + \text{jets}$	$80 < p_T^{\text{jet}} < 120$	558.4	725.9	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$120 < p_T^{\text{jet}} < 170$	108.0	140.4	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$170 < p_T^{\text{jet}} < 300$	30.12	39.16	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$300 < p_T^{\text{jet}} < 470$	2.139	2.781	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$470 < p_T^{\text{jet}} < 800$	0.2119	0.2755	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$800 < p_T^{\text{jet}} < 1400$	0.007078	0.009201	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$1400 < p_T^{\text{jet}} < 1800$	0.00004510	0.00005863	$\sim 2\text{M}$
PYTHIA	$\gamma + \text{jets}$	$p_T^{\text{jet}} > 1800$	0.000001867	0.000002427	$\sim 2\text{M}$
PYTHIA	WW	-	33.6	54.83 (NLO)	$\sim 10\text{M}$
PYTHIA	WW	$\hat{p}_T > 500$	0.005235	0.008543 (NLO)	$\sim 1\text{M}$
PYTHIA	WZ	-	12.6	33.2 (NLO)	$\sim 10\text{M}$
PYTHIA	WZ	$\hat{p}_T > 500$	0.001695	0.004345 (NLO)	$\sim 1\text{M}$
PYTHIA	ZZ	-	5.20	17.7 (NLO)	$\sim 10\text{M}$
PYTHIA	ZZ	$\hat{p}_T > 500$	0.001065	0.001700 (NLO)	$\sim 1\text{M}$
MADGRAPH	$W \rightarrow \ell\nu$	-	30400	see 8.2.1	$\sim 80\text{M}$
MADGRAPH	$W \rightarrow \ell\nu$	$50 < p_T(W) < 70$	811.2	see 8.2.1	$\sim 53\text{M}$
MADGRAPH	$W \rightarrow \ell\nu$	$70 < p_T(W) < 100$	428.9	see 8.2.1	$\sim 22\text{M}$
MADGRAPH	$W \rightarrow \ell\nu$	$p_T(W) > 100$	228.9	see 8.2.1	$\sim 14\text{M}$
PYTHIA	QCD	$30 < \hat{p}_T < 50$	66290000.0	-	$\sim 6\text{M}$
PYTHIA	QCD	$50 < \hat{p}_T < 80$	8149000.0	-	$\sim 6\text{M}$
PYTHIA	QCD	$80 < \hat{p}_T < 120$	1034000.0	-	$\sim 300\text{K}$
PYTHIA	QCD	$120 < \hat{p}_T < 170$	156300.0	-	$\sim 1\text{M}$
PYTHIA	QCD	$170 < \hat{p}_T < 300$	34140.0	-	$\sim 1\text{M}$
PYTHIA	QCD	$300 < \hat{p}_T < 470$	1760.0	-	$\sim 1\text{M}$
PYTHIA	QCD	$470 < \hat{p}_T < 600$	113.9	-	$\sim 1\text{M}$
PYTHIA	QCD	$600 < \hat{p}_T < 800$	27.0	-	$\sim 1\text{M}$
PYTHIA	QCD	$800 < \hat{p}_T < 1000$	3.55	-	$\sim 500\text{K}$
PYTHIA	QCD	$1000 < \hat{p}_T < 1400$	0.7378	-	$\sim 500\text{K}$
PYTHIA	QCD	$1400 < \hat{p}_T < 1800$	0.03352	-	$\sim 500\text{K}$
PYTHIA	QCD	$\hat{p}_T > 1800$	0.001829	-	$\sim 500\text{K}$

Table A.2. Overview of the SSM W' signal Monte Carlo samples. The samples were generated with PYTHIA6 using the START53_V7A – v1 frontier conditions with the CTEQ6L1 PDF set and the S10 pileup scenario. For each mass point, 20000 events were produced. The full dataset name for each of the samples is /WprimeToENu.M-[.].TuneZ2star_8TeV-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM, where [.] has to be replaced with the W' mass. The NNLO k-factors and cross-sections are calculated for $\sqrt{s} = 8$ TeV and are taken from [108].

