IEKP-KA/2014-10

Karlsruher Institut für Technologie

Search for the Higgs Boson Produced via Vector-Boson Fusion in the Decay Channel $H \to \tau\tau$

Joram Berger

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
KIT – Karlsruher Institut für Technologie

genehmigte

DISSERTATION

von

Dipl.-Phys. Joram Berger
aus Tübingen

Tag der mündlichen Prüfung: 27. Juni 2014

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Wim de Boer

"Great are the works of the LORD,
studied by all who delight in them."

Psalm 111:2

Contents

Chapter 1

Introduction

The search for the Higgs boson was the field of particle physics that received the most attention in recent years. In 2012 the Higgs boson was discovered at the Large Hadron Collider (LHC) at CERN. Following the discovery, the Nobel Prize in Physics was awarded to François Englert and Peter Higgs in 2013, who laid the theoretical foundation for the understanding of the origin of the mass of elementary particles with the mechanism of electroweak symmetry breaking.

Part of the Higgs physics programme of the CMS Collaboration is the search for Higgs decays to τ-leptons. After the discovery, which was based on the bosonic decay channels, the proof for the predicted coupling of the Higgs boson to down-type fermions was yet to be established. In this thesis, the vector boson fusion (VBF) production mode of the Higgs boson is studied. While the sensitivity of the bosonic decay channels relies on the gluon fusion production mode, which has the largest Higgs boson production cross section according to the Standard Model, the VBF channel is the most important production mode in the $H \to \tau\tau$ search. This is because the Higgs boson branching ratio to τ-leptons is larger than for the bosonic channels, in case that the Higgs boson is light. This provides a sufficient number of events to study the VBF production mode. Moreover, the clear signature of the VBF Higgs topology can be exploited for the analysis in a dedicated event selection.

Two jets in the forward region are the main signature of VBF events. In addition, the region between those two jets does typically not contain hadronic activity, except for hadronic Higgs decays. Thus, a detailed understanding of the hadronic activity in an event in signal and background events as well as a precise measurement of the detector response to jets is of utmost importance for the study of this topology and the VBF event selection.

The decays of τ-leptons involve neutrinos which are not detectable in the experiment. Instead they manifest themselves as missing energy which is calculated from the four-momenta of all reconstructed objects in the event including jets. Therefore, small uncertainties on the jet energy scale indirectly improve the resolution of the missing transverse energy. This helps to improve the reconstruction of the ditau system from

the kinematics of the visible decay products and the missing transverse energy.

Chapter 2 introduces the theoretical foundations of this work. An overview of the Standard Model of particle physics is given, including the Higgs mechanism. The LHC and the CMS detector are described in Chapter 3. This description comprises the event and physics object reconstruction, the analysis workflow and the software framework developed and used for this thesis.

Chapter 4 is dedicated to detailed systematic studies of the jet energy measurement and the calibration of the jet energy scale. These measurements are part of the CMS-wide jet energy corrections and are based on the Z-jet balancing in $Z(\rightarrow \mu\mu) + \text{jet}$ events. This is a promising topology for a data-driven calibration due to the high accuracy of the Z boson reconstruction in CMS and the large production cross section of the Z bosons at the LHC.

While the understanding of the VBF production mode and its measurement is relevant for all $H \rightarrow \tau\tau$ channels as described in the first part of Chapter 5, the second part of this chapter is dedicated to the analysis of the subsequent decay to muons in the $H \rightarrow \tau\tau \rightarrow \mu\mu$ subchannel. Finally, the results of the combined $H \rightarrow \tau\tau$ search are presented with a particular focus on the contribution of the VBF production mode.

Chapter 2

The Standard Model of Particle Physics

The increasing knowledge about subatomic structures and new particles was accompanied by the evolution of quantum mechanics and special relativity. The discovery of new particles and phenomena required new and adapted theories. Similarly, theoretical predictions could be confirmed or excluded by the experiment. This interplay between theoretical predictions and experimental discoveries lead to a theoretical description of particle physics that is now called the Standard Model of particle physics.

2.1 Introduction to the Standard Model

The Standard Model of particle physics describes the fundamental constituents of matter and their interactions. The aim is to describe all phenomena observed in nature in a consistent way from basic principles. This description combines quantum field theory with the theory of special relativity, as introduced in Section 2.2, and starts from basic symmetry principles.

2.1.1 Fundamental interactions

All observed interactions between particles in nature can be reduced to four different forces: gravitation, the electromagnetic interaction, the weak interaction and the strong interaction. While we experience the former two in our every-day life, the latter two have a very short interaction length and are therefore only observed in nuclear and particle physics.

No theory, describing all four forces consistently, has been found yet. Gravitation is not included in the Standard Model. It is extremely weak on small scales and thus negligible for collider physics experiments.

The three remaining interactions are mediated by gauge bosons. Gauge bosons are the quanta of gauge fields described by quantum field theory. The electromagnetic interaction is mediated by a single vector boson, the photon. The mediators of the

Force	range	couples to	vector boson				
	r/m		Name	m/GeV	Q/e	T_3	J^P
Electromagnetism	∞	el. charge	γ	0	0	0	1^-
Weak interaction	10^{-13}	weak charge	$W^\pm$	80.40	±1	±1	1
			Z	91.19	0	0	1
Strong interaction	10^{-15}	colour	g	0	0	±1	1^-

Table 2.1: The interactions of the Standard Model are shown along with the associated gauge bosons. For each interaction particle, the mass m in GeV, the electric charge Q/e, the third component T_3 of the weak isospin and the spin J together with the parity P are denoted in a condensed form [1].

weak interaction are the W and Z bosons. Gluons mediate the strong interaction. In contrast to the photon, which has no electric charge, the $W^\pm$ and Z bosons carry the charge of the weak interaction and gluons carry the strong charge called *colour*. Therefore, these vector bosons participate themselves in the interaction they mediate. Table 2.1 gives a short overview of the interactions of the Standard Model as well as the corresponding particles and their properties.

2.1.2 Fundamental particles

The basic constituents of matter are fermions carrying half-integer spin. There are two classes of fermions, each containing six particles: quarks, which interact strongly, and leptons which do not experience the strong force. An overview of quarks and leptons is given in Table 2.2. Every particle has an associated antiparticle with same properties but the opposite sign of quantum numbers.

Fermions	generation			electr.	colour	weak isospin		spin
	1	2	3	charge		T	T_3	
Leptons	ν_e	ν_μ	ν_τ	0	—	$\frac{1}{2}$	$+\frac{1}{2}$	$\frac{1}{2}$
	e	μ	τ	-1			$-\frac{1}{2}$	
Quarks	u	c	t	$+2/3$	r, g, b	$\frac{1}{2}$	$+\frac{1}{2}$	$\frac{1}{2}$
	d	s	b	$-1/3$			$-\frac{1}{2}$	

Table 2.2: Fundamental particles of the Standard Model that represent the constituents of matter.

Quarks The nucleons, forming the atomic nucleus, are made of *quarks*. The set of six different quarks can be further subdivided into three generations. Each generation contains two quarks that carry a fraction of $2/3$ and $-1/3$ of the elementary charge, respectively. The six quarks can be distinguished by their flavour, which is a set of quantum numbers that are conserved in all interactions except the weak interaction, where flavour-changing currents exist.

The eigenstates of the weak interaction $|d'\rangle$, $|s'\rangle$ and $|b'\rangle$ differ from the mass eigenstates of those down-type quarks. The CKM-matrix describes the transformation between the two representations. It is named after Cabbibo, Kobayashi and Maskawa [2, 3].

Quarks have never been observed as free particles. As a consequence of the Pauli principle, the existence of the Δ^{++} resonance, composed of three valence u-quarks, demands an additional charge called *colour*. It can adopt three values named *red*, *green* and *blue*. Each colour has its anti-colour. A combination of three particles containing all three colours is colour-neutral (white) as well as a colour and anti-colour combination. The strong force rises with the distance between coloured particles. Therefore, free particles can only be colour-neutral compositions. This phenomenon is called *confinement* and it explains the observation of the bound states of quarks which are called *hadrons*. These comprise *mesons*, which are quark anti-quark combinations $q\bar{q}$, and *baryons* composed of three quarks qqq. The proton is the only stable one with a lifetime of $\tau_p > 1.6 \times 10^{32}$ a [4].

Leptons The six fermions that do not experience strong interaction are called *leptons*. The three electrically charged leptons are the electron e, the muon μ and the tau-lepton τ. As it is the case for quarks, the mass of these leptons increases from generation to generation. A neutrino is associated with each of the charged leptons, called ν_e, ν_μ and ν_τ. They are neutral as their name suggests.

Each generation has a specific quantum number, L_ℓ, which is conserved in all interactions and so is the lepton number $L = L_e + L_\mu + L_\tau$. The mass eigenstates of the neutrinos differ from their interaction eigenstates leading to experimentally observed oscillations between different neutrino flavours. These oscillations do not conserve the individual lepton numbers of each generation, L_e, L_μ, L_τ, but only the lepton number L.

In this thesis, the term *leptons* (*electrons, muons, tau-leptons*) comprises particles ℓ^- ($\mathrm{e}^-, \mu^-, \tau^-$) as well as anti-particles ℓ^+ ($\mathrm{e}^+, \mu^+, \tau^+$) of the charged leptons. Hadronic decay products from τ-leptons, reconstructed from the detector information as a single object are called hadronic taus τ_h.

2.1.3 Parameters and theories beyond the Standard Model

The equations of the Standard Model contain several parameters which can not be derived by theoretical considerations and are to be determined experimentally. These 18 parameters can be chosen as: 9 masses of fermions (leaving neutrinos massless), 4 coupling constants, 4 CKM mixing angles, 1 vacuum angle, the Higgs mass and the vacuum expectation value.

To this day, no discrepancy between the predictions of the Standard Model and collider physics experiments have been discovered. But cosmological problems as well as theoretical considerations, like the hierarchy problem, led to extensions of the Standard Model like Supersymmetry and further theories beyond the Standard Model.

2.1.4 Unit conventions

The physical units in this work are based on the International System of Units (SI) [5], including the associated non-SI units electron volt (eV) and and barn (b) in conjunction with the usual SI prefixes. Physical constants and properties of the fundamental particles are taken from CODATA [6] and the Particle Data Group [1]. In the frame of special relativity and quantum mechanics, the particle physics convention is adopted to use units in which the speed of light c and the reduced Planck constant $\hbar$ are set to unity,

$$c = \hbar = 1\,.$$

2.2 The Standard Model as a quantum field theory

Fundamental particles with the same quantum numbers are indistinguishable. As such they are considered as an excitation of a field which is a function of a point in space-time x^μ.

Following the calculus of variation of classical mechanics, the Lagrangian L is a function of generalized coordinates and their first time derivatives. The equations of motion are deduced by minimizing the action $S = \int L\,\mathrm{d}t$. The latter can be shown to be equivalent to solving the Euler-Lagrange equations. In analogy to classical mechanics, the Lagrangian density

$$\mathcal{L}(\boldsymbol{\phi}, \partial_\mu \boldsymbol{\phi}) \qquad \text{with} \quad \partial_\mu \boldsymbol{\phi} = \frac{\partial \boldsymbol{\phi}}{\partial x^\mu} \tag{2.1}$$

is defined in quantum field theory which is a function of the fields $\boldsymbol{\phi}$ and their derivatives $\partial_\mu \boldsymbol{\phi}$ which take the place of the generalized coordinates. The Lagrangian density is a Lorentz-scalar quantity. The action to be minimized is given by

$$S = \int \mathrm{d}^4x\; \mathcal{L}(\boldsymbol{\phi}, \partial_\mu \boldsymbol{\phi})\,. \tag{2.2}$$

Following again the calculus of variations, the equations of motion are derived from the equivalent of the classical Euler-Lagrange Equation,

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \,. \tag{2.3}$$

The appropriate Lagrangian density is an axiomatic statement of quantum field theory. It is formed in a way that the desired field equations are reproduced. The free fields are described by the following homogeneous differential equations.

- The Lagrangian density of a scalar field $\phi(x)$ (spin $s = 0$) is given by

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi^*)(\partial^\mu \phi) - \frac{1}{2} m^2 \phi^* \phi \tag{2.4}$$

  leading to the Klein-Gordon equation

$$\left(\partial_\mu \partial^\mu - m^2\right)\phi = 0 \,. \tag{2.5}$$

- The Lagrangian density of fermion fields $\psi(x)$ (spin $s = \frac{1}{2}$) is

$$\mathcal{L} = \mathrm{i}\bar{\psi}\partial\!\!\!/\psi - m\bar{\psi}\psi \,. \tag{2.6}$$

  The following conventions are used: $\partial\!\!\!/ = \gamma^\mu \partial_\mu$ where γ^μ are the Dirac matrices and $\bar{\psi} = \psi^\dagger \gamma^0$, with $\psi^\dagger$ being the adjoint of the spinor ψ. This leads to the Dirac equation – the equation of motion for spinor fields:

$$(\mathrm{i}\partial\!\!\!/ - m)\psi = 0 \,. \tag{2.7}$$

- The Lagrangian density of a four-vector field $A^\mu(x)$ (spin $s = 1$) is called Proca Lagrangian

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} m^2 A^\nu A_\nu \,. \tag{2.8}$$

  where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the field strength tensor. This results in the equation of motion for vector bosons with the mass m leading to the source-free Maxwell equations

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \,. \tag{2.9}$$

Instead of directly using these Lagrangians for every process of interest, the Feynman rules are derived which offer a pattern for the determination of physical quantities like cross sections and decay rates.

2.3 Local gauge invariance

In classical physics, electric and magnetic fields, $\vec{E}$ and $\vec{B}$, can be written as derivatives of the scalar potential ϕ and the vector potential $\vec{A}$

$$\vec{B} = \vec{\nabla} \times \vec{A}\,, \qquad \vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t}\,. \tag{2.10}$$

As only these derivatives of the potentials are physically measurable quantities, a transformation of $A_\mu = (\phi, \vec{A})$ of the form

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \chi(x)\,, \tag{2.11}$$

involving the scalar field $\chi(x)$, does not change the electromagnetic fields.

