Measurement of the Jet Momentum Resolution and Search for a light Standard Model Higgs Boson in the $H(b\bar{b})W(\ell\nu)$ Channel with the CMS Detector at the LHC

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT) genehmigte

DISSERTATION

von

Dipl.-Phys. Hauke Held
aus Oldenburg (Oldb.)

Tag der mündlichen Prüfung: 20. April 2012

Referent: Prof. Dr. Th. Müller, Institut für Experimentelle Kernphysik
Karlsruher Institut für Technologie (KIT)

Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik
Karlsruher Institut für Technologie (KIT)

meiner Familie

Abstract

The Higgs boson is the last particle predicted by the Standard Model which remains undetected. Its potential discovery was a main objective of the construction of the Large Hadron Collider (LHC) at CERN. Its exclusion would necessitate the existence of new physics beyond the Standard Model. A search for the Higgs boson decaying into two bottom quarks in association with the production of a leptonically decaying W boson is presented based on pp collision data recorded with the CMS experiment in 2011, corresponding to an integrated luminosity of $\mathcal{L} = 4.65\,\mathrm{fb}^{-1}$. Events are selected requiring the presence of an isolated charged lepton (electron or muon), missing transverse energy and two b-jets, which are clustered with the anti-$\mathrm{k_T}$ jet algorithm at first. The search is performed in a boosted event topology, where both the W boson and the Higgs boson candidates have high momenta and move back-to-back in the transverse detector plane. Artificial Neural Networks are employed to discriminate signal and background events. No significant excess of signal events is observed and upper $\mathrm{CL_s}$ 95% confidence level exclusion limits are calculated relative to the Standard Model prediction for different Higgs boson trial masses $m(H)$:

$m(H)$ [GeV/c^2]	110	115	120	125	130	135
$\mathrm{CL_s}$ expected	4.66	5.05	6.13	7.22	8.94	10.9
$\mathrm{CL_s}$ observed	4.45	6.33	6.65	7.49	8.95	11.6

The analysis is performed twice, based on different reconstruction techniques of the $H \to b\bar{b}$ decay. First, standard anti-$\mathrm{k_T}$ jets are used, then a fat-, sub- and filterjet clustering algorithm specifically designed for the reconstruction of boosted objects is applied. The latter approach shows possible improvements in the expected search sensitivity of $2 - 10\%$ and is considered for the 2012 data-taking period.

The precise knowledge of the jet p_T resolution is important for analyses with jets in the final state, and it is needed for the understanding of the mass resolution of the jet-based Higgs boson candidates. The first measurement of the jet p_T resolution at CMS applying the data-driven dijet asymmetry method to the 2010 pp dataset, corresponding to an integrated luminosity of $\mathcal{L} = 35.9\,\mathrm{pb}^{-1}$, is presented. The resolution in data is found to be systematically poorer compared to simulation. A resolution scaling method is applied to events from simulation to account for this discrepancy. The jet p_T resolutions are key ingredients to the measurements of the inclusive jet and b-jet cross sections, which are summarized.

Zusammenfassung

Das Higgs-Boson ist das letzte innerhalb des Standardmodells vorhergesagte Elementarteilchen, das bisher nicht nachgewiesen werden konnte. Seine potenzielle Entdeckung war einer der Hauptgründe für den Bau des Large Hadron Colliders (LHC) am CERN. Sein experimenteller Ausschluss würde die Existenz von neuer Physik jenseits des Standardmodells bedingen. Die vorliegende Arbeit stellt die Suche nach einem leichten Standardmodell-Higgs-Boson vor, das in Assoziation mit einem leptonisch zerfallenden W-Boson produziert wird und in zwei b-Jets zerfällt. Dazu wird ein pp-Kollisionsdatensatz analysiert, der mit dem Compact Muon Solenoid (CMS) Experiment im Jahr 2011 aufgezeichnet wurde und einer integrierten Luminosität von $\mathcal{L} = 4.65\,\mathrm{fb}^{-1}$ entspricht. Die Ereignisselektion basiert auf einem isolierten Lepton (Elektron oder Myon), fehlender transversaler Energie sowie zwei b-Jets, die zunächst mit dem anti-k_T-Jet-Algorithmus rekonstruiert werden. Zur Unterdrückung von Untergründen wird die Suche auf Ereignisse eingeschränkt, in denen sich die W-Boson- und Higgs-Boson-Kandidaten mit hohen Impulsen in der transversalen Detektorebene voneinander entfernen. Die Trennung von Signal- und Untergrundereignissen geschieht mit Hilfe Neuronaler Netze. In den Daten wird kein signifikanter Ereignisüberschuß beobachtet. Dementsprechend werden Ausschlußgrenzen gemäß eines 95% Konfidenzintervalls mit der CL_s-Methode relativ zur Standardmodell-Vorhersage für verschiedene Higgs-Boson-Massen-Hypothesen berechnet:

$m(H)$ [GeV/c^2]	110	115	120	125	130	135
CL_s erwartet	4.66	5.05	6.13	7.22	8.94	10.9
CL_s beobachtet	4.45	6.33	6.65	7.49	8.95	11.6

Zusätzlich wird die Analyse basierend auf einem Jet-Substruktur-Algorithmus durchgeführt, der speziell für die Rekonstruktion von hochenergetischen Zerfällen entwickelt wurde. Die Verwendung dieses Algorithmus ermöglicht eine Verbesserung der erwarteten Sensitivität der Suche um $2 - 10\%$ und ist damit potenziell interessant für die Analyse der im Jahr 2012 zu erwartenden Daten.

Die genaue Kenntnis der Jet-p_T-Auflösung ist wichtig für Analysen mit Jets im Endzustand und hat besondere Bedeutung für das Verständnis der Massen-Auflösung der Higgs-Boson-Kandidaten. Die erste Messung der Jet-p_T-Auflösung am CMS-Experiment anhand der Dijet-Asymmetrie-Methode wird auf Basis des pp-Datensatzes des Jahres 2010 entsprechend einer integrierten Luminosität von $\mathcal{L} = 35.9\,\mathrm{pb}^{-1}$ präsentiert. Die Messung in Datenereignissen zeigt eine systematisch schlechtere Auflösung im Vergleich zur Simulation. Diese Diskrepanz kann durch geeignete Skalierung der Simulationsereignisse korrigiert werden. Die Jet-p_T-Auflösungen sind besonders wichtige Bestandteile der Messungen der inklusiven Jet- und b-Jet-Wirkungsquerschnitte, deren Ergebnisse ebenfalls diskutiert werden.

Contents

1. Introduction

The grand opening of the Large Hadron Collider (LHC) underneath the Franco-Swiss border has generated vast public interest in the field of elementary particle research. Today, scientific news from CERN quickly spread well beyond the realms of the high energy physics community. This extraordinary attention originates from a widely held expectation:

Elementary particle physics is at the cross-roads. The Standard Model (SM) of particle physics has been tremendously successful in explaining the fundamental interactions between elementary particles for the last decades. However, as a consequence of the massiveness of matter, the Standard Model predicts the existence of the Higgs boson; the last particle to remain yet undetected.

The discovery of this elusive particle is one of the main objectives of the construction of the LHC and the collaboration of thousands of researches from all over the world. Therefore, the experimental proof of its existence would mark a great success for the validation of the Standard Model and would be a huge leap towards a better understanding of nature. On the other hand, the experimental exclusion of the Higgs boson might be considered an equally important "discovery", since the absence of the Higgs boson immediately necessitates the existence of new physics beyond the Standard Model.

The fundamental question of the existence of the Higgs boson is central to the work presented in this thesis, which can be divided into three main parts: theoretical framework, experimental setup and statistical methods describe the basis of the thesis, followed by the presentation of highlights from jet-physics measurements with 2010 collision data, which should be seen as pre-requisites for the main part, the search for a light Standard Model Higgs boson in the channel $H(b\bar{b})W(\ell\nu)$.

The following Chapter 2 describes the mathematical formulation of the Standard Model with a particular emphasis on the development of the Higgs mechanism. In addition, it gives an overview of current theoretical and experimental constraints to the search for the Higgs boson.

Chapter 3 introduces the experimental setup. It starts with an overview of the LHC accelerator chain and the four main LHC experiments. A detailed description is given for the Compact Muon Solenoid (CMS) experiment, which is used to collect the data presented the following.

The raw electronic signals from the detector cannot be analyzed without further

processing. The reconstruction of higher level physics objects using the CMS software framework is described in Chapter 4. A special emphasis within this chapter is put on the specification of different jet clustering algorithms, which are of great importance for the different studies presented in the following.

The efficient analysis of large amounts of data relies on multivariate methods, such as artificial Neural Networks, to classify different types of events. In addition, the interpretation of the analysis results is based on statistical tests. These methods are specified in detail in Chapter 5.

The understanding of (b-)jets plays an important role in the context of the search for a Higgs boson decaying into two b-jets. In this sense, the first measurement of the jet energy resolution of the CMS experiment with 2010 collision data is a necessary pre-requisite to the Higgs boson search, but also an important ingredient to all other analyses with jets in the final state. Chapter 6 presents this measurement using the dijet asymmetry method and shows results from the related inclusive (b-)jet cross section measurements.

Chapter 7 describes the search for a light Standard Model Higgs boson in the decay channel $H(b\bar{b})W(\ell\nu)$. The analysis is performed twice: First using the current default CMS jet reconstruction and then employing a dedicated sub-/filterjet algorithm proposed by theorists. The results of both approaches are evaluated and compared in terms of 95% confidence level exclusion limits.

Finally, the results of the Higgs boson search are compared to other public results in the concluding Chapter 8.

2. The Higgs Boson in the Standard Model

The Standard Model is the theoretical framework describing elementary particles and their interactions. Formed in the course of the 20^{th} century it is one of the most successful theoretical concepts of modern times. With the technological progress over the last 50 years it has been experimentally validated to ever higher precisions, with the exception of the important discovery of the elusive Higgs boson.

This chapter starts with a brief theoretical overview of the Standard Model and introduces its mathematical formulation with a particular emphasis on the development of the Higgs mechanism, which is needed to assert the coherence of the theory and explains the generation of particle masses. The second section summarizes the theoretical expectations on a Higgs search at the LHC, in particular their implications on this thesis.

2.1. The Standard Model of Elementary Particle Physics

2.1.1. Theoretical Overview

In the universe as we see it today, four fundamental forces mediate the interactions between 12 different matter particles. These forces are the electromagnetic force between electric charges, the weak force responsible e. g. for the β decay associated with the weak isospin, the strong force coupling to the so called color charge and the gravitational force between massive objects. The *Standard Model of Particle Physics* (SM) is the unified theoretical description of the first three and their interplay with the matter particles. The following phenomenological introduction is based on Reference [1].

Possessing half-integral spin, all matter particles are point-like *fermions*, which do not show any experimental evidence for further underlying structure. They are grouped into six colorless *leptons* and six colored *quarks* as summarized in Table 2.1. Each of them has a corresponding *anti-particle* which possesses equal mass, but opposite quantum numbers, namely electric charge, color and third component of the weak isospin. Both leptons and quarks can be ordered into three *families* or *generations*. While the particles of the first generation are the constituents of the ordinary matter surrounding us, the particles of the second and third generation solely appear in nature in high energy cosmic rays interacting with the atmosphere of the earth or in artificial particle collisions in high energy physics laboratories.

Table 2.1.: Fundamental particles in the Standard Model: leptons and quarks

fermions	generation 1	2	3	electric charge [e]	color charge	weak isospin (T_3)
leptons	ν_e	ν_μ	ν_τ	$-$	$-$	$+1/2$
	e	μ	τ	-1	$-$	$-1/2$
quarks	u	c	t	$+2/3$	r,g,b	$+1/2$
	d	s	b	$-1/3$	r,g,b	$-1/2$

Within the SM, the interactions between particles are described as an exchange of force mediating spin 1 *(vector-)bosons*. Table 2.2 summarizes their basic properties.

The electromagnetic force is mediated by photons which couple to the electric charge. Since they are massless objects the electromagnetic interaction has infinite range.

The bosons of the weak interaction, the $W^\pm$ and Z^0, are massive objects with a corresponding small lifetime and an interaction range of $O(10^{-3}\,\text{fm})$, which couple to the weak isospin. In the SM electromagnetic and weak interactions are unified to the *electroweak* force.

Finally, eight massless gluons are the mediating particles of the strong interaction which couple to color: *hadrons* are composite quark objects, which may be subgrouped into bound two-quark states which are called *mesons* and three-quark states which are called *baryons*. The experimental observation of the Δ^{++} baryon, consisting of three u quarks, motivates the introduction of the quantum number *color* for the quarks with the values blue, green and red in order to preserve Pauli's exclusion principle [2] for fermions. Within the SM, only color neutral objects can exist as free particles. This results in a phenomenon called *confinement*: quarks never occur isolated. If they are forced apart, as for example in an experimental particle collider, the energy of the color-field rises to values that allow for the creation of new bound quark-antiquark states and leads to *hadronization*. Gluons are themselves color-charged with color and anti-color and do thus self-interact, resulting in a short interaction range of the strong interaction.

In the following natural units $\hbar = c = 1$ are used for the sake of simplicity unless otherwise noted.

2.1.2. The Lagrange Formulation of Relativistic Field Theory

Mathematically speaking the Standard Model is formulated as a *gauge invariant relativistic quantum field theory* (QFT) [4]. In the SM, particles and interactions are described by fields which act as operators on particle states in the quantum mechanical Hilbert space.

Table 2.2.: Fundamental forces in the Standard Model. Numbers are taken from [3].

fundamental force	mediating particle	mass	quantum number (J^P)	electric charge [e]
electromagnetic	photon (γ)	–	1^-	–
weak	Z^0	91.19 GeV	1	–
weak	$W^\pm$	80.40 GeV	1	± 1
strong	8 gluons (g)	–	1^-	–

Lagrange Formalism

In classical mechanics the most general statement about dynamics is given through the *principle of least action* [5], where the action S is usually written as the time integral over the Lagrangian L

$$S = \int L(q_i, \dot{q}_i, t)dt$$

with n generalized coordinates q_i and the Lagrangian for a closed system with kinetic energy T and potential energy U reads

$$L = T(\dot{q}_i, t) - U(q_i, \dot{q}_i, t).$$

Minimization of the action $\delta S = 0$ then leads to the *Euler-Lagrange* equations of motion.

Proceeding to relativistic field theory it is helpful to introduce the Lagrangian density $\mathcal{L}$ as a function of complex fields $\phi_i(\mathbf{x})$ and their derivatives $\partial_\mu \phi_i \equiv \frac{\partial \phi_i}{\partial x^\mu}$ and to rewrite the action as a spatial integral [6]

$$S = \int Ldt = \int \mathcal{L}(\phi_i, \partial_\mu \phi_i) d^4x.$$

In analogy to classical mechanics minimization of the action $\delta S = 0$ yields the Euler-Lagrange equations of motion for fields,

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\mathcal{L}}{\partial \phi_i} = 0.$$

Application of the Lagrangian formalism is particularly suited to relativistic quantum field theory, because it accommodates Lorentz invariance and causality.

Lagrange Densities of the Standard Model Particles

Based on their spin, the SM contains three classes of particles, each of them being described by a different type of field [4].

Spin 0 particles are described by *scalar fields* $\phi(\mathbf{x})$. The Lagrange density for a free scalar particle of mass m reads

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - m^2 \phi^\dagger \phi, \tag{2.1}$$

leading to the *Klein-Gordon equation* as equation of motion

$$(\Box + m^2)\phi = 0,$$

with $\Box \equiv \partial_\mu \partial^\mu$. A complex scalar field with $\phi^\dagger \neq \phi$ has two degrees of freedom which can be interpreted as particle and antiparticle.
Spin 1 bosons are described by *vector fields* $A_\mu(\mathbf{x})$. Assuming a particle mass m, the Lagrangian density for a free spin 1 particle is given by

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{m^2}{2} A_\mu A^\mu \tag{2.2}$$

with field strength

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.3}$$

The corresponding equation of motion is known as the *Proca equation*

$$\left[(\Box + m^2) g^{\mu\nu} - \partial^\mu \partial^\nu\right] A_\nu = 0,$$

where $g_{\mu\nu}$ denotes the metric tensor as $\mathrm{diag}(1, -1, -1, -1)$.
Spin 1/2 fermions, such as e. g. the electron, are described by four-component *Dirac spinor fields* $\psi(\mathbf{x})$. The Lagrange density for a free fermion with mass m is written as

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \tag{2.4}$$

where $\bar{\psi} = \psi^\dagger \gamma^0$ relates Dirac spinors with the Dirac γ-matrices defined in [6]. The first part of this expression is identified with the kinetic term of the Lagrangian and the second part with the mass term. The Lagrangian yields the *Dirac equation* as the equation of motion

$$(i\gamma^\mu \partial_\mu - m)\psi = 0.$$

2.1.3. Gauge Theories

One of the key ingredients to the formulation of the SM is the concept of *gauge theories*, which is also known in classical electrodynamics [7]: the magnetic field $\vec{B}$ and the electric field $\vec{E}$ can be defined in terms of vector potentials $\vec{A}$ and scalar potentials Φ,

$$\vec{B} = \vec{\nabla} \times \vec{A} \quad \text{and} \quad \vec{E} = -\vec{\nabla}\Phi - \frac{\partial \vec{A}}{\partial t}$$

The vector potential $\vec{A}$ in this equation is arbitrary in the way that some scalar, differentiable function $\Lambda(\mathbf{x})$ may be added. The magnetic field $\vec{B}$ and the electric field $\vec{E}$ remain unchanged under the simultaneous transformation

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \Lambda(\mathbf{x}) \quad \text{with} \quad A_\mu = \left(\Phi, \vec{A}\right). \tag{2.5}$$

This transformations is called a *gauge transformation* and the invariance of the fields under such a transformation is called *gauge invariance*. Application of the principle of gauge invariance to QFT yields the structure of the interactions between fermions and bosons, as well as boson self-interactions [4].

Abelian Gauge Theory - QED

In the case of the free fermion Lagrangian (2.4), the requirement that the global abelian $U(1)$ *unitary group* symmetry is preserved, that is

$$\psi(\mathbf{x}) \to \psi'(\mathbf{x}) = e^{i\alpha}\psi(\mathbf{x})$$

for arbitrary real numbers α, may be extended to the postulation of invariance under *local gauge transformations* where $\alpha \to \alpha(\mathbf{x})$ is an arbitrary real function of space-time,

$$\psi(\mathbf{x}) \to \psi'(\mathbf{x}) = e^{i\alpha(\mathbf{x})}\psi(\mathbf{x}).$$

However, the derivative term in the fermion Lagrangian (2.4) is not invariant under such a transformation:

$$\partial_\mu \psi'(\mathbf{x}) = \partial_\mu \left(e^{i\alpha(\mathbf{x})} \cdot \psi(\mathbf{x})\right) = \underbrace{e^{i\alpha(\mathbf{x})} i\psi(\mathbf{x})\partial_\mu \alpha(\mathbf{x})}_{\neq 0} + e^{i\alpha(\mathbf{x})}\partial_\mu \psi(\mathbf{x})$$

The principle of gauge invariance necessitates the introduction of a vector field A_μ and the simultaneous *minimal substitution* of the derivative ∂_μ with the *covariant derivative* D_μ:

$$\partial_\mu \to D_\mu = \partial_\mu - ieA_\mu, \tag{2.6}$$

where e is a free parameter which is identified with the coupling strength in the following. In addition, the vector field has to transform as

$$A_\mu(\mathbf{x}) \to A'_\mu(\mathbf{x}) = A_\mu(\mathbf{x}) + \frac{1}{e}\partial_\mu \alpha(\mathbf{x}). \tag{2.7}$$

The required transformations (2.6) and (2.7) are the quantum mechanical analogy to (2.5). As a consequence the *invariant Lagrange density* of a free fermion describes an additional interaction of the vector field with the electromagnetic current $j^\mu = e\bar{\psi}\gamma^\mu\psi$, defining e as the coupling strength of the interaction,

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - j^\mu A_\mu.$$

It is notable that the postulation of local gauge invariance results in a new vector field A_μ, which has to be massless, since a mass term $\frac{1}{2}m^2 A^\mu A_\mu$ would violate the gauge invariance. The introduced vector field A_μ is not yet a dynamical field, because the Lagrangian misses a kinetic term. To complete the Lagrangian density for quantum electrodynamics (QED) such a term is invoked from classical electrodynamics and the final Lagrangian reads

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - e\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu},$$

with $F^{\mu\nu}$ as defined in (2.3). The vector field A_μ thus becomes the massless spin 1 photon obeying Maxwell's equations.

In addition, *Noether's theorem* [8] states, that each symmetry transformation which leaves a physical system unchanged results in a conserved quantity. An example of this is the homogeneity of time leading to the conservation of energy [5]. Correspondingly in QED, the electric charge e is identified with the conserved *quantum number* which results from the invariance under local $U(1)$ transformations.

Non-Abelian Gauge Theories - Generalization

In the SM, local gauge symmetry is a fundamental principle [6] and its generalized form is used to derive the structure of the interactions [4]. In analogy to the case of QED, the non-interacting system is described by a *multiplet* $\Psi = (\psi_1, ..., \psi_n)^T$ of fermion fields with mass m and the free fermion Lagrangian (2.4) then reads

$$\mathcal{L}_0 = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi \quad \text{with} \quad \bar{\Psi} = (\bar{\psi}_1, ..., \bar{\psi}_n).$$

In general, $\mathcal{L}_0$ is invariant under global transformations of the form

$$\Psi(\mathbf{x}) \to U(\alpha^1, ..., \alpha^N)\Psi(\mathbf{x}),$$

where U stands for unitary matrices from n-dimensional representations of the non-Abelian *Lie group* G of rank N and $\alpha^1, ..., \alpha^N$ denote N real parameters. Apart from QED, experimentally observed cases in the SM are representations of the *special unitary groups* $SU(N)$ with $G = SU(2)$ (weak interaction) and $G = SU(3)$ (strong interaction), where the fermion fields $\psi_1, ..., \psi_n$ form the fundamental representations with $n = 2$ and $n = 3$ respectively. For $U \ni SU(N)$ the transformation matrices can be rewritten as

$$U = e^{i\kappa_a T_a},$$

with κ_a being arbitrary real coefficients and T_a denoting the $N^2 - 1$ linearly independent generators of the Lie group which form the Lie algebra

$$[T_a, T_b] = i f_{abc} T_c.$$

In this expression f_{abc} denotes the structure constants, which are characteristic real numbers for the group. Groups with non-vanishing f_{abc} are called non-Abelian groups.

As in QED, the extension of the global gauge symmetry to a local symmetry necessitates the introduction of vector gauge fields G_μ^a, with $a = 1, ..., N^2 - 1$ for each linearly independent generator T_a. The minimal substitution to a covariant derivative then reads

$$\partial_\mu \to D_\mu = \partial_\mu - igT_a G_\mu^a,$$

where g represents the coupling of the gauge field (similar to the electric charge e in QED). In this way a set of $N^2 - 1$ gauge fields $G_\mu^a(\mathbf{x})$ enter the Lagrangian and induce an interaction term

$$\mathcal{L}_0 \to \mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\text{int}} \quad \text{with} \quad \mathcal{L}_{\text{int}} = -g\bar{\Psi}\gamma^\mu T_a \Psi G_\mu^a,$$

containing interactions of the currents $j_a^\mu = g\bar{\Psi}\gamma^\mu T_a \Psi$ with the gauge fields.

The Lagrangian is completed in analogy to QED by taking the kinetic term $-\frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}$ of the gauge fields into account. The final $SU(N)$ Lagrangian density then reads

$$\mathcal{L}_0 = i\bar{\Psi}\gamma^\mu \partial_\mu \Psi - m\bar{\Psi}\Psi - g\bar{\Psi}\gamma^\mu T_a \Psi G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu},$$

where the field strength is defined as $G_{\mu\nu}^a \equiv \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g f_{abc} G_\mu^b G_\nu^c$.

The Electroweak Interaction - $SU(2) \times U(1)$

The weak interaction is responsible for the transitions of the fundamental fermions, e. g. quarks and leptons. Experimental observations suggest they appear in generations with *left-handed doublets* and *right-handed singlets* as fundamental representations of the group $SU(2) \times U(1)$ [9–11]. Consequently, the particles can be classified according to the quantum numbers of the $SU(2)$ symmetry, the *weak isospin* T and its third component T_3, as well as the *weak hypercharge* Y of the $U(1)$ symmetry. They are connected via the Gell-Mann-Nishijima relation to the electric charge Q:

$$Q = T_3 + \frac{Y}{2}.$$

For the following it is useful to define handedness through the *chirality* of the fermion fields,

$$\psi_L \equiv \frac{1 - \gamma_5}{2}\psi \quad \text{and} \quad \psi_R \equiv \frac{1 + \gamma_5}{2}\psi, \tag{2.8}$$

with $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$. For massless particles the chirality is equal to the normalized projection of the spin to the momentum vector, the *helicity*. Left-handed fields

Table 2.3.: Fermion multiplets in the electroweak sector. One distinguishes the left-handed doublets marked with L and the right-handed singlets denoted with R. The prime on the quarks indicates the difference between mass- and flavor eigenstates, as explained later on in Section 2.1.4.

fermions	generation 1	2	3
leptons	$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_{\mathrm{L}}$ $(e^-)_{\mathrm{R}}$	$\begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_{\mathrm{L}}$ $(\mu^-)_{\mathrm{R}}$	$\begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_{\mathrm{L}}$ $(\tau^-)_{\mathrm{R}}$
quarks	$\begin{pmatrix}u\\ d'\end{pmatrix}_{\mathrm{L}}$ $(u)_{\mathrm{R}}$ $(d')_{\mathrm{R}}$	$\begin{pmatrix}c\\ s'\end{pmatrix}_{\mathrm{L}}$ $(c)_{\mathrm{R}}$ $(s')_{\mathrm{R}}$	$\begin{pmatrix}t\\ b'\end{pmatrix}_{\mathrm{L}}$ $(t)_{\mathrm{R}}$ $(b')_{\mathrm{R}}$

possess a weak isospin of $T = 1/2$ and thus form doublets, while right-handed fields have $T = 0$ and appear in singlets, as shown in Table 2.3.

In fact, $SU(2) \times U(1)$ is interpreted as the group of gauge transformations of the unified electromagnetic and weak, or short *electroweak*, interactions which leave the free fermion Lagrangian (2.4) invariant. The group has four generators, thus, also four gauge bosons are predicted.

Requiring local gauge invariance leads to the introduction of three vector fields W_μ^a and one scalar field B^0. The simultaneous minimal substitution leads to two different covariant derivatives for left- and right-handed fields,

$$\partial_\mu \to D_\mu^{L,R} = \partial_\mu - ig_2 T_a^{L,R} \cdot W_\mu^a + ig_1 \frac{Y}{2} B_\mu^0 \quad \text{with} \quad T_a^L = \frac{1}{2}\sigma_a,\ T_a^R = 0, \tag{2.9}$$

with two independent gauge coupling constants g_2 for the $SU(2)$ and g_1 for $U(1)$, as well as T_a as the generators of $SU(2)$, with $\sigma_{1,2,3}$ being the *Pauli matrices* and Y as the generator of $U(1)$. The final Lagrangian of the fermion-gauge interactions and the pure gauge fields can then be written as

$$\mathcal{L} = \sum_j \bar{\psi}_L^j i\gamma^\mu D_\mu^L \psi_L^j + \bar{\psi}_{R,\pm}^j i\gamma^\mu D_\mu^R \psi_{R,\pm}^j - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu}^0 B_0^{\mu\nu},$$

where $\psi_{R,\pm}$ indicates the difference between up and down type fermions and j runs over the three lepton and quark generations. The physical fields $W^\pm, Z^0$ and A_μ of the gauge bosons observed in experiments are found to be linear combinations

of the fields W_μ^a and B^0, e. g. shown in Ref. [12]:

$$\begin{aligned} W^\pm &= \frac{1}{\sqrt{2}}(W^1 \mp iW^2) \\ Z^0 &= -B^0 \sin\theta_W + W^3 \cos\theta_W \\ A_\mu &= B^0 \cos\theta_W + W^3 \sin\theta_W, \end{aligned}$$

where θ_W is called the *Weinberg angle* or the *weak mixing angle* [3],

$$\sin^2\theta_W \approx 0.232.$$

By introducing weak isospin ladder operators $T^\pm \equiv \frac{1}{2}(T_1 + iT_2)$ the fermion-gauge interactions can also be expressed by the physical fields, each of them giving rise to a characteristic interaction term,

$$\begin{aligned} j^{\mathrm{em}}(A_\mu) &\sim e\bar{\psi}\gamma^\mu\psi A_\mu \\ j^{\mathrm{CC}}(W^\pm) &\sim g\bar{\psi}\gamma^\mu(1-\gamma^5)(T^+W^+ + T^-W^-)\psi \\ j^{\mathrm{NC}}(Z^0) &\sim \frac{g}{\cos\theta_W}\bar{\psi}\gamma^\mu(g_V - g_A\gamma^5)\psi Z^0. \end{aligned}$$

Photons couple to fermions through the *electromagnetic current* j^{em} with a strength proportional to the electric charge e. Because the term $(1-\gamma^5)/2$ is defined in (2.8) as the operator projecting on the left-handed part of a Dirac particle, $W^\pm$ bosons exclusively couple to left handed fermion fields in the *weak charged current* j^{CC}. The strength of fermion interactions with Z bosons through the *weak neutral current* j^{NC} depends on the weak isospin T_3 and the particle handedness. For a fermion with charge q_f (in units of e) the vector and axial-vector couplings are given as

$$g_V \equiv T_3 - 2q_f \sin^2\theta_W \quad \text{and} \quad g_A \equiv T_3.$$

In contrast to the W bosons, the neutral currents of the weak interaction which are mediated by the Z bosons possess a small right handed contribution for $q_f \neq 0$.

In summary, the description of the unified electroweak interaction in terms of the $SU(2) \times U(1)$ symmetry groups explains the fundamental electromagnetic and weak forces as exchanges of mediating vector bosons with fermions. However, this description cannot explain the masses of the W and Z bosons (see Table 2.2). They require an additional description, which will be explained in Section 2.1.4.

The Strong Interaction - $SU(3)$

Based on the same principles previously discussed, the strong interaction is formulated as a gauge theory called *quantum chromodynamics* (QCD) using the symmetry group $G = SU(3)$. In QCD, the quarks form the free fermions and occur in *three different color states*. For each quark type $q = u, d, ...$, the triplets of fermion

fields $\Psi = (q_r, q_g, q_b)^T$ form the fundamental representations of the group. $SU(3)$ has eight generators,

$$T_a = \frac{1}{2}\lambda_a, \quad a = 1, ..., 8,$$

which are expressed of the eight 3×3 *Gell-Mann matrices* λ_a.

Requiring gauge invariance, the eight generators lead to eight *gluon* gauge fields G^a_μ. The Lagrangian density then assumes the form

$$\mathcal{L}_{\text{QCD}} = \bar{\Psi}(i\gamma^\mu - m)\Psi - g_s\bar{\Psi}\gamma^\mu T^a \Psi G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a,$$

with the field strength $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu$ and mass m as a free parameter for each triplet Ψ. g_s denotes the strong coupling constant, which can be expressed in terms of the more commonly used finestructure constant of the strong interaction $\alpha_s = \frac{g_s^2}{4\pi}$.

The main difference between QED and QCD lies in the fact, that the gluon fields are themselves colored objects, whereas the photon itself is electrically neutral. This results in gluon self interactions, which lead to an energy dependence of the coupling $\alpha_s(Q^2)$. Only at high energies does α_s decrease to small numerical values, such that the quarks can be considered "free". Therefore this effect is called *asymptotic freedom*.

If, on the other hand, the two quarks are driven apart, the energy of the connecting fields rises dramatically and allows only for bound quark-antiquark states. This phenomenon is referred to as *confinement*. At some point, e. g. after particle collisions, the energy becomes high enough to lead to the creation of new quark-antiquark pairs which recombine with the original quarks. This effect is called *hadronization* and explains why single quarks cannot be directly observed in nature. Instead, the best one can do experimentally is to approximate them by clustering sprays of collimated particles arising from hadronization into *jets*. This is further described in Chapter 4.

2.1.4. Spontaneous Symmetry Breaking and the Higgs Mechanism

Until this point, particle masses of bosons and fermions are not yet accounted for in the mathematical formulation of the Standard Model. However, some of the bosons and fermions in the SM are obviously massive. The widely accepted way of introducing particle masses within the SM is the use of spontaneous symmetry breaking in the Higgs mechanism, which was originally proposed by Anderson, Brout, Englert, Guralnik, Hagen, Higgs and Kibble [13–16]. This is explained in the following for the gauge bosons. The related Yukawa couplings generating the fermion masses will be briefly mentioned afterwards.

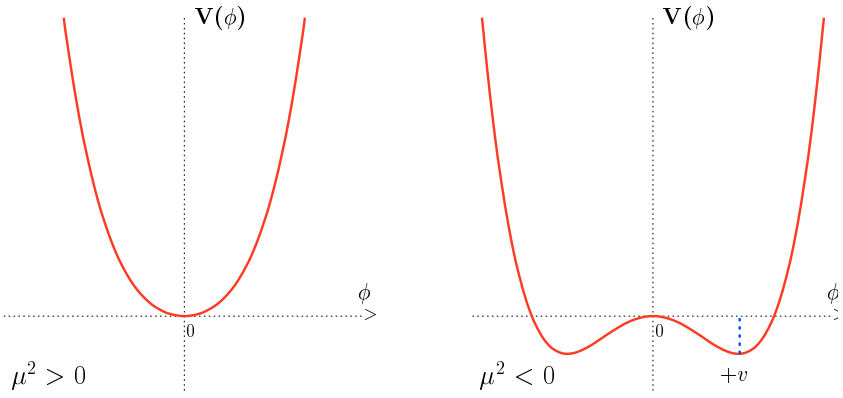


Figure 2.1.: The characteristic Higgs potential taken from [17]. For $\mu^2 > 0$ only one global minimum exists at $\phi = 0$. For $\mu^2 < 0$ multiple minima exist, one of which is chosen during the spontaneous symmetry breaking.

Global Symmetry Breaking

In order to explain the masses of the gauge bosons with the Higgs mechanism one has to introduce an additional complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$, the Higgs field, with a potential $V(\phi)$ [18]. Requiring invariance under global $U(1)$ gauge symmetry $\phi \to e^{i\alpha}\phi$, a simple Lagrangian leading to spontaneous symmetry breaking possesses a kinetic term T and a potential V including ϕ^4 self-interaction ($\lambda > 0$):

$$\mathcal{L} = T - V = \partial_\mu \phi^* \partial^\mu \phi - (\mu^2 \phi^* \phi + \lambda \cdot (\phi^* \phi)^2). \tag{2.10}$$

Figure 2.1 illustrates $V(\phi)$ in one dimension. For $\mu^2 > 0$ the potential $V(\phi)$ has a minimum at $\phi = 0$. For $\mu^2 < 0$, however, one finds a circle of minima in the (ϕ_1, ϕ_2)-plane with radius ν

$$\phi_1^2 + \phi_2^2 = \nu^2 \quad \text{with} \quad \nu^2 = -\frac{\mu^2}{\lambda}$$

Spontaneous symmetry breaking means that it is completely legitimate to choose one particular classical solution, e. g. $\phi_1 = \nu$ and $\phi_2 = 0$, out of the possible minima. However, perturbative calculations should involve expansions around this classical minimum. It is thus convenient to rewrite ϕ in terms of additional small perturbative, real fields η, ξ,

$$\phi(x) = \frac{1}{\sqrt{2}} \left(\nu + \eta(x) + i\xi(x)\right), \tag{2.11}$$

followed by a Taylor expansion around the minimum in the Lagrangian (2.10), which yields

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \eta)^2 + \frac{1}{2}(\partial_\mu \xi)^2 + \mu^2 \eta^2 + \text{const.} + \text{higher terms in } \eta, \xi. \tag{2.12}$$

Assuming a mass term of the form $-\frac{1}{2}m_\eta^2\eta^2$, this Lagrangian can be interpreted in terms of a massive field η with a generated mass $m_\eta = \sqrt{-2\mu^2}$ and a massless scalar *Goldstone boson* ξ. Goldstone bosons generally emerge whenever a continuous symmetry of a physical system is spontaneously broken, according to the *Goldstone theorem* [19].

Local Symmetry Breaking

However, in nature these additional massless bosons have never been observed. Furthermore, W and Z are massive vector bosons, not massive scalars. These aspects are addressed by requiring local gauge symmetry. Reconsidering the example of $U(1)$, the Lagrangian (2.10) has to be invariant under $\phi \to e^{i\alpha(x)}\phi$. This necessitates the introduction of an additional gauge field A_μ and a minimal substitution,

$$D_\mu = \partial_\mu - ieA_\mu \quad \text{with} \quad A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\alpha.$$

Insertion and repetition of the ground state expansion (2.11) in terms of η, ξ for $\mu^2 < 0$ from the previous section yields the analogon to (2.12),

$$\begin{aligned}\mathcal{L}'' =& \frac{1}{2}(\partial_\mu\xi)^2 + \frac{1}{2}(\partial_\mu\eta)^2 - \lambda\nu^2\eta^2 + \frac{1}{2}e^2\nu^2 A_\mu A^\mu \\ & - e\nu A_\mu\partial^\mu\xi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \text{higher terms.}\end{aligned} \tag{2.13}$$

In addition to the kinematic term, this Lagrangian now describes a massive scalar field η with $m_\eta = \sqrt{2\lambda}\nu$, still possesses a massless scalar Goldstone particle ξ, but also introduces the sought for massive vector boson A_μ with $m_{A_\mu} = e\nu$. Nevertheless, the coupling of the vector boson A_μ with the massless scalar field ξ still indicates problems with this description.

The Higgs Mechanism

The central idea of the Higgs mechanism is to choose a particular local gauge transformation in such a way, that the problematic massless Goldstone boson ξ in (2.13) is absorbed into the longitudinal polarization of the massive vector field A_μ. This is achieved for example by using

$$\phi \to \frac{1}{\sqrt{2}}\left(\nu + h(x)\right)e^{i\theta(x)/\nu} \quad \text{and} \quad A_\mu \to A_\mu + \frac{1}{e\nu}\partial_\mu\theta$$

as a particular choice of gauge with real fields h and $\theta(x)$. This is in fact an alternative parameterization of (2.11), preserving the number of freedoms, but introducing $i\theta$ as the phase of $U(1)$. Due to the invariance of the Lagrangian under $U(1)$ transformations θ is thus absorbed in the following.

Insertion to the original Lagrangian (2.10) with the covariant derivative leads to

$$\begin{aligned}\mathcal{L}''' =& \frac{1}{2}(\partial_\mu h)^2 - \lambda\nu^2 h^2 + \frac{1}{2}e^2\nu^2 A_\mu^2 - \lambda\nu h^3 - \frac{1}{4}\lambda h^4 \\ &+ \frac{1}{2}e^2 A_\mu^2 h^2 + \nu e^2 A_\mu^2 h - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \end{aligned} \tag{2.14}$$

This exemplary Lagrangian for a model with a $U(1)$ gauge symmetry finally describes solely two fields: a massive vector boson A_μ and its interaction with the massive scalar *Higgs field* h.

Spontaneous Symmetry Breaking of the Local $SU(2) \times U(1)$ Gauge Theory

The Higgs mechanism itself is a general concept and can also be applied to the electroweak $SU(2) \times U(1)$ gauge theory, where it is needed to generate the mass for the W and Z bosons. From experiment it is known that the photon is massless. The spontaneous symmetry breaking should thus only affect the $SU(2)$ part of the theory, leaving the electromagnetic $U(1)$ subgroup unbroken. To achieve this, one introduces an isospin doublet of complex scalar fields with hypercharge $Y = 1$,

$$\Phi(x) = \begin{pmatrix}\phi^+(x)\\ \phi^0(x)\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1(x) + i\phi_2(x)\\ \phi_3(x) + i\phi_4(x)\end{pmatrix}.$$

Assuming local gauge invariance and accordingly using the minimal substitution towards a covariant derivative as prescribed in (2.9) for $T = \frac{1}{2}$ and $Y = 1$,

$$D_\mu = \partial_\mu - ig_2\frac{\sigma_a}{2}W_\mu^a + i\frac{g_1}{2}B_\mu,$$

the corresponding Higgs field Lagrangian is written as

$$\mathcal{L}_\mathrm{H} = T - V = (D_\mu\Phi)^\dagger(D^\mu\Phi) - \mu^2\Phi^\dagger\Phi - \lambda\cdot(\Phi^\dagger\Phi)^2. \tag{2.15}$$

Focusing on the case of $\mu^2 < 0$ as in the previous passage, one finds that the potential V in $\mathcal{L}_\mathrm{H}$ has a minimum for $|\Phi|$ with $\Phi^\dagger\Phi \equiv \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda}$. One particular choice $\phi_3^2 = -\frac{\mu^2}{\lambda} \equiv \nu$ leads to the *vacuum expectation value* (vev)

$$\langle\Phi_0\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \nu\end{pmatrix}.$$

Again, the field Φ has to be expanded around this ground state Φ_0. Applying the Higgs mechanism and choosing a particular gauge to absorb massless Goldstone bosons, the fluctuations from the vacuum can be described in terms of four real fields $\theta_1, \theta_2, \theta_3$ and h,

$$\Phi(x) = e^{iT_a\theta_a(x)/\nu}\begin{pmatrix}0\\ \frac{\nu + h(x)}{\sqrt{2}}\end{pmatrix}, \tag{2.16}$$

where T_a denote the generators of the $SU(2)$ group. Insertion of (2.16) into the Lagrangian (2.15) gives, in analogy to the previous paragraph, rise to a massive Higgs field with mass $m_H = \sqrt{2\lambda}\nu$. The parameter m_H is a free parameter in the SM and the search for the Higgs boson, corresponding to the experimental determination of its value, is one of the main objectives of this thesis.

In addition, the new Lagrangian contains a term of the Higgs-gauge field interactions, which gives rise to the mass terms of the vector bosons. It was already mentioned in the previous section, that the $SU(2) \times U(1)$ eigenstates, W^a_μ and B_μ, do not correspond to the mass eigenstates of the experimentally observed $W^\pm$ and Z^0. Diagonalization of the interaction term with respect to the mass eigenstates yields the masses of the vector bosons, $m_W = \frac{1}{2} g_2 \nu$ and $m_Z = \frac{1}{2}\sqrt{g_1^2 + g_2^2}\nu$. With the masses of the W and Z the aforementioned Weinberg angle can be written as $\cos\theta_W = \frac{m_W}{m_Z}$.

Fermion Masses - the Yukawa Sector

The Higgs field Φ is also used to describe the masses of leptons and quarks. They are, however, not generated in the same way as the $SU(2)$ gauge bosons, but introduced into the SM via additional *Yukawa interactions* between the fermion fields and the Higgs field.

For the leptons, the Yukawa term in the Lagrangian is given by

$$\mathcal{L}_{\mathrm{Y}}^{\mathrm{leptons}} = -\sum_{l=e,\mu,\tau} G_l \bar{L}_l \Phi l_R + h.c. \quad \text{with} \quad L_l = \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L$$

with L_l and l_R being the $SU(2)$ isospin doublets and singlets, as previously explained in Table 2.3 and G_l denoting the Yukawa couplings, which are free parameters in the SM. A particular characteristic of the form of the Higgs field (2.16) is the attribution of mass through the Yukawa couplings solely to the charged leptons. In this description neutrinos are assumed to be massless particles, despite experimental evidence of e. g. neutrino oscillations [20, 21], which can only be explained with massive neutrinos. The masses of the charged leptons are found to be proportional to the Yukawa coupling, $m_l = G_l \frac{\nu}{\sqrt{2}}$.

Quark masses are explained in a similar, albeit slightly more complicated way. In the quark sector, masses also have to be attributed to the right-handed particles. Furthermore, as Table 2.3 already indicates, the $SU(2) \times U(1)$ eigenstates of the down-type quarks are not identical to their mass eigenstates. The electroweak eigenstates d', s' and b' are found to be linear combinations of the mass eigenstates d, s and b. By convention, they are rotated with the unitary 3×3 *Cabibbo-Kobayashi-Maskawa* (CKM) matrix [22],

$$\begin{pmatrix} |d\rangle \\ |s\rangle \\ |b\rangle \end{pmatrix} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} |d'\rangle \\ |s'\rangle \\ |b'\rangle \end{pmatrix},$$

where the CKM matrix elements $|V_{ij}|$ correspond to the electroweak transition probabilities from one quark i into another j, currently given in [3] as

$$\begin{pmatrix} |V_{\mathrm{ud}}| & |V_{\mathrm{us}}| & |V_{\mathrm{ub}}| \\ |V_{\mathrm{cd}}| & |V_{\mathrm{cs}}| & |V_{\mathrm{cb}}| \\ |V_{\mathrm{td}}| & |V_{\mathrm{ts}}| & |V_{\mathrm{tb}}| \end{pmatrix} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.00347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{pmatrix}.$$

Using the electroweak eigenstate notations the Yukawa term for the quarks can be written in analogy to the leptons,

$$\mathcal{L}_{\mathrm{Y}}^{\mathrm{quarks}} = -\sum_{i,j=1}^{3} G_{ij}^{d} \bar{Q}_i \Phi d'_j - \sum_{i,j=1}^{3} G_{ij}^{u} \bar{Q}_i \Phi^c u_j + h.c.,$$

with Q_i denoting the three left-handed quark isospin doublets and u_j, d'_j representing the up- and down-type right-handed quark isospin singlets, respectively, and Φ^c standing for the charge-conjugated field. The Yukawa term leads to mass matrices $m_{ij}^{(u,d)} = G_{ij}^{(u,d)} \frac{\nu}{\sqrt{2}}$ of the quarks which are defined separately for up- and down-type quarks by the strength of the three-dimensional matrices G^d and G^u, which again are free parameters of the SM. The masses of the physical quarks m_q can be retrieved by diagonalization of these mass matrices. As simultaneous diagonalization of $m^{(u)}$ and $m^{(d)}$ is not possible it is a convention to rotate the down-type quarks into the weak eigenstate using the CKM matrix.

2.1.5. Particle Production and Feynman Rules

Having defined particle content and the Lagrangian densities within the Standard Model, Quantum Field Theory provides the mathematical tools to calculate quantitative predictions, e. g. the probability to find a particle interaction leading to a transition of a given initial state $|i\rangle$ into a final state $|f\rangle$.

The S-matrix formalism allows the computation of the transition amplitude $\mathcal{A}_{i\to f}$ from $|i\rangle$ to $|f\rangle$ via

$$\mathcal{A} = \langle f | \, S \, | i \rangle \, . \tag{2.17}$$

Following Fermi's golden rule [23] the transition probability is proportional to the square of its Matrix elements $\sim |\mathcal{M}_{fi}|^2$. In general, for a $2 \to N$ scattering process with initial states $|i\rangle$, $|j\rangle$ and final state $|f\rangle$, the total *cross section* can be written [6] as

$$d\sigma_{i,j\to f} = \frac{1}{2E_i E_j |\nu_i - \nu_j|} |\mathcal{M}_{fi}|^2 d\Pi_N, \tag{2.18}$$

with the energy of the incoming partons $E_{i,j}$, the relative beam velocity $|\nu_i - \nu_j|$ and the phase space factor

$$d\Pi_N = \left(\prod_{f=1}^{N} \frac{d^3 p_f}{(2\pi)^3} \frac{1}{2E_f} \right) (2\pi)^4 \delta^{(4)}(p_i + p_j - \sum_{f=1}^{N} p_f).$$

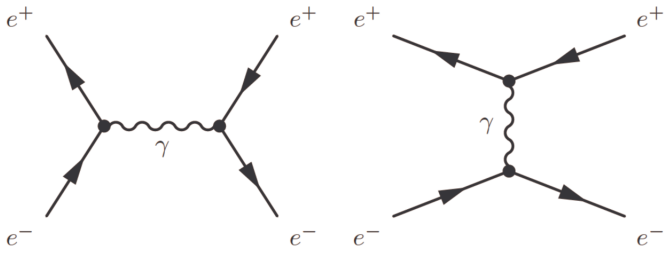


Figure 2.2.: Example of Feynman diagrams for leading order e^+e^- scattering. Each line or vertex corresponds to a characteristic term in the calculation of the process amplitude. All possible diagrams have to be summed for the final result. The number of vertices indicates the order of the calculation.

Calculation of the cross section thus results in the evaluation of the matrix elements $|\mathcal{M}_{fi}|$. The S-matrix in (2.17) corresponds to the time-evolution operator in quantum mechanics. The matrix elements can thus be calculated using a time dependant perturbation series, the so called Dyson series [24], which sums up all possible interaction paths leading from the initial to the final state.

Feynman graphs, such as for example shown in Figure 2.2 for e^+e^- scattering, are graphical representations of the different Dyson terms which have to be summed. The example shows the two leading order diagrams and can be interpreted in terms of the *Feynman rules*, which then prescribe the calculation. Each straight line corresponds to the propagation of an electron or positron in space-time (where time evolves from left to right and space vertically), introduced into the calculation as a characteristic propagator term. The interaction is mediated by a photon, which again introduces a propagator term. The intersection of lines is called a vertex and the related term goes proportionally to a multiplication with the coupling constant (assuming small couplings), as prescribed by the corresponding interaction term in the Lagrangian. Additional Feynman graphs with one more vertex contribute to the calculation in next-to-leading order (NLO) of perturbation theory and usually contain loops of the participating particles. The precision of theoretical predictions improves with higher orders. The complexity of the calculation, however, also increases dramatically. Most theory predictions used throughout this thesis thus correspond to NNLO and NLO calculations.

At the Large Hadron Collider (LHC) the colliding particles are protons, which themselves are composite objects consisting of three fundamental valence quarks, a sea of virtual quark-antiquark pairs surrounding them and gluons which bind them together. In order to make quantitative statements, e. g. about the rate of the Higgs boson production, it is necessary to take the proton's partonic substructure into account. Because of the running of the strong coupling constant $\alpha_s(Q^2)$ to high values which do not allow the use of perturbation theory, it is not possible to calculate the cross section of a pp collision into some multi-particle final state X directly with (2.18). Instead one uses the *factorization ansatz* and separates the

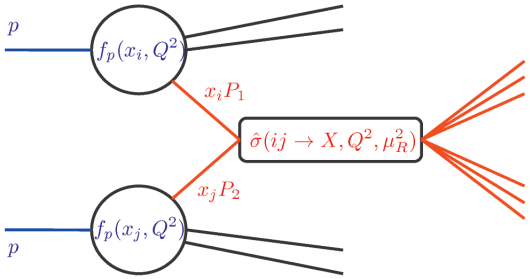


Figure 2.3.: Illustration of the parton model and the factorization ansatz. The interaction of the colliding partons is calculated as a hard process in the partonic cross section $\hat{\sigma}$ and folded with the initial parton PDFs f_p.

calculation into the short-range *hard scattering process* between two partons, which usually happens with a high momentum transfer Q^2 and can thus be treated perturbatively; and a long-range *low energy process* which describes the extraction of the scattering partons from the original protons. Figure 2.3 schematically describes this: two partons of the incoming protons p scatter and each of them carries a certain momentum fraction $x \equiv \frac{p^{\text{parton}}}{p^{\text{proton}}}$ which goes into the perturbative calculation of the hard process. The extraction probability for each scattering parton is determined experimentally by *parton distribution functions* (PDFs) $f_p(x, Q^2)$, which depend on the momentum fraction x and the total momentum transfer Q^2. In the factorization ansatz the total cross section is calculated by integrating the partonic cross sections $\hat{\sigma}$ of the hard process with the sum of all possible initial parton PDFs:

$$\sigma(pp \to X) = \sum_{i,j} \int dx_i f_p(x_i, \mu_F^2) \int dx_j f_p(x_j, \mu_F^2) d\hat{\sigma}(ij \to X, \mu_F^2, \mu_R^2)$$

The factorization ansatz depends on the energy scale μ_F^2 at which the factorization is introduced. This so called factorization scale is somewhat arbitrary and depends on the process in question. It is often set to the momentum transfer $\mu_F^2 = Q^2$ of the hard scattering process. μ_R denotes the renormalization scale which is needed to regularize the theory after the choice of the energy scale. The parton distribution functions f_p have been measured in several deep-inelastic scattering experiments and are combined by different collaborations. Noteworthy in the context of this thesis are the CTEQ6 [25] and the MSTW08 [26] PDFs. Figure 2.4 shows an example of two PDFs for different factorization scales. The dominance of the valence quarks u, d is clearly visible for high x, whereas for low x the gluon contribution dominates the PDFs.

Figure 2.5 summarizes the cross sections of the main physics processes at the center-of-mass energies of the Tevatron and the LHC (currently $\sqrt{s} = 7\,\text{TeV}$). The total cross section is many orders of magnitudes higher than the interesting physics processes, e. g. σ_{Higgs}. Noteworthy are also the production cross sections of the most important background processes to the Higgs boson search presented later, like σ_{b} and $\sigma_{\text{W,Z}}$.

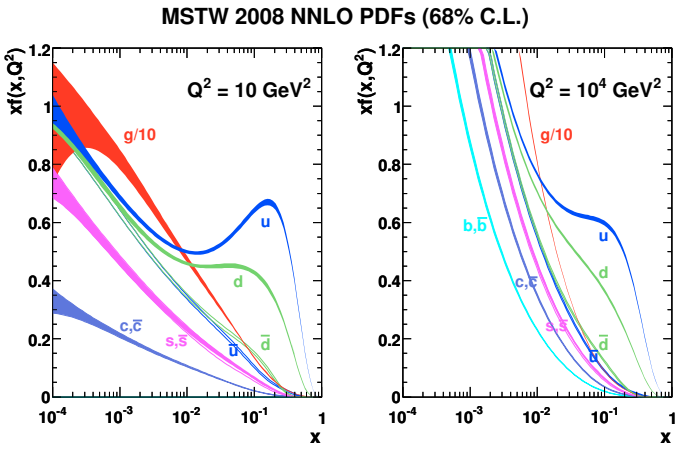


Figure 2.4.: Example of a parton distribution function at low and high Q^2 [27] including the associated one-sigma (68%) confidence level uncertainty bands. Note the dominance of u and d valence quarks at high values of x and the steep rise of the gluon contributions towards lower x.

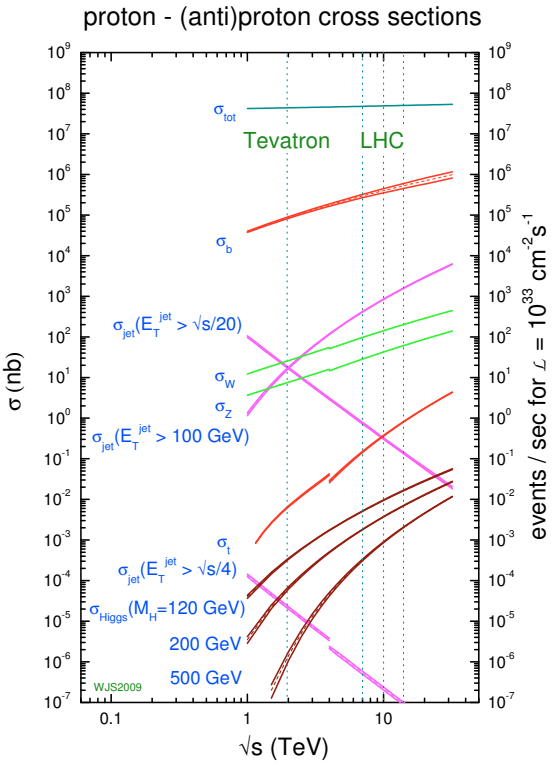


Figure 2.5.: Overview of Standard Model cross sections taken from [27]. The difference of many orders of magnitude between the production of well known processes and the production of interesting and possibly new physics, e. g. σ_{Higgs}, is one of the main challenges of the Higgs search presented in this thesis.

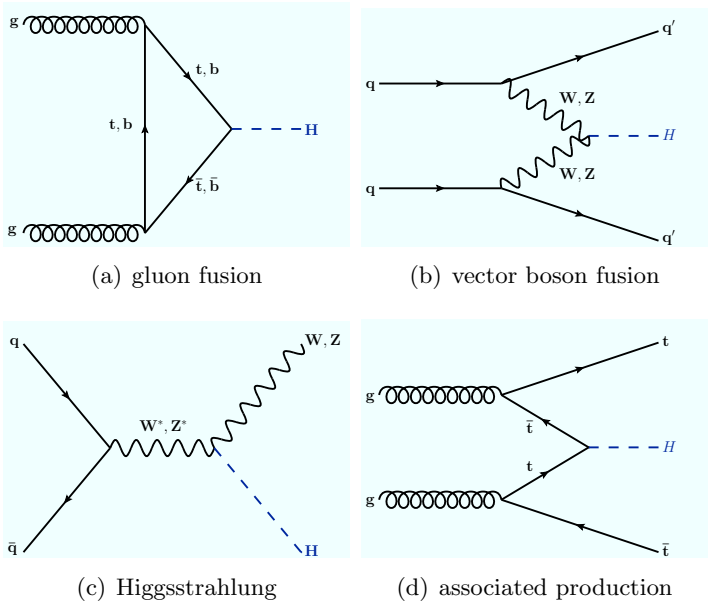
2.2. The Higgs Boson

The Higgs boson is the particle associated with the Higgs field h which generates the mass of the SM particles. It is the only particle which is predicted by the Standard Model but has not yet been observed in experiments.

This section starts with a description of the different Higgs boson production mechanisms at the LHC, followed by a summary of the possible decay modes. A particular focus is given to the decay of the Higgs boson into two b quarks. The section closes with an overview of current theoretical and experimental constraints on the search for the Higgs boson.

2.2.1. Higgs Boson Production at the LHC

At the LHC there are four main production mechanisms for the SM Higgs boson. Figure 2.6 summarizes them according to their importance in terms of cross section.



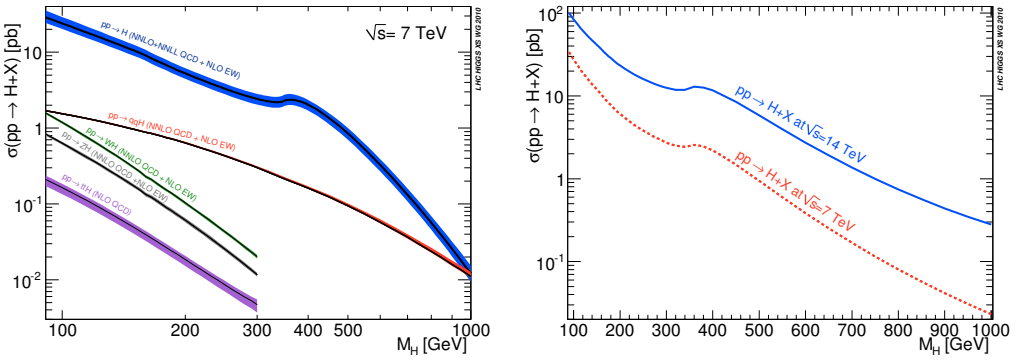
(a) gluon fusion (b) vector boson fusion

(c) Higgsstrahlung (d) associated production

Figure 2.6.: Higgs boson production mechanisms at the LHC.

The *gluon fusion* process, depicted in Figure 2.6(a), is the most dominant production process at the LHC. The gluons couple to a virtual quark triangle loop. Due to the proportionality of the Higgs coupling to quark masses and the high mass of the top quark this process is dominated by virtual top exchanges at the vertices.

The second largest contribution to the Higgs boson production cross section results from the *vector boson fusion* process in 2.6(b). The reduction in cross section with respect to gluon fusion is roughly a factor of ten, as can also be seen in



(a) (b)

Figure 2.7.: SM Higgs boson cross sections, taken from [29]. The left graph describes the cross sections of the different Higgs boson production mechanisms as a function of the Higgs boson mass. The right graph illustrates the rise in cross section when going from $\sqrt{s} = 7\,\text{TeV}$ to $\sqrt{s} = 14\,\text{TeV}$.

Figure 2.7(a). However, the distinct topology of this production mechanism makes it very attractive in some of the Higgs boson search channels [28].

The Higgs production via *Higgsstrahlung* as shown in 2.6(c) is even lower. Two quarks form a virtual vector boson which then radiates the Higgs boson. This leads to characteristic signatures with e. g. leptonically decaying vector bosons on the one hand side and the Higgs boson on the other. The Higgs boson search presented in this thesis relies on this production mechanism.

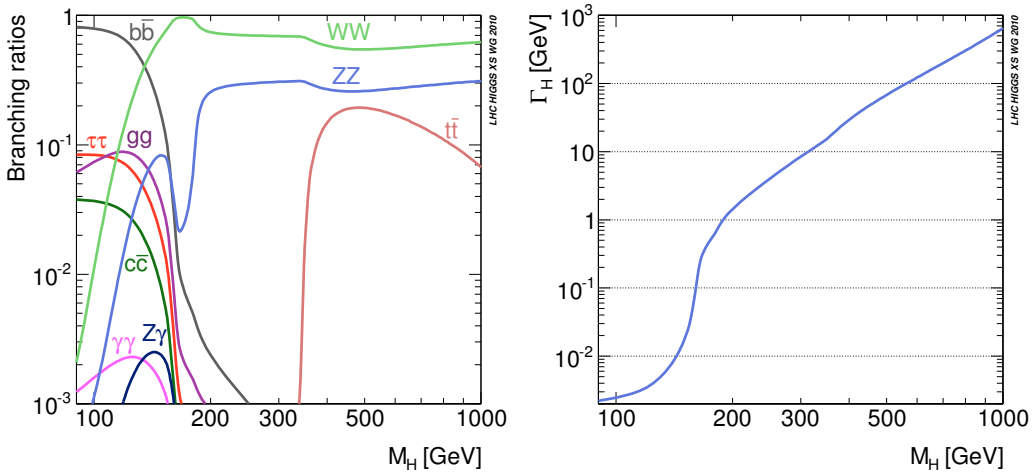
Figure 2.6(d) shows the *associated production* of a Higgs boson with the top quark. This process contributes the smallest fraction to the Higgs boson cross section, but may play an important role for the measurement of the coupling of the top quark to the Higgs boson.

Figure 2.7(a) summarizes the contributions of the four production mechanisms to the total LHC SM Higgs boson cross section for the current center-of-mass energy. The right graph, Figure 2.7(b), illustrates the difference predicted by theory for the Higgs boson cross section going from $\sqrt{s} = 7\,\text{TeV}$ to $\sqrt{s} = 14\,\text{TeV}$.

2.2.2. Higgs Boson Decay Modes

The branching ratios of the Higgs boson into different decay modes vary widely with the (yet unknown) Higgs boson mass, as shown in Figure 2.8(a). At low Higgs boson masses m_H, the dominant decay mode is $H \to b\bar{b}$, on which the Higgs boson search described in this thesis is based. Another, experimentally accessible, low mass modes is $H \to \tau\tau$ [30]. The decay mode $H \to \gamma\gamma$ contributes less than a factor 100 in branching ratio compared to $H \to b\bar{b}$, but is experimentally very clean and thus considered as a very promising discovery mode at low $m(H)$.

The Higgs boson decays in ZZ and WW also give important contributions to a



(a) (b)

Figure 2.8.: SM Higgs boson branching ratios (left) and decay widths (right), taken from [29]. The branching ratio for the $b\bar{b}$ decay mode of interest in this thesis is dominant for light Higgs boson masses to about 135 GeV. In this region, Higgs boson widths of only a few MeV are expected.

low Higgs mass search and their branching ratios dominate the Higgs decays above a Higgs boson mass threshold of two times the Z mass. The Higgs boson decay into two heavy top quarks is possible only at high values of $m_H > 2m_t$.

Figure 2.8(b) shows the decay width of the Higgs boson as a function of its mass. At low Higgs boson masses one expects a Higgs boson mass of only a few MeV, i. e. a very narrow resonance.

2.2.3. Constraints on Higgs Bosons

There are several constraints on the mass of the Higgs bosons which are motivated by theory. In addition, indirect theory constraints stem from the consideration of additional experimental precision measurements of SM parameters. Furthermore direct searches at LEP, the Tevatron and the LHC disfavor the existence of the SM Higgs boson in certain mass regions.

Theoretical Constraints

It is theoretically desirable to introduce the Higgs mechanism into the SM in a way such that the theory is well behaved and valid up to some high energy scales, e. g. the reduced Planck scale $M_P \sim \mathcal{O}(10^{18}\,\text{GeV})$ by which some new physics introducing gravity should appear [31].

Based on this assumption one theoretical argument introducing an upper boundary to the Higgs boson mass is often referred to as *triviality* [32]. Its origin lies

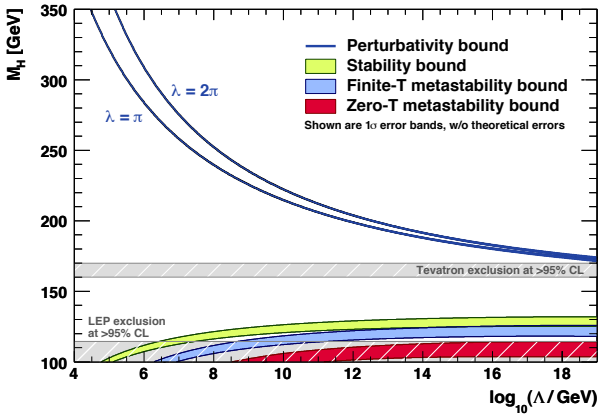


Figure 2.9.: Theoretical constrains on m_H, taken from [31], as a function of the scale Λ where the constraints set in. The two perturbativity boundary lines indicate the theoretical uncertainty on this upper mass constraint. The stability bound constraints the lower Higgs boson mass region. The preferred theoretical value for m_H lies between these boundaries implying the validity of the SM up to high scales Λ, possibly even the Planck scale M_P.

in the intrinsic running of the coupling $\lambda(\mu)$ of a scalar field, which increases as a function of the scale Λ and thus eventually becomes non-perturbative at the so called Landau pole Λ_T, unless one assumes the trivial solution $\lambda = 0$.

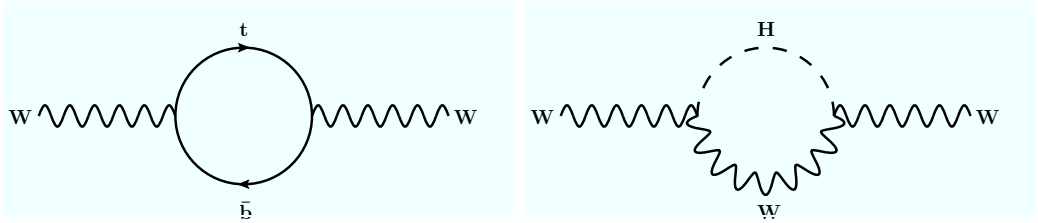
In the SM, the renormalization group equations (RGE) [6] are used to account for the scale dependence of the couplings and their application to the Higgs fields yields a well-behaved theory only for some Higgs boson masses depending on the scale Λ. For larger Higgs boson masses the RGE lead into the non-perturbative regime [31]. This relation between the upper boundary for m_H and the maximum allowed scale Λ can be calculated and is illustrated in the blue *perturbativity bound* lines in Figure 2.9. Thus, in the case of large Higgs boson masses, either a new non-perturbative theory has to enter at the scale Λ or new physics will be necessary below Λ to cancel the blow-up from the running of $\lambda(\mu)$ and preserve the validity of perturbation theory.

In addition it is possible to deduce an upper bound of the Higgs boson mass by studying the longitudinal WW and ZZ *scattering* [33]. Postulation of unitarity of the S matrix ($S^\dagger S = 1$) in (2.17) and decomposition into partial waves in the three scattering modes $W_L^+W_L^- \to W_L^+W_L^-$, $Z_LZ_L \to Z_LZ_L$ and $W_L^+W_L^- \to Z_LZ_L$ lead to [34]

$$m_H^2 \leq \frac{8\pi\sqrt{2}}{5G_F} \approx (780\,\mathrm{GeV})^2,$$

with G_F denoting the Fermi constant.

On the other hand, *vacuum stability* introduces a theoretical lower boundary [35] for the Higgs boson masses. For small m_H the evaluation of the RGE result in



(a) virtual top loop (b) virtual Higgs loop

Figure 2.10.: Examples of additional contributions of NLO processes to the W mass.

negative self couplings of the Higgs field for some scales Λ implying that the electroweak vacuum expectation value could be only a local minimum and that a deep new vacuum should exist at scales larger than Λ. In this case the universe would be unstable unless new physics at scales $< \Lambda$ would prevent the existence of such a new vacuum. This lower stability boundary is depicted in green in Figure 2.9.

In summary, the theoretical considerations presented here suggest a light SM Higgs boson within the perturbativity and stability boundaries in order to preserve the SM up to high scales, or the introduction of (yet undiscovered) new physics mitigating the effects leading to these boundaries.

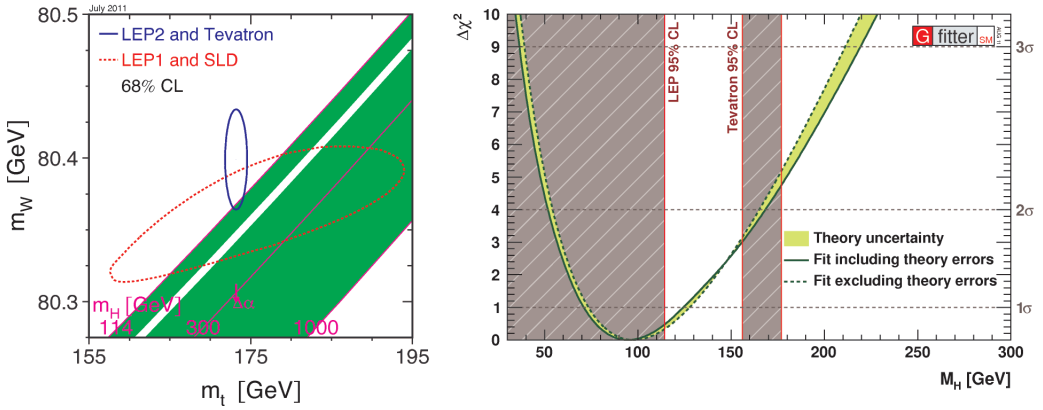
Indirect Constraints from Electroweak Precision Measurements

Additional constraints on the Higgs boson mass come from electroweak precision measurements. In the Standard Model the masses of W boson and top quark, m_W and m_t respectively, are fundamental parameters with a well defined relation to the expected Higgs boson mass: m_W is calculated [36] as

$$m_W = \sqrt{\frac{\pi\alpha_{\text{em}}}{\sqrt{2}G_F}} \cdot \frac{1}{\sin\theta_W\sqrt{1-\Delta r}} \quad \text{with} \quad \Delta r \sim f(m_t^2, \log m_H),$$

where α_{em} denotes the fine-structure constant, G_F stands for the Fermi constant and θ_W is the Weinberg angle. The term $1/\sqrt{1-\Delta r}$ introduces the addition of radiative corrections to the tree level calculation, which are illustrated in Figure 2.10. This correction is found to be proportional to the square of the top quark mass [37] and goes logarithmically with the Higgs boson mass, introducing a direct relation between m_W, m_t and m_H. Figure 2.11(a) shows 68 % confidence contours in the (m_W, m_t)-plane and expected correlations for different Higgs boson masses m_H. Figure 2.11(b) combines all electroweak measurements into one global fit and deduces the most probable value for m_H in terms of a χ^2 test.