$m_{W'}(\text{GeV})$	σ_{LO} (pb)	k-factor	σ_{NNLO} (pb)
300	113.5	1.350	153.2
500	16.48	1.363	22.46
700	4.28	1.351	5.782
900	1.471	1.347	1.981
1100	0.5881	1.331	0.7828
1300	0.2588	1.317	0.3408
1500	0.1193	1.293	0.1543
1700	0.05781	1.257	0.07267
1900	0.02958	1.230	0.03638
2000	0.02123	1.214	0.02577
2100	0.01547	1.199	0.01855
2200	0.01127	1.194	0.01346
2300	0.008387	1.172	0.009830
2400	0.00622	1.164	0.007240
2500	0.004725	1.140	0.005387
2600	0.003574	1.152	0.004117
2700	0.002692	1.153	0.003104
2800	0.002102	1.145	0.002407
2900	0.001651	1.148	0.001895
3000	0.001319	1.151	0.001518
3100	0.001056	1.178	0.001244
3200	0.000869	1.187	0.001032
3300	0.0007067	1.207	0.0008530
3400	0.0005965	1.220	0.0007277
3500	0.0005071	1.242	0.0006298
3700	0.0003691	1.278	0.0004717
4000	0.0002507	1.331	0.0003337

Table A.3. Overview of the SSMO / SSMS W' signal Monte Carlo samples. The samples were generated with MADGRAPH using the START53_V7A – v1 frontier conditions with the CTEQ6L1 PDF set and the S10 pileup scenario. For the samples with $m_{W'} > 3000$ GeV and $m_{W'} = 300$ GeV the frontier conditions START53_V7C – v1 were used instead. For each mass point the simulation is divided into three subsamples with help of the electron p_T . Each of the subsamples contains 100000 events. The subsamples are divided by the p_T of the electron. The full dataset name for each of the samples is $/[.]_8\text{TeV}/\text{Summer12_DR53X-PU_S10.START53.V7[.]-v1/AODSIM}$, where the first $[.]$ has to be replaced with the name tag shown in the table and the second one with A or C depending on the frontier conditions which were used.