Similarly, a gauge invariance can be imposed in quantum field theories. The phase of a complex field does not change the expectation values. For a constant phase, this is called global gauge invariance. If the field $\psi(x)$ transforms as

$$\psi(x) \to \psi'(x) = \mathrm{e}^{iq\Lambda(x)} \cdot \psi(x)\,,$$

with the local phase $q\Lambda(x)$, the derivative of the transformed field

$$\partial_\mu \psi(x) \to \partial_\mu \psi'(x) = \mathrm{e}^{iq\Lambda(x)} \partial_\mu \psi(x) + iq\psi(x) \cdot \partial_\mu \Lambda(x) \tag{2.12}$$

contains an extra term that spoils the gauge invariance. In order to make it invariant under such local phase transformations, the minimal substitution of ∂_μ by the covariant derivative D_μ is given by

$$\partial_\mu \to D_\mu = \partial_\mu - iqA_\mu\,.$$

where $A_\mu(x)$ introduces a new degree of freedom and is called gauge field. By using this derivative, we obtain

$$D_\mu \psi(x) = \partial_\mu \psi(x) - iqA_\mu \psi(x) \tag{2.13}$$

which transforms as

$$D'_\mu \psi'(x) = \mathrm{e}^{iq\Lambda(x)} \cdot \partial_\mu \psi(x) + iq\mathrm{e}^{iq\Lambda(x)} \cdot \psi(x) \cdot \partial_\mu \Lambda(x) - iqA'_\mu \psi(x) \cdot \mathrm{e}^{iq\Lambda(x)}\,. \tag{2.14}$$

By comparing these two equations, it can be deduced that A_μ must transform as

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \Lambda\,.$$

to ensure local gauge invariance. The Lagrangian then reads

$$\mathcal{L}_\psi = \mathrm{i}\bar{\psi}\partial\!\!\!/_\mu \psi - m\bar{\psi}\psi + iqA_\mu \bar{\psi}\gamma^\mu \psi\,.$$

where the first two terms describe the free fermion and the third term the coupling of the fermion to the new gauge field. A further term $-\frac{1}{4}F_{\mu\nu}F_{\mu\nu}$ can be introduced as the kinetic term of the new field A_μ. However, the corresponding mass term $\frac{1}{2}m^2 A_\mu A_\mu$ would again spoil the gauge invariance. Thus, the new field has to remain massless. Hence, the requirement of local gauge invariance leads to the introduction of a new massless field A_μ with the corresponding transformation rule.

2.4 The weak interaction

The weak interaction mediates transitions between fermions. Following the experimental observation that the W boson couples only to left-handed fermions and right-handed anti-fermions, the fundamental constituents of matter can be arranged in doublets of left-handed and singlets of right-handed fermions. In analogy to other spin systems, the weak isospin T is introduced as the generator of the $SU(2)$ group that describes the weak interaction. The weak hypercharge Y is defined by

$$Y = 2\left(\frac{Q}{e} - T_3\right) \tag{2.15}$$

where Q is the charge of the particle, e the proton charge and T_3 the third component of the weak isospin. Y is the generator of the unitary group $U(1)_Y$ and it is conserved in weak interactions.

The left-handed doublets ($T = \frac{1}{2}$) of leptons and quarks are:

$$L_i = \begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_{\mathrm{L}} \quad \begin{matrix}T_3 = +\frac{1}{2}\\ T_3 = -\frac{1}{2}\end{matrix} \quad Y = -1 \tag{2.16}$$

$$Q_i = \begin{pmatrix}u\\ d\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}c\\ s\end{pmatrix}_{\mathrm{L}}, \quad \begin{pmatrix}t\\ b\end{pmatrix}_{\mathrm{L}} \quad \begin{matrix}T_3 = +\frac{1}{2}\\ T_3 = -\frac{1}{2}\end{matrix} \quad Y = +\tfrac{1}{3} \tag{2.17}$$

Right-handed fermions are represented by singlets ($T = 0$).

$$\ell_{\mathrm{R},i} = e^-_{\mathrm{R}}, \quad \mu^-_{\mathrm{R}}, \quad \tau^-_{\mathrm{R}} \quad T_3 = 0 \quad Y = -1 \tag{2.18}$$

$$u_{\mathrm{R},i} = u_{\mathrm{R}}, \quad c_{\mathrm{R}}, \quad t_{\mathrm{R}} \quad T_3 = 0 \quad Y = -\tfrac{4}{3} \tag{2.19}$$

$$d_{\mathrm{R},i} = d_{\mathrm{R}}, \quad s_{\mathrm{R}}, \quad b_{\mathrm{R}} \quad T_3 = 0 \quad Y = -\tfrac{2}{3} \tag{2.20}$$

To ensure the conservation of the weak isospin in a flavour-changing process, the W bosons must be part of an isospin triplet ($T = 1$). The coupling constant of this triplet in weak interactions is denoted as g_2. Additionally, a weak isospin singlet B is introduced ($T = 0$) with a coupling constant g_1.

Local gauge transformations form the corresponding groups $SU(2)_L$ and $U(1)_Y$. The generators $\vec{T}$ and Y of these groups play the role of the electromagnetic charge

in the aforementioned example of gauge transformations. In the same way as for the Lagrangian of quantum electrodynamics, imposing local gauge invariance leads to the emergence of gauge bosons.

The experimentally observable gauge bosons are linear combinations of these four fields $\vec{W}$ and B:

$$A_\mu = \quad B_\mu \cos\theta_\mathrm{W} + W^3_\mu \sin\theta_\mathrm{W} \tag{2.21}$$

$$Z_\mu = -B_\mu \sin\theta_\mathrm{W} + W^3_\mu \cos\theta_\mathrm{W} \tag{2.22}$$

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp W^2_\mu\right) \tag{2.23}$$

The weak mixing angle θ_W is obtained on the condition that the photon field does not couple to neutrinos. It is also called *Weinberg angle* and it links the electric charge and the weak coupling constants. As will be shown in Section 2.6, it also relates the masses of the weak gauge bosons.

$$e = g_1 \cdot \cos\theta_\mathrm{W} = g_2 \cdot \sin\theta_\mathrm{W} \tag{2.24}$$

$$m_W = m_Z \cdot \cos\theta_\mathrm{W} \tag{2.25}$$

Its value [1] at the scale of the Z boson mass is

$$\sin^2\theta_\mathrm{W} = 0.231\,26(5)\,. \tag{2.26}$$

The weakness of the weak interaction is not a consequence of the coupling term but results from the masses of the gauge bosons in the denominator of the propagator term.

2.5 The Lagrangian of the Standard Model

The strong interaction is described by the non-abelian group $SU(3)_C$ where C stands for the colour charge. The coupling constant is denoted g_3. Eight gauge bosons, called gluons, arise from the requirement of local gauge invariance. This completes the group structure of the Standard Model, $SU(3)_C \times SU(2)_L \times U(1)_Y$.

For each group, field strength tensors can be defined for the boson fields as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_3 f^{abc} G^b_\mu G^c_\nu \qquad a,b,c = 1,\ldots,8 \tag{2.27}$$

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2 \varepsilon^{abc} W^b_\mu W^c_\nu \qquad a,b,c = 1,2,3 \tag{2.28}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.29}$$

where ε^{abc} is the total antisymmetric tensor, which follows from the commutation relations of the generators of the non-abelian $SU(2)_L$ group, and f^{abc} the corresponding structure constants of the $SU(3)_C$ group. The $U(1)_Y$ is an abelian group and the

commutator $[Y, Y] = 0$. The three generators of the $SU(2)_L$ are $\vec{T} = \frac{\vec{\sigma}}{2}$, where $\vec{\sigma}$ are the Pauli matrices. For the $SU(3)_C$, the generators $\tilde{T}_a$ are the Gell-Mann matrices multiplied by $\frac{1}{2}$.
The full covariant derivative D_μ is given by

$$D_\mu = \partial_\mu - \mathrm{i}g_3\tilde{T}_a G_\mu^a - \mathrm{i}g_2 T_a W_\mu^a - \mathrm{i}g_1 \frac{Y}{2} B_\mu \tag{2.30}$$

and its contraction with the γ matrices is defined as $\not{D} = D_\mu \gamma^\mu$. Using these definitions, the Standard Model Lagrangian $\mathcal{L}_{\mathrm{SM}}$ is given by

$$\begin{aligned}\mathcal{L}_{\mathrm{SM}} = &-\tfrac{1}{4} G^a_{\mu\nu} G_a^{\mu\nu} - \tfrac{1}{4} W^a_{\mu\nu} W_a^{\mu\nu} - \tfrac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ &+ \bar{L}_i \mathrm{i}\not{D} L_i + \bar{\ell}_{\mathrm{R},i} \mathrm{i}\not{D} \ell_{\mathrm{R},i} + \bar{Q}_i \mathrm{i}\not{D} Q_i + \bar{u}_{\mathrm{R},i} \mathrm{i}\not{D} u_{\mathrm{R},i} + \bar{d}_{\mathrm{R},i} \mathrm{i}\not{D} d_{\mathrm{R},i} + \mathcal{L}_{\mathrm{Higgs}}\end{aligned}$$

where $\mathcal{L}_{\mathrm{Higgs}}$ denotes the Higgs sector of the Standard Model which will be described in the following section.

2.6 Electroweak symmetry breaking

The electroweak unification can not explain the occurrence of massive gauge bosons. Adding explicit mass terms to the Lagrangian spoils the local gauge invariance and makes the theory non-renormalizable. This contradicts the experimental evidence for non-vanishing masses of fermions and of the gauge bosons of the weak interaction. This problem can be solved by the Brout-Englert-Higgs mechanism [7–12] which is built upon local gauge invariance and electroweak symmetry breaking. This description follows [13].

The principle of symmetry breaking can be explained using a simple example: the Lagrangian of a real scalar field ϕ with the potential V

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - V(\phi) \qquad \text{with} \quad V(\phi) = \mu^2 \phi^2 + \lambda \phi^4 \tag{2.31}$$

is symmetric under a reflection ($x \to -x$), due to the absence of cubic terms. The parameter λ is required to be positive to ensure the existence of a global minimum. For positive values of μ^2, the minimum of the potential is at $\phi = 0$ as illustrated in Figure 2.1a. However, in case $\mu^2 < 0$ the minima of the potential are located at

$$\phi_{\min} = v = \pm\sqrt{-\frac{\mu^2}{2\lambda}}$$

and not at zero (cf. Fig. 2.1b). The vacuum expectation value $\langle 0|\Phi|0\rangle$ is one of those minima v. For a physical interpretation, this Lagrangian has to be expanded around the vacuum expectation value. The Lagrangian of the new field $\sigma = \phi - v$ reads

$$\mathcal{L} = \frac{1}{2}\partial_\mu \sigma \partial^\mu \sigma + 2\mu^2 \sigma^2 - 2\sqrt{-2\lambda\mu^2}\sigma^3 - \lambda\sigma^4 + \text{const.} \tag{2.32}$$

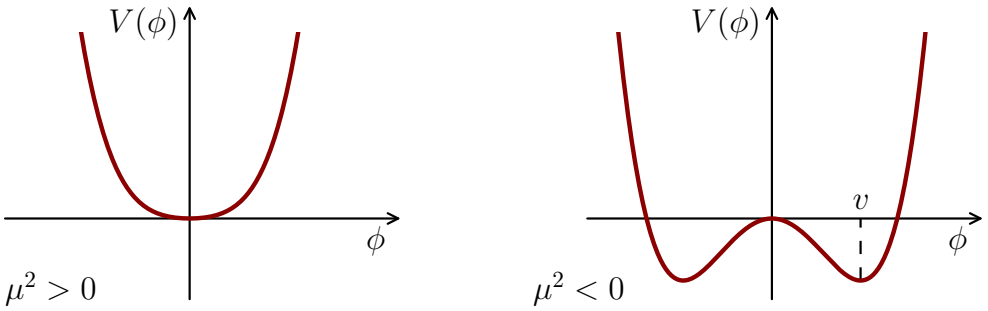


Figure 2.1: The quartic symmetric potential V of a scalar field ϕ for $\mu^2 > 0$ (*left*) and for $\mu^2 < 0$ (*right*).

This can now be interpreted as the Lagrangian of a scalar field of mass $m = 2\sqrt{-\mu^2}$ and self-interactions. The reflection symmetry of the original potential is broken in the ground state of the system as can be seen in Figure 2.1 or from the presence of the cubic term in Eq. (2.32).

2.6.1 The Higgs mechanism in the Standard Model

In case of the non-abelian $SU(2)$ group of the Standard Model, three degrees of freedom are required to generate the masses of the weak gauge bosons. The simplest choice for the Higgs field is a $SU(2)$ doublet of complex scalar fields

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{2.33}$$

The Lagrangian of this scalar doublet

$$\mathcal{L}_{\mathrm{H}} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 , \tag{2.34}$$

containing a kinetic term and the symmetric potential, is added to the Lagrangian of the Standard Model. This potential again has a vacuum expectation value different from zero which can be chosen to be

$$\langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$$

with the same value of v as before.

The expansion of the first term in the Lagrangian $\mathcal{L}_{\mathrm{H}}$ around this minimum is

$$|D_\mu \Phi|^2 = \left| \left(\partial_\mu - \mathrm{i} g_2 T_a W_\mu^a - \mathrm{i} g_1 \frac{1}{2} B_\mu \right) \Phi \right|^2 \tag{2.35}$$

$$= \frac{1}{2} \left| \begin{pmatrix} \partial_\mu - \frac{\mathrm{i}}{2}(g_2 W_\mu^3 + g_1 B_\mu) & -\frac{\mathrm{i}}{2} g_2 (W_\mu^1 - \mathrm{i} W_\mu^2) \\ -\frac{\mathrm{i}}{2} g_2 (W_\mu^1 + \mathrm{i} W_\mu^2) & \partial_\mu + \frac{\mathrm{i}}{2}(g_2 W_\mu^3 - g_1 B_\mu) \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \right|^2 \tag{2.36}$$

$$= \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8} g_2^2 (v+H)^2 |W_\mu^1 + \mathrm{i} W_\mu^2|^2 + \frac{1}{8}(v+H)^2 |g_2 W_\mu^3 - g_1 B_\mu|^2 \tag{2.37}$$

The masses of the gauge bosons, as defined in Eq. (2.23), can be read from bilinear terms in Eq. (2.37). They are

$$m_W = \frac{1}{2} v g_2 \,, \qquad m_Z = \frac{1}{2} v \sqrt{g_1^2 + g_2^2} \,, \qquad m_A = 0 \,.$$

In conclusion, the introduction of a scalar field and a symmetric potential leads to a new ground state around which the theory has to be expanded. This expansion yields mass terms for the weak gauge bosons while leaving the photon – and the gluons which are not touched – massless.