In summary, the indirect electroweak precision measurements suggest a light Standard Model Higgs boson with $m_H = 80.359\,\text{GeV}$ [39]. This most probable value is however already excluded by the direct searches presented in the following.



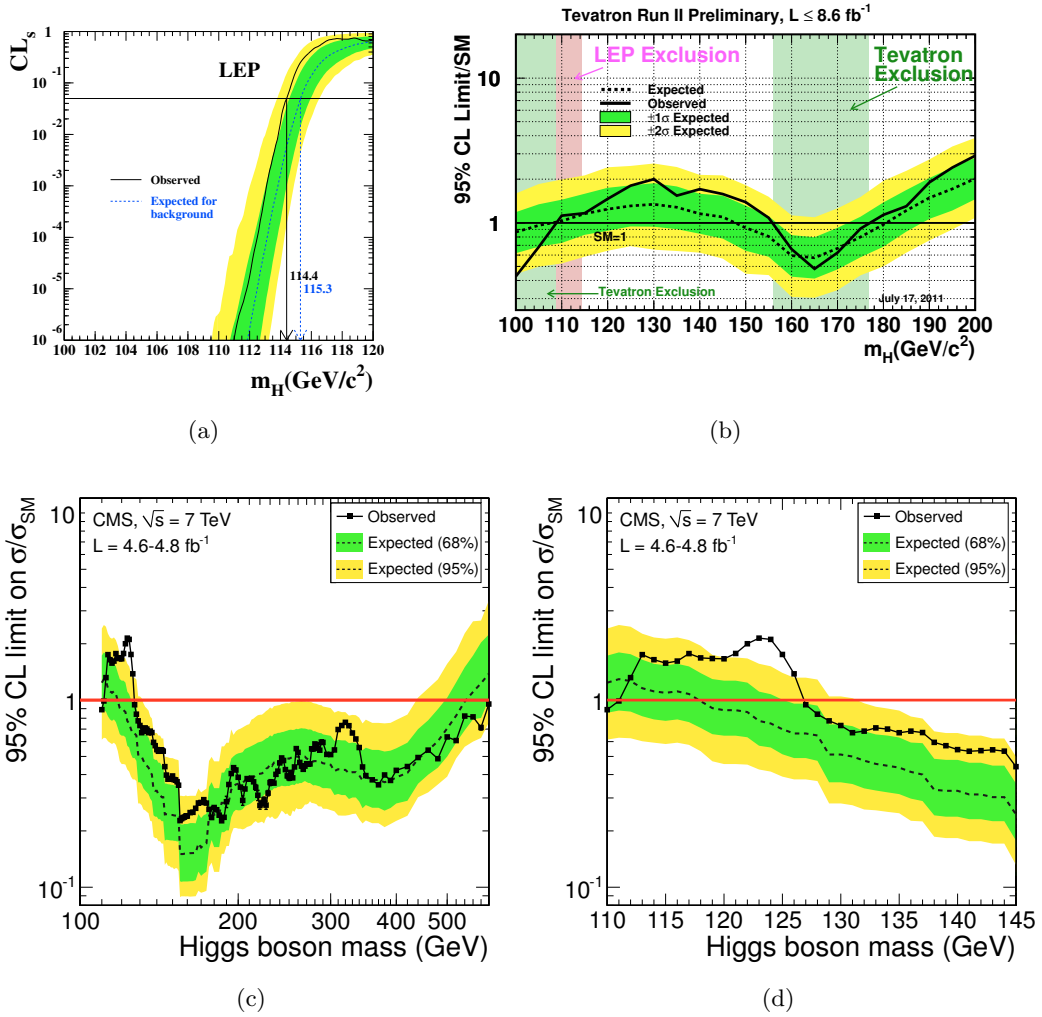
(a) (b)

Figure 2.11.: Indirect electroweak constraints on m_H based on precision measurements. The left Figure, taken from [38], shows the 68 % confidence bands of measurements to m_t and m_W and the correlation to the Higgs boson mass m_H. The right plot, taken from [39], illustrates the most probable value of m_H after a global χ^2 fit to the EWK measurements. Both Figures suggest a light SM Higgs boson at the threshold to the LEP exclusions.

Direct Constraints from Higgs Boson Searches

Direct searches for the Higgs boson have been performed in various channels at LEP [40], the Tevatron [41] and most recently at the LHC. Figure 2.12 summarizes the status quo of these searches. While LEP experiments have excluded the SM Higgs boson in the light mass region $m_H \leq 114.4\,\text{GeV}$ at a 95 % confidence level, first the Tevatron and now also the LHC experiments currently exclude a wide portion of the higher Higgs boson masses. The combined 95% C.L. exclusion region published by the CMS collaboration [42] covers the mass range from $127 - 600\,\text{GeV}$ and reports an excess of events with a local significance of 3.1σ for Higgs boson mass hypothesis of 124 GeV. The corresponding exclusion intervals from the ATLAS collaboration [43] cover Higgs boson masses of $113 - 115\,\text{GeV}$, $131 - 237\,\text{GeV}$ and $251 - 468\,\text{GeV}$.

It should be mentioned that partial results from the Higgs search studied in this thesis are already included in the combined CMS limits presented here. Furthermore, given all these indications suggesting a light Standard Model Higgs boson, the HW search channel described later turns out to be particularly promising for the next months to come in order to cast light on the small but tricky mass range left for the Higgs boson to hide.



(a) (b)

(c) (d)

Figure 2.12.: Direct constraints on $m(H)$ based on measurements at (a) LEP [40], (b) the Tevatron [41] and (c), (d) CMS [42]. All figures show the median expected limit with a dashed line and the corresponding $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands. The final interpretation of these graphs is based on the black solid lines depicting the observed limits: if they go below $\sigma/\sigma_{\mathrm{SM}} = 1$ or $\mathrm{CL_s} = 0.05$ the corresponding Higgs boson mass hypothesis is excluded at 95% confidence level. Figure (a) shows the $\mathrm{CL_s}$ value as a function of the Higgs boson mass. A value of $\mathrm{CL_s} = 0.05$ corresponds to a 95% C.L. exclusion on $\sigma/\sigma_{\mathrm{SM}}$, as shown in the other figures. The LEP exclusion ends at 114 GeV. Figure (b) shows how the Tevatron experiments extended this exclusion range. Figure (c) shows the most recent result based on a CMS combination and Figure (d) shows the same limits, but zooms into the light Higgs boson mass region. In summary, these direct searches suggest a light Standard Model Higgs boson above the LEP limit of 114 GeV and below the CMS limit of 127 GeV.

3. The Compact Muon Solenoid at the Large Hadron Collider

Research in elementary particle physics requires large scale experimental setups and is thus generally performed in international collaborations. The data used throughout this thesis has been collected with the Compact Muon Solenoid Experiment at the Large Hadron Collider, which is located on the CERN property in Geneva, Switzerland.

This chapter starts with an introduction of CERN and is followed by detailed descriptions of the Large Hadron Collider particle accelerator and the Compact Muon Solenoid particle detector experiment.

3.1. CERN

The European Organization for Nuclear Research - CERN (formerly known as "Conseil Européen pour la Recherche Nucléaire") is an international organization operating one of the world's largest laboratories for fundamental scientific research. It is located near Geneva on the Franco-Swiss border. CERN was founded in the aftermath of the Second World War in 1954, and its primary purpose was to reestablish european science and promote international cooperation in nuclear research [44]. Born originally from economic necessity of the post-war years, the concept of a joined international effort has proven itself to be very successful over time, enabling significant discoveries, such as e. g. the one of Z and W bosons in 1983 [45], but also technology spin-offs, such as the invention of the world wide web in 1990 [46].

Today, CERN is run by 20 member states and puts a particular focus on elementary particle physics, hosting specialized facilities such as the Large Hadron Collider (LHC) and bringing together scientists from all over the world in international collaborations like the Compact Muon Solenoid (CMS) collaboration, in order to study the constituents of our universe and their interactions.

3.2. The Large Hadron Collider

The Large Hadron Collider is the world's largest and most powerful particle collider, built to discover or exclude the Standard Model Higgs boson and new physics beyond the Standard Model. To achieve the necessary new energy regime for these searches with respect to previous collider experiments, namely the Large Electron-

Positron Collider (LEP) [47] at CERN and the Tevatron [48] at the Fermi National Laboratory, the LHC was built with state-of-the-art superconducting technology for design center-of-mass collision energies of up to $\sqrt{s} = 14\,\text{TeV}$ and aiming at peak luminosities of $L \sim 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. While the majority of the available operation time of the LHC is devoted to proton-proton collisions, which will be exclusively discussed in the following, it should be noted that there is also a rich heavy-ion collision program which has - already in the first year of heavy-ion operation - yielded very interesting results, e. g. the observation of "jet quenching" by ATLAS [49] and CMS [50].

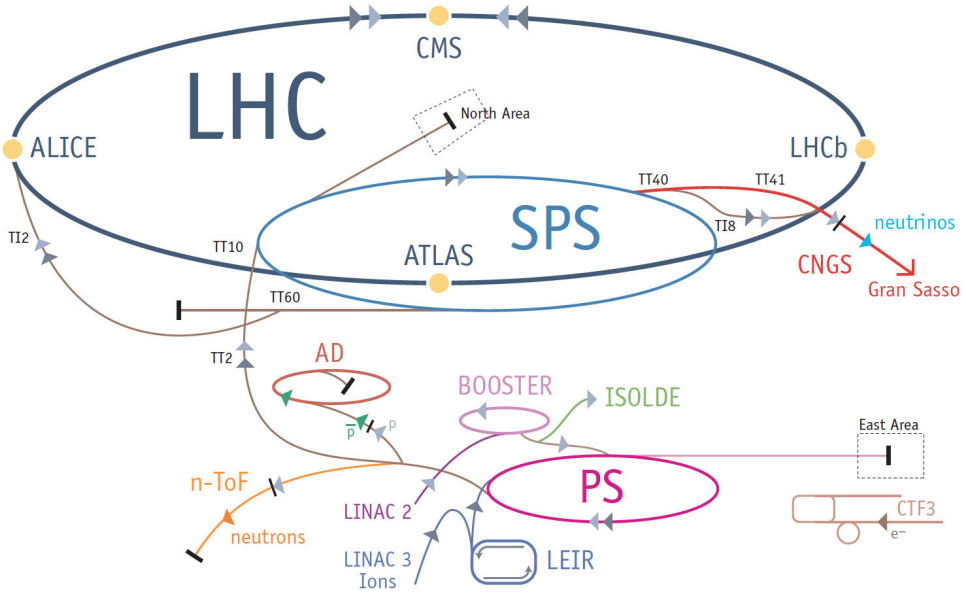


Figure 3.1.: The CERN accelerator complex, taken from [51]. Protons are pre-accelerated successively to higher energies in the Linac2, the Booster, the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS) before injection into the Large Hadron Collider (LHC), where they are finally accelerated up to 7 TeV.

3.2.1. Proton Production and Pre-Acceleration Chain

Precedent to an injection into the LHC main ring the protons have to be produced from hydrogen and then pre-accelerated successively to the LHC injection energy of 450 GeV. The full accelerator chain is schematically depicted in Figure 3.1. Most of the pre-acceleration structure has already existed before the construction of the LHC and is a testimony to more than half a century of accelerator-based particle research at CERN. Only small modifications were needed to accommodate it within the new accelerator chain. A full description of the different components is given in [52], a summary of the main aspects follows.

The *Proton Source* at the beginning of the accelerator chain is a simple bottle of

hydrogen gas. The gas is led into a duoplasmatron, which creates a current of free electrons that interacts with the gas and strips off the electrons from the hydrogen atoms. The hydrogen ions, or protons, are then led into the *Radio Frequency Quadrupole (RFQ)*. This very short linear accelerator with a length of only 1.75 m is used to focus the protons and accelerate them to an initial energy of 750 keV. The proton beam is picked up by the *Linear Accelerator (Linac2)*, another linear radio-frequency (RF) accelerator. The Linac2 uses electromagnetic cavities to accelerate the protons from 750 keV to 50 MeV over a length of 30 m, prior to injection to the *proton synchrotron booster (PSB)*. The booster is the first circular accelerator in the accelerator chain with a diameter of 50 m. In four vertically super-seeded rings the proton energy is increased to 1.4 GeV before injection into the *Proton Synchrotron (PS)*. The PS, originally built back in the 1950's, is a circular accelerator, which picks up the protons from the booster and accelerates them in a ring of 628.3 m circumference to an energy of 25 GeV. Next, the *Super Proton Synchrotron (SPS)* is the largest pre-accelerator with a circumference of 6912 m. The SPS consists of 744 bending dipole magnets and 216 focusing quadrupole magnets and raises the kinetic energy of the protons from 25 GeV to the LHC injection energy of 450 GeV in 21.6 s.

Table 3.1.: Overview of proton energies at different stages in the acceleration chain.

Accelerator	Type	Proton energy
RFQ	linear	750 keV
Linac2	linear	50 MeV
PS Booster	circular	1.4 GeV
PS	circular	25 GeV
SPS	circular	450 GeV
LHC	circular	$3.5 - 7$ TeV

3.2.2. LHC Main Ring

The Large Hadron Collider is a superconducting hadron accelerator and collider, situated roughly 100 m beneath surface in the pre-existing 26.7 km long tunnel of the former Large Electron Positron Collider (LEP). The switch from an electron and positron to a proton accelerator allows for much higher attainable center-of-mass collision energies, as protons suffer significantly lower energy losses due to synchrotron radiation on their circular orbit around the ring compared to electrons, due to their much higher mass.

Within the LHC, 16 superconducting high frequency cavities accelerate the protons to nominal beam energies of up to 7 TeV per beam. In order to countervail the centrifugal force experienced by the particles in the ring, 1323 superconducting dipole magnets are employed providing a bending magnetic field of up to 8.33 T.

They have to be operated at a temperature below 2 K (−271.15° C), requiring cooling with superfluid helium which is provided through a complex cryogenic infrastructure. In addition to the cavities and the dipoles, more than 7000 normal and superconducting magnets are used to focus the proton beams.

Even though the LHC has been designed for 7 TeV operation per beam corresponding to 8.33 T dipole magnetic fields, the electrical current inside the magnets is currently limited to a corresponding energy of 3.5 TeV per beam, or center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$, which essentially eliminates the risk of another severe accident as the one which occured in the beginning of the LHC operation in late 2008 [53]. As a conductor between two LHC magnets lost its superconductivity, the resulting increase in temperature led to the explosion of about 6 tons of superfluid helium. In consequence, 37 magnets were severily damaged and had to be replaced, many more were displaced. The two beam vacuum tubes were polluted with soot over a distance of 2.8 km. The costly repairs and the recommissioning took more than one year. While several temporary measures have been taken to assert the safety and reliability of the LHC at lower collision energies, a long shutdown is foreseen for the end of 2012 to replace several components before restarting with the higher design beam energies.

As the LHC is operated with particles of the same charge the use of one single beam pipe is not feasible. The counter-clockwise circulating proton beams are thus kept in separated beam pipes except for dedicated interaction points (IPs), where they run through each other. It is noteworthy, however, that the two seperated beam pipes share the same magnet structures due to constraints on space in the tunnel. This required an additional effort in the magnet design.

The protons in the LHC circulate around the ring in well-defined and separated bunches, which are a consequence of the radio frequency acceleration scheme in the pre-accelerators and the LHC. The LHC is designed to accommodate up to 2808 bunches, with each bunch containing roughly 10^{11} protons and with a bunch spacing of 25 ns. Current operation schemes use a slightly larger bunch spacing of 50 ns. The bunch size itself is not constant around the ring. While the proton bunches stretch up to a few centimeters long far from the interaction regions, they are squeezed to up to 16µm in the vicinity of the collision points to increase the luminosity.

Luminosity

The *instantaneous luminosity* L is, with the center-of-mass collision energy, one of the characteristic collider quantities of interest to data analysts, as it is directly proportional to the number of generated events per second for a given production cross section σ_{prod} [52]:

$$\dot{N} = \frac{\partial N_{\text{evt}}}{\partial t} = L \cdot \sigma_{\text{prod}}.$$

Table 3.2.: Overview of the LHC beam parameters as in the design report and currently achieved [52, 54, 55].

Parameter	unit	value 2010	value 2011	design
proton energy	TeV	3.5	3.5	7
protons per bunch	–	$1.2 \cdot 10^{11}$	$1.45 \cdot 10^{11}$	$1.5 \cdot 10^{11}$
number of bunches	–	368	1380	2808
bunch spacing	ns	150	50	25
normalised transverse emittance	µm rad	$2.4 - 4$	$1.9 - 2.4$	3.75
beta function β^* at IP	m	3.5	1.5	0.55
instantaneous luminosity	$\mathrm{cm}^{-2}\mathrm{s}^{-1}$	$2 \cdot 10^{32}$	$3.5 \cdot 10^{33}$	$1 \cdot 10^{34}$

It depends solely on the beam parameters and, assuming a Gaussian distribution of the beam particles, can be written as

$$L = \frac{N_1 N_2 f_{\mathrm{rev}} \gamma}{4\pi \epsilon_{\mathrm{norm}} \beta^*} F,$$

where N_1 and N_2 denote the number of particles in each beam, f_{rev} corresponds to the beam revolution frequency and γ is the relativistic gamma factor. The normalized transverse beam emitance ϵ_{norm} is a measure of the area occupied by beam particles, the beta function β^* describes the envelope of beam oscillations at the collision point. Together they are a measure of the beam sizes at the IPs and are subject to fine tuning by the LHC operators prior to stable beam collisions. The geometric luminosity reduction factor,

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2},$$

is a result of the beam crossing angle θ_c at the interaction point. σ_z is the standard deviation of the bunch length and σ^* the transverse standard deviation of the beam size. The formula assumes round beams with identical beam parameters, as well as $\sigma_z \ll \beta$.

The search for rare processes, such as the production of Higgs bosons, in LHC collisions therefore requires not only high center-of-mass energies, but also high beam intensities resulting in high instantaneous luminosities. The instantaneous luminosity is not constant over time, but depends on the beam parameters at the start of a LHC fill, which is the period between beam injection and beam dump. It decreases during the fill due to degrading emittance, beam-beam interactions and other effects. Information of the current luminosity profile is saved by the CMS experiment for so called luminosity sections of 23 seconds, which correspond to roughly 2^{18} beam orbits.

The integration over time of the instantaneous luminosity L leads to the total number of generated events

$$N_{\text{evt}} = \int L \cdot \sigma_{\text{prod}} \mathrm{d}t \equiv \mathcal{L} \cdot \sigma_{\text{prod}}.$$

The *integrated luminosity* $\mathcal{L}$ defined here is generally used to quantify the amount of accumulated data, which has been delivered and, considering data taking efficiencies, recorded by the detector experiments.

Experiments at the LHC

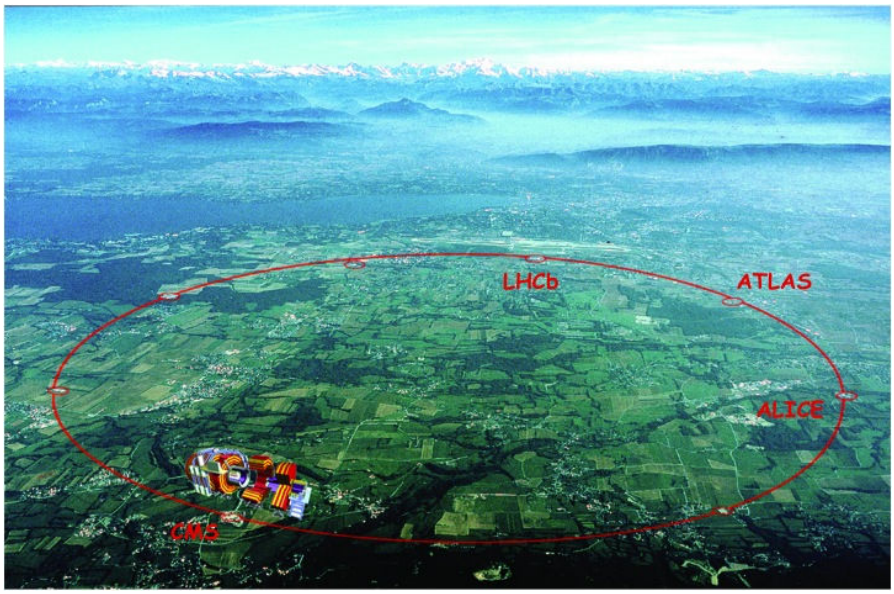


Figure 3.2.: LHC aerial view with locations of the four main detector experiments, adapted with modifications from [56]. CMS (in the front) is located in France, near the Jura mountains. One can see Geneva International Airport, downtown Geneva around the lake and the Mount-Blanc Alps in the back.

The Large Hadron Collider has four dedicated interaction points, one for each of the four main experiments hosted at the LHC, as shown in Figure 3.2. Two of the experiments, ATLAS (A Toroidal LHC ApparatuS) [57], near the CERN main site in Meyrin, and CMS (Compact Muon Solenoid), on the other side in France, were designed as general purpose experiments with a broad range of different physics measurements in mind. ALICE (A Large Ion Collider Experiment) [58] has a dedicated design to study heavy ion collisions in particular. The asymmetric LHCb (Large Hadron Collider beauty) experiment [59] is specialized in b-physics, i. e. studying collision events which involve the decay of B hadrons.

Two additional and smaller experiments, namely TOTEM (TOTal Elastic and diffractive cross section Measurement) [60] and LHCf (Large Hadron Collider for-

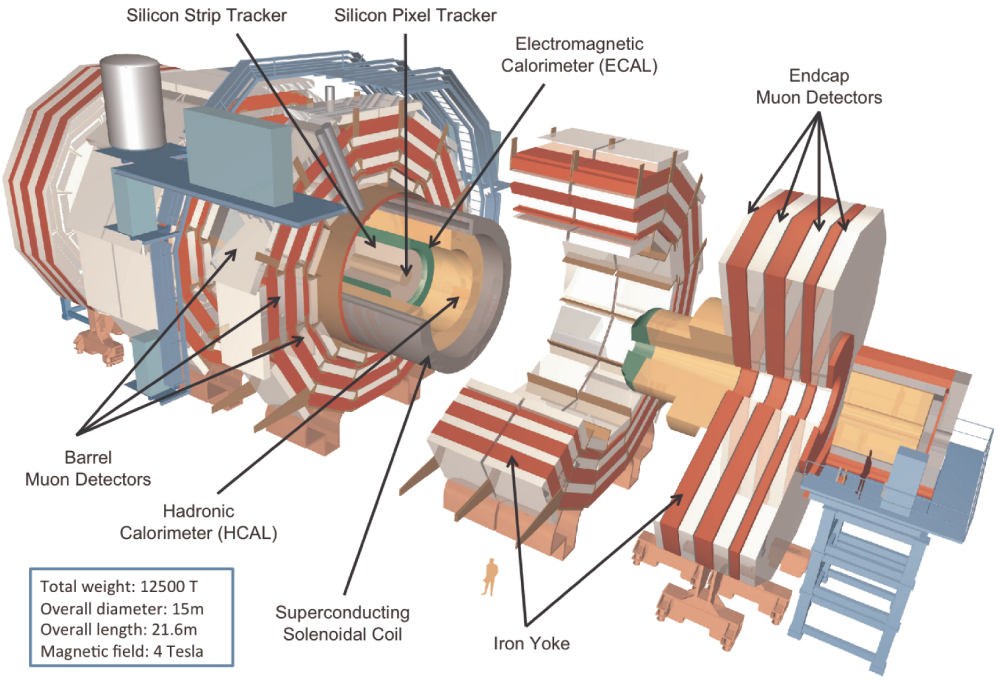


Figure 3.3.: Schematic overview of the Compact Muon Solenoid (CMS) detector, adapted from [62] with modifications. Note the symmetric and almost hermetic design around the beam pipe.

ward) [61] are highly specialized experiments to measure the total cross section and the decay of neutral pions respectively. They are located near the CMS and ATLAS experiments.

The data used throughout this theses has been recorded with the CMS experiment. A description of its main features follows in the next section.

3.3. The CMS Experiment

The Compact Muon Solenoid detector is designed to detect particles in pp and heavy ion collisions. As a general purpose detector the design is not aimed at one particular physics analysis or decay mode, but to enable a broad variety of different measurements, among them the search for the elusive Higgs boson in various final states.

As the name suggests, CMS is compact in design and built to provide especially good muon detection and resolution. It follows the tradition of symmetrical detectors with its different sub-detectors, loosely categorized here into tracking, calorimetry and muon identification, being placed around the beam pipe. Hermetic endcaps around the barrel region provide an overall detection coverage of nearly 4π.

Distinct features of the CMS detector are a superconducting solenoid providing a magnetic field of up to 4 T in the tracking and calorimeter systems, as well as its modular design, which allows for re-opening and accessing of different detector slices for repairs during technical stops of the LHC. A schematic overview of the different sub-detectors is given in Figure 3.3. A fully detailed description of all detector aspects is given in the technical design report [63]. They are summarized in the following.

3.3.1. Coordinate System

The CMS coordinate system is sketched in Figure 3.4. It is a right handed system with its origin at the center of the CMS detector. The x coordinate points to the center of the LHC ring and the y coordinate points upwards. The plane perpendicular to the beam directions is called the *transverse plane*. In proton proton collisions not all of the kinetic energy of the protons is absorbed into the collision products, as the colliding partons inside the protons carry only fractions of the total proton energies. As a result, the parton center-of-mass systems are boosted in a longitudinal directions along the beam pipe by unknown amounts. Energies and momenta defined in the transverse plane, however, are conserved in the collisions.

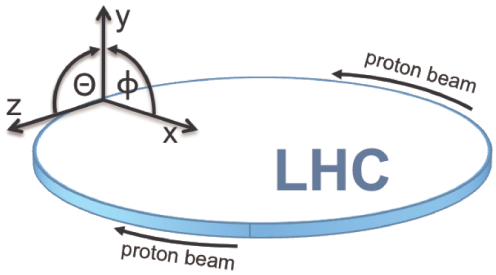


Figure 3.4.: CMS coordinate system

Since the proton beam is not polarized, the particle decays are expected to be invariant under rotations. Because of this, cylindrical coordinates (r, ϕ) are generally used in the transverse plane and a polar angle θ (measured from the positive z axis) is used to describe the longitudinal direction. For practical purposes the polar angle θ is often substituted by the *rapidity*

$$Y \equiv \frac{1}{2} \ln \left[\frac{E + p\cos(\theta)}{E - p\cos(\theta)} \right], \tag{3.1}$$

where $(E, \vec{p})$ is the energy-momentum four vector of the particle. The rapidity Y is invariant under boosts along the z-axis. The *pseudorapidity*,

$$\eta \equiv -\ln\left(\tan(\theta/2)\right),$$

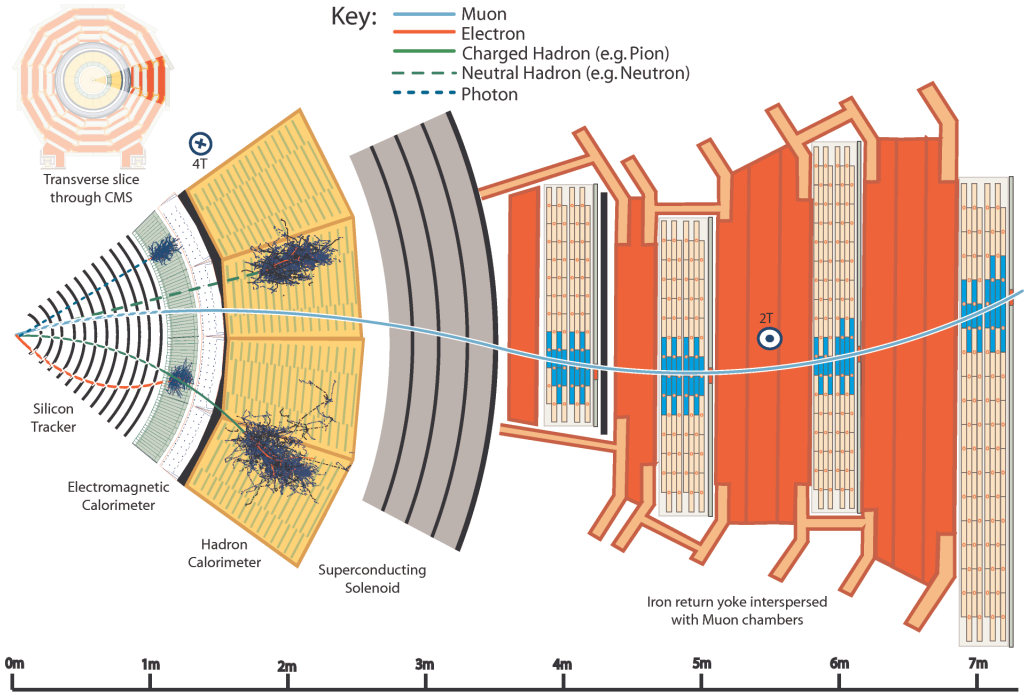


Figure 3.5.: Detection of particles in CMS [62], sketched in a transverse slice of the detector. Different particles are measured and identified by their characteristic behaviour under the influence of a magnetic field in the different sub-detectors.

is an approximation of Y valid for high-energy products of particle decays and exact for massless particles.

Spatial separation in the detector between any two particles is quantified in terms of

$$\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2}.$$

3.3.2. Inner Tracking System

The inner tracking system lies in the heart of the CMS detector. It is the first subsystem a particle encounters traversing the detector (see Figure 3.5). The design of of the CMS tracker was driven by the expectation of about 1000 particles from more than 20 overlapping pp collisions passing the tracker volume for each crossing, i. e. every 25 ns, at LHC design operation. This led to requirements on robust, efficient and precise reconstruction of charged particle trajectories on the physics side, as well as radiation hardness and a minimal material budget on the hardware side. The CMS tracker is composed of a pixel detector and a silicon microstrip tracker.

Both detectors make use of semiconducting materials and are based on p-on-n type silicon sensors: Charged particles traversing the semiconducting material lead to electron-hole pairs through ionization. The electrons and holes drift to the n and p crystals, respectively, creating measurable currents. The positions of the pixel detector and the microstrip tracker inside CMS are shown in Figure 3.6.

Pixel Detector

The pixel system is the innermost part of the CMS detector and closest to the interaction region. With a pixel cell size of $100 \times 150\,\mu\text{m}^2$ it contributes significantly to a small impact parameter resolution and very good track resolutions in both $r-\phi$ and z directions, with a single point spatial resolution in the range of $15 - 20\,\mu\text{m}$. This is of special importance to this thesis, as it enables efficient secondary vertex reconstruction, which is needed e. g. for the identification of jets originating from bottom quarks (see also Section 4.3.7). The pixel system covers a pseudorapidity range of $-2.5 < \eta < 2.5$. It consists of three straight barrel layers with 53 cm of length and radii of 4.4, 7.3 and 10.2 cm from the beam pipe, as well as two turbine-like tilted endcap disks extending from ~ 6 to 15 cm in radius at $z = \pm 34.5$ and $z = \pm 46.5$ cm on each side. In total, the barrel pixel detector contains ~ 48 million pixels covering a total area of $0.78\,\text{m}^2$ and the endcap pixel disks contain ~ 18 million pixels covering $0.28\,\text{m}^2$. Due to the vicinity to the interaction region a radiation tolerant design has been chosen for the pixel system. At nominal LHC luminosity operation the lifetime is expected to be at least 2 years.

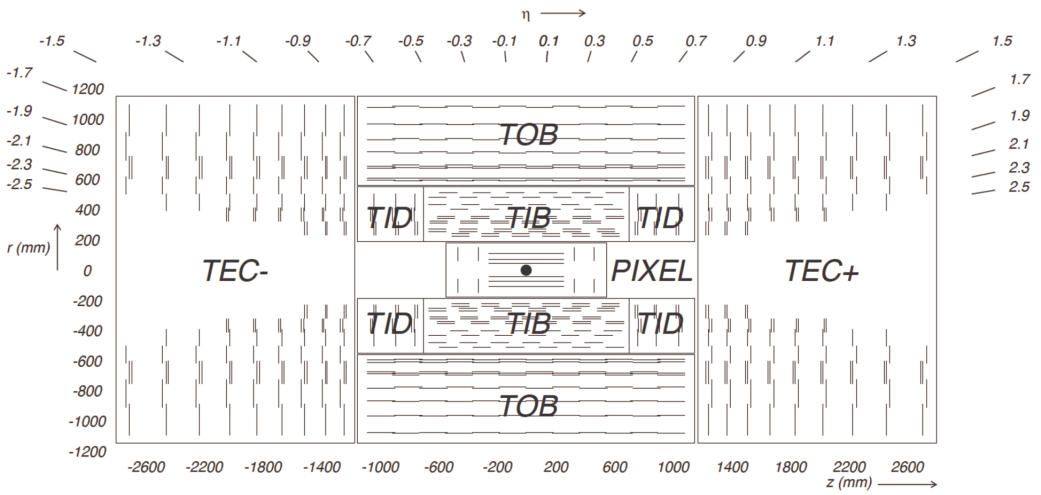


Figure 3.6.: Acceptance of the subsystems of the CMS tracker, taken from [63]. The tracker consists of pixel and silicon strip subsystems (acronyms are defined in the text). The overall tracking coverage lies within a pseudorapidity of $|\eta| < 2.5$.

Silicon Microstrip Tracker

At larger distances from the interaction region, where the particle flux per bunch crossing decreases, the use of the less expensive silicon microstrip technology is feasible with acceptable occupancies. By segmenting the p and n sides of the sensor into tilted strips and using separate read-outs for each, the position of the charged particles is determined with good resolution in all three dimensions. Multiple strip layers then allow for the reconstruction of the exact particle trajectories and, in conjunction with the magnetic field, the determination of the particle momenta. The silicon tracker is cylindrical in shape and placed around the pixel system in radii between 20 cm and 116 cm from the beam. Three different subsystems provide an overall tracking coverage within $|\eta| < 2.5$. The Tracker Inner Barrel (TIB) and Disks (TID) extend in radius to 55 cm and are composed of 4 barrel layers, supplemented by 3 disks at each end. They are surrounded by the Tracker Outer Barrel (TOB), which has a radius of 116 cm and consists of 6 barrel layers of micro-strip sensors, extending in z between ± 118 cm. Beyond this z range the tracker barrel detectors are enclosed by the Tracker EndCaps (TEC$\pm$), where the sign indicates the longitudinal position in the detector. They cover the region $124\,\text{cm} < |z| < 282\,\text{cm}$ and $22.5\,\text{cm} < |r| < 113.5\,\text{cm}$. Each of the endcap trackers is composed of 9 disks. Depending on the distance from the interaction region the single point spatial resolution of the CMS silicon strip tracker ranges from 23 to 53 µm in (r, ϕ) direction and between 230 and 530 µm in z direction. In total, it consists of about 9.3 million strips and is, with roughly 200 m^2 of active silicon area, the largest silicon tracker built to this date.

3.3.3. Calorimeter

The purpose of the calorimeter is to measure the energy of photons, electrons and hadrons by inducing secondary particle showers that produce light in scintillating materials. The CMS detector features an electromagnetic calorimeter (ECAL), with a pre-shower system in front of the end-caps, and a hadronic calorimeter (HCAL).

Electromagnetic Calorimeter

The electromagnetic calorimeter is built around the inner tracking system (see Figure 3.3) and is made of 61200 scintillating lead tungstate ($PbWO_4$) crystals in the central barrel part and enclosed by 7324 crystals in each of the two endcaps. Electromagnetically interacting particles, such as photons and electrons, are stopped in the lead tungstate crystals and produce cascades of secondary particles through bremsstrahlung and electron-positron pair production, thus emitting characteristic blue-green scintillation light. Avalanche photo-diodes (APDs) are used in the barrel region and vacuum photo-triodes (VPTs) in the endcaps to collect the scintillation light, which is a direct measure for the energy of the incoming particle. The advantage of $PbWO_4$ crystals with respect to other materials is the high material

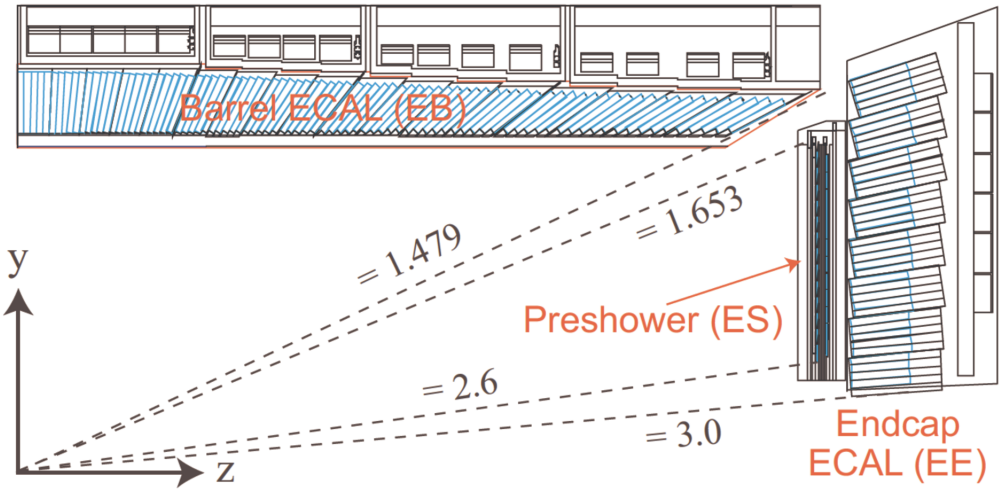


Figure 3.7.: Schematic overview of the acceptances of the ECAL systems, taken from [64].

density ($8.28\,\mathrm{g/cm^3}$), which leads to short radiation lengths ($X_0 = 0.89\,\mathrm{cm}$) and a small Molière radius (2.2 cm), which quantifies the transverse dimension of the shower and indicates a fine granularity. Lead tungstate crystals are radiation-hard and fast with a scintillation decay time on the order of the nominal LHC bunch crossing time: about 80% of the light is emitted within 25 ns. Figure 3.7 shows the acceptance of the electromagnetic calorimeter. The central barrel part of the ECAL (EB) covers $|\eta| < 1.479$, with a granularity that is 360-fold in ϕ and (2×85)-fold in η. This translates into front face cross sections of $22 \times 22\,\mathrm{mm^2}$. The crystal length of 230 mm corresponds to a radiation length of $\approx 25.8 \cdot X_0$, so that an overwhelming fraction of the electromagnetic energy should be well contained within the EB. The endcaps (EE) cover the pseudorapidity region from $1.479 < |\eta| < 3.0$ with front face cross sections of $28.62 \times 28.62\,\mathrm{mm^2}$ and crystal lengths of 220 mm, or $\approx 24.7 \cdot X_0$ radiation lengths. A sampling preshower calorimeter (ES) is installed in front of the ECAL endcaps in the region of $1.653 < |\eta| < 2.6$ to help identify photons from π^0 hadrons and to distinguish these from incoming single photons (as e. g. in $H \to \gamma\gamma$). It consists of two lead radiator layers which initiate electromagnetic showers from photons and electrons and silicon strip layers behind each radiator to measure the deposited energies and shower profiles.

The energy resolution of a calorimeter can be parametrized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{3.2}$$

where S stands for the stochastic term, which comprises event-to-event fluctuations. N denotes the noise term, which may be due to electronics or digitization noise and C stands for the constant term with its dominant contributions coming from

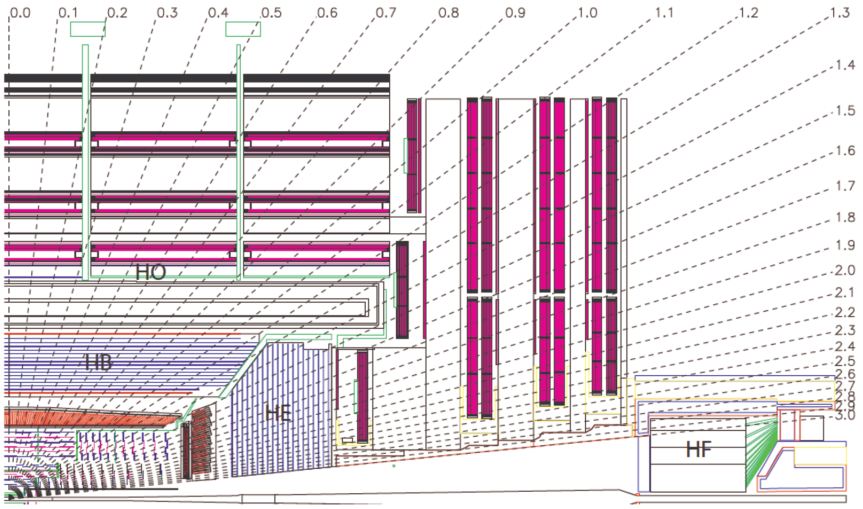


Figure 3.8.: Schematic overview over the acceptance of the HCAL systems, taken from [63].

the non-uniformity of the light collection, crystal inter-calibration errors or leakage of energy from the back of the crystal. E is the energy in GeV. Typical values for the ECAL energy resolution found in test beam data are $N = 0.12\%$, $S = 2.8\%$ and $C = 0.30\%$. This "NSC" formula, and in particular the constant term, will be of importance in Chapter 6.

Hadronic Calorimeter

The electromagnetic calorimeter is surrounded by a brass/scintillator sampling hadron calorimeter with a coverage of up to $|\eta| < 3.0$, as shown in Figure 3.8. Because the nuclear interaction length of the brass absorber $\lambda_I^{\mathrm{HCAL}}$ is much larger than the typical radiation length in the ECAL ($\lambda_I^{\mathrm{HCAL}} \approx 16.42\,\mathrm{cm} \approx 18.5 \cdot X_0^{\mathrm{ECAL}}$), hadrons typically deposit most of their energy in the hadronic calorimeter, which contains up to $\approx 11\lambda_I^{\mathrm{HCAL}}$. The incoming hadrons induce cascades of secondary particles by subsequent inelastic scattering with the material. Most of the shower energy is absorbed into the 50 mm thick copper absorber plates and only fractions are detected in the interleaved layers of 4 mm thick scintillating material. Wavelength shifting optical fibres are used to transport the scintillation light to hybrid photo-diodes (HPDs). The total hadron energy is estimated through these sample energy depositions, thus the name "sampling calorimeter". In comparison to the ECAL crystals the sampling method yields a more coarse energy resolution. The HCAL is segmented into a barrel system (HB) with an acceptance of $|\eta| < 1.3$ and endcap systems (HE) which cover the pseudorapidity range from $1.3 < |\eta| < 3.0$. It is complemented by an outer hadron calorimeter (HO), which is designed to catch hadron showers leaking out of HB. Due to the worse resolution this system is not used in this thesis, but its shielding power is essential to prevent high-energy

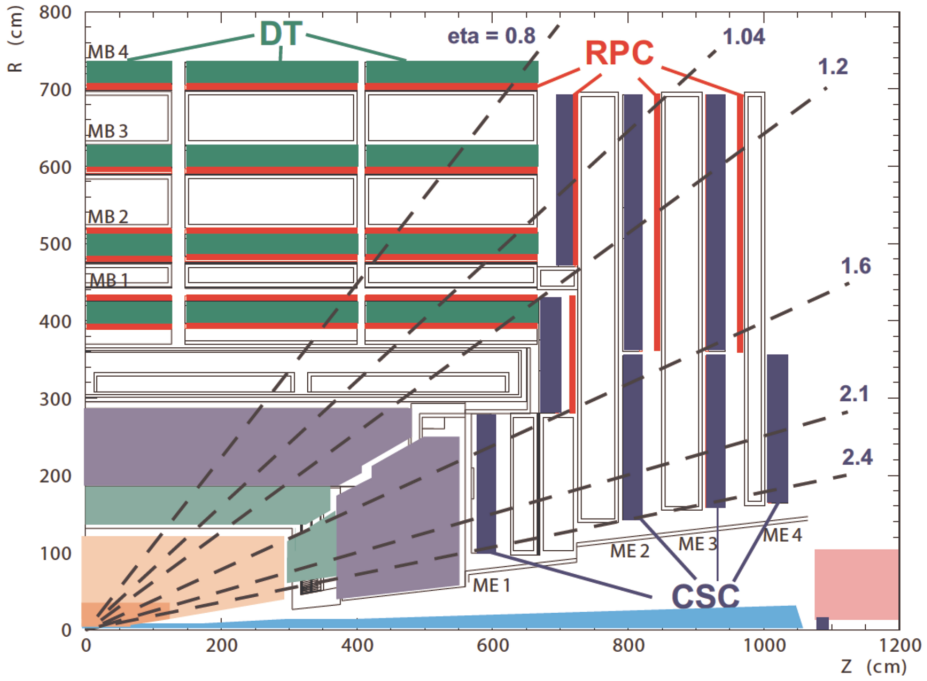


Figure 3.9.: Schematic overview over the acceptance of the muon systems, taken from [64].

jets and hadrons from entering into the muon stations. Beyond $|\eta| = 3$ forward hadron calorimeters (HF) have been installed on each side at $|z| = 11.2\,\mathrm{m}$ from the interaction point to extend the energy flow measurement of the HCAL to up to $|\eta| < 5$ and provide a hermetic coverage. The HF is located in a region of very high particle fluxes. As a result it has been designed with steel absorber structures and quartz fibers as active scintillating material. Photomultipliers are used to detect the Cherenkov light generated by particle showers in the quartz fibers.

3.3.4. Muon Spectrometer

The importance of the muon system to the experiment is already implied in its middle name. Most of the pp collisions at the LHC are of little interest, but the presence of leptons often indicates rare physics events. Muons are minimally ionizing particles meaning they only deposit small amounts of ionization energy in materials. They are thus easily detectable, because they are the only charged particles which produce hits in the tracker and are likely to traverse the calorimeter, the solenoidal coil and to reach the muon stations. The muon system has three functions: muon identification, momentum measurement and event triggering. It consists of three different types of gaseous detectors, which are interleaved with the iron return yoke of the solenoid (see Figure 3.3). In the barrel region, drift tube chambers (DT) cover the pseudorapidity range to $|\eta| < 1.2$, as depicted in

Figure 3.9. These chambers contain an Ar/CO_2 gas mixture, which is ionized by traversing charged particles. The ions are collected and read out by a system of wires with applied voltages and the particles' trajectories are calculated from their drift times to the wires. Cathode strip chambers (CSC) are used in the two endcap regions of CMS and extend the muon system coverage to $|\eta| < 2.4$. They are multi-proportional $Ar/CO_2/CF_4$ gas chambers, which apply narrower spatial wire separation and higher voltage than the DTs and thus do not rely on the drift times for the momentum measurements. The CSCs can operate at high rates and in the large and non-uniform field environment of the endcaps. Complimentary to the DT and CSC muon stations, resistive plate chambers (RPCs) have been mounted in both the barrel and endcap regions to provide an independent, fast and fine-grained trigger with a sharp p_T threshold. The RPCs are $C_2H_2F_4/iC_4H_{10}/SF_6$ gaseous parallel-plate detectors, which are able to tag an ionising event in a much shorter time than the 25 ns between two nominal LHC bunch crossings.

3.3.5. Trigger and Data Acquisition System

One of the most important challenges of the CMS experiment is posed by the vast amount of events produced via well established processes during LHC operations. Since it is technically impossible to permanently store the up to one billion pp interactions produced per second at the LHC, this background has to be rejected in a fast and efficient way, while making sure to keep as many of the interesting physics processes as possible. The nominal bunch crossing time of 25 ns corresponds to a crossing frequency of ≈ 40 MHz, where depending on the luminosity multiple pp collisions per crossing may occur.

Contrary to previous hadron collider experiments, CMS uses not three but only two trigger levels to achieve the large, necessary data rate reduction of $\approx 10^6$. The ultra fast hardware based Level 1 Trigger (L1) produces a maximum output of 100 kHz and the software based High Level Trigger (HLT) with its state of the art ≈ 8000 CPU-core filter farm reduces the data stream to a maximum of a few 100 Hz. This is explained in the following.

Level 1 Trigger

The Level 1 Trigger (L1) is based on custom-designed, programmable hardware. Figure 3.10 illustrates the L1 trigger decision process. Coarsely segmented data from the calorimeters and muon systems are logically evaluated in steps of increasing regional width to extract the L1 trigger decision, while buffering the full detector data in pipelined memory in the detector front-end electronics.

High Level Trigger and Data Acquisition

All events which pass the L1 trigger are sent from the detector front-ends to the HLT filter farm in the surface building at Point 5. The network is capable of

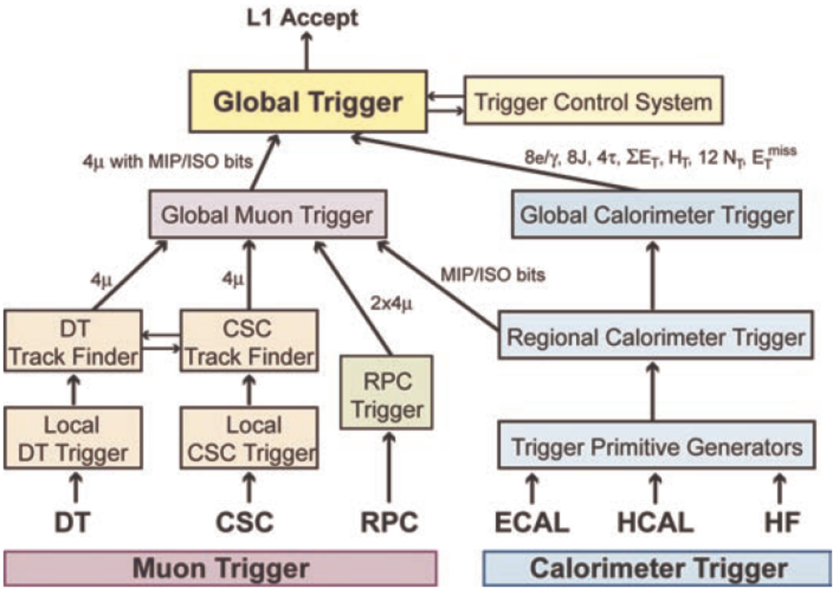


Figure 3.10.: Overview of the Level 1 Trigger architecture, taken from [63].

sustaining $\approx 100\,\mathrm{GByte/s}$ coming from about 650 data sources, as depicted in Figure 3.11. The builder network, which is sliced into 8 autonomous systems capable of handling 12.5 kHz of input each, assembles full events from the event fragments of the different data sources and sends them into the filter system. The filter farm runs a light-weight and slightly faster version of the offline CMS reconstruction software. This HLT reconstruction software is as sophisticated as the full offline reconstruction, but it is applied in stages such that an event can be exposed to only a fraction of the full event reconstruction computing time if it is discarded early in the trigger selection. Events which pass the HLT conditions are automatically send to the CERN Tier 0, where they are stored and further processed.

The concept of a fully software driven high level trigger has already proven itself to be very successful. An important advantage of using software is that it can be adapted quickly (even during operation) to cope with the rapid evolution of the LHC beam parameters. In terms of hardware, the overall filtering power of the HLT can be flexibly increased to rising needs either by extension of the filter farm (more off-the-shelf PC nodes) or by replacement with better nodes.

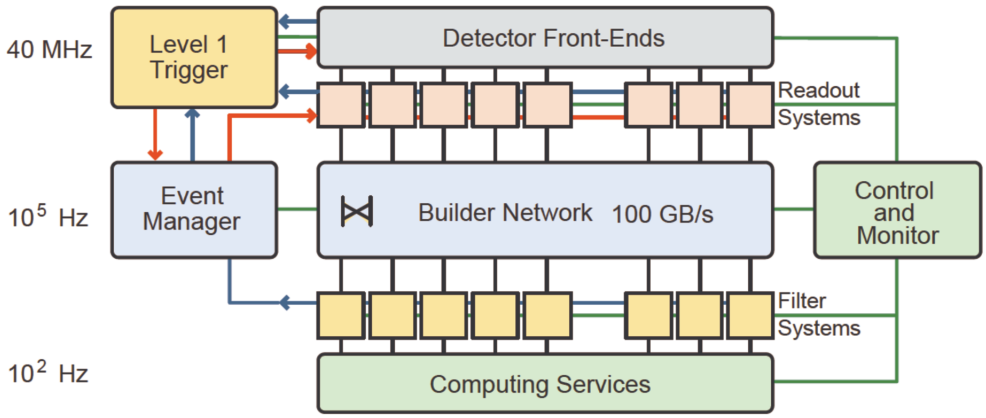


Figure 3.11.: Schematic overview of the data acquisition framework and the high level trigger, taken from [63].

4. Event Generation, Simulation and Reconstruction

Data analysis in particle physics is usually based on the detailed comparison of a theoretical expectation with the collision data measurement from the detector.

The expectation for this comparison is often derived from Monte Carlo simulation methods. Therefore, the first section of this chapter summarizes the simulation of the hard scattering process, the hadronization and showering models and introduces the different Monte Carlo generators used in this thesis.

Furthermore, in order to perform a valid comparison, the interaction of the simulated particles with the detector materials and the magnetic field has to be taken into account. The second section thus explains this important detector simulation step.

The CMS detector does not detect physics objects such as muons or jets themselves, but rather provides "raw" information from the various sub-detectors, e. g. hits in the silicon tracker and muon chambers or energy deposits in the calorimeters. Dedicated algorithms are therefore used to reconstruct physics objects in order to make comprehensive simulation-data comparisons based on these information.

A peculiarity of the CMS event reconstruction is the ParticleFlow (PF) approach [65]. In contrast to the conventional reconstruction procedure (e. g. [66]), this algorithm simultaneously categorizes physical objects based on the full detector information. In the third section of this chapter an explanation of the event reconstruction using the ParticleFlow approach and the ParticleFlow definitions of the physics objects used in this thesis is given.

The event reconstruction may be spoiled by temporary hardware noise or problems during data-taking. Thus, the last section provides a short description of the data quality management employed to select the "good" collision data which is used exclusively in this thesis.

4.1. Event Generation

The processes occurring in pp collisions at the LHC are of quantum mechanical and therefore probabilistic nature. It is thus convenient to rely on Monte Carlo methods [67] to simulate expectations from theory. Monte Carlo methods are algorithms which make use of random sampling and are thus ideal to estimate non-deterministic processes. As already shown in Figure 2.5, the Higgs boson production processes which are of particular interest in this thesis, but also the background processes which lead to similar experimental signatures, are expected

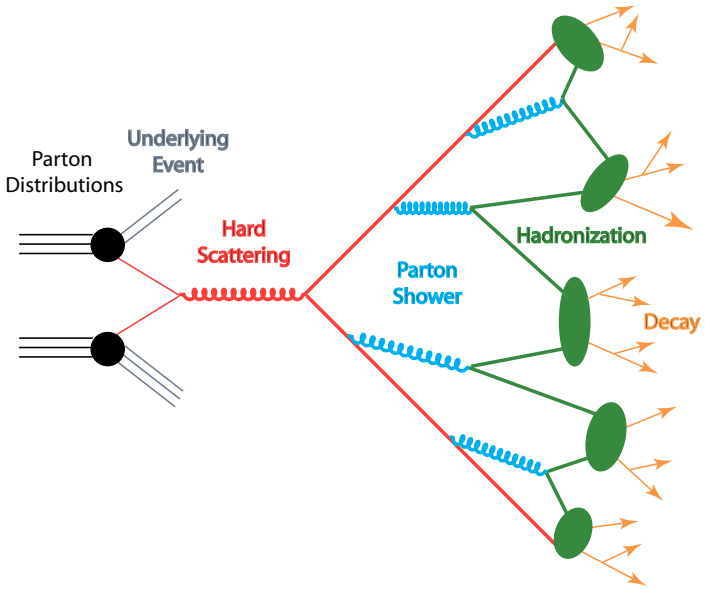


Figure 4.1.: Different phases of the event generation using Monte Carlo simulation, taken from [68]. Radiation of electrically or color charged partons emerging from the hard scattering process is simulated in the parton showering step. Color-neutral objects are formed during the hadronization and short-lived particles are further decayed.

to occur with a low rate in ordinary pp collisions. Therefore, dedicated sets of Monte Carlo samples with prescriptions for particular processes are generated and usually combined to compare to detector data.

The event generation with Monte Carlo methods consists of several steps, which are illustrated in Figure 4.1: First, the hard scattering process is simulated according to the relevant Feynman diagrams and the corresponding probabilities, as already described in Section 2.1.5. Furthermore, additional radiation from initial and final state partons with electric or color charge is taken into account in the parton shower step of the simulation. Finally, the particles emerging from the parton shower are used to form color-neutral particles in the hadronization step and are, if necessary, decayed into long-lived, "stable" particles, which are not expected to decay further within the detector.

Hard Scattering Process

The pp collisions in the LHC are partially characterized by the PDFs, which describe the momentum fractions of the different partons. The hard scattering itself occurs at high momentum transfer scales and allows for the calculation of the partonic cross section of the hard scattering process using perturbation theory. Events are generated randomly within the accessible phase space following these perturbative calculations. Decays of resonances, such as e. g. the Higgs, or the W and Z bosons, are included in this simulation of the hard process. Decays of secondary

unstable particles are not included in this part of the simulation and occur further downstream in the simulation chain.

Parton Shower

In addition, photons and soft gluons can be radiated from any electrically or color charged parton emerging in the hard scattering process and they themselves can split up again into pairs of quarks and anti-quarks. Since these radiation processes may lead to significant corrections to the overall event topology, it is imperative to treat them in a reliable way. Two different approaches exist to simulate the initial and final state radiation.

The matrix element method is based on the evaluation of the relevant Feynman diagrams of the process in question including real gluon or photon radiation at leading order and uses the resulting transition probabilities to randomly generate events. Using Feynman diagrams the event generation with the matrix element method accounts for the kinematics, as well as the spin and helicity structure of a process and furthermore describes interference effects of multiple diagrams. The order of the perturbative calculation (e. g. leading order or next-to-leading order) corresponding to the simulation can in principle be defined by adding relevant diagrams, but is limited by the increasing complexity of the diagram determination. In addition, at lower scales the perturbative series does not converge anymore due to the running of the strong coupling constant α_s.

An alternative method, the so-called parton shower method, therefore avoids this by using an arbitrary number of splittings of one parton into two new particles. This successive radiation and splitting procedure is parametrized using the Dokshitzer-Gibov-Lipatov-Altarelli-Parisi (DGLAP) [69–71] equation evolution, leading to simulated parton showers of the final state partons from the hard process and from the initial state partons [10].

The two approaches are often combined, such that the matrix element method is used to simulate the event in leading order (LO) or next-to-leading order (NLO) up to a process dependant cut-off energy scale, which is used to separate the simulation of high energetic partons within the matrix elements and to treat the low energetic radiation with the parton showers. In order to correctly combine the two approaches different dedicated matching algorithms have been developed. The matching schemes which are most frequently used are the CKKW [72] scheme, which is based on matching the transverse momenta of the partons, and the MLM [73] scheme, which is based on angular matching.

Hadronization & Decay into Stable Particles

The particles emerging from the shower are still elementary particles. Due to the confinement property of quantum chromodynamics quarks and gluons, however, cannot exist as free particles. Several models have thus been developed to simulate the formation of color-neutral bound states from these particles. Since the detailed

description of this hadronization process is not feasible starting from the QCD Lagrangian, these hadronization models are mainly based on phenomenological descriptions. Prominent examples are the Lund string fragmentation [74] and the cluster hadronization model [75].

In addition, many of the particles evolving from the hadronization simulation have a lifetime which is short with respect to the distance they travel through the CMS detector. Their decay into stable particles, i. e. particles with a lifetime long enough to fully traverse the detector, is simulated according to the known branching fractions of these different mesons and baryons.

Underlying Event and Pile-Up

The scattering process in a typical pp collision does not only consist of exactly the two partons from the aforementioned hard scattering process. Proton remnants, which are themselves color-connected, may also interact and do influence the overall event topology. In addition, secondary hard interactions may take place at the same time between two partons of the proton remnants. All of these additional processes are summarized under the name of underlying event and are accounted for in the Monte Carlo simulation [76].

Since the LHC employs proton bunches with high proton densities and currently uses bunch spacings of the order of 50 ns in order to achieve high instantaneous luminosities, multiple proton-proton collisions per bunch crossing usually occur. These additional pp interactions are called pile-up (PU) events. In-time pile-up describes additional interactions resulting from pile-up events in the same bunch crossing. Due to the finite response time of the detector electronics, pile-up events can also be produced in adjacent bunch crossings. These PU contributions are referred to as out-of-time pile-up. While PU did not play a significant role in the 2010 data taking period, its importance drastically increased in 2011 and both contributions, in-time and out-of-time PU events are mixed into the MC simulation to account for this effect.

4.1.1. Event Generators

Several MC generators have been used throughout this thesis for the simulation of the signal and background processes of the Higgs boson search presented in Chapter 7, but also for the measurement of the jet energy resolutions in Chapter 6. An overview of their general properties is thus given in the following.

Pythia

Pythia 6.4 [77] is a multi-purpose MC package which can be used to generate a very wide range of Standard Model and Beyond the Standard Model processes. It provides algorithms to simulate the hard scattering process, the parton shower, hadronization, particle decays and the underlying event. A particular strength of Pythia 6.4 is the simulation of soft QCD radiation, which can not be calculated

using perturbation theory. The simulation of the hard scattering process is based on leading order calculations only, but other MC generators providing the simulation of a higher leading order hard process, e. g. based on the matrix element method, can be interfaced to Pythia 6.4 to use its showering and hadronization routines. By default, Pythia 6.4 employs the aforementioned Lund string fragmentation model to simulate the hadronization step of the event generation. Various free parameters allow for a fine tuning of the different simulation routines to achieve good data and MC simulation agreement. Complete sets of optimized and fixed parameters are therefore referred to as generator tunes in the following.

Herwig++

HERWIG++ [78] is an alternative multi-purpose event generator to Pythia 6.4, again providing a full set of routines simulating SM and BSM processes, showering, hadronization, particle decays and the underlying event. Its roots are based on the old Fortran version of HERWIG [79], but HERWIG++ has been completely rewritten in C++ for the LHC era. One of the main differences to Pythia 6.4 is the use of a cluster hadronization model instead of the Lund string model. As with Pythia 6.4, other MC generators can be interfaced to HERWIG++ to use its showering and hadronization routines.

Madgraph

MadGraph [80] is a leading order matrix element generator. It is designed to cover most of the SM physics and also provides many additional BSM processes. Event generation using MadGraph is a two step procedure. MadGraph itself determines all relevant LO Feynman diagrams of a specified process including real gluon or photon radiation at LO. The MadEvent package, which comes with MadGraph, is then used for the actual event generation and randomly draws events according to the differential cross sections determined by MadGraph. The MadGraph package does not contain routines for showering or hadronization and is thus usually interfaced to Pythia 6.4, to take care of these downstream in the overall simulation process.

Powheg

Powheg [81, 82] is an event generator which provides certain selected physics processes with NLO precision, such as production of single top quarks [83] and Higgs boson production via Higgsstrahlung [84], which are both of importance in the Higgs boson search described in chapter 7. One general problem of NLO generators is the danger of over-counting when matching the NLO event generator to a LO parton shower generator, such as e. g. HERWIG++ or Pythia 6.4. The Powheg method employs a special technique to overcome this problem. It generates the hardest emissions from the NLO process first and prior to applying a dedicated subtraction formalism in order to solve the over-counting issues. This

implies that POWHEG needs to be interfaced with p_T- or angular-ordered parton shower algorithms. Virtuality-ordered showers are not supported.

4.2. Detector Simulation

The event generation described previously does not yet account for any interaction of the "generator particles" with the detector material or the magnetic field. A dedicated detector simulation based on GEANT 4 [85, 86] allows for the detailed modelling of the CMS detector geometry and material budget and contains routines for the simulation of the full detector response to the different particles. GEANT 4 simulates electromagnetic and hadronic interactions and models the trajectories of the particles within CMS's magnetic field. The simulation not only includes additional showering of the generator particles interacting with the detector material and bremsstrahlung, the radiation of photons from charged particles, but also comprises the simulation of the electronic signals produced by particles traversing the different sub-systems of CMS. The detector simulation thus models the full detector response to a generated event and ensures full comparability between events from MC simulation and events from pp collision data.

4.3. ParticleFlow Event Reconstruction

In order to perform a comprehensive data analysis using MC simulation and collision data, the physics objects, such as leptons or jets, have to be reconstructed from the the raw detector information, such as hits in the tracker or energy deposits in the calorimeters. In contrast to the conventional object-by-object reconstruction, the ParticleFlow approach employed by CMS aims at the simultaneous reconstruction of all stable particles (in PF categories muons, electrons, photons and charged and neutral hadrons) in the event based on the full available sub-detector information. A detailed description of the algorithm is given in [65], summaries of the algorithm performance on data are given in [87–89].

The PF algorithm basically consists of three main components. In a first step fundamental reconstructed objects, such as tracks and calorimeter clusters, are formed and stored in an intermediate collection (PFElements). These elements are then linked by the algorithm based on the respective distances into so-called building blocks (PFBlocks) which are finally used to form the particle candidates (PFCandidates).

4.3.1. Tracks and Primary Vertices

Tracks indicate the reconstructed trajectory of charged particles traversing the tracker. They are essential ingredients (PFElements) to the PF algorithm and allow - through their bending in the magnetic field - the precise determination of the particle momenta and the sign of their electrical charge. The default track reconstruction

algorithm employed in CMS is the Combinatorial Track Finder (CTF) [90–92]. The CTF may be run "inside-out" (standard) starting with the innermost pixel layers and growing towards the last tracker layer or "outside-in" (e. g. for lepton tracks), starting with the outermost silicon-strip tracker and working towards the innermost pixel layer. In both configurations, the four main conceptual steps are the same. Initial track candidates are seeded by each pair of hits in two given layers. Starting from each of these seeds, pattern recognition is used to propagate trajectory candidates iteratively from layer to layer using a combinatorial Kalman filter [93,94] and taking into account effects such as multiple scattering, energy loss or missing hits, until the last layer is reached. The number of possible combinations (and thus the CPU footprint) during this step is constrained by requiring quality criteria, such as a minimal number of valid hits, a maximal number of missing hits and a χ^2 compatibility test of hits matching the predicted trajectories. The large collection of resulting track candidates is then cleaned for ambiguities, e. g. track candidates sharing the same seeds or multiple hits along their trajectories, based on the number of valid hits and the track quality. The optimal track parameters of the selected tracks are finally obtained by refitting the tracks inside-out and outside-in. In order to reconstruct all tracks, the CTF is run several times, each time removing the hits which belong to already selected high-purity tracks of previous iterations.

The collection of high-purity tracks can be used to define candidates of primary vertices. The primary vertex (PV) indicates the origin of the collision in the detector and is important for the interpretation of the rest of the event. In LHC collisions, where multiple interactions per bunch crossing occur (pile-up), one expects more than one good primary vertex and the number of reconstructed PVs can be used as an estimator of the number of (in-time) pile-up events. The reconstruction of primary vertices is based on the Adaptive Vertex Fitting (AVF) [95] method, a modified version of the Kalman filter. The tracks assigned to each of the vertex candidates are weighted based on their compatibility χ^2, assigning lower weights to more incompatible tracks. The primary vertex candidates are then refitted iteratively using these weights. The final classification as good primary vertex is based on the weighted sum of tracks belonging to the PV. The collection of good primary vertices is finally sorted by the weighted p_{T}^2 sum of the associated tracks; the one with the highest sum is usually selected as the primary vertex of the hard process in the event, while the others are associated to pile-up.

Further descriptions of track and primary vertex reconstruction, as well as detailed performance results on collision data, are given in [96].

4.3.2. Calorimeter Clustering

The second important ingredient for the PF algorithm are calorimeter clusters [65], which are computed separately in the different ECAL and HCAL sub-detectors. The ParticleFlow clustering algorithm defines a grid of cells based on the respective detector granularity and identifies cluster seeds as cells with an energy exceeding a specific threshold. For each of the cells topological clusters are then accumulated

iteratively by adding adjacent cells with energy thresholds of two standard deviations over the noise in the respective sub-detectors. This corresponds to 80 MeV in the ECAL barrel and up to 300 MeV in the endcaps. In the HCAL, the clustering threshold amounts to up to 800 MeV of energy due to the coarse granularity. Starting from the topological clusters so called ParticleFlow clusters are then derived iteratively by a weighted sharing of all cell energies based on the cell-cluster distances. The detailed prescription and weight function can be found in [97].

4.3.3. Linking Algorithm

Particles traversing the CMS detector most likely give rise to several PFElements. An electron for example would lead to the reconstruction of a charged-particle track and one or multiple calorimeter clusters. The various reconstructed PFElements are therefore connected through the linking algorithm into so called PFBlocks in order to fully reconstruct the particles and avoid double counting in different sub-detectors.

Links between tracks and calorimeter clusters are obtained in a three step procedure. First, each track is extrapolated from the last measured hit in the tracker to the PFElements in the ECAL pre-shower detector. Then, the extrapolation is extended into the ECAL at a depth corresponding to the expected maximum of a typical electron shower. Finally, the track is evaluated in the HCAL at a calorimeter depth of one interaction length λ, which is typical for a hadron shower. If the extrapolated track lies within the different cluster boundaries the respective links are established. The presence of gaps or possible cracks between the calorimeter cells is taken into account by extending the cluster boundaries by one cell. The quality of each link is defined by the algorithm in terms of the distance in the (η, ϕ)-plane of the extrapolated track position and the calorimeter cluster position. In addition, energy from bremsstrahlung photons emitted by electrons is recollected by fitting tangents to tracks linked with ECAL clusters in each tracker layer and linking them to additional clusters in the ECAL as potential photon energy contributions. Links between any two calorimeter clusters are established when the cluster positions of the calorimeters with finer granularity lie within the cluster boundaries of calorimeters with worse granularity. A link between a charged-particle track and a track in the muon systems is established when a global fit between the two returns a good χ^2 (see also Section 4.3.4).

Based on the quality of each of these links the PF algorithm produces PFBlocks of linked PFElements which are used to categorize and reconstruct the different PFCandidates. This is explained in detail in the following Sections. An important aspect of the PF approach is the fact that for each reconstructed PFCandidate, the corresponding PFElements are removed from the corresponding PFBlocks for the next reconstruction step to avoid energy double counting. It should also be noted that the fine granularity in the different CMS sub-detectors ensures that most blocks are only made of a small number of elements, leading to an algorithm performance which is to some extent independent of the event complexity and

allows reconstruction even in high luminosity (pile-up) environments.

4.3.4. Muons

Muons are minimally ionizing charged particles and thus have a rather clean signature in the CMS detector. The reconstruction of the PFCandidates thus starts with the evaluation of PFBlocks involving links to the muon systems. As already mentioned, links of tracker tracks and muon tracks are based on global χ^2 fits [64,89,98] and the corresponding muon candidates are referred to as "global muons" within CMS. Each global muon candidate gives rise to a PF muon if the combined momentum is compatible with a measurement based solely on the tracker within three standard deviations. If a PF muon is reconstructed the corresponding PFElements are removed from the PF collections. Furthermore, average muonic energy depositions in the HCAL (ECAL) are estimated based on cosmic muon measurements to be 3 (0.5) GeV, and subtracted from matching calorimeter clusters with an uncertainty of $\pm 100\%$.