Sample	σ_{LO} (pb)	Sample	σ_{LO} (pb)
wprime_oppsign_e.M300_g1_pt150to100	74.4	wprime_samesign_e.M300_g1_pt150to100	74.4
wprime_oppsign_e.M300_g1_pt1100to200	77.3	wprime_samesign_e.M300_g1_pt1100to200	77.3
wprime_oppsign_e.M300_g1_pt1200to1d5	0.03	wprime_samesign_e.M300_g1_pt1200to1d5	0.03
wprime_oppsign_e.M600_g1_pt1100to300	9.38	wprime_samesign_e.M600_g1_pt1100to300	9.38
wprime_oppsign_e.M600_g1_pt1300to600	0.468	wprime_samesign_e.M600_g1_pt1300to600	0.468
wprime_oppsign_e.M600_g1_pt1600to1d5	2.38e-05	wprime_samesign_e.M600_g1_pt1600to1d5	2.38e-05
wprime_oppsign_e.M800_g1_pt1100to300	3.01	wprime_samesign_e.M800_g1_pt1100to300	3.01
wprime_oppsign_e.M800_g1_pt1300to600	1.36	wprime_samesign_e.M800_g1_pt1300to600	1.36
wprime_oppsign_e.M800_g1_pt1600to1d5	0.000117	wprime_samesign_e.M800_g1_pt1600to1d5	0.000117
wprime_oppsign_e.M1000_g1_pt1250to500	0.768	wprime_samesign_e.M1000_g1_pt1250to500	0.768
wprime_oppsign_e.M1000_g1_pt1500to1000	0.0507	wprime_samesign_e.M1000_g1_pt1500to1000	0.0507
wprime_oppsign_e.M1000_g1_pt11000to1d5	8.36e-07	wprime_samesign_e.M1000_g1_pt11000to1d5	8.36e-07
wprime_oppsign_e.M1200_g1_pt1250to500	0.245	wprime_samesign_e.M1200_g1_pt1250to500	0.245
wprime_oppsign_e.M1200_g1_pt1500to1000	0.16	wprime_samesign_e.M1200_g1_pt1500to1000	0.16
wprime_oppsign_e.M1200_g1_pt11000to1d5	2.21e-06	wprime_samesign_e.M1200_g1_pt11000to1d5	2.21e-06
wprime_oppsign_e.M1400_g1_pt1250to500	0.12	wprime_samesign_e.M1400_g1_pt1250to500	0.12
wprime_oppsign_e.M1400_g1_pt1500to1000	0.101	wprime_samesign_e.M1400_g1_pt1500to1000	0.101
wprime_oppsign_e.M1400_g1_pt11000to1d5	5.75e-06	wprime_samesign_e.M1400_g1_pt11000to1d5	5.75e-06
wprime_oppsign_e.M1600_g1_pt1250to500	0.0777	wprime_samesign_e.M1600_g1_pt1250to500	0.0777
wprime_oppsign_e.M1600_g1_pt1500to1000	0.0573	wprime_samesign_e.M1600_g1_pt1500to1000	0.0573
wprime_oppsign_e.M1600_g1_pt11000to1d5	1.6e-05	wprime_samesign_e.M1600_g1_pt11000to1d5	1.6e-05
wprime_oppsign_e.M1800_g1_pt1250to500	0.0603	wprime_samesign_e.M1800_g1_pt1250to500	0.0603
wprime_oppsign_e.M1800_g1_pt1500to1000	0.0321	wprime_samesign_e.M1800_g1_pt1500to1000	0.0321
wprime_oppsign_e.M1800_g1_pt11000to1d5	5.81e-05	wprime_samesign_e.M1800_g1_pt11000to1d5	5.81e-05
wprime_oppsign_e.M2000_g1_pt1250to500	0.0526	wprime_samesign_e.M2000_g1_pt1250to500	0.0526
wprime_oppsign_e.M2000_g1_pt1500to1000	0.0175	wprime_samesign_e.M2000_g1_pt1500to1000	0.0175
wprime_oppsign_e.M2000_g1_pt11000to1d5	0.000888	wprime_samesign_e.M2000_g1_pt11000to1d5	0.000888
wprime_oppsign_e.M2200_g1_pt1250to500	0.0483	wprime_samesign_e.M2200_g1_pt1250to500	0.0483
wprime_oppsign_e.M2200_g1_pt1500to1000	0.0085	wprime_samesign_e.M2200_g1_pt1500to1000	0.0085
wprime_oppsign_e.M2200_g1_pt11000to1d5	0.00248	wprime_samesign_e.M2200_g1_pt11000to1d5	0.00248
wprime_oppsign_e.M2400_g1_pt1250to500	0.0458	wprime_samesign_e.M2400_g1_pt1250to500	0.0458
wprime_oppsign_e.M2400_g1_pt1500to1000	0.00514	wprime_samesign_e.M2400_g1_pt1500to1000	0.00514
wprime_oppsign_e.M2400_g1_pt11000to1d5	0.00184	wprime_samesign_e.M2400_g1_pt11000to1d5	0.00184
wprime_oppsign_e.M2500_g1_pt1250to500	0.0449	wprime_samesign_e.M2500_g1_pt1250to500	0.0449
wprime_oppsign_e.M2500_g1_pt1500to1000	0.00424	wprime_samesign_e.M2500_g1_pt1500to1000	0.00424
wprime_oppsign_e.M2500_g1_pt11000to1d5	0.00147	wprime_samesign_e.M2500_g1_pt11000to1d5	0.00147
wprime_oppsign_e.M2600_g1_pt1250to500	0.0442	wprime_samesign_e.M2600_g1_pt1250to500	0.0442
wprime_oppsign_e.M2600_g1_pt1500to1000	0.00364	wprime_samesign_e.M2600_g1_pt1500to1000	0.00364
wprime_oppsign_e.M2600_g1_pt11000to1d5	0.00115	wprime_samesign_e.M2600_g1_pt11000to1d5	0.00115
wprime_oppsign_e.M2800_g1_pt1250to500	0.043	wprime_samesign_e.M2800_g1_pt1250to500	0.043
wprime_oppsign_e.M2800_g1_pt1500to1000	0.0029	wprime_samesign_e.M2800_g1_pt1500to1000	0.0029
wprime_oppsign_e.M2800_g1_pt11000to1d5	0.00068	wprime_samesign_e.M2800_g1_pt11000to1d5	0.00068
wprime_oppsign_e.M3000_g1_pt1250to500	0.0422	wprime_samesign_e.M3000_g1_pt1250to500	0.0422
wprime_oppsign_e.M3000_g1_pt1500to1000	0.00251	wprime_samesign_e.M3000_g1_pt1500to1000	0.00251
wprime_oppsign_e.M3000_g1_pt11000to1d5	0.000394	wprime_samesign_e.M3000_g1_pt11000to1d5	0.000394
wprime_oppsign_e.M3200_g1_pt1250to500	0.0416	wprime_samesign_e.M3200_g1_pt1250to500	0.0416
wprime_oppsign_e.M3200_g1_pt1500to1000	0.00228	wprime_samesign_e.M3200_g1_pt1500to1000	0.00228
wprime_oppsign_e.M3200_g1_pt11000to1d5	0.00023	wprime_samesign_e.M3200_g1_pt11000to1d5	0.00023
wprime_oppsign_e.M3600_g1_pt1250to500	0.0406	wprime_samesign_e.M3600_g1_pt1250to500	0.0406
wprime_oppsign_e.M3600_g1_pt1500to1000	0.00202	wprime_samesign_e.M3600_g1_pt1500to1000	0.00202
wprime_oppsign_e.M3600_g1_pt11000to1d5	8.91e-05	wprime_samesign_e.M3600_g1_pt11000to1d5	8.91e-05
wprime_oppsign_e.M4000_g1_pt1250to500	0.04	wprime_samesign_e.M4000_g1_pt1250to500	0.04
wprime_oppsign_e.M4000_g1_pt1500to1000	0.00189	wprime_samesign_e.M4000_g1_pt1500to1000	0.00189
wprime_oppsign_e.M4000_g1_pt11000to1d5	4.66e-05	wprime_samesign_e.M4000_g1_pt11000to1d5	4.66e-05