2.6.2 Fermion masses

Explicit mass terms for fermions would also break the local gauge invariance of the Standard Model. However, fermion masses can be generated by couplings between the Higgs doublet and the fermions, using the same Higgs doublet as before. These Yukawa couplings for a lepton of the i-th generation has the form

$$\mathcal{L}_{\mathrm{Y}}^{\ell_i} = \lambda \bar{\psi}_{\mathrm{L}} \Phi \psi_{\mathrm{R}} + h.\,c. \tag{2.38}$$

The expansion around the ground state yields

$$= -\frac{1}{\sqrt{2}} \lambda_{\ell,i} \bar{L}_i \begin{pmatrix} 0 \\ v + H \end{pmatrix} \ell_{\mathrm{R},i} + h.\,c. \tag{2.39}$$

$$= -\frac{\lambda_{\ell,i}}{\sqrt{2}} (v+H) \bar{\ell}_{\mathrm{L},i} \ell_{\mathrm{R},i} \tag{2.40}$$

from which the mass term can be read to be $m_{\ell_i} = v \lambda_{\ell,i} / \sqrt{2}$. For up-type fermions, the components of the isodoublet are exchanged by using $\tilde{\Phi} = i \sigma_2 \Phi^*$ instead of Φ. In case of an up-type quark in generation i, the Yukawa term is then

$$\mathcal{L}_Y^{u_i} = -\frac{\lambda_{u_i}}{\sqrt{2}} (v+H) \bar{u}_{\mathrm{L},i} u_{\mathrm{R},i} \tag{2.41}$$

The mass of all fermions is thus given by

$$m_f = \lambda_f \cdot \frac{v}{\sqrt{2}} \,. \tag{2.42}$$

This means that mass is no intrinsic property of particles, but it is generated through the coupling to the Higgs field.

2.6.3 The Higgs boson

The introduction of the Higgs mechanism has the additional consequence that a new particle appears in the theory. The covariant derivative Eq. (2.37) contains the kinematic term $\frac{1}{2}(\partial_\mu H)^2$ of the boson H. The mass of this boson is a result of the expansion of the Higgs potential around the chosen ground state. Making use of $-\mu^2 = 2\lambda v^2$, the potential

$$V = \mu^2(v+H)^2 + \lambda(v+H)^4 \tag{2.43}$$

$$= 4\lambda v^2 H^2 + 4\lambda v H^3 + \lambda H^4 + \text{const.} \tag{2.44}$$

contains the mass term of a boson with the mass

$$m_H = 2\sqrt{2\lambda}\, v = 2\sqrt{-\mu^2} \,.$$

The higher order terms in H represent the Higgs self-interactions. Similarly, the couplings of the Higgs boson to fermions f and gauge bosons V can be derived from Eq. (2.40) and (2.41)

$$g_{Hff} = i\frac{m_f}{v} \,, \qquad g_{HVV} = -2i\frac{m_V^2}{v} \,, \qquad g_{HHVV} = -2i\frac{m_V^2}{v^2} \tag{2.45}$$

where the coefficients arise from the Feynman rules for these vertices. This Standard Model Higgs boson is a massive spin-0 particle with a positive eigenvalue to the parity operator $J^P = 0^+$.

2.7 The Higgs boson at the LHC

Decades after the Higgs theory was established, the Higgs boson was discovered by the collaborations of ATLAS and CMS in 2012 [14, 15]. The possibilities to produce Higgs bosons at a hadron collider as well as the analysed decay channels will be discussed in this section.

2.7.1 Higgs production at the LHC

At the Large Hadron Collider (LHC), the Standard Model Higgs boson can be produced in mainly four channels: the gluon fusion, the vector boson fusion, the vector boson associated production and the top pair associated production [16]. The corresponding Feynman diagrams are shown in Figure 2.2. The Higgs production cross sections of all four channels for centre-of-mass energies of 7 TeV and 8 TeV are shown in Figure 2.3.

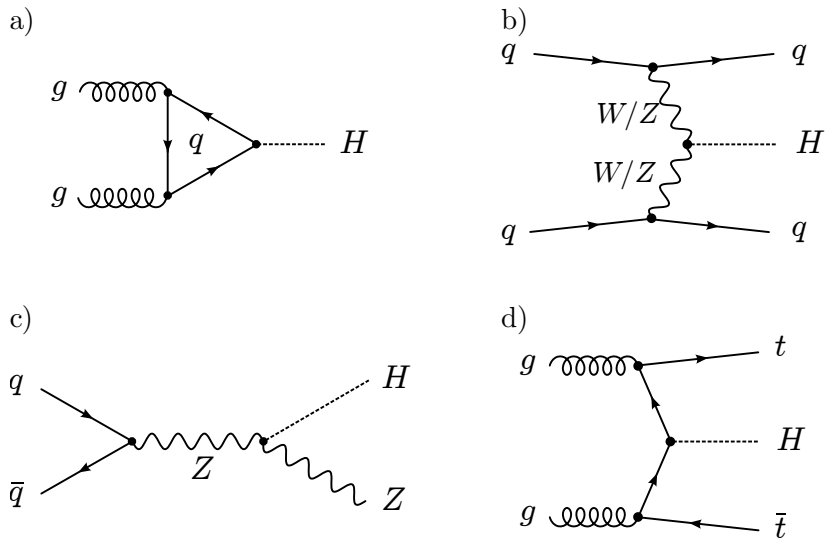


Figure 2.2: The four main Higgs production channels at the LHC: a) gluon fusion, b) vector boson fusion, c) vector boson associated production and d) $t\bar{t}$ associated production.

Gluon fusion Two gluons from the colliding protons can couple to the Higgs boson via a fermion loop. The top quark, being the heaviest quark, has the largest loop contribution. The gluon fusion channel has by far the largest production cross section in LHC pp collisions. The leading order cross section is enhanced by next-to-leading order (NLO) contributions of roughly the same size [18, 19]. The next-to-next-to-leading order (NNLO) calculations lead to an additional increase of $10-20\,\%$ [20–22]. Further small contributions come from three-loop virtual corrections, taking into account the finite quark masses [23]. The gluon fusion cross section [17] for a Higgs boson at the mass of $m_H = 125.3\,\text{GeV}$ at a centre-of-mass energy of 7 TeV and 8 TeV

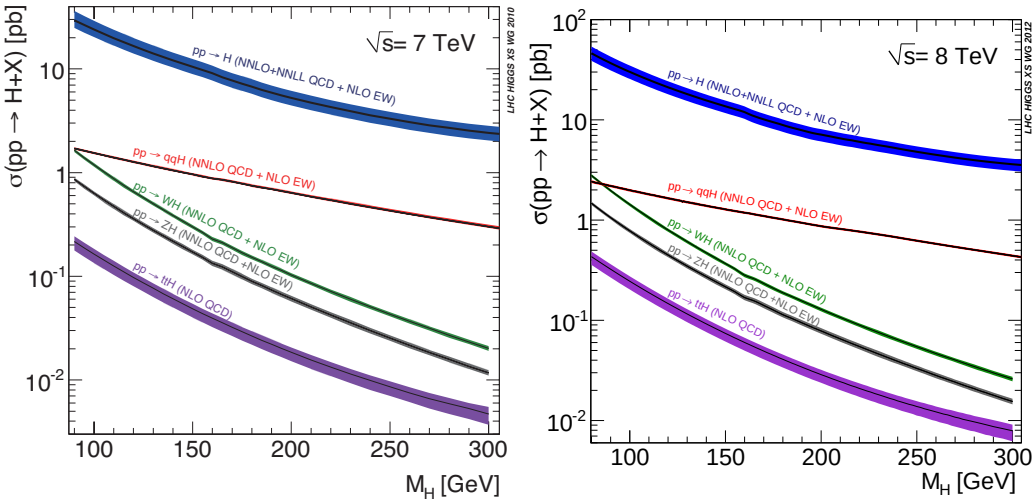


Figure 2.3: Production cross sections of the four main Higgs production channels at the LHC for a for a centre-of-mass energy of 7 TeV and 8 TeV. Gluon fusion ($pp \to H$) has the largest cross section, followed by Vector Boson Fusion ($pp \to qqH$), vector boson associated production ($pp \to WH$ and $pp \to ZH$) and top pair associated production ($pp \to t\bar{t}H$) The production cross sections are shown in the low mass range on which the Higgs search was focussed in the last years [17].

is calculated at NNLO QCD and NLO electroweak. They amount to

$$\sigma^{7\,\mathrm{TeV}}_{gg\to H} = 15.06\,\mathrm{pb}\ ^{+7.1}_{-7.8}\,\%\ (\mathrm{QCD\ scale})\ ^{+7.6}_{-7.1}\,\%\ (\mathrm{PDF} + \alpha_\mathrm{S})\ \text{and} \tag{2.46}$$

$$\sigma^{8\,\mathrm{TeV}}_{gg\to H} = 19.18\,\mathrm{pb}\ ^{+7.2}_{-7.8}\,\%\ (\mathrm{QCD\ scale})\ ^{+7.5}_{-6.9}\,\%\ (\mathrm{PDF} + \alpha_\mathrm{S}). \tag{2.47}$$

Uncertainties arise from the introduction of renormalization and factorisation scales of quantum chromodynamics (QCD) perturbation theory and from the limited precision of the PDFs (cf. Section 2.8) and the strong coupling constant α_S. This production channel has no distinctive experimental signature and can only be exploited in searches with clearly identifiable Higgs boson decay products.

Vector boson fusion The second largest production cross section is exhibited by the vector boson fusion (VBF) Higgs production. Two quarks from the colliding protons radiate a massive vector boson. These two bosons fuse to produce the Higgs final state. The matrix element squared is proportional to [24]:

$$|\mathcal{A}_\mathrm{VBF}|^2 \propto \frac{1}{\left(2p_{q_1}\cdot p_{q_1'} + m_V^2\right)^2} \cdot \frac{1}{\left(2p_{q_2}\cdot p_{q_2'} + m_V^2\right)^2} \cdot \hat{s}\, m_{jj}^2 \tag{2.48}$$

where m_{jj} is the invariant mass of the two jets resulting from the outgoing quarks, m_V is the mass of the exchanged vector boson and $q_{1,2}$ are the incoming and $q'_{1,2}$ are the outgoing quarks. This matrix element is dominated by the contribution in the forward region, where the dot-products are small [24], resulting in a large rapidity difference between the two, called the *rapidity gap*. The two outgoing quarks of the matrix element hadronise and form two jets, detected predominantly in the forward region. These two jets, called *tagging jets*, are a distinguishing feature of VBF events. In a typical VBF event selection, the two jets with the largest transverse momentum are used to identify these events.

The factor m_{jj} in the matrix element enhances the cross section for high values of the invariant dijet mass. Compared to background events in an event selection, this results in wide tails in the m_{jj} distribution. Furthermore, no coloured particle is exchanged between the two quarks at leading order. This leads to a different gluon radiation pattern in VBF events compared to the major backgrounds. The missing hadronic activity between the two jets is a characteristic signature forming the basis of a central jet veto [25].

NLO corrections give only small enhancements of 5 % to 10 % [26]. Electroweak corrections decrease the total cross section by about 5 % [27, 28]. NNLO calculations with QCD scale uncertainties of around 0.2 % to 0.3 % show how well this production channel is understood. The vector boson fusion cross sections [17] for a Higgs boson at the mass of $m_H = 125.3\,\text{GeV}$ are

$$\sigma^{7\,\text{TeV}}_{\text{VBF}} = 1.219\,\text{pb} \pm 0.3\,\%\ (\text{QCD scale})\ ^{+2.5}_{-2.1}\,\%\ (\text{PDF} + \alpha_\text{S})\ \text{and} \tag{2.49}$$

$$\sigma^{8\,\text{TeV}}_{\text{VBF}} = 1.576\,\text{pb} \pm 0.2\,\%\ (\text{QCD scale})\ ^{+2.6}_{-2.8}\,\%\ (\text{PDF} + \alpha_\text{S}). \tag{2.50}$$

Vector boson associated production Higgs bosons can be radiated from a W or Z boson. The cross section of these processes are small but they can be a useful handle to identify the Higgs boson decaying to $\gamma\gamma$ or $b\bar{b}$. These signatures are also exploited in a dedicated analysis in $H \to \tau\tau$ channel. Due to the quark-antiquark initial state, this process has a larger cross section at $p\bar{p}$ colliders. It is therefore the dominant Higgs production process at the Tevatron [16].

The cross sections [17] of W and Z associated productions of a Higgs boson at the mass of $m_H = 125.3\,\text{GeV}$ are calculated at NNLO QCD and NLO electroweak. They are

$$\sigma^{7\,\text{TeV}}_{WH} = 0.5736\,\text{pb} \pm 0.9\,\%\ (\text{QCD scale}) \pm 2.6\,\%\ (\text{PDF} + \alpha_\text{S}), \tag{2.51}$$

$$\sigma^{7\,\text{TeV}}_{ZH} = 0.3325\,\text{pb} \pm 2.9\,\%\ (\text{QCD scale}) \pm 2.7\,\%\ (\text{PDF} + \alpha_\text{S}), \tag{2.52}$$

$$\sigma^{8\,\text{TeV}}_{WH} = 0.6988\,\text{pb} \pm 1.0\,\%\ (\text{QCD scale}) \pm 2.4\,\%\ (\text{PDF} + \alpha_\text{S})\ \text{and} \tag{2.53}$$

$$\sigma^{8\,\text{TeV}}_{ZH} = 0.4123\,\text{pb} \pm 3.1\,\%\ (\text{QCD scale}) \pm 2.5\,\%\ (\text{PDF} + \alpha_\text{S}). \tag{2.54}$$

$t\bar{t}$ associated production The Higgs boson can also be radiated off a top quark. This process offers the possibility to measure the top Yukawa coupling [29]. Even though the latter is large, corresponding to the high mass of the top-quark, the heavy $t\bar{t}H$ final state is kinematically suppressed. The production cross section of the $t\bar{t}$ associated Higgs production poses a major challenge to the dedicated LHC Higgs searches. Furthermore, the analysis of this channels suffers from a large $t\bar{t}jj$ background. The analysis of jet substructures is used to significantly reduce combinatorical backgrounds [16].