4.3.5. Electrons

Electrons also show a characteristic signature in the CMS detector. For the lighter electrons the tracker with its large silicon material budget acts like an additional pre-shower detector leading to a significant energy loss of the electrons through the radiation of bremsstrahlung. High-energetic bremsstrahlung photon emission may cause kinks in the electron trajectories, such that the conventional track reconstruction is not able to follow the electron path. Therefore a dedicated electron track reconstruction is employed [99] for electron candidates. Two complementary methods [89, 100] are used to reconstruct PF electrons, the "ECAL-driven" reconstruction which works well for high p_{T} electrons and the "tracker-driven" reconstruction which is advantageous for non-isolated and low p_{T} electrons.

The ECAL-driven reconstruction aggregates all the bremsstrahlung photon and electron energy deposits into so called "super-clusters", which are narrow regions in η but rather large along ϕ (the direction of bremsstrahlung emission), and uses the super-cluster as a seed for the electron track reconstruction.

The tracker-driven reconstruction method exploits the characteristic electron trajectories with respect to other particles. Each track in a relevant PFBlock is pre-identified with a multi-variate analysis [101] and then - depending on the result - refitted as an electron track candidate to follow the trajectory all the way into the ECAL. A final selection of the tracker-driven electron candidate is then performed based on a combination of tracking and calorimetric variables.

Each of the reconstructed electrons from any of the two methods gives rise to a PF electron and the corresponding tracks and calorimeter clusters (including clusters identified as bremsstrahlung photons) are removed from the PFBlock. The most recent detailed technical information about the PF electron reconstruction implementations are given in [102].

4.3.6. Photons and Hadrons

After the muon and electron reconstruction, the remaining PFElements in a PFBlock may be attributed to charged hadrons, photons or neutral hadrons. In principle, the detection of charged hadrons is based on the links of the remaining tracks to calorimeter clusters and the reconstruction of neutral particles is finally based on the comparison of the remaining excess calorimeter cluster energies in the ECAL (photons) and HCAL (neutral hadrons). However, since the CMS calorimeters are non-compensating and the ECAL is calibrated for photons (with a substantially different response for hadrons) and the HCAL is calibrated assuming pions with energies of 50 GeV (not interacting in the ECAL), the PF algorithms performs a dedicated recalibration [65] of all remaining calorimeter cluster energies in order to correctly compare calorimeter energies and detect neutral particles. More specifically, after the calorimeter cluster recalibration, each of the remaining tracks in a given PFBlock gives rise to a PF charged hadron. In order to improve the energy resolution, the tracker momentum measurement is used and the charged hadron energy is determined under the charged pion mass hypothesis. The PFElements corresponding to the charged hadrons are then removed from the PFBlock. If the recalibrated energy in the ECAL and HCAL clusters linked to tracks from the previously reconstructed charged hadron is larger than the total associated charged particle energy, the energy excess is used to give rise to a PF photon. If an additional energy excess remains, this gives rise to a PF neutral hadron. The remaining calorimeter clusters in the ECAL and HCAL, which were not linked to any tracks, give rise to PF photons and PF neutral hadrons, respectively. In the latter case only the HCAL calorimeter clusters are recalibrated correspondingly.

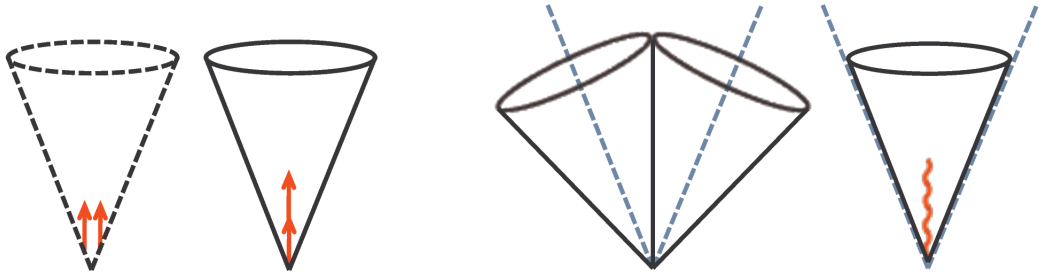
At this point all PF physics objects are reconstructed and can be used as input to higher level physics objects such as jets or missing transverse energy (MET).

4.3.7. Jets

Most of the final state particles in the pp collisions of the LHC are quarks and gluons. Due to the nature of the strong force they immediately hadronize into sprays of collimated mesons and baryons. Jet algorithms are thus used to re-group adjacent particles (hadrons, as well as non-isolated leptons and photons) observed in the detector into jets and associate their energies and directions with the initial partons. Jets are therefore extremely important to relate experimental observations with theory predictions.

Jet Algorithms

There are several different jet algorithms employed in CMS analyses and all of them come with their implementations in the FastJet [103] software package. One usually differentiates between two general types of algorithms, simple cone based algorithms (such as e. g. SISCone [104]) which identify the energy flow within cones of specific radii and sequential clustering algorithms which are based on



(a) Violation of collinear safety (b) Violation of infrared safety

Figure 4.2.: Illustration of collinear and infrared safety: A collinear splitting should not lead to a difference in the number of reconstructed jets (a), nor should a soft emission (e. g. pile-up or underlying event) change the definition of the jet (b).

pair-wise successive recombinations according to a distance parameter. The three algorithms used in this thesis belong to the latter group: the k_T [105, 106] jet algorithm, the anti-k_T [107] jet algorithm and the Cambridge/Aachen (CA) [108] jet algorithm. All of these algorithms fulfill the two fundamental requirements imposed by theory [109], collinear-splitting safety and infrared safety, which are explained in Figure 4.2.

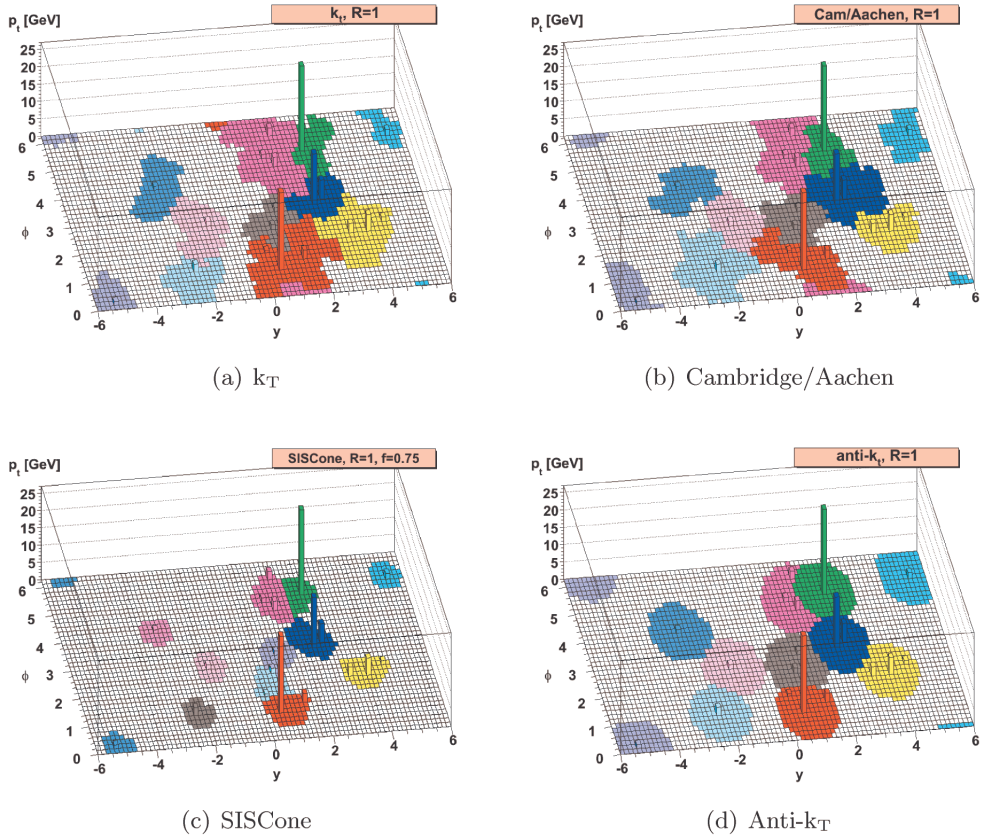
The three sequential clustering algorithms are based on the distance d_{ij} between each pair of particles,

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2},$$

and the distance between any particle i and the beam

$$d_{iB} = p_{T,i}^{2p}.$$

After all distances d_{ij} and d_{iB} are calculated the algorithm finds the smallest and vectorially combines the two particles i and j into a new particle i' if $d_{ij} < d_{iB}$, otherwise the particle i is removed from the list and called a jet. This prescription is repeated until all available particles are clustered into jets. The parameter R^2 scales the distance d_{ij} with respect to d_{iB} such that any pair of final jets is separated at least by R. The parameter p governs the relative power of energy versus geometrical scales in order to distinguish the three different algorithms. The parameter choice $p = 1$ corresponds to the k_T algorithm, which was already extensively used at LEP. It is used in this thesis to estimate the average pile-up contamination per event as explained later when describing the Jet Energy Corrections. The Cambridge/Aachen (CA) algorithm uses $p = 0$. This means it is based solely on the geometrical distance between particles. The CA algorithm has been found to be very well suited for the analysis of jet sub-structures [110]. The rather new anti-k_T algorithm uses $p = -1$ and produces cone-shaped jets (see Figure 4.3). It is the default CMS analysis jet algorithm (with $R = 0.5$).



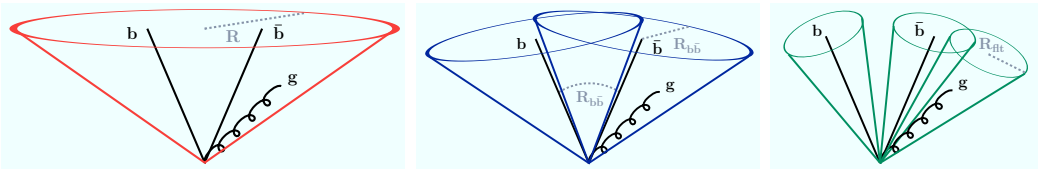
(a) k_T (b) Cambridge/Aachen

(c) SISCone (d) Anti-k_T

Figure 4.3.: Clustering properties of different jet algorithms, taken from [107]. The k_T algorithm (a) is a classic sequential clustering algorithm. The Cambridge/Aachen (b) algorithm has been found to perform best in resolving jet sub-structure. SISCone (c) is an infrared and collinear safe cone algorithm, but uses a lot of computation resources for high numbers of input objects. The anti-k_T algorithm (d) is used as default jet algorithm within CMS. It provides the advantages of the sequential clustering algorithms and produces cone-shaped jets.

Fat-, Sub- and Filterjets

Apart from the standard jet algorithms the Higgs boson search presented in Chapter 7 also employs a dedicated Subjet/Filter (SJF) algorithm which has been specifically developed to reconstruct heavily boosted objects [111], such as a high energetic Higgs boson decaying into two b-jets. The SJF algorithm is designed for highly efficient reconstruction of both b partons and - if required - leading order gluon radiation using "fat jets". It adapts flexibly to the momentum dependence of the angle between the decay products of the Higgs boson, denoted $\Delta R(j_1, j_2)$ in the following, by resolving "subjets". Furthermore, it addresses contamination due to the underlying event and pile-up giving optimal mass resolution by reclustering



(a) fat jets (b) subjets (c) filter jets

Figure 4.4.: Illustration of the Subjet/Filter algorithm, adapted from [111]. A fat jet (a) is reconstructed using the CA jet algorithm with $R = \pi/2$. Two subjets (b) are then extracted from the fat jet taking into account the momentum dependence of the angle between the Higgs boson decay products, $R_{b\bar{b}} \equiv \Delta R(j_1, j_2)$. Three filter jets (c) are obtained by reclustering with a smaller radius $R_{\text{filt}} = \min(0.3, R_{b\bar{b}}/2)$ in order to reject contamination form the underlying event and pile-up.

"filter jets". The different stages of the SJF algorithm are depicted in Figure 4.4. A fat jet is reconstructed using the Cambridge/Aachen jet algorithm with a large distance parameter of $R = \pi/2$. The successive recombinations are then iteratively undone to peel off two subjets requiring a significant mass-drop and symmetric splitting:

1. Break the jet j into two subjets by undoing the last clustering step. They are labeled such that $m_{j_1} > m_{j_2}$.
2. If there is a significant mass drop defined by $m_{j_1} < \mu \cdot m_j$, and if the splitting is not too asymmetric, $y = \frac{\min(p^2_{\text{T},j_1}, p^2_{\text{T},j_2})}{m_j^2} \cdot \Delta R^2(j_1, j_2) > y_{\text{cut}}$, associate the jet j with the Higgs boson neighbourhood and exit the loop. The mass drop requirement is parameterized by μ, the asymmetry parameter is indicated with y_{cut}.
3. Otherwise redefine jet j to be equal to j_1 and go back to step 1.

The implementation used in this thesis follows the theoretical considerations in [111] and uses the suggested mass drop cut of $\mu = 0.67$ and an asymmetry parameter of $y_{\text{cut}} = 0.09$. The cut on y is expected to reduce asymmetric configurations which might generate significant jet masses in non-b- or single-b-jets due to soft gluon divergence. In addition it is advantageous to also filter the Higgs boson neighbourhood defined by the fat jet by re-running the CA jet algorithm with a significantly smaller radius $R_{\text{filt}} = \min(0.3, \Delta R(j_1, j_2)/2)$. The three hardest filter jets (corresponding to two b-partons and leading order gluon radiation) that appear can then be taken to describe the Higgs boson candidate, while contamination from the underlying event or pile-up is rejected.

Calo- and Jet-Plus-Track-Jets

The input to all the jet clustering algorithms described above are generic four-vectors describing the directions and energies of the input objects. These may

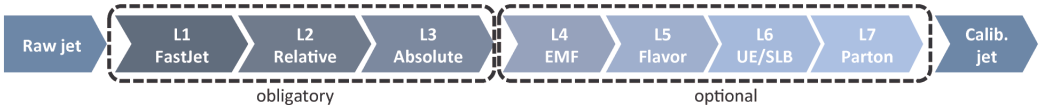


Figure 4.5.: The factorized jet energy correction approach: Each level corrects for a different jet-p_{T} bias. The first three correction levels are obligatory, while the latter correct for analysis specific effects.

be compososite objects, such as the PFCandidates previously described, or e. g. energy deposit measurements in the electromagnetic and hadronic calorimeter cells. Therefore, one differentiates not only between the jet algorithms, but also between the input objects used for the jet clustering.

Apart from PFJets, i. e. jets clustered with jet algorithms using PFCandidates as inputs, two other types of jets are of importance in Chapter 6. CaloJets and Jet-Plus-Track (JPT) jets are built using the same jet algorithms, but use different input objects. CaloJets and JPTJets were extensively used as alternatives to PFJets in the beginning of the CMS data taking, especially in 2010, to cross-validate the PF reconstruction algorithm.

CaloJets are reconstructed from energy deposits in the electromagnetic and hadronic calorimeter cells, which are combined into projective calorimeter towers. A calorimeter tower in the detector barrel ($|\eta| < 1.4$) corresponds to the unweighted sum of one single HCAL cell and, due to the finer granularity of the ECAL, a 5×5 ECAL crystals block. In the forward detector region thresholds are applied to individual calorimeter cells when building the calorimeter towers in order to suppress noise from the readout electronics and contributions from pile-up [112].

JPTJets [113] combine CaloJets with additional information from the tracking system to improve the p_{T} response and resolution of the jets by measuring charged particles in the tracker. The JPT algorithm associates charged particle tracks with the CaloJets based on spatial seperation in (η, ϕ) between the CaloJet axis and the track momentum at the interaction vertex, correcting thus for bending in the magnetic field. The CaloJets are then corrected for using the more precise charged particle track measurements.

Jet Energy Corrections

The purpose of the jet energy correction (JEC) is to relate, on average, the energy of a raw detector jet to the energy of a corresponding true particle jet as expected from MC simulation. This correction is required for comparisons with theory predictions. Several mechanisms cause the measured energy of the detector jet to deviate from its real energy, such as the non-linearity of the calorimeters, saturation effects and others. They are described in the following for each correction step.

The deviation is usually quantified in terms of the jet response

$$\mathcal{R} = \frac{p_{\mathrm{T}}^{\mathrm{jet}}}{p_{\mathrm{T}}^{\mathrm{ref}}}\,, \tag{4.1}$$

where the super-script "jet" labels the measured detector jet and "ref" stands for the true generator particle reference jet. CMS employs a factorized approach with three obligatory correction levels to adjust for the different mechanisms [114].

L1 FastJet: Electronics noise and contamination from pile-up contribute to the measurement of the raw jet. In this thesis, this is corrected for on an event-by-event basis by estimating an average energy density ρ per unit area [115] in the detector and subtracting it according to the area of the jet.

L2 Relative: The non-uniformity of the CMS sub-detectors leads to variations of the jet response with respect to the pseudorapidity η. Large samples of simulated events dominated by QCD activity (QCD MC) are used to derive η-dependant correction factors to account for this effect. The inverse of the average response $\langle R\rangle^{-1}$ in each bin of η is then used as correction factor to the jet momentum.

L3 Absolute: The non-linearity of the CMS calorimeters leads to an additional variation of the jet response as a function of the jet p_{T}. Again, QCD MC simulation is used to derive p_{T}-dependant correction factors.

An additional correction *L2L3 Residual* is applied to data taking into account the fact, that the L2 and L3 corrections are determined solely on MC. It adjusts the momentum of the jets for small residual differences observed in data/MC comparison studies using the dijet p_{T} balancing method for the relative jet energy correction and γ/Z+jet balancing techniques for the absolute jet energy correction. Furthermore, the use of PF particles as input objects to the jets allows for an additional filtering of pile-up contributions. The *Charged Hadron Subtraction* (CHS) correction filters PF particles associated to pile-up vertices (in z, preserving particles from secondary vertices, such as b-decays) out of the particle list before reconstructing any higher level physics objects. Other available Monte Carlo jet energy corrections, which have not been used in this thesis, are the L4 through L7 corrections, as sketched in Figure 4.5. These correction levels are currently not recommended as default (and are thus not officially maintained) by the CMS JEC group. They are thus not applied in most of the current CMS analyses.

Jet Energy Resolution

The precise knowledge of the jet energy resolution (JER) is very important to analyses with jets in the final state. As the measurement of the JER in CMS comprises a significant part of the work towards this thesis it is described in more detail in Chapter 6.

Tagging of b-Jets

The search for a light Higgs boson in the decay $H \to b\bar{b}$ requires very good identification of b-jets which originate from the hadronization of b-quarks in the final state. There are several algorithms, so called "b-taggers", employed by CMS which discriminate b-jets from light quark or gluon jets. Most of them are based on the lifetime of the B hadron which is longer than for light hadrons and allows for the reconstruction of a 3D-displaced secondary vertex (SV) from the B hadron decay. Other b-tagging algorithms use the relative momentum of non-isolated soft leptons from semi-leptonic B hadron decays with respect to the jet direction for b-tagging. Two algorithms are used in the studies presented in the following chapters, the Secondary Vertex High-Purity tagger (SSVHP) and the Combined Secondary Vertex (CSV) tagger.

The Secondary Vertex High-Purity tagger [116] requires the reconstruction of at least one secondary vertex with at least three associated tracks in the event. The discriminator value of this tagger is then computed from the significance of the three-dimensional decay length between the primary and secondary vertex. The SSVHP tagger is a rather simple b-tagger which is easy to calibrate and which was thus predominantly used in the first year of LHC data-taking.

The Combined Secondary Vertex (CSV) tagger [116] is the most advanced b-tagger currently employed in CMS physics analyses. Input to the CSV are not only information about the secondary vertex, but also additional lifetime information and various track information. This allows for effective b-tagging even if no secondary vertex is reconstructed and yields a very high b-tag efficiency, which is not limited by the SV reconstruction efficiency. The different input variables are evaluated in a likelihood ratio which is applied twice, first to discriminate between b- and c-jets and between b- and light jets, and then, with a weighted average, to produce a final CSV b-tag discriminator decision.

4.3.8. Missing Transverse Energy

Neutrinos are special particles in the sense that they do not interact with any of the sub-detectors. Therefore they have to be inferred indirectly from the remaining energy imbalance in the transverse plane of the detector after all PF particles have been reconstructed. This imbalance, which is called missing transverse energy (MET), is calculated from the negative sum over all reconstructed ParticleFlow candidates [114, 117],

$$\vec{\not{E}}_{\mathrm{T}} = -\sum_i (E_i \sin\theta_i \cos\phi_i \hat{\mathbf{x}} + E_i \sin\theta_i \sin\phi_i \hat{\mathbf{y}}) = \not{E}_x \hat{\mathbf{x}} + \not{E}_y \hat{\mathbf{y}}\,,$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ are the unit vectors along the x and y axis as shown in Figure 3.4. The magnitude of the missing transverse energy vector, $|\vec{\not{E}}_{\mathrm{T}}|$, is abbreviated with $\not{E}_{\mathrm{T}}$ in the following and is often chosen to quantify the neutrino transverse momentum. A related variable extensively used in the Higgs boson search is the $\not{E}_{\mathrm{T}}$ significance,

which is obtained by dividing the $\not\!\!E_{\mathrm{T}}$ by the square root of the scalar sum of the transverse energies of all PF particles in the event.

Closely related to the missing transverse energy $\not\!\!E_{\mathrm{T}}$ is a physics object referred to as PFMHT. It is computed in analogy to $\not\!\!E_{\mathrm{T}}$, but is based on PFJets above a p_{T} threshold of 30 GeV as opposed to all particles. Due to the strict low latency requirements of the High Level Trigger, PFMHT is used as a fast alternative to $\not\!\!E_{\mathrm{T}}$ in the online reconstruction.

4.4. Data Quality Management

A key ingredient to successful data analysis is to ensure constant good conditions during the collision data taking in order to provide compatible sets of data for a comparison with simulated events. Unfortunately, data taking conditions do change due to various possible reasons and the data taking may be spoiled, e. g. because of temporary electronic noise in single read-out channels or problems in a whole sub-detector component. Since it is very difficult for individual analysts to spot the sometimes very subtle problems which may (and do) occur during data taking, CMS has a centralized Data Quality Management (DQM) [118] group of experts who constantly check the quality of the data. Based on these checks, the DQM group regularly publishes a so-called "JSON" list of good data samples which are the input for the studies presented in this thesis.

5. Multivariate Methods and Statistical Tools

The physics analyses presented in the following chapters are making use of three important statistical and multivariate concepts: Parameter estimation using the maximum likelihood estimator method, construction of exclusion limits with the CL_s and Bayesian approaches and event classification using Neural Networks.

This chapter gives an overview over these methods and their implementations in the theta statistical inference framework [119] and the NeuroBayes package [120, 121].

5.1. Maximum Likelihood Parameter Estimation

The maximum likelihood estimator (MLE) method allows the calculation of the best estimator $\hat{a}$ of parameters describing the underlying probability density function (p.d.f.) of a data sample [122]: Consider a set of N measurements $X = \{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N\}$. The underlying p.d.f. $f(\vec{x}_i|\vec{a})$, which is the probability to measure $\vec{x}_i$ for a given $\vec{a}$, has to be known and normalized ($\int_\Omega f(\vec{x}|\vec{a})d\vec{x} = 1$ for all $\vec{a}$). If the $\vec{x}_i$ are statistically independent, the joint probability density function for X is obtained by taking the product of the individual p.d.f.s. In this case the *Likelihood Function*,

$$\mathcal{L}(\vec{a}) = f(\vec{x_1}|\vec{a}) \cdot f(\vec{x_2}|\vec{a}) \cdots f(\vec{x_N}|\vec{a}) = \prod_{i=1}^{N} f(\vec{x_i}|\vec{a}),$$

gives the probability to find X depending on the choice of parameters $\vec{a}$. The calculation of the intuitively "best" parameter estimator $\hat{a}$ thus results in the maximization of $\mathcal{L}(\vec{a})$. For computational reasons, it is expedient to work with the logarithm of the likelihood function, since the logarithm is monotonous and does not change the position of the extrema. The product in the likelihood function can then be rewritten as a sum

$$\log \mathcal{L}(\vec{a}) = \log \prod_{i=1}^{N} f(\vec{x_i}|\vec{a}) = \sum_{i=1}^{N} \log f(\vec{x_i}|\vec{a}).$$

This is also called the *log-likelihood method*.

Because most optimization algorithms are available as minimizers for historical reasons, one often employs the negative log-likelihood, $-\log \mathcal{L}(\vec{a})$, and searches for

the minimum, instead of the maximum, to calculate the best parameter estimator $\hat{a}$. The MLE method is implemented in the statistics framework theta, which is discussed in Section 5.3.

5.2. Exclusion Limits

In the absence of signal events in a high energy physics search, e. g. for Higgs boson production in association with a vector boson, the degree of confidence in the signal smallness is statistically quantified using exclusion limits. Various different methods for the derivation of exclusion limits exist. In the following the CL_s exclusion limit prescription, which is the current default at CMS, and the Bayesian approach, which is a computationally fast alternative, are discussed. Both of them are applied in Chapter 7 using their implementation in the theta framework, which is discussed in Section 5.3.

CL_s Exclusion Limit

The primary technique for deriving exclusion limits in this thesis uses the CL_s [123, 124] method in its LHC-like modified frequentist form [125]. It is based on the profile likelihood test statistic

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data} \,|\, \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data} \,|\, \hat{\mu}, \hat{\theta}_{\hat{\mu}})} \quad \text{with } 0 \leq \hat{\mu} \leq \mu.$$

In the realm of this thesis, the likelihood function for limit setting is always given by the product of the binned likelihoods of each (statistically independent) analysis channel, written as

$$\mathcal{L}(\text{data} \,|\, \mu, \theta) = \text{Poisson}\left(n \,|\, \mu \cdot s(\theta) + b(\theta)\right) \cdot \pi_\theta(\theta),$$

where the Poisson term stands for the product of probabilities to observe n_i events in each bin i,

$$\prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i}.$$

In this test statistic definition "data" refers to the actual experimental observation (n_i) or pseudo-data drawn from toy experiments, which are discussed in the following paragraph. The parameter μ represents the signal strength modifier, $\mu \equiv \sigma/\sigma_{\text{SM}}$, and θ stands for the set of nuisance parameters of the model. The prior $\pi_\theta(\theta)$ describes the probability distributions of the nuisance parameters, e. g. Gaussian distributions. The best estimator value of the signal strength modifier is denoted with $\hat{\mu}$. In elementary particle searches, the range of μ is usually restricted to a physically meaningful regime, e. g. μ is not allowed to be negative. The conditional maximum likelihood estimate of all nuisance parameters for a fixed signal

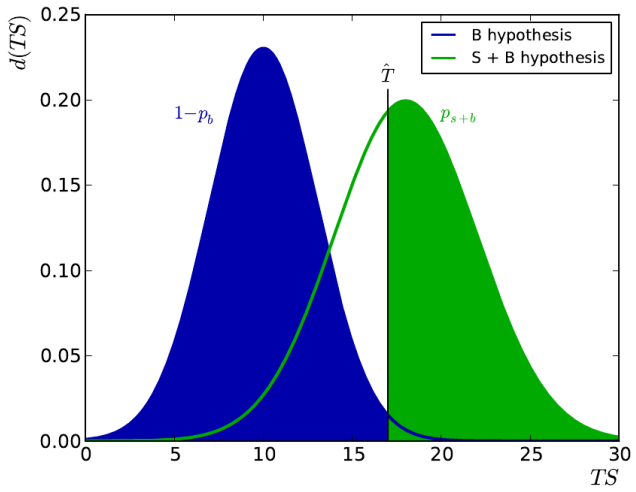


Figure 5.1.: Illustration of the definition of the CL_s values, taken from the theta documentation [119]: With test statistic (TS) distributions for the signal-plus-background $(s+b)$ and the background-only (b) hypotheses, the CL_s value for a measurement with test statistic value $\hat{T}$ is defined as $p_\mu/(1-p_b)$, where $p_\mu \equiv 1 - p_{s+b}$.

strength μ is denoted $\hat{\theta}_\mu$. The symbols n_i, s_i and b_i indicate the numbers of observed, expected signal and expected background rates in each bin i, respectively.

With these definitions, the derivation of an observed CL_s limit is a multi-step procedure: For various signal strength μ (including the background only hypothesis with $\mu = 0$) the observed test statistic value $\tilde{q}_\mu^{\text{obs.}}$ is calculated. In addition, the probability density functions $f(\tilde{q}_\mu \,|\, \mu, \theta)$ and $f(\tilde{q}_\mu \,|\, 0, \theta)$ are constructed for the signal-plus-background hypothesis and for the background-only hypothesis by generating toy Monte Carlo pseudo-data.

The $CL_s(\mu)$ value is then defined as the ratio of the two related probabilities,

$$CL_s(\mu) \equiv \frac{p_\mu}{1 - p_b}.$$

These p-values are given by

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs.}} \mid \text{signal} + \text{background}) = \int_{\tilde{q}_\mu^{\text{obs.}}}^{\infty} \int_\theta f(\tilde{q}_\mu \,|\, \mu, \theta) \cdot \pi_\theta(\theta)\, d\theta\, d\tilde{q}_\mu,$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs.}} \mid \text{background-only}) = \int_{\tilde{q}_\mu^{\text{obs.}}}^{\infty} \int_\theta f(\tilde{q}_\mu \,|\, 0, \theta) \cdot \pi_\theta(\theta)\, d\theta\, d\tilde{q}_\mu.$$

The interpretation of the CL_s value depends on μ: For $\mu = 1$ and a $CL_s \leq \alpha$, one states that the model in question is excluded with a $(1-\alpha)$ CL_s confidence level (C.L.) at the nominal model signal strength. Consequently, to quote e. g. an observed 95% C.L. upper limit on a given Higgs model the value of μ (the

signal strength with respect to the Standard Model prediction) is adjusted until $\mathrm{CL_s} = 0.05$ is reached and denoted $\mu^{95\%CL}$ in the following.

The derivation of the corresponding expected limit and its $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands follows in principle the same procedure: Assuming the background-only hypothesis a large set of pseudo-data is generated following the prescription given above. For each toy experiment the $\mathrm{CL_s}$ and $\mu^{95\%CL}$ values are calculated as if they were real data. From the distribution of toy $\mu^{95\%CL}$ values a cumulative distribution is then derived from which the 50% intersection indicates the median expected limit value. The $\pm 1\sigma$ (68%) uncertainty bands are given by the 16% and 84% quantiles and the $\pm 2\sigma$ (95%) bands are given by the crossings at 2.5% and 97.5%, respectively.

Bayesian Exclusion Limit

In contrast to frequentist statistics, Bayesian statistics is based on the *degree of belief* into a given statistics model and needs a-priori assumptions on the probability of possible outcomes of a data measurement. Exclusion limits using the Bayesian approach are used in this thesis as a cross-check to $\mathrm{CL_s}$ exclusions and are thus only briefly described in the following. They are discussed in more detail in [3,126].

Bayes theorem is used to calculate the posterior probability density function $L(\mu)$ on the signal strength modifier μ [125],

$$L(\mu) = \frac{1}{C} \int_\theta p(\mathrm{data} \,|\, \mu s + b) \rho_\theta(\theta) \pi_\mu(\mu) d\theta,$$

where the probability density functions $\rho_\theta(\theta)$ describe the belief in the scale of the uncertainties on expected signal and background rates before the actual analysis of the data. The prior $\pi_\mu(\mu)$ is a probability density which reflects the a-priori assumptions on possible outcomes for μ. If the model probability $p(\mathrm{data} \,|\, \mu s + b)$ decreases rapidly for large μ, the actual form of π_μ is not important. In most cases it is therefore legitimate to choose a flat prior $\pi_\mu = 1$ for $\mu \geq 0$ and zero otherwise. The constant C is used to normalize the function $L(\mu)$ to unity. It should be mentioned, that the integration over $d\theta$ in $L(\mu)$ is also referred to as marginalization of the uncertainties.

In the end, the 95% C.L. upper limit on μ is obtained by numerical integration until the posterior probability reaches 95%,

$$0.95 = \int_0^{\mu^{95\%CL}} L(\mu) d\mu.$$

Both types of exclusion limits are implemented in the theta framework which is introduced in the following Section.

5.3. The theta Statistical Inference Framework

The theta Statistical Inference Framework [119] is designed, implemented and distributed by Jochen Ott. It includes several template-based modules to calculate likelihood-based quantities on large (pseudo-) datasets and provides implementations of common methods for hypothesis testing and exclusion limit derivation. theta has already been successfully used for various published CMS data measurements, including the top cross section measurements [127] and the top charge asymmetry measurement [128].

5.4. Artificial Neural Networks

Artificial Neural Networks (NNs) are a powerful tool for event classification in multivariate data analysis (MVA). In general, multivariate analysis techniques allow for the identification and exploitation of complex correlations between sets of discriminating variables, i. e. variables which differ in signal and background events in one way or the other. In contrast to a basic cut-based separation of signal and background, where requirements are applied only to projections on individual axes of the discriminating variable space, the exploitation of correlations often allows for gains in signal efficiencies and purities.

These correlations and relationships have to be learned by any multivariate tool prior to its application to classification tasks. In the case of Neural Networks this learning procedure is also referred to as training and is described in the following after a general introduction to the Neural Network topology.

Neural Network Topology

Based on the principles of information processing in the brain, Neural Networks are learning systems of interconnected nodes, which change their structure while learning patterns from training data. The Neural Networks used in this thesis consist of multiple input nodes and one output node. The purpose of each node is to connect its N input values $\vec{x}$ with sets of individual weights $\vec{w}$ and to calculate the node output as

$$\mathrm{NN}_{\mathrm{out}} = \mathcal{S}\left(\sum_{i=1}^{N} w_i \cdot x_i\right),$$

using the symmetric sigmoid function $\mathcal{S}$ to map the result to the interval $[-1, 1]$,

$$\mathcal{S}(x) = \frac{2}{1 + e^{-x}} - 1.$$

In general, a Neural Network can consist of multiple layers of nodes, where the first layer is called input layer, the last layer is the output layer and all layers in between are referred to as hidden layers. A three layer feed-forward Neural

Network structure is used for classification tasks in this thesis, meaning that the output values of the input layer are fed as inputs to the nodes of the hidden layer. The node of the output layer then combines the outputs from the hidden layer into the final $\mathrm{NN_{out}}$ discriminator value,

$$\mathrm{NN_{out}} = \mathcal{S}\left(\sum_j w_j^{(2\to3)} S\left(\sum_k w_{jk}^{(1\to2)} \cdot x_k\right)\right).$$

x_k stands for the components of the input vector $\vec{x}$ and $w_{jk}^{(1\to2)}$ denotes the weights between the input and the hidden layer. Similarly, $w_j^{(2\to3)}$ indicates the weights between the hidden and the output layer.

Neural Network Training

The training of a Neural Network is performed on a dedicated training sample of events where the true class of signal or background is known for each event. During the training the weights of the Neural Network are adjusted using an iterative backpropagation algorithm [129, 130]. At the beginning of a training, the node weights are randomly distributed. The $\mathrm{NN_{out}}$ value of each event is then computed and compared to the true target value by evaluation of a cost function which quantifies the deviation. The basic principle of backpropagation is to successively minimize this deviation. A common cost function is the entropy function,

$$\mathcal{E} = \sum_i \log\left(\frac{1}{2}\left(1 + t_i \cdot NN_{\mathrm{out},i}\right)\right).$$

t_i denotes the true target value and $NN_{\mathrm{out},i}$ the calculated Neural Network output for each data event $\vec{x}_i$. The minimization of the cost function with respect to the node weights can be achieved with the method of steepest descend.

Neural Network Classification

Once a Neural Network is trained, its structure is fixed and it may be applied to classify any data sample providing the same input variables.

5.5. The NeuroBayes Package

All neural networks used in this thesis are based on the NeuroBayes (NB) software package [120, 121]. NeuroBayes was first implemented by Michael Feindt and members of the Institut für Experimentelle Kernphysik at the Karlsruhe Institute of Technology (KIT). Since 2002 it is being maintained and distributed by the spin-off company Phi-T in Karlsruhe.

An important feature of the NeuroBayes package is the variable pre-processing performed prior to the actual Neural Network training in order to improve the learning capabilities and the robustness of the network. This pre-processing includes a flattening of the input variable distributions to ensure roughly equal bin populations and a subsequent decorrelation (principal component analysis) of all variables, such that the network only has to learn the remaining non-linear structures of the data. During the pre-processing, continuous input variables are regularized using spline fits in order to minimize the risk of learning from statistical outliers. Discrete input variables can be specified and treated accordingly as ordered or unordered classes.

One key aspect of the NeuroBayes package is the determination of the statistical relevance of individual network connections or entire nodes. Statistically insignificant connections or nodes are removed during the Neural Network training making it robust against statistical fluctuations in the training data.

It should be mentioned, even though it is not explicitly used in the following, that the NeuroBayes output may be interpreted as a Bayesian a posteriori signal probability, if the numbers of signal and background events used during the Neural Network training correspond to the a priori signal and background probabilities.

6. First Measurement of the Jet Energy Resolution at CMS

The precise knowledge of the jet p_{T} resolution is an important ingredient to any analysis with jets in the final state. It is required to relate the quantities measured in the detector to the true properties of the underlying processes. Consequently, the size of the jet p_{T} resolution is also of great importance for the mass resolution of the jet-based Higgs boson candidates in the search presented in Chapter 7.

This chapter summarizes the first measurement of the jet p_{T} resolution at CMS applying the data-driven dijet asymmetry method to the full 2010 pp dataset.

In the end of this chapter, two exemplary results of 2010 jet physics measurements are given which make heavy use of the jet p_{T} resolution: the measurement of the inclusive jet and b-jet cross sections, which were performed in close collaboration with the jet p_{T} resolution measurement.

6.1. Selection of Events

This section gives an overview of the samples of simulated events and primary datasets of pp collision data used in the resolution measurement. In addition, the construction of the data sample using dijet average triggers, as well as the event selection are described. Additional information and technical details on the event selection are given in Appendix A.

6.1.1. QCD MC Samples

The jet p_{T} resolution measurement uses samples of QCD events which have been simulated with Pythia 6.4 and Geant 4 as part of the central `Fall10` MC production campaign. The samples are generated using the CMS default MC tune Z2, which is a re-tune of Z1 [131] with CTEQ6L1 [25] as PDF (for "tune" see also Section 4.1.1). All samples are listed in the appendix in Table A.1 together with the number of events and cross sections.

In order to obtain a roughly constant number of events over the whole phase space, the samples are sliced into mutually exclusive $\hat{p}_{\mathrm{T}}$ (leading parton momentum) bins during the event generation. The full p_{T} spectrum of the simulated events is then reassembled by weighting all events with the fraction of the cross section and the total number of events in each exclusive sample respectively and adding them together. The resulting spectrum is shown in Figure 6.1.

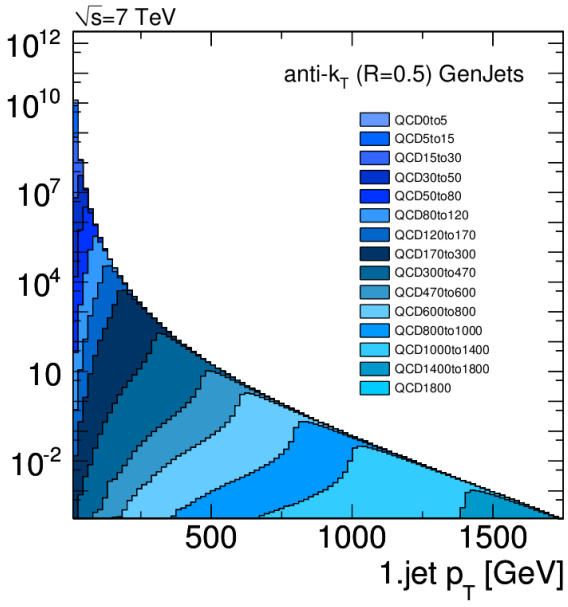


Figure 6.1.: Spectrum of the leading generator jet p_{T} reconstructed from all QCD samples listed in Table A.1 after weighting each event according to the cross section and the total number of events in the sample.

6.1.2. 2010 Jet Data

Data from the `JetMET`, `JetMETTau` and `Jet` primary datasets (PDs) is merged to form the full recorded 2010 LHC jet dataset, as summarized in the appendix in Table A.2. Primary datasets refer to collections of data with related triggers. The official CMS list of certified luminosity blocks for each run is considered to select only events which have been validated by the data quality management teams. The total integrated luminosity of the data after this good run selection amounts to $35.9\,\mathrm{pb}^{-1}$.

6.1.3. Dijet Average Trigger

From this entire available collision data sample all events are considered which fire one of the `HLT_DiJetAveXXU` dijet average triggers. The post-script `XXU` indicates the existence of several of such triggers with different uncorrected dijet average transverse momentum thresholds. Table 6.1 lists all triggers used in this analysis. The dijet triggers select events based on the average dijet p_{T},

$$p_{\mathrm{T}}^{\mathrm{ave}} \equiv \frac{1}{2}\left(p_{\mathrm{T}}^{\mathrm{jet}\,1} + p_{\mathrm{T}}^{\mathrm{jet}\,2}\right),$$

of the two leading uncorrected jets reconstructed by the High Level Trigger (HLT).

Over the course of 2010 the different dijet triggers have been heavily prescaled in order to cope with the rapid rise in the instantaneous luminosity delivered by

Table 6.1.: Trigger turn-ons for the uncorrected dijet average triggers with respect to $p_{\mathrm{T}}^{\mathrm{ave}}$ for CaloJets, JPTJets and PFJets, taken from [132]. The highest turn-on in each row was used to define common bin boundaries for all three jet types. As calorimeter jets are used in the 2010 HLT menus, higher turn-on points are found for the track assisted jet reconstruction types.

trigger	CaloJets	JPTJets	PFJets	Max
HLT_MinBias	-	-	-	
HLT_DiJetAve15U	38 GeV	43 GeV	43 GeV	43 GeV
HLT_DiJetAve30U	59 GeV	66 GeV	70 GeV	70 GeV
HLT_DiJetAve50U	86 GeV	96 GeV	100 GeV	100 GeV
HLT_DiJetAve70U	111 GeV	124 GeV	127 GeV	127 GeV
HLT_DiJetAve100U	147 GeV	165 GeV	168 GeV	168 GeV
HLT_DiJetAve140U	196 GeV	220 GeV	214 GeV	220 GeV

the LHC. In order to retain the maximum number of events for the resolution measurement the data is divided into mutually exclusive $p_{\mathrm{T}}^{\mathrm{ave}}$ bins. Each of the bins is defined by one of the dijet triggers and a corresponding $p_{\mathrm{T}}^{\mathrm{ave}}$ range for which the particular trigger is fully efficient, according to the measurements of the trigger turn-ons. A trigger turn-on is defined as the lowest value of $p_{\mathrm{T}}^{\mathrm{ave}}$ for which the trigger efficiency is above 99 %. The trigger efficiency is evaluated dividing the number of jets passing the trigger and a suitable reference trigger as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$. As a reference trigger for HLT_DiJetAve15U the minimum bias high level trigger HLT_MinBias is chosen. For HLT_DiJetAve30U and higher the next lower dijet trigger, e. g. HLT_DiJetAve15U, is used as a reference trigger respectively. Table 6.1 lists the trigger turn-ons and the corresponding $p_{\mathrm{T}}^{\mathrm{ave}}$ lower boundaries for all three triggers. The upper bin boundaries are given by the lower boundary of the next highest trigger respectively.

6.1.4. Event Selection

For both data and simulation samples, the presence of a well reconstructed primary vertex (PV) with $|z(\mathrm{PV})| < 24.0\,\mathrm{cm}$ and three or more associated tracks is required. The radial distance of the primary vertex has to lie within the beam pipe $\rho(\mathrm{PV}) < 2.0\,\mathrm{cm}$. In the case of collision data, only certified events are considered. Events are selected if they contain at least two jets ("dijet events") which are back-to-back in azimuth, $|\Delta\phi| > 2.7$. To further enhance the dijet event purity, the leading jet in each event is required to fulfill a set of requirements on basic jet reconstruction properties referred to as "loose jet ID" requirements [133, 134]. They are used to suppress unphysical jets or jets from electronic noise which may mimic the dijet event topology in combination with the presence of a physical jet. Events are rejected if $\not{E}_{\mathrm{T}}/\sum E_{\mathrm{T}} < 0.5$ and $\not{E}_{\mathrm{T}} > 100$ GeV to remove effects from anomalous calorimeter noise, beam halo or cosmic-ray backgrounds mimicking real jets.

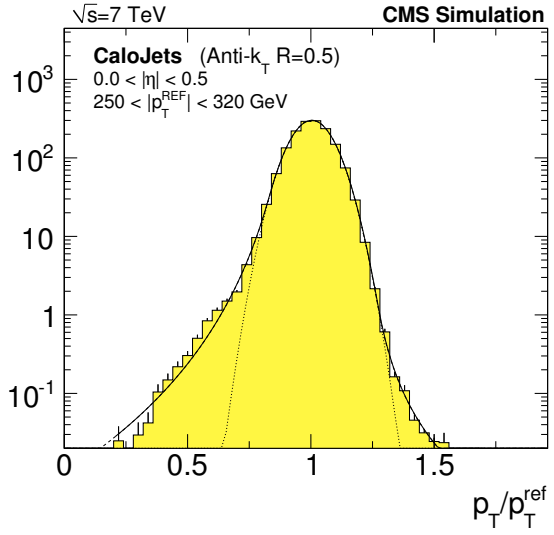


Figure 6.2.: Illustration of a jet response distribution [114]. The core of the distribution is fitted iteratively with a Gauß in this example, the tails are described by a double-sided Crystal ball function.

6.1.5. Jet Energy Correction

The official L2 and L3 jet energy corrections for the 2010 dataset based on Fall10 MC truth have been applied to both MC and data. The data has been additionally corrected for a residual jet energy scale bias. As the number of pile-up interactions in the 2010 dataset is small, no L1 pile-up correction is applied.

6.2. Jet Resolution in MC Truth

The overall strategy of the CMS jet resolution determination is based on truth information of simulated events. Their derivation was part of the work for this thesis [135] and is summarized in this Section. The resolution expectation from MC is then compared to a measurement using the data-driven asymmetry method, which is described in the next Section, and corrected for residual differences.

The anti-k$_\mathrm{T}$ jet clustering algorithm with distance parameter $R = 0.5$ is applied to all stable generator particles excluding neutrinos to form generator jets. They are referred to as reference jets in the following, with $p_\mathrm{T}^\mathrm{ref}$ denoting their transverse momenta. The reference jets are then unambiguously matched to the different types of reconstructed jets (CaloJets, JPT-Jets or PFJets) using the Euclidean distance in the η-ϕ-plane of the detector (see also Section 3.3.1):

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < \Delta R_\mathrm{max}.$$

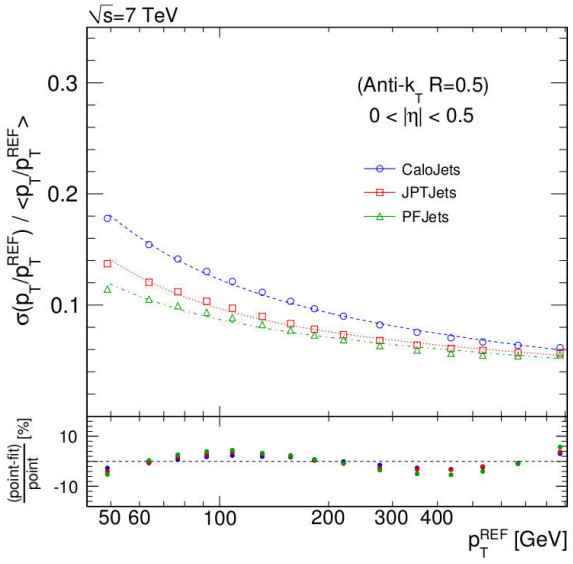


Figure 6.3.: Illustration of the jet p_{T} resolution from MC truth as a function of $p_{\mathrm{T}}^{\mathrm{ref}}$, taken from [132]. The jet resolution is shown for Calo-, JPT- and PFJets in the very central detector region with $|\eta| < 0.5$.

The matching parameter ΔR_{max} is chosen as $\Delta R_{\mathrm{max}} = 0.25$ for all three jet types following studies of its impact on the resolution. For this value the resolution is stable against small variations of ΔR_{max}. In addition, only the two pairs with the highest transverse momenta reference jets are considered. The jet response $\mathcal{R}$ is then defined for each jet pair as

$$\mathcal{R} = \frac{p_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{ref}}}.$$

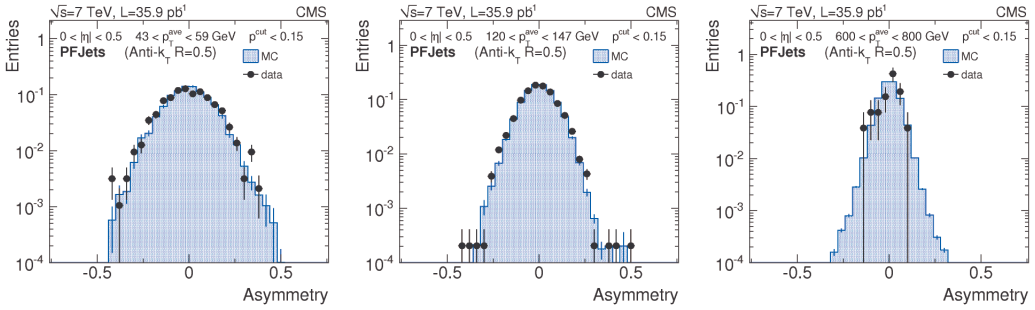
An example of such a jet response distribution in a specific part of the phase space is shown in Figure 6.2. The width of this distribution is interpreted as the jet p_{T} resolution and quantified using its standard deviation,

$$\sigma_{\mathcal{R}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathcal{R}_i - \langle \mathcal{R} \rangle\right)^2.$$

Figure 6.3 shows the relative MC truth jet resolutions of Calo-, JPT- and PFJets as a function of $p_{\mathrm{T}}^{\mathrm{ref}}$. All three can be described by a fit to a variation of the standard NSC formula [64],

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \sqrt{\mathrm{sgn}(N) \cdot \left(\frac{N}{p_{\mathrm{T}}}\right)^2 + S^2 \cdot p_{\mathrm{T}}^{(m-1)} + C^2}, \tag{6.1}$$

where N refers to the "noise", S to the "stochastic" and C to the "constant" term. These terms describe the response of the calorimeters, as previously explained in



(a) (b) (c)

Figure 6.4.: Examples of the dijet asymmetry distribution in the central detector region $|\eta| < 0.5$ for (a) low, (b) medium and (c) high dijet average p_{T} bins [114]. Note the decreasing width of the distributions with higher transverse momentum.

Section 3.3.3. The additional parameter m is introduced and the negative sign for the noise S is allowed [135] to significantly improve the fits at low p_{T} for the jet types which include tracking information, while retaining a similar functional form as the classic one used for calorimeter jets. For these jet types the classic NSC description is not sufficient anymore. The tracker has a better momentum resolution with respect to the calorimeters and the combination of the two in the track assisted jet reconstruction types necessitates a modified p_{T}-dependence.

6.3. Dijet Asymmetry Method

The asymmetry method is a data-driven technique to measure the jet p_{T} resolution by exploiting momentum conversation in the transverse plane of dijet events. It has been developed at the DØ experiment at the Tevatron [136] and was first established within the CMS collaboration for calorimeter jets based on simulation studies [137].

In the summer of 2010, the asymmetry method has been successfully applied for the first time to LHC collision data from the CMS commissioning era [112, 138] as part of the work for this thesis and then updated to the full 2010 dataset [114, 132]. The results of this update are presented in the following.

Overview

The asymmetry A of events with at least two jets is defined as

$$A \equiv \frac{(p_{\mathrm{T}}^{\mathrm{jet}\,1} - p_{\mathrm{T}}^{\mathrm{jet}\,2})}{(p_{\mathrm{T}}^{\mathrm{jet}\,1} + p_{\mathrm{T}}^{\mathrm{jet}\,2})},$$

where $p_{\mathrm{T}}^{\mathrm{jet}\,1}$ and $p_{\mathrm{T}}^{\mathrm{jet}\,2}$ refer to the randomly ordered transverse momenta of the two leading jets. The variance of the asymmetry, σ_A, can be expressed as

$$\sigma_A^2 = \left|\frac{\partial A}{\partial p_{\mathrm{T}}^{\mathrm{jet}\,1}}\right|^2 \cdot \sigma^2(p_{\mathrm{T}}^{\mathrm{jet}\,1}) + \left|\frac{\partial A}{\partial p_{\mathrm{T}}^{\mathrm{jet}\,2}}\right|^2 \cdot \sigma^2(p_{\mathrm{T}}^{\mathrm{jet}\,2}).$$

Assuming $p_{\mathrm{T}} \equiv \left\langle p_{\mathrm{T}}^{\mathrm{jet}\,1} \right\rangle = \left\langle p_{\mathrm{T}}^{\mathrm{jet}\,2} \right\rangle$ and $\sigma(p_{\mathrm{T}}) \equiv \sigma(p_{\mathrm{T}}^{\mathrm{jet}\,2}) = \sigma(p_{\mathrm{T}}^{\mathrm{jet}\,2})$, the fractional jet p_{T} resolution can be expressed as

$$\left(\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right) = \sqrt{2}\sigma_A \ . \tag{6.2}$$

This equation implies an idealized topology of two jets with exactly compensating transverse momenta which is spoiled in realistic collision events by the presence of extra activity, e.g. from additional soft radiation or the underlying event. Consequently, the resulting asymmetry distributions are broadened, and the width of the asymmetry distributions systematically overestimates the true jet resolution. But even in the ideal situation of only two particle jets and no soft radiation the two jets are not necessarily balanced. Fragmentation effects cause some energy to be showered outside the jet cone ("out of cone radiation"), which results in a p_{T} imbalance at particle level and broadens the jet resolution. The width of the asymmetry distribution is thus a convolution of these different contributions

$$\sigma_A = \sigma_{\mathrm{intrinsic}} \oplus \sigma_{\mathrm{imbalance}}.$$

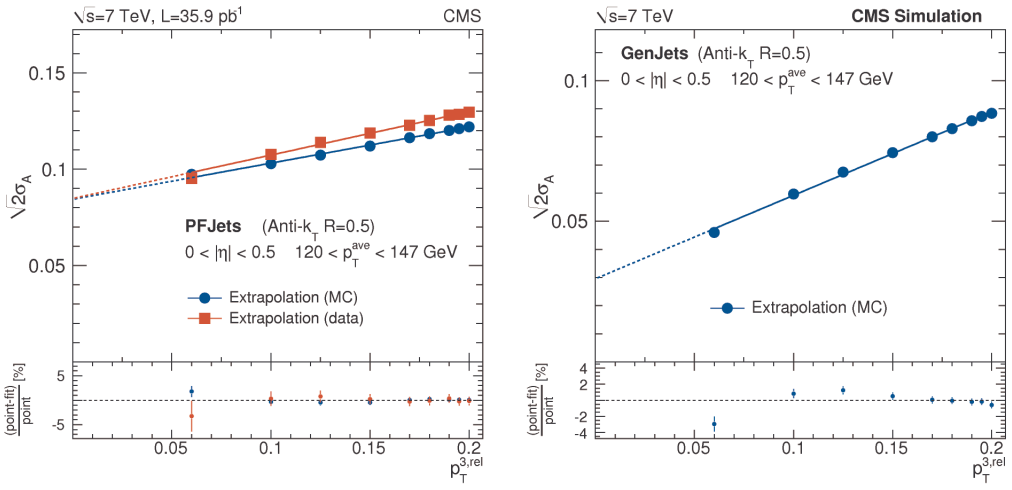
Several corrections are applied to remediate these different effects.

Soft Radiation Correction

To account for soft radiation in dijet events, the asymmetry distributions of the sample are evaluated multiple times for decreasing amounts of extra activity in each η and $p_{\mathrm{T}}^{\mathrm{ave}} \equiv (p_{\mathrm{T}}^{\mathrm{jet}\,1} + p_{\mathrm{T}}^{\mathrm{jet}\,2})/2$ bin, and the jet resolution is extracted by extrapolating to zero. The amount of additional event activity is hereby quantified with the relative momentum of the third jet in the event with respect to the average dijet momentum,

$$p_{\mathrm{T}}^{3,\,\mathrm{rel.}} \equiv \frac{p_{\mathrm{T}}^{\mathrm{jet}\,3}}{p_{\mathrm{T}}^{\mathrm{ave}}}.$$

Figure 6.5 illustrates this extrapolation procedure for PFJets and generator particle jets. Points corresponding to a certain $p_{\mathrm{T}}^{3,\,\mathrm{rel.}}$ only go into the extrapolation if their corresponding absolute $p_{\mathrm{T}}^{\mathrm{jet}\,3}$ lies in an interval of (6 GeV, 150 GeV) depending on the respective $p_{\mathrm{T}}^{\mathrm{ave}}$ bin. The extrapolation is observed to be in very good agreement with the measured points for all three algorithms.



(a) (b)

Figure 6.5.: Illustration of the extra event activity extrapolation for PFJets and generator jets ("GenJets"), taken from [114]. The extrapolation to zero extra activity corrects for additional radiation in the event spoiling the dijet balance.

Further improvement of the soft radiation correction is achieved by using a smoothing technique to mitigate the impact of statistical outliers. The "slope" of the extrapolation, defined as

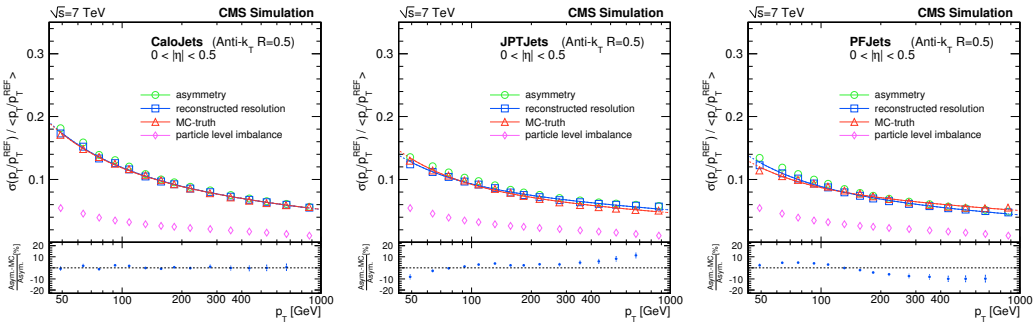
$$\mathrm{k_{soft}} \equiv \frac{\sigma(p_\mathrm{T}^{3,\,\mathrm{rel.}} \to 0)}{\sigma(p_\mathrm{T}^{3,\,\mathrm{rel.}} = 0.15)},$$

is calculated and plotted against the mean $p_\mathrm{T}^\mathrm{ave}$. The mean $p_\mathrm{T}^\mathrm{ave}$ values are defined by the mean of the the $p_\mathrm{T}^\mathrm{ave}$ distributions in the respective $p_\mathrm{T}^\mathrm{ave}$- and $|\eta|$-bins. In the formula, $\sigma(p_\mathrm{T}^{3,\,\mathrm{rel.}} \to 0)$ denotes the extrapolated width at zero extra event activity and $\sigma(p_\mathrm{T}^{3,\,\mathrm{rel.}} = 0.15)$ the width at a stable working point which is accessible in low and high $p_\mathrm{T}^\mathrm{ave}$ bins. The resulting graph is then fitted with a logarithmic polynomial

$$\mathrm{k_{soft}}(p_\mathrm{T}) = p_0 + p_1 \cdot \log(0.01 \cdot p_\mathrm{T}) + p_2 \cdot \log^2(0.01 \cdot p_\mathrm{T}) + p_3 \cdot \log^3(0.01 \cdot p_\mathrm{T})$$

to smooth fluctuations in the bin by bin extrapolation described above and illustrated in Figure 6.5. In this empirically motivated $\mathrm{k_{soft}}$-fit, the parameter p_3 is fixed to zero for reconstructed jets, but improves the description of the slightly different $\mathrm{k_{soft}}$ shape in the case of generator jets. Further details on this are given in Ref. [132].

The corrected asymmetry width is finally obtained by an evaluation of the fit at a given mean $p_\mathrm{T}^\mathrm{ave}$ value (described above) and multiplication with the width of the corresponding stable working point.



(a) (b) (c)

Figure 6.6.: Application of the asymmetry method to simulated dijet events [114]. For (a) CaloJets, (b) JPTJets and (c) PFJets the extrapolated reconstructed (green) and particle level (magenta) asymmetries are shown. The final jet p_{T} resolution (blue) is then compared to the MC truth expectation (red).

Particle Level Imbalance Correction

Physics effects such as parton showering and hadronization, where particles are emitted outside the clustered particle jet lead to a p_{T} imbalance at particle level. To account for its contribution to the measured jet p_{T} resolution, the asymmetry method is applied to generator jets with the same distance parameter ($R = 0.5$) which are reconstructed from stable MC particles (indicated with "REC"). The observed particle level resolution (indicated with "GEN") is subtracted in quadrature:

$$\left(\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right)^2 = \left(\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right)^2_{\mathrm{REC}} - \left(\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}\right)^2_{\mathrm{GEN}} . \tag{6.3}$$

Resolution from the Asymmetry Method

Figure 6.6 illustrates the different steps of the asymmetry procedure for (a) CaloJets, (b) JPTJets and (c) PFJets respectively. The raw asymmetry derived from the extrapolation of the reconstructed asymmetry is shown in green, the estimation of the intrinsic resolution from the application to generator jets is shown in magenta, and the combination of the two to the final asymmetry result is shown in blue. The jet p_{T} resolution is fitted with the modified NSC formula (6.1) and compared to the resolution from MC truth.

The Asymmetry Method at higher Pseudorapidities

For two jets in different pseudorapidity regions the resolution is different because the relevant sub-detectors are not the same. However, it is possible to calculate the resolution of one jet if the resolution of the other jet is already known [139]. From

(6.2) one derives the relation for the second jet p_{T} resolution to be

$$\frac{\sigma(p_{\mathrm{T}}^{\mathrm{jet}\,2})}{p_{\mathrm{T}}} = \sqrt{2 \cdot \sigma_{\mathrm{A}_{1,2}}^2 - \left(\frac{\sigma(p_{\mathrm{T}}^{\mathrm{jet}\,1})}{p_{\mathrm{T}}}\right)^2},$$

where the notation $\sigma_{\mathrm{A}_{1,2}}$ reflects the fact that only one of the two leading jets lies within the central bin. One can thus measure the asymmetry in events with one jet in the central and one in the $|\eta|$-bin under study, instead of requiring both jets to be in the forward $|\eta|$-bin. This would yield fewer events and therefore a less precise measurement.

6.4. Systematic Uncertainties

There are several dominating sources of systematic uncertainty inherent to the dijet asymmetry method. They are described in detail in the following:

Systematic Uncertainty of the Soft Radiation Correction

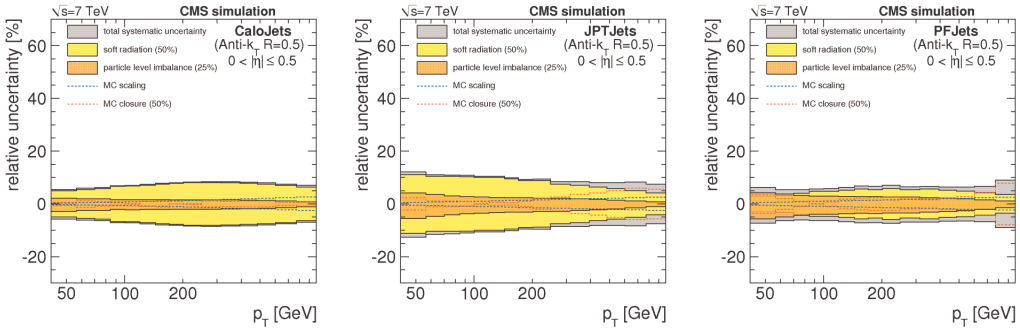
The linear fits of the extrapolation to zero are the basis for the correction to the asymmetry due to extra event activity. Even though the fit residuals suggest very good agreement for all jet types and p_{T}- and η-ranges, the linear extrapolation assumes a certain behavior for very low event activity which cannot be measured. The linear extrapolation is thus evaluated at half-the-distance between the stable working point at $p_{\mathrm{T}}^{3,\,\mathrm{rel.}} = 0.15$ and the difference to the full extrapolation to zero is assigned as systematic uncertainty.

Systematic Uncertainty of the Particle Level Imbalance

The intrinsic particle level resolution due to out-of-cone showering and fragmentation effects can only be studied with simulated events and is therefore subject to potentially large systematic uncertainties. The scale of the particle level imbalance is varied by $\pm 25\%$. The impact of this variation is expressed by subtracting 75% and 125% of the nominal particle jet resolution given in (6.3) in quadrature.

Systematic Uncertainty of the MC Closure

The dijet asymmetry method is capable of measuring the true jet p_{T} resolution, as shown by the above comparisons to the evaluation based on MC truth information. Half of the residual deviations between asymmetry and MC truth are taken as a systematic uncertainty, while no bias correction is derived based on the good level of agreement.



(a) (b) (c)

Figure 6.7.: Relative systematic uncertainties for (a) CaloJets, (b) JPTJets and (c) PFJets in the central detector region [114]. The particle level imbalance uncertainty is shown in opaque orange, the solid yellow contribution corresponds to the uncertainty from the soft radiation variation and the dashed red line depicts the impact from the remaining differences in the MC closure. The dashed blue line visualizes the contribution of the uncertainty on the additional constant term which is added to the MC resolution. The grand total, calculated as the square sum of the different contributions, is given in grey. Note that while the relative uncertainty for the different jet types is rather similar, the absolute size decreases with better resolution (e. g. PFJets).

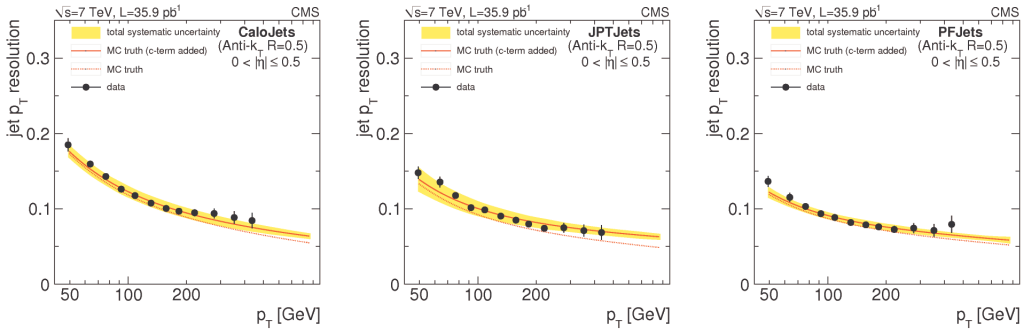
Systematic Uncertainty due to MC Resolution Scaling

By comparing the resolution measured from the asymmetry method in data with the resolution measured in MC simulations an additional constant term is fitted to describe the difference observed between data and simulation (Section 6.5). The MC resolution is scaled to correct for this difference. The statistical uncertainty on an additional constant term which we fit in data is taken and being treated as a systematic uncertainty of the scaled MC resolution.

Overview of the uncertainties

Figure 6.7 shows the size of the different systematic uncertainties as a function of $p_{\mathrm{T}}^{\mathrm{ave}}$. While the contributions of the different sources of systematic uncertainty to the grand total vary with p_{T} and $|\eta|$, the overall relative systematic uncertainty is found to be within 10% in the central detector region $0.0 < |\eta| < 1.5$. It increases to 15% for $1.5 < |\eta| < 2.0$, 25% for $2.0 < |\eta| < 2.5$ and 30% for $2.5 < |\eta| < 3.0$.

The influence of the presence of additional collisions due to pile-up has been assessed by dividing the data sample into subsamples with exactly one reconstructed primary vertex and more than one primary vertices. No significant deviations with respect to the inclusive measurement are observed in the 2010 dataset.



(a) (b) (c)

Figure 6.8.: Bias-corrected data measurements of the jet p_{T} resolution in the central detector region ($|\eta| < 0.5$) compared to MC truth before (red dashed line) and after correction for the measured discrepancy between data and simulation (red solid line) [114] for (a) CaloJets, (b) JPTJets and (c) PFJets.

6.5. Results

The full asymmetry method as described above is applied to both data and MC simulation and a systematically poorer resolution is found in data compared to the simulation. Figures 6.8(c) and 6.9(a)-6.9(f) show the measured resolutions for PFJets in different bins of η.

MC/Data Resolution Discrepancy

The observed discrepancy in data and MC simulation resolution has been studied in great detail as a function of p_T and $|\eta|$. Its likely source is an imperfect HCAL-intercalibration [114]. While the discrepancy is expected to be reduced by a re-calibration of the calorimeters in the future, the simulation is manually corrected for the observed discrepancy by analysts in the meantime. Studies showed that the MC/data discrepancy can be quantified by fitting an additional constant term from data. To obtain this term, the resolution is fitted in MC simulation using the modified NSC formula (6.1) in a first step. In a second step all of the parameters are fixed to the MC value and an additional free parameter corresponding to the constant term is added to the function. The additional constant term is then fitted and added to the MC truth parameterization to correspond to the data. Figures 6.8(c) and 6.9(a)-6.9(f) show the improvement obtained when comparing the agreement of bias-corrected data measurements to MC truth with an additional constant term and to the MC truth without this term.