Table A.4. Overview of Monte Carlo samples used for the HNC CI analysis. The first sample is the combined $W + CI$ sample. The second sample is a W only sample generated with the same generator settings, which is used to subtract the W contribution from the combined sample. Each sample was generated with PYTHIA6 using the START53_V7A – v1 frontier conditions with the CTEQ6L1 PDF set and the S10 pileup scenario. The full dataset name is $/[.]_TuneZ2star_8\text{TeV-pythia6}/\text{Summer12_DR53X-PU_S10.START53.V7A-v1/AODSIM}$, where $[.]$ has to be replaced with the sample tag given in the table below.

Sample	σ_{LO} (pb)	N_{ev}
CItoENu_Lambda-3000_ptmin300	0.5691	100000
WTtoENu_MSUB166_Pt-300toInf	0.0185	1000000

Table A.5. Lower and upper bounds used for the Bayesian limits in the SSMS and SSMO scenario together with the number of expected background and signal events as well as the number of measured data events.

SSMO					
$m_{W'}$ (GeV)	M_T^{\min} (GeV)	M_T^{\max} (GeV)	N_{bg}	N_{sg}	N_{data}
300.0	220	380	20600.0	883000.0	21529
600.0	220	810	23200.0	118000.0	24420
800.0	220	990	23300.0	41500.0	24485
1000.0	570	1290	443.0	11000.0	478
1200.0	570	1450	446.0	5280.0	483
1400.0	570	1770	448.0	2640.0	484
1600.0	570	2150	449.0	1380.0	484
1800.0	570	2370	449.0	767.0	485
2000.0	570	2490	449.0	449.0	485
2200.0	570	2610	449.0	281.0	485
2400.0	570	2970	449.0	188.0	485
2600.0	570	3030	449.0	137.0	485
2800.0	570	3210	449.0	104.0	485
3000.0	570	3450	449.0	79.5	485
3200.0	570	3690	449.0	70.5	485
3600.0	570	3930	449.0	51.0	485
4000.0	570	4000	449.0	42.5	485
SSMS					
$m_{W'}$ (GeV)	M_T^{\min} (GeV)	M_T^{\max} (GeV)	N_{bg}	N_{sg}	N_{data}
300.0	220	1530	23300.0	895000.0	24508
600.0	240	3030	17300.0	95500.0	18278
800.0	300	2970	7210.0	29500.0	7711
1000.0	570	3690	449.0	9820.0	485
1200.0	570	3690	449.0	4310.0	485
1400.0	630	3810	281.0	1890.0	281
1600.0	720	3630	147.0	864.0	151
1800.0	870	3510	55.6	401.0	57
2000.0	1020	3030	23.2	188.0	22
2200.0	1170	3090	10.3	91.0	13
2400.0	1290	3150	5.6	43.0	7
2600.0	1490	3390	2.1	21.0	1
2800.0	1650	3570	1.0	9.8	1
3000.0	1810	3690	0.5	4.6	1
3200.0	2000	3870	0.2	2.1	1
3600.0	2370	3930	<0.1	0.4	0
4000.0	2730	4000	<0.1	0.1	0