The top-pair associated production cross sections [17] for a Higgs boson at the mass of $m_H = 125.3\,\text{GeV}$ are

$$\sigma_{\text{VBF}}^{7\,\text{TeV}} = 0.0857\,\text{pb}\,{}^{+3.2}_{-9.3}\,\%\ (\text{QCD scale})\ \pm 8.4\,\%\ (\text{PDF} + \alpha_S)\ \text{and} \tag{2.55}$$

$$\sigma_{\text{VBF}}^{8\,\text{TeV}} = 0.1283\,\text{pb}\,{}^{+3.8}_{-9.3}\,\%\ (\text{QCD scale})\ \pm 8.1\,\%\ (\text{PDF} + \alpha_S). \tag{2.56}$$

2.7.2 Higgs decay modes

As shown in Section 2.6.2, the Higgs couplings to gauge bosons and fermions rise quadratically and linearly with their mass. Therefore, the branching ratio of the Higgs boson strongly depends on its own mass relative to the masses of the decay products. As the Higgs boson mass is the only free parameter involved, the branching ratios can be calculated for different Higgs boson masses as shown in Figure 2.4 for the main decay channels at the LHC. For a given Higgs boson mass, the heaviest

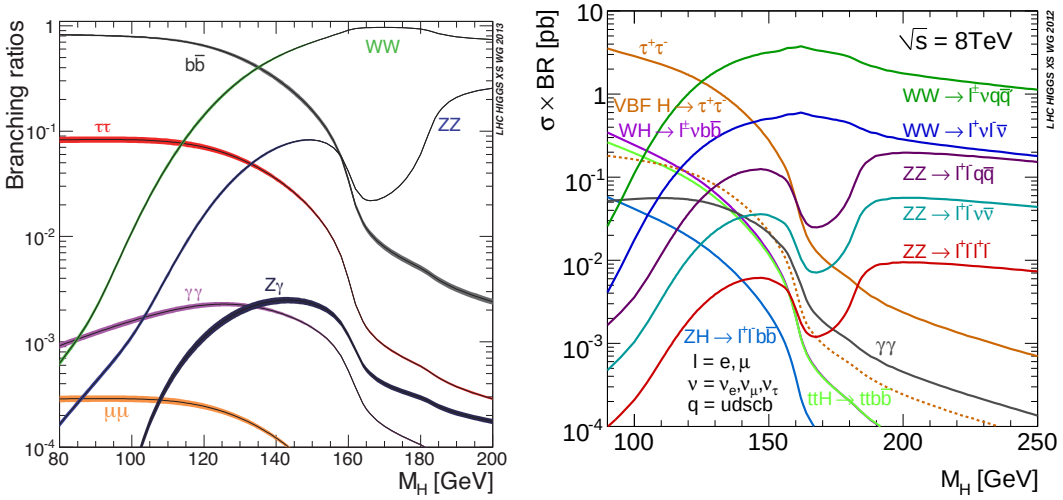


Figure 2.4: Higgs branching ratios (*left*) and Higgs cross section at $\sqrt{s} = 8\,\text{TeV}$ times branching ratio (*right*). The fermionic channels show the highest branching ratios at low Higgs boson masses [17].

particles accessible in phase space have the highest branching ratio. While fermionic decays ($H \to b\bar{b}$ and $H \to \tau\tau$) dominate at low Higgs boson masses, the $H \to WW$ and $H \to ZZ$ channels peak at their respective production threshold.

Decay mode j	**branching ratio** $\mathcal{B}$ in %	**relative error**
$b\bar{b}$	57.2	$^{+3.24\,\%}_{-3.30\,\%}$
WW	21.9	$^{+4.22\,\%}_{-4.16\,\%}$
gg	8.54	$^{+10.20\,\%}_{-9.95\,\%}$
$\tau\tau$	6.27	$^{+5.70\,\%}_{-5.63\,\%}$
$c\bar{c}$	2.89	$^{+12.20\,\%}_{-12.19\,\%}$
ZZ	2.72	$^{+4.22\,\%}_{-4.16\,\%}$
$\gamma\gamma$	0.228	$^{+4.91\,\%}_{-4.89\,\%}$

Table 2.3: All decay modes of the Higgs boson with a branching ratio above 0.2 %. The total decay width is $\Gamma_H = (4.11 \pm 0.16)\,\text{MeV}$ [17].

For the Higgs boson found at $m_H = 125.3\,\text{GeV}$, the branching ratios $\mathcal{B}$ are listed in Table 2.3. $H \to b\bar{b}$ is the dominant channel. However, these events are extremely hard to distinguish from the QCD background at a hadron collider. The Higgs boson mass being located in the transition between the fermionic low mass region and the vector boson dominated higher masses, the $H \to WW$ channel has already a higher branching ratio than $H \to \tau\tau$, but the one of the $H \to ZZ$ channel is still lower. In the subsequent decay into two pairs of charged leptons, the latter channel provides extremely clean signatures, outweighing the smaller branching ratio. For that reason, this channel is also called the golden channel.

The branching ratio of $H \to \gamma\gamma$ is again one order of magnitude smaller. Being a massless particle, the photon does not couple to the Higgs boson directly. However, similar to the gluon fusion production channel, the Higgs can decay to a pair of photons via a loop of charged massive particles, such as the W boson and heavy quarks. The clear signal of two high energetic photons makes $H \to \gamma\gamma$ one of the discovery channels of the Higgs boson. The Feynman diagrams of the five main Higgs decay modes at the LHC are shown in Figure 2.5.

After the discovery of the Higgs boson in a subset of these channels, it is crucial to prove the Standard Model characteristics of this newly found particle by discoverying all measurable Higgs boson decays and by studying the branching ratios of all these decay channels. In addition, the spin and parity properties have to be measured.

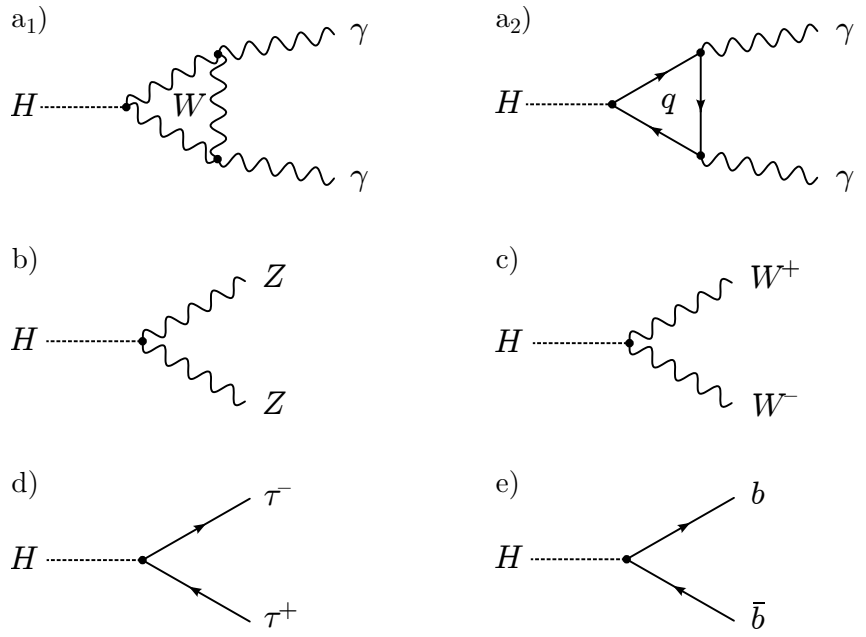


Figure 2.5: The five main Higgs decay modes at the LHC: a) two contributions to the diphoton final state including bosonic and fermionic loops; b, c) bosonic decay modes ZZ and WW; d, e) fermionic decays into pairs of τ-leptons or b-quarks.

2.7.3 Statistical inference

The results in the search for the Higgs boson are combined statistically. A profile likelihood ratio is used to quantify the compatibility of the measured data with a given hypothesis. It is defined by

$$\lambda(\mu) = \frac{L(data \mid \mu, \tilde{\theta}(\mu))}{L(data \mid \hat{\mu}, \hat{\theta})} \tag{2.57}$$

where μ is the signal strength modifier that quantifies the signal contribution as a function of the Standard Model expectation. It is the only parameter of interest in this statistical analysis. $\tilde{\theta}(\mu)$ represents the nuisance parameters that maximise the likelihood L for a given μ. $\hat{\mu}$ and $\hat{\theta}$ are the parameters at which the likelihood reaches its global maximum. By construction, the likelihood ratio is $\lambda(\mu) \in [0, 1]$.

The likelihood function is the product of Poisson probabilities in each bin of the analysis and the pdfs of all nuisance parameters, representing the systematic uncertainties. The latter usually follow a log-normal distribution and are sampled using toy Monte-Carlo. All sources of uncertainties are either treated as uncorrelated or as fully correlated, which allows the likelihood functions to have factorised form.

Exclusion limits are calculated using a modified frequentist construction called the CL_s method [30–32]. The test statistic is defined by

$$q_\mu = -2 \ln \lambda(\mu) \tag{2.58}$$

with $0 \leq \hat{\mu} \leq \mu$. The lower constraint avoids unphysical values as the signal rate is non-negative. The upper constraint guarantees a one-sided confidence interval as needed for upper limits. This way, upward fluctuations are not treated as evidence against the signal hypothesis. High values of q_μ represent an increasing disagreement of the observed data with the given hypothesis. The test statistic is usually calculated using Monte-Carlo methods.

The probability to obtain the observed number of events or more under the signal+background hypothesis is

$$CL_{s+b}(\mu) = \mathcal{P}(q_\mu \geq q_\mu^{\text{obs}} \mid \mu, \hat{\theta}_\mu^{\text{obs}}) = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \mid \mu)\, \mathrm{d}q_\mu \tag{2.59}$$

where q_μ^{obs} is the observed value of the test statistic for a given μ. The probability of observing the same amount under the background-only hypothesis is

$$1 - CL_b(\mu) = \mathcal{P}(q_\mu \geq q_\mu^{\text{obs}} \mid 0, \hat{\theta}_0^{\text{obs}}) = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \mid 0)\, \mathrm{d}q_\mu \,. \tag{2.60}$$

The $CL_s(\mu)$ is the ratio of these two tail probabilities,

$$CL_s(\mu) = \frac{CL_{s+b}}{1 - CL_b} \,. \tag{2.61}$$

The Higgs boson is excluded at a $(1 - \alpha)$ confidence level (CL) if $CL_s(\mu = 1) \leq \alpha$. Similarly, an upper limit on μ can be calculated by varying μ until $CL_s(\mu) = \alpha$ is reached. Usually, a confidence level of 95 % is requested ($\alpha = 0.05$).

In order to compare the observed results with the expectation from the background-only hypothesis, exclusion limits can be calculated with $\mu = 0$ for reference. The median of this distribution is then called the expected limit and the $\pm 1\sigma$ and $\pm 2\sigma$ quantiles of the distribution are used to quantify the expected fluctuation around the median value. The same can be done for the signal+background hypothesis setting the signal strength modifier to $\mu = 1$.

In the search for a new particle, excesses need to be quantified. To this end, the test-statistic is defined as the negative log-likelihood ratio of the background hypothesis

$$q_0 = -2 \ln \lambda(0) \qquad \text{with } \hat{\mu} \geq 0 \,. \tag{2.62}$$

Based on this test-statistic, the p-value is defined as

$$p_0 = \mathcal{P}(q_0 \geq q_0^{\text{obs}}) = \int_{q_0^{\text{obs}}}^{\infty} f(q_0 \mid 0)\, \mathrm{d}q_0 \tag{2.63}$$

being a measure of how likely the observed deviation is an upward fluctuation of the background-only hypothesis. The significance Z is calculated using the convention of a one-sided Gaussian tail

$$p = \frac{1}{2}\Big[1 - \operatorname{erf}\big(Z/\sqrt{2}\big)\Big]$$

where erf is the Gaussian error function. By convention, a p-value above 3σ is considered evidence for the investigated signal process and 5σ is called a discovery.

2.8 Parton distribution functions

Protons are composite particles and therefore only its constituents take part in the hard interactions in high energy proton-proton collisions. The proton is consists of quarks and gluons which are also called by the collective term for these strongly interacting particles, *partons*. While the valence quarks determine the quantum numbers and thus most properties of the proton, this picture is not complete. Gluons and virtual quark-antiquark pairs (*sea-quarks*) can also be found in hadrons. The substructure is described by the parton distribution functions (PDFs) $f_i(x, Q^2)$ as a function of the energy scale Q of the process and the momentum fraction x carried by parton i.

For LHC analyses, the PDF4LHC [33] group recommends the usage of a selection of NNLO PDF sets from different groups for calculations and the estimation of uncertainties: CT10 [34], MSTW [35] and NNPDF [36]. Figure 2.6 contains the proton PDFs of the NNPDF Collaboration. As expected for the proton, the probability density for valence up-quarks u_{val} is twice as large as for the valence down-quark d_{val}. The sea-quarks q_{sea} are the sum of all non-valence quarks or anti-quarks. Those as well as the gluons carry mostly smaller fractions of the proton momentum.

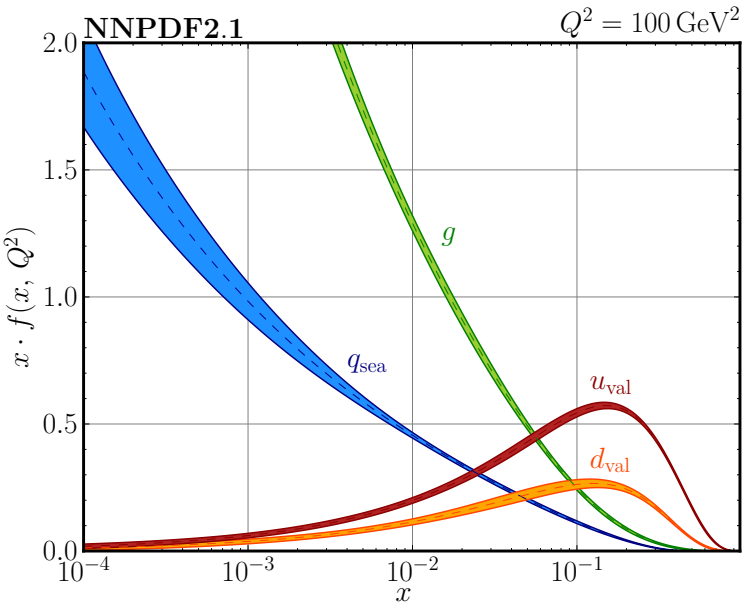


Figure 2.6: Visualization of the proton PDF of the NNPDF Collaboration [36] for different parton flavours at $Q^2 = 100\,\mathrm{GeV}^2$ (Figure produced using LHAPDF [37]).

Chapter 3

The Compact Muon Solenoid at the Large Hadron Collider

Founded in 1954, the European Organisation for Nuclear Research (CERN) has a long tradition of successful particle physics research. In particular the peaceful exploration of subatomic physics in growing international collaborations has been the main task since the beginning. Located near Geneva, on the border between France and Switzerland, CERN attracted a large number of physicist of international repute and hosted a series of large scale particle physics experiments. Over the decades, spectacular discoveries were made and some of them even led to Nobel Prizes as in 1984 when Carlo Rubbia and Simon van der Meer received the Nobel Prize for their contribution to the discovery of the W and Z boson the year before [38]. In 1992, Georges Charpak was awarded the Nobel Prize in physics for the invention of the multiwire proportional chamber in 1968 [39]. Many other forerunners of modern particle physics worked at CERN.