Resolution Scaling

In order for the jets in simulation to resemble the resolution measured in data, the measured data/MC offset described in the previous subsection is used as part

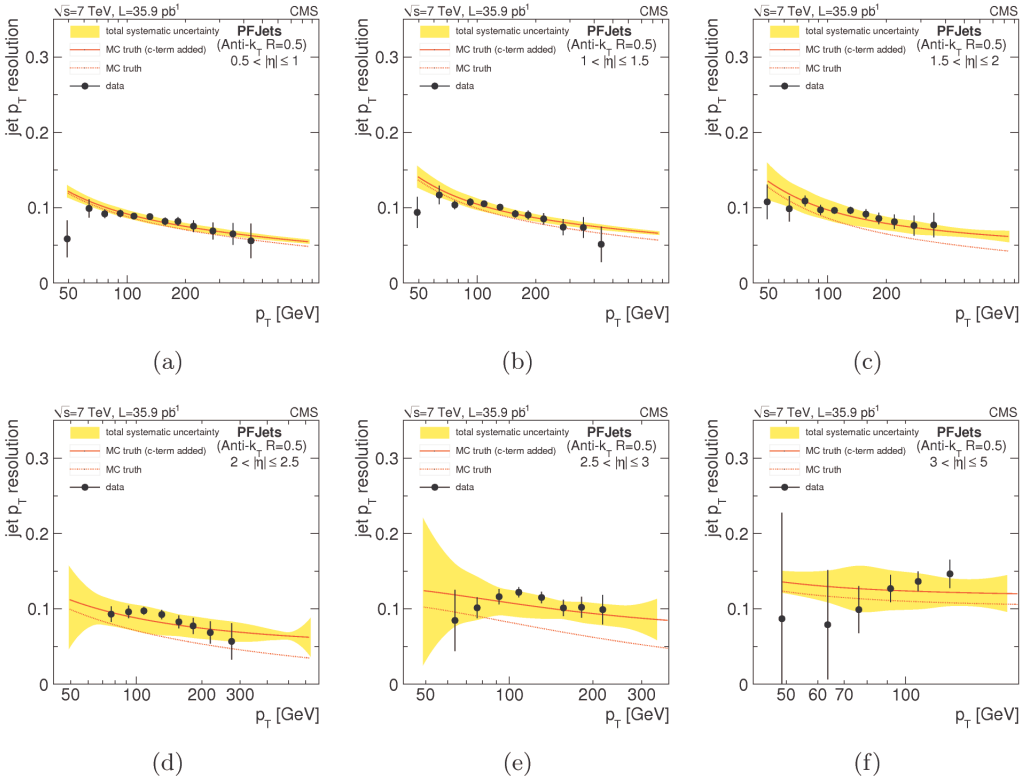


Figure 6.9.: Bias-corrected data measurements of the PFJet resolution in various forward detector regions compared to MC truth expectation. [114]

of the following scaling prescription. For each reconstructed jet the corresponding generator jet is matched unambiguously using the closest ΔR. The p_{T} difference of the two is then multiplied with the data/MC offset factor and added relative to the reconstructed jet p_{T} to correct for the difference.

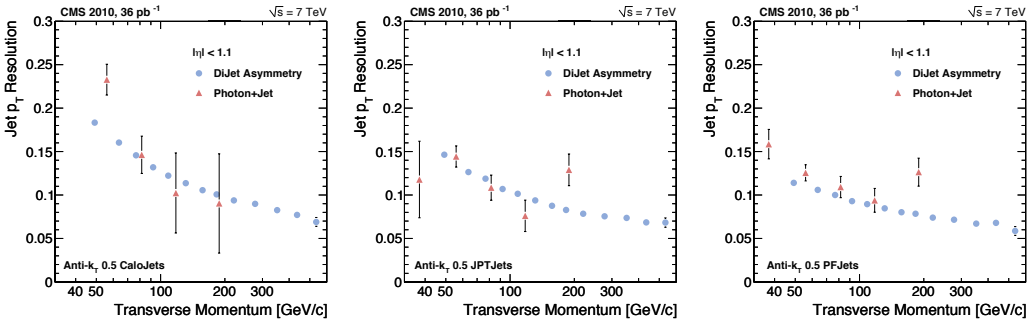
In contrast to other correction approaches, such as Gaussian smearing, the scaling approach has the advantage that it is deterministic and can be used to adjust the jet resolution in simulation to be both worse and better. The number of unmatched jets is usually very small and hey are not treated within the scaling approach.

This prescription is also used for the determination of jet related systematic uncertainties, as e. g. described in the Higgs boson search in Section 7.5.9.

Comparison with the γ+Jets Measurement

An alternative measurement of the jet p_{T} resolution has been performed using γ+jets events [114]. This measurement is expected to be sensitive at very low jet transverse momenta where dijet events cannot be easily triggered and is used to validate the dijet measurement at medium jet p_{T}'s where both methods overlap.

Instead of exploiting the p_{T} balance of two jets, this methods benefits from



(a) (b) (c)

Figure 6.10.: Comparison of the jet p_{T} resolution measured in dijet and γ+jets events in the central detector region ($|\eta| < 1.1$) [114] for (a) CaloJets, (b) JPTJets and (c) PFJets.

the excellent response of the electromagnetic calorimeter to measure photons in events where one photon is produced in association with one or more jets. In analogy to the dijet asymmetry method the γ+jets method uses an extrapolation to zero extra event activity to correct for radiation in the event spoiling the ideal γ+jet balance. Figure 6.10 shows the comparison of the measurements using the dijet asymmetry and the γ+jets in the central detector region for Calo-, JPT- and PFJets. The two methods yield consistent results, even though the statistical uncertainty on the γ+jets measurement is still large in comparison to the dijet asymmetry measurement in $\mathcal{L} = 36\,\mathrm{pb}^{-1}$.

6.6. Measurement of the Inclusive (b-) Jet Cross Section

The precise knowledge of the jet p_{T} resolution is important for any analysis with jets in the final state. In the following, two results of analyses which use the resolution as a key ingredient are presented, the measurements of the inclusive jet and b-jet cross sections. These analyses were carried out in collaboration with the jet p_{T} resolution measurement and extended beyond to further contributions and cross checks, which are, however, beyond the scope of this thesis.

Nevertheless, the general concepts and main results are summarized in the following for completeness. Further details are found in the References [140–142].

Measurement of the Inclusive Jet Cross Section

The measurement of the inclusive jet cross section [143] is an important test of the standard model at hadron colliders [144, 145]. Deviations from theoretical predictions using perturbative quantum chromodynamics could indicate new phenomena beyond the standard model, and the measurement is also very sensitive to the jet energy scale and resolution and is thus an excellent cross check of both [146].

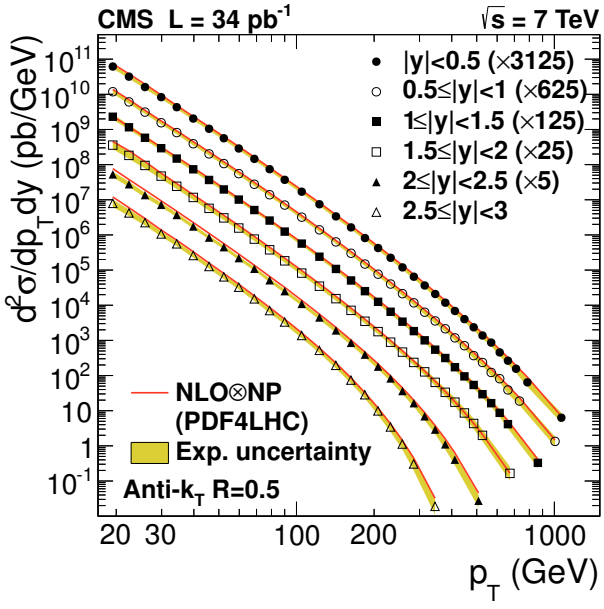


Figure 6.11.: Inclusive jet cross section in 6 bins of rapidity $|y|$ compared to next-to-leading order pQCD theory predictions [143]. The jet p_{T} resolution is a key ingredient to the unfolding needed to compare the detector measurement to theory.

Measurements at the Tevatron [147–149] and from ATLAS [150] indicate agreement with theory predictions for jet transverse momenta of up to 700 GeV.

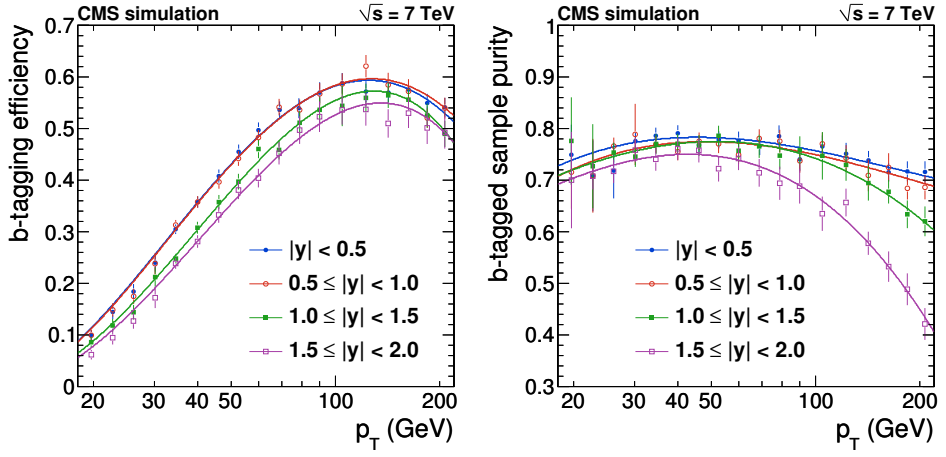
The double-differential jet production cross section is defined as

$$\frac{d^2\sigma_{\mathrm{jet}}}{dp_{\mathrm{T}}dy} = \frac{N_{\mathrm{jet}}}{\Delta p_{\mathrm{T}}\Delta y \cdot \epsilon_{\mathrm{jet}} \cdot \mathcal{L}}, \tag{6.4}$$

where y denotes the rapidity (3.1), N_{jet} is the number of jets per bin, Δp_{T} and Δy are the bin widths of p_{T} and y respectively, $\mathcal{L}$ stands for the total integrated luminosity and ϵ_{jet} is a factor summarizing the event selection and jet reconstruction efficiencies.

One particular challenge of this measurement is to account for jet p_{T} resolution effects in the steeply falling jet p_{T} spectrum (see also Figure 6.1). Due to resolution, reconstructed jets migrate into p_{T} bins different to the underlying true particle p_{T}. Significantly more migrations take place from lower to higher p_{T} bins due to the steeply falling nature of the underlying spectrum, and the measured distribution is therefore distorted and must be corrected accordingly. This is done using an Ansatz method. The true particle jet p_{T} spectrum is modelled by a phenomenological power-law parameterization motivated by the parton model [143]:

$$f(p_{\mathrm{T}};\alpha,\beta,\gamma) = N_0 \cdot p_{\mathrm{T}}^{-\alpha} \cdot \left(1 - \frac{1}{\sqrt{s}} 2p_{\mathrm{T}} \cosh(y_{\mathrm{min}})\right)^{\beta} \cdot \exp^{-\frac{\gamma}{p_{\mathrm{T}}}} \tag{6.5}$$



(a) (b)

Figure 6.12.: Continuous parameterization of the b-tagging efficiency (a) and purity (b) [151] used for the inclusive b-jet cross section measurement in different rapidity bins $|y|$. The parameterization have been cross-checked using data-driven techniques.

In this parameterization N_0 is a normalization factor, y_{min} is the low-edge of the rapidity bin y and $\alpha\,,\beta\,,\gamma$ are fit parameters. The $\alpha-$parameter describes the overall p_{T} evolution, the $\beta-$term a kinematic cutoff at high jet p_{T} and γ fits the low p_{T} saturation of the spectrum. This parameterization is smeared as

$$F(p_{\text{T}}^{\text{meas}}) = \int_0^\infty f(p_{\text{T}}^{\text{ref}}, y; \alpha, \beta, \gamma) g(p_{\text{T}}^{\text{meas}} - p_{\text{T}}^{\text{ref}}, y; \sigma) dp_{\text{T}}^{\text{ref}},$$

where $f(p_{\text{T}}^{\text{ref}}, y; \alpha, \beta, \gamma)$ is the unsmeared Ansatz (6.5) and $g(p_{\text{T}}, y; \sigma)$ is an area-normalized Gaussian centered around zero with a width corresponding to the jet p_{T} resolution $\sigma = (\sigma_{p_{\text{T}}}/p_{\text{T}}) \cdot p_{\text{T}}^{\text{ref}}$. In this expression, the resolution is corrected for the data/MC discrepancy as described in the previous Sections before being fitted to the experimental data. The p_{T} resolution correction C_{smear} is then determined bin-by-bin by taking the ratio of the function (6.5) to the smeared function

$$C_{\text{smear}} = \frac{f(p_{\text{T}})}{F(p_{\text{T}})}$$

and applied to the measured data.

Figure 6.11 shows the fully corrected inclusive jet cross sections for $\mathcal{L} = 34\,\text{pb}^{-1}$ as a function of p_{T} between 18 and 1100 GeV for several bins of rapidity y. Within the experimental and theoretical uncertainties the measurement is found to be consistent with the pQCD predictions.

Measurement of the Inclusive b-Jet Cross Section

The measurement of the inclusive jet cross section described above has been applied to a smaller subset of the data ($\mathcal{L} = 60\,\mathrm{nb}^{-1}$) with a particular focus on jets arising from the hadronization of b-quarks, so called b-jets [151]. As already mentioned in Section 4.3.7, the identification of b-jets is often based on the long B-hadron lifetime within the b-jet and the reconstruction of a secondary vertex associated with its decay. In this measurement, b-jets are identified using the secondary vertex high-purity tagger (SSVHP).

A good understanding of the $b\bar{b}$-production rate and the underlying reconstruction properties of b-jets are desirable pre-requisites for the search of a Higgs boson decaying into two b-quarks, which is described in the following Chapter 7.

The efficiencies and purities of the SSVHP b-tagger have to be taken into account when, in analogy to (6.4), defining the double-differential b-jet production cross section

$$\frac{d^2\sigma_{b-\mathrm{jet}}}{dp_\mathrm{T}dy} = \frac{N_\mathrm{jet} \cdot f_\mathrm{b}}{\Delta p_\mathrm{T} \Delta y \cdot \epsilon_\mathrm{b} \cdot \epsilon_\mathrm{jet} \cdot \mathcal{L}}, \tag{6.6}$$

where ϵ_b and f_b are the b-tagging efficiency and purity respectively. Figure 6.12 show the continuous parameterization of the b-tagging efficiency and purity derived from simulated MC events for the SSVHP as functions of p_T and according to the $|y|$-binning of the cross section measurement.

The jet p_T resolution correction is done similar to the inclusive cross section. Figure 6.13 shows the fully corrected b-jet cross sections as a function of p_T in 4 bins of rapidity. Reasonable agreement is found between data and next-to-leading order theory prediction within the uncertainties.

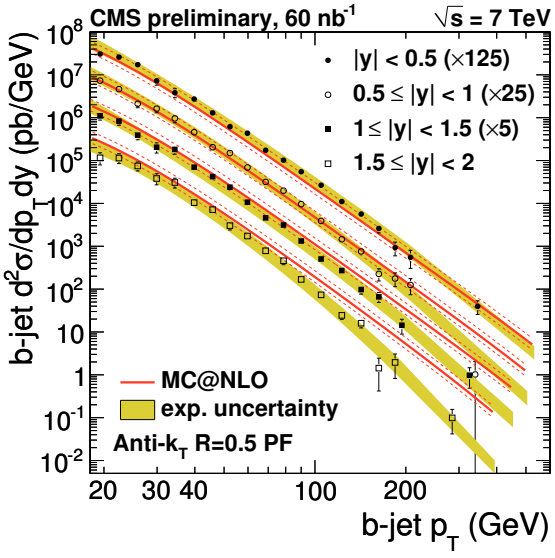


Figure 6.13.: Inclusive b-jet cross section in four bins of rapidity $|y|$ [151].

7. Search for a light Standard Model Higgs Boson

The search for a Standard Model Higgs boson in the associated $H(b\bar{b})W(\ell\nu)$ channel is challenging due to the large background contributions, but attracts attention because it is one of only a few channels which are sensitive to a light Higgs boson in the mass range of 110 GeV to 135 GeV.

This chapter starts with an introduction to the HW analysis strategy, followed by descriptions of the samples, triggers and objects used throughout the search.

The analysis is then presented both using a cut-based approach and a Neural Network approach. The muon and electron channels are analyzed separately, and their combination is presented as well.

The entire set of analyses is then performed using the Fat-, Sub- and Filterjet (SJF) algorithm originally proposed for the boosted Higgs search as introduced in Section 4.3.7 and the results are presented and compared to the standard reconstruction.

7.1. Boosted HW Analysis Strategy

A light Standard Model Higgs boson is predicted to predominantly decay into a pair of b-quarks, as was shown in Figure 2.8. Due to the large inclusive b-jet cross section (see Section 6.6), however, the dominant $gg \to H \to b\bar{b}$ production channel is impossible to trigger and detect. One potential solution to this problem, which was proposed a long time ago, is to search for $H \to b\bar{b}$ decays in associated VH Higgsstrahlung production. Despite the lower production cross section, Higgsstrahlung production has an important advantage: in addition to the two b-jets from the Higgs boson, events from Higgsstrahlung have leptons and/or missing transverse energy from the W or Z boson decays in the final state leading to a distinct signature in the detector, as shown in the Feynman diagrams in Figure 7.4.

The feasibility of extracting a HW signal from LHC collision event samples has already been studied in the pre-collision era and, because of the huge $b\bar{b}$ background contributions, was found to be "very difficult (...) even under the most optimistic assumptions" by the ATLAS collaboration [152]. Most likely for the same reason, this search channel was omitted entirely in the analogous CMS feasibility studies [153]. The optimism connected with it revived only recently, after the proposal of studying this search in a boosted event topology [111], where both the Higgs boson and the vector boson have large momenta and move back-to-back in the transverse detector plane. Figure 7.1 shows how this boost requirement effectively

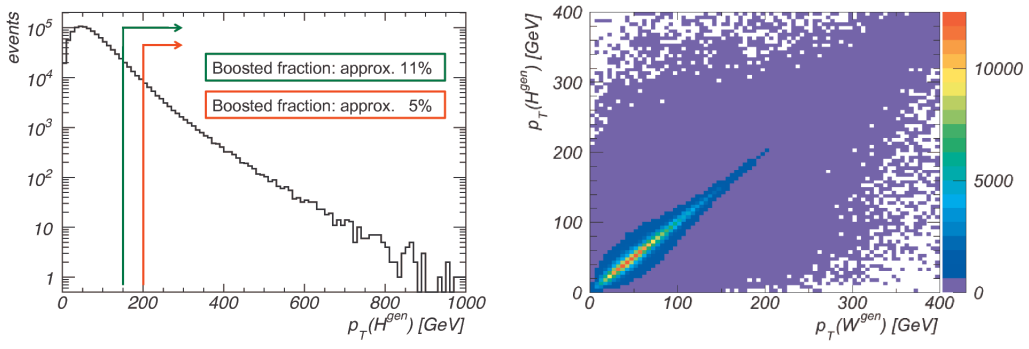


Figure 7.1.: Visualization of the boost requirement in simulated HW signal events ($m(H) = 125\,\text{GeV}$). Left: fractions of signal events remaining after a 150 GeV and a 200 GeV boost requirement. Right: correlation between the p_T of the Higgs and the W bosons.

reduces the signal phase space to less than 10% of the total cross section, depending on the specific p_T requirement. It has been argued by the authors of [111] however that the boost requirement also implies significant compensating advantages. In terms of signal it causes the decay products to be central in the detector and thus allows for the best possible reconstruction and b-tagging using the tracking system. In terms of background it reduces the contribution from $b\bar{b}$ events. Also, as can be seen in the Feynman diagrams of the background processes in Figure 7.5, it is e. g. in $\text{t}\bar{\text{t}}$ + jets events impossible to produce a high p_T $b\bar{b}$ system without a significant amount of additional jet activity. And the $H(b\bar{b})Z(\nu\nu)$ channel becomes a feasible detection channel, which however is not covered by this thesis.

The basic event selection starts with the reconstruction of a leptonically decaying W boson and the reconstruction of a Higgs boson candidate decaying into two b-jets. A key component of the analysis is the boost requirement for the Higgs boson candidate and the W boson.

Background contributions to the selected sample arise from W + jets production, singly and pair-produced top quark events, diboson production and residual QCD multijet events. The most important of these contributions are W+light jets, $\text{t}\bar{\text{t}}$ and $\text{W} + \text{b}\bar{\text{b}}$ jets. The W + light jets category refers to events with a W boson produced in association with jets arising from the hadronization of light (up, down, strange and charm) quarks as well as gluons and is therefore abbreviated as W + udscg jets. Consequently, the $\text{W} + \text{b}\bar{\text{b}}$ jets category indicates events with a W boson produced in association with one or more b-quark. For these three processes high purity control regions (CR) are defined such that they do not overlap with the signal sample, but yield reasonably high numbers of events. They are used to validate the simulated shapes of key variables in data and to determine their normalizations using data-driven scale factors. Other backgrounds are taken from simulation.

Discriminating variables are used to extract signal and background yields which

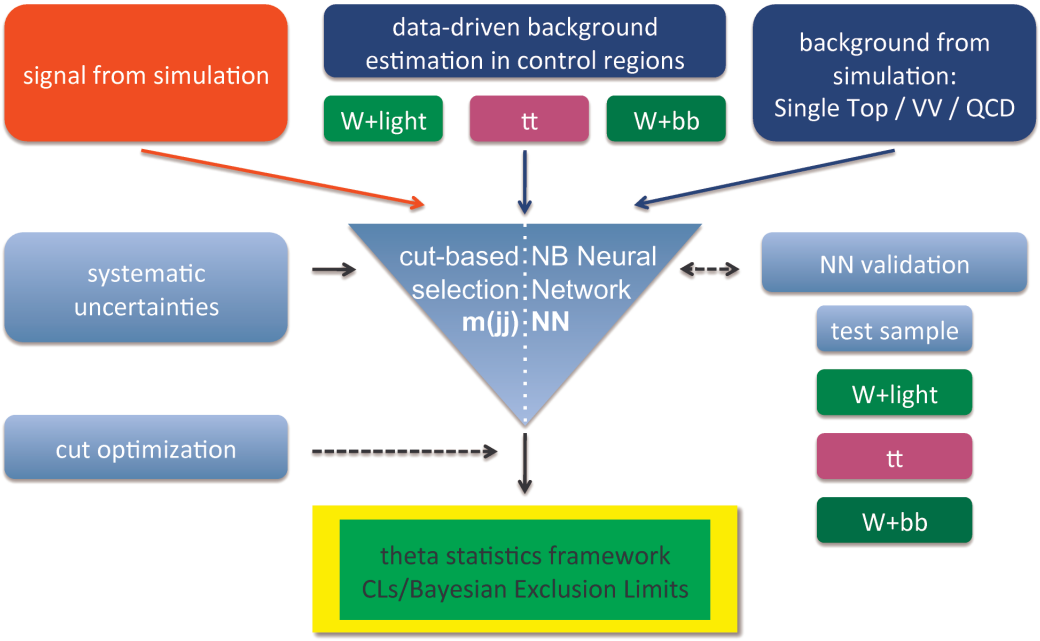


Figure 7.2.: Overview of the analysis strategy illustrating the interplay of the different ingredients to the HW search. This scheme is applied twice: once using standard jet reconstruction and then repeated with jets reconstructed with the sub-/filterjet algorithm.

are compared to data first using a "cut-and-count" technique ("m_{jj} analysis") and then a more sophisticated multivariate Neural Network approach ("NN analysis"). The yields in data and simulation are used for statistical inference including a full treatment of systematic uncertainties. Upper 95% confidence level (C.L.) Bayesian and CL_s exclusion limits are determined for the muon and electron channels separately and then combined. The analysis strategy is illustrated in Figure 7.2.

The analysis is performed twice, based on different reconstruction techniques for the $H \to b\bar{b}$ decay. First, standard anti-k_T ($R = 0.5$) jets are used ("standard analysis"). Then, the dedicated SJF jet algorithm originally proposed in [111] is applied, which advertises higher signal efficiency, better background rejection and higher Higgs boson mass resolution. This analysis is referred to as "SJF analysis".

Figure 7.3 shows the correlation between the expected opening angle of the b-quarks originating from the Higgs boson and the transverse momentum of the Higgs boson. The decay products move closer to each other with increasing transverse momentum of the Higgs boson. The lower and upper red lines indicate the thresholds for the Higgs boson reconstruction using two standard jets with $R = 0.5$ or one fat SJF jet with the technically possible maximum clustering parameter $R = \pi/2$, respectively. They illustrate that in the boosted Higgs boson regime above $p_T(H^{gen}) > 150\,\text{GeV}$ both analysis techniques yield similarly high efficiencies for the reconstruction of the $H \to b\bar{b}$ signal.

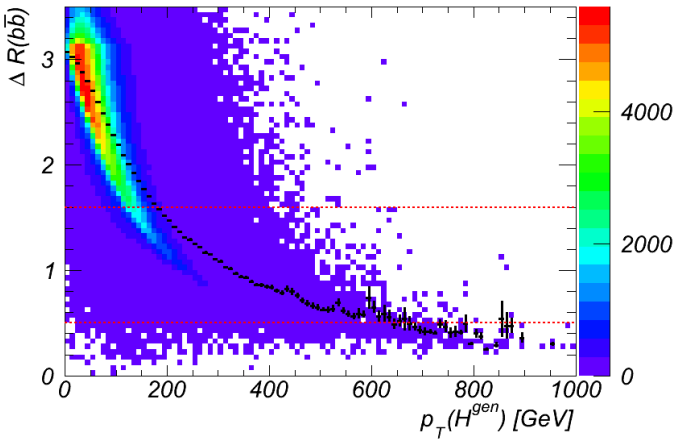


Figure 7.3.: Correlation between the Higgs boson momentum and the $b\bar{b}$ opening angle in a sample of generated HW signal events with $m(H) = 125\,\text{GeV}$. The decay products move closer to each other with increasing Higgs boson momentum. However, as indicated by the lower red line, even in the boosted regime above $p_{\text{T}}(H^{\text{gen}}) > 150\,\text{GeV}$ they are well separated above $\Delta R(b, \bar{b}) > 0.5$. This corresponds to the minimal distance of two anti-k_{T} jets with a clustering parameter $R = 0.5$ used in the standard jet reconstruction. The $H \to b\bar{b}$ decays can thus be reconstructed with a high efficiency in the boosted Higgs boson regime using two standard jets. The upper red line on the other hand shows that it is also possible to capture these events with a so called fat jet with a clustering parameter of $R = \pi/2$, corresponding to the technically possible maximum of the SJF algorithm. The distribution of events enclosed by the two red lines illustrates that both techniques yield similar $H \to b\bar{b}$ decay reconstruction efficiencies.

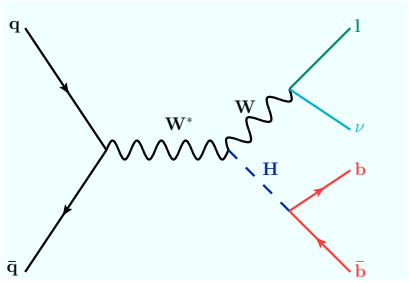
7.2. Monte Carlo, Data Samples and Triggers

Monte Carlo Samples

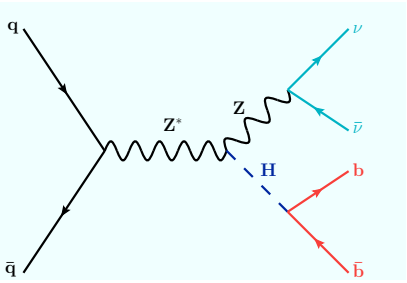
Several different generators are used to simulate samples of signal and background events. Table B.1 in Appendix B summarizes the generators, event numbers and cross sections for each individual sample and gives additional technical information on the generator parameters.

The signal samples are produced using Powheg and are scaled to next-to-next-to-leading order (NNLO) cross sections [29]. The $t\bar{t}$ + jets and V + jets samples are generated with MadGraph and scaled to next-to-next-to-leading logarithm (NNLL) [154] and next-to-leading order (using k-factors in analogy to [155]) respectively. Single top events are simulated using Powheg and correspond to NNLO and NNLL [156–158], depending on the production channel. The diboson and QCD samples are made with Pythia 6.4. The VV samples are normalized to NLO cross sections provided in [159]. The QCD samples are scaled using the leading order simulation cross section.

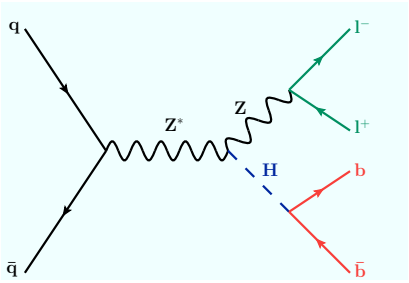
All samples contain an admixture of additional pile-up events to properly describe



(a) HW signal channel

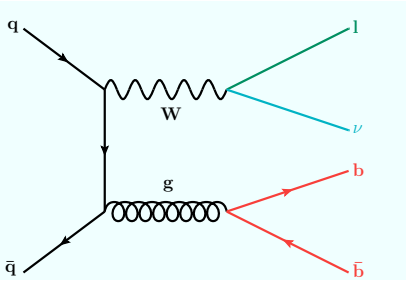


(b) HZ invisible channel

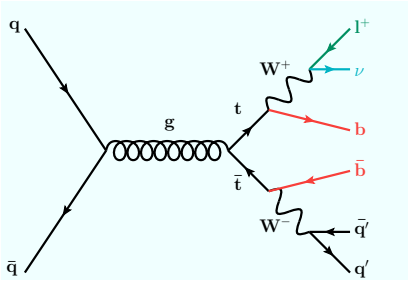


(c) HZ leptonic channel

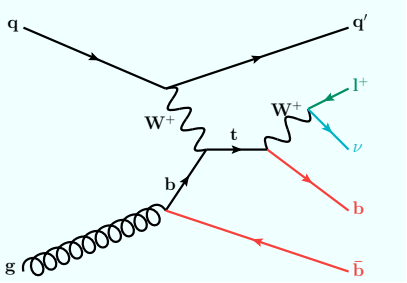
Figure 7.4.: Feynman diagrams for the HV signal.



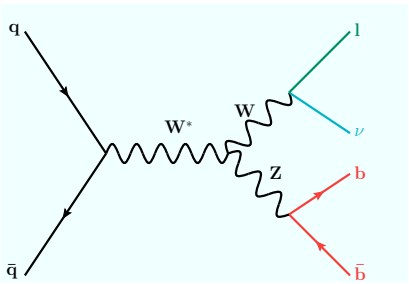
(a) W + jets background



(b) $t\bar{t}$ + jets background



(c) Single top background



(d) Diboson background

Figure 7.5.: Important backgrounds to the HW search. The color code indicates similarities of these final states and the HW signal.

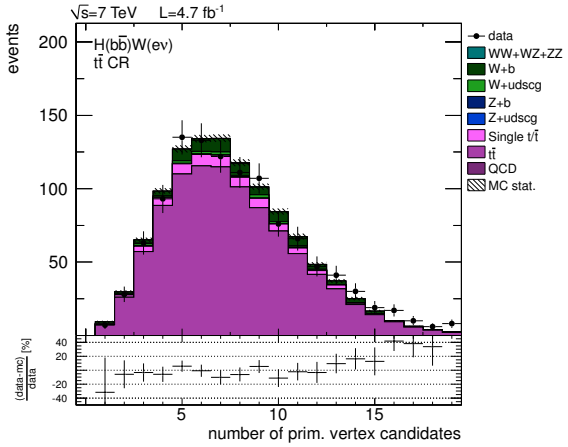


Figure 7.6.: Distribution of the number of primary vertices in data and simulation in the $t\bar{t}$ control region (CR) (for the definition of the $t\bar{t}$ CR see Section 7.5.6) of the electron channel. The simulation is reweighted to the pile-up profile expected in the full 2011 LHC dataset.

the high luminosity event conditions at the LHC. The samples are reweighted according to the pile-up profile expected in collision data [160]. Figure 7.6 shows the distribution of the number of primary vertices in data and samples of reweighted simulation events.

Data Samples & Triggers

The analysis presented in the following is based on the full 2011 dataset and corresponds to an integrated luminosity of $\mathcal{L} = 4.65\,\mathrm{fb}^{-1}$ after good run selection.

The data for the muon channel $H(b\bar{b})W(\mu\nu)$ analysis is taken exclusively from the Single Muon primary dataset (PD). Table B.2 in Appendix B summarizes the different dataset blocks in detail. A large set of different single muon triggers with and without online isolation requirements is used to cope with the steep rise in instantaneous luminosity provided by the LHC and to maximize the trigger efficiency over the full range of the data taking period. The explicit trigger paths with the corresponding run ranges are listed in Table B.3. This trigger selection corresponds to the official recommendation of the CMS Hbb group. The corresponding trigger, reconstruction and identification efficiencies which are applied to the samples of simulated events have been determined centrally within the CMS Higgs physics analysis group (PAG) using the tag-and-probe method and are taken from [155, 161].

The electron channel $H(b\bar{b})W(e\nu)$ relies on single electron triggers for the first part of the 2011 data taking campaign. For the second, larger part of the data sample the electron plus hadronic activity (ElectronHad) primary dataset is used. An overview over the different blocks is given in Table B.2. Table B.3 summarizes the

official trigger recommendation. Single electron triggers with tight online isolation criteria are used for the data from the single electron primary dataset. Hadronic cross triggers with tight online electron isolation are used for the data from the ElectronHad PD: in addition to single isolated electrons, they require additional online jets and online PFMHT (see Section 4.3.8). The determination of the corresponding trigger efficiencies [155, 161] uses multiple tag-and-probe steps. The single electron efficiency is calculated using the standard tag-and-probe method. For the cross triggers, the single jet reconstruction efficiency is then determined in events triggered on electrons with reconstructed jets as tags. In a third step the final efficiency is determined in events with electrons and jets as a function of the reconstructed $\not{E}_{\mathrm{T}}$.

7.3. Physics Objects and Vector Boson Reconstruction

The HW search presented in the following is based exclusively on physics objects reconstructed with the ParticleFlow algorithm as described in full detail in Chapter 4. The effect of pile-up is minimized by applying the Charged Hadron Subtraction correction (described in Section 4.3.7) during the event reconstruction consistently for collision data and samples of simulated events.

Vector Boson Reconstruction

The W boson is reconstructed from its decay products, one isolated muon or electron and the missing transverse energy $\not{E}_{\mathrm{T}}$ measured in the event, which characterizes the escaping neutrino from the W boson decay. The transverse momentum $p_{\mathrm{T}}(W)$ and the transverse mass $m_{\mathrm{T}}(W)$ of the W candidate are thus computed as

$$p_{\mathrm{T}}(W) = \sqrt{(\not{E}_x + p_x(\ell))^2 + (\not{E}_y + p_y(\ell))^2}$$
$$m_{\mathrm{T}}(W) = \sqrt{(\not{E}_{\mathrm{T}} + p_{\mathrm{T}}(\ell))^2 - p_{\mathrm{T}}(W)^2}.$$

As the $\not{E}_{\mathrm{T}}$ variable only provides information on the transverse magnitude of the missing energy, the flight direction of the W is calculated with a quadratic equation based on a W mass hypothesis of $m(W) = 80.43\,\mathrm{GeV}$.

7.4. Event Pre-Selection

All collision data events are required to be included in the list of certified good luminosity blocks provided by the CMS data quality management group. In addition, events from both collision data and from simulation are required to fulfill the following baseline selection criteria:

Good Primary Vertex Selection

All events considered in the HW search are required to have at least one good reconstructed primary vertex: the vertex fit must be associated with at least three tracks and the radial distance of the PV has to lie within $\rho(\mathrm{PV}) < 0.2\,\mathrm{cm}$ of the beam axis. To further reject non-collision or beam-related backgrounds the z coordinate of the PV is required to fall within the luminous collision region of $|z(\mathrm{PV})| < 24\,\mathrm{cm}$.

Muon Identification Selection

Muon candidates have to fulfill the tight selection criteria recommended by the Vector Boson Task Force [98]. They have to be reconstructed as global and tracker muons and the track associated to the global muon has to have a normalized χ^2 fit value below 10. In addition the track has to have a total of at least eleven tracker hits with one or more of them in the pixel sub-detector. At least one valid muon chamber hit is required in the final track fit with matched muon segments in at least two muon stations. The impact parameter of the track in the transverse plane of the detector has to be less than $|d_{xy}| < 2\,\mathrm{mm}$. In addition the relative lepton isolation, defined as

$$\mathrm{I}^{\mathrm{rel.}} \equiv \frac{\sum_i \left(p_{\mathrm{T},i}(\text{charged hadrons}) + p_{\mathrm{T},i}(\text{neutral hadrons}) + p_{\mathrm{T},i}(\text{photons})\right)}{p_{\mathrm{T}}(\text{lepton})}, \quad (7.1)$$

is required to be smaller than $\mathrm{I}^{\mathrm{rel.}} < 0.15$. The sum $\sum_i p_{\mathrm{T},i}$ is calculated from PF particles whose flight directions lie within a cone of radius $\Delta R < 0.4$ around the muon track. The muon has to lie within the tracker coverage of $|\eta| < 2.4$. Furthermore, an offline p_{T} requirement of $p_{\mathrm{T}}(\mu) > 20\,\mathrm{GeV}$ is applied to the leading muon candidate.

Electron Identification Selection

Electron candidates are required to fulfill the Vector Boson Task Force 80% working point (WP80) [100, 162] within a pseudorapidity range of $|\eta| < 2.5$, excluding the ECAL gap regions of $1.4442 < |\eta| < 1.5666$: below $|\eta| < 1.4442$ the spatial matchings between the candidate electron track and the corresponding ECAL supercluster have to satisfy $\Delta\phi < 0.8$ and $\Delta\eta < 0.007$. The supercluster η width has to be $\sigma_{i\eta i\eta} < 0.01$ and the so called hadronic leakage variable describing the fraction of the hadronic to the electromagnetic energy contributions is required to be less than $H/E < 0.12$. For $|\eta| > 1.5666$, these values change to $\Delta\phi < 0.7$, $\Delta\eta < 0.009$, $\sigma_{i\eta i\eta} < 0.03$ and $H/E < 0.15$. In addition, the transverse impact parameter of the track has to be less than $|d_{xy}| < 2\,\mathrm{mm}$ and the relative isolation (7.1) has to satisfy $\mathrm{I}^{\mathrm{rel.}} < 0.15$, where again the sum $\sum_i p_{\mathrm{T},i}$ is calculated within a cone of $\Delta R < 0.4$. Finally, an offline p_{T} threshold of $30\,\mathrm{GeV}$ is applied to the leading electron candidate.

Additional Lepton Veto

Additional isolated leptons are vetoed in order to reject backgrounds from Z boson decays and similar processes.

Jet Identification Selection

Two types of jets are used in the following, depending on the type of analysis. For jets clustered with the anti-k_T jet algorithm the official loose jet ID as described in [134] is required. For the sub- and filterjets described in the second part of this chapter there is no official jet identification recommendation. Therefore, the same anti-k_T jet ID requirements are applied to the SJF jets in order to reject fake and mismeasured jets. The effect of this is observed to be small.

7.5. Standard Jet Reconstruction Analysis

This section summarizes the HW search using standard jet reconstruction. The description of the analysis using sub-/filterjets is given in Section 7.6.

7.5.1. Standard Jets

The standard analysis is based on jets clustered with the anti-k_T algorithm with distance parameter $R = 0.5$. The jets are corrected with the L1FastJet pile-up correction based on the average pile-up energy density ρ and the CMS default L2Relative and L3Absolute jet energy corrections (see also Section 4.3.7). For data events the L2L3Residual correction is applied.

The measurement of the jet energy resolution described in Chapter 6 has shown that the samples of simulated events have to be bias-corrected to match the jet resolution observed in data. All jets in simulated events are thus scaled by 5% in the central detector ($|\eta| < 2.5$) and 10% in the forward region, following the prescription given in Section 6.5. Figure 7.7 illustrates the impact of the bias correction on the resolution of the reconstructed Higgs boson mass.

The Combined Secondary Vertex (CSV) b-tagger, described in Section 4.3.7, is used to identify jets arising from the fragmentation and hadronization of b-quarks.

7.5.2. Higgs Boson Candidate Reconstruction

The Higgs boson candidate is reconstructed from the pair of central jets whose dijet system, defined as the four vector sum $\vec{j}_1 + \vec{j}_2$, yields the highest transverse momentum (as an estimator of $p_T(H)$) in the event. To be considered, each of the jets has to fulfill a p_T threshold of 30 GeV and has to be well contained within the tracker region ($|\eta| < 2.4$) for subsequent b-tagging to be applicable.

This particular prescription to construct the Higgs bosons is found to be as good or better compared to alternative methods, e. g. the selection of the two jets with the highest CSV sum or pairing the highest p_T jets of the events. It yields a high

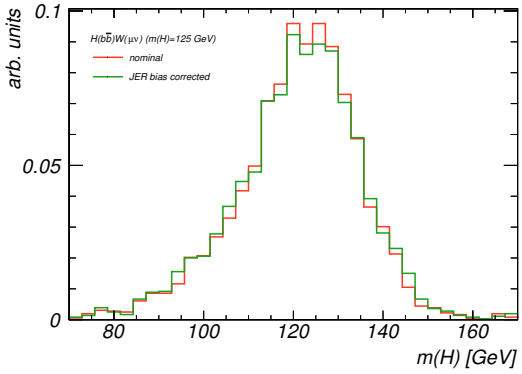


Figure 7.7.: Reconstructed Higgs boson mass before (red) and after (green) the necessary jet energy resolution bias correction described in Section 6.5 in a simulated signal sample with a generated Higgs boson mass of $m(H) = 125\,\text{GeV}$. This distribution is obtained requiring the signal selection described in Section 7.5.5. The Higgs boson mass resolution using anti-k_T ($R = 0.5$) PFJets is approximately 11% for the nominal simulated sample and degrades by roughly one percent as a result of the correction.

reconstruction efficiency and rejection of wrong combinations in signal events, while suppressing signal-like events from $t\bar{t}$ events [163].

An example of the reconstructed Higgs boson mass is given in Figure 7.7 for a simulated signal sample with a generated Higgs boson mass of $m(H) = 125\,\text{GeV}$. The predicted mass resolution using the highest $p_T(jj)$ reconstruction method is approximately 12% after the jet energy resolution bias correction.

7.5.3. Signal and Background Characteristics

HW signal

Signal events are characterized by the presence of a W boson recoiling at high transverse momentum against two b-jets with an invariant mass in the range of $110 < m(H) < 135\,\text{GeV}$, as illustrated in the Feynman diagrams in Figure 7.4. The dijet $p_{T,jj}$ spectrum is predicted to be harder than e. g. the V + jets background spectrum. Because of the required kinematic boost of the W boson and the dijet system, the combined WH system tends to have a large invariant mass. As a result, the vector boson and the dijet system are central in the detector and move back-to-back with an opening angle between the two of them which peaks at $\Delta\varphi = \pi$. Events contain exactly one isolated lepton from the W boson decay, and additional jets only arise from radiation or pile-up. Figure 7.8 shows some exemplary distributions of simulated signal events.

Various background processes have been identified, which yield configurations in the detector similar to the signal.

V + jets

The V + jets process, illustrated in Figure 7.5, includes W and Z boson production in association with one or more jets. Topologically, this background looks very much like the signal process, but has a softer p_{T} spectrum and a more sharply falling dijet mass distribution, as well as decay characteristics which differ from the signal, such as the azimuthal opening angle $\Delta\varphi$ between the W boson and the dijet system or their difference in pseudorapidity, as shown in Figure 7.8. After application of the full signal selection, $\mathrm{W} + \mathrm{b\bar{b}}$ events contribute more than 80% of the total V + jets background, the rest stems mostly from W + udscg with two mistags.

Top Quarks

Production of pairs of top quarks, as well as singly produced top quarks via the t-channel, the s-channel and the tW-channel (see Figure 7.5) are another source of background to the HW search. Decays of top quarks involve one or two real W boson decays, and in the case of $\mathrm{t\bar{t}}$ events, at least two b-jets.

$\mathrm{t\bar{t}}$ events differ from the HW signal mainly in two ways: they contain significantly more high energetic jets, and the opening angle between the two b-jets tends to be large.

Single top events are much harder to discriminate against, but the cross section is low enough such that this background contribution can be estimated from the MC prediction.

Dibosons

Production of vector boson pairs, in particular WZ, is yet another background. WZ events with a leptonically decaying W boson and a hadronic Z boson decay to $Z \to b\bar{b}$ are irreducible and can create resonant dijet systems with masses close to the potential signal. They can only be distinguished from the signal due to the difference between the Z and Higgs boson masses, relying very much on good mass resolution and therefore jet energy resolution.

QCD Multijet Events

The background to the unboosted $H \to b\bar{b}$ search arising from $b\bar{b}$ production and QCD multi-jet events faking an isolated lepton and b-jets is found to be very low after the boost requirement. The reconstruction of a high p_{T} W boson requires the presence of an isolated lepton with at least $p_{\mathrm{T}}(\ell) > 20\,\mathrm{GeV}$ and significant $\not{E}_{\mathrm{T}}$, as well as a b-tagged high p_{T} dijet system moving into the opposite transverse direction. These requirements suppress QCD events such that their very low rate is estimated from simulation without additional modifications. Additional studies within the Hbb group indicate that the background contribution of QCD to the boosted VH search is negligible [163].

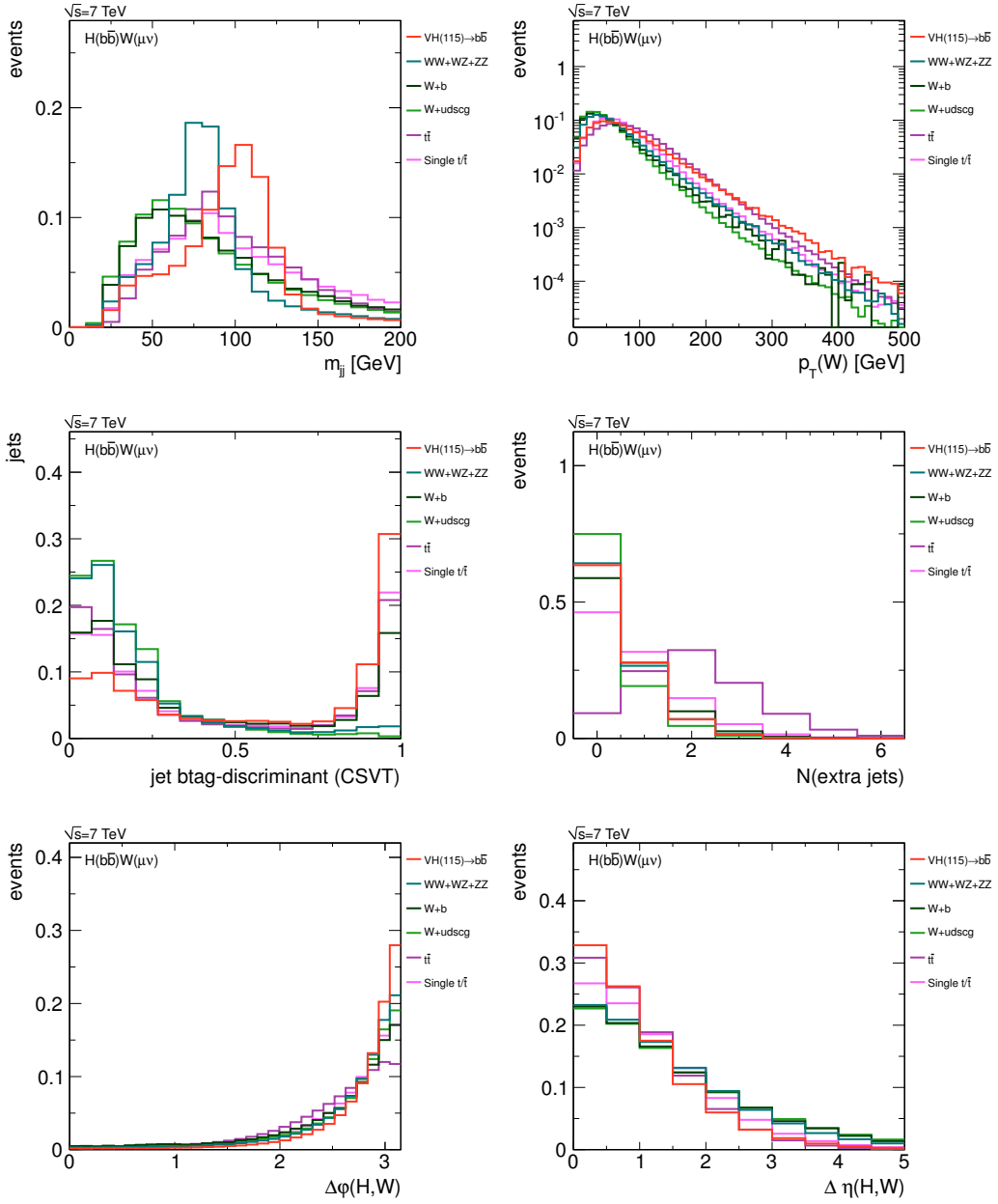


Figure 7.8.: Examples of discriminating variable distributions using simulated signal and background events, from the top left to the bottom right: the Higgs boson candidate (dijet) mass, the momentum of the W boson, b-tagging, the number of additional central jets in the event and the $\Delta\varphi$ and $\Delta\eta$ separation between the reconstructed Higgs boson candidate and the W boson. In these distributions the muon pre-selection is applied and events are normalized to unit area.

7.5.4. Definition of Variables

Based on these considerations and Figure 7.8 it is useful to introduce the following variables, which will be used to define the different analysis regions and help to discriminate signal versus background events:

- $p_{\mathrm{T}}(\ell)$: lepton transverse momentum.
- $\not\!\!E_{\mathrm{T}}$: transverse missing energy based on PF particles.
- $\not\!\!E_{\mathrm{T}}$ significance: computed as the ratio of the missing transverse energy and the square root of the sum of the transverse energies.
- $p_{\mathrm{T}}(W)$: transverse momentum of the W boson candidate.
- $m_{\mathrm{T}}(W)$: transverse mass of the W boson candidate.
- m_{jj}: mass of the Higgs boson candidate built from the pair of central jets ($|\eta| < 2.4$) above a p_{T} threshold of 30 GeV which has the highest dijet p_{T}.
- $p_{\mathrm{T},jj}$: p_{T} of the dijet Higgs boson candidate.
- N(extra jets): additional central jets in the event above a p_{T} threshold of 20 GeV.
- CSV: continuous output of the Combined Secondary Vertex b-tagger for the first and second jet of the dijet Higgs boson candidate.
- $p_{\mathrm{T}}(j_i)$: transverse momentum of the first or the second jet of the dijet Higgs boson candidate.
- $\Delta R(j, j)$: opening angle between the two Higgs boson candidate jets.
- $\Delta\varphi(H, W)$: azimuth opening angle between the W boson and the Higgs boson candidate.
- $\Delta\eta(H, W)$: absolute difference in η between the W boson and the Higgs boson candidate.

7.5.5. Event Selection

The selection of signal events is kept identical to the selection described in the CMS paper [164], which had been optimized maximizing the Punzi figure of merit [165],

$$\mathcal{S} \equiv \frac{S}{1.5 + \sqrt{B}}, \tag{7.2}$$

which performs particularly well for searches where the ratio of signal to background events is small. In this definition, S stands for the number of signal events expected from simulation and B for the number of expected background events.

Table 7.1.: Selection criteria for the final m_{jj} analysis for the muon and electron channels after the event pre-selection specified in Section 7.4. The second part of the table summarizes the specific mass window cut for each simulated Higgs boson mass point and corresponds to roughly plus/minus one standard deviation of the expected Higgs boson mass resolution.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}	GeV	> 0	> 0
$p_{\mathrm{T},jj}$	GeV	> 165	> 165
$p_{\mathrm{T}}(W)$	GeV	> 160	> 160
$\max(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.898	> 0.898
$\min(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.4	> 0.4
$\Delta\varphi(H, W)$	–	> 2.95	> 2.95
N(extra jets)	–	0	0
$\not{E}_{\mathrm{T}}$	GeV	–	35
$m_{jj}(110)$	GeV	$95 - 125$	$95 - 125$
$m_{jj}(115)$	GeV	$100 - 130$	$100 - 130$
$m_{jj}(120)$	GeV	$105 - 135$	$105 - 135$
$m_{jj}(125)$	GeV	$110 - 140$	$110 - 140$
$m_{jj}(130)$	GeV	$115 - 145$	$115 - 145$
$m_{jj}(135)$	GeV	$120 - 150$	$120 - 150$

Table 7.2.: Selection criteria for the NN analysis training region after the pre-selection specified in Section 7.4. The first part imposes minimal requirements on the boost and b-tagging to obtain a signal enriched NN training region. The second part summarizes the final selection criteria on the NN discriminator, described in detail in Section 7.5.8.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}	GeV	$(0, 250)$	$(0, 250)$
$p_{\mathrm{T},jj}$	GeV	> 150	> 150
$p_{\mathrm{T}}(W)$	GeV	> 150	> 150
$\max(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.4	> 0.4
$\min(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.4	> 0.4
$\not{E}_{\mathrm{T}}$	GeV	–	35
$\mathrm{NN}_{\mathrm{out}}(110)$	–	> 0.65	> 0.75
$\mathrm{NN}_{\mathrm{out}}(115)$	–	> 0.65	> 0.80
$\mathrm{NN}_{\mathrm{out}}(120)$	–	> 0.60	> 0.65
$\mathrm{NN}_{\mathrm{out}}(125)$	–	> 0.60	> 0.75
$\mathrm{NN}_{\mathrm{out}}(130)$	–	> 0.75	> 0.70
$\mathrm{NN}_{\mathrm{out}}(135)$	–	> 0.80	> 0.75

m_{jj} Analysis

Table 7.1 summarizes the selection criteria (also referred to as "cuts" in the following) defining the final m_{jj} analysis signal selection. The reconstruction of a dijet Higgs boson candidate is asserted by the $m_{jj} > 0\,\text{GeV}$ requirement. In addition, the signal region is defined by the kinematic boost on the dijet system and the W boson, as well as the large azimuthal opening angle $\Delta\varphi$ between the two. Both jets forming the Higgs boson candidate have to be tagged and the larger of the two b-tags has to satisfy the tight CSV working point provided by the b-tagging and vertexing (BTV) physics object group [166]. The optimal value for the lower b-tag is found to be 0.4, a value between the loose and medium CSV working points. In order to minimize background contributions from $\mathrm{t\bar{t}}$ events the extra jet activity is required to be minimal. An additional cut on $\not\!\!E_{\mathrm{T}}$ is applied in the electron channel to suppress residual QCD events.

NN Analysis

For the Neural Network (NN) analysis the selection criteria of the m_{jj} analysis are loosened to obtain a significantly larger sample with signal-like characteristics ("NN training region"). A slightly lower kinematic boost is required for the dijet system and the W boson. Furthermore, CSV tags are required for both Higgs boson candidate jets in order to reduce the number of $\mathrm{W}+\mathrm{udscg}$ background events to a reasonable level. In the electron channel an additional $\not\!\!E_{\mathrm{T}}$ requirement is necessary to reject residual QCD multijet events. Any further cuts on the event topology are dropped and the respective variables are used within the Neural Network, as will be discussed in Section 7.5.8. The detailed selection criteria defining the NN analysis region are given in Table 7.2.

7.5.6. Data-driven Estimation of Backgrounds

Any meaningful statement on the existence of HW signal events requires precise understanding of the background contributions and their uncertainties. By making modifications to the signal selection requirements, mutually exclusive control regions are defined for the three dominant background contributions, $\mathrm{W}+\mathrm{udscg}$ jets, $\mathrm{t\bar{t}}$ and $\mathrm{W}+\mathrm{b\bar{b}}$ jets, such that the control samples contain large numbers of events and are dominated by the single process which is to be studied, while retaining the majority of the characteristics of the signal sample. The control regions can therefore be used to validate the shape of key variables with data and to extract data-driven background normalization factors which are extrapolated and used to estimate the background rates after the signal selection.

W+udscg Control Region

A high purity $\mathrm{W}+\mathrm{udscg}$ jets control sample is obtained by not requiring any b-tags in the dijet system, as shown in Table 7.3. To increase the number of events in this

and the following control regions, the boost requirement of the dijet system and the W boson is reduced, the constraint on the additional jet activity is relaxed and the cut on $\Delta\varphi(H, W)$ is dropped in comparison to the signal selection. In order to suppress residual QCD multijet events in the control regions which gain importance when e. g. no b-tags are required an additional cut on the $\not\!\!E_{\mathrm{T}}$ significance is applied. These differences in the control region definitions with respect to the signal selection are covered by systematic uncertainties on the scale factors which are described at the end of this sub-section.

Table 7.3.: Definition of the W + udscg jets background control region. The key to a high purity control sample is the omission of the b-tag requirements for the dijet system.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}	GeV	$(0, 250)$	$(0, 250)$
$p_{\mathrm{T},jj}$	GeV	> 150	> 150
$p_{\mathrm{T}}(W)$	GeV	> 150	> 150
$\max(\mathrm{CSV1}, \mathrm{CSV2})$	–	–	–
$\min(\mathrm{CSV1}, \mathrm{CSV2})$	–	–	–
N(extra jets)	–	< 2	< 2
$\not\!\!E_{\mathrm{T}}$ significance	–	> 2	> 2

Figures 7.9 - 7.10 show key variable distributions of the muon channel in this background control region. In these distributions the scale factors for W + udscg jets, $\mathrm{t\bar{t}}$ and W + $\mathrm{b\bar{b}}$ jets whose determination is described in the following sub-sections, are already applied and the number of simulated events is normalized to the number of events in data in order to facilitate the comparison of shapes. Good agreement between simulated events and data is found for all variables. Remaining differences which exceed the purely statistical uncertainty shown in the plots are accounted for by additional systematic uncertainties on the scale factors, which are described at the end of this sub-section.

Additional variables and the same distributions for the electron channel are included in Appendix B in Figures B.1 through B.4.

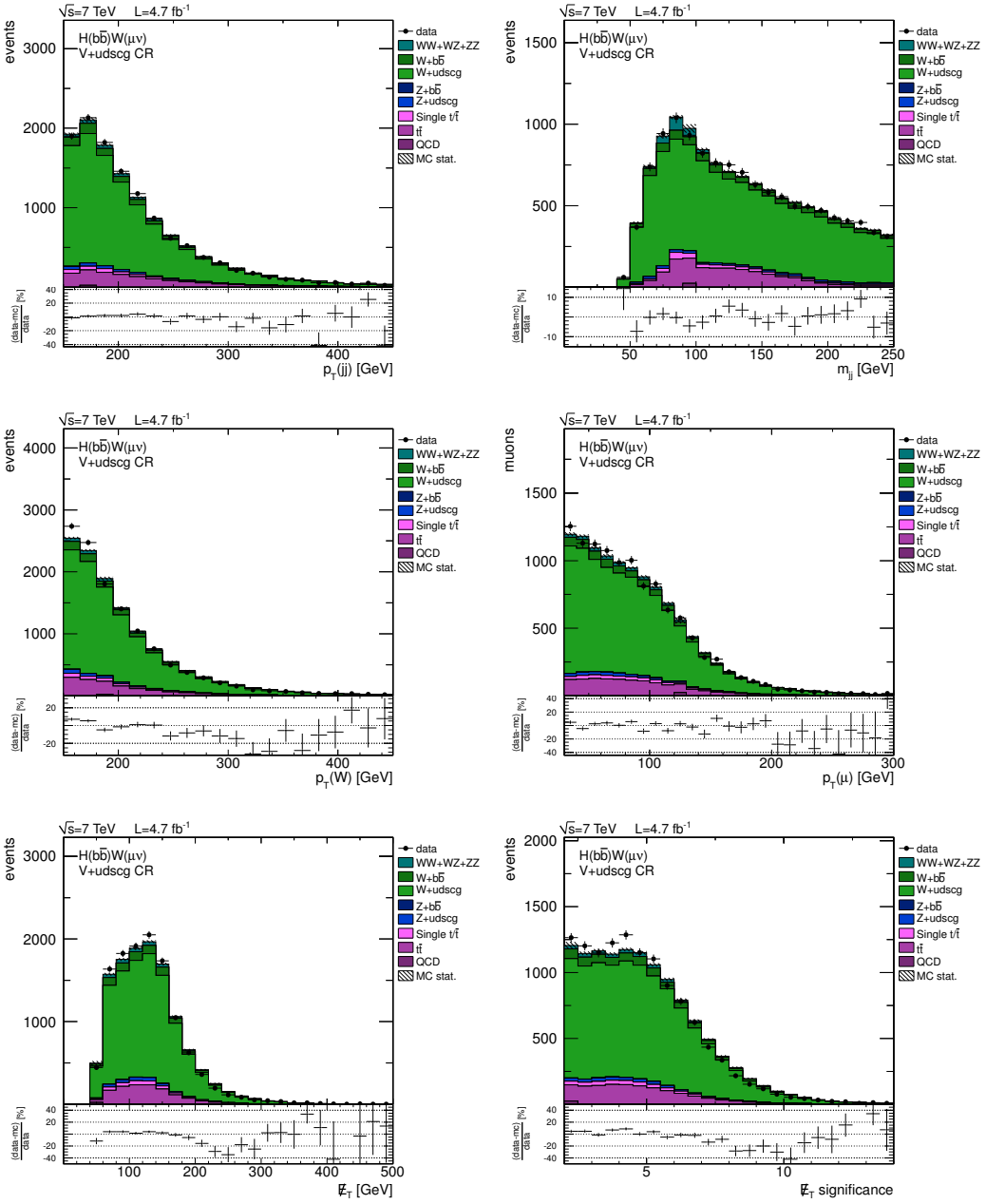


Figure 7.9.: W + udscg jets control region distributions for the muon channel: the p_{T} of the dijet system and the corresponding dijet mass, the p_{T} of the reconstructed W boson and muon, as well as the $\not{E}_{\mathrm{T}}$ and $\not{E}_{\mathrm{T}}$ significance distributions. Scale factors have been applied and the number of simulated events is normalized to data in order to facilitate the shape comparison. The same set of distributions for the electron channel is presented in Figure B.2 in the appendix.

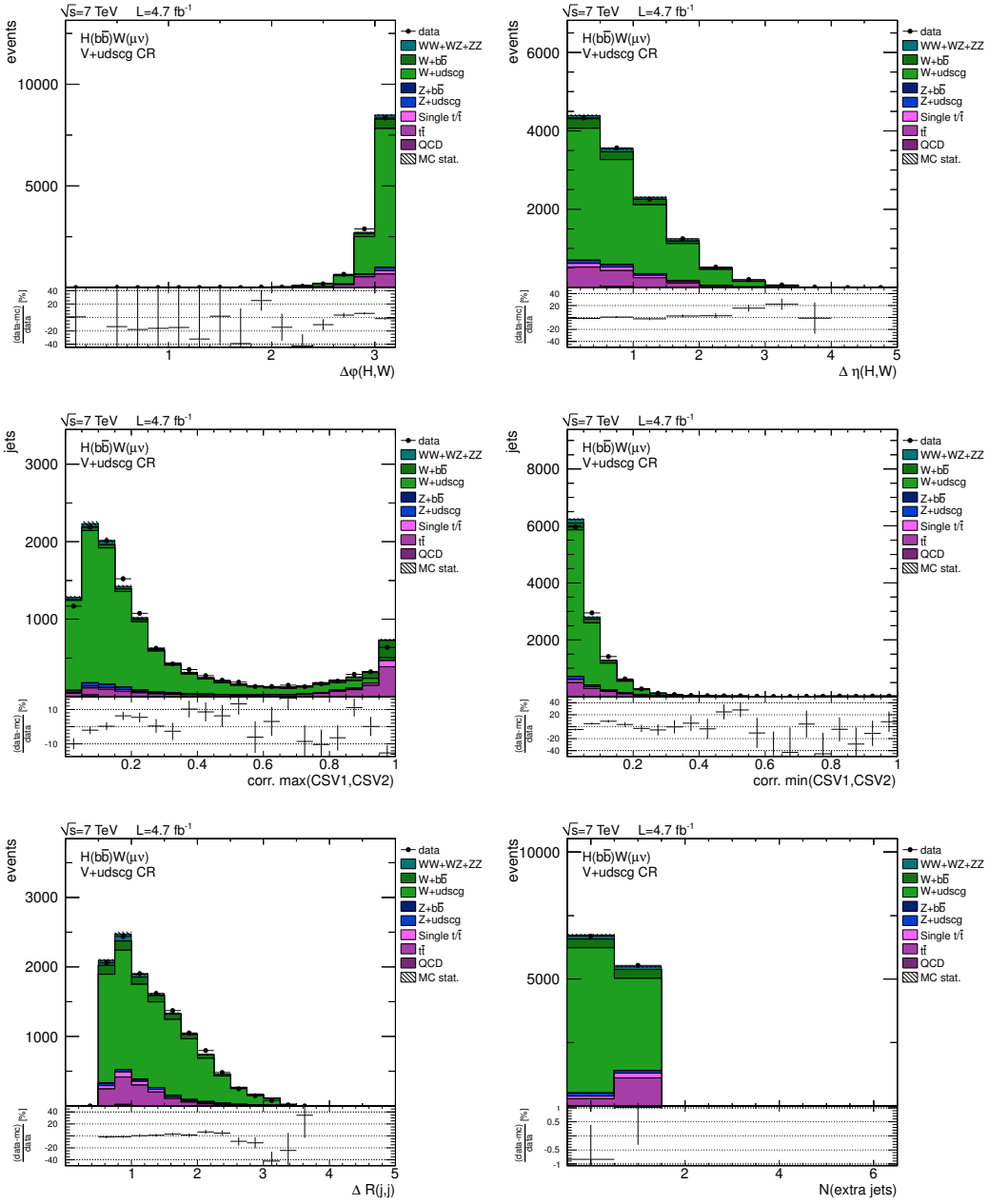


Figure 7.10.: W + udscg jets control region distributions for the muon channel (cont.): the azimuthal opening angle between the dijet system and the W boson, their difference in η, the CSV b-tag discriminator outputs for the dijet constituents, the opening angle ΔR between them and the number of additional central jets in the event. Scale factors have been applied and the number of simulated events is normalized to data. The same set of distributions for the electron channel is presented in Figure B.3 in the appendix.

$t\bar{t}$ Control Region

Given the presence of at least one real W boson and one or more b-jets, it follows that the production of top quark pairs introduces another important background to this analysis. Typically, $t\bar{t}$ events involving one isolated lepton have a high jet multiplicity and one or more b-jets, as can be seen from the corresponding Feynman diagram shown in Figure 7.5. A $t\bar{t}$ control region can thus be constructed by requiring a boost on the W and the dijet system, one tight CSV b-tag and an inversion of the minimal additional jet criterion with respect to the signal region. The detailed selection is summarized in Table 7.4.

Table 7.4.: Definition of the $t\bar{t}$ background control region. The most important handle on this background is the number of additional central jets in the event. It is reversed with respect to the signal region to at least 2 in order to obtain a high purity $t\bar{t}$ sample. One tight CSV b-tag on the Higgs boson candidate jets is necessary to reject contributions from W + udscg jets events.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}	GeV	$(0, 250)$	$(0, 250)$
$p_{\mathrm{T},jj}$	GeV	> 150	> 150
$p_{\mathrm{T}}(W)$	GeV	> 150	> 150
$\max(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.898	> 0.898
$\min(\mathrm{CSV1}, \mathrm{CSV2})$	–	–	–
$N(\text{extra jets})$	–	> 1	> 1
$\not{E}_{\mathrm{T}}$ significance	–	> 2	> 2

In order to present both electron and muon channels, Figures 7.11 - 7.12 show exemplary distributions of key variables in the $t\bar{t}$ background control region for the electron channel. Good agreement between data and simulated events is found in all variables. The corresponding plots for the muon channel and additional distributions for the electron channel are included in Figures B.5 through B.8 in the appendix.

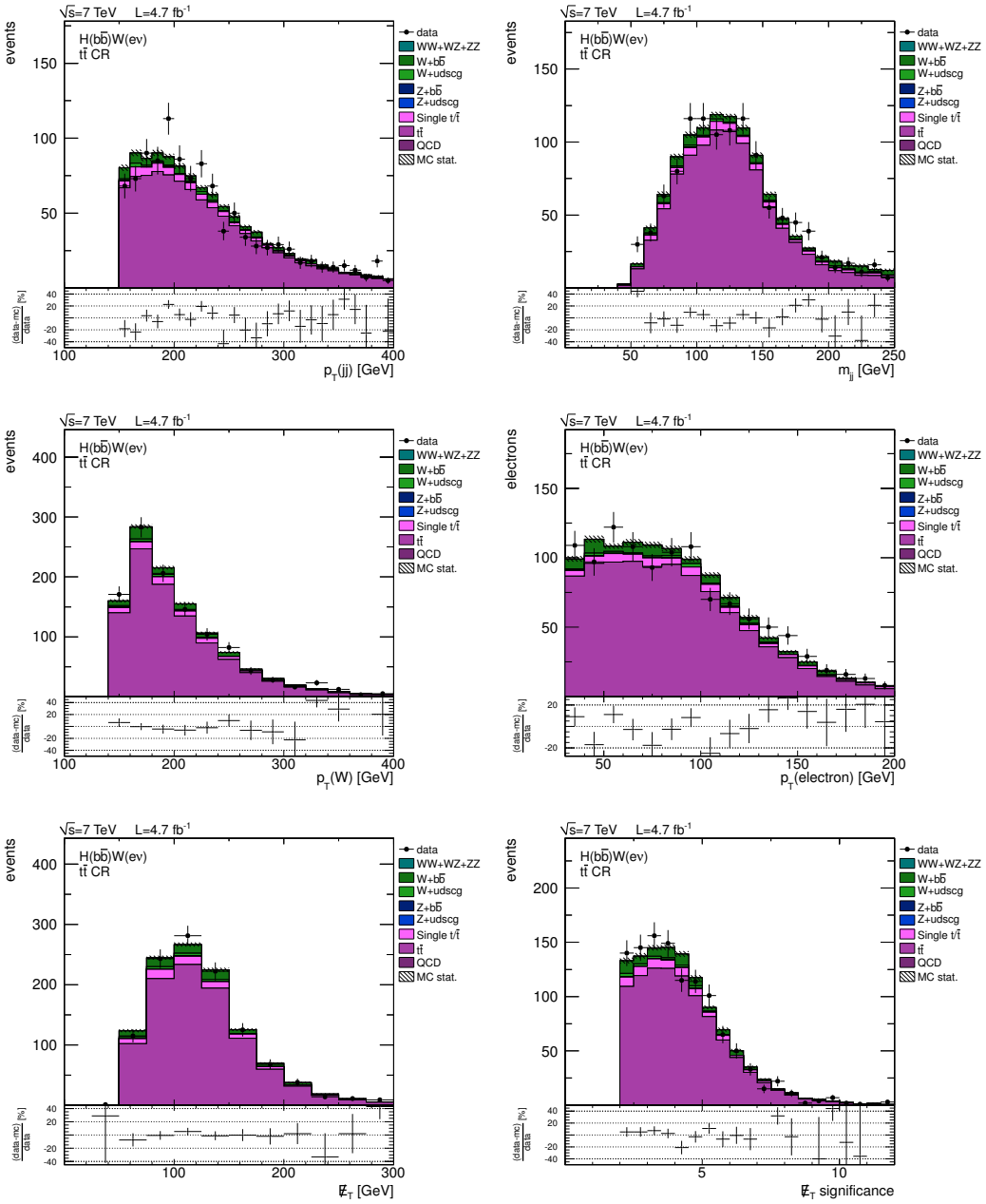


Figure 7.11.: $t\bar{t}$ control region distributions in the electron channel: the p_{T} of the dijet system and the corresponding dijet mass, the p_{T} of the reconstructed W boson and electron, as well as the $\not{E}_{\mathrm{T}}$ and $\not{E}_{\mathrm{T}}$ significance distributions. Scale factors have been applied and the number of simulated events is normalized to data. The same set of distributions for the muon channel is presented in Figure B.5 in the Appendix.