Bibliography

- [1] CMS Collaboration, *Search for new physics in final states with a lepton and missing transverse energy, Paper in preparation CMS EXO-12-060* (2013) [URL].
- [2] G. M. Fraser, *The particle century*. IOP, Bristol, 1998.
- [3] G. Altarelli, *Collider Physics within the Standard Model: a Primer*, [arXiv:1303.2842].
- [4] D. Forero, M. Tortola, and J. Valle, *Global status of neutrino oscillation parameters after Neutrino-2012*, *Phys.Rev.* **D86** (2012) 073012, [arXiv:1205.4018].
- [5] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys.Lett.* **12** (1964) 132–133.
- [6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys.Rev.Lett.* **13** (1964) 321–323.
- [7] Beringer, J. et al. (Particle Data Group), *Review of Particle Physics (RPP)*, *Phys.Rev.* **D86** (2012) 010001.
- [8] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys.Rev.Lett.* **30** (1973) 1346–1349.
- [9] D. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, *Phys.Rev.* **D8** (1973) 3633–3652.
- [10] P. A. M. Dirac, *The Quantum Theory of the Electron*, *Proc.R.Soc.Lond* **A117** (1928) 610–624.
- [11] S. Weinberg, *A Model of Leptons*, *Phys.Rev.Lett.* **19** (1967) 1264–1266.
- [12] S. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl.Phys.* **22** (1961) 579–588.
- [13] A. Salam and J. C. Ward, *Partial Symmetries of Weak Interactions*, *Phys.Lett.* **13** (1964) 168–171.
- [14] CMS Collababoration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys.Lett.* **B716** (2012) 30–61, [arXiv:1207.7235].
- [15] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys.Lett.* **B716** (2012) 1–29, [arXiv:1207.7214].
- [16] A. Martin, W. Stirling, R. Thorne et al., *Parton distributions for the LHC*, *Eur.Phys.J.* **C63** (2009) 189–285, [arXiv:0901.0002].
- [17] M. Lindner, T. Ohlsson, and G. Seidl, *Seesaw mechanisms for Dirac and Majorana neutrino masses*, *Phys.Rev.* **D65** (2002) 053014, [hep-ph/0109264].
- [18] R. Mohapatra, *Unification and Supersymmetry: The Frontiers of Quark-Lepton Physics*. Graduate Texts in Contemporary Physics. Springer, 2003.
- [19] R. N. Mohapatra and J. C. Pati, *Left-Right Gauge Symmetry and an Isoconjugate Model of CP Violation*, *Phys.Rev.* **D11** (1975) 566–571.
- [20] D. Kim, Y. Oh, and S. C. Park, *W' in new physics models at the LHC*, [arXiv:1109.1870].

- [21] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for new heavy vector bosons in $p\bar{p}$ colliders*, *Z.Phys.* **C45** (1989) 109.
- [22] A. Papaefstathiou, *Phenomenological aspects of new physics at high energy hadron colliders*, [[arXiv:1108.5599](#)].
- [23] CMS Collaboration, *Search for leptonic decays of W' bosons in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **1208** (2012) 023, [[arXiv:1204.4764](#)].
- [24] ATLAS Collaboration, *ATLAS search for a heavy gauge boson decaying to a charged lepton and a neutrino in pp collisions at $\sqrt{s} = 7$ TeV*, *Eur.Phys.J.* **C72** (2012) 2241, [[arXiv:1209.4446](#)].
- [25] CDF Collaboration, *Search for a new heavy gauge boson W' with event signature electron + missing transverse energy in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. D* **83** (2011) 031102.
- [26] D0 Collaboration, *Search for W' bosons decaying to an electron and a neutrino with the D0 detector*, *Phys.Rev.Lett.* **100** (2008) 031804, [[arXiv:0710.2966](#)].
- [27] D0 Collaboration, *Search for W' Boson Resonances Decaying to a Top Quark and a Bottom Quark*, *Phys.Rev.Lett.* **100** (2008) 211803, [[arXiv:0803.3256](#)].
- [28] E. Eichten, K. D. Lane, and M. E. Peskin, *New Tests for Quark and Lepton Substructure*, *Phys.Rev.Lett.* **50** (1983) 811–814.
- [29] F. Schneider, *Search for New Physics in Final States with One Muon and Missing Transverse Energy with CMS Data*, Master's thesis, RWTH-Aachen University, Dec., 2012. [[URL](#)].
- [30] K. Lane, F. Paige, T. Skwarnicki et al., *Simulations of supercollider physics*, *Phys.Rept.* **278** (1997) 291–371, [[hep-ph/9412280](#)].
- [31] CDF Collaboration, *Search for quark lepton compositeness and a heavy W' boson using the $e\nu$ channel in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys.Rev.Lett.* **87** (2001) 231803, [[hep-ex/0107008](#)].
- [32] CMS Collaboration, *Search for new physics in final states with a lepton and missing transverse energy in pp collisions at the LHC*, *Phys.Rev.* **D87** (2013) 072005, [[arXiv:1302.2812](#)].
- [33] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001.
- [34] *LEP design report*. CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.
- [35] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [36] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [37] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.
- [38] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.
- [39] C. Lefevre, *LHC: the guide (English version). Guide du LHC (version anglaise)*, (Feb., 2009), [[URL](#)].