In August 1989, the Large Electron-Positron Collider (LEP) became operational. Electroweak precision measurements were carried out in e^+e^- collisions starting from $\sqrt{s} = 90\,\text{GeV}$ [40]. It was used until 2000 and then dismantled and removed from its underground location to make way for the LHC. Even though the search for the Higgs boson was an important part of the LEP programme, it was not found due to the limited energetic reach of the accelerator and its experiments.

Not only the immediate physical problems, but also the technical challenges physicists face during the planning and construction of particle accelerators led to significant achievements, particularly in the field of vacuum technology, superconductivity and computer science. The World Wide Web [41], which this year celebrates its 25^{th} anniversary, is probably the most famous and most widely used offspring of the research environment at CERN. Today, CERN hosts the largest collider experiment ever built, both in terms of the size and the centre-of-mass energy: the Large Hadron Collider (LHC).

3.1 The Large Hadron Collider

The LHC reuses the 27 km tunnel of the LEP collider [42]. It is a ring accelerator for protons and heavy ions, providing an unprecedented particle energy and luminosity. The search for the Higgs boson was driving the design of both the accelerator and the detectors. Besides that, the research programme comprises the full spectrum of particle physics: from QCD, electroweak and top physics measurements over flavour physics and quarkonia to searches for supersymmetry, exotica and further particle physics phenomena beyond the Standard Model.

At four interaction points, the accelerated particle beams can be brought to collision. Large underground caverns around these interaction points host the main LHC experiments: The biggest two experiments, ATLAS [43] and CMS [44], are general-purpose detectors that cover all of the aforementioned fields of physics. The two independent experiments allow for the cross-confirmation of any new discovery. The LHCb Experiment [45] focusses on measurements of CP violation and rare decays in flavour physics. Most of the time is dedicated to pp collisions, but the LHC also provided heavy ion collisions for several weeks in each year of running. The ALICE Experiment [46] is especially built to analyse these in order to produce and to explore a quark-gluon plasma. Three further experiments are located at the LHC ring sharing the interaction points with the larger four. The TOTEM[1] [47] and the LHCf Experiment [48] use detectors positioned on either side of the CMS and the ATLAS detector, respectively. TOTEM is dedicated to forward physics, a precise measurement of the total pp cross section and to the study of the proton structure, while LHCf uses particles from forward scattering to simulate cosmic rays. Another smaller experiment named MoEDAL[2], located in the Vertex Locator cavern of the LHCb experiment, is used to search for magnetic monopoles. Further experiments are connected with CERN either by their location on the CERN sites near Geneva or by using particle beams from CERN accelerators.

A chain of linear accelerators and synchrotrons serves to pre-accelerate the proton bunches before filling them into the main LHC ring as sketched in Figure 3.1. After passing the linear accelerator LINAC2, the proton energy is 50 MeV. The energy is increased to 1.4 GeV by the Booster and subsequently by the Proton Synchrotron (PS) to 25 GeV. The last pre-acceleration step is carried out by the Super Proton Synchrotron (SPS) from which the particles are injected into the LHC at an energy of 450 GeV [49].

The main LHC ring has eight radio-frequency cavities along the beam pipe to accelerate particles. As the particles are only accelerated at certain positions and periodic time intervals according to the phase of the electric field in those cavities,

[1] TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC
[2] Monopole & Exotics Detector at the LHC

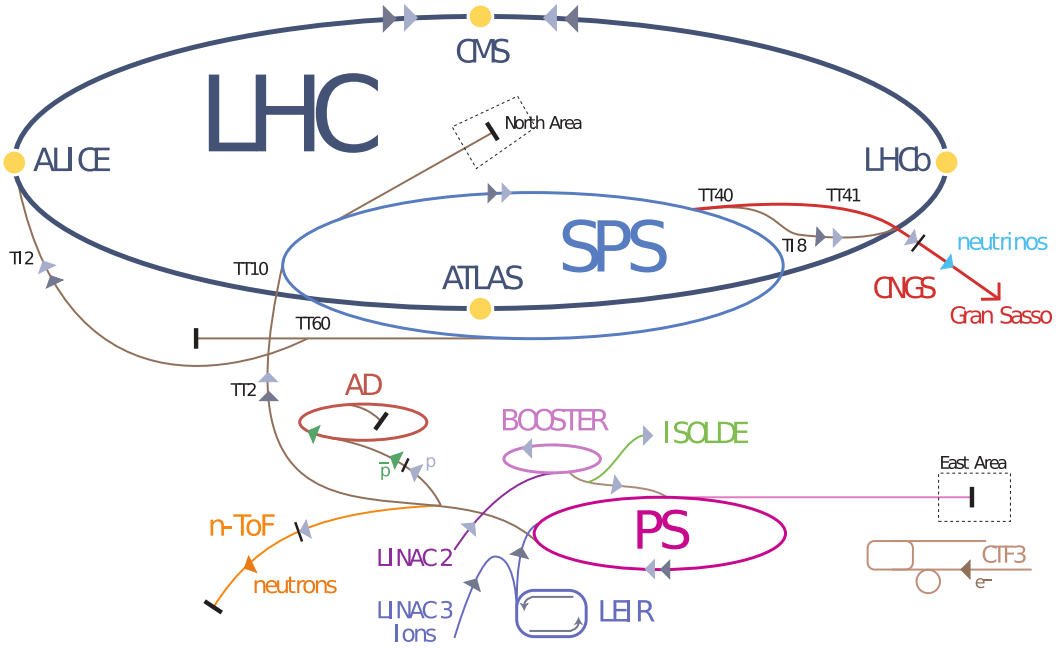


Figure 3.1: Overview of the accelerator chain. Protons or heavy ions are pre-accelerated in a series of linear accelerators and synchrotrons before being filled into the main LHC ring accelerator [49].

each proton beam is subdivided into bunches. Each beam pipe can hold 2808 bunches and a typical bunch contains more than 10^{11} protons. In this configuration, the time interval between two bunches at a point on the beam axis is 25 ns. The Lorentz force of 1232 superconducting dipole magnets keeps the particles on a circular track inside the beam pipe of 3 cm diameter. Superconducting magnets surround the beam pipe which are cooled using superfluid helium at 1.9 K. The pressure inside the beam pipe is around 100 nPa and even 1 nPa at the interaction points to avoid collisions with gas molecules. The beam is focussed by a complex system of quadrupol and higher order magnets as well as collimators.

After the construction phase, first collisions were envisaged in 2009. A momentous incident, caused by a finite resistance in one of the junctions between the superconducting magnets of the LHC delayed the start of physics measurements by one year. Additional fuses were installed at all such junctions to avoid similar electrical breakdowns and quenching of the magnets in future. In November 2009, the LHC produced first collisions. The centre-of-mass energy $\sqrt{s}$ was successively increased, setting new records in high energy physics, reaching 7 TeV in February 2010. The collisions in 2011 were produced at this energy. This key figure was increased to

8 TeV for the LHC run in 2012. A list of the most important LHC parameters is given in Table 3.1, showing the design values as planned in 2006 compared to the values achieved in 2011 and 2012.

Parameter		design (2006)	2011	2012
Energy per nucleon	E/TeV	7	3.5	4
Magnetic dipole field	B/T	8.33	4.17	4.76
Instantaneous luminosity	$\mathcal{L}/10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$	10	3.7	7.7
Bunch separation	t_b/ns	25	50	50
Number of bunches	n_b	2808	1380	1380
Number of particles per bunch	$N_\mathrm{p}/10^{11}$	1.15	1.45	1.7
Normalized emittance	ε_n/µm	3.75	2.4	2.5
γ-factor	γ	7460	3730	4262.9
β-value at interaction point	$\beta^\star$/m	0.55	1.0	0.6
RMS beam radius at IP	σ/µm	16.7		
Integrated luminosity (deliv.)	$L_\mathrm{del}/\mathrm{fb}^{-1}$	0.04	6.1	23.1
Integrated luminosity (rec.)	$L_\mathrm{rec}/\mathrm{fb}^{-1}$	0.036	5.0	19.7
Number of collisions/crossing	$\bar{n}_\mathrm{c}$	19	17	37
Stored beam energy	E_b/MJ	362	110	140

Table 3.1: Design specifications of the LHC [50] compared to parameters used during the LHC runs in 2011 and 2012 [51, 52] for proton-proton collisions. This is only an overview as some parameters varied over the run periods.

While most of the time is spent for the machine setup, the time in which the LHC produced events that were recorded by the experiments increased from 63 days and 10 h (23.5 %) in 2011 to 73 d and 17 h (28.4 %) in 2012. After the current shut-down, the proton-proton collisions will be resumed in 2015 at a centre-of-mass energy of 13 TeV.

3.1.1 Luminosity

The expected event rate, $\dot{N}$, for any process can be factorized in the luminosity L, provided by the accelerator, and the cross section of the process σ,

$$\dot{N} = L \cdot \sigma \,.$$

The luminosity can be expressed as

$$L = \frac{n_1 n_2 f_\mathrm{rev}}{A_\mathrm{eff}} = \frac{n_p^2 n_\mathrm{b} f_\mathrm{rev} \gamma F}{4\pi \varepsilon_n \beta^\star}$$

where n_1 and n_2 are the number of particles in the colliding beams, n_p is the number of particles per bunch and n_b is the number of bunches in a beam, f_rev the revolution frequency, γ the relativistic gamma factor and F the geometric luminosity reduction factor. The effective collision area of the two beams is given by the normalized emittance ε_n and the value of the betatron function at the interaction point $\beta^\star$.

To obtain the total number of expected events, the total integrated luminosity $\mathcal{L}$ can be calculated as

$$N = \int L \cdot \sigma \, \mathrm{d}t = \mathcal{L} \cdot \sigma \, .$$

This requires a constant beam energy and hence a constant cross section. Tuning of beam parameters, as summarized in Table 3.1, increased the instantaneous luminosity dramatically over the past years as can be seen in Figure 3.2. The integrated luminosity of this period is shown in Figure 3.3.

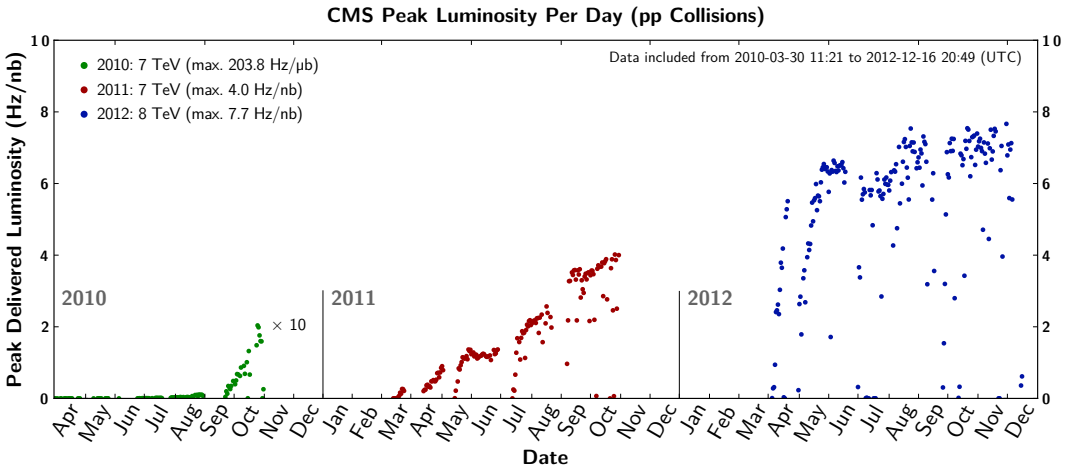


Figure 3.2: Peak luminosity per day delivered to the CMS Experiment by the LHC. It shows both the dramatic increase of the luminosity within the run periods and between the years. While the conditions were relatively stable in 2011, the picture is different for the 2012 run period [52].

The luminosity is measured by each experiment at its interaction point. Two methods are used in CMS for real-time measurements of the relative instantaneous luminosity: The HF method is based on the determination of the particle flux in the hadron forward calorimeter. The alternative method measures the luminosity by counting the number of vertices reconstructed in the pixel tracker of the CMS detector. In the vast majority of run periods, these two measurements agree as can be seen in Figure 3.4 for the 2012 run period.

The absolute luminosity measurement uses van-der-Meer scans in separate runs. The uncertainty on the result is 2.5 % (syst.) + 0.5 % (stat.) for the 2012 data set [53].

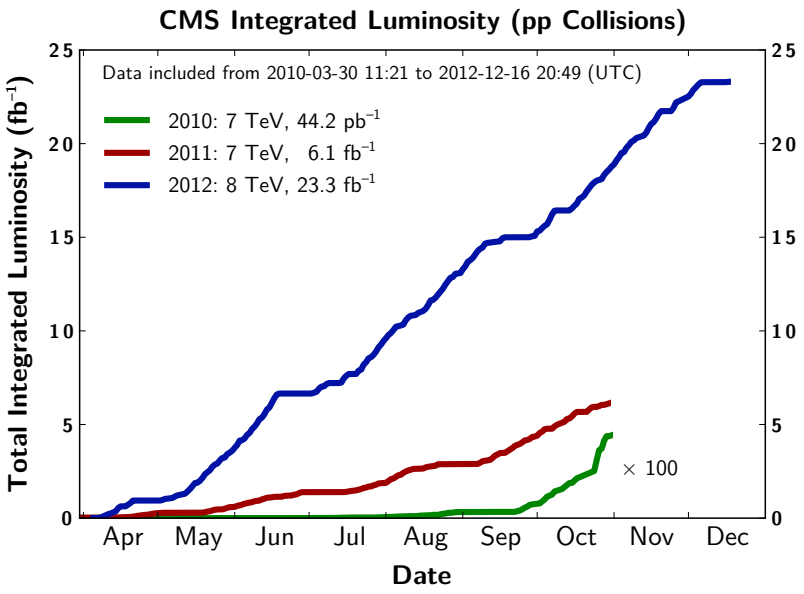


Figure 3.3: Integrated luminosity in pp collisions delivered to the CMS Experiment in 2010, 2011 and 2012 [52].