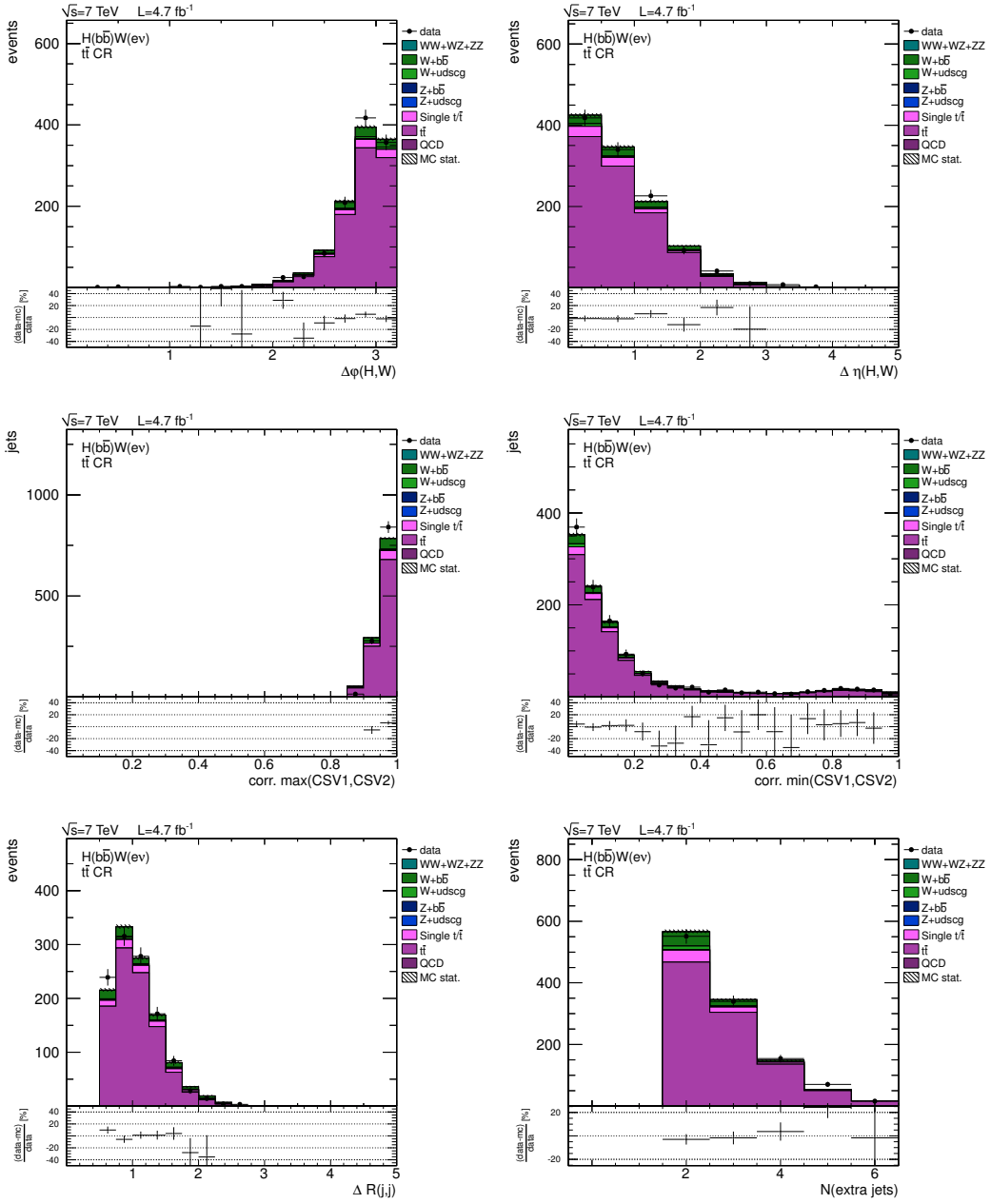


Figure 7.12.: $t\bar{t}$ control region distributions in the electron channel (cont.): the azimuthal opening angle between the dijet system and the W boson, their difference in η, the CSV b-tag discriminator outputs for the dijet constituents, the opening angle ΔR between them and the number of additional central jets in the event. Scale factors have been applied and the number of simulated events is normalized. The same set of distributions for the muon channel is presented in Figure B.6 in the appendix.

$\mathrm{W} + \mathrm{b\bar{b}}$ Control Region

The background contribution from events with a W boson and additional jets arising from the fragmentation of one or more b-quarks is the most difficult one to be constrained directly from data. Topologically, this background is nearly indistinguishable from single top production and in terms of event rate it is hard to separate from $\mathrm{t\bar{t}}$ production. Therefore it is impossible to construct a high purity $\mathrm{W} + \mathrm{b\bar{b}}$ control region with large statistics like the ones discussed previously. However, the construction of a $\mathrm{W} + \mathrm{b\bar{b}}$ enriched control sample which is not entirely pure, but allows for a scale factor determination with reasonable uncertainties, is feasible.

A control region which is orthogonal to the signal region is achieved by reversing the boost requirement, meaning that the p_{T} thresholds of the W boson and the dijet system are required to be below 150 GeV instead of above. The reversal of the kinematic boost leads to event characteristics which differ from the signal sample. The effect of this is taken into account when estimating the systematic uncertainties (as discussed in the end of this sub-section), such that the scale factor is valid within its uncertainties in the boosted regime. Table 7.5 summarizes the exact $\mathrm{W} + \mathrm{b\bar{b}}$ control region definition. As the $\mathrm{W} + \mathrm{b\bar{b}}$ category comprises events with one or more b-quarks, one tight b-tag is required to suppress contributions from W + udscg events. The reversal of the kinematic boost requires a harder constraint on the $\not{E}_{\mathrm{T}}$ significance in the electron channel with respect to the other control region definitions in order to suppress residual QCD events.

Table 7.5.: Definition of the $\mathrm{W} + \mathrm{b\bar{b}}$ jets background control region. A reversal of the kinematic boost of the W boson and the dijet system is used to assert orthogonality to the signal region. The minimal extra jet requirement reduces the $\mathrm{t\bar{t}}$ contribution as much as possible. In this low p_{T} regime a slightly higher cut on the $\not{E}_{\mathrm{T}}$ significance is needed in the electron channel with respect to the boosted control regions in order to remove residual QCD multijet events.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}	GeV	$(0, 250)$	$(0, 250)$
$p_{\mathrm{T},jj}$	GeV	< 150	< 150
$p_{\mathrm{T}}(W)$	GeV	< 150	< 150
$\max(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.898	> 0.898
$\min(\mathrm{CSV1}, \mathrm{CSV2})$	–	–	–
$N(\text{extra jets})$	–	0	0
$\not{E}_{\mathrm{T}}$ significance	–	> 2	> 3

Figures 7.13 - 7.14 show distributions of the key analysis variables in the $\mathrm{W} + \mathrm{b\bar{b}}$ control region for the muon channel. Good agreement between data and simulated events is found in all variables. Additional plots for the muon channel and the same distributions for the electron channel are included in the Appendix B in Figures B.9 through B.12.

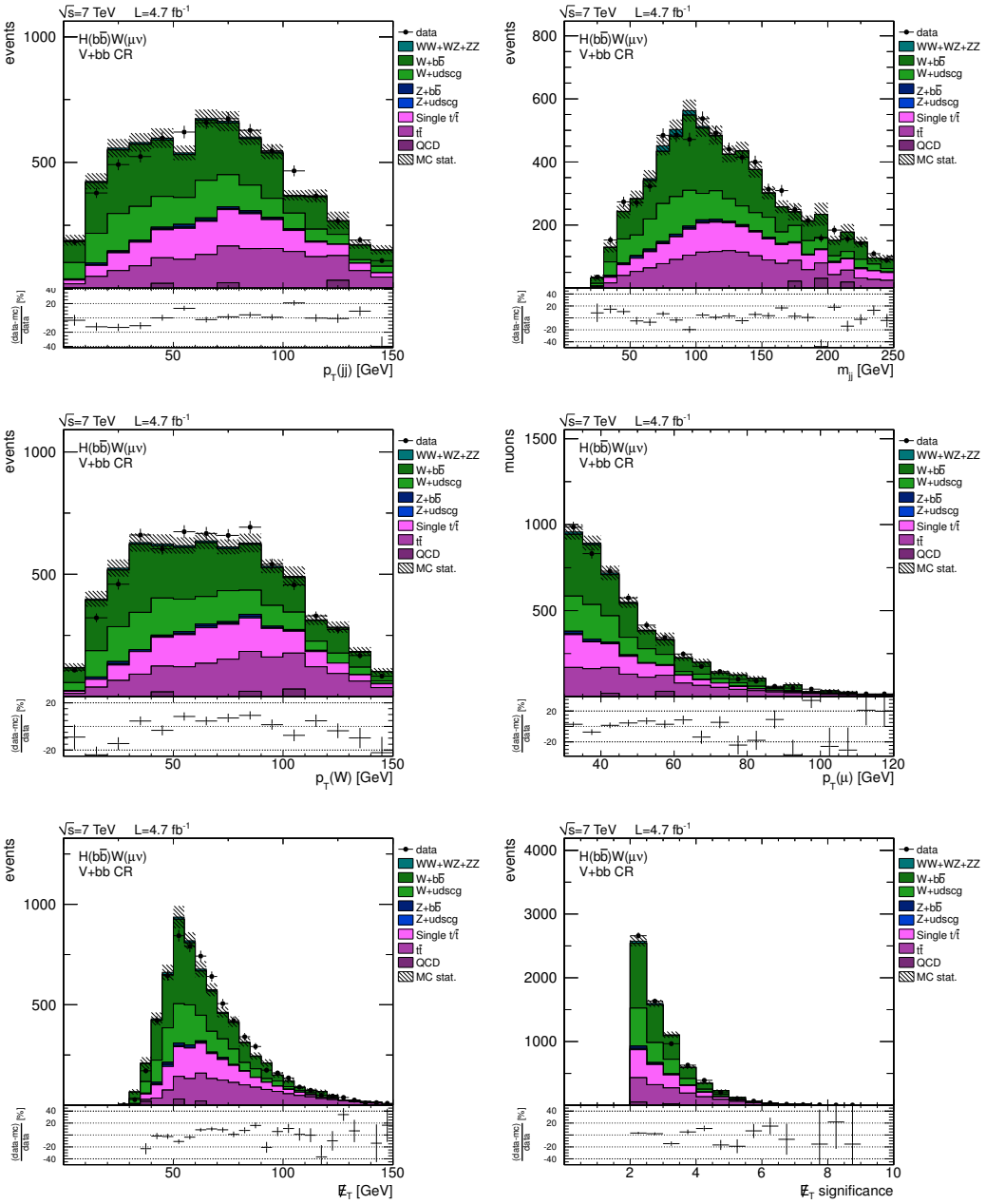


Figure 7.13.: $W + b\bar{b}$ jets control region distributions for the muon channel: the p_T of the dijet system and the corresponding dijet mass, the p_T of the reconstructed W boson and muon, as well as the $\not{E}_T$ and $\not{E}_T$ significance distributions. Scale factors have been applied and the number of simulated events is normalized to data. The same set of distributions for the electron channel is presented in Figure B.10 in the appendix.

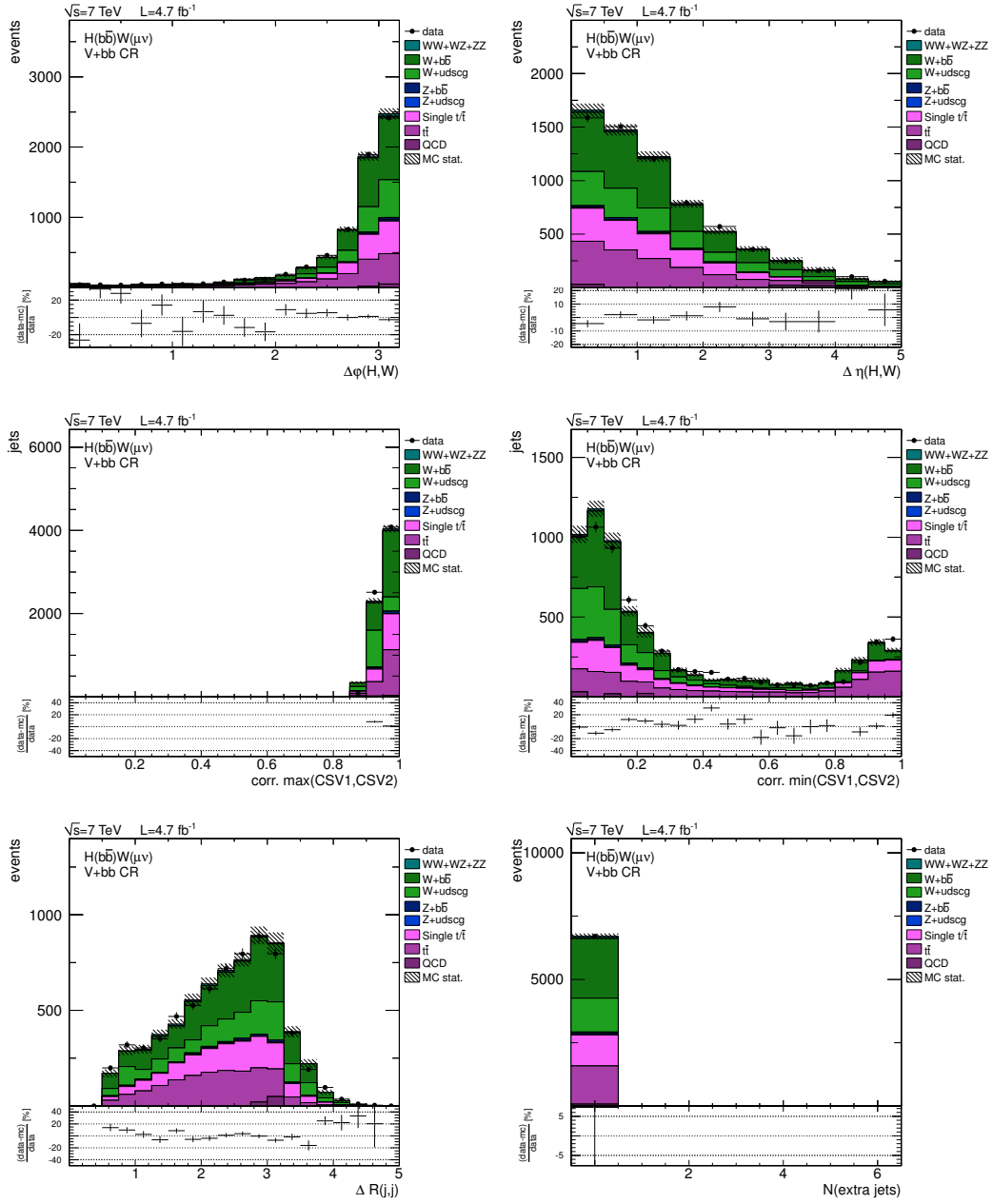


Figure 7.14.: W + $b\bar{b}$ jets control region distributions for the muon channel (cont.): the azimuthal opening angle between the dijet system and the W boson, their difference in η, the CSV b-tag discriminator outputs for the dijet constituents, the opening angle ΔR between them and the number of additional central jets in the event. Scale factors have been applied and the number of simulated events is normalized to data. The same set of distributions for the electron channel is presented in Figure B.11 in the appendix.

Table 7.6.: Summary of the data/MC scale factors. The values are the best parameter estimators of a simultaneous maximum likelihood fit to the three control regions in both the electron and the muon channel. The statistical uncertainties (stat.) on the fit results are very small compared to the systematic uncertainties (syst.) which are discussed in the next sub-section.

scale factor	value		(stat.)		(syst.)
$\mathrm{W} + \mathrm{udscg}$	0.8846	$\pm$	0.0088	$\pm$	0.1327
$\mathrm{t\bar{t}}$	0.9616	$\pm$	0.0231	$\pm$	0.0962
$\mathrm{W} + \mathrm{b\bar{b}}$	1.9263	$\pm$	0.0768	$\pm$	0.3853

Scale Factor Determination Using a Maximum Likelihood Fit

Following the methodology for parameter estimation described in Section 5.1, a maximum likelihood fit to the event yields in data and MC is used to obtain the data/MC scale factors. The fit is performed simultaneously in all three control regions and in both the electron and the muon channel.

The priors for the $\mathrm{W} + \mathrm{udscg}$, $\mathrm{t\bar{t}}$ and $\mathrm{W} + \mathrm{b\bar{b}}$ scale factors are assumed to be flat allowing a fit range within $(0, \infty)$. The priors of the nuisance parameters describing the additional backgrounds, such as single top, are fixed to their nominal values. The uncertainties of these additional backgrounds are not included as Gaussian constraints into the fit, because of the dilution of the $\mathrm{W} + \mathrm{b\bar{b}}$ control region with other backgrounds. Single top for example is anti-correlated with the $\mathrm{W} + \mathrm{b\bar{b}}$ event yield and makes up a large component of the $\mathrm{W} + \mathrm{b\bar{b}}$ enriched control region. Inclusion of its uncertainty into the likelihood fit would overestimate the uncertainty in the final signal region due to the cancelling effect of single top and $\mathrm{W} + \mathrm{b\bar{b}}$ moving into opposite directions. The omission of these correlations in the scale factor determination and their effect on the fit uncertainties is negligible with respect to an additional and in comparison very large systematic uncertainty on top of the simple fit errors which is discussed in the next sub-section.

Systematic Uncertainty on the Scale Factor Determination

The determination of the data/MC scale factors is based on a particular set of definitions for the control regions, i. e. the selection criteria specified in Tables 7.3 through 7.5. The details of the definitions, meaning the exact values of the selection requirements, should not make a large difference and their variation yields a systematic uncertainty. In addition, data/MC agreement in the distributions of key analysis variables is good throughout the various control regions, but small remaining shape differences can bias the determination of the scale factors. Finally, correlations between the three control region backgrounds and other contributions, such as single top, are not yet accounted for and thus not included in the uncer-

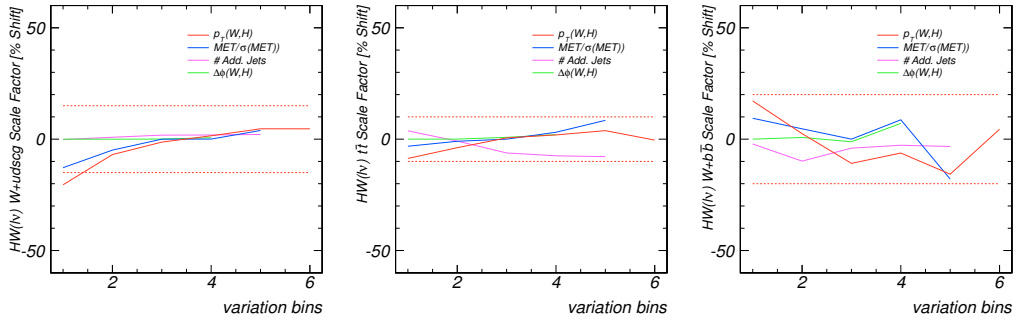


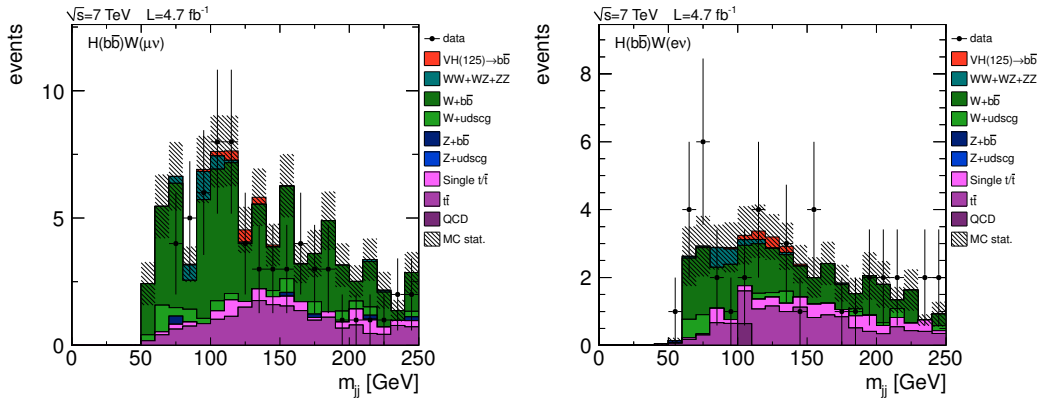
Figure 7.15.: Systematic uncertainties on the background scale factors: the kinematic boost requirement, the $\not\!\!E_{\mathrm{T}}$ significance, the number of additional jets in the event and opening angle between W boson and dijet system are varied (see text) within the definitions of the background control regions. To illustrate the effect of the variations in different variables comprehensively, the relative shift of the fit results is shown as a function of "variation bins" instead of the true units of the corresponding variable (e. g. $p_{\mathrm{T}}(H)(\,\mathrm{GeV})$): the x-axis thus shows the sequence of different cuts for the tested variables. For each control region, the maximum variation indicated by the red dotted lines is taken as a conservative estimate of additional systematic uncertainty on the respective scale factor.

tainties retrieved from the maximum likelihood fit.

In order to account for all these effects, the uncertainty derived from the maximum likelihood fit is inflated accordingly: Figure 7.15 shows the relative scale factor shifts in the three control regions as a function of varying requirements of the kinematic boost (red solid line), the $\not\!\!E_{\mathrm{T}}$ significance, the number of additional jets in the events and the opening angle between the W boson and the Higgs boson candidate. This is expected to be a conservative approach, as for example the boost thresholds of the W and the dijet system are scanned from $0-120\,\mathrm{GeV}$ in the unboosted $\mathrm{W}+\mathrm{b\bar{b}}$ control region and from $100-200\,\mathrm{GeV}$ in the case of the $\mathrm{W}+\mathrm{udscg}$ and $\mathrm{t\bar{t}}$ control regions, well exceeding the $150-165\,\mathrm{GeV}$ boost required in the final signal selections. The blue solid line illustrates the effect of a variation of the $\not\!\!E_{\mathrm{T}}$ significance in the control region definitions. It is varied from $0-4$. Compared to the first two, the effect of changing the number of additional jets required in the different control regions (magenta line) from $0-4$ is rather small. The same holds for variations of the opening angle between the W boson and the Higgs boson candidate (green line) from 0 to 3.

The overall additional systematic uncertainty on each data/MC scale factor is estimated by taking the envelope of the maximal relative shift: 15% systematic uncertainty are taken for the $\mathrm{W}+\mathrm{udscg}$ scale factor and 10% for the $\mathrm{t\bar{t}}$ scale factor. The uncertainty for the $\mathrm{W}+\mathrm{b\bar{b}}$ scale factor is broadened from an estimated 15% to 20% to additionally account for the dilution of this control sample, e. g. by single top events.

The full systematic uncertainty values estimated here are added in quadrature



(a) (b)

Figure 7.16.: Higgs boson candidate mass distribution in the signal region for the (a) muon channel and (b) electron channel. The distributions are dominated by backgrounds. The Higgs boson signal predicted by the Standard Model for $m(H) = 125\,\text{GeV}$ is represented by the red histogram. The final cut-and-count selection includes small sliding mass windows of $\pm 15\,\text{GeV}$ around the assumed Higgs boson masses.

to the fit uncertainties to obtain the full uncertainty on the data/MC scale factors.

7.5.7. Cut-based Analysis

In the previous sub-sections the good agreement of the variable distributions in samples of events from collision data and MC simulation has been shown in three different background control regions.

After the m_{jj} signal selection summarized in Table 7.1, the final predicted rates of signal and background events are from simulation, corrected for the measured scale factors from Table 7.6. Figure 7.16 shows the Higgs boson candidate mass distributions after the m_{jj} signal selection, but before the application of the sliding mass window cut. Figure 7.16(a) illustrates this distribution in the muon channel. It is dominated by backgrounds, with events from $\text{W}+\text{b}\bar{\text{b}}$ production accounting for the largest contribution. The uncertainty due to the limited available statistics of simulation events is indicated by the black shaded area and covers the distribution's spikiness. The Higgs boson signal predicted by the Standard Model for a Higgs boson mass of 125 GeV is represented by the red histogram. Figure 7.16(b) shows the same distribution for the electron channel. It is noteworthy that the total event yield in the electron channel is predicted to be smaller with respect to the muon channel, due to tighter selection criteria imposed by the triggers and a harder offline selection, which is needed to suppress fake electrons.

The final expected and observed event yields after all m_{jj} analysis signal selection requirements are summarized in Table 7.7 for the muon channel and in Table 7.8 for the electron channel. In addition, the expected significances for each channel

Table 7.7.: Final event yield of the m_{jj} analysis in the muon channel. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	0.64	0.64	0.43	0.26	0.06	0.25
W + $b\bar{b}$	14.41	12.59	11.52	10.74	9.18	7.31
Z + jets	0.00	0.00	0.00	0.00	0.00	0.00
$t\bar{t}$	3.54	3.66	3.96	4.39	4.88	4.85
Single top	1.05	1.20	1.26	1.33	1.15	1.00
VV	1.14	0.67	0.35	0.15	0.12	0.05
QCD	0.00	0.00	0.00	0.00	0.00	0.00
Background	20.79	18.76	17.51	16.87	15.39	13.46
Signal	1.81	1.57	1.38	1.12	0.90	0.69
Data	20	20	18	15	16	10
Significance	0.30	0.27	0.24	0.20	0.17	0.13

Table 7.8.: Final event yield of the m_{jj} analysis in the electron channel. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	0.31	0.31	0.66	0.66	0.66	0.49
W + $b\bar{b}$	4.80	3.95	3.71	3.84	4.08	3.30
Z + jets	0.00	0.00	0.00	0.00	0.00	0.00
$t\bar{t}$	2.87	3.18	3.20	3.23	3.25	3.28
Single top	0.69	0.71	0.82	0.79	0.70	0.80
VV	0.49	0.30	0.18	0.19	0.07	0.07
QCD	0.64	0.64	0.00	0.00	0.00	0.00
Background	9.79	9.10	8.58	8.72	8.76	7.93
Signal	1.26	1.11	0.92	0.76	0.64	0.46
Data	6	6	6	7	6	4
Significance	0.27	0.25	0.21	0.17	0.14	0.11

and mass point are given based on the figure of merit defined in equation (7.2).

The statistical evaluation of these numbers is shown in Section 7.5.10, after a discussion of the uncertainties given in Section 7.5.9. But first, the multivariate approach to this analysis using Neural Networks is presented in the following.

7.5.8. Neural Network Analysis

The fundamental idea of the Neural Network based approach to the HW search is to drop as many of the hard selection criteria from the m_{jj} analysis as possible and to combine them - after a proper NN training - in one optimized cut on the Neural Network discriminator, which is denoted $\mathrm{NN_{out}}$ in the following. In this way, correlations and non-linear relationships between the different variables are considered in the final signal selection, leading to higher signal efficiencies.

The Neural Networks employed in this thesis rely on the NeuroBayes package. Details on the NN method in general and technical aspects of the NNs used are described in Chapter 5.

Neural Network Training

The general NN analysis strategy is based on a NN training sample which is selected using significantly looser cuts with respect to the m_{jj} analysis, such that it contains a large number of events with signal-like characteristics for the NN training. Features of this sample were motivated in Section 7.5.5, the detailed definition is summarized in Table 7.2.

Several of the variables defined in the m_{jj} analysis have been studied as discriminating input variables to the Neural Networks. Based on their discrimination power, eight of them are chosen as final inputs to the NN. They are listed in Table 7.9.

The most important input variables, based on the relative weights of the input nodes determined by NeuroBayes, are the number of additional central jets in the event, the mass of the Higgs boson candidate and the CSV discriminator output for the second Higgs boson jet. This ranking reflects the dominant background contributions in the NN training sample. After requiring two b-tags, $\mathrm{t\bar{t}}$ is the dominant background contribution, which depends strongly on the event jet multiplicity. The $\mathrm{W + b\bar{b}}$ background is very similar to the signal, but tends to have lower dijet invariant mass scales than HW signal events. In addition, most of the events passing the NN training selection do have a rather good b-tag discriminator value for the leading jet. The Neural Networks thus take a lot of discrimination power from the b-tag of the second leading Higgs boson jet.

The prefix "corr." of the CSV variables indicates that the CSV discriminator output has been corrected or "reshaped" in the simulated samples to account for the b-tagging efficiency difference between data and simulation. Three discrete working points of the b-taggers have been evaluated by the BTV group and precise estimations of efficiency and mistag rates, including systematic uncertainties, are

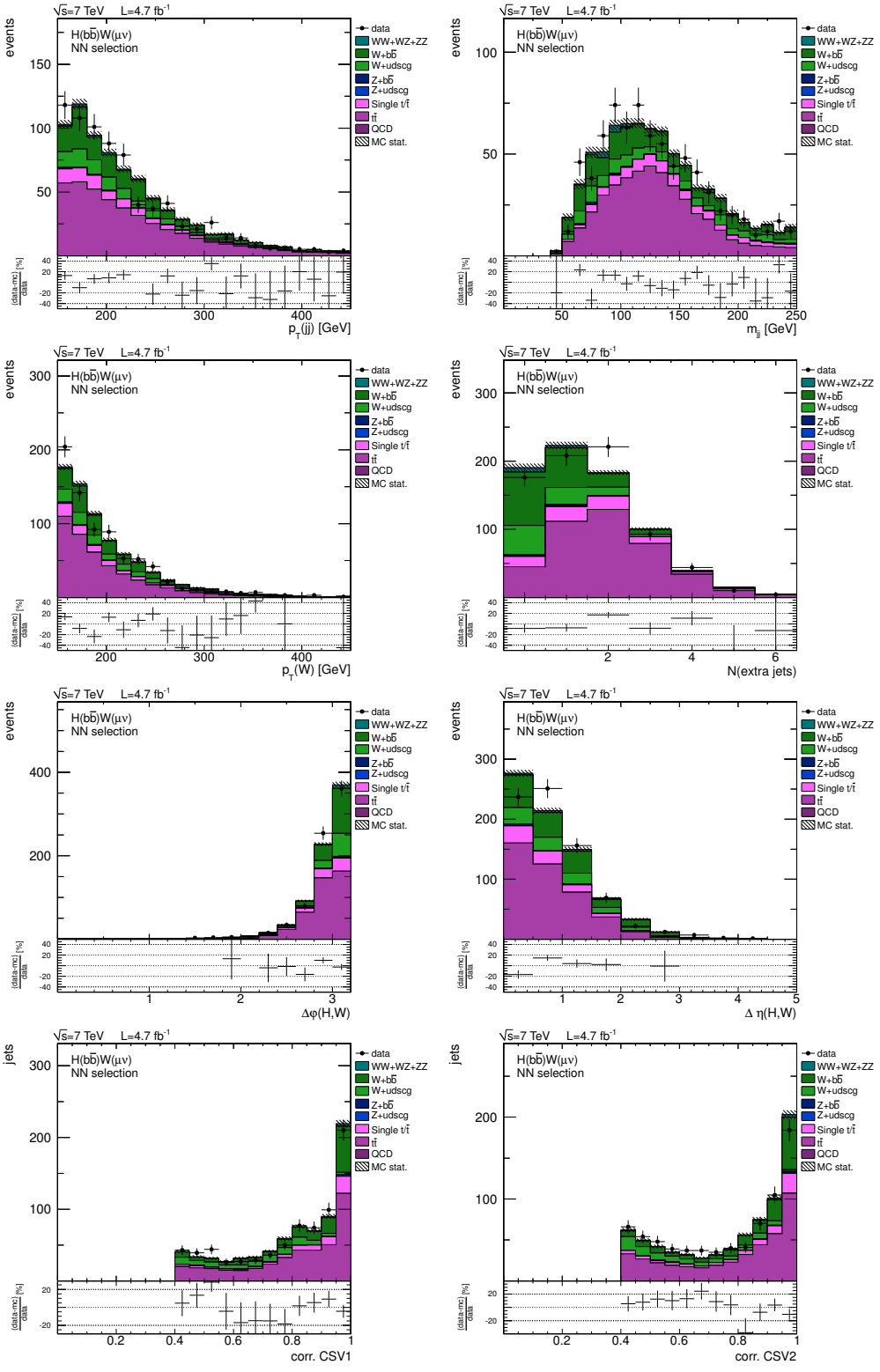


Figure 7.17.: Distributions of NN input variables in the muon channel after the NN training selection described in the first part of Table 7.2. The number of simulated events is normalized to the number of events in data.

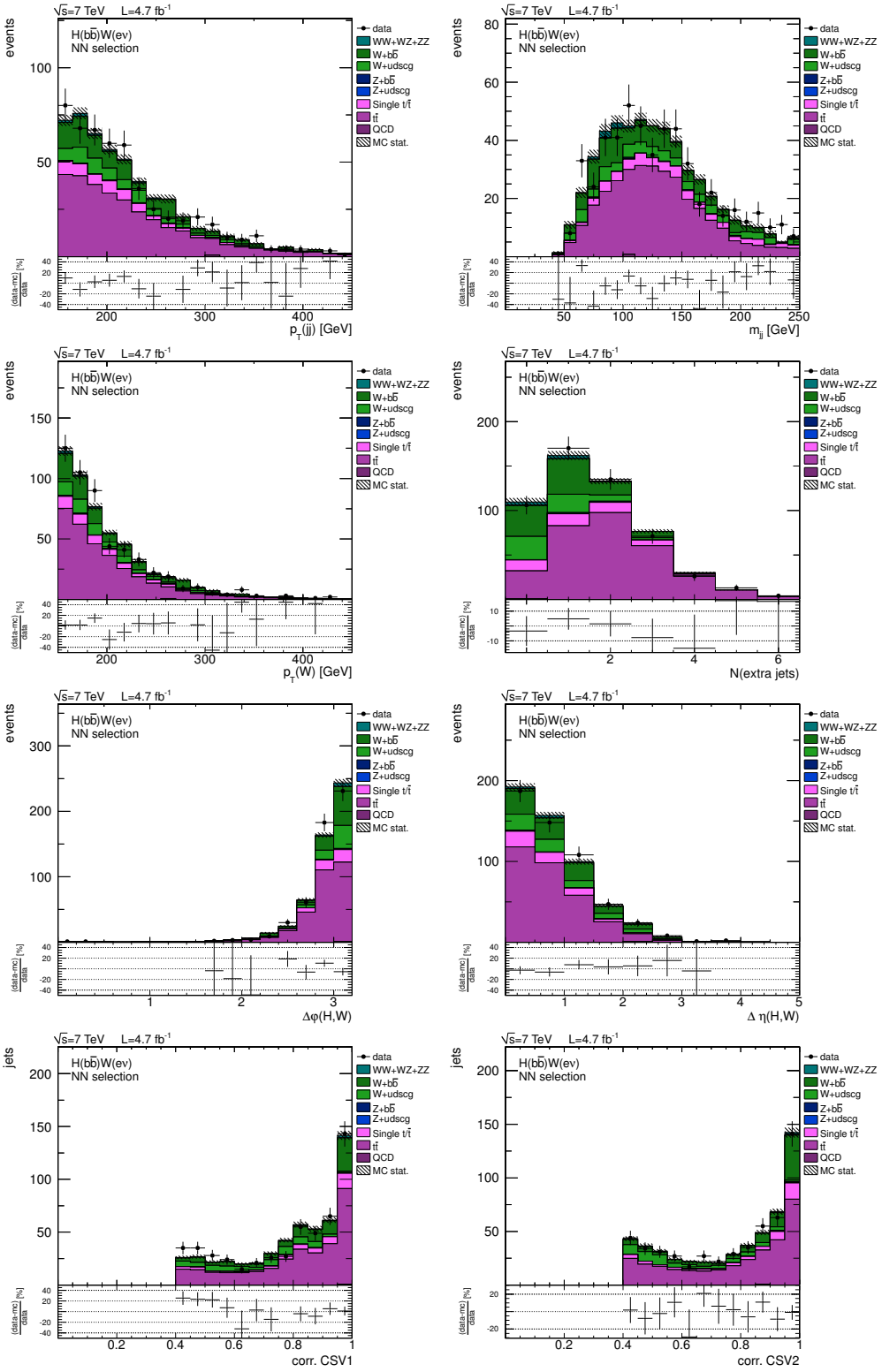


Figure 7.18.: Distributions of NN input variables in the electron channel after the NN training selection described in the first part of Table 7.2. The number of simulated events is normalized to the number of events in data.

Table 7.9.: Summary of the NN input variables. They have been identified as providing the most signal vs. background discrimination power and stable training results. The variables are sorted according to their importance to the NN classification as determined by NeuroBayes.

Ranking	Variable
1.	N(extra jets)
2.	m_{jj}
3.	corr. CSV2
4.	$p_{\mathrm{T}}(W)$
5.	corr. CSV1
6.	$\Delta\eta(H,W)$
7.	$\Delta\varphi(H,W)$
8.	$p_{\mathrm{T},jj}$

provided. In the m_{jj} analysis these efficiencies are taken into account as event weights which are applied to simulation events to correct for the measured differences with respect to data. For the NN analysis however, this information is needed not for a set of discrete working points but continuously for the whole range of possible discriminator outputs. This is achieved by the following recommendation of the BTV group to evaluate a reshaping function $f(x)$, which is based on interpolations between the measurements at discrete working points [155].

Figures 7.17 - 7.18 show the data and MC distributions for the NN input variables in the muon and electron channel, respectively, after applying the NN training selection. In these distributions the number of events in simulation is normalized to the number of events observed in data to facilitate the comparison of shapes. Good agreement within the statistical uncertainties is observed in all distributions. In particular, data and simulation agree very well for the CSV discriminator distributions after applying the above described corrections to the simulated events.

Twelve Neural Networks are trained on MC simulation, one for each trial Higgs boson mass point in the muon and electron channel, in order to maximize the possible discrimination power. In each training, the simulated background samples are weighted relative to their SM prediction. The signal sample is reweighted such that the numbers of background and signal events in the NN training are roughly the same.

Figure 7.19 shows an exemplary NN training output for the muon channel and a generated Higgs boson mass of $m(H) = 125\,\mathrm{GeV}$. The left distribution describes background and signal events as a function of the $\mathrm{NN_{out}}$ discriminator. An excellent discrimination behaviour is observed. The right plot illustrates how the signal purity increases linearly as a function of $\mathrm{NN_{out}}$ thresholds within statistical fluctuations. This nice linear increase is expected for a well-behaved NN training:

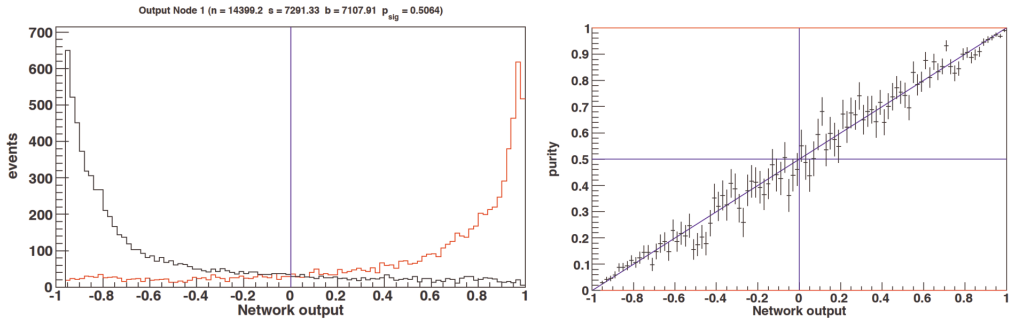


Figure 7.19.: Example of a NN training output in the muon channel for a generated Higgs boson mass of $m(H) = 125\,\text{GeV}$. The left plot shows the distribution of signal (red) and background (black) events as a function of NN_{out} and illustrates the excellent discrimination power. The right plot describes the linear rise of the signal purity as a function of increasing requirements on the NN_{out} discriminator and is an indicator for the "goodness" of the training: deviations from the diagonal would suggest problems during the training process. These two exemplary distributions are characteristic for all twelve NN trainings performed in this part of the analysis.

A requirement of $\text{NN}_{\text{out}} > 0$ for example, which coincides with the Neural Network being indecisive whether events with this discriminator are to be classified background- or signal-like, leads by construction to a signal purity of 50%.

The NN output distributions of the other 11 trained NNs show very similar characteristics as the one presented here, both in the muon and in the electron channel.

Neural Network Validation

The validation of the Neural Network trainings and the cross-checks to exclude chances of overtraining include two components: cross-validation of the NN trainings on a dedicated test sub-sample and application of the NN trainings in the three background control regions previously defined.

For the cross-validation, the training samples are sub-divided into two parts. 80% of the events are used for the NN training, while the remaining 20% are evaluated in-situ in each of the training iterations. The classification errors of the dedicated training and test samples are found to be of the same order over the whole duration of the trainings and thus give no indication for possible overtraining.

An additional and completely independent cross check includes the evaluation of each trained NN in the different background control regions. From first principles, the NNs are not expected to fully describe these regions, because they were trained in a different phase space and are thus based on different sets of training information. However, good reproduction of the classification in simulated events is observed when applying the NN to collision data is found for all networks. This is particularly astonishing for the $\text{W}+\text{b}\bar{\text{b}}$ control region, where the NNs are applied

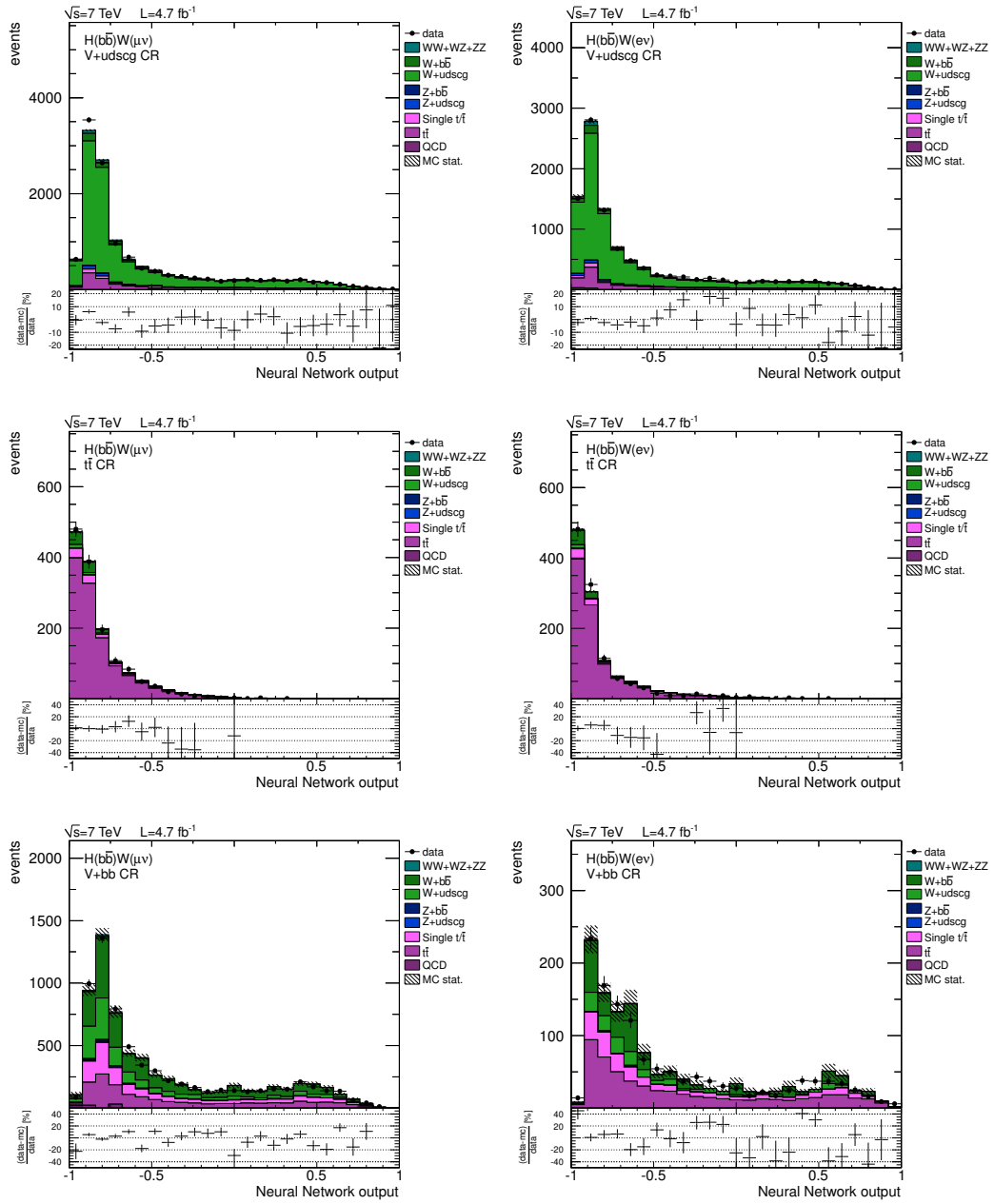


Figure 7.20.: Distributions of the NN_{out} discriminator in the different control regions, shown here exemplarily for the $NN_{out}(125)$ networks. The left row of plots shows the NN discriminator in the $W+udscg$, $t\bar{t}$ and $W+b\bar{b}$ control regions, respectively. The number of simulated events is normalized to the number of events observed in data to facilitate the comparison of shapes. The right row shows the same distributions in the electron channel for reference. The characteristic shape of each background in simulated events is fully reproduced in the data, giving confidence that the NNs are neither overtrained nor biased in events from collision data.

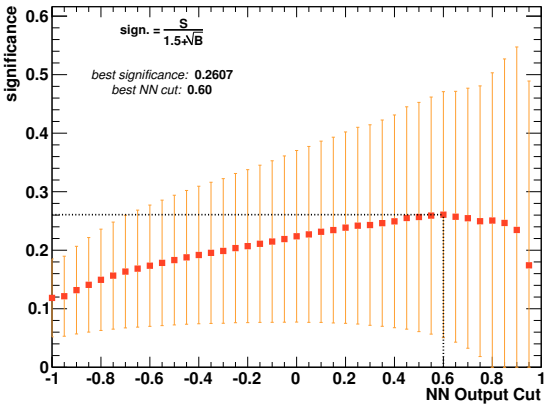


Figure 7.21.: Illustration of the $\mathrm{NN_{out}}$ cut optimization. For each trained NN the optimal discriminator requirement is calculated by maximizing the expected significance using the figure of merit defined in (7.2). This example shows the optimization of the $\mathrm{NN_{out}}(125)$ network in the muon channel.

to a low-p_{T} regime sample after being trained on a sample requiring a kinematic boost.

Figure 7.20 exemplarily shows these control distributions in the muon channel (left) and the electron channel (right) for the $\mathrm{NN_{out}}(125)$ networks, corresponding to a simulated signal MC sample with a generated Higgs boson mass of 125 GeV. The distributions from simulated events are again normalized to the event yield observed in data.

The positive cross-validation and the good reproduction of the NNs in the control regions gives sufficient confidence to apply them also in the signal region.

Neural Network Results

The final evaluation of the NN analysis corresponds to the m_{jj} analysis. Each NN discriminator cut is optimized individually using the figure of merit defined in equation (7.2), as illustrated in Figure 7.21. The final requirements on the different discriminators vary between $0.55 - 0.80$ and are summarized for each network in the second part of Table 7.2.

Table 7.10 gives the muon channel overview over the predicted signal and background yields for the different trial Higgs boson masses, as well as the observed event yield in data, after all selection criteria. The corresponding significance values given in the table are based on the signal and background yields as expected from simulation. Table 7.11 provides the same information for the electron channel.

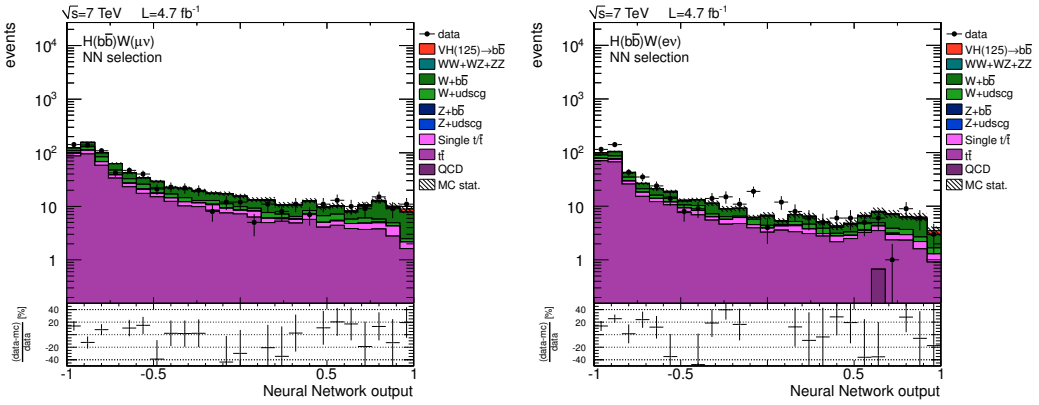
In terms of expected significance, improvements are found with respect to the same values for the m_{jj} analysis given in Tables 7.7 and 7.8. For the muon channel,

Table 7.10.: Final event yield of the NN analysis in the muon channel. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	3.82	3.01	3.07	3.70	1.38	0.86
W + $b\bar{b}$	24.04	20.27	18.86	22.38	15.20	8.79
Z + jets	0.60	0.31	0.13	0.30	0.30	0.20
$t\bar{t}$	14.45	12.72	12.03	15.73	9.47	7.06
Single top	4.76	3.97	3.82	5.26	3.19	2.05
VV	2.82	1.85	1.53	1.51	0.48	0.32
QCD	0.00	0.00	0.00	0.00	0.00	0.00
Background	50.49	42.14	39.43	48.88	30.0248	19.29
Signal	3.45	2.92	2.49	2.21	1.54	1.02
Data	61	60	51	54	33	25
Significance	0.40	0.37	0.32	0.26	0.22	0.17

Table 7.11.: Final event yield of the NN analysis in the electron channel. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	1.50	1.33	2.85	1.75	1.90	1.39
W + $b\bar{b}$	8.58	6.96	12.24	7.54	10.18	6.80
Z + jets	0.17	0.17	0.28	0.00	0.12	0.00
$t\bar{t}$	5.02	4.22	10.34	5.91	7.50	5.65
Single top	1.65	1.75	2.49	1.84	2.29	1.97
VV	1.07	0.59	0.94	0.38	0.31	0.28
QCD	0.67	0.00	0.67	0.00	0.00	0.00
Background	18.65	15.02	29.81	17.44	22.30	16.09
Signal	1.92	1.53	1.79	1.21	1.09	0.73
Data	14	11	25	18	21	14
Significance	0.33	0.28	0.26	0.21	0.18	0.13



(a) muon channel (b) electron channel

Figure 7.22.: Distributions of the NN_{out} discriminator in the NN signal region, shown here exemplarily for the $NN_{out}(125)$ networks. The left distribution (a) shows the muon channel and the right (b) shows the electron channel. In both distributions the number of simulated events is normalized to the luminosity prediction. For the final event selection the optimized cuts on the NN_{out} discriminator, given in Table 7.2, are applied.

the number of events in data exceeds the MC simulation prediction for all trial Higgs boson mass points. In the electron channel on the other hand, the reverse trend is observed. It should be stressed that the different mass points in each channel are highly correlated and systematic trends are not unexpected, because the mass windows which are corresponding to approximately one standard deviation of the mass resolution for two neighboring trial masses overlap.

The full interpretation of these values requires a coherent treatment of the involved uncertainties, which are discussed in the following sub-section.

7.5.9. Systematic Uncertainties

Several sources of uncertainties to the HW search are identified and categorized into three groups: Theory uncertainties cover effects on the expected event yields due to scale and pdf uncertainties, as well as corrections to electroweak (EWK) and QCD calculations in events with a kinematic boost. Reconstruction uncertainties refer to various systematic effects introduced by the physics object reconstruction, such as the variation of the jet energy resolution within its uncertainties as measured in Chapter 6. Sample uncertainties include the uncertainties on the measured scale factors, uncertainty estimates for the remaining backgrounds and statistical uncertainties from the finite sizes of the samples of simulated events.

Furthermore, two types of uncertainties are distinguished here: Some of the systematic effects studied merely have an effect on the rate of certain processes (e. g. luminosity measurement). The corresponding event yields are therefore directly scaled by the expected rate changes. These uncertainties are referred to as “rate

Table 7.12.: Overview of systematic uncertainties. Shown are the relative changes in event rates to the signal and to the various background processes. The values vary depending on each particular process and channel.

Uncertainty	Rate change
Luminosity	4.5%
Signal cross section	4%
Theory correction p_{T}-boost (EWK)	10%
Theory correction p_{T}-boost (QCD)	10%
Lepton efficiency and triggers (per lepton)	2%
Jet Energy Scale	up to 10%
Jet Energy Resolution	up to 10%
Unclustered $\not{E}_{\mathrm{T}}$	3%
b-tagging	up to 15%
Background scale factors	$10 - 20\%$
Single top and diboson samples	30%
Monte Carlo statistics (per sample)	up to 30%

uncertainties" in the following. The other uncertainty type is called "shape uncertainty" and it includes systematic effects which change each event individually (e. g. a jet-by-jet variation of the jet energies). They are evaluated by constructing systematically modified samples of events prior to a re-application of the selection criteria in the case of the m_{jj} analysis or a re-application (not training) of the Neural Network with a subsequent cut on the discriminator in the case of the NN analysis.

All of these uncertainties, which are discussed in detail in the following and which are summarized in Table 7.12, are included in the statistical inference and calculation of upper exclusion limits as nuisance parameters.

Luminosity and Theory Uncertainties

The integrated luminosity is measured centrally for the LHC experiments. In this analysis the luminosity measurement is used to normalize the samples of simulated events to prediction. The uncertainty on the LHC luminosity measurement amounts to 4.5% [167]. A rate uncertainty, meaning a global variation of all simulated event yields, of $\pm 4.5\%$ is thus included in the exclusion limit calculation.

The HW signal cross sections from Higgsstrahlung production are calculated at next-to-next-to-leading order accuracy. The total uncertainty on this theory calculation is estimated to be of the order of 4% [29], including the effects of scale and pdf variations.

The HW search presented in this thesis is performed in a kinematically boosted

regime, requiring high transverse momenta of the Higgs boson. Potential differences between the theoretical p_{T} spectrum of the Higgs boson used in Monte Carlo generators and of Higgs events in actual data could introduce systematic effects in the signal acceptance. Two calculations are available that estimate possible electroweak radiative [168, 169] and QCD [170] corrections to VH production in the boosted regime. The estimated effect to HW signal events from the electroweak corrections for a boost of 150 GeV is approximately 10%. The QCD corrections account for another estimated 10% uncertainty.

Reconstruction Uncertainties

The muon and electron efficiencies for trigger, reconstruction and identification are determined in data using tag-and-probe techniques and are provided centrally from the Hbb group. The uncertainties on the corresponding scale factors are determined from the statistical errors of these efficiencies and are reported to lie within 2% [155] for each, muons and electrons.

The uncertainties on the jet energy scale (JES) are provided as functions of jet p_{T} and η by the CMS JetMET physics object group [114]. In this analysis, they are evaluated on a jet-by-jet basis, scaling the p_{T} of each jet and creating new systematic samples of simulated events according to up and down variations of the JES. Consequently, the JES uncertainty is introduced into the statistical inference as a shape uncertainty. This implies that the change in event yield is not one single number, but varies for each channel and process. The overall change in event yield due to the JES variation in the m_{jj} analysis is found to be approximately $7-10\%$ for the main processes. In the NN analysis this effect is reduced to about $3-6\%$. This increased robustness of the NN analysis with respect to the m_{jj} approach can be explained by the combination of several variables into one discriminator. Systematic variations in one variable often leave other uncorrelated input variables unchanged and thus have a reduced impact on the final event selection.

For the jet energy resolution, recently updated uncertainty values based on 2011 collision data are taken with respect to the values presented in Chapter 6 for the systematic variation [171]: the p_{T} resolution of central jets lying within $|\eta| < 1.5$ is varied by 10% using the previously described resolution scaling technique, for jets within $1.5 \leq |\eta| < 2.5$ an uncertainty of 15% is applied and all other jets are varied by 20%. The uncertainty is included as shape uncertainty. Its net effect on the event yields is found to be roughly of the same order as the JES uncertainty, again having a larger effect in the m_{jj} analysis than in the NN analysis.

The uncertainty on the calibration of the unclustered $\not{E}_{\mathrm{T}}$, i. e. the missing transverse energy associated with particles in the event which are not clustered into jets, has been evaluated by the Hbb group. The uncertainty is determined to be 3% across all processes and channels.

Two different methods are used in this thesis to apply the data/MC scale factors for b-tagging provided by the BTV physics object group [166]. In the m_{jj} analysis, events are reweighted according to performance studies carried out for

pre-defined working points. Consequently, to evaluate the systematic uncertainty on the b-tagging, the event reweighting is repeated with scale factors shifted by their uncertainties for b-tags, which are associated with jets arising from the fragmentation of b-, and c-quarks, to create systematic b-tagging samples. The uncertainty for mistagging, which is associated to the erratic tagging of jets arising from the fragmentation of light quarks, is estimated independently by shifting the corresponding mistag scale factors by their uncertainties to create systematic mistag samples. The rate change observed from the b-tagging and mistagging variations is found to be up to 15%, depending on the process. For the NN analysis the CSV reshaping approach is used. To estimate the uncertainty the CSV reshaping function is rederived for systematically shifted b-tagging scale factors and accordingly for the shifted mistagging scale factors. These shifted functions are then used consistently as inputs to the NN to create systematic templates for the b-tagging and mistagging uncertainties.

Signal and Background Sample Uncertainties

The uncertainties on the dominant backgrounds in the signal region are estimated directly from data. For the W + udscg jets background contribution the total uncertainty on the scale factor is taken as 15%, including the error from the maximum likelihood fit given in Table 7.6 and the corresponding systematic uncertainty shown in Figure 7.15. The total $t\bar{t}$ scale factor is found to be slightly smaller and amounts to 10%. The dominating background contribution in the signal region comes from $W + b\bar{b}$ events. This background is hard to constrain in data and consequently the total uncertainty on its scaling factor is larger with respect to the others and estimated to be 20%. Z + jets events are only a minor background to the HW search. The uncertainty on the Z + jets event yield is estimated to be of the same order as W + jets events and is taken to be 20%. This coincides with the uncertainty estimate for $Z + b\bar{b}$ events reported in related searches for Higgs boson production in association with a Z boson [155].

Additional background contributions from single top and diboson production are taken exclusively from simulation. The rate uncertainty on simulated events from single top production is estimated as 30% which corresponds approximately to the uncertainty of the CMS cross section measurement [172]. The uncertainty on the number of events from diboson production is estimated to be similar and set to 30% as well.

An important contribution to the uncertainties comes from the finite size of the samples of simulated events which were available for this analysis. Therefore, each nominal template is shifted according to its statistical bin uncertainties in order to obtain a set of systematic templates describing the statistical uncertainties of the MC samples. The observed change in event yields ranges from 3% in the case of the signal samples to up to 30% for some background templates, depending on channel and process.

All of these uncertainties are taken into account in the statistical evaluation of the Higgs boson search presented in the following.

7.5.10. Statistical Evaluation and Results

The final event yields of the m_{jj} analysis are given in Table 7.7 for the muon channel and in Table 7.8 for the electron channel. For the NN analysis the corresponding numbers are given in Table 7.10 and in Table 7.11, respectively.

None of the channels indicate a significant excess of signal-like events. Therefore, 95% C.L. CL_s exclusion limits are calculated relative to the Standard Model HW predictions based on the statistical prescription given in Section 5.2. The uncertainties described in Section 7.5.9 are fully included in this calculation as nuisance parameters of the statistical model.

The exclusion limit calculation is performed seperataly for each analysis channel and then also for the combinations of the muon and electron channels. Figure 7.23 gives an overview of the final results of the HW search using standard jets.

The top row shows the muon channel exclusion limits. Figure 7.23(a) shows the result for the NN analysis and can be directly compared to the m_{jj} result in Figure 7.23(b). In each plot, the dotted line indicates the median expected limit for the search. A comparison of both lines indicates an improved expected limit for the NN analysis with respect to the m_{jj} analysis. Table 7.13 summarizes the exact values for the expected and observed limit and can be used to quantify this improvement in the muon channel to 3 to 11% depending on the trial Higgs boson mass points. The green and yellow bands in both plots illustrate the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the expected limit, respectively. The uncertainty bands are found to be smaller for the NN analysis, which, as previously mentioned, shows that the neural networks are less sensitive to systematic variations than the m_{jj} analysis with its hard selection criteria. The observed limits in both analyses are found to be well covered by the uncertainty bands.

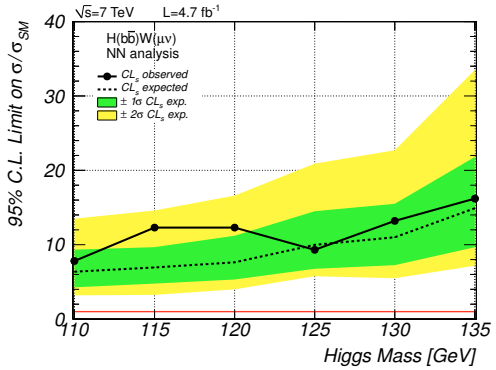
The corresponding limits for the electron channel are given in Figures 7.23(c) and 7.23(d), again for the NN and m_{jj} analyses, respectively. Their sensitivity is found to be of the same order as the limits from the muon channel. It should be noted that for two of the trial Higgs boson mass points the m_{jj} analysis is found to produce slightly better expected limits than the NN analysis. For the others, the use of the NN improves the expected limits. In general, the observed limits in the electron channel are found to be stronger with respect to the muon channel over the whole Higgs boson mass range, which can be explained with the high correlation between different trial Higgs boson mass points.

The combination of the electron and muon channels is shown in Figure 7.23(e) for the NN analysis and in Figure 7.23(f) for the m_{jj} analysis. The combination of the two channels significantly improves the expected and observed limits and also reduces the size of the uncertainty bands. It should be noted that fluctuations in the observed limits of the two NN analyses are smoothened out in the combination.

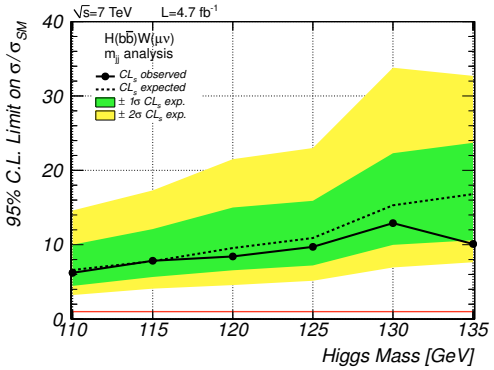
In conclusion for the HW search using standard jets the combined result for the muon and the electron channel using the NN analysis approach gives the optimal results in terms of expected limit and width of the corresponding uncertainty bands and yields expected upper limits from 4.66 at $m(H) = 110\,\text{GeV}$ to 10.9 at $m(H) = 135\,\text{GeV}$. The corresponding observed limits are found to be of the same order and within the expected uncertainties, ranging from values 4.45 for very low Higgs boson masses to 11.6 at $m(H) = 135\,\text{GeV}$.

Table 7.13.: Summary of CL_s limits in the analysis using standard jets. For each channel and the combination the individual CL_s expected and observed values are shown for the NN analysis and the m_{jj} analysis.

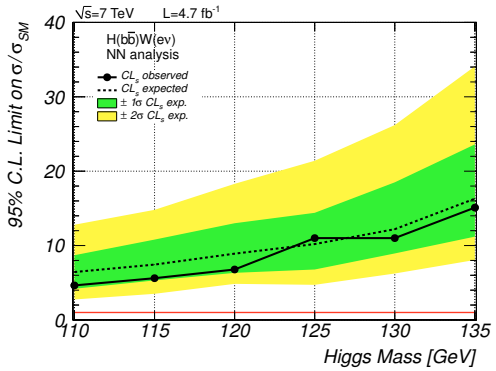
$m(H)$	muons				electrons				combination			
	NN		m_{jj}		NN		m_{jj}		NN		m_{jj}	
	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.
110 GeV	6.37	7.81	6.60	6.22	6.44	4.64	6.67	3.70	4.66	4.45	4.47	3.08
115 GeV	6.94	12.3	7.78	7.84	7.44	5.61	6.59	5.28	5.05	6.33	5.53	5.09
120 GeV	7.63	12.3	9.55	8.42	8.91	6.78	10.4	5.66	6.13	6.65	6.91	5.78
125 GeV	9.98	9.31	10.9	9.71	10.2	11.0	10.5	6.32	7.22	7.49	7.44	5.18
130 GeV	11.0	13.2	15.3	12.9	12.2	11.0	14.7	9.12	8.94	8.95	10.7	8.88
135 GeV	14.9	16.2	16.8	10.1	16.3	15.1	15.2	8.09	10.9	11.6	11.0	6.60



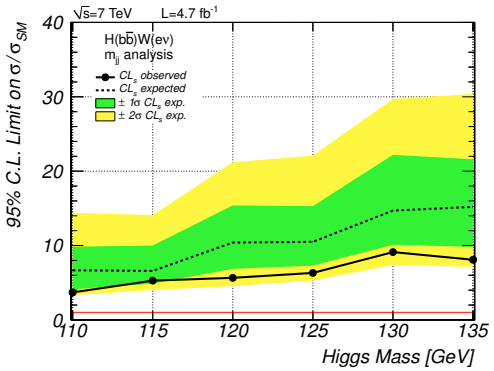
(a) Muon channel - NN analysis



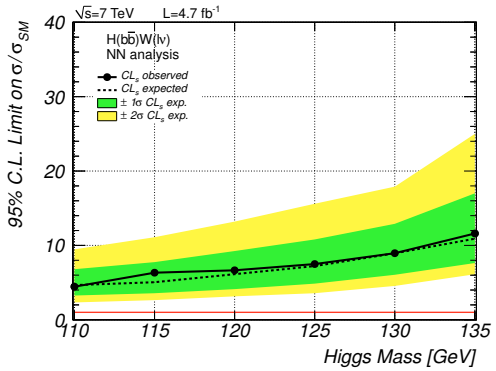
(b) Muon channel - m_{jj} analysis



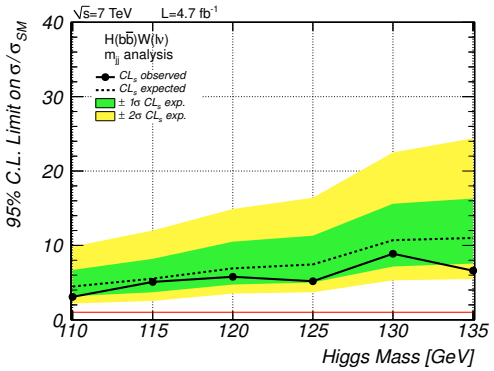
(c) Electron channel - NN analysis



(d) Electron channel - m_{jj} analysis



(e) Combination - NN analysis



(f) Combination - m_{jj} analysis

Figure 7.23.: CL_s exclusion limits using standard jets, for the NN analysis (left) and the m_{jj} analysis (right). The top row shows the muon channel result, the middle row the electron channel and the bottom row gives the exclusion limits for the combined HW search. The corresponding Bayesian limits are given in Figure B.13 in the appendix for comparison.

7.6. Fat-, Sub- and Filter-Jet Reconstruction Analysis

After the presentation of the HW search using the default CMS jet reconstruction one important question remains: Can the dedicated sub-/filterjet algorithm [111] originally proposed for this boosted Higgs boson analysis further improve the sensitivity of the analysis?

In order to address this question the standard analysis is performed again based on the alternative jet reconstruction.

In contrast to the previous section, the focus changes in the following from a detailed description of each individual analysis ingredient to a description of the differences when using sub- and filterjets. Furthermore, a particular emphasis is put on the Neural Network approach to allow for a comparison to the optimal result found in the standard analysis. The cut-based m_{jj} approach in the SJF analysis is included merely as a sanity check.

7.6.1. Sub- & Filterjets

The SJF analysis employs sub- and filterjets (see also Section 4.3.7) which are reconstructed from fat jets with a clustering parameter of $R = \pi/2$, the allowed maximum in the FastJet package. The motivation for choosing this large clustering parameter with respect to the theory proposal ($R = 1.2$) is given by the boost requirement of 150 GeV, which is suggested by the optimization of the standard analysis. Figure 7.3 illustrates the correlation between the opening angle of the two b-quarks and the transverse momentum of the Higgs boson and shows that the allowed maximum of $R = \pi/2$ is a good fit to capture the $H \to b\bar{b}$ decay in this boosted regime.

The clustering of additional energy, e. g. from pile-up, due to the large clustering parameter is addressed by the dynamic $\Delta R(j_1^{\text{sub}}, j_2^{\text{sub}})$ selection for high boosts and by the small jet radii of the filterjets for arbitrary boosts: one of the fundamental intrinsic properties of the SJF algorithm, as described in Section 4.3.7, is the dynamic choice of the actual $\Delta R(j_1^{\text{sub}}, j_2^{\text{sub}})$ based on the kinematics of the reconstructed objects. This means that the actual separation is $\Delta R(j_1^{\text{sub}}, j_2^{\text{sub}}) < \pi/2$ in most of the events. Filterjets are reclustered with small jet radii defined by $R_{\text{flt}} = \min(0.3, \Delta R(j_1, j_2)/2)$.

The SJF algorithm is by construction enabled to resolve two jets with a small opening angle beyond the limitations of a standard jet algorithm with $R = 0.5$. It should be noted however that only a very small fraction of signal events, typically those with boosts larger than 600 GeV (Fig. 7.3), have jets from the Higgs decay so close together that they will be merged by standard jet reconstruction. Moreover, in the electron channel, these events are likely to be missing from the SJF sample as well due to the trigger requirement of two central (standard) $R = 0.5$ anti-k_T jets.

Both sub- and filterjets are corrected with the L1FastJet pile-up correction based on the average pile-up energy density ρ. Since there are no dedicated jet energy

corrections for the SJF jets, the CMS default L2Relative and L3Absolute corrections for anti-k_T ($R = 0.5$) (AK5) jets are applied to the sub- and filterjets. In addition, for data events the obligatory L2L3Residual correction (AK5) is applied. Correcting the energy scale of SJF jets with AK5 correction parameters systematically leads to an overcorrection of subjets and an underestimation of the energy of filterjets as a result of their respective jet areas. The effect of the miscalibration on this sensitivity study, however, is small, since the corrections are applied to data and simulation consistently and the goal is merely to separate signal from background and not to measure e. g. the actual mass.

In analogy to the corrections of the jet energy scale, the sub- and filterjets in the samples of simulated events are bias-corrected for jet energy resolution using the prescription for the standard jets as described in Section 7.5.1.

The CSV b-tagger is applied to the two leading filterjets using the standard jet working points and efficiencies provided by the BTV physics object group. It is reasonable to assume that the performance measurements of the b-tagging derived from standard jet reconstructions are applicable to filter jets, since the algorithm used to match charged tracks to the jet is identical.

In summary, the points above describe the *basic working hypothesis* of the SJF analysis: in the absence of dedicated correction parameters (JES, JER, b-tagging) for sub- and filterjets the corresponding standard jet parameters are used assuming the effect on this study of a possible sensitivity improvement to be minor.