- [40] CMS Collaboration, *CMS Luminosity - Public Results (rev. 100)*, CMS twiki page, (2013), [URL].
- [41] S. Beauceron, *The CMS High Level Trigger, PoS ICHEP2012* (2013) 505.
- [42] I. Bird, K. Bos, N. Brook et al., *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005.
- [43] CMS Collaboration, *Electron Reconstruction in CMS, CMS Note 2006/40* (2006) [URL].
- [44] CMS Collaboration, *Electron Reconstruction in CMS, CMS Analysis Note AN-09-164* (2009) [URL].
- [45] HEEP Group, *HEEP Electron ID (rev. 65)*, CMS twiki page, (2012), [URL].
- [46] CMS Collaboration, *Electron Reconstruction in the CMS Electromagnetic Calorimeter, CMS Note 2001/34* (2001) [URL].
- [47] CMS Collaboration, *Review of clustering algorithms and energy corrections in ECAL, CMS Internal Note IN-10-008* (2010) [URL].
- [48] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting, Nucl.Instrum.Meth. A262* (1987) 444–450.
- [49] P. Billoir, *Progressive track recognition with a Kalman like fitting procedure, Comput.Phys.Commun. 57* (1989) 390–394.
- [50] W. Adam, R. Frühwirth, A. Strandlie et al., *Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC, J.Phys.G:Nucl.Part.Phys. 31* (2005) N9, [physics/0306087].
- [51] C. Charlot, *The gsf track reconstruction*, EGM Global Event Description Workshop, Paris, (Jan., 2013), [URL].
- [52] CMS Collaboration, *Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV, JINST 8* (2013) P09009, [arXiv:1306.2016].
- [53] R. Paramatti, *The standard ecal superclustering*, EGM Global Event Description Workshop, Paris, (Jan., 2013), [URL].
- [54] M. Peruzzi, *Summary of the electron and photon energy corrections*, EGM Global Event Description Workshop, Paris, (Jan., 2013), [URL].
- [55] CMS Collaboration, *Missing transverse energy performance of the CMS detector, JINST 6* (2011) P09001, [arXiv:1106.5048].
- [56] CMS Collaboration, *Performance of Missing Transverse Energy Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector, CMS Physics Analysis Summary PAS-JME-12-002* (2013) [URL].
- [57] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET, CMS Physics Analysis Summary PAS-PFT-09-001* (2009) [URL].
- [58] CMS Collaboration, *Electron Reconstruction within the Particle Flow Algorithm, CMS Analysis Note AN-10-034* (2010) [URL].

- [59] CMS Collaboration, *Introduction to Jet Energy Corrections at CMS (rev. 3)*, CMS twiki page, (2011), [URL].
- [60] JetMET Physics Object Group, *Formulae for pfMET Type-I and Type-II Corrections (rev. 12)*, CMS twiki page, (2013), [URL].
- [61] S. Thüer, *Search for heavy charged gauge bosons in the decay channel $W' \rightarrow \mu\nu$ with CMS data*, Master's thesis, RWTH-Aachen University, Apr., 2012. [URL].
- [62] J.-F. Schulte, *Search for heavy charged gauge bosons in the decay channel $W' \rightarrow e\nu$ with CMS data*, Master's thesis, RWTH-Aachen University, June, 2012. [URL].
- [63] W. Bender, *Studie zum Nachweis schwerer, geladener Eichbosonen in leptonischen Zerfallskanälen mit dem CMS-Detektor*, Diplomarbeit, RWTH-Aachen University, Oct., 2008. [URL].
- [64] S. Knutzen, *Search for heavy vector bosons W' in the tau decay channel using pp collisions at 8 TeV*, Master's thesis, RWTH-Aachen University, Mar., 2013. [URL].
- [65] S. Erdweg, *Search for dark matter and W' in the final state with a muon and missing transverse energy with CMS*, Master's thesis, RWTH-Aachen University, Dec., 2013. [URL].
- [66] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.
- [67] C. Magass and others, *ACSusyAnalysis Software Package (rev. 105)*, CMS twiki page, (2013), [URL].
- [68] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, *Nucl.Instrum.Meth.* **A389** (1997) 81–86.
- [69] GEANT4 Collaboration, *GEANT4: A Simulation toolkit*, *Nucl.Instrum.Meth.* **A506** (2003) 250–303.
- [70] GEANT4 Collaboration, *Geant4 developments and applications*, *IEEE Trans.Nucl.Sci.* **53** (2006) 270.
- [71] F. Cossutti, P. Lenzi, N. Naziridis et al., *PREP: Production and reprocessing management tool for CMS*, *J.Phys.Conf.Ser.* **396** (2012) 052023, [URL].
- [72] P. Lenzi, S. Padhi, G. Gomez et al., *Standard Model Cross Sections for CMS at 8 TeV (rev. 27)*, CMS twiki page, (2013), [URL].
- [73] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **0605** (2006) 026, [hep-ph/0603175].
- [74] M. Olschewski, *W' and contact interaction background determination*, CMS Exotica Non-Hadronic Meeting, (Aug., 2012), [URL].
- [75] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **0206** (2002) 029, [hep-ph/0204244].
- [76] C. Carloni Calame, G. Montagna, O. Nicrosini et al., *Precision electroweak calculation of the production of a high transverse-momentum lepton pair at hadron colliders*, *JHEP* **0710** (2007) 109, [arXiv:0710.1722].