3.1.2 Pile-up interactions

As the processes that have not been studied by previous collider experiments are relatively rare (cf. Fig. 2.3), the LHC was designed to deliver the highest possible luminosity. The triggers then make a preselection on interesting event signatures while the majority of strongly interacting collisions is not even written out. This high luminosity comes at the price of multiple proton-proton interactions within the same event, called pile-up (PU). In all events, the hard interaction, that is the reason for triggering and analysing it, is overlaid by a large amount of additional interactions, mostly containing low energetic jets.

The vertices of pile-up interaction are reconstructed from tracks pointing to them. These vertices, including the vertex of the hard interaction, are called primary vertices in contrast to vertices arising from the decay of unstable particles with measurable decay length, called secondary vertices. The main interaction vertex of the hard interaction is identified as the one with the highest $\sum p_{\mathrm{T}}^2$ over all associated tracks. The reconstruction efficiency for pile-up vertices is about 70 %.

Figure 3.5 illustrates the response of the calorimeters over time which is read out in time slices (TS) of 25 ns. Hence, the interval between two collisions in 2012 corresponds to two time slices. This information is integrated for the evaluation of the event. Pile-up activity within the window of two time slices is called *in-time pile-up*. Activity in the detector, which is read out for the reconstruction of an event

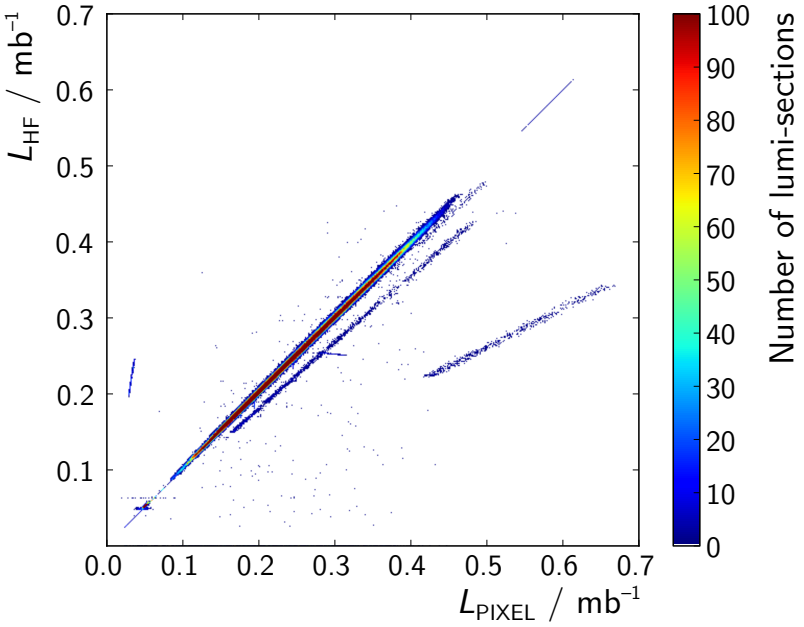


Figure 3.4: Correlation of the luminosity measurements in 2012 using the HF and the pixel luminosity method. For the vast majority of lumi-sections, both measurements agree. However, some lumi-sections exhibit an ambiguous lumiosity measurement and are thus rejected in the analysis of pile-up quantities.

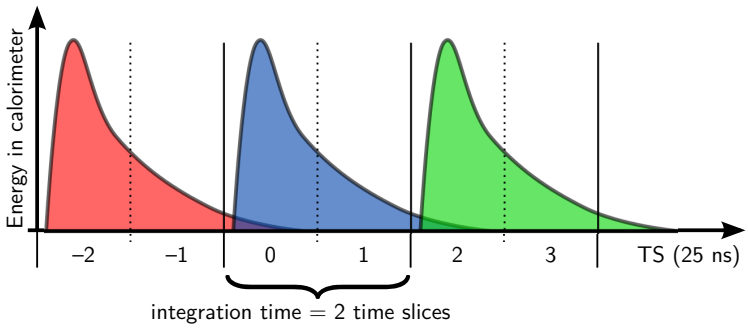


Figure 3.5: Calorimeter response curves and time slices [54]. The signal of the calorimeters is integrated over two time slices of 25 ns each. This reduces the impact of neighbouring interactions compared to a longer integration time.

but originates from interactions of a preceding or following bunch crossing, is called *out-of-time pile-up* (OOT). It can be further distinguished between OOT PU from the bunch crossing before (early OOT PU) and from the bunch crossing after (late OOT PU) the current event.

Compared to 2011, the integration window has been reduced to two time slices for the data taking in 2012. This reduced the influence of out-of-time pile-up significantly and late out-of-time pile-up is almost completely rejected [54].

3.2 The CMS Experiment

The Compact Muon Solenoid (CMS) is one of the two general purpose detectors at the LHC located at one of the interaction points near the village of Cessy in France. It is designed to measure the properties of the particles emerging from proton-proton and heavy ion collisions. Its name emphasises three of most important features, the compact design, the particular design for muon measurement and the solenoid, providing a homogeneous magnetic field in the inner detector. The experiment was designed, built and is run by a collaboration of 3638 physicists [55] who also analyse the data, striving for new physical results. This section is based on the technical design reports [50, 56] and follows the description in [57].

Figure 3.6 gives an overview of the composition of the CMS detector in layers around the interaction point. A more realistic impression is provided by the photograph in Figure 3.7 in which the different components of the central detector are marked. In addition, the impact of different particle types within the detector is visualized.

After the introduction of the CMS coordinate system in the following section, the detector subsystems are described from the inner components outwards. A longitudinal section through the CMS detector is provided in Figure 3.11 showing all major components to scale.

3.2.1 The coordinate system

CMS uses a right-handed Cartesian coordinate system with its origin inside the detector at the nominal interaction point. The x-axis points towards the centre of the LHC ring and the y-axis points upwards. Consequently, the orientation of the z-axis is such that the axes form a right-handed coordinate system and points therefore westwards to the Jura mountains, tangential to the beam pipe. According to customary conventions for polar coordinate systems, the azimuthal angle ϕ is measured from the x-axis in the x-y-plane and the polar angle θ is measured from the positive z-direction.

In high energy physics, the quantities rapidity y and pseudorapidity η are prevalent instead of the polar angle, as differences in rapidities are invariant under Lorentz-

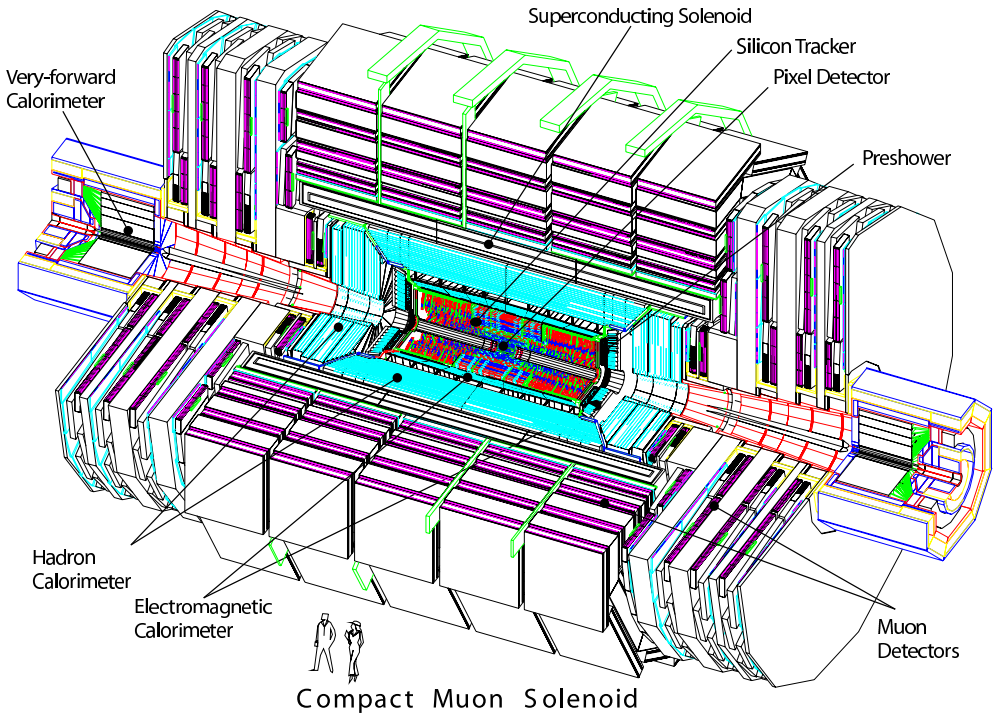


Figure 3.6: Schematic view of the CMS detector. The CAD drawing reveals the multilayer design centred around the nominal interaction point which is located inside the beam pipe. Two humans are depicted as a scale reference [56].

boosts in the $\pm z$-direction. The rapidity y is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) . \tag{3.1}$$

In addition to the rapidity, the pseudorapidity η is also widely used, as the latter does not depend on the mass of the particle but only on the polar angle θ. Rapidity and pseudorapidity converge for massless particles or comparatively high energies

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) . \tag{3.2}$$

The angular distance ΔR between two particles coming from a vertex at the origin of the coordinate system is defined by

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} . \tag{3.3}$$

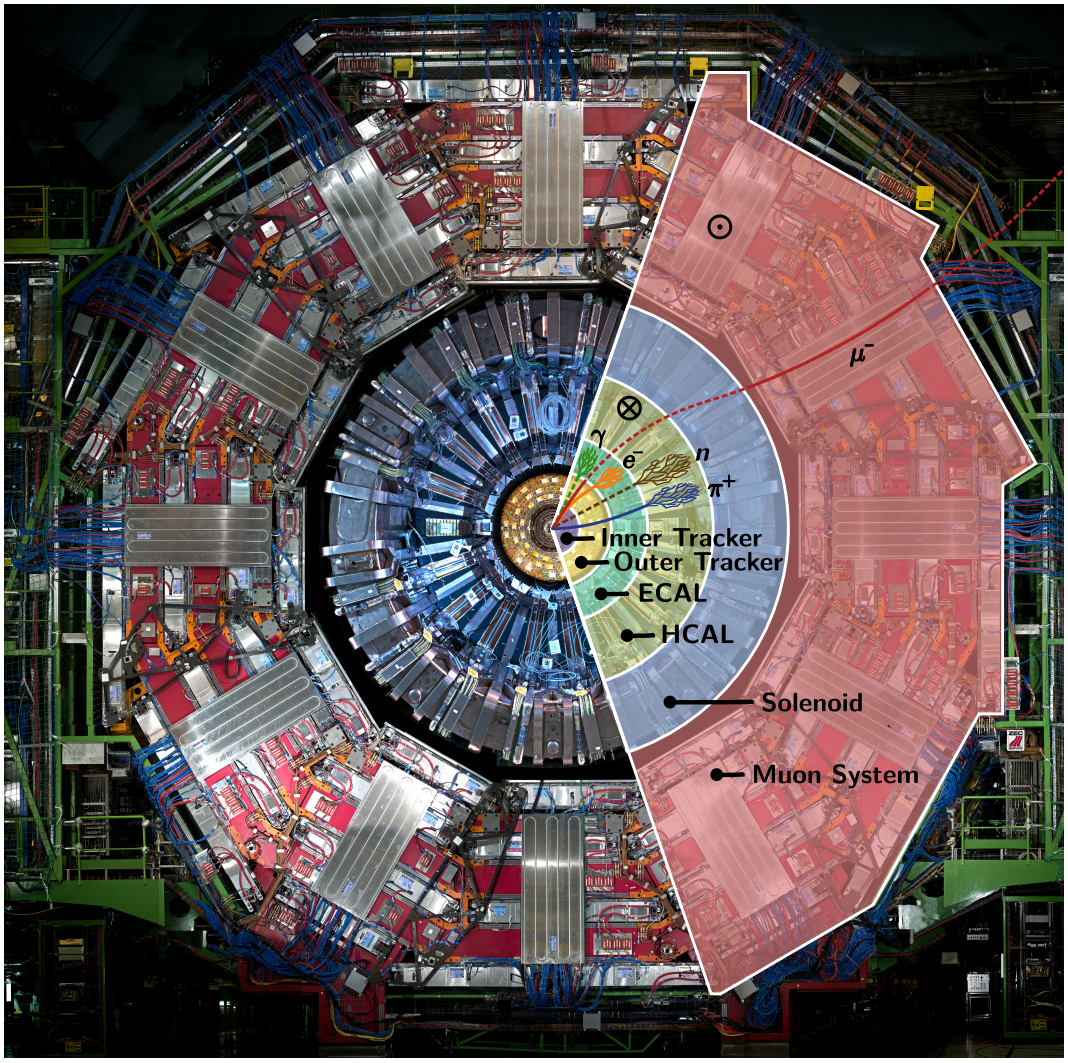


Figure 3.7: The CMS detector. *Components*: A schematic division into detector components is presented in the right half of the picture (• inner tracker, • outer tracker, • electromagnetic calorimeter, • hadronic calorimeter, • solenoid, • muon system). The direction of the magnetic field is shown inside the solenoid (⊗) and in the return yoke (⊙). *Particles*: The path of particles is indicated by dashed (invisible track) and solid (visible track) lines for different particle classes: photons γ, muons μ, electrons e, neutral hadrons (neutron n) and charged hadrons (pion π^+). This image is composed of several photos to provide a single cross section through the detector (photograph taken from [58]).

Instead of using Cartesian coordinates, components of four-momenta in the laboratory frame are often stated in terms of the mass, azimuthal angle, pseudorapidity and transverse momentum of the object,

$$(m, \phi, \eta, p_{\mathrm{T}})\,. \tag{3.4}$$

3.2.2 The inner tracking system

Located as close as possible to the interaction point, the tracking system represents the innermost part of the CMS detector. It is constructed to measure precisely the outgoing charged particle tracks from LHC collisions. This information is used to determine the direction of a particle, its transverse momentum – exploiting its mass hypothesis – and to identify the vertices, the particles originate from. The tracking system has a length of 5.8 m and a diameter of 2.5 m. It covers the pseudorapidity range of $|\eta| < 2.5$. The inner tracking system comprises two sub-systems: the pixel tracker and strip tracker as shown in Figure 3.8.