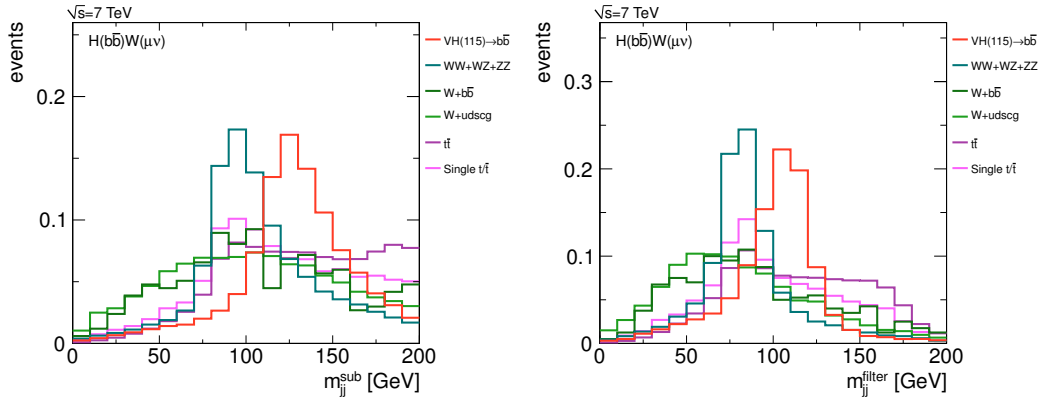
7.6.2. Higgs Boson Candidate Reconstruction

The choice on how to reconstruct the Higgs boson candidate is simplified with respect to the standard analysis. A necessary requirement is the reconstruction of a fat jet and the associated identification of jet substructure by the SJF algorithm. Furthermore, to assert the quality of the substructure reconstruction, the two leading filterjets are, in analogy to the standard jet analysis, required to be central ($|\eta| < 2.4$) and have to exceed a p_T threshold of 30 GeV. If these basic requirements are fulfilled in an event, both the di-subjet system and the tri-filterjet system describe the Higgs boson candidate.

For the m_{jj} approach in the SJF analysis, the tri-filterjet system is choosen as Higgs boson candidate, as it yields the better mass resolution and, due to the small filterjet radii, is found to be less impacted by pile-up. It should be noted that the latter will be of increasing importance for the analysis of future LHC collision data, where the number of additional pile-up events is expected to increase.

In the NN analysis one does not have to pick one of the two, but can use the information of both jet types for discrimination.

An example of the reconstructed Higgs boson candidate masses based on sub- and filterjets is given in Figure 7.24 in red for signal events with a generated Higgs boson mass of 115 GeV. The mass resolution for the reconstructed tri-filterjet system after applying the signal selection criteria described in Table 7.14 is found to be approximately 12%. This is comparable to the mass resolution of the standard jet



(a) (b)

Figure 7.24.: Distribution of Higgs boson mass distributions in the SJF analysis for simulated signal and background events. (a) shows the invariant mass of the di-subjet system and (b) the mass of the tri-filterjet system. The sizable reduction in jet area moving from sub- to filterjets results in the shift of mass spectra. The tri-filterjet mass is expected to yield optimal mass resolution and pile-up resiliency. The distributions are shown after the event pre-selection for muons described in Section 7.4 and a p_T boost requirement of 150 GeV on the W boson in order to assert that the SJF algorithm runs on boosted objects. The distributions are normalized to unit area. Comparison with the corresponding anti-k_T jet distribution shown in Figure 7.8 exhibits one key advantage of the properties inherent to the SJF algorithm. The backgrounds are shaped differently with respect to the standard analysis and systematically shifted to lower invariant masses.

approach. Figure 7.24 also illustrates the change of the reconstructed signal mass peak when moving from sub- to filterjets. This change is a result of the application of the standard jet correction parameters and the difference in jet area between the two SJF jet types. It is taken into account when defining the selection criteria in the following.

7.6.3. Signal and Background Characteristics

The general validity of the considerations on the characteristics of signal events and the different backgrounds does not change by switching the jet reconstruction algorithm.

It should be emphasized however that the sub- and filterjet algorithm is found to change the shape of the backgrounds, in particular V + jets, in an advantageous way. Comparison of the background shapes of the invariant mass distributions in the standard analysis given in Figure 7.8 with the corresponding sub- and filterjet distributions depicted in Figure 7.24 illustrates this: the invariant masses of the backgrounds systematically shift to lower values.

7.6.4. Definition of Variables

In addition to the variables already defined for the standard analysis in Section 7.5.4 it is useful to introduce the following variables for the SJF analysis:

- m_{jj}^{flt}: mass of the Higgs boson candidate built from the two or three filterjets arising from the two b-quarks and leading order radiation.
- m_{jj}^{sub}: mass of the Higgs boson candidate built from the two subjets.
- $p_{\mathrm{T},jj}^{\mathrm{flt}}$: p_{T} of the tri-filterjet Higgs boson candidate.
- $p_{\mathrm{T},jj}^{\mathrm{sub}}$: p_{T} of the di-subjet Higgs boson candidate.
- CSV: continuous output of the Combined Secondary Vertex b-tagger for the first and second filterjet in the event.
- $\Delta R(j_1^{\mathrm{flt}}, j_2^{\mathrm{flt}})$: opening angle between the first and second filterjet.
- $\Delta R(j_1^{\mathrm{flt}}, j_3^{\mathrm{flt}})$: opening angle between the first and third filterjet.
- $\Delta R(j_1^{\mathrm{sub}}, j_2^{\mathrm{sub}})$: opening angle between the two subjets.

7.6.5. Event Selection

m_{jj} Analysis

The event selection of the m_{jj} analysis using the SJF jet reconstruction is not optimized with respect to the significance defined in (7.2). It is instead chosen as close as possible to the standard jet analysis as it only serves as a basic sanity check to the NN analysis in the following.

As motivated in Section 7.6.2, the tri-filterjet system is chosen to describe the Higgs boson candidate in the SJF analysis. The selection criteria from the standard jet analysis given in Table 7.1 are translated into selection criteria for the SJF analysis given in Table 7.14 by introducing filterjet variables instead of the standard jet variables. The dedicated mass window cuts for each trial Higgs boson mass point are re-defined to account for the offset of the filterjet energy scale: corrections derived for standard anti-$\mathrm{k_T}$ ($R = 0.5$) jets are applied to filterjets which are clustered with $R \leq 0.3$ and therefore carry systematically less energy. To account for this, the mass windows are adjusted according to the filterjet mass distribution found in simulation.

NN Analysis

The event selection for the NN analysis changes only minimally with respect to the standard analysis. The first part of Table 7.15 shows the selection criteria for events used in the NN training ("training region"). Again, the standard jet variables are substituted with the corresponding filterjet variables in order to define a large

Table 7.14.: Selection criteria for the final SJF m_{jj} analysis for the muon and electron channel after the event pre-selection specified in Section 7.4. The selection criteria are defined identical to the standard jet analysis, with the exception that the standard jet variables are exchanged with the corresponding filterjet variables.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}^{flt}	GeV	> 0	> 0
$p_{\text{T},jj}^{\text{flt}}$	GeV	> 165	> 165
$p_{\text{T}}(W)$	GeV	> 160	> 160
$\max(\text{CSV1}, \text{CSV2})$	–	> 0.898	> 0.898
$\min(\text{CSV1}, \text{CSV2})$	–	> 0.4	> 0.4
$\Delta\varphi(H^{\text{flt}}, W)$	–	> 2.95	> 2.95
$N(\text{extra jets})$	–	0	0
$\not{E}_{\text{T}}$	GeV	–	35
$m_{jj}^{\text{flt}}(110)$	GeV	$93 - 120$	$93 - 120$
$m_{jj}^{\text{flt}}(115)$	GeV	$96 - 125$	$96 - 125$
$m_{jj}^{\text{flt}}(120)$	GeV	$100 - 130$	$100 - 130$
$m_{jj}^{\text{flt}}(125)$	GeV	$104 - 134$	$104 - 134$
$m_{jj}^{\text{flt}}(130)$	GeV	$108 - 138$	$108 - 138$
$m_{jj}^{\text{flt}}(135)$	GeV	$111 - 142$	$111 - 142$

Table 7.15.: Selection criteria for the SJF NN analysis training region after the pre-selection specified in Section 7.4. The first part describes the NN training selection. The second part summarizes the final re-optimized selection criteria on the SJF NN discriminator.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}^{flt}	GeV	$(0, 250)$	$(0, 250)$
$p_{\text{T},jj}^{\text{flt}}$	GeV	> 150	> 150
$p_{\text{T}}(W)$	GeV	> 150	> 150
$\max(\text{CSV1}, \text{CSV2})$	–	> 0.4	> 0.4
$\min(\text{CSV1}, \text{CSV2})$	–	> 0.4	> 0.4
$\not{E}_{\text{T}}$	GeV	–	35
$\text{NN}_{\text{out}}(110)$	–	> 0.60	> 0.65
$\text{NN}_{\text{out}}(115)$	–	> 0.60	> 0.60
$\text{NN}_{\text{out}}(120)$	–	> 0.70	> 0.60
$\text{NN}_{\text{out}}(125)$	–	> 0.65	> 0.60
$\text{NN}_{\text{out}}(130)$	–	> 0.80	> 0.65
$\text{NN}_{\text{out}}(135)$	–	> 0.75	> 0.65

sample of events with signal-like characteristics. The second part summarizes the re-optimized selections for the discriminator NN_{out} of the SJF Neural Networks. Further details on the new NN topology and training are given later in Section 7.6.8.

7.6.6. SJF Algorithm Validation in Background Control Regions

A key aspect of the SJF analysis presented in this thesis is to validate the sub- and filterjet algorithm in data and simulation. For this purpose important sub- and filterjet variables are studied in background control regions which are defined in analogy to the standard analysis with the exception of the $W + b\bar{b}$ control region. It is re-defined in a kinematically boosted regime, because the SJF algorithm is specifically designed to reconstruct boosted objects and does not work properly in the low p_T regime used in the standard analysis.

W + udscg and $t\bar{t}$ Control Regions

After substituting standard jet variables with filter jet variables, the control region definitions for W + udscg and $t\bar{t}$ are identical to the standard analysis. The W + udscg control region is thus still characterized by the omission of b-tagging requirements and a $t\bar{t}$ enriched sample is obtained by reversing the requirement on the event jet multiplicity with respect to the signal selection given in Table 7.14. The detailed definitions for these two background control regions in the SJF analysis are summarized in Table 7.16.

Table 7.16.: Definition of the W + udscg jets and $t\bar{t}$ background control regions using SJF jet reconstruction. The control regions (CR) are defined in analogy to the standard analysis summarized in Table 7.3 and Table 7.4.

		W + udscg CR		$t\bar{t}$ CR	
variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}^{flt}	GeV	$(0, 250)$	$(0, 250)$	$(0, 250)$	$(0, 250)$
$p_{T,jj}^{\text{flt}}$	GeV	> 150	> 150	> 150	> 150
$p_T(W)$	GeV	> 150	> 150	> 150	> 150
max(CSV1, CSV2)	–	–	–	> 0.898	> 0.898
min(CSV1, CSV2)	–	–	–	–	–
N(extra jets)	–	< 2	< 2	> 1	> 1
$\not{E}_T$ significance	–	> 2	> 2	> 2	> 2

Figure 7.25 shows the distributions for selected filterjet variables in the W+udscg control region for the muon channel. In these distributions the scale factors from the standard analysis (discussed in the end of this sub-section) for W + udscg jets, $t\bar{t}$ and $W + b\bar{b}$ have been applied. Good agreement is found between data and simulation for all variables. Furthermore, important variable distributions for

subjets and additional topological distributions involving the tri-filterjet system, which are used in the following, are presented in Figure B.14 in the appendix. The corresponding distributions for the electron channel are given in Figures B.15 - B.16.

In order to present the electron channel as well and to show that the $t\bar{t}$ background is equally well described, Figure 7.26 shows the distributions for key filterjet variables in the $t\bar{t}$ control region for the electron channel. Again, good agreement is found between data and simulation. Figure B.19 shows additional important distributions in the electron channel and the corresponding plots for the muon channel are presented in Figures B.17 -B.18. All of them are included in the appendix.

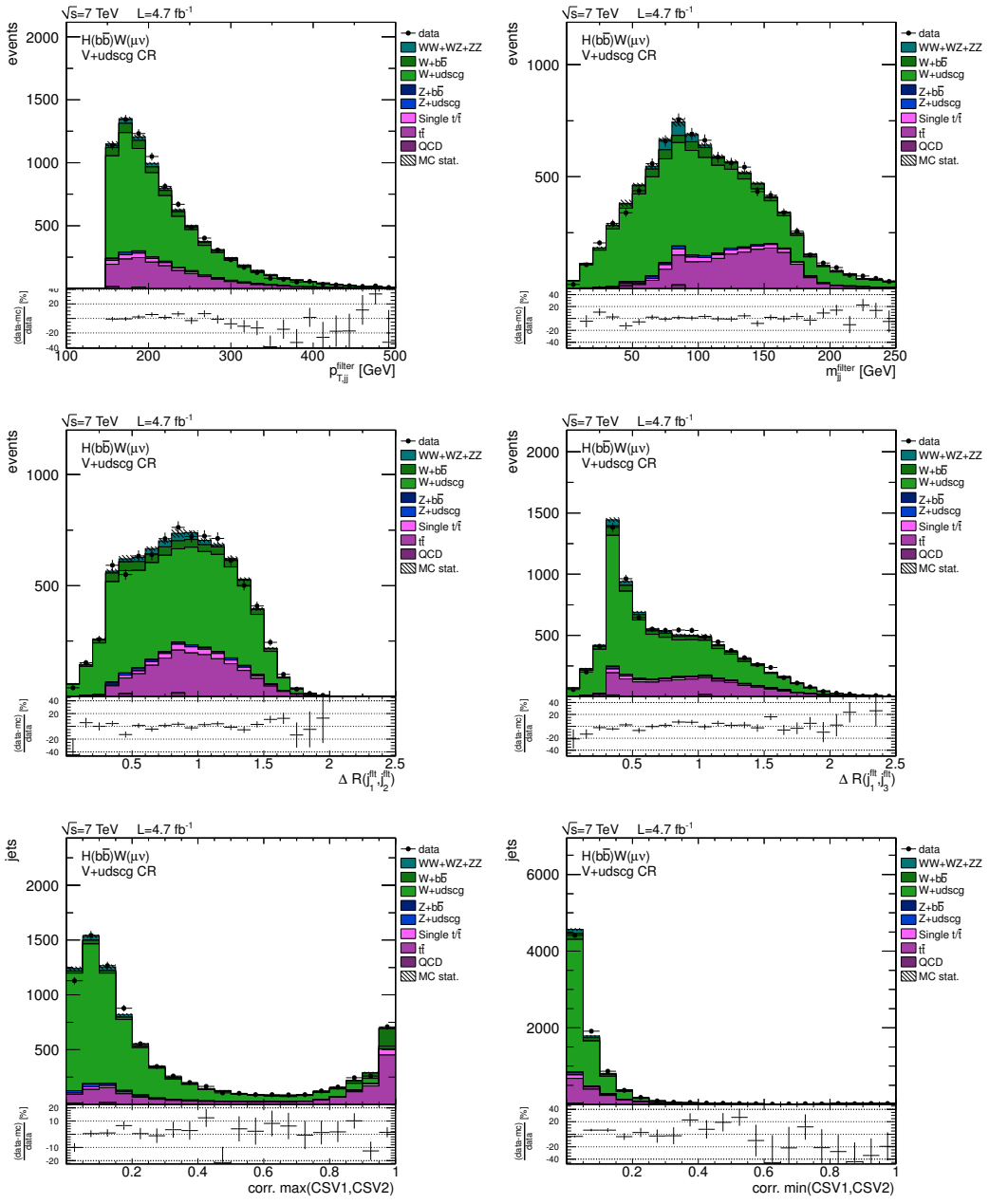


Figure 7.25.: W+udscg jets control region distributions for the muon channel using SJF jet reconstruction: the p_{T} of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data in order to facilitate the shape comparison. The same set of distributions for the electron channel is presented in Figure B.15 in the appendix.

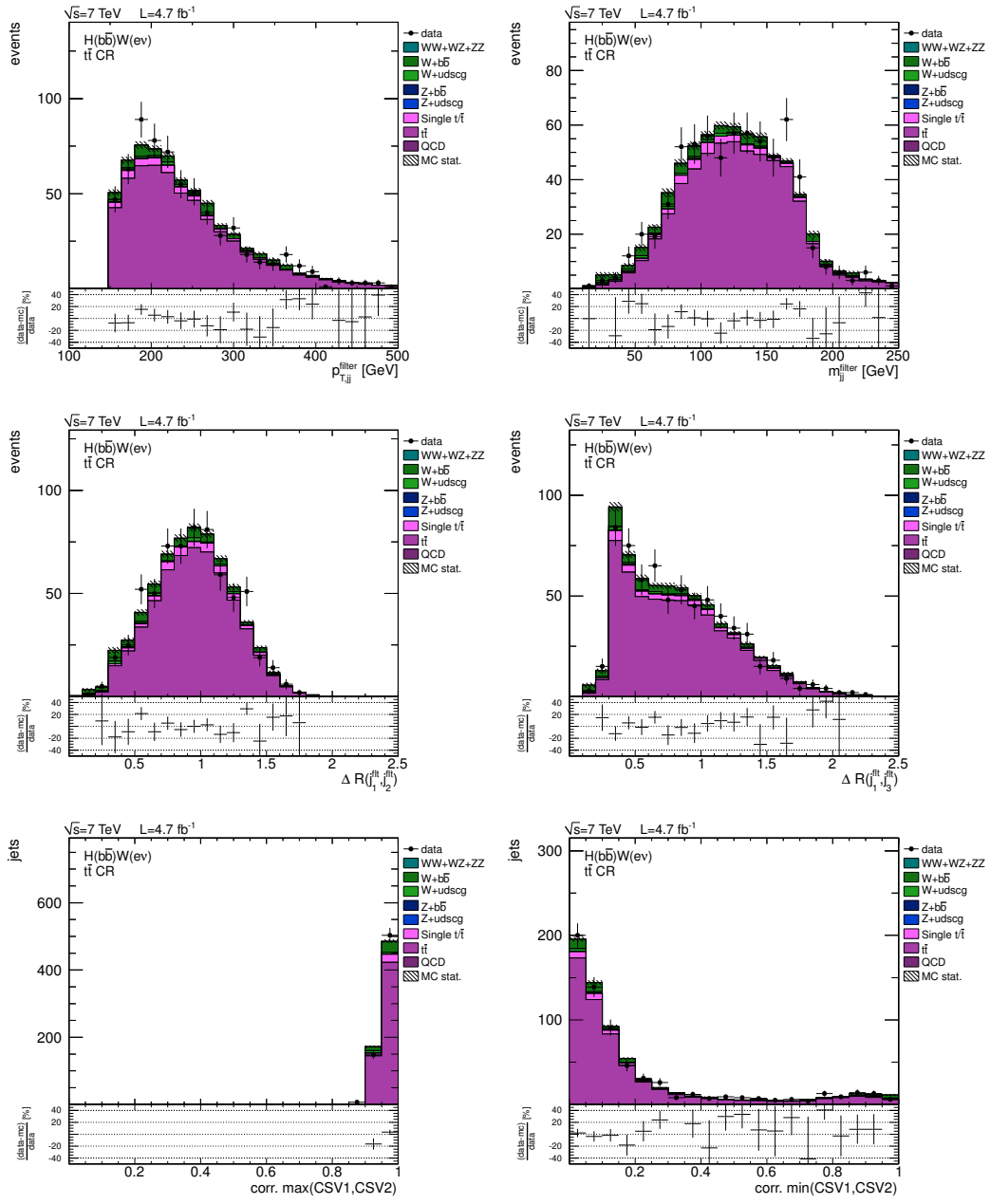


Figure 7.26.: $t\bar{t}$ control region distributions for the electron channel using SJF jet reconstruction: the p_T of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data. The same set of distributions for the muon channel is presented in Figure B.17 in the appendix.

$\mathrm{W} + \mathrm{b\bar{b}}$ Control Region

The definition of the $\mathrm{W}+\mathrm{b\bar{b}}$ background control region cannot be translated directly from the standard analysis to the SJF analysis, because the sub-/filterjet algorithm is specifically designed to be applied to boosted objects. Its behaviour is thus not well defined in the low p_T regime used in the standard analysis $\mathrm{W} + \mathrm{b\bar{b}}$ control region. In order to validate the algorithm on $\mathrm{W}+\mathrm{b\bar{b}}$ background events the control region is defined in a moderately boosted regime, where both the W boson and the tri-filterjet Higgs candidate have high transverse momenta. The corresponding p_T threshold is lowered to 110 GeV with respect to the usual boost requirement in order to maximize the number of remaining $\mathrm{W}+\mathrm{b\bar{b}}$ events. Overlap with potential signal events is prevented by constraining the events to be in a tri-filterjet mass sideband requiring $m_{jj}^\mathrm{flt} < 100\,\mathrm{GeV}$. Background contributions from W + udscg events are reduced by applying two b-tag requirements corresponding to the NN training selection. In addition, the $\mathrm{t\bar{t}}$ background is decreased by selecting events with less than two additional jets in the event.

The number of remaining events in the $\mathrm{W}+\mathrm{b\bar{b}}$ control region is much smaller with respect to the standard analysis, but sufficient for the validation of key variables. Furthermore, moving the $\mathrm{W}+\mathrm{b\bar{b}}$ control region into a boosted regime and requiring two b-tags has the important advantage that it allows for an independent cross check of the correct $\mathrm{W} + \mathrm{b\bar{b}}$ scale factor determination in the standard analysis.

The detailed $\mathrm{W} + \mathrm{b\bar{b}}$ control region definitions are summarized in Table 7.17.

Table 7.17.: Definition of the $\mathrm{W} + \mathrm{b\bar{b}}$ jets background control region using SJF jet reconstruction. In order to test the sub-/filterjet reconstruction on rapidly moving objects a moderately boosted mass sideband control region is defined in contrast to the low p_T regime used in the standard jet analysis. Furthermore, two relatively loose b-tags are required in analogy to the NN training selection to reduce $\mathrm{W}+$udscg contributions. The $\mathrm{t\bar{t}}$ contribution is minimized by constraining the number of additional jets in the event to be less than two.

variable	unit	$H(b\bar{b})W(\mu\nu)$	$H(b\bar{b})W(e\nu)$
m_{jj}^flt	GeV	$(0, 100)$	$(0, 100)$
$p_{\mathrm{T},jj}^\mathrm{flt}$	GeV	> 110	> 110
$p_\mathrm{T}(W)$	GeV	> 110	> 110
$\max(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.400	> 0.400
$\min(\mathrm{CSV1}, \mathrm{CSV2})$	–	> 0.400	> 0.400
N(extra jets)	–	< 2	< 2
$\not{E}_\mathrm{T}$ significance	–	> 2	> 3

Figure 7.27 shows selected filterjet variable distributions in the $\mathrm{W} + \mathrm{b\bar{b}}$ control region for the muon channel. Good agreement is found between data and simulation for all variables. Furthermore, important variable distributions for subjets and

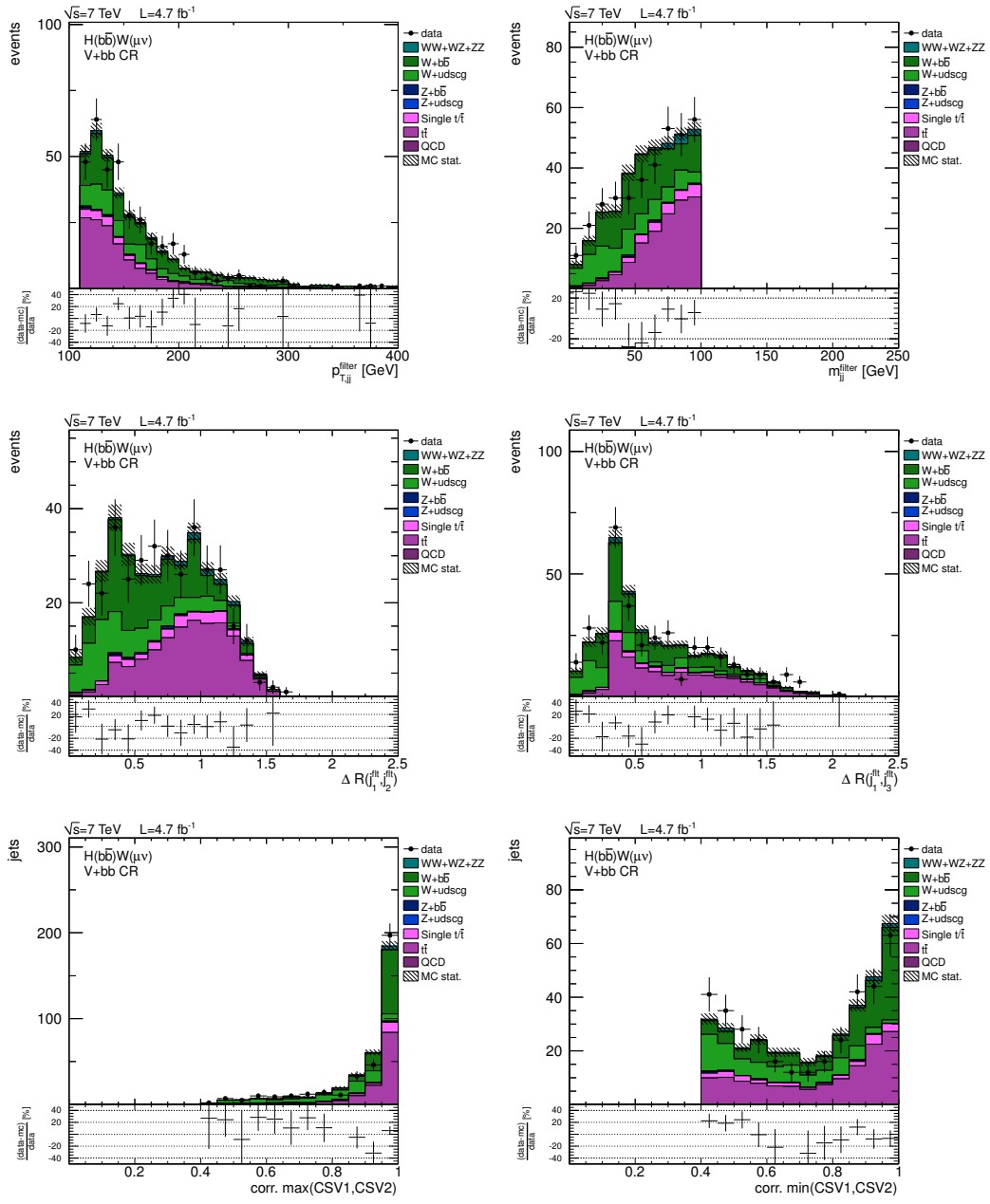
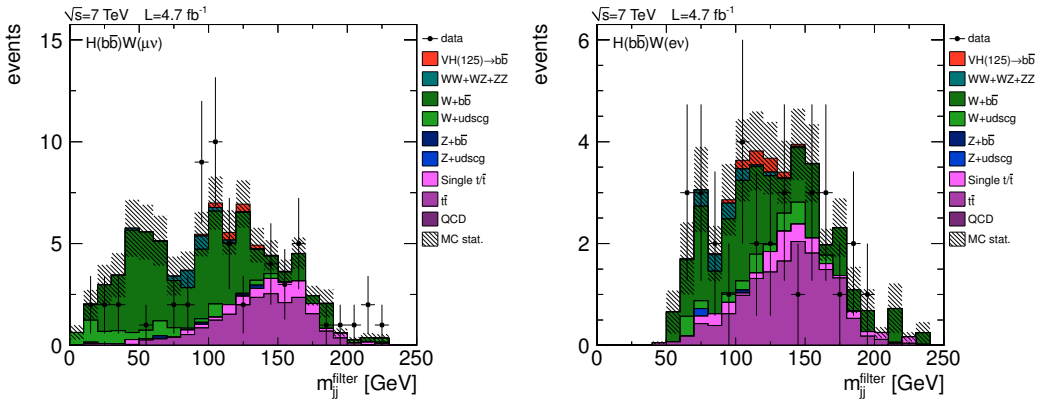


Figure 7.27.: $W + b\bar{b}$ jets control region distributions for the muon channel using SJF jet reconstruction: the p_T of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data. The same set of distributions for the electron channel is presented in Figure B.21 in the appendix.



(a) (b)

Figure 7.28.: Higgs boson candidate mass distribution in the signal region using SJF jet reconstruction for the (a) muon channel and (b) electron channel. The distributions are dominated by backgrounds. Simulated events for an expected Higgs boson signal with $m(H) = 125\,\text{GeV}$ are included in red.

additional topological variables involving the tri-filterjet system, which are used in the following, are presented in Figure B.20 in the appendix. The corresponding distributions for the electron channel are given in Figures B.21 - B.22.

Scale Factors

The background scale factors derived in the standard analysis are not expected to change with the reconstruction method. This is checked by repeating the simultaneous maximum likelihood fit to the control regions described in Section 7.5.6. The fit values are found to be compatible with the standard jet results given in Table 7.6 within the uncertainties. Therefore, the scale factors from the standard analysis are also used in the SJF analysis to be consistent in the final comparison of the methods.

7.6.7. Cut-based Analysis

The invariant tri-filterjet mass distributions in the signal region for both the muon and electron channel are presented in Figure 7.28. The distributions are dominated by background events. Simulated events for an expected Standard Model Higgs boson signal with $m(H) = 125\,\text{GeV}$ are included in red. The black shaded bands indicate the statistical uncertainty due to the small number of remaining simulated events after the signal region selection. Comparison with Figure 7.16 showing the corresponding distributions using standard jet reconstruction illustrates the previously mentioned change in background shape. Again, the total event yield in the electron channel is smaller with respect to the muon channel because of tighter

Table 7.18.: Final event yield of the m_{jj} analysis in the muon channel using SJF jet reconstruction. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	0.82	0.66	0.69	0.67	0.27	0.32
W + $b\bar{b}$	10.66	10.30	11.56	11.57	9.37	8.10
Z + jets	0.10	0.00	0.11	0.11	0.28	0.28
$t\bar{t}$	3.43	4.07	4.75	5.28	5.53	6.18
Single top	0.77	0.95	1.04	1.28	1.32	1.64
VV	0.80	0.6251	0.35	0.25	0.23	0.17
QCD	0.00	0.00	0.00	0.00	0.00	0.00
Background	16.58	16.61	18.49	19.15	16.99	16.69
Signal	1.43	1.38	1.20	1.01	0.79	0.61
Data	21	21	17	12	8	7
Significance	0.26	0.25	0.21	0.17	0.14	0.11

Table 7.19.: Final event yield of the m_{jj} analysis in the electron channel using SJF jet reconstruction. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	0.71	0.67	0.68	0.68	0.85	0.70
W + $b\bar{b}$	4.89	5.04	5.04	4.03	3.75	4.30
Z + jets	0.06	0.06	0.06	0.06	0.00	0.00
$t\bar{t}$	2.80	3.17	3.73	3.83	4.39	4.72
Single top	0.24	0.28	0.55	0.73	1.07	1.21
VV	0.41	0.37	0.33	0.26	0.16	0.12
QCD	0.00	0.00	0.00	0.00	0.00	0.00
Background	9.12	9.59	10.39	9.59	10.22	11.05
Signal	1.12	1.07	0.88	0.74	0.62	0.45
Data	6	8	8	6	7	7
Significance	0.25	0.23	0.19	0.16	0.13	0.09

online selection criteria imposed by the electron triggers and tighter offline selection criteria needed to suppress events from fake electrons.

For the final m_{jj} analysis selection, event yields expected from simulation and observed in data are evaluated in mass windows around the predicted Higgs boson signal, as described in Table 7.14. The final yields are summarized in Table 7.18 and Table 7.19 for the muon and electron channel, respectively. In addition, the tables provide the values for the expected significance based on equation (7.2). They show that the largest fraction of observed events is expected to come from $\mathrm{W} + \mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}}$ background contributions.

Comparison with the corresponding event yields based on standard jet reconstruction (Tables 7.7, 7.8) shows that the expected event yields for background and signal are slightly lower in the SJF analysis. Consequently, the expected significances are also found to be smaller. The number of observed events in data is found to be of the same order, but a slightly more pronounced deficit of events in data is observed for the high trial mass points in the muon channel.

The SJF m_{jj} analysis selection criteria have been translated directly from the standard analysis and are not re-optimized with respect to (7.2). The intention of the m_{jj} analysis using sub-/filterjets is to serve as a basic cross check before applying multivariate analysis techniques to obtain an optimized result. In this sense, the m_{jj} analysis using standard jets and the m_{jj} analysis using sub-/filterjets are found to give consistent results.

7.6.8. Neural Network Analysis

In the following the complete Neural Network analysis is performed based on the SJF jet reconstruction including new trainings and thereby optimization. Therefore, the results will allow a direct comparison of the performance of the two jet reconstruction approaches.

Neural Network Training

Several discriminating variables have been studied as inputs to the Neural Network. Based on their discrimination power determined by NeuroBayes, thirteen variables are chosen as final inputs. They are summarized in Table 7.20 according to their relative importance.

As in the standard analysis, the invariant mass of the tri-filterjet system, the number of additional central jets in the event and the value of the CSV b-tagging discriminator of the second filterjet are found to be the most important inputs to the NN. The NN training yields that the invariant masses of both the subjet and filterjet system contribute significant discrimination power. Additional tests on the correlation between sub- and filterjet information include the study of the difference between the invariant di-subjet and tri-filterjet masses. Figure 7.29 exemplarily shows such a distribution in the $\mathrm{W} + \mathrm{udscg}$ control region for the electron channel. The good agreement of this difference in simulation and data gives confidence that

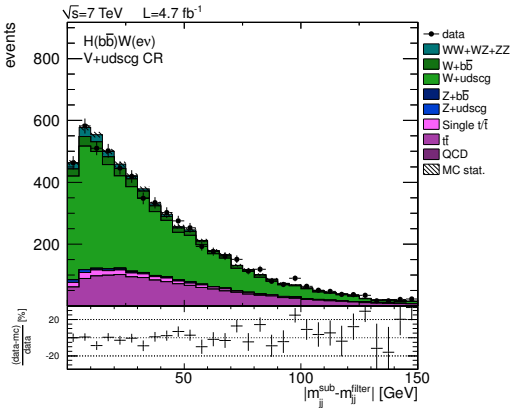


Figure 7.29.: Illustration of the correlation between sub- and filterjet information, exemplarily shown in the W + udscg control region for the electron channel. The difference of the invariant di-subjet and tri-filterjet masses in simulated events with respect to data agrees well, and is exploited in the Neural Network analysis.

the use of both sub- and filterjet information in the NN is feasible and improves the discrimination power.

The distributions of all input variables after the NN training selection given in Table 7.15 are presented in Figures B.23 - B.26 in the appendix. Good agreement between data and simulation is observed in both the muon and electron channels.

In analogy to the standard analysis, 12 NNs are trained on simulation using the input variables given in Table 7.20. The NN training distributions (discriminator and purity) are found to be very similar to the ones from the standard analysis, and are therefore not shown.

Neural Network Validation

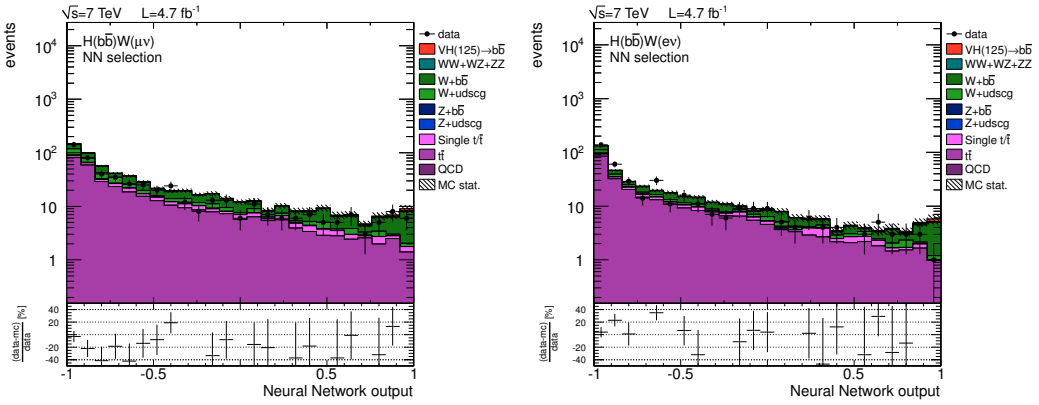
The validation of the SJF Neural Network trainings is done by repeating the full methodology described for the standard analysis. Exemplary $\mathrm{NN_{out}}$ distributions in the different control regions for both muon and electron channels are presented in Figure B.27 in the appendix for the Neural Network trainings associated with a generated Higgs boson mass of 125 GeV. The discriminator shapes are found to agree very well in data and simulation and give no indication for overtraining. The other 11 trainings lead to similar $\mathrm{NN_{out}}$ distributions in the control regions.

Neural Network Results

Figure 7.30 exemplarily shows the distribution of the $\mathrm{NN_{out}}(125)$ discrinimator after the NN selection for the muon and electron channel. In direct analogy to the standard analysis the final evaluation is based on a set of cuts which are re-optimized for each NN with respect to (7.2). Their values are summarized in the second part of Table 7.15.

Table 7.20.: Summary of the NN input variables using SJF jet reconstruction. They have been identified as providing the most signal vs. background discrimination power and stable training results. The variables are sorted according to their importance to the NN classification.

Ranking	Variable
1.	m_{jj}^{flt}
2.	$N(\text{extra jets})$
3.	corr. CSV2
4.	m_{jj}^{sub}
5.	corr. CSV1
6.	$p_{\mathrm{T}}(W)$
7.	$\Delta\varphi(H^{\mathrm{flt}}, W)$
8.	$p_{\mathrm{T},jj}^{\mathrm{sub}}$
9.	$p_{\mathrm{T},jj}^{\mathrm{flt}}$
10.	$\Delta R(j_1^{\mathrm{flt}}, j_2^{\mathrm{flt}})$
11.	$\Delta\eta(H^{\mathrm{flt}}, W)$
12.	$\Delta R(j_1^{\mathrm{sub}}, j_2^{\mathrm{sub}})$
13.	$\Delta R(j_1^{\mathrm{flt}}, j_3^{\mathrm{flt}})$



(a) muon channel

(b) electron channel

Figure 7.30.: Distributions of the $\mathrm{NN_{out}}$ discriminator in the NN signal region using SJF jet reconstruction, shown here exemplarily for the $\mathrm{NN_{out}}(125)$ networks. The left distribution (a) shows the muon channel and the right (b) shows the electron channel. In both distributions the number of simulated events is normalized to the luminosity prediction. For the final event selection the optimized cuts on the $\mathrm{NN_{out}}$ discriminator, given in Table 7.15, are applied.

Table 7.21 gives an overview of the muon channel with its predicted signal and background yields, as well as the observed event yields in data, after applying all selection criteria. The significance values are based on the numbers of signal and background, as expected from MC simulation. Table 7.22 provides the same information for the electron channel.

Comparison with the corresponding numbers of the standard analysis (Tables 7.10, 7.11) shows reduced expected background yields. The actual size of this reduction, however, depends on the mass point and channel. For signal events, the rates are observed to be slightly smaller as well. In combination, the values for the expected significances in the SJF NN analysis are slightly reduced with respect to the standard analysis for all mass points.

The event yields in data show a small deficit with respect to the expected event yields in both the muon and electron channel.

7.6.9. Systematic Uncertainties

The evaluation of systematic uncertainties is done in full analogy to Section 7.5.9 and the description of the individual contributions is therefore not repeated here. For the sub- and filterjets the evaluation of the uncertainties of the jet energy scale and resolution as well as b-tagging relies on the parameters derived using standard anti-k_T ($R = 0.5$) jets.

The sizes of the different uncertainty contributions are found to be similar to the ones in the standard analysis.

7.6.10. Statistical Evaluation and Results

For the HW search using the SJF jet reconstruction approach, the final event yields of the m_{jj} analysis are given in Table 7.18 for the muon channel and in Table 7.19 for the electron channel. The corresponding numbers for the NN analysis are given in Table 7.21 and Table 7.22, respectively.

In analogy to the standard jet analysis, 95% C.L. CL_s exclusion limits are calculated relative to the Standard Model HW predictions. Figure 7.31 illustrates the final results for the two SJF analysis approaches in each channel and for the combination.

Comparison of the expected median limits indicated with the dashed black line of the NN analysis (left row) and the m_{jj} analysis (right row) shows a significant increase of expected search sensitivity using the NN approach. The reason for this are the final signal selection criteria, which are not specifically re-optimized with respect to (7.2) for the SJF m_{jj} analysis.

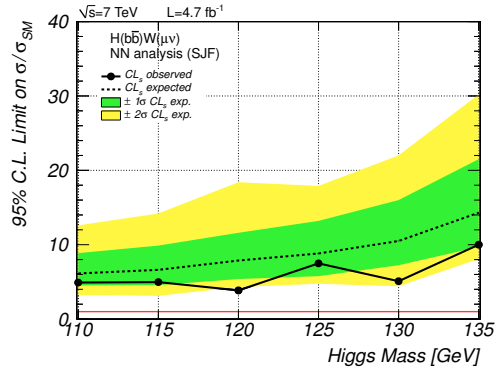
The observed limits indicated with the black solid lines are generally found to be lower than the corresponding expected limits in both the electron and muon channel. This is a direct consequence of the deficit of observed data events which was already discussed earlier. Overall, the observed limits are found to be consistent with the uncertainty bands of the expected limits.

Table 7.21.: Final event yield of the NN analysis in the muon channel using SJF jet reconstruction. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

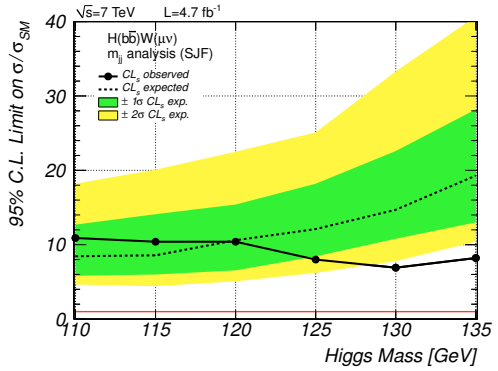
Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	3.26	2.70	2.06	1.67	0.61	0.91
W + $b\bar{b}$	18.52	17.70	15.20	13.69	9.50	10.44
Z + jets	0.20	0.20	0.00	0.00	0.00	0.00
$t\bar{t}$	10.15	11.14	9.48	9.40	5.91	8.10
Single top	2.25	2.81	2.01	1.72	1.10	2.08
VV	2.72	2.20	1.09	1.04	0.45	0.45
QCD	0.00	0.00	0.00	0.00	0.00	0.00
Background	37.08	36.74	29.84	27.54	17.57	21.98
Signal	2.89	2.62	2.08	1.74	1.18	0.94
Data	37	33	21	26	15	18
Significance	0.38	0.35	0.30	0.26	0.21	0.15

Table 7.22.: Final event yield of the NN analysis in the electron channel using SJF jet reconstruction. For each trial Higgs boson mass point the predicted number of background and signal events is given and compared to the observed event yield from data. Data/MC scale factors are applied. The value for the expected significance is based on equation (7.2).

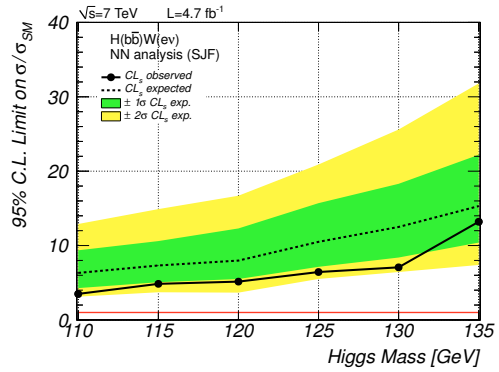
Process	H(110)	H(115)	H(120)	H(125)	H(130)	H(135)
W + udscg	1.34	1.86	2.24	1.52	1.39	1.82
W + $b\bar{b}$	8.04	8.07	9.31	9.38	8.15	7.67
Z + jets	0.00	0.00	0.00	0.00	0.00	0.00
$t\bar{t}$	5.88	6.65	8.46	7.42	7.14	6.94
Single top	1.47	1.01	1.74	1.22	1.37	1.79
VV	1.55	1.15	1.02	0.73	0.52	0.46
QCD	0.00	0.00	0.06	0.00	0.00	0.00
Background	18.29	18.73	22.78	20.28	18.58	18.68
Signal	2.01	1.74	1.62	1.28	1.03	0.76
Data	12	14	18	15	16	15
Significance	0.35	0.30	0.26	0.21	0.18	0.13



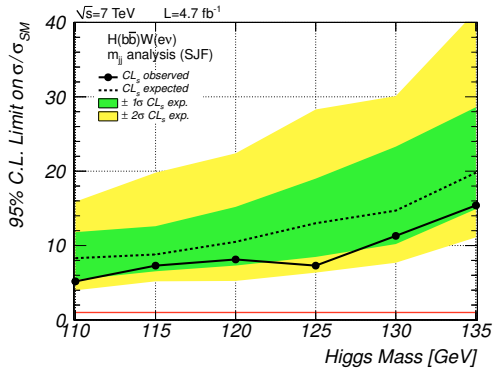
(a) Muon channel - NN analysis



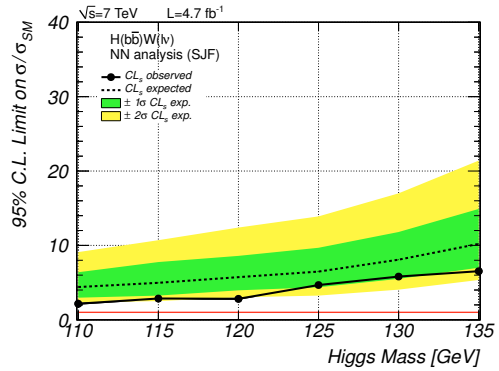
(b) Muon channel - m_{jj} analysis



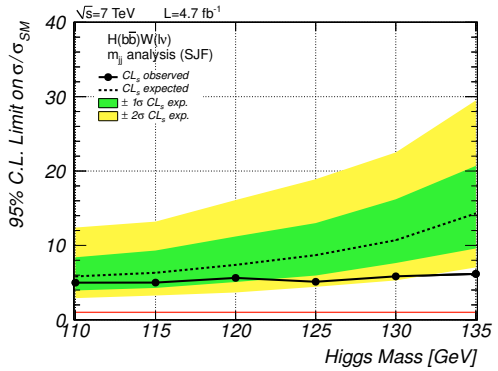
(c) Electron channel - NN analysis



(d) Electron channel - m_{jj} analysis



(e) Combination - NN analysis



(f) Combination - m_{jj} analysis

Figure 7.31.: CL_s exclusion limits using SJF jet reconstruction, for the NN analysis (left) and the m_{jj} analysis (right). The top row shows the muon channel results, the middle row the ones for the electron channel and the bottom row gives the exclusion limits for the combined HW search. The corresponding Bayesian limits are given in Figure B.28 in the appendix for comparison.

Table 7.23.: Summary of CL_s limits in the analysis using SJF jet reconstruction. For each channel and the combination the individual CL_s expected and observed values are shown for the NN analysis and the m_{jj} analysis.

$m(H)$	muons				electrons				combination			
	NN		m_{jj}		NN		m_{jj}		NN		m_{jj}	
	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.	obs.
110 GeV	6.13	4.91	8.45	10.9	6.32	3.50	8.30	5.18	4.40	2.16	5.83	4.99
115 GeV	6.62	4.97	8.57	10.4	7.32	4.85	8.80	7.30	4.97	2.86	6.32	5.00
120 GeV	7.87	3.86	10.6	10.4	7.98	5.15	10.5	8.13	5.73	2.82	7.38	5.63
125 GeV	8.82	7.49	12.1	8.00	10.5	6.44	13.0	7.30	6.48	4.67	8.69	5.11
130 GeV	10.5	5.09	14.7	6.91	12.5	7.07	14.7	11.3	8.08	5.81	10.7	5.84
135 GeV	14.3	10.0	19.3	8.21	15.3	13.2	19.8	15.4	10.2	6.53	14.3	6.17

The detailed values for both the NN and m_{jj} analyses are presented in Table 7.23. They are the basis for the following comparison of the two central results of this chapter: the combined expected and observed limits based on the NN analyses using standard jets and SJF jet reconstruction.

7.7. Comparison of the Search Sensitivities

The $H(b\bar{b})W(\ell\nu)$ search presented in this chapter has been performed using two fundamentally different approaches to the jet reconstruction:

Section 7.5 describes the search using the standard anti-k_T ($R = 0.5$) jet reconstruction which is the current default at the CMS experiment. In this standard analysis the Higgs boson candidate is based on the reconstruction of two separated standard jets. The standard analysis section specifies the details of the different analysis ingredients and finds the Neural Network approach to yield better results with respect to the m_{jj} analysis approach.

Section 7.6 highlights the differences when employing the originally proposed sub-/filterjet reconstruction algorithm. In this SJF analysis the Higgs boson candidate is based on the (filtered) jet substructure identified within one fat jet ($R = 1.6$) capturing the potential decay products of the Higgs boson.

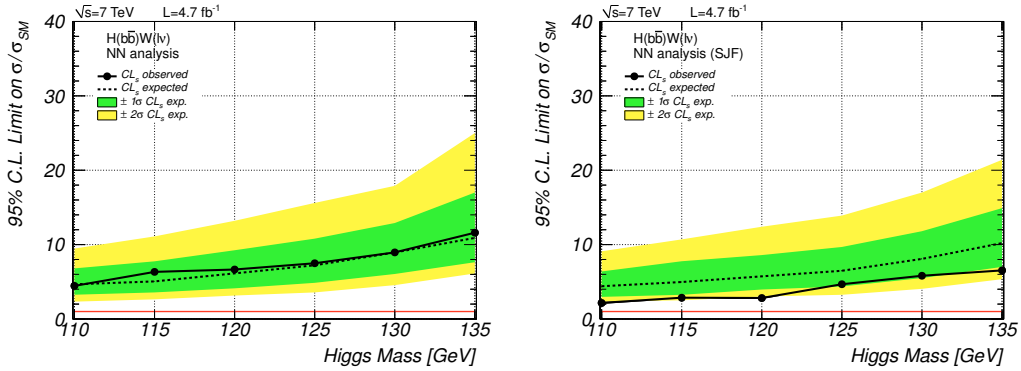
In each case the NN analysis is considered to compare the search sensitivity.

Figure 7.32 illustrates the two main results of both analyses, i. e. the CL_s 95% C.L. exclusion limits for the $H(b\bar{b})W(\ell\nu)$ combination using Neural Networks.

Figure 7.32(a) shows the standard jet result. The expected limit is found to lie in the interval $(4.7, 10.9)$ from a Higgs boson mass hypothesis of 110 GeV up to 135 GeV. The corresponding observed limits lie within $(4.5, 11.6)$.

Figure 7.32(b) depicts the NN exclusion limit using SJF jet reconstruction. The expected limit is found to lie within the interval of $(4.4, 10.2)$. The observed limit is found to be even more stringent compared to the standard analysis, lying within $(2.2, 6.5)$. Since different events are selected in data in the two analyses and the mass points in each analysis are highly correlated through the NN, this systematically lower observed limit is not unexpected. Both observed limits are found to be fully covered by the uncertainty on the expected limits, as indicated by the green and yellow $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands.

The expected median limit values in each mass point, on the other hand, quantify the expected difference in search sensitivity. Relative improvements of $2 - 10\%$ are found when moving from the standard to the sub-/filterjet reconstruction algorithm. This is summarized in Table 7.24.



(a) Standard jet NN Combination (b) Sub-/filterjet NN Combination

Figure 7.32.: Comparison of CL_s exclusion limits employing (a) standard jet reconstruction and (b) SJF jet reconstruction for the $H(b\bar{b})W(\ell\nu)$ combination using the NN analysis approach. An improvement of the search sensitivity (expected limit) of $2-10\%$ is observed moving from standard to sub- and filterjets. Different events are selected in data which explains the difference of the observed exclusion limits. It should be noted that the trial Higgs boson mass points are highly correlated in each analysis.

Table 7.24.: Comparison of the expected CL_s limits employing standard jet reconstruction and sub-/filterjet reconstruction for the $H(b\bar{b})W(\ell\nu)$ combination using the NN analysis approach. The expected limit values quantify the search sensitivity of the two reconstruction methods. Slight improvements of $2-10\%$ are expected when moving from standard to sub-/filterjets. The values given here are a summary of Table 7.13 (standard jets) and Table 7.23 (sub-/filterjets).

CL_s expected	$H(110)$	$H(115)$	$H(120)$	$H(125)$	$H(130)$	$H(135)$
Standard jets	4.66	5.05	6.13	7.22	8.94	10.9
Sub-/filterjets	4.40	4.97	5.73	6.48	8.08	10.2
Rel. difference	5.5%	1.6%	6.5%	10.2%	9.6%	6.4%

8. Conclusion & Outlook

The question of the existence of the Higgs boson has been central to elementary particle physics over the last decades. In this timeline, the start of the Large Hadron Collider at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ in March 2010 marks the advent of a new era to the search for this elusive particle, allowing for yet unprecedented collision energies and - possibly - Higgs boson production rates. Since the beginning of LHC operations, there has been rapid progress in the field of direct searches for the Higgs boson and the remaining possible Higgs boson mass region has been reduced dramatically. Today we know that the Standard Model Higgs boson is either very light, very heavy - or it does not exist.

The studies presented in this thesis highlight the work done from the commissioning of the first LHC collisions to the search for a light Standard Model Higgs boson in the channel $H(b\bar{b})W(\ell\nu)$.

The precise knowledge of the jet energy resolution is essential for all physics analyses with jets in the final state. The first measurement of the jet energy resolution of the CMS experiment using the dijet asymmetry method has shown a systematic discrepancy of the resolution description in simulation with respect to the collision data. Based on this measurement, a bias-correction is recommended to all CMS physics analyses to rescale the simulations such that they account for the discrepancy. The uncertainty on the jet energy resolutions was evaluated to $10 - 30\%$, making it an important source of systematic uncertainty to many jet related studies, including the Higgs boson search.

The jet energy resolution measurement is also key ingredient to the measurement of the inclusive jet cross section measurement. This study is an important test of the understanding of the underlying jet properties and a sensitive probe for new physics in the jet spectrum. The measurement has been found to agree with Standard Model theory predictions. It was extended to the measurement of the inclusive b-jet cross section, which represents an invaluable testbed to any analysis with b-jets in the final state.

The search for the Higgs boson in the channel $H(b\bar{b})W(\ell\nu)$ was performed in a light Higgs boson mass region from 110 to 135 GeV. The key to a successful reduction of the dominant backgrounds in this channel is the requirement of a kinematic boost, i. e. selecting only events where both the W boson and the Higgs boson have high transverse momenta and move back-to-back in the central detector plane. Using the anti-k_T jet reconstruction algorithm and employing artificial Neural Networks

to discriminate between the remaining backgrounds and the expected signal, CL_s 95% confidence level exclusion limits have been calculated, since no excess of signal events was observed in data. Based on this analysis, one expects to exclude Higgs boson signals at a 95% C.L. in an interval of $(4.7, 10.9)$ times the Standard Model for mass hypotheses ranging from 110 to 135 GeV. The corresponding limits observed in 2011 LHC data were found to be within $(4.5, 11.6)$.

The analysis was performed twice, based on different reconstruction techniques of the $H \to b\bar{b}$ decay. First, standard anti-k_T jets were used. Then, a dedicated fat-/sub- and filterjet algorithm was applied to study possible improvements to the search sensitivity. Improvements of the order of $2 - 10\%$ to the expected CL_s 95% C.L. exclusion limit were found when using sub-/filterjet information in the Neural Network analysis instead of the default anti-k_T jets.

The studies are compatible with the currently published CMS $VH(b\bar{b})$ result [164], which combines the 5 analysis channels with the vector bosons decaying via $W \to \mu\nu$, $W \to e\nu$, $Z \to \nu\nu$, $Z \to \mu\mu$ and $Z \to ee$. The expected combined limits lie within an interval of $(2.7, 6.7)$ times the Standard Model prediction. The corresponding observed limit range is $(3.1, 7.5)$.

Other important search channels in the light Higgs boson mass region include $H \to \gamma\gamma$ [173] which finds expected limits within $(1.5, 2.0)$ from 110 to 140 GeV and already disfavors the Higgs boson from 127 to 131 GeV at 95% C.L. The channel $H \to WW$ [174] is sensitive to Higgs boson masses as low as 110 GeV. It disfavors the SM Higgs boson in a mass range from $129 - 270$ GeV, but the expected and observed limits rise drastically for lower trial Higgs boson masses to values of ~ 9 times the Standard Model at 110 GeV. The Higgs boson decay into τ leptons, $H \to \tau\tau$ [175], is sensitive to Higgs boson masses from 110 to 145 GeV and yields expected limits of $(2.6, 4.8)$ in this mass range. The corresponding observed limit is found to lie within $(3.5, 6.3)$.

Outlook

Given the current theoretical and experimental evidence suggesting a light Higgs boson and the sensitivity of the $H(b\bar{b})W(\ell\nu)$ mode in this mass regime, this search channel will play an important role in the direct Higgs boson searches of the forthcoming months. In this sense, there are two possible future scenarios.

The $H(b\bar{b})W(\ell\nu)$ search contributes to the *discovery* of a light Standard Model Higgs boson. In this case the analysis strategy will have to be further optimized with respect to two important aspects to provide the necessary significance for a subsequent study of the Higgs boson properties with more collision data: reducing the systematic uncertainties and increasing the signal to background ratio, e. g. by including further discriminating variables. To give two examples: assuming a Higgs boson mass of 125 GeV, an optimized $S/B \sim 0.2$ and 10% background uncertainty,

a significance above 1 sigma should be within reach using the current 2012 LHC data projections of $\mathcal{L} = 20\,\text{fb}^{-1}$. In the most optimistic case, assuming $S/B \sim 0.2$ and 5% background uncertainty, it should be possible to measure the cross section with a significance above 2 sigma in a data sample corresponding to $\mathcal{L} = 60\,\text{fb}^{-1}$. Without the aforementioned optimizations higher significances cannot be achieved, even by increasing the amount of analyzed data.

The $H(b\bar{b})W(\ell\nu)$ search contributes to the full exclusion of the Standard Model Higgs Boson. This scenario should be considered a *"discovery"* as well, as it necessitates the existence of new physics beyond the Standard Model.

A. Additional Information on the Jet Energy Resolution Measurement

This Appendix provides additional technical details on the jet p_{T} resolution measurement presented in Chapter 6.

A.1. QCD MC Samples

Samples of simulated QCD events using PYTHIA 6.4 are used in the jet p_{T} resolution measurement. They are processed with the full GEANT 4 based CMS detector simulation as part of the central `Fall11` MC campaign using a CMS reconstruction software (CMSSW) version `CMSSW_3_8_4_patch2`. As simulation tune the CMS default Z2 tune is chosen, which is a re-tune of Z1 [131] with CTEQ6L1 [25] as PDF. All samples are listed in Table A.1 together with the number of events and cross sections. The samples are generated in mutually exclusive $\hat{p}_{\mathrm{T}}$ bins. The full MC spectrum is obtained by weighting the events with the fraction of the cross section and the total number of events in each exclusive sample respectively.

A.2. 2010 Jet Data

Data from the `JetMET, JetMETTau` and `Jet` primary datasets is merged to form the full recorded 2010 LHC jet dataset. A summary is given in Table A.2. The official CMS list of certified luminosity blocks for each run is being used to select only events which have been validated by the data quality management teams:

- `Cert_136033-149442_7TeV_Nov4ReReco_Collisions10_JSON.txt`

The total integrated luminosity of the data after this good run selection amounts to $35.9\,\mathrm{pb}^{-1}$.

Table A.1.: Summary of the 2010 PYTHIA 6.4 QCD datasets. The datasets are produced with upper and lower $\hat{p}_T$ boundaries (exclusive samples). For the last dataset the upper $\hat{p}_T$ boundary is left open.

MC sample	events	σ [pb]
/QCD_Pt_0to5_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	103819	4.844e+10
/QCD_Pt_5to15_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	433782	3.675e+10
/QCD_Pt_15to30_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2832553	8.159e+08
/QCD_Pt_30to50_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2161010	5.311e+07
/QCD_Pt_50to80_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2343234	6.358e+06
/QCD_Pt_80to120_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2549698	784900
/QCD_Pt_120to170_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2455344	115100
/QCD_Pt_170to300_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2795687	24260
/QCD_Pt_300to470_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	2890492	1168
/QCD_Pt_470to600_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	1897914	70.21
/QCD_Pt_600to800_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	1843206	15.57
/QCD_Pt_800to1000_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	1974880	1.843
/QCD_Pt_1000to1400_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	1017955	0.3318
/QCD_Pt_1400to1800_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	1013161	0.01086
/QCD_Pt_1800_TuneZ2_7TeV_pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO	527179	0.0003575

Table A.2.: Summary of the different jet primary datasets (PD). The data corresponds to the full luminosity delivered by LHC in 2010 and recorded by CMS.

era	PD	dataset name
2010A	JetMET	/JetMET/Run2010A-Nov4ReReco_v1/RECO
2010A	JetMETTau	/JetMETTau/Run2010A-Nov4ReReco_v1/RECO
2010B	Jet	/Jet/Run2010B-Nov4ReReco_v1/RECO

B. Additional Information on the Higgs Boson Search

This appendix provides additional technical details on the search for a light Standard Model Higgs boson in the channel $H(b\bar{b})W(l\nu)$ presented in Chapter 7.

B.1. Technical Details on Simulated Samples

A summary of simulated signal and background samples used in the Higgs boson search is given in Table B.1 including technical details.

B.2. Technical Details on Data Samples and Triggers

A summary of the different primary datasets used in the Higgs boson search are given in Table B.2.
Supplementary information on the corresponding trigger selections are presented in Table B.3.

B.3. Supplementary Distributions

Supplementary control distributions, which are needed to for a coherent presentation of the HW analysis using standard jet reconstruction and sub-/filterjets, are shown in Figures B.1 - B.28.

Table B.1.: Summary of the simulated signal and background samples used in the Higgs boson search. Signal cross sections are based on NNLO calculations and are taken from [29]. The $t\bar{t}$ cross section (NNLL) is taken from [154]. Single top cross sections (NNLO and NNLL) are based on [156–158]. The V+jets samples use the LO simulation cross sections and are scaled to the NLO order using the k-factors in analogy to [155]. The VV samples use the NLO cross sections provided by [159]. The QCD sample cross sections are taken from the Pythia 6.4 simulation output. The default CMS PDF set CTEQ6L1 [25] is used with all samples. The Pythia 6.4 samples use the Z2 underlying event tune [131]. All samples contain admixtures of pile-up according to the `Fall11-PU_S6` profile.

Sample	Events	Cross-Section [pb]	MC-Filter	K-Factor	Luminosity [fb^{-1}]
$H(b\bar{b},110)W(l\nu)$ (Powheg)	1094448	0.211304	1	1	5179.49
$H(b\bar{b},115)W(l\nu)$ (Powheg)	1100000	0.172121	1	1	6390.85
$H(b\bar{b},120)W(l\nu)$ (Powheg)	1100000	0.13775	1	1	7985.48
$H(b\bar{b},125)W(l\nu)$ (Powheg)	1097224	0.107103	1	1	10244.6
$H(b\bar{b},130)W(l\nu)$ (Powheg)	1025000	0.0799938	1	1	12813.5
$H(b\bar{b},135)W(l\nu)$ (Powheg)	1025000	0.0573211	1	1	17881.7
$H(b\bar{b},110)Z(ll)$ (Powheg)	218778	0.035514	1	1	6160.33
$H(b\bar{b},115)Z(ll)$ (Powheg)	220000	0.0291949	1	1	7535.56
$H(b\bar{b},120)Z(ll)$ (Powheg)	220000	0.0235421	1	1	9344.96
$H(b\bar{b},125)Z(ll)$ (Powheg)	216760	0.0183991	1	1	11781
$H(b\bar{b},130)Z(ll)$ (Powheg)	218884	0.0138289	1	1	15828
$H(b\bar{b},135)Z(ll)$ (Powheg)	220000	0.00998188	1	1	22039.9
$H(b\bar{b},110)Z(\nu\nu)$ (Powheg)	219247	0.0703429	1	1	3116.83
$H(b\bar{b},115)Z(\nu\nu)$ (Powheg)	220000	0.0578266	1	1	3804.48
$H(b\bar{b},120)Z(\nu\nu)$ (Powheg)	220000	0.0466301	1	1	4717.98
$H(b\bar{b},125)Z(\nu\nu)$ (Powheg)	220000	0.0364433	1	1	6036.77
$H(b\bar{b},130)Z(\nu\nu)$ (Powheg)	219422	0.0273911	1	1	8010.7
$H(b\bar{b},135)Z(\nu\nu)$ (Powheg)	219352	0.0197712	1	1	11094.5
$t\bar{t}$+jets (MadGraph)	56961377	165	1	1	345.22
t (tW-channel) (Powheg)	814390	7.87	1	1	103.48
$\bar{t}$ (tW-channel) (Powheg)	323401	7.87	1	1	41.0929
t (t-channel) (Powheg)	3900171	41.92	1	1	93.0384
$\bar{t}$ (t-channel) (Powheg)	1944826	22.65	1	1	85.8643
t (s-channel) (Powheg)	259971	3.19	1	1	81.4956
$\bar{t}$ (s-channel) (Powheg)	137980	1.44	1	1	95.8194
$W(l\nu)$+jets (MadGraph)	81265381	24380	1	1.277	2.61024
$W(l\nu)$+jets ($p_T(W) > 100\,\text{GeV}$) (MadGraph)	8047094	212.5	1	1.2	31.5572
$Z(ll)$+jets (MadGraph)	35655408	2289	1	1.299	11.9914
$Z(ll)$+jets ($p_T(Z) > 100\,\text{GeV}$) (MadGraph)	1137280	25.1	1	1.15	39.4
WW (Pythia 6.4)	4185916	43	1	1	97.3469
WZ (Pythia 6.4)	4265243	18.2	1	1	234.354
ZZ (Pythia 6.4)	4191045	5.9	1	1	710.347
QCD (μ−enriched) (Pythia 6.4)	18630241	2.966e+08	0.0002855	1	0.22
QCD ($\hat{p}_\text{T}:\ 120-170\,\text{GeV}$) (Pythia 6.4)	6127528	115000	1	1	0.0532829
QCD ($\hat{p}_\text{T}:\ 170-300\,\text{GeV}$) (Pythia 6.4)	6220160	24300	1	1	0.255974
QCD ($\hat{p}_\text{T}:\ 300-470\,\text{GeV}$) (Pythia 6.4)	6432669	1170	1	1	5.49801
QCD ($\hat{p}_\text{T}:\ 470-600\,\text{GeV}$) (Pythia 6.4)	3990085	70.2	1	1	56.8388
QCD ($\hat{p}_\text{T}:\ 600-800\,\text{GeV}$) (Pythia 6.4)	4245695	15.6	1	1	272.16
QCD ($\hat{p}_\text{T}:\ 800-1000\,\text{GeV}$) (Pythia 6.4)	4053888	1.84	1	1	2203.2
QCD ($\hat{p}_\text{T}:\ 1000-1400\,\text{GeV}$) (Pythia 6.4)	2093222	0.332	1	1	6304.89

Table B.2.: Summary of the different primary datasets (PD). The data corresponds to the full luminosity delivered by LHC in 2011 and recorded by CMS. The total integrated luminosity after the good luminosity selection amounts to $\mathcal{L} = 4.65\,\mathrm{fb}^{-1}$ in both the muon and the electron channel.

Mode	Era	Primary Dataset
$W(\mu\nu)H(b\bar{b})$	2011A	/SingleMu/Run2011A-May10ReReco-v1/AOD
	2011A	/SingleMu/Run2011A-05Aug2011-v1/AOD
	2011A	/SingleMu/Run2011A-PromptReco-v4/AOD
	2011A	/SingleMu/Run2011A-PromptReco-v6/AOD
	2011B	/SingleMu/Run2011B-PromptReco-v1/AOD
$W(e\nu)H(b\bar{b})$	2011A	/SingleElectron/Run2011A-May10ReReco-v1/AOD
	2011A	/ElectronHad/Run2011A-05Aug2011-v1/AOD
	2011A	/ElectronHad/Run2011A-PromptReco-v4/AOD
	2011A	/ElectronHad/Run2011A-PromptReco-v6/AOD
	2011B	/ElectronHad/Run2011B-PromptReco-v1/AOD

Table B.3.: Summary of the trigger selection used in the HW search. Multiple triggers in the same run range are connected by a a logical "or". The trigger selection and the corresponding efficiencies were centrally produced in the CMS Hbb Higgs sub-group and have been taken from [155, 161].