- [77] H.-L. Lai, M. Guzzi, J. Huston et al., *New parton distributions for collider physics*, *Phys.Rev.* **D82** (2010) 074024, [[arXiv:1007.2241](#)].
- [78] J. Pumplin, D. Stump, J. Huston et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **0207** (2002) 012, [[hep-ph/0201195](#)].
- [79] G. Balossini, G. Montagna, C. M. Carloni Calame et al., *Combination of electroweak and QCD corrections to single W production at the Fermilab Tevatron and the CERN LHC*, *JHEP* **1001** (2010) 013, [[arXiv:0907.0276](#)].
- [80] S. Brensing, S. Dittmaier, M. Kramer et al., *Radiative corrections to W^- boson hadroproduction: Higher-order electroweak and supersymmetric effects*, *Phys.Rev.* **D77** (2008) 073006, [[arXiv:0710.3309](#)].
- [81] S. Alioli, P. Nason, C. Oleari et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **0909** (2009) 111, [[arXiv:0907.4076](#)].
- [82] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur.Phys.J.* **C71** (2011) 1547, [[arXiv:1009.2450](#)].
- [83] CMS Physics Validation Group, *Pileup Scenarios (rev. 6)*, CMS twiki page, (2012), [[URL](#)].
- [84] M. Hildreth, *Pileup reweighting*, CMSDASia - CMS Data Analysis School in Taipei, Taiwan, (Sept., 2012), [[URL](#)].
- [85] CMS Physics Validation Group, *Pileup studies (rev. 32)*, CMS twiki page, (2013), [[URL](#)].
- [86] Egamma Physics Object Group, *E/gamma HLT Algorithm work 2012 (rev. 14)*, CMS twiki page, (2012), [[URL](#)].
- [87] JetMET Physics Object Group, *MET Optional Filters (rev. 62)*, CMS twiki page, (2013), [[URL](#)].
- [88] CMS Collaboration, *Search for heavy narrow dilepton resonances in pp collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV*, *Phys.Lett.* **B720** (2013) 63–82, [[arXiv:1212.6175](#)].
- [89] C. Seez, *Shower shape variables (rev. 8)*, CMS twiki page, (2011), [[URL](#)].
- [90] Egamma Physics Object Group, *E/gamma Commissioning and physics deliverables (rev. 16)*, CMS twiki page, (2012), [[URL](#)].
- [91] B. Clerbaux and L. Thomas, *HEEP selection efficiencies estimated from data (Tag & Probe method) (rev. 19)*, CMS twiki page, (2013), [[URL](#)].
- [92] L. Thomas, *Update on the HEEP ID efficiency and scale factor*, CMS Egamma Meeting, (Jan., 2013), [[URL](#)].
- [93] CMS Collaboration, *Search for new physics in the single lepton + MET final states with the full 2012 dataset at $\sqrt{s} = 8$ TeV*, *CMS Analysis Note* **AN-12-423** (2012) [[URL](#)].
- [94] CMS Collaboration, *Measuring Electron Efficiencies at CMS with Early Data*, *CMS Physics Analysis Summary* **PAS-EGM-07-001** (2008) [[URL](#)].
- [95] CMS Collaboration, *Search for a Higgs boson decaying into two photons in proton-proton collisions recorded by the CMS detector at the LHC*, *CMS Analysis Note* **AN-11-129** (2011) [[URL](#)].