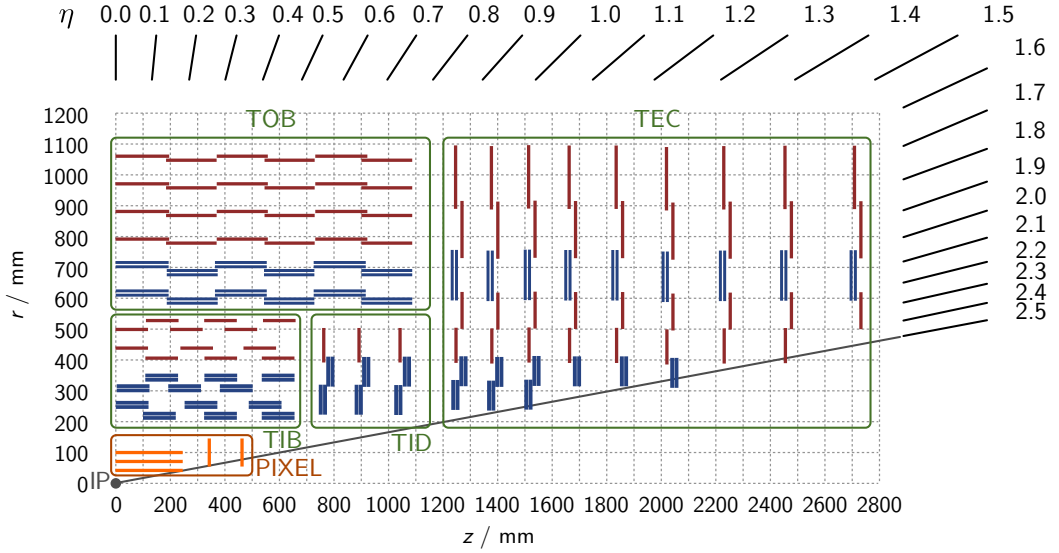


Figure 3.8: One quadrant of a longitudinal section of the inner tracking detector of CMS along the r-z-plane. The components of the silicon pixel detector are shown in orange. The strip detector comprises four components: The Tracker Inner Barrel (TIB) is complemented by the Tracker Inner Disks (TID). These two are surrounded by the Tracker Outer Barrel (TOB). High η ranges are covered by the Tracker End Cap (TEC) up to $\eta = 2.5$. The red components represent single-sided silicon strip modules, while the double-sided modules are shown in blue (adapted from [59] and [44]).

The *pixel tracker* consists of three concentric layers of pixel modules forming a barrel which are complemented by two disks on each side. The location requires particularly radiation-resistant components. About a thousand particles pass the pixel tracker per bunch crossing, which yields an occupancy of $\mathcal{O}(10^{-4})$ per pixel and bunch crossing [56]. The pixel sizes are 100 µm × 150 µm in r-ϕ and z, respectively. By taking advantage of the large Lorentz effect, the pixel tracker reaches a resolution of 10 µm × 20 µm. This high spatial resolution is needed for a precise determination of the vertex positions and the required momentum resolution. The latter is as good as 1 % for a track of a 10 GeV particle.

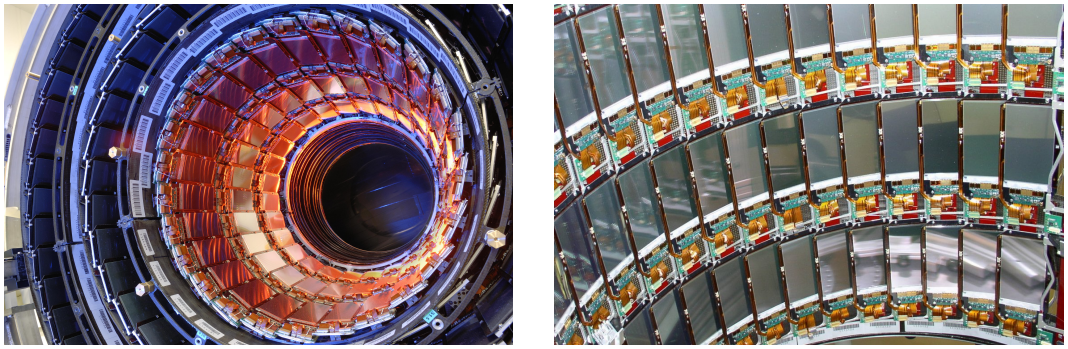


Figure 3.9: One half of the CMS inner tracker barrel (TIB) and a closer view of its silicon strip modules [60].

The *strip tracker* measures the particle tracks with a reduced resolution ($\leq$ 23 µm), reflecting the smaller particle flux at larger distances from the interaction point. The one-dimensional strip layout allows the tracker to be constructed at lower costs. 24 240 silicon microstrip sensors cover an area of 200 m^2. The tracker is run at an operating temperature of around −20 °C.

3.2.3 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) serves to measure the energy of electrons, positrons and photons. A particle, entering the electromagnetic calorimeter, results in an electromagnetic shower, caused by bremsstrahlung and pair production. The energy of the particle is deposited in the calorimeter material via Compton scattering and the photo-electric effect. The energy deposit is proportional to the energy of the original particle. The absorber material is made out of 75 848 lead tungstate ($PbWO_4$) crystals. This material was chosen because of its short radiation length $X_0 = 8.8$ mm and the small Molière radius of $R_M = 22$ mm. This is also reflected in its high density of $\varrho = 8.3\,\mathrm{g/cm^3}$. Furthermore, lead tungstate shows a fast light emittance of 80 % during the time between two bunch crossings (25 ns). Avalanche

photodiodes are attached to these scintillating crystals in the barrel. In the endcaps, vacuum phototriodes are used, which are more radiation resistant.

The electromagnetic calorimeter comprises two parts: the barrel (EB) for a pseudorapidity range up to $|\eta| = 1.479$ and the two endcaps (EE), covering $|\eta|$-values from 1.479 to 3.0. For a better discrimination of photons against neutral pions, a pre-shower device (SE) is mounted in front of the ECAL endcap. Silicon strip sensors with a resolution of 2 mm, placed behind two planes of lead, are able to distinguish single photons from pion decays into photon pairs which would not be possible with the typical crystal size of 30 mm.

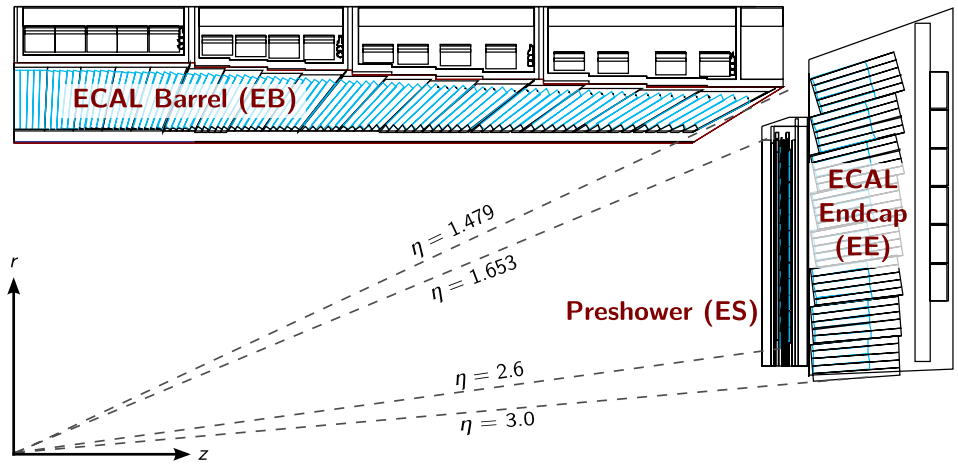


Figure 3.10: The electromagnetic calorimeter of the CMS detector [56] showing the barrel component and the endcap with the additional preshower device.

3.2.4 The hadronic calorimeter

The hadronic calorimeter (HCAL) is a sampling calorimeter surrounding the aforementioned parts of the CMS detector. The compact design is a consequence of the layout as a sampling calorimeter and permits the placement of the main parts of both calorimeters inside the solenoid. Brass has been chosen as the absorber material of the HCAL, as it is non-magnetic and has a relatively short interaction length of $\lambda_I = 16\,\text{cm}$. The large fraction of passive material is cost efficient. On the other hand, it requires detailed in-situ calibrations and the energy resolution is reduced compared to what is possible with homogeneous calorimeters. The entire subsystem can be divided into four parts, which are explained in the following paragraphs:

Hadron Barrel (HB) The HB is the part of the hadronic calorimeter with the highest resolution ($\Delta\eta \times \Delta\phi = 0.087 \times 0.087$). It is therefore the optimal region for a

data-driven calibration of the jet energy scale, which will be explained in Chapter 4.2. For η between 1.305 and 1.392, the barrel overlaps with the endcap. The thickness of the HCAL amounts to $7 - 11$ interaction lengths and should therefore be sufficient to stop nearly all hadrons in the calorimeter.

Hadron Outer (HO) However, not all hadrons can be stopped within the limited range of the hadron barrel. For this reason, a small part of the HCAL is located outside the solenoid, called the hadron outer barrel. This part serves as a so called "tail catcher" and measures the energy of hadron showers penetrating the magnetic coil. It spans a pseudorapidity range up to $|\eta| = 1.26$ which is comparable to the reach of the HB.

Hadron Endcap (HE) With 14 additional calorimeter towers, the hadron endcap increases the coverage to the region $1.3 < |\eta| < 3.0$. The ϕ-segmentation, which is 5° near the HB, changes to 10° for the outermost towers. The η-segmentation increases to 0.35 at the highest η. As this component of the hadronic calorimeter extends close to the beam pipe, is has to be particularly radiation hard. Continuous losses of the detector response have to be monitored and corrected for during data taking.

Hadron Forward (HF) To cover even higher values of the pseudorapidity in the range $2.8 < |\eta| < 5.2$, the hadron forward calorimeter is located at $z = \pm 11.2\,\mathrm{m}$ from the interaction point close to the beam pipe. Jets with very high pseudorapidities and the hadronization products of the beam remnants are detected with this system. A square grid of quartz fibres is inserted into grooves in the steel plates which serve as absorber material. Cherenkov light, emitted in the quartz fibres, produces the signal in the photomultipliers. The HF is 1.65 m long and divided into 900 towers.

3.2.5 The superconducting solenoid

The superconducting solenoid magnet is eponymous for the CMS detector. It provides a homogeneous magnetic field in the inner detector with a magnetic flux density of $B = 3.8\,\mathrm{T}$. It ensures the bending of particle tracks, and hence permits particle momenta to be measured. Muons particularly require a good momentum resolution in order to distinguish and identify high energetic muons of opposite charge. The 2168 turns of the solenoid are made of NbTi, which is superconducting at the operation temperature of about 4 K. The conductors carry a current of 19.5 kA. The total energy stored in the magnetic field amounts to 2.7 GJ. The support structure of the outer detector, made of steel, serves as a return yoke for the magnetic field outside the solenoid and provides a magnetic flux density of $B \approx 2.4\,\mathrm{T}$.

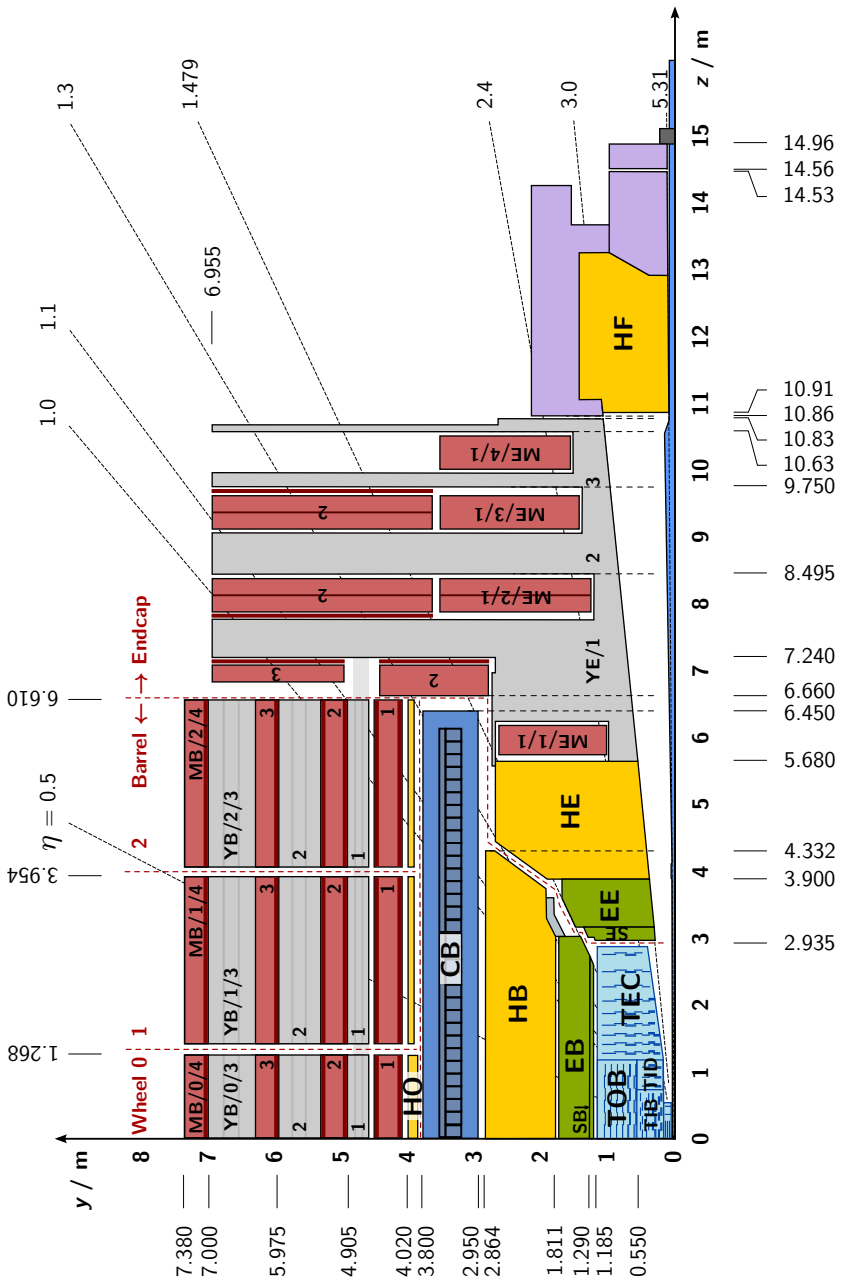


Figure 3.11: Longitudinal section of the CMS detector in the y-z plane. It shows the tracking detector close to the nominal interaction point at $(0, 0)$ and the electromagnetic (EB, EE) and hadronic (HB, HE, HO, HF) calorimeters. The coil of the solenoid magnet (CB) surrounds the inner barrel. The iron return yoke (YB, YE) is interleaved with the muon chambers (MB, ME). Resistive plate chambers are marked in dark red.

3.2.6 The muon system

The last characteristic feature of the CMS detector is the muon system surrounding the aforementioned detector components. Since muons give a clear signature to many event topologies to be observed at LHC, the CMS detector was designed with a precise muon reconstruction [56]. The muon system also provides the ability to trigger on muons over almost the entire rapidity range up to $|\eta| = 2.4$.