Mode	Run Range	Trigger
$W(\mu\nu)H(b\bar{b})$	160431 – 173198	HLT_IsoMu17_v*
		HLT_IsoMu20_v*
		HLT_IsoMu24_v*
		HLT_Mu30_v*
		HLT_Mu40_v*
	173198 – 175832	HLT_IsoMu20_v*
		HLT_IsoMu24_v*
		HLT_IsoMu24_eta2p1_v*
		HLT_Mu40_eta2p1_v*
	175832 – 178390	HLT_IsoMu20_v*
		HLT_IsoMu24_v*
		HLT_IsoMu24_eta2p1_v*
		HLT_IsoMu30_eta2p1_v*
		HLT_Mu40_v*
		HLT_Mu40_eta2p1_v*
	178390 – 180252	HLT_IsoMu24_v*
		HLT_IsoMu24_eta2p1_v*
		HLT_IsoMu30_eta2p1_v*
		HLT_Mu40_v*
		HLT_Mu40_eta2p1_v*
$W(e\nu)H(b\bar{b})$	160431 – 163869	HLT_Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v*
	163869 – 178390	HLT_Ele17_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CentralJet30_CentralJet25_PFMHT15_v*
		HLT_Ele22_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CentralJet30_CentralJet25_PFMHT20_v*
		HLT_Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CentralJet30_CentralJet25_PFMHT20_v*
		HLT_Ele30_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_DiCentralJet30_PFMHT25_v*
	178390 – 180252	HLT_Ele27_WP80_DiCentralPFJet25_PFMHT15_v*

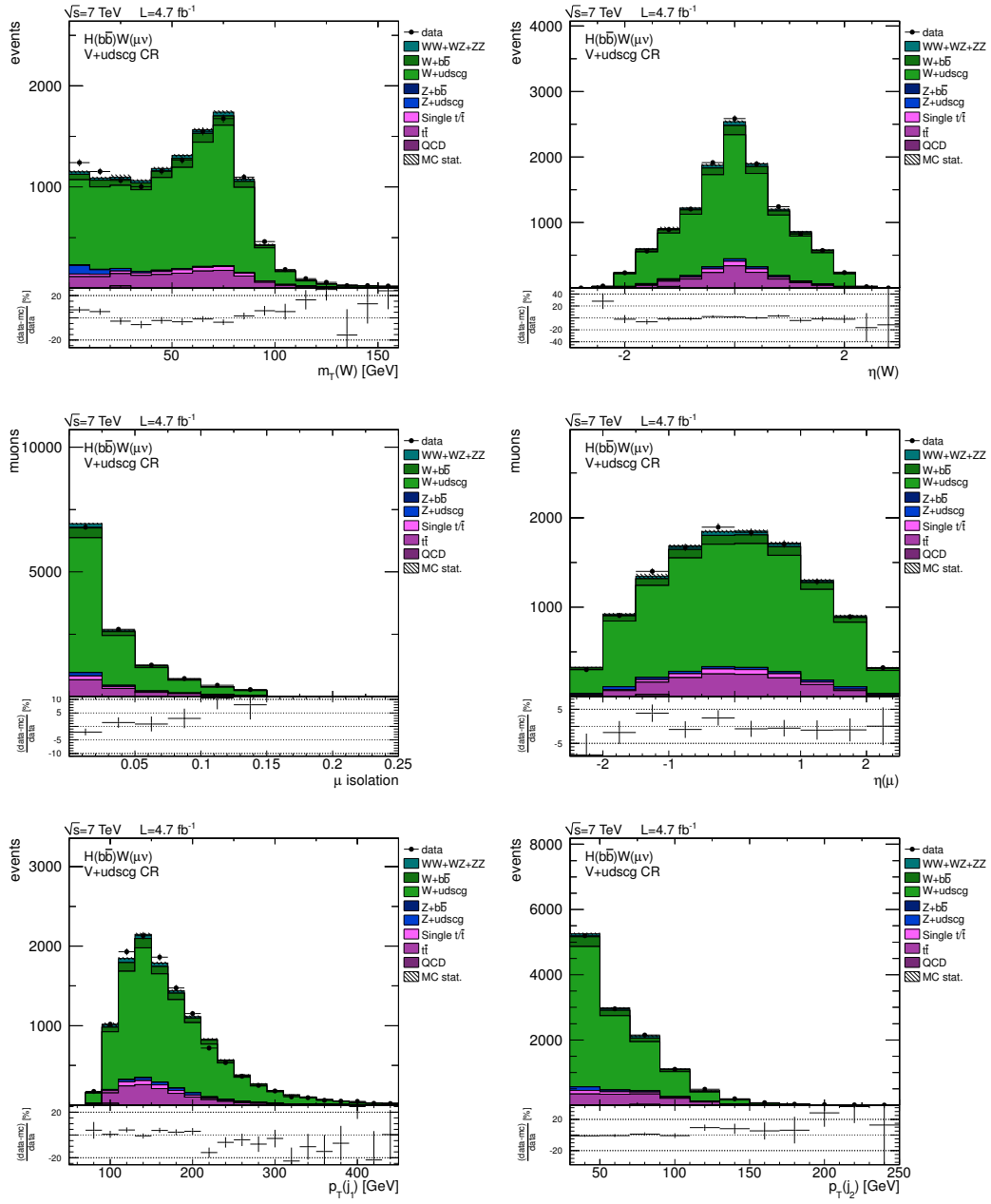


Figure B.1.: Additional W + udscg jets control region distributions for the muon channel: the transverse mass and pseudorapidity of the W boson, the isolation and pseudorapidity of the muon and the p_{T} of the first and second Higgs candidate jet. Scale factors have been applied and the number of simulated events is normalized to data in order to facilitate the shape comparison.

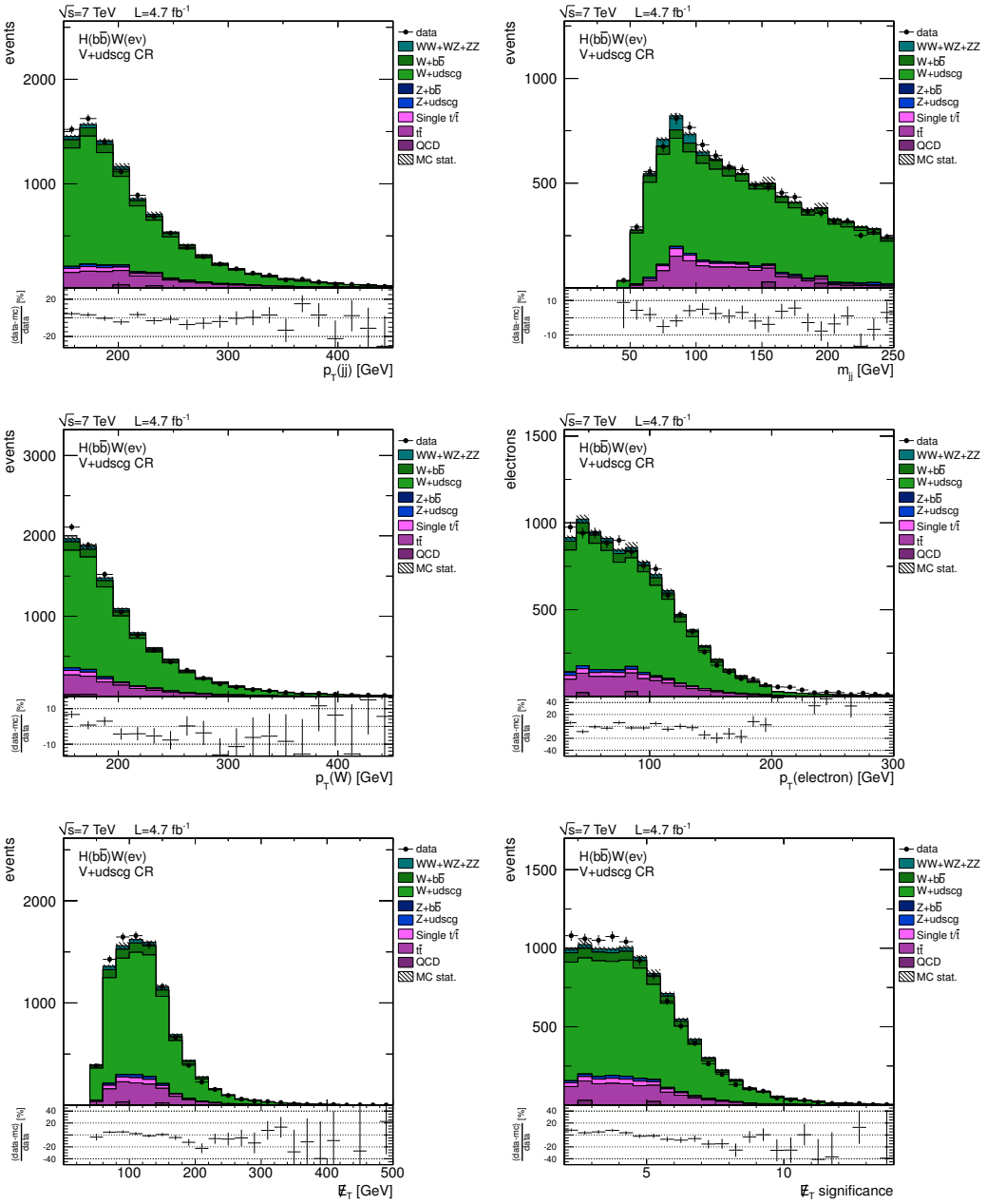


Figure B.2.: W + udscg jets control region distributions for the electron channel: the p_{T} of the dijet system and the corresponding dijet mass, the p_{T} of the reconstructed W boson and electron, as well as the $\not{E}_{\mathrm{T}}$ and $\not{E}_{\mathrm{T}}$ significance distributions. Scale factors have been applied and the number of simulated events is normalized to data.

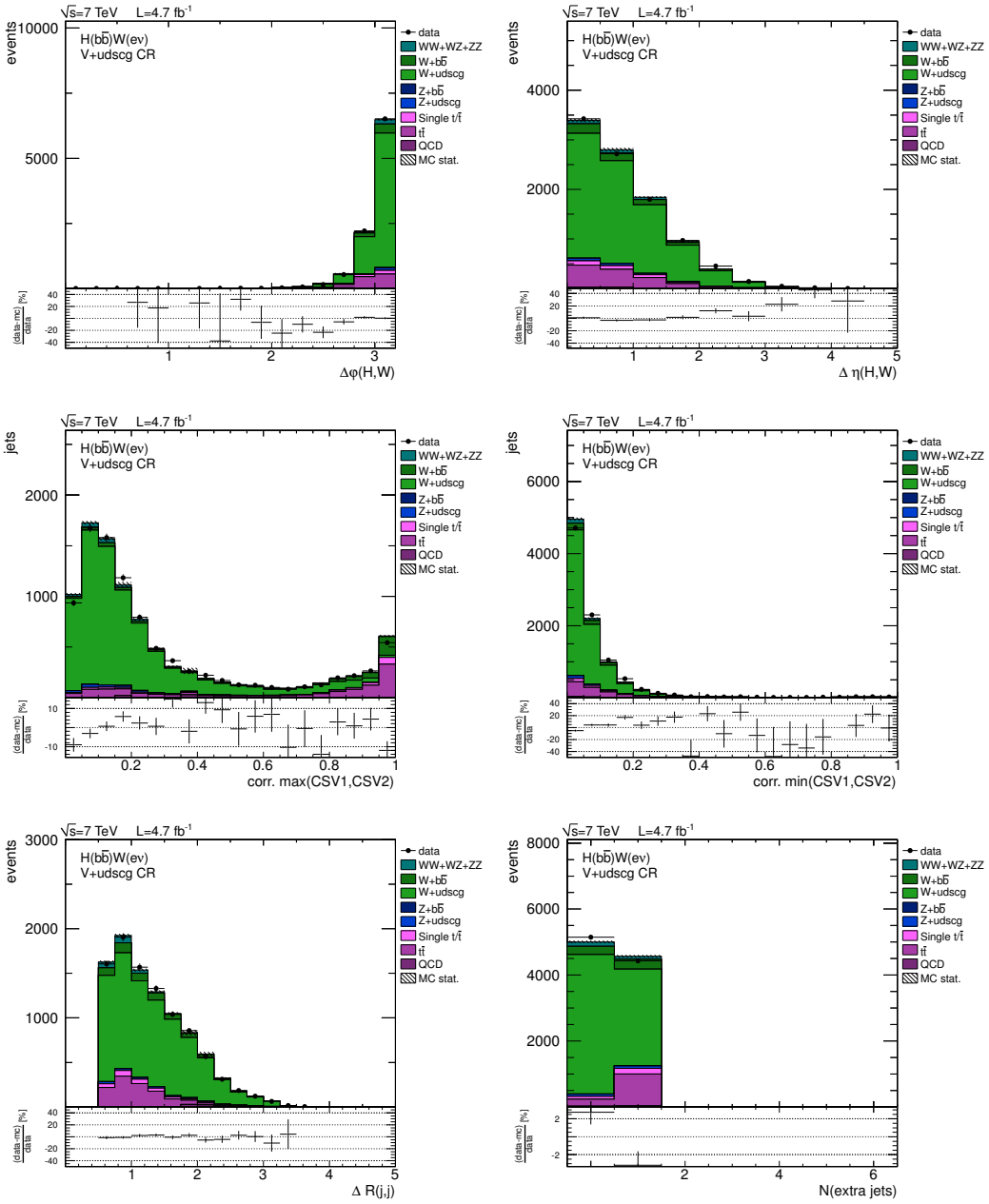


Figure B.3.: W + udscg jets control region distributions for the electron channel (cont.): the azimuthal opening angle between the dijet system and the W boson, their difference in η, the CSV b-tag discriminator outputs for the dijet constituents, the opening angle ΔR between them and the number of additional central jets in the event. Scale factors have been applied and the number of simulated events is normalized to data.

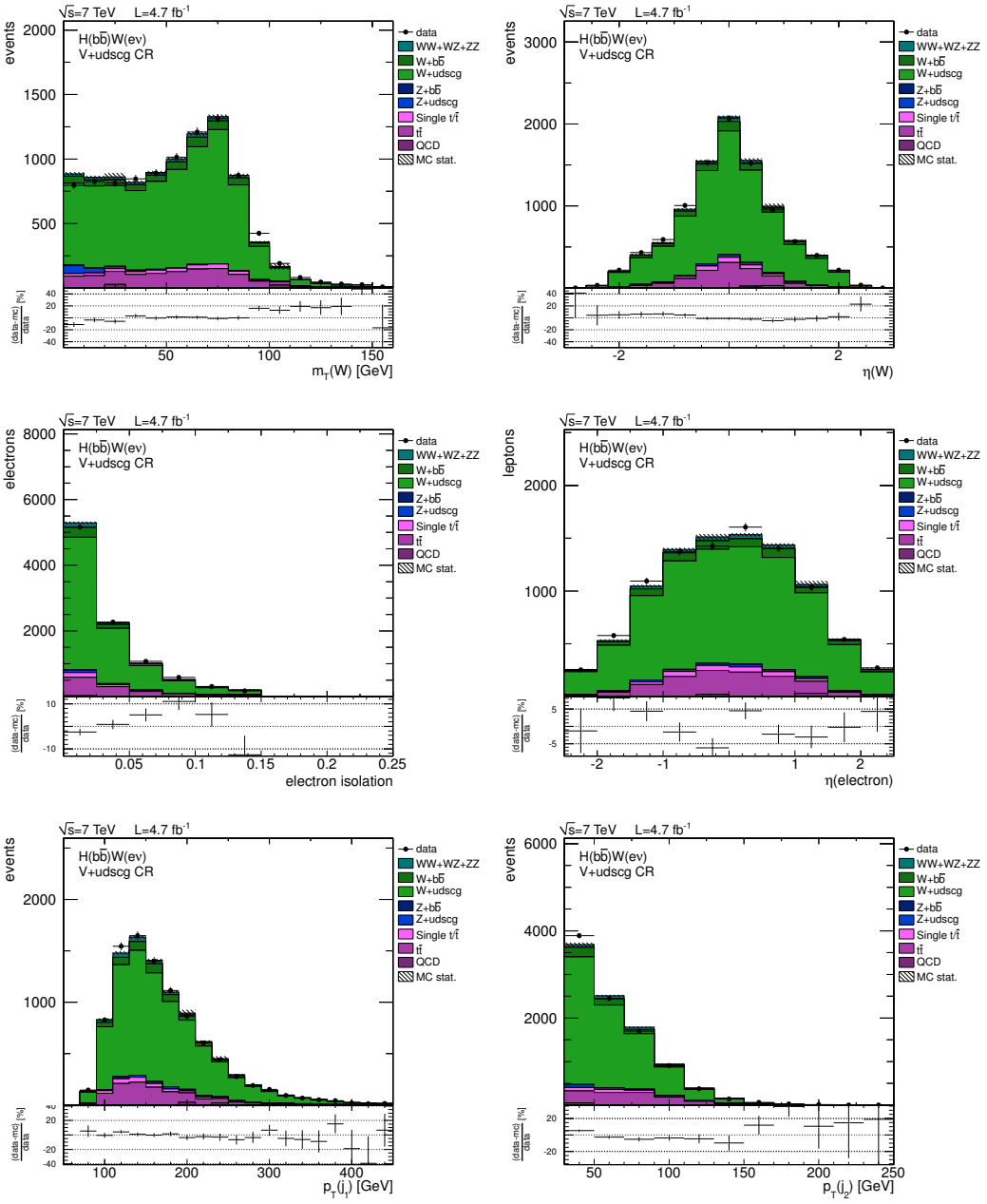


Figure B.4.: Additional W+udscg jets control region distributions for the electron channel: the transverse mass and pseudorapidity of the W boson, the isolation and pseudorapidity of the electron and the p_{T} of the first and second Higgs candidate jet. Scale factors have been applied and the number of simulated events is normalized to data.

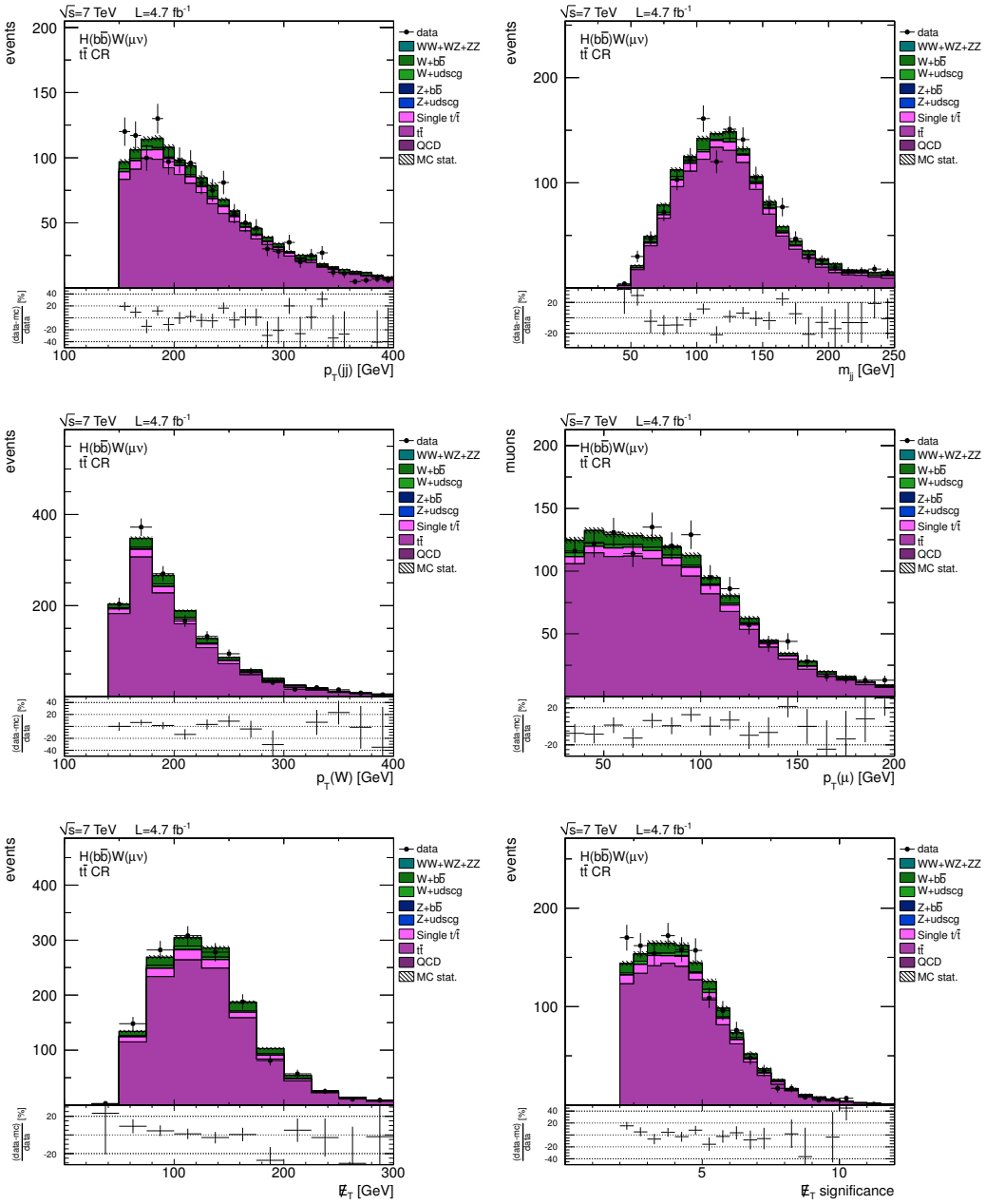


Figure B.5.: $t\bar{t}$ + jets control region distributions for the muon channel: the p_T of the dijet system and the corresponding dijet mass, the p_T of the reconstructed W boson and muon, as well as the $\not{E}_T$ and $\not{E}_T$ significance distributions. Scale factors have been applied and the number of simulated events is normalized to data.

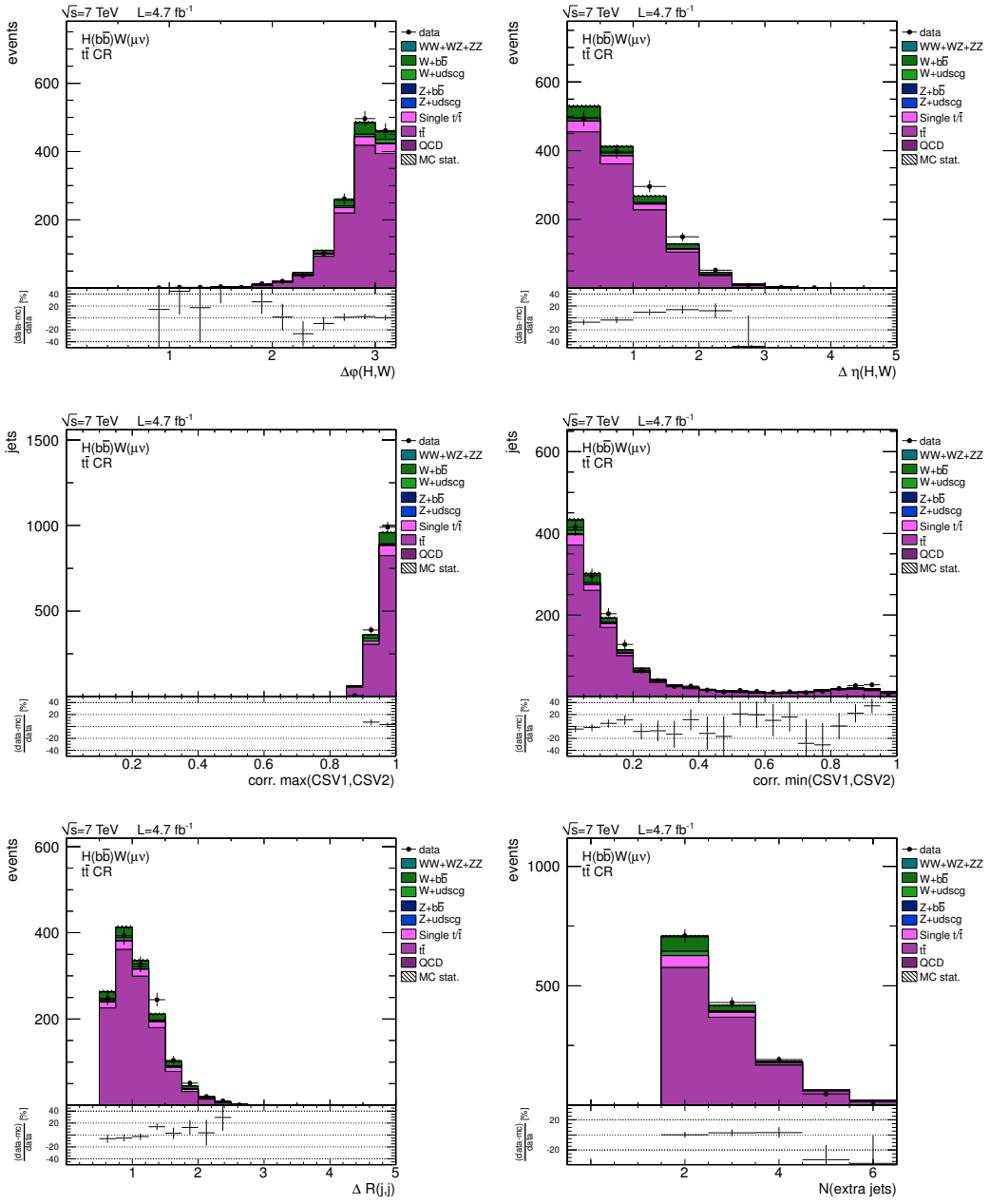


Figure B.6.: $t\bar{t}$ + jets control region distributions for the muon channel (cont.): the azimuthal opening angle between the dijet system and the W boson, their difference in η, the CSV b-tag discriminator outputs for the dijet constituents, the opening angle ΔR between them and the number of additional central jets in the event. Scale factors have been applied and the number of simulated events is normalized to data.

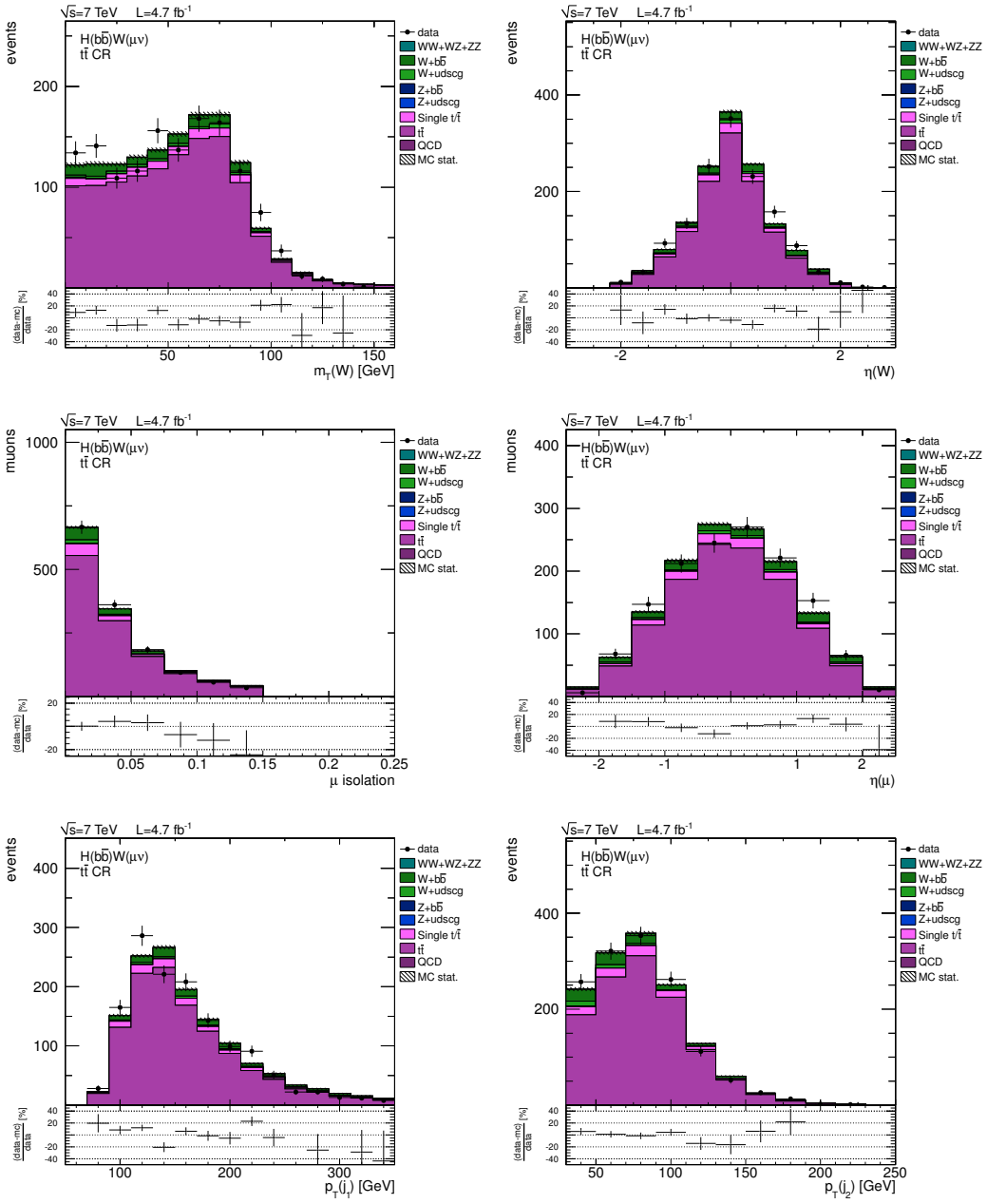


Figure B.7.: Additional $t\bar{t}$ + jets control region distributions for the muon channel: the transverse mass and pseudorapidity of the W boson, the isolation and pseudorapidity of the muon and the p_T of the first and second Higgs candidate jet. Scale factors have been applied and the number of simulated events is normalized to data.

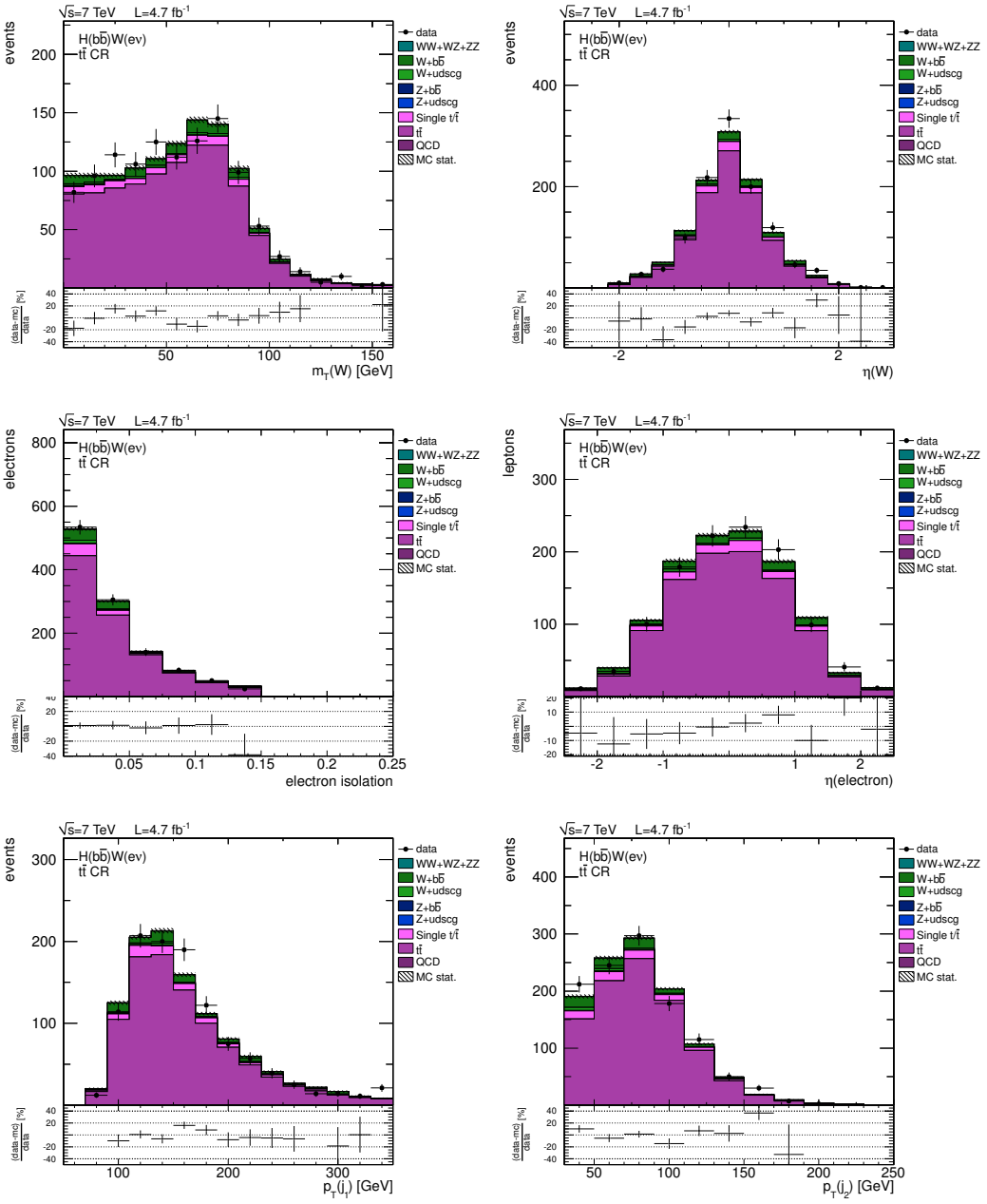


Figure B.8.: Additional $t\bar{t}$ + jets control region distributions for the electron channel: the transverse mass and pseudorapidity of the W boson, the isolation and pseudorapidity of the electron and the p_T of the first and second Higgs candidate jet. Scale factors have been applied and the number of simulated events is normalized to data.

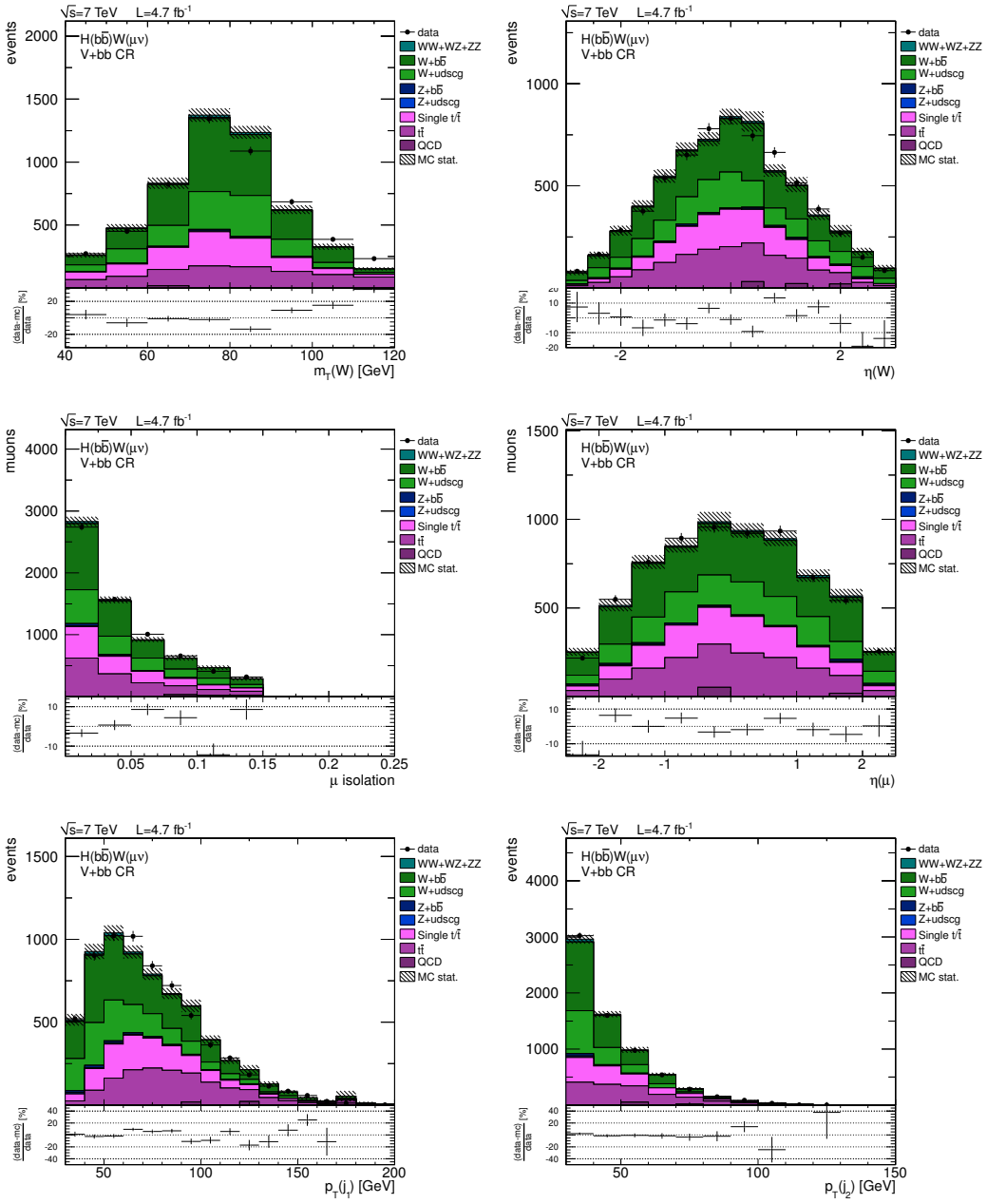


Figure B.9.: Additional $\mathrm{W} + \mathrm{b\bar{b}}$ jets control region distributions for the muon channel: the transverse mass and pseudorapidity of the W boson, the isolation and pseudorapidity of the muon and the p_{T} of the first and second Higgs candidate jet. Scale factors have been applied and the number of simulated events is normalized to data.

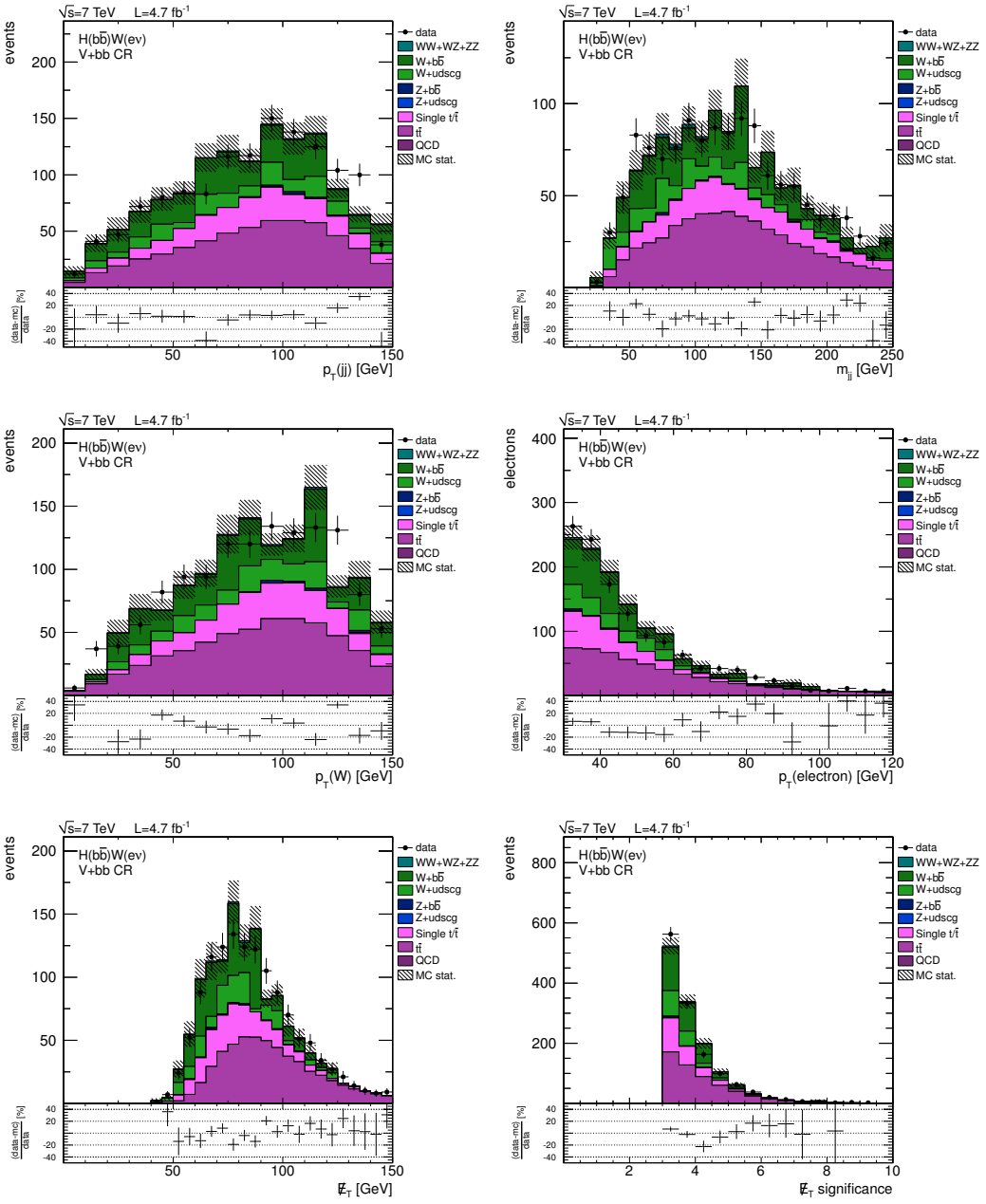


Figure B.10.: $W + b\bar{b}$ jets control region distributions for the electron channel: the p_T of the dijet system and the corresponding dijet mass, the p_T of the reconstructed W boson and electron, as well as the $\not{E}_T$ and $\not{E}_T$ significance distributions. Scale factors have been applied and the number of simulated events is normalized to data.

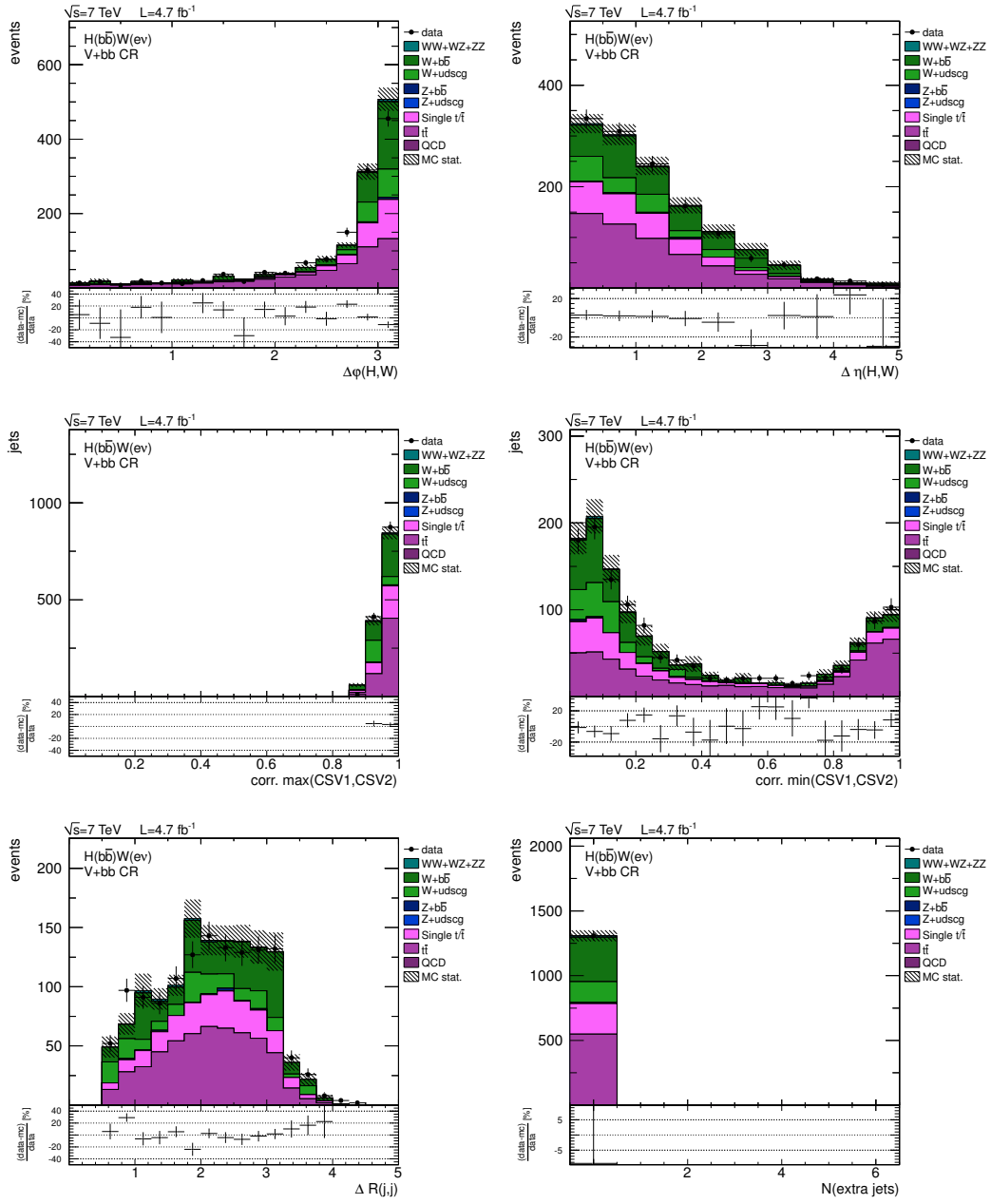


Figure B.11.: $W + b\bar{b}$ jets control region distributions for the electron channel (cont.): the azimuthal opening angle between the dijet system and the W boson, their difference in η, the CSV b-tag discriminator outputs for the dijet constituents, the opening angle ΔR between them and the number of additional central jets in the event. Scale factors have been applied and the number of simulated events is normalized to data.

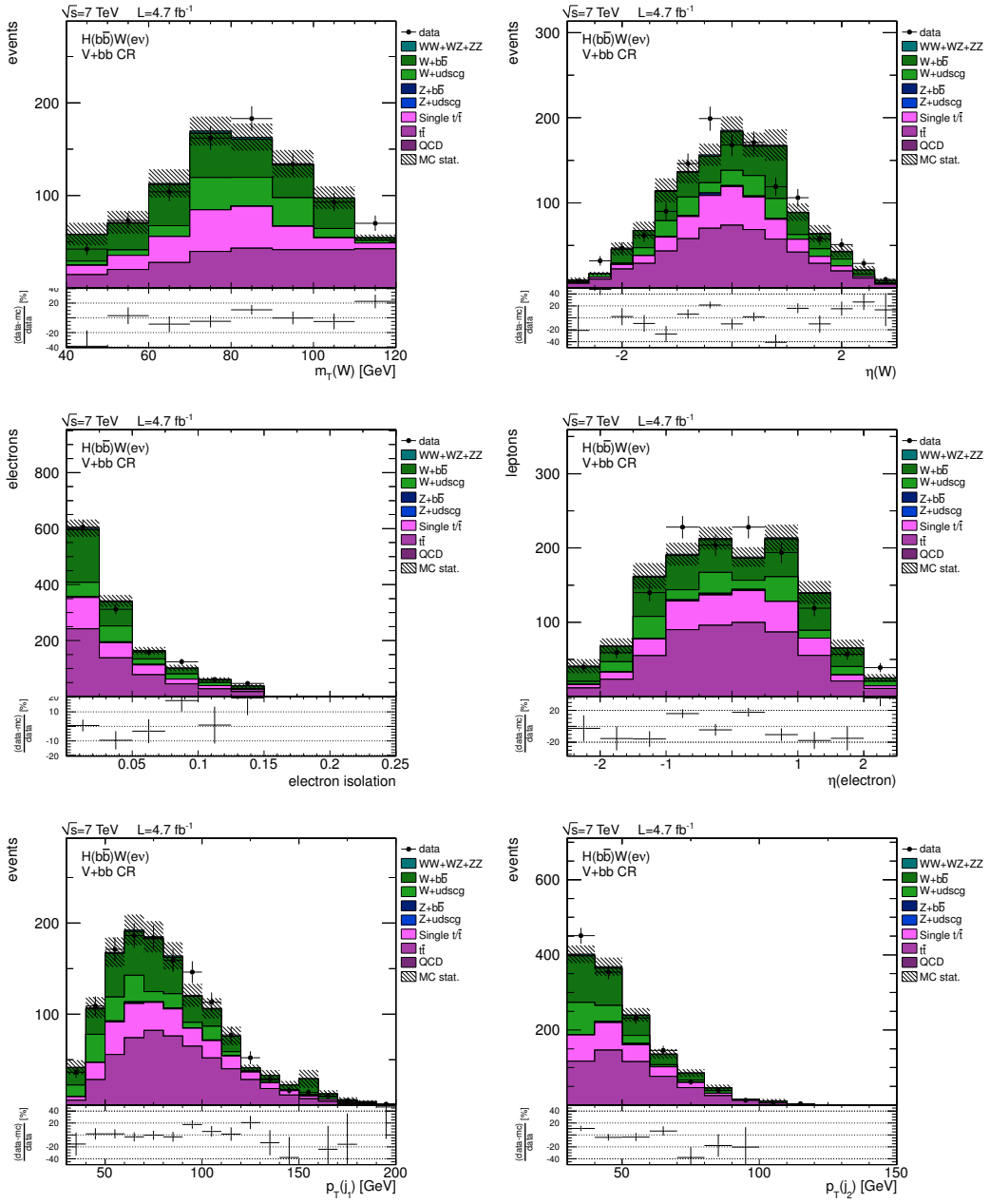
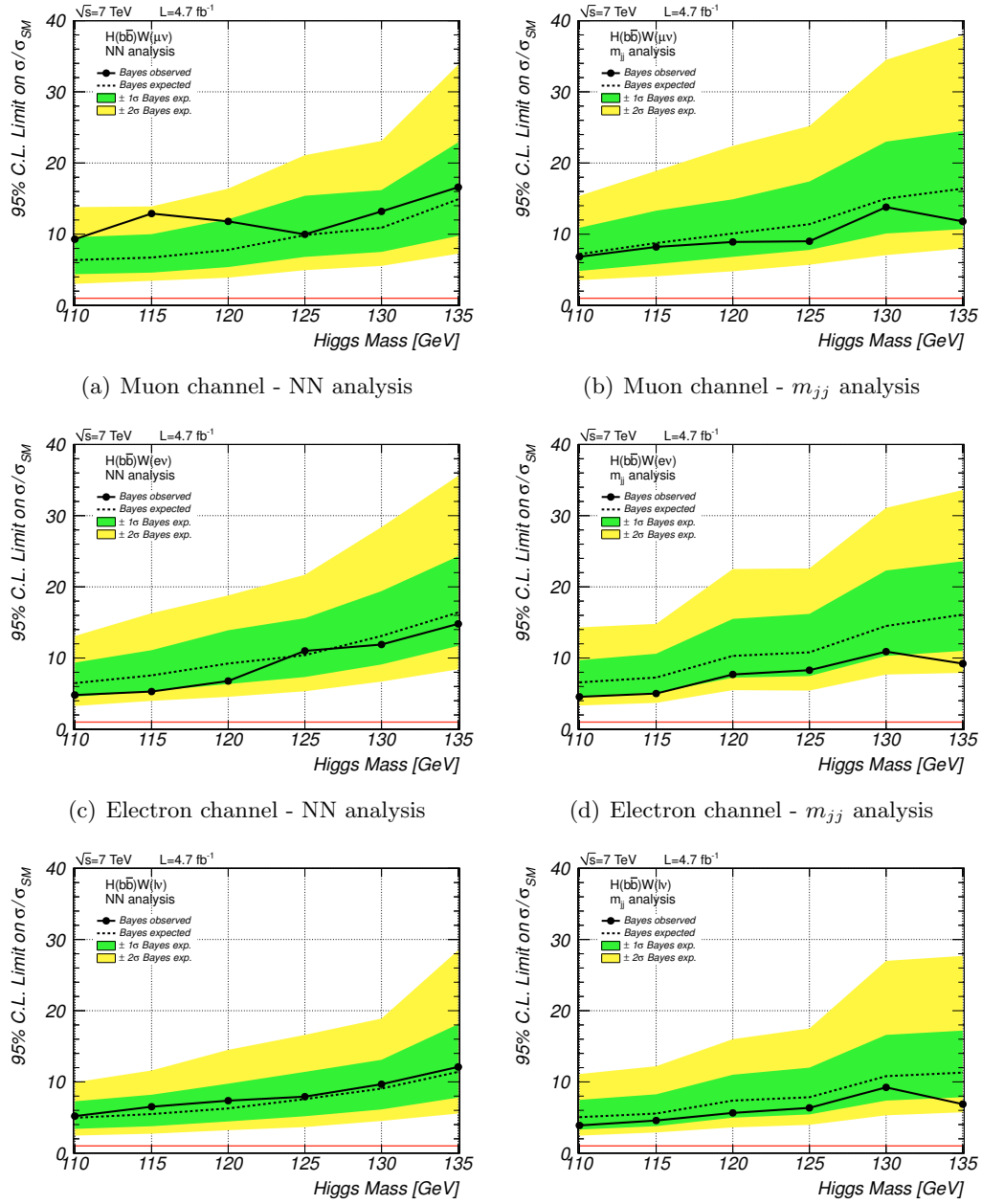


Figure B.12.: Additional $W + b\bar{b}$ jets control region distributions for the electron channel: the transverse mass and pseudorapidity of the W boson, the isolation and pseudorapidity of the electron and the p_T of the first and second Higgs candidate jet. Scale factors have been applied and the number of simulated events is normalized to data.



(a) Muon channel - NN analysis

(b) Muon channel - m_{jj} analysis

(c) Electron channel - NN analysis

(d) Electron channel - m_{jj} analysis

(e) Combination - NN analysis

(f) Combination - m_{jj} analysis

Figure B.13.: Bayesian exclusion limits using standard jets, for the NN analysis (left) and the m_{jj} analysis (right). The top row shows the muon channel result, the middle row the electron channel and the bottom row gives the exclusion limits for the combined HW search. The Bayesian limits are found to be in very good agreement with the corresponding CL_s exclusion limits.

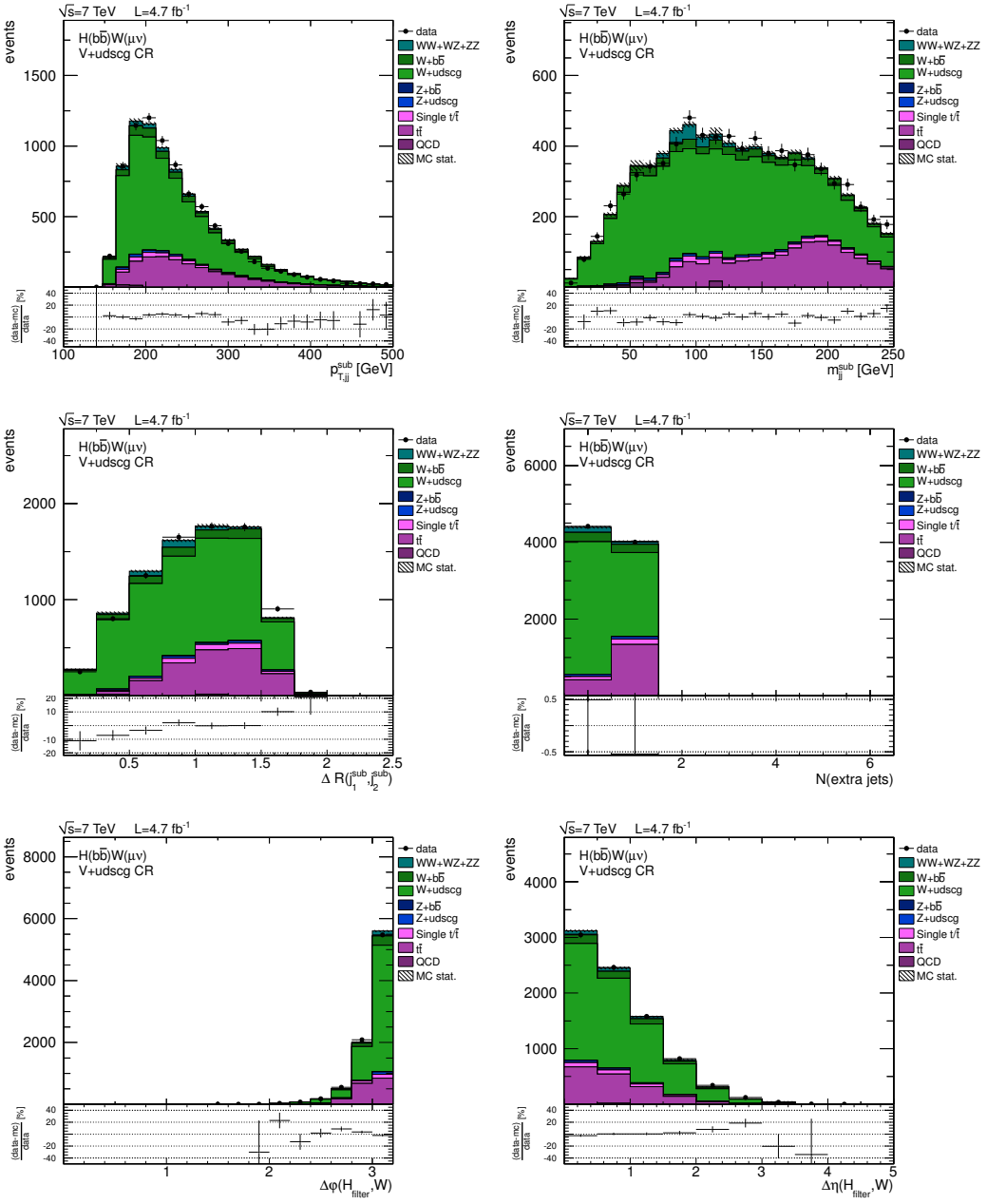


Figure B.14.: Additional W + udscg jets control region distributions for the muon channel using SJF jet reconstruction: the p_{T} of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data in order to facilitate the shape comparison.

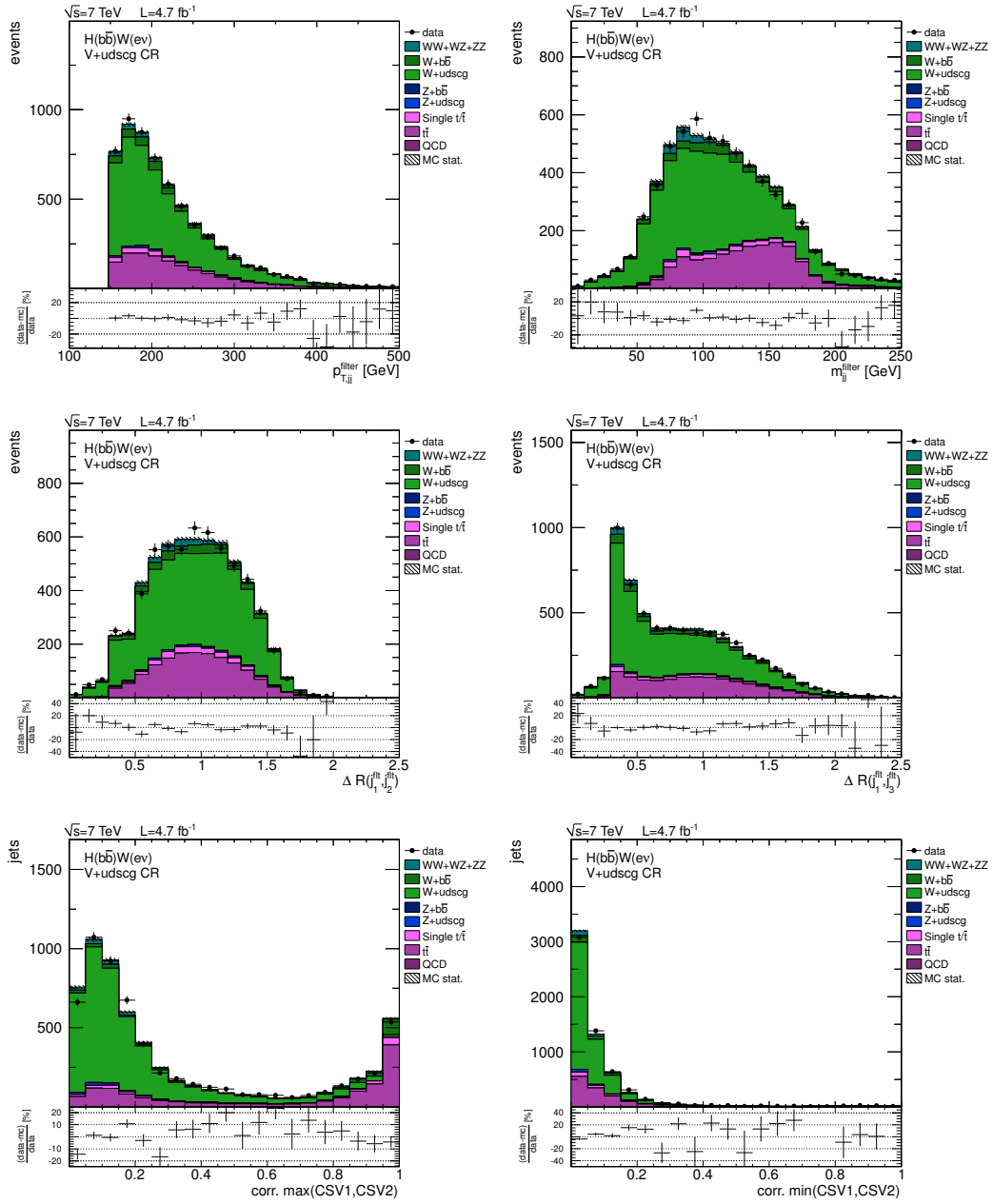


Figure B.15.: W + udscg jets control region distributions for the electron channel using SJF jet reconstruction: the p_{T} of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data.

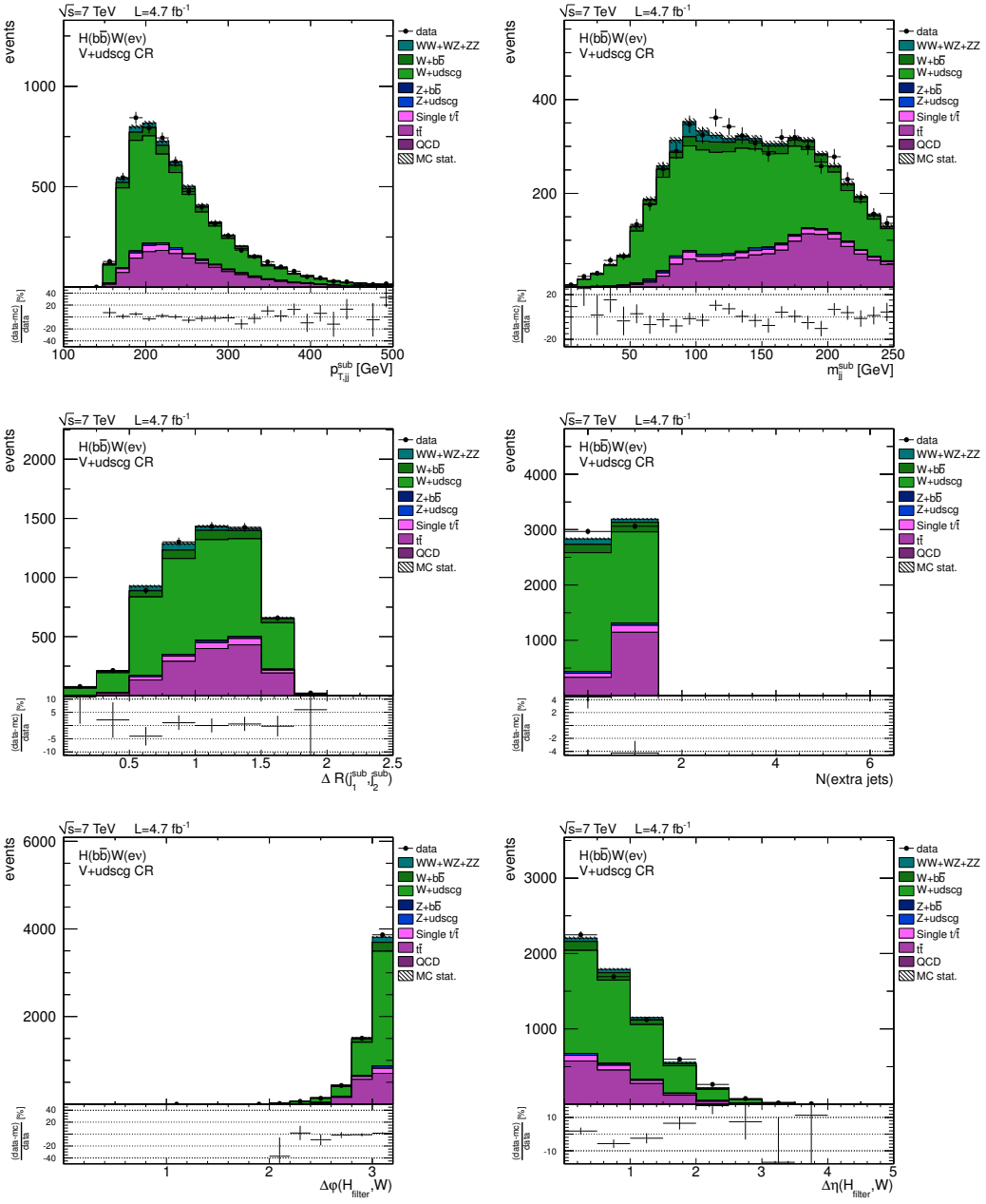


Figure B.16.: Additional W+udscg jets control region distributions for the electron channel using SJF jet reconstruction: the p_{T} of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

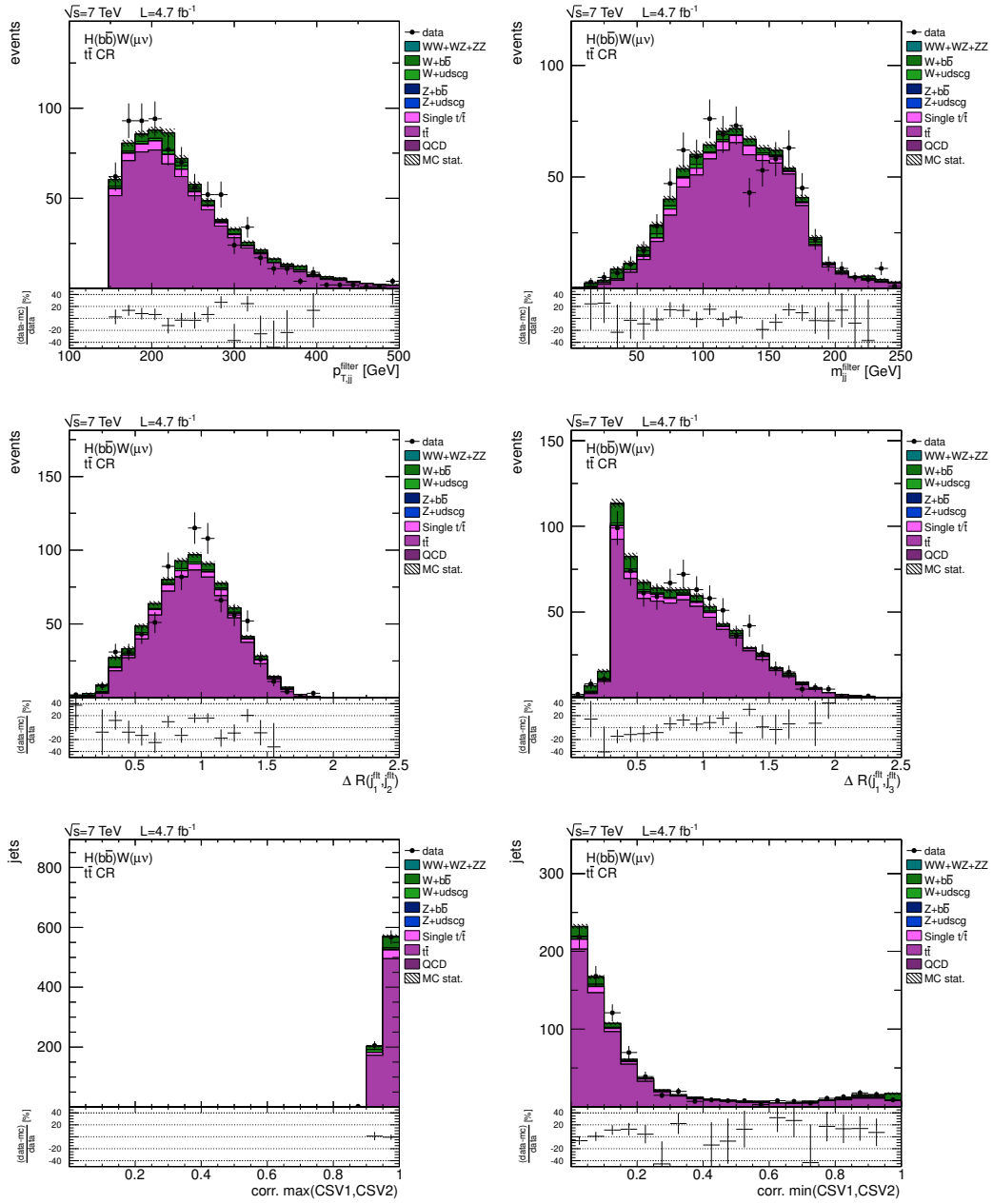


Figure B.17.: $t\bar{t}$ control region distributions for the muon channel using SJF jet reconstruction: the p_T of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data.