- [96] CMS Collaboration, *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*, *CMS Physics Analysis Summary* **PAS-LUM-13-001** (2013) [URL].
- [97] CMS Collaboration, *Search for High Mass Resonances Decaying to Electron Pairs at 8 TeV with the Full 2012 dataset*, *CMS Analysis Note* **AN-12-415** (2012) [URL].
- [98] CMS Collaboration, *ECAL Detector Performance, 2011 Data*, *CMS Detector Performance Note* **DP-12-007** (2012) [URL].
- [99] JetMET POG, *Official Prescription for calculating uncertainties on Missing Transverse Energy (MET) (rev. 20)*, CMS twiki page, (2012), [URL].
- [100] CMS Collaboration, *PAT Tools (rev. 52)*, CMS twiki page, (2013), [URL].
- [101] Physics Validation Group, *Utilities for Accessing Pileup Information for Data (rev. 13)*, CMS twiki page, (2012), [URL].
- [102] D. Bourilkov, R. C. Group, and M. R. Whalley, *LHAPDF: PDF use from the Tevatron to the LHC*, [hep-ph/0605240].
- [103] S. Alekhin, S. Alioli, R. D. Ball et al., *The PDF4LHC Working Group Interim Report*, [arXiv:1101.0536].
- [104] A. Vicini, *Practical Implementation of the PDF4LHC recipe*, [URL].
- [105] M. Whalley, D. Bourilkov, and R. Group, *The Les Houches accord PDFs (LHAPDF) and LHAGLUE*, [hep-ph/0508110].
- [106] A. Martin, W. Stirling, R. Thorne et al., *Uncertainties on $\alpha(S)$ in global PDF analyses and implications for predicted hadronic cross sections*, *Eur.Phys.J.* **C64** (2009) 653–680, [arXiv:0905.3531].
- [107] R. D. Ball, V. Bertone, S. Carrazza et al., *Parton distributions with LHC data*, *Nucl.Phys.* **B867** (2013) 244–289, [arXiv:1207.1303].
- [108] CMS Collaboration, *PDF Uncertainties and K-factor for the W' Search at 8 TeV Collisions*, *CMS Analysis Note* **AN-12-172** (2012) [URL].
- [109] CMS Collaboration, *Documentation of the RooStats-based statistics tools for Higgs PAG (rev. 106)*, CMS twiki page, (2013), [URL].
- [110] L. Moneta, K. Belasco, K. S. Cranmer et al., *The RooStats Project*, *PoS* **ACAT2010** (2010) 057, [arXiv:1009.1003].
- [111] S. A. Schmitz, *Search for Large Spatial Extra Dimensions with Dimuon Events from 7 TeV pp collisions at CMS*. PhD thesis, RWTH-Aachen University, July, 2013. [URL].
- [112] A. L. Read, *Modified frequentist analysis of search results (The $CL(s)$ method)*, *CERN-OPEN-2000-205* (2000).
- [113] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in summer 2011*, *ATL-PHYS-PUB-2011-011, ATL-COM-PHYS-2011-818, CMS-NOTE-2011-005* (2011).
- [114] CMS Collaboration, *Search for a heavy gauge boson W' in the final state with an electron and large missing transverse energy in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys.Lett.* **B698** (2011) 21–39, [arXiv:1012.5945].

- [115] J. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, [arXiv:1103.0354].
- [116] J. Alwall, P. Demin, S. de Visscher et al., *MadGraph/MadEvent v4: The New Web Generation*, *JHEP* **0709** (2007) 028, [arXiv:0706.2334].

Danksagung

Zum Abschluss dieser Arbeit möchte ich mich bei den Personen bedanken, die mich auf dem Weg durch das Studium und die Masterarbeit begleitet und unterstützt haben. Als erstes möchte ich mich bei meiner Familie bedanken, deren finanzielle, aber vor allem auch moralische Unterstützung mir dieses Studium erst ermöglicht haben.

Ein ganz besonderer Dank gebührt Dr. Kerstin Hoepfner, die durch intensive Betreuung, Beratung und Korrekturlesen zum Gelingen dieser Masterarbeit beigetragen hat.

Prof. Dr. Hebbeker danke ich für die Möglichkeit, an einem Thema aktueller Forschung zu arbeiten und diese Masterarbeit am III. Physikalischen Institut anzufertigen. Darüber hinaus danke ich Prof. Dr. Wiebusch für die Zweitkorrektur.

Des Weiteren bedanke ich mich bei der gesamten CMS Arbeitsgruppe für die gute Arbeitsatmosphäre und die Hilfe bei zahlreichen physikalischen und technischen Problemen. Ein besonderer Dank gilt hierbei aktuellen und ehemaligen Kollegen aus der W' Arbeitsgruppe: Sören Erdweg, Simon Knutzen, Mark Olschewski, Klaas Padeken, Fabian Schneider, Jan-Frederik Schulte und Sebastian Thüer. Simon Knutzen und Klaas Padeken danke ich darüber hinaus für das Korrekturlesen dieser Arbeit.

Zuletzt möchte ich mich bei meiner Mutter, meiner Schwester Sophie, meiner Freundin Kathrin und meinen Freunden bedanken, die mich in den stressreichen Phasen der Masterarbeit ertragen und unterstützt haben.

Selbstständigkeitserklärung

Hiermit erkläre ich schriftlich, dass ich die vorliegende Arbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Zitate wurden als solche kenntlich gemacht.

Aachen, den 16. Dezember 2013

Philipp Millet