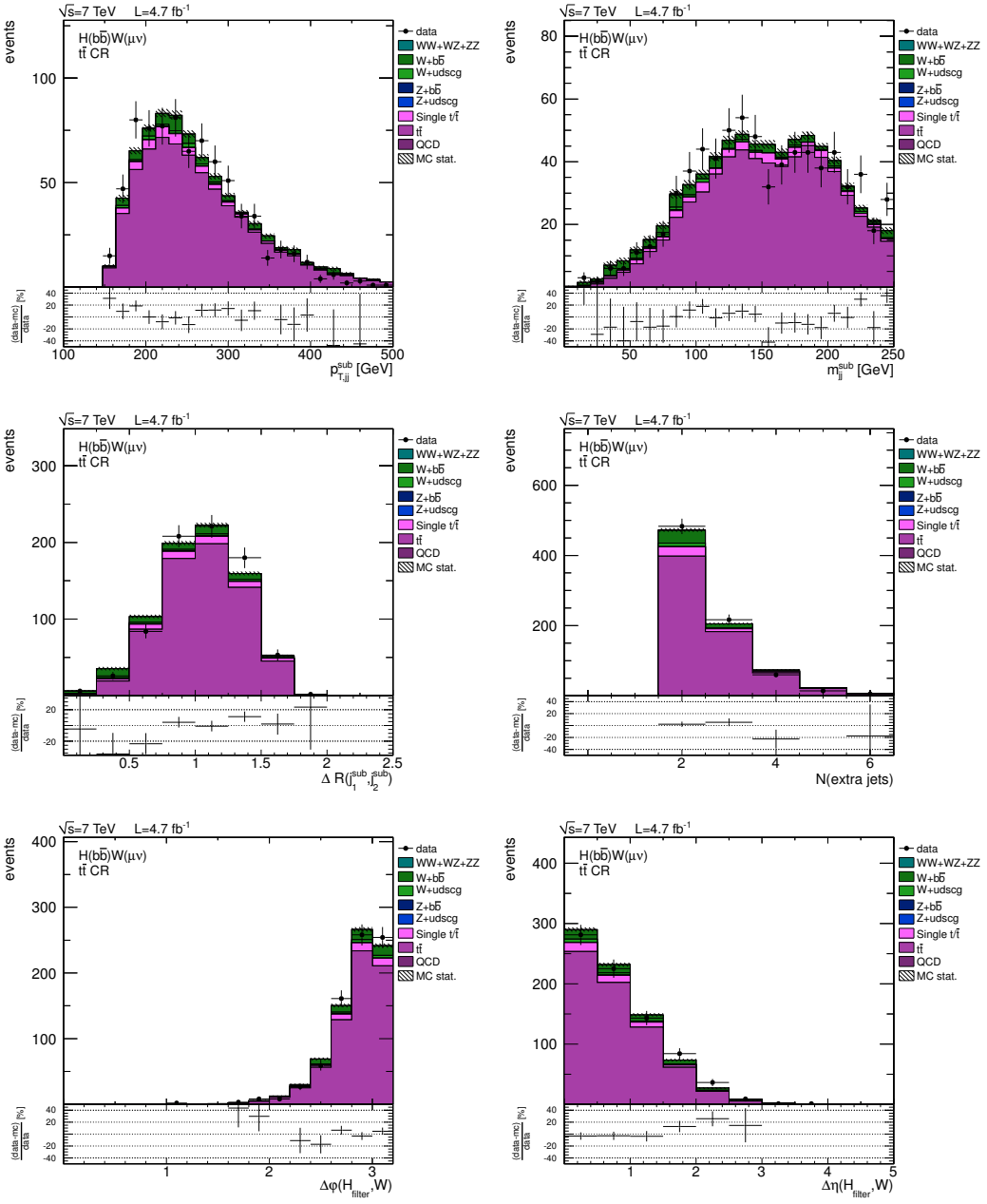


Figure B.18.: Additional $t\bar{t}$ control region distributions for the muon channel using SJF jet reconstruction: the p_T of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

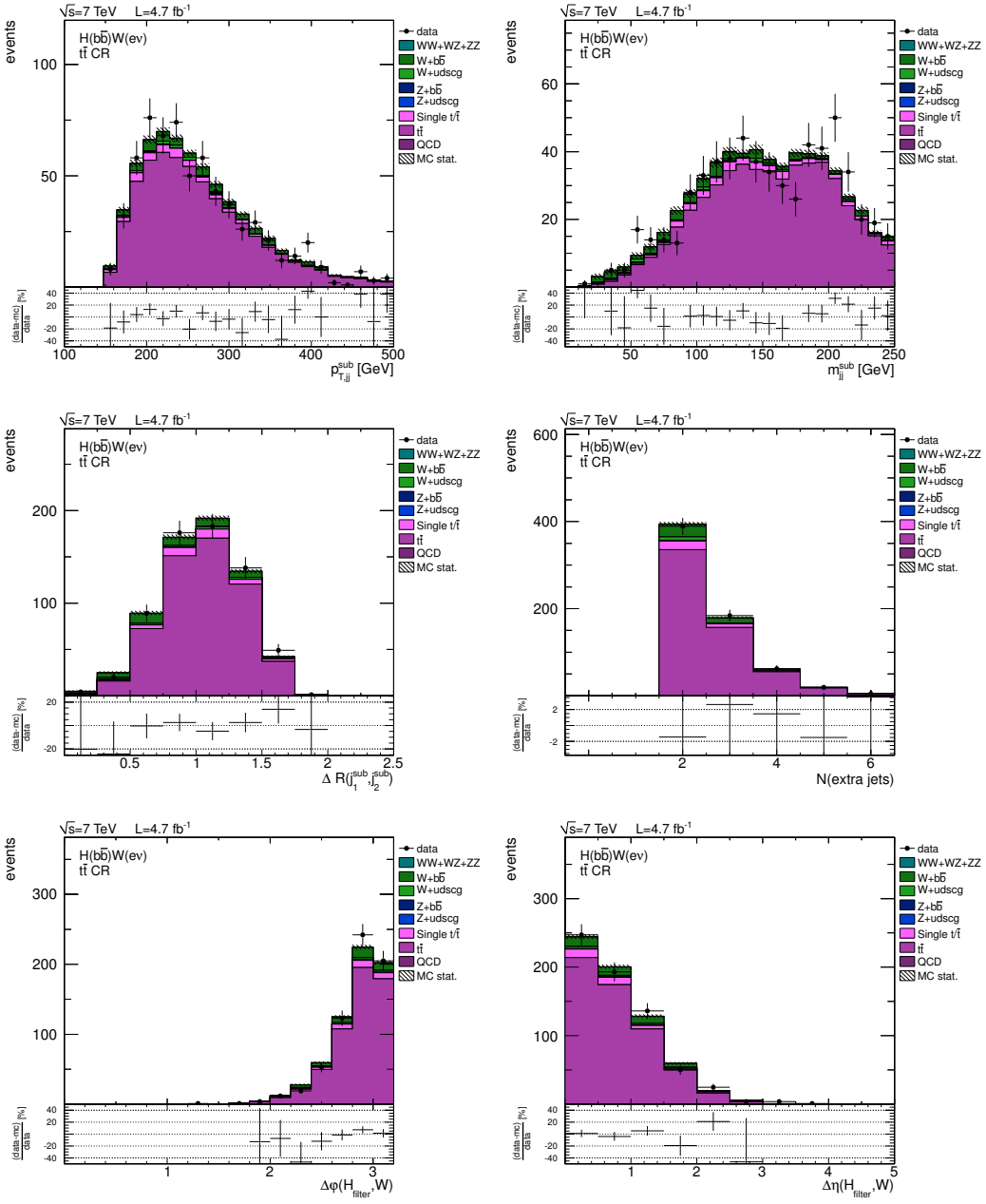


Figure B.19.: Additional $t\bar{t}$ control region distributions for the electron channel using SJF jet reconstruction: the p_T of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

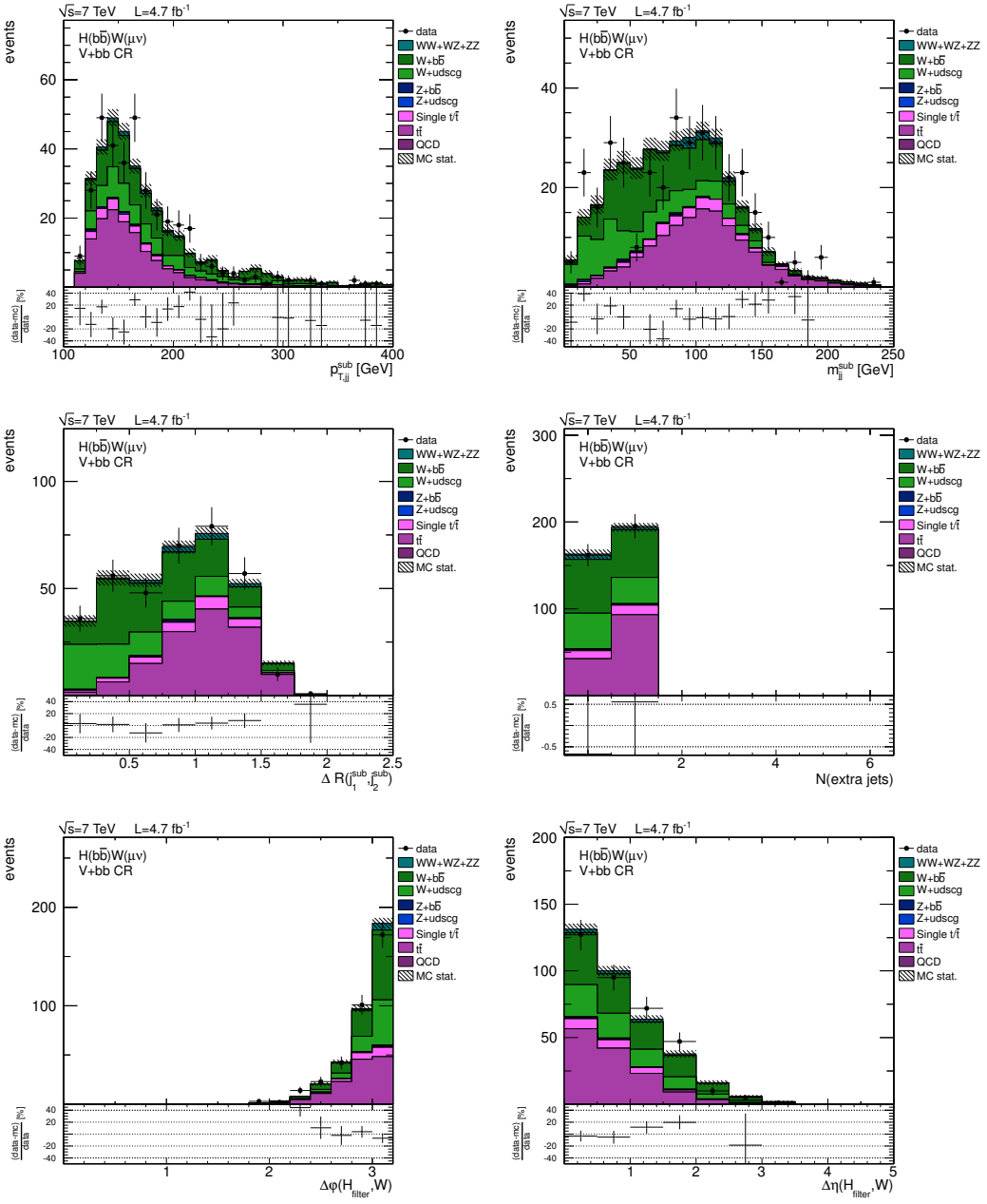


Figure B.20.: Additional W+b$\bar{\text{b}}$ jets control region distributions for the muon channel using SJF jet reconstruction: the p_{T} of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

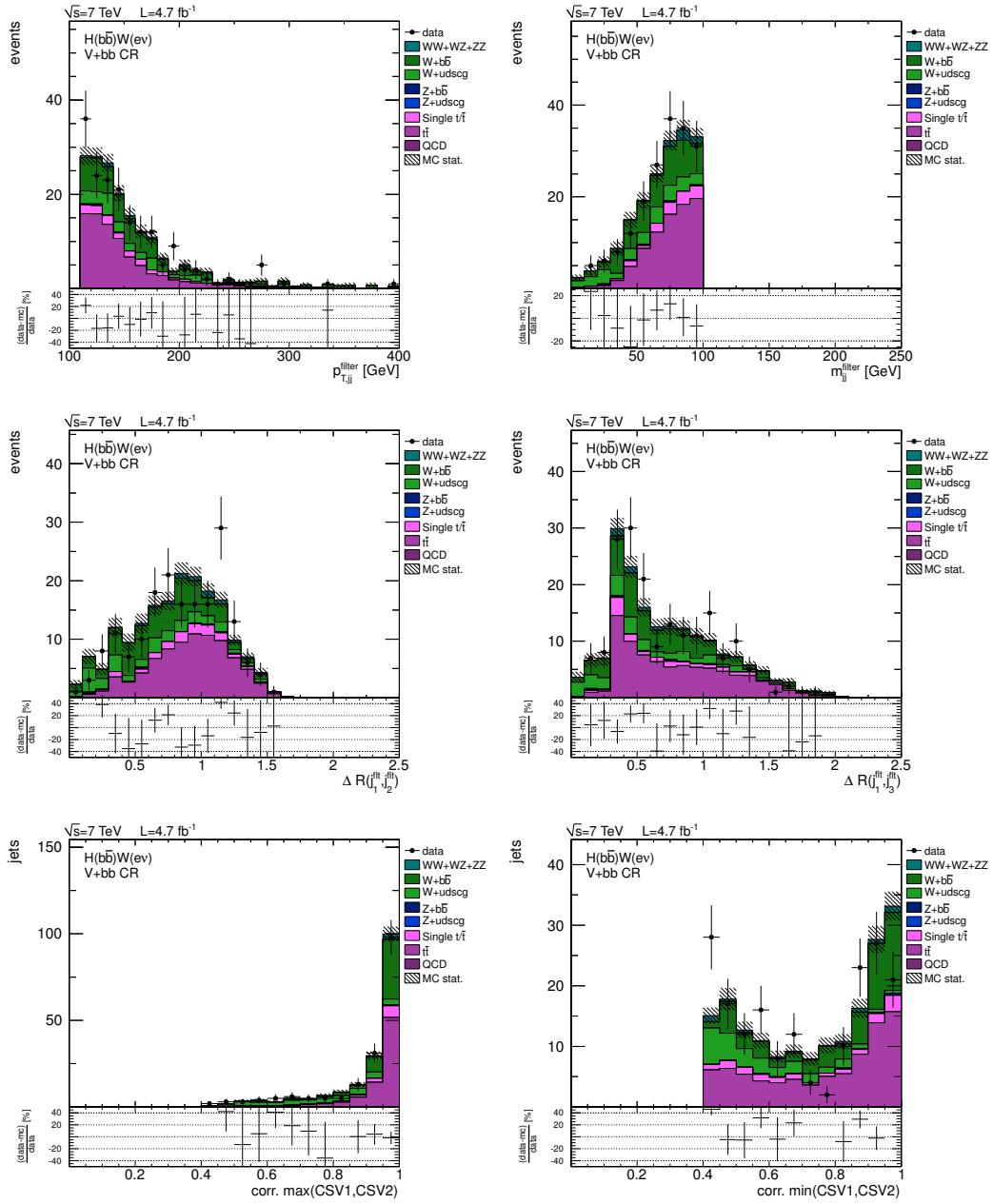


Figure B.21.: $W + b\bar{b}$ jets control region distributions for the electron channel using SJF jet reconstruction: the p_T of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data.

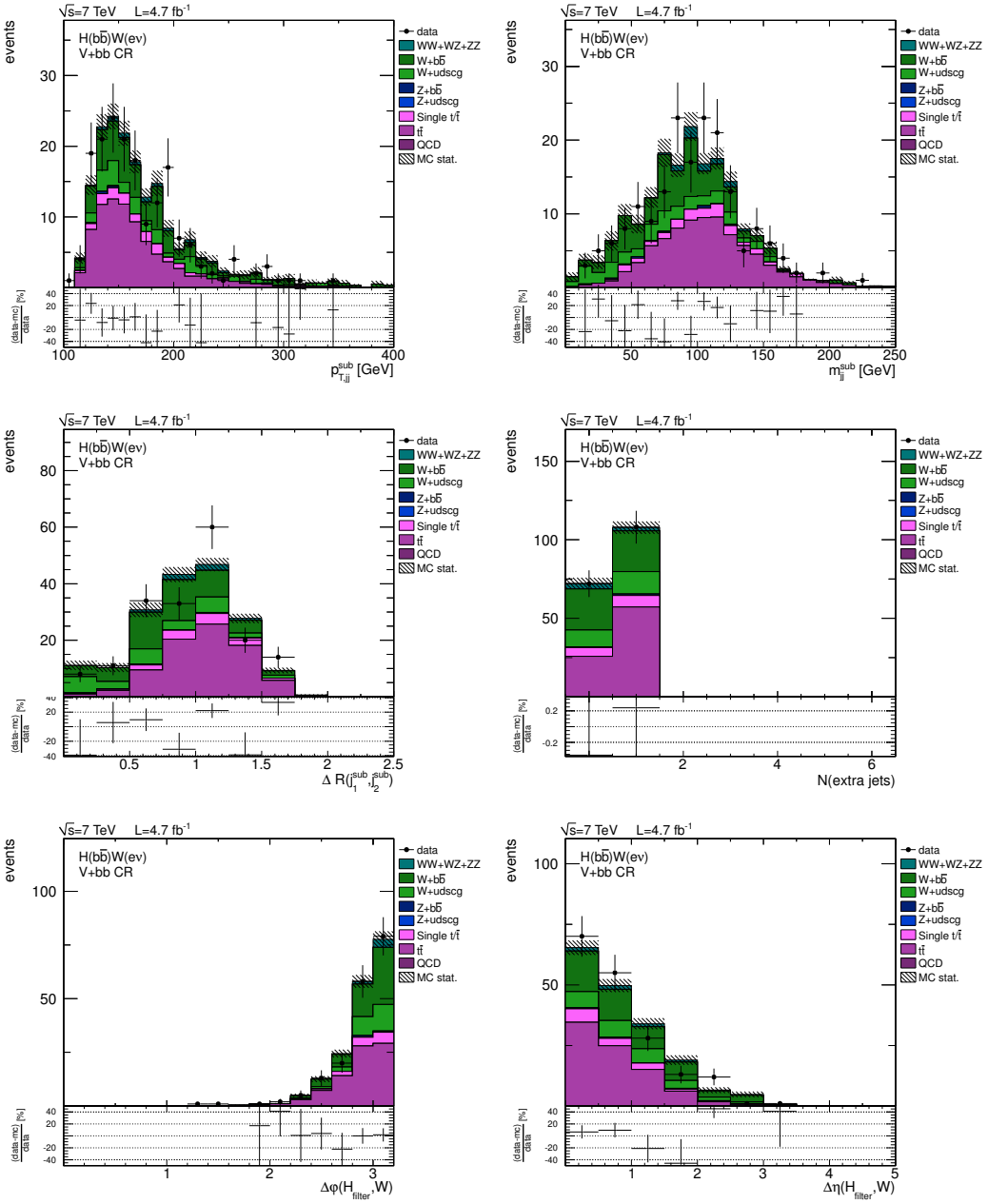


Figure B.22.: Additional $W + b\bar{b}$ jets control region distributions for the electron channel using SJF jet reconstruction: the p_T of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

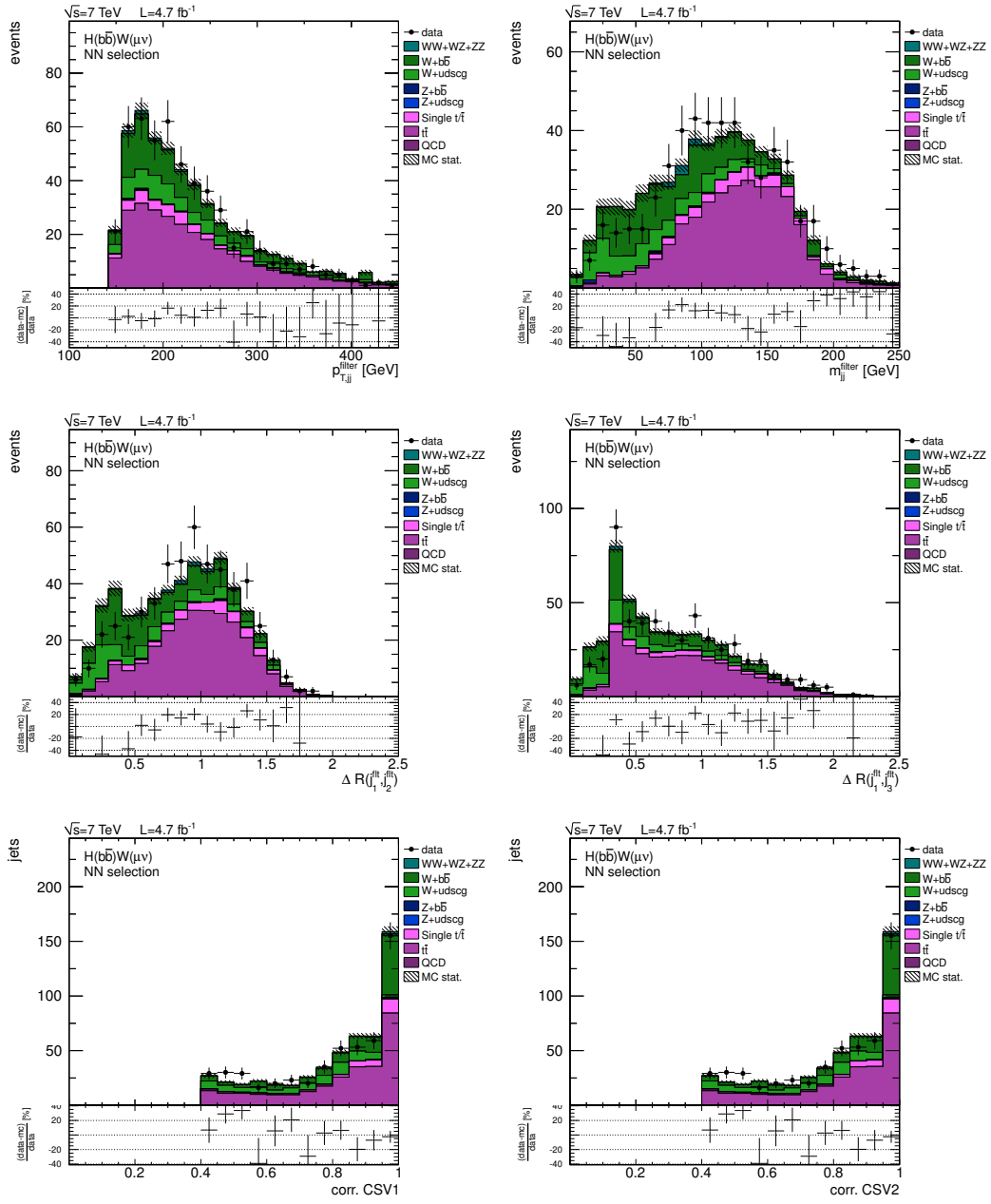


Figure B.23.: Distributions of NN input variables for the muon channel using SJF jet reconstruction: the p_{T} of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data.

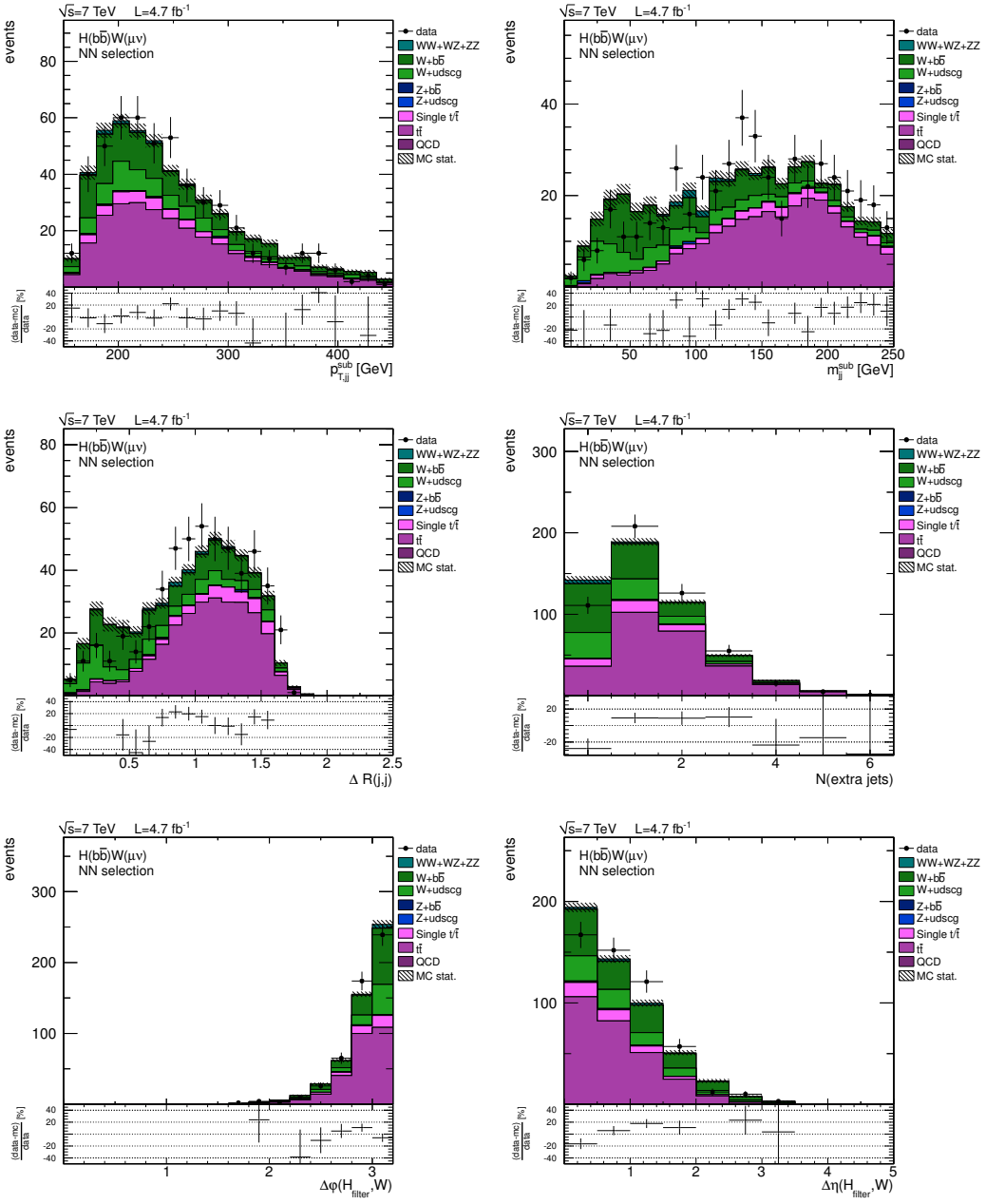


Figure B.24.: Distributions of NN input variables for the muon channel using SJF jet reconstruction (cont.): the p_{T} of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

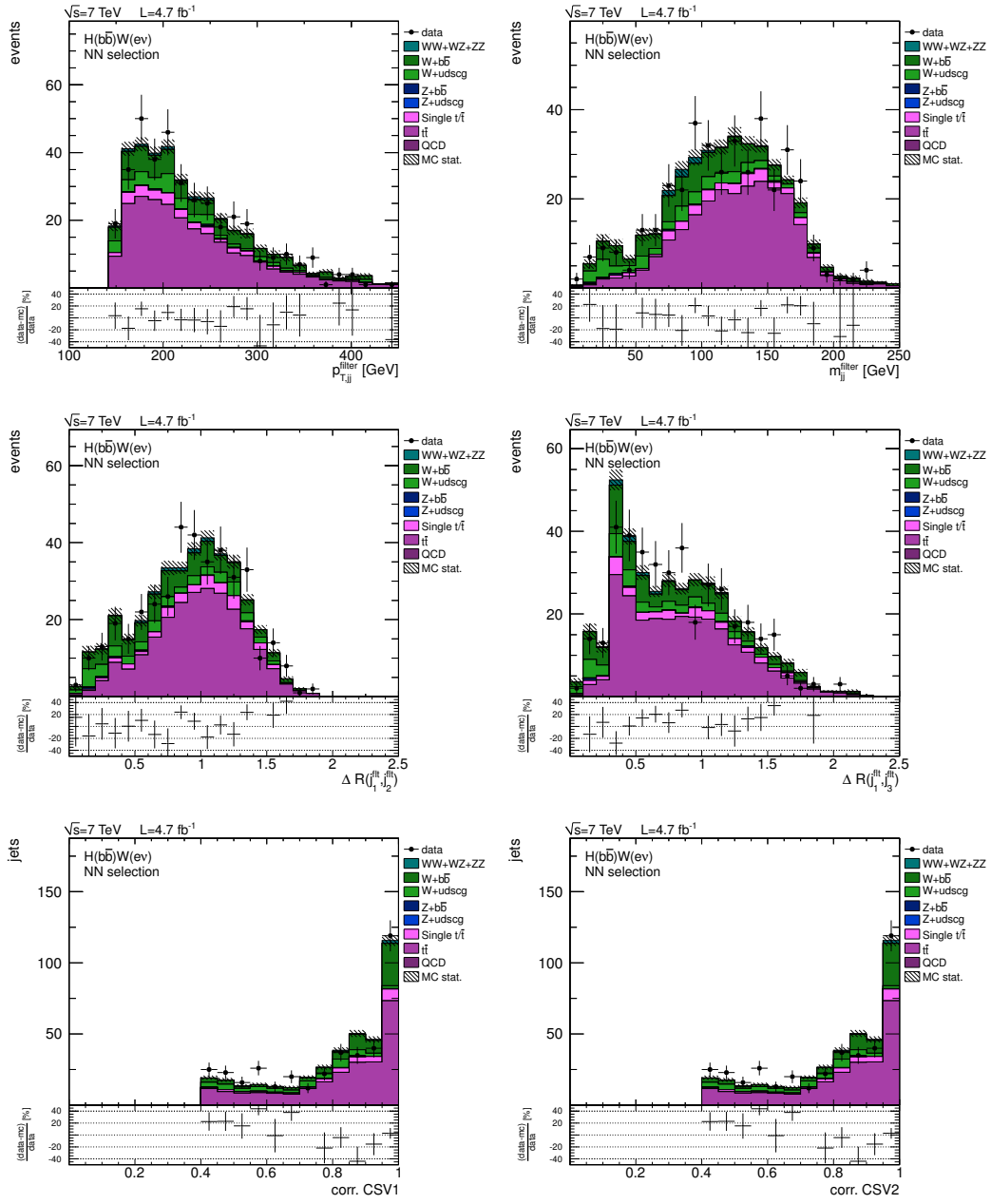


Figure B.25.: Distributions of NN input variables for the electron channel using SJF jet reconstruction: the p_{T} of the tri-filterjet system and the corresponding invariant mass, the angular separation between the first and second, as well as the first and third filterjet and the CSV b-tagging discriminators associated with the first and second filterjet. Scale factors have been applied and the number of simulated events is normalized to data.

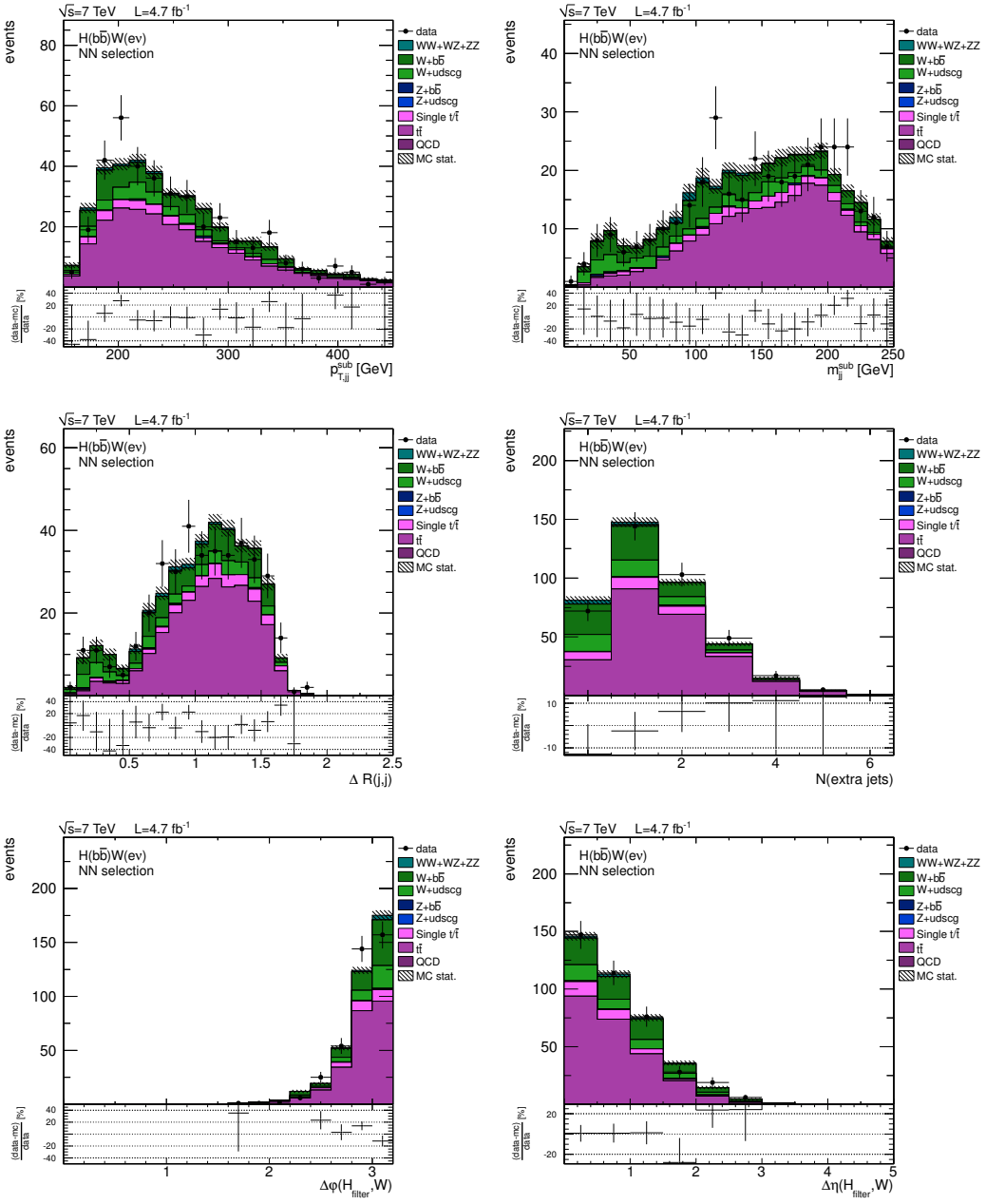


Figure B.26.: Distributions of NN input variables for the electron channel using SJF jet reconstruction (cont.): the p_{T} of the di-subjet system and the corresponding invariant mass, the angular separation of the two subjets, the number of additional jets in the event, the azimuthal opening angle between the tri-filterjet Higgs candidate and the W boson and the difference in η between the two. Scale factors have been applied and the number of simulated events is normalized to data.

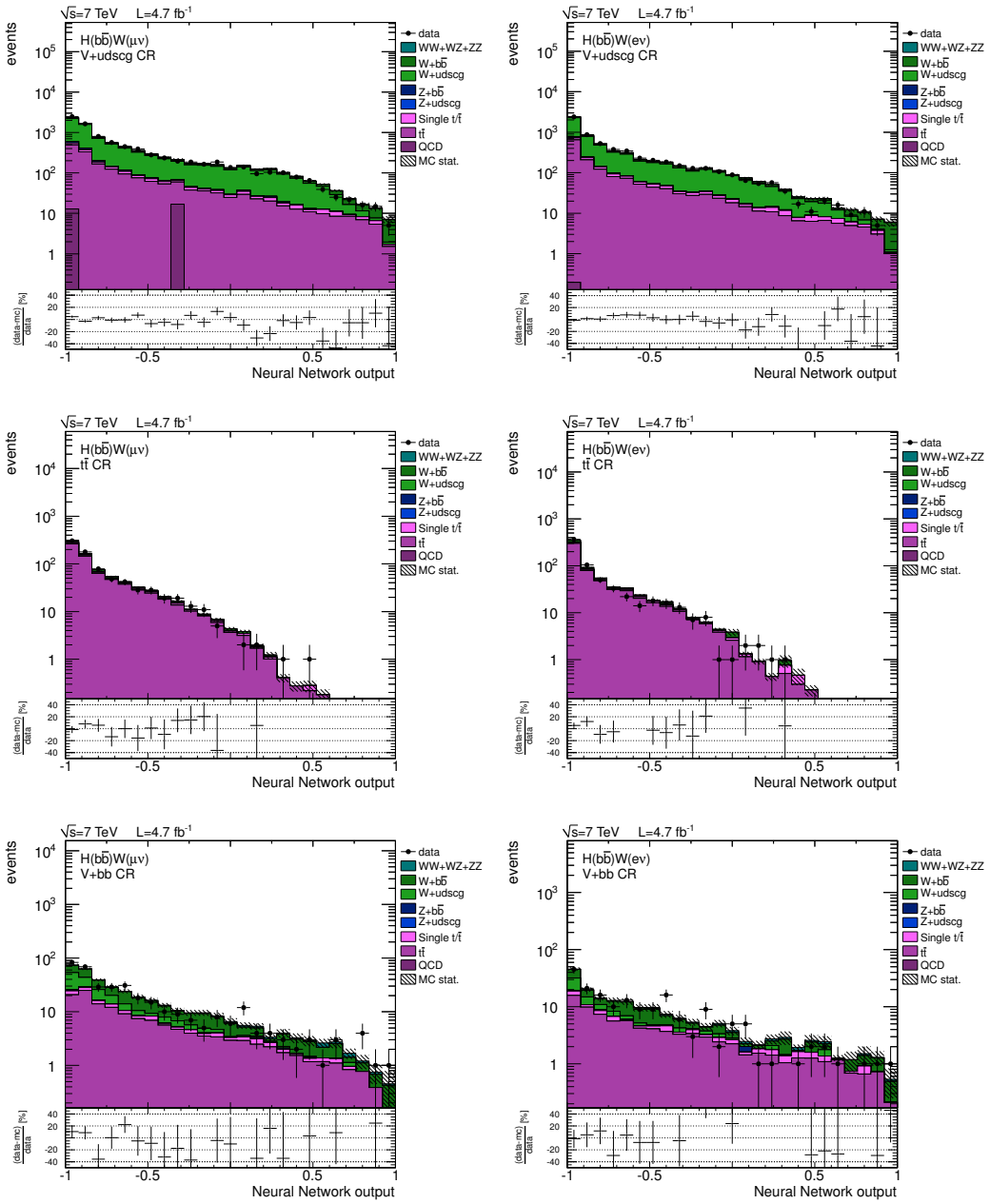
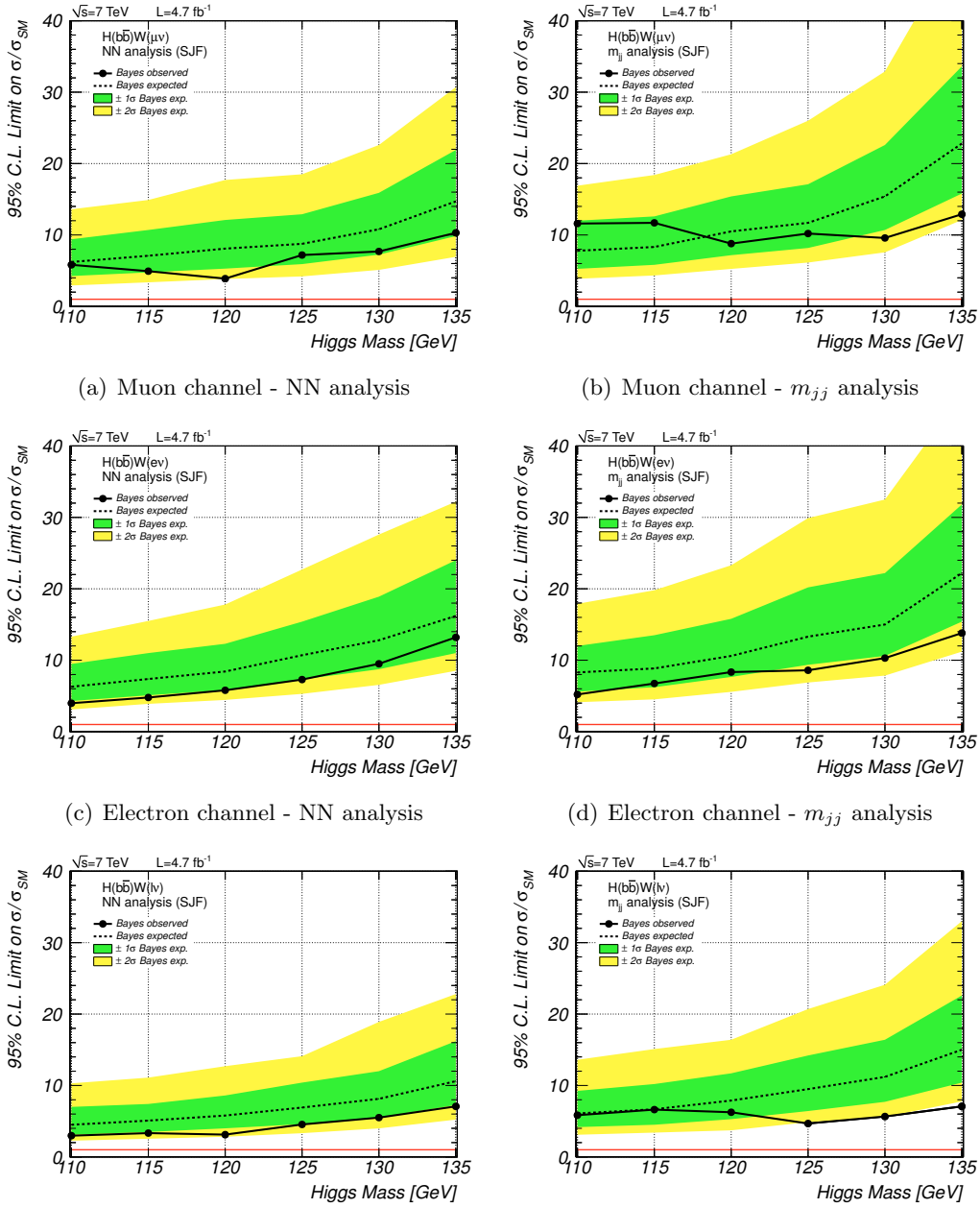


Figure B.27.: Distributions of the NN_{out} discriminator in the different control regions using SJF jet reconstruction, shown here exemplarily for the $NN_{out}(125)$ networks. The left row of plots shows the NN discriminator in the W + udscg, $t\bar{t}$ and W + $b\bar{b}$ control regions, respectively. The number of simulated events is normalized to the number of events observed in data to ease the shape comparison. The right row shows the same distributions in the electron channel for reference. The characteristic shape of each background in simulated events is fully reproduced in the data, giving confidence that the NNs are neither overtrained nor biased in events from collision data.



(a) Muon channel - NN analysis

(b) Muon channel - m_{jj} analysis

(c) Electron channel - NN analysis

(d) Electron channel - m_{jj} analysis

(e) Combination - NN analysis

(f) Combination - m_{jj} analysis

Figure B.28.: Bayesian exclusion limits using SJF jet reconstruction, for the NN analysis (left) and the m_{jj} analysis (right). The top row shows the muon channel result, the middle row the electron channel and the bottom row gives the exclusion limits for the combined HW search. The Bayesian limits are found to be in very good agreement with the corresponding CL_s exclusion limits.

List of Figures

List of Tables

Bibliography

[1] Povh, Rith, Scholz, and Zetsche, *Teilchen und Kerne*. Springer, 6th ed., 2006.

[2] W. Pauli, "The Connection Between Spin and Statistics," *Phys. Rev.*, vol. 58, pp. 716–722, Oct 1940.

[3] K. Nakamura *et al.*, "The Review of Particle Physics," *J. Phys.*, vol. G37, no. 075021, 2010 and 2011 partial update for the 2012 edition.

[4] W. Hollik, "Quantum field theory and the standard model," 2010. arXiv:1012.3883.

[5] L. D. Landau and E. M. Lifschitz, *Lehrbuch der Theoritischen Physik - Band 1. Mechanik*. Verlag Harri Deutsch, 2004.

[6] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory*. Westview Press, 1995.

[7] J. D. Jackson, *Classical Electrodynamics*. John Wiley and Sons, Inc., 3rd ed., 1998.

[8] E. Noether, "Invariante Variationsprobleme," *Nachr. d. König. Ges. d. Wiss. zu Göttingen*, pp. 235–257, 1918.

[9] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[10] A. Salam, *Elementary Particle Theory*. Stockholm: Almquist and Wiksells, 1969.

[11] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry," *Phys. Rev. D*, vol. 2, pp. 1285–1292, Oct 1970.

[12] P. Schmüser, *Feynman-Graphen und Eichtheorien für Experimentalphysiker*. Springer, 1st ed., 1988.

[13] P. W. Anderson, "Plasmons, gauge invariance, and mass," *Phys. Rev.*, vol. 130, pp. 439–442, Apr 1963.

[14] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.

[15] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.

[16] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[17] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model," *Phys. Rept.*, vol. 457, pp. 1–216, 2008. hep-ph/0503172.

[18] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley and Sons, 1st ed., 1984.

[19] J. Goldstone, A. Salam, and S. Weinberg, "Broken symmetries," *Phys. Rev.*, vol. 127, pp. 965–970, Aug 1962.

[20] Super-Kamiokande Collaboration, "Evidence for Oscillation of Atmospheric Neutrinos," *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, Aug 1998.

[21] SNO Collaboration, "Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory," *Phys. Rev. Lett.*, vol. 87, p. 071301, Jul 2001.

[22] M. Kobayashi and T. Maskawa, "*cp*-violation in the renormalizable theory of weak interaction," *Progress of Theoretical Physics*, vol. 49, no. 2, pp. 652–657, 1973.

[23] P. A. M. Dirac, "The Quantum Theory of the Emission and Absorption of Radiation," *Royal Society of London Proceedings Series A*, vol. 114, pp. 243–265, Mar. 1927.

[24] F. J. Dyson, "The *s* matrix in quantum electrodynamics," *Phys. Rev.*, vol. 75, pp. 1736–1755, Jun 1949.

[25] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis," *Journal of High Energy Physics*, vol. 2002, no. 07, p. 012, 2002.

[26] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Parton distributions for the lhc," *The European Physical Journal C - Particles and Fields*, vol. 63, pp. 189–285, 2009. 10.1140/epjc/s10052-009-1072-5.

[27] Martin-Stirling-Thorne-Watt Parton Distribution Functions, `http://projects.hepforge.org/mstwpdf/`.

[28] C. Hackstein, "Searches for the Higgs Boson at the LHC Based on its Couplings to Vector Bosons," 2011. Dissertation, urn:nbn:de:swb:90-235955.

[29] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," CERN, Geneva, 2011. CERN-2011-002.

[30] M. Zeise, "Study of Z Boson Decays into Pairs of Muon and Tau Leptons with the CMS Detector at the LHC," 2011. Dissertation, urn:nbn:de:swb:90-239075.

[31] J. Ellis, J. Espinosa, G. Giudice, A. Hoecker, and A. Riotto, "The probable fate of the Standard Model," *Physics Letters B*, vol. 679, no. 4, pp. 369 – 375, 2009.

[32] J. E. David and Callaway, "Triviality pursuit: Can elementary scalar particles exist?," *Physics Reports*, vol. 167, no. 5, pp. 241 – 320, 1988.

[33] C. Kolda and H. Murayama, "The higgs mass and new physics scales in the minimal standard model," *Journal of High Energy Physics*, vol. 2000, no. 07, p. 035, 2000.

[34] W. Marciano, G. Valencia, and S. Willenbrock, "Renormalization-group-improved unitarity bounds on the higgs-boson and top-quark masses," *Phys. Rev. D*, vol. 40, pp. 1725–1729, Sep 1989.

[35] J. Casas, J. Espinosa, and M. Quiros, "Standard model stability bounds for new physics within LHC reach," *Physics Letters B*, vol. 382, no. 4, pp. 374 – 382, 1996.

[36] A. Sirlin, "Radiative corrections in the $\mathrm{SU}(2)_L \times \mathrm{U}(1)$ theory: A simple renormalization framework," *Phys. Rev. D*, vol. 22, pp. 971–981, Aug 1980.

[37] P. Schieferdecker, "Measurement of the Top Quark Mass at DØ Run II with the Matrix Element Method in the Lepton+Jets Final State," 2005. Dissertation, urn:nbn:de:bvb:19-44217.

[38] M. W. Grunewald, "Precision Electroweak Measurements and Constraints on the Standard Model," Dec 2010. CERN-PH-EP-2010-095, updated results from the 2011 Summer Conferences taken from `http://lepewwg.web.cern.ch/LEPEWWG/`.

[39] Gfitter Group, "Updated status of the global electroweak fit and constraints on new physics," Jul 2011. CERN-OPEN-2011-033.

[40] G. Abbiendi, "Search for the Standard Model Higgs Boson at LEP," *Phys. Lett. B*, vol. 565, pp. 61–75. 22 p, Mar 2003.

[41] The TEVNPH Working Group for the CDF and D0 Collaborations, "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to $8.6\,\mathrm{fb}^{-1}$ of Data," Sep 2011.

[42] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV.," Feb 2012. Comments: Submitted to Physics Letters B.

[43] ATLAS Collaboration, "Combination of Higgs Boson Searches with up to 4.9fb^{-1} of pp Collision Data Taken at $\sqrt{s} = 7$ TeV with the ATLAS Experiment at the LHC," Dec 2011. ATLAS-CONF-2011-163.

[44] "Convention for the establishment of a European organization for nuclear research: Paris, 1st July, 1953 : as amended. Convention pour l'établissement d'une Organisation européenne pour la Recherche nucléaire. Paris, le 1er juillet 1953 : telle qu'elle a été modifiée," 1971. CERN document.

[45] C. Rubbia, "Experimental observation of the intermediate vector bosons W+, W-, and Z0," *Rev. Mod. Phys.*, vol. 57, no. CERN-OPEN-94-004, pp. 699–722, 1994. Nobel lecture.

[46] "The World Wide Web: Proposal for a HyperText Project", `http://www.w3.org/Proposal.html`.

[47] "The Large Electron-Positron Collider", `http://public.web.cern.ch/public/en/research/lep-en.html`.

[48] "The Tevatron", `http://www.fnal.gov/pub/science/accelerator/index.html`.

[49] ATLAS Collaboration, "Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS Detector at the LHC," *Phys. Rev. Lett.*, vol. 105, p. 252303 . 19 p, Nov 2010.

[50] CMS Collaboration, "Observation and studies of jet quenching in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV," Feb 2011. CMS-PAS-HIN-10-004.

[51] C. Lefevre, "LHC: the guide (english version)." Oct 2006.

[52] L. Evans and P. Bryant, "LHC Machine," *Journal of Instrumentation*, vol. 3, no. 08, p. S08001, 2008.

[53] L. Rossi, "Superconductivity: its role, its success and its setbacks in the Large Hadron Collider of CERN," *Superconductor Science and Technology*, vol. 23, no. 3, p. 034001, 2010.

[54] LHC Beam Commissioning Working Group and W. Herr, "LHC machine commissioning and the near future", `http://lhc-commissioning.web.cern.ch/lhc-commissioning/presentations/2010/lhc-status-split-oct10.pdf`.

[55] J. Wenninger, "LHC status report at the Run Coordination Workshop 2011 in Madrid", `http://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=123738`.

[56] CMS Collaboration, "Photos from the cms photo book." CMS Collection, Aug 2008.

[57] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[58] K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC," *JINST*, vol. 3, p. S08002, 2008.

[59] A. A. Alves *et al.*, "The LHCb Detector at the LHC," *JINST*, vol. 3, p. S08005, 2008.

[60] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, no. 08, p. S08007, 2008.

[61] O. Adriani *et al.*, "Technnical Proposal for the CERN LHCf Experiment: Measurement of Photons and Neutral Pions in the Very Forward Region of LHC," Sep 2005. CERN-LHCC-2005-032.

[62] CMS Outreach, "The Media Center", `https://cms.web.cern.ch/cms/Media/Images/`.

[63] CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST*, vol. 3 S08004, p. 361, 2008.

[64] G. L. Bayatian *et al.*, "CMS Physics Technical Design Report Volume I: Detector Performance and Software," 2006. CERN-LHCC-2006-001.

[65] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}," Apr 2009. CMS-PAS-PFT-09-001.

[66] H. Held, "A Search for the Decay $B_s^0 \to K^{*\pm}K^{\mp}$ at the CDF II Experiment using Artificial Neural Networks," 2009. Diplomarbeit, IEKP-KA/2009-3.

[67] N. Metropolis and S. Ulam, "The Monte Carlo Method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[68] M. Renz, "Erste Messung des Wirkungsquerschnitts der Top-Quark-Paarproduktion bei $\sqrt{s} = 7$ TeV im Elektron+Jets Kanal mit dem CMS-Experiment," 2011. Dissertation, urn:nbn:de:swb:90-236748.

[69] V. Gribov and L. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov.J.Nucl.Phys.*, vol. 15, pp. 438–450, 1972.

[70] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.*, vol. B126, p. 298, 1977.

[71] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics," *Sov. Phys. JETP*, vol. 46, pp. 641–653, 1977.

[72] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, "QCD Matrix Elements + Parton Showers," *Journal of High Energy Physics*, vol. 2001, no. 11, p. 063, 2001.

[73] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-pair production in hadronic collisions," *Journal of High Energy Physics*, vol. 2007, no. 01, p. 013, 2007.

[74] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics," *Physics Reports*, vol. 97, no. 2-3, pp. 31 – 145, 1983.

[75] B.R. Webber, "A QCD model for jet fragmentation including soft gluon interference," *Nuclear Physics B*, vol. 238, no. 3, pp. 492 – 528, 1984.

[76] M. Heinrich, "A Jet Based Approach to Measuring Soft Contributions to Proton-Proton Collisions with the CMS Experiment," 2011. Dissertation, urn:nbn:de:swb:90-218398.

[77] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *Journal of High Energy Physics*, vol. 2006, no. 05, p. 026, 2006.

[78] M. Bähr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson, M. Seymour, A. Sherstnev, and B. Webber, "Herwig++ physics and manual," *The European Physical Journal C - Particles and Fields*, vol. 58, pp. 639–707, 2008. 10.1140/epjc/s10052-008-0798-9.

[79] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. Seymour, and B. Webber, "HERWIG 6.5 Release Note," Oct 2002. arXiv:hep-ph/0210213.

[80] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer, "MadGraph/MadEvent v4: the new web generation," *Journal of High Energy Physics*, vol. 2007, no. 09, p. 028, 2007.

[81] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms," *Journal of High Energy Physics*, vol. 2004, no. 11, p. 040, 2004.

[82] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *Journal of High Energy Physics*, vol. 2007, no. 11, p. 070, 2007.

[83] E. Re, "Single-top production with the POWHEG method," *PoS*, vol. DIS2010, p. 172, 2010.

[84] K. Hamilton, P. Richardson, and J. Tully, "A positive-weight Next-to-Leading Order Monte Carlo simulation for Higgs boson production," *Journal of High Energy Physics*, vol. 2009, no. 04, p. 116, 2009.

[85] S. A. et al., "Geant4: A simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.

[86] J. A. et al., "Geant4 developments and applications," *IEEE Transactions on Nuclear Science*, vol. 53, no. 1, pp. 270 – 278, 2006.

[87] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction with the first LHC collisions recorded in the CMS detector," Mar 2010. CMS-PAS-PFT-10-001.

[88] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV," Jul 2010. CMS-PAS-PFT-10-002.

[89] CMS Collaboration, "Commissioning of the Particle-Flow event reconstruction with leptons from J/Psi and W decays at 7 TeV," Jul 2010. CMS-PAS-PFT-10-003.

[90] W. Adam, B. Mangano, T. Speer, and T. Todorov, "Track Reconstruction in the CMS tracker," Dec 2006. CMS-NOTE-2006-041.

[91] S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, "Track reconstruction, primary vertex finding and seed generation with the Pixel Detector," Jan 2006. CMS-NOTE-2006-026.

[92] T. Speer and W. Adam and R. Frühwirth and A. Strandlie and T. Todorov and M. Winkler, "Track reconstruction in the CMS tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 559, no. 1, pp. 143 – 147, 2006.

[93] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2-3, pp. 444 – 450, 1987.

[94] M. Winkler and T. Todorov, "A comparative study of track reconstruction methods in the context of CMS physics," May 2002. CMS-THESIS-2002-015.

[95] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting," Mar 2007. CMS-NOTE-2007-008.

[96] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions," Jul 2010. CMS-PAS-TRK-10-005.

[97] CMS Particle Flow Physics Object Group, "Particle Flow Reconstruction of Jets, Taus, and MET," 2009. CMS-AN-2009-039.

[98] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV," Jul 2010. CMS-PAS-MUO-10-002.

[99] W. Adam, R. Früwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC," *Journal of Physics G: Nuclear and Particle Physics*, vol. 31, no. 9, p. N9, 2005.

[100] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV," Aug 2010. CMS-PAS-EGM-10-004.

[101] M. Pioppi, "Electron Pre-Identification in the Particle Flow Framework," 2008. CMS-AN-2008-032.

[102] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi, "Electron Reconstruction within the Particle Flow Algorithm," 2009. CMS-AN-2009-034.

[103] M. Cacciari and G. P. Salam, "Dispelling the N**3 myth for the k(t) jet-finder," *Phys. Lett.*, vol. B641, pp. 57–61, 2006. `http://fastjet.fr`.

[104] G. P. Salam and G. Soyez, "A practical seedless infrared-safe cone jet algorithm," *Journal of High Energy Physics*, vol. 2007, no. 05, p. 086, 2007.

[105] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions," *Nucl. Phys.*, vol. B406, pp. 187–224, 1993.

[106] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev. D*, vol. 48, pp. 3160–3166, Oct 1993.

[107] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

[108] Yu.L. Dokshitzer and G.D. Leder and S. Moretti and B.R. Webber, "Better jet clustering algorithms," *Journal of High Energy Physics*, vol. 1997, no. 08, p. 001, 1997.

[109] G. P. Salam, "Elements of QCD for hadron colliders," p. 56, Nov 2010. arXiv:1011.5131.

[110] CMS Collaboration, "Study of Jet Substructure in pp Collisions at 7 TeV in CMS ," Feb 2011. CMS-PAS-JME-10-013.

[111] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider," *Phys. Rev. Lett.*, vol. 100, p. 242001, Jun 2008.

[112] CMS Collaboration, "Jet Performance in pp Collisions at $\sqrt{s} = 7\,\text{TeV}$," Jul 2010. CMS-PAS-JME-10-003.

[113] CMS Collaboration, "Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS," Aug 2009. CMS-PAS-JME-09-002.

[114] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS," *Journal of Instrumentation*, vol. 6, no. 11, p. P11002, 2011.

[115] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Physics Letters B*, vol. 659, no. 1-2, pp. 119 – 126, 2008.

[116] CMS Collaboration, "Algorithms for b Jet Identification in CMS," Jul 2009. CMS-PAS-BTV-09-001.

[117] CMS Collaboration, "Missing transverse energy performance of the CMS detector," *Journal of Instrumentation*, vol. 6, no. 09, p. P09001, 2011.

[118] L. Tuura, A. Meyer, I. Segoni, and G. D. Ricca, "CMS data quality monitoring: Systems and experiences," *Journal of Physics: Conference Series*, vol. 219, no. 7, p. 072020, 2010.

[119] J. Ott, Th. Müller, and J. Wagner-Kuhr, "Theta - A Framework for Template-based Modeling and Inference," 2010. `http://theta-framework.org/`.

[120] M. Feindt, "A Neural Bayesian Estimator for Conditional Probability Densities," 2004. arXiv:physics/0402093.

[121] M. Feindt and U. Kerzel, "The NeuroBayes neural network package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 559, no. 1, pp. 190 – 194, 2006.

[122] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*. Teubner Verlag, 1998.

[123] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)," 2000. CERN-OPEN-2000-205.

[124] T. Junk, "Confidence level computation for combining searches with small statistics," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 434, no. 2-3, pp. 435 – 443, 1999.

[125] CMS Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011," 2011. CMS-NOTE-2011-005.

[126] J. Ott, "Studien zur Suche nach schweren in Top-Quark-Paare zerfallenden Teilchen mit dem CMS-Detektor," 2009. Diplomarbeit, IEKP-KA/2009-20.

[127] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV using the kinematic properties of events with leptons and jets," *The European Physical Journal C - Particles and Fields*, vol. 71, pp. 1–27, 2011.

[128] CMS Collaboration, "Measurement of the charge asymmetry in top-quark pair production in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Physics Letters B*, vol. 709, no. 1-2, pp. 28 – 49, 2012.

[129] Y. LeCun, "Learning Processes in an Asymmetric Threshold Network," *Disordered systems and biological organization*, pp. 233–240, 1986.

[130] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Parallel distributed processing: Explorations in macrostructure of cognition*, pp. 318–362, 1986.

[131] R. Field, "Early LHC Underlying Event Data - Findings and Surprises," 2010. arXiv:1010.3558.

[132] H. Held, P. Schieferdecker, and M. Voutilainen, "Update of the Measurement of the Jet p_{T} Resolution in $\sqrt{s} = 7\,\mathrm{TeV}$ Collision Data with the Asymmetry Method," 2010. CMS-AN-2010/371.

[133] A. Harel and P. Schieferdecker, "Calorimeter Jet Quality Criteria for the First CMS Collision Data, and preparations for calibrating their efficiencies," 2009. CMS-AN-2009/087.

[134] N. Saoulidou, "Particle Flow Jet Identification Criteria," 2010. CMS-AN-2010/003.

[135] R. Ciesielski, H. Held, P. Schieferdecker, and M. Zielinski, "Jet Energy Resolutions Derived from QCD Simulation for the Analysis of First $\sqrt{s} = 7\,\mathrm{TeV}$ Collision Data," 2010. CMS-AN-2010/121.

[136] DØ Collaboration, "High p_{T} jets in $p\bar{p}$ collisions at $\sqrt{s} = 630\,\mathrm{GeV}$ and $1800\,\mathrm{GeV}$," *Phys.Rev.D64*, vol. 032003, 2000.

[137] CMS Collaboration, "Measurement of the Jet Energy Resolutions and Jet Reconstruction Efficiency at CMS," Jul 2009. CMS-PAS-JME-09-007.

[138] C. Dragoui, H. Held, P. Schieferdecker, and M. Zielinski, "Measurement of the Jet p_{T} Resolution in $\sqrt{s} = 7\,\mathrm{TeV}$ Collision Data with the Asymmetry Method," 2010. CMS-AN-2010/134.

[139] M. Voutilainen, "Measurement of the Inclusive Jet Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96\,\text{TeV}$," *Helsinki Institue of Physics Internal Report Series*, Aug 2008. Dissertation, HIP-2008-05.

[140] M. Voutilainen, F. Pandolfi, N. Saoulidou, P. Schieferdecker, H. Held, J. Komaragiri, D. Martschei, S. Honc, Th. Müller, M. Feindt, S. de Visscher, M. Chadwick, A. Rizzi, W. Adam, M. Kaya, and O. Kaya, "Measurement of the Inclusive Jet and b-Jet Production in pp Collisions at $\sqrt{s} = 7\,\text{TeV}$ using Particle Flow," 2010. CMS-AN-2010/128.

[141] P. Bhat, R. Cavanaugh, D. Duggan, J. Haas, A. Heister, H. Held, M. Kaya, O. Kaya, O. Kodolova, P. Lawson, L. Lebolo, A.-M. Magnan, D. Mason, A. Nikitenko, A. Oehler, F. Pandolfi, H. Prosper, K. Rabbertz, K. Rose, N. Saoulidou, P. Schieferdecker, F. Stober, M. Voutilainen, and J. Weng, "Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 7\,\text{TeV}$," 2010. CMS-AN-2010/439.

[142] M. Feindt, H. Held, S. Honc, J. Komaragiri, D. Martschei, Th. Müller, P. Schieferdecker, and J. Wagner-Kuhr, "Extraction of Heavy Flavour Content using a Template Fitting Procedure for the Inclusive b-Jet cross section measurement at $\sqrt{s} = 7\text{TeV}$," 2010. CMS-AN-2010/174.

[143] CMS Collaboration, "Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV," *Phys. Rev. Lett.*, vol. 107, p. 132001, Sep 2011.

[144] UA2 Collaboration, "Observation of very large transverse momentum jets at the CERN $p\bar{p}$ collider," *Physics Letters B*, vol. 118, no. 1-3, pp. 203 – 210, 1982.

[145] UA1 Collaboration, "Hadronic jet production at the CERN proton-antiproton collider," *Physics Letters B*, vol. 132, no. 1-3, pp. 214 – 222, 1983.

[146] A. Oehler, "Strategy for an Initial Measurement of the Inclusive Jet Cross Section with the CMS Detector," 2009. Dissertation, urn:nbn:de:swb:90-149218.

[147] CDF Collaboration, "Measurement of the inclusive jet cross section using the k_{T} algorithm in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF II detector," *Phys. Rev. D*, vol. 75, p. 092006, May 2007.

[148] CDF Collaboration, "Measurement of the inclusive jet cross section at the Fermilab Tevatron $p\bar{p}$ collider using a cone-based jet algorithm," *Phys. Rev. D*, vol. 78, p. 052006, Sep 2008.

[149] DØ Collaboration, "Measurement of the Inclusive Jet Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.*, vol. 101, p. 062001, Aug 2008.

[150] ATLAS Collaboration, "Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector," *The European Physical Journal C - Particles and Fields*, vol. 71, pp. 1–59, 2011.

[151] CMS Collaboration, "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7\,\text{TeV}$," Jul 2010. CMS-PAS-BPH-10-009.

[152] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 2.* Technical Design Report ATLAS, Geneva: CERN, 1999.

[153] CMS Collaboration, "CMS physics Technical Design Report, Volume II: Physics Performance," *J. Phys. G*, vol. 34, no. CERN-LHCC-2006-021. CMS-TDR-008-2, pp. 995–1579. 669 p, 2006.

[154] N. Kidonakis, "Higher-order corrections to top-antitop pair and single top quark production," 2009. arXiv:0909.0037.

[155] J. Berryhill, T. Boccali, P. Bortignon, S. Das, M. de Gruttola, M. Fisher, I. Furic, C. Grab, H. Held, J. Konigsberg, D. L. Pegna, N. Mohr, M. Mooney, J. Olsen, A. Rizzi, P. Schieferdecker, and S. Zenz, "Search for the Standard Model Higgs Boson Produced in Association with a W or Z and Decaying to Bottom Quarks," 2011. CMS-AN-2011-430.

[156] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production," *Phys. Rev. D*, vol. 83, p. 091503, May 2011.

[157] N. Kidonakis, "Next-to-next-to-leading logarithm resummation for s-channel single top quark production," *Phys. Rev. D*, vol. 81, p. 054028, Mar 2010.

[158] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-," *Phys. Rev. D*, vol. 82, p. 054018, Sep 2010.

[159] CMS Collaboration, "", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSections`.

[160] CMS Collaboration, "Pile-Up Reweighting Instructions", `https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupReweighting`.

[161] CMS Higgs Physics Analysis Group, "Trigger Strategies for Higgs Searches," 2011. CMS-AN-2011-065; comment: corresponding trigger efficiency functions provided by A. Rizzi and D. Lopes-Pegna in private communicaton.

[162] CMS Vector Boson Task Force, "Electron ID Working Points", `https://twiki.cern.ch/twiki/bin/view/CMS/VbtfEleID2011`.

[163] J. Berryhill, T. Boccali, P. Bortignon, M. Chen, S. Das, M. de Gruttola, M. Fisher, I. Furic, C. Grab, H. Held, J. Konigsberg, K. Mishra, M. Mooney, J. Olsen, D. L. Pegna, A. Rizzi, P. Schieferdecker, and D. Wright, "Search for the Standard Model Higgs Boson Produced in Association with a W or Z and Decaying to Bottom Quarks," 2011. CMS-AN-2011-240.

[164] CMS Collaboration, "Search for the standard model Higgs boson decaying to bottom quarks in pp collisions at $\sqrt{s} = 7\,\text{TeV}$," *Physics Letters B*, vol. 710, no. 2, pp. 284 – 306, 2012.

[165] G. Punzi, "Sensitivity of Searches for New Signals and Its Optimization," *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, p. 79, 2003. arXiv:physics/0308063.

[166] CMS Collaboration, "Measurement of the b-tagging efficiency using $\text{t}\bar{\text{t}}$ events," 2011. CMS-PAS-BTV-11-003.

[167] CMS Collaboration, "Measurement of the CMS Luminosity," 2010. CMS-PAS-EWK-10-004.

[168] M. L. Ciccolini, S. Dittmaier, and M. Krämer, "Electroweak radiative corrections to associated *WH* and *ZH* production at hadron colliders," *Phys. Rev. D*, vol. 68, p. 073003, Oct 2003.

[169] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, "EW corrections to Higgs strahlung at the Tevatron and the LHC with HAWK," *Proceedings in Science*, Oct 2011. PoS(EPS-HEP2011)235.

[170] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated Higgs-W-Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at NNLO," *Physical Review Letters*, vol. 107, p. 152003, Oct. 2011.

[171] CMS Collaboration, "Jet Resolutions from Dijet Events Status Report", `https://indico.cern.ch/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=146429` and private communication with M. Zielinski.

[172] CMS Collaboration, "Measurement of the t-Channel Single Top Quark Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV," *Phys. Rev. Lett.*, vol. 107, p. 091802, Aug 2011.

[173] CMS Collaboration, "Search for a Higgs boson decaying into two photons in the CMS detector," 2011. CMS-PAS-HIG-11-030.

[174] CMS Collaboration, "Search for the Higgs Boson Decaying to W^+W^- in the Fully Leptonic Final State," 2011. CMS-PAS-HIG-11-024.

[175] CMS Collaboration, "Search for Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7\,\text{TeV}$," 2011. CMS-PAS-HIG-11-029.

Danksagung

Mein herzlicher Dank gilt Prof. Dr. Thomas Müller für die Aufnahme in seine Arbeitsgruppe, die gute Betreuung meiner Arbeit und die kontinuierliche Unterstützung während der letzten drei Jahre. Seine Entscheidung, mir die Möglichkeit eines 2-jährigen Forschungsaufenthalts am CERN zu eröffnen, habe ich als sehr großen Vertrauensbeweis empfunden, und sehe sie als einen wesentlichen Faktor zum erfolgreichen Gelingen dieser Arbeit an. Für die Übernahme des Korreferats danke ich Prof. Dr. Günter Quast.

Besonders hervorheben möchte ich den Dank an meinen Betreuer Philipp Schieferdecker für die tolle, freundschaftliche Zusammenarbeit während der letzten Jahre. Seine fröhlich-fordernde Art Wissenschaft zu erleben ist sehr erfrischend und hat mich auch in den intensiven Forschungsphasen immer wieder stark motiviert. Dass ich mich selbst nach seiner Neuorientierung in die Wirtschaft zu jeder Zeit bei Fragen auf ihn verlassen konnte ist alles andere als selbstverständlich und ich rechne es ihm sehr hoch an.

Zum Gelingen dieser Arbeit haben weitere Personen in großem Maße beigetragen. Für die gute Zusammenarbeit und die vielen hilfreichen Diskussionen auf dem Gebiet der Jet-Messungen danke ich meinem CERN-Kollegen Mikko Voutilainen. Desweiteren danke ich Andrea Rizzi, David Lopes-Pegna und Jim Olsen für die Kooperation im Bereich der Higgs-Physik. Ganz besonderer Dank gilt meinen EKP-Kollegen Manuel Zeise und Daniel Martschei, die sich als verlässliche Ratgeber immer viel Zeit für Fragen und Diskussionen genommen haben. Jochen Ott danke ich herzlich für seine tatkräftige Unterstützung der statistischen Auswertertungen mit seinem theta-Framework, das er eigens für die Higgs-Suche erweitert hat, und Fred Stober für seine kompetenten Hilfen mit grid-control. Meinen Higgs-Mitstreitern Benedikt Maier, Christian Böser, Thorsten Chwalek und Simon Fink danke ich sehr für die gute Zusammenarbeit und die vielfältige Unterstützung bei der Analyse. Jeannine Wagner-Kuhr bin ich dankbar für ihre Hilfestellungen und die Koordination meiner Arbeit mit den verschiedenen Arbeitsgruppen.

Ich danke meinen Kollegen Philipp Schieferdecker, Manuel Zeise, Thorsten Chwalek, Thomas Peiffer, Jochen Ott, Benedikt Maier und Christian Böser für die kritische Lektüre meiner Arbeit und für die hilfreichen Verbesserungsvorschläge.

Für die prima Arbeitsatmosphäre am CERN und in Karlsruhe möchte ich meiner Arbeitsgruppe und meinen Institutskollegen ganz herzlich danken, und hier im Besonderen meinen Zimmerkameraden Jens Hansen, Benedikt Maier, Steffen Röcker, Philipp Schieferdecker, Danilo Piparo und Manuel Zeise.

Ohne den Einsatz von Cheng-Ju Stephen Lin wäre diese Arbeit vermutlich nie begonnen worden. Er hat mich im Sommer 2007 ganz unkompliziert als Sommerstudent eingeladen und mit seinem Enthusiasmus mein Interesse an der Teilchenphysik geweckt. Desweiteren freue ich mich über das anhaltende Interesse meines früheren Lehrers Gerd Model, der sich auch heute noch bei jedem Besuch im Norden über den Fortschritt meiner physikalischen Studien unterrichten lässt.

Überaus glücklich schätze ich mich für die große Unterstützung während der letzten Jahre durch meine Freunde und Familie. Ganz herzlich möchte ich mich daher bei den 'Physikern & Co.' für die vielen schönen, gemeinsamen Zeiten seit dem ersten Semester bedanken. Desweiteren danke ich ganz besonders meinen Großeltern, meinen Geschwistern und meinen Eltern für ihren Rückhalt und ihre Unterstützung. Für ihr Vertrauen, ihre Unterstützung, ihre Freundschaft danke ich Esther von ganzem Herzen.