The muon system uses three different detection methods: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The barrel muon chambers ($|\eta| < 1.2$) are drift tubes. The five dodecagonal wheels of the steel support construction are interleaved with drift chambers, arranged in four rings (stations). With ten extra modules, this results in a total amount of 250 drift chambers. They are filled with a mixture of Argon and CO_2. The muon endcaps ($0.9 < |\eta| < 2.4$) consist of cathode strip chambers (CSC), arranged in four disks on each side of the barrel with one or three rings each. The 540 chambers are read out in 500 000 channels. The position resolution is about 1 mm, the time resolution is about 15 ns and the association to the right bunch crossing has an efficiency of more than 98 %. 612 resistive plate chambers (RPC) serve as an independent and complementary system for the first level trigger, covering the range up to $|\eta| < 2.1$. They comprise one sensitive layer with two gaps per chamber and have a time resolution of 3 ns.

3.2.7 The trigger and data acquisition system

During the data taking from 2010 to 2012, the collision rate was 20 MHz. At design luminosity (Tab. 3.1), two bunches will cross every 25 ns, i. e., with a frequency of 40 MHz. Since current hardware is not capable of storing data at that rate, a cascade of triggers drastically reduces the event rate to a manageable level. For this reason, the physically most interesting events are chosen by the trigger system, which is the first step in the delicate event selection process. It comprises two levels, the Level-1 trigger (L1) and the High Level Trigger system (HLT):

The *Level-1 trigger* is a very fast hardware trigger and is located near the detector. Wherever possible, the hardware is adaptable to future needs in the event selection by programmable logic circuits. The data are cached in hardware buffers until the L1 trigger decides, based on individual information of the tracker, the calorimeters or the muon system, whether the event is passed on to the HLT or discarded. The L1 trigger reduces the event rate to 100 kHz.

The *High Level Trigger* is a software based trigger system and runs on a processor farm. It has access to the whole event's raw data and is therefore able to combine information from different detector components and to use more complex algorithms. Thus, the event rate is reduced to 150 Hz. With an event size of 1.5 MB on average, 225 MB/s of raw data are written on storage and are at hand for later analysis.

Just as the quality of the analyses performed at CERN is based on the aforementioned properties of the detector, it also largely depends on the ensuing computing, to which the next section is dedicated.

3.3 Reconstruction of physical objects

The Particle Flow reconstruction [61, 62] used by CMS serves to identify particles in the event and combine the information of all detector subsystems. This information contains the tracks and vertices, reconstructed from the hits in the tracking detector, the clusters of energy deposits in the electromagnetic and hadronic calorimeters, divided into cells following the tower structure of the calorimeters, and the tracks in the muon system. The reconstructed particles are grouped into five classes: muons, electrons, photons, charged and neutral hadrons.

In each step, the identified particle, including all its traces in the detector, is removed from the list of input objects and the energy deposits are recalculated. This ensures that all energy is assigned to exactly one particle and thus avoids double-counting of energy contributions. The procedure starts with the muons as they can be clearly identified by their tracks in the muon system. Other tracks are combined with the energy deposit they point to. In case this energy is found mostly in the ECAL, these particles are classified as electrons. The corresponding energy is determined using multivariate regression techniques. The remaining tracks pointing to HCAL dominated clusters are identified as charged hadrons. Neutral particles are reconstructed from the remaining calorimeter clusters. Similarly to the charged particles, photons and neutral hadrons are distinguished according to their energy deposits. The algorithm results in a set of Particle Flow candidates and the missing energy calculated from the sum of all measured particles. Short-lived particles have to be identified and reconstructed in the analysis by their decay products. Neutrinos escape undetected and thus contribute to the missing energy.

3.3.1 Muon reconstruction

The kinematics of low energetic muons is measured precisely by the inner tracking system. For high energetic muons ($p_{\mathrm{T}} \gtrsim 200\,\mathrm{GeV}$) the momentum resolution is improved by up to an order of magnitude, compared to the resolution only determined from the inner tracker, by combining the information of the tracker and the muon system. This is achieved mainly due to the larger leaver arm when including hits in the muons system in the track fit. This is shown in Figure 3.12.

The charge of a muon is determined from the curvature of the track with over 99 % confidence level [63]. To achieve this, a good alignment and the knowledge about the relative positions between the muon chambers and the inner tracker elements are essential. In particular, the jet energy correction, described in Section 4.2, and the

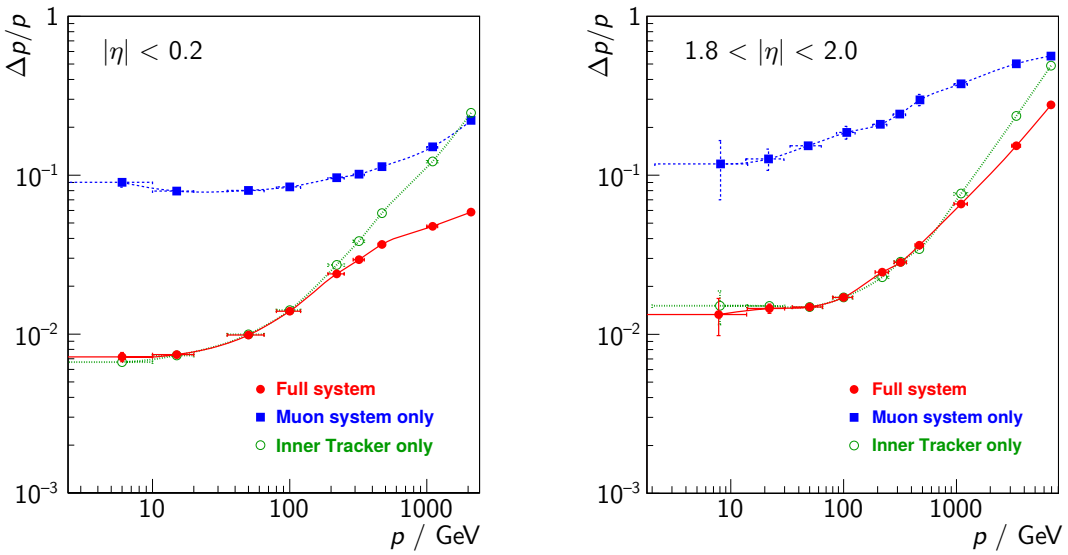


Figure 3.12: Muon resolution in the barrel (*left*) and in the endcap (*right*). In both cases, the momentum resolution is improved by the combined fit of the full system compared to tracks reconstructed from hits in the inner tracker or the muon system only [56].

Higgs search in the $H \to \tau\tau \to \mu\mu$ channel, described in Section 5.3, depends on the precise reconstruction of muons. The performance of the muon system can also be illustrated by the invariant mass spectrum of dimuon events shown in Figure 3.13.

3.3.2 Jet reconstruction and jet algorithms

A jet is a collimated stream of particles emerging form gluons or quarks in high energy collisions. Partons produced in such a collision radiate additional partons which eventually form hadrons that can be detected in the experiment. This process is described by the parton shower and hadronization in Monte-Carlo generators. The four-vector of a jet is defined by the result of a *jet alogrithm* which clusters the particles which are nearly collinear to a single object.

Jet algorithms can be applied on different input objects: four-vectors of generated particles (in this case the jets are called particle jets), tracks in the inner detector (track jets) and energy deposits in calorimeter towers (calorimeter jets). Neutrinos are excluded from the list of input objects of particle flow jets to be comparable to reconstructed jets. Muons are included and are usually treated by appropriate isolation criteria. Besides those, CMS defines jets which include the track momentum in addition to the calorimeter deposits (Jet-Plus-Track jets, JPT) and Particle Flow jets (PF) which are clustered from Particle Flow candidates. The latter is the default

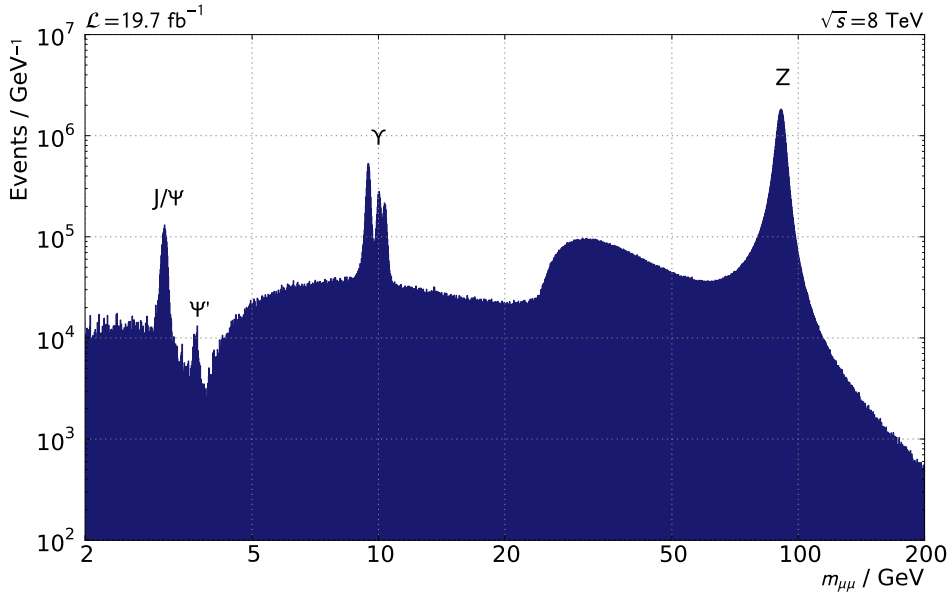


Figure 3.13: Distribution of the invariant mass of all dimuon pairs in 2012 data, passing the identification and isolation criteria (15.9 million events). The Z peak is located at 91 GeV and its width is 4.5 GeV, hence already in the order of the natural Z boson width. Different prescales of the available triggers results in several plateau structures of different height.

jet type in CMS in line with the Particle Flow reconstruction of all other objects. A further distinction can be made for reconstructed jets depending on the selection of the input objects: Regular jets perform the clustering on all of the input objects whereas Charged Hadron Subtraction (CHS) jets exclude particles from pile-up (Sec. 3.1.2) from the clustering. This method is described in Section 4.4.1. The FASTJET package [64, 65] is used for jet clustering.

Whether a particle is split into two collinear particles or not should not influence the result of a jet clustering algorithm. This property is called collinear safety. Equally, a jet is called infrared safe if additional soft input particles do not influence the clustering. These two cases are illustrated in Figure 3.14.

Jet algorithms can be divided into two classes: cone type jet algorithms and sequential clustering algorithms. *Cone type jet algorithms* have a constant distance measure in the η-ϕ space and hence a fixed jet cone. The strive for infrared and collinear safety led to a long series of cone type algorithms, differing in the seeding mechanism and the clustering procedure. The Seedless Infrared-Safe Cone Algorithm (SIS-Cone) [66] is one of the newest cone type algorithms. It is both collinear and

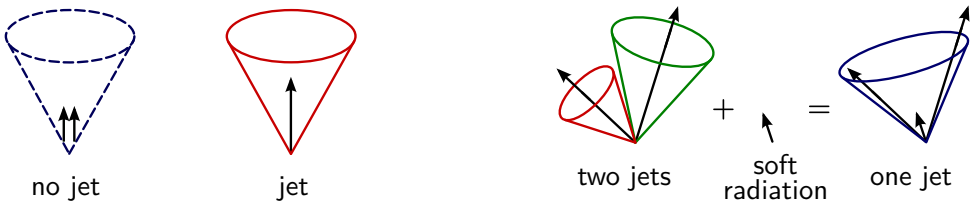


Figure 3.14: Collinear and infrared safety of jet algorithms.

infrared safe. However, it is only rarely used in recent analyses. The issue of infrared and collinear safety is addressed by construction by *sequential clustering alogrithms* which are described in the following paragraph.

k_T, Cambridge/Aachen and Anti-k_T Algorithms In contrast to the aforementioned cone-based algorithms, the k_T algorithm [67, 68], the Cambridge/Aachen algorithm and the Anti-k_T algorithm [69], which are most used within the CMS Collaboration, belong to the class of sequential recombination algorithms. They do not produce a jet of a fixed geometrical shape, as the distance measure has no purely geometrical definition, but is weighted by the transverse momentum of the input objects. The relative influence of both criteria on the distance measure is ruled by a continuous parameter q, which sets the power of the transverse momentum scale relative to the geometrical distance [69].

Distances between entities (particles or proto-jets) are given by d_{ij} and distances between an entity and the beam are denoted as d_{iB}.

$$\begin{aligned} d_{ij} &= \min(p_{\mathrm{T},i}^{2q}, p_{\mathrm{T},j}^{2q}) \, \frac{\Delta R_{ij}^2}{R^2} \\ d_{iB} &= p_{\mathrm{T},i}^{2q} \end{aligned} \tag{3.5}$$

where the angular distance is defined by ΔR_{ij}^2 is the angular distance between the objects i and j in the η-ϕ space and $p_{\mathrm{T},i}$ is the transverse momentum of the object i.

The sequential recombination algorithms also start with a list of entities. The procedure searches for the smallest distance. If it is a distance between two entities i and j, they are combined, whereas i is called a jet and moved from the object list to the list of jets, if d_{iB} is the smallest distance. For the inclusive jet clustering, the distances are recalculated and this procedure is repeated until there are no entities left. Finally, the list of jets contains all the jets found with the algorithm. This list is returned as output of the algorithm. By default, the combination of two objects implies four-vector addition. The sum of all jet momenta is thus equal to the momentum sum of all input objects.

In case of $q = 1$, one retrieves the k_T algorithm which is used since 1992 [67, 68]. Values q between 0 and 1 result in very similar jets to the ones clustered by the k_T algorithm as the ordering between particles is preserved.

If $q = 0$, Eq. (3.5) represents the distance measures of the Cambridge/Aachen algorithm. It is used to cluster jets with very large jet size parameters followed by the analysis of subjets using the k_T algorithm. Such studies of the jet substructure are mainly done to identify decay products of highly boosted objects.

The Anti-k_T algorithm ($q = -1$) is used the most within the CMS Collaboration. Even though the Anti-k_T algorithm is not starting with a fixed cone, the resulting jet shapes are fairly cone-like (cf. Fig. 3.15) and the boundaries of the catchment areas are particularly resilient with respect to soft radiation. The Anti-k_T algorithm with the jet resolution parameter $R = 0.5$ is the default in CMS and it will be used uniquely throughout this work even though all mentioned algorithms have been studied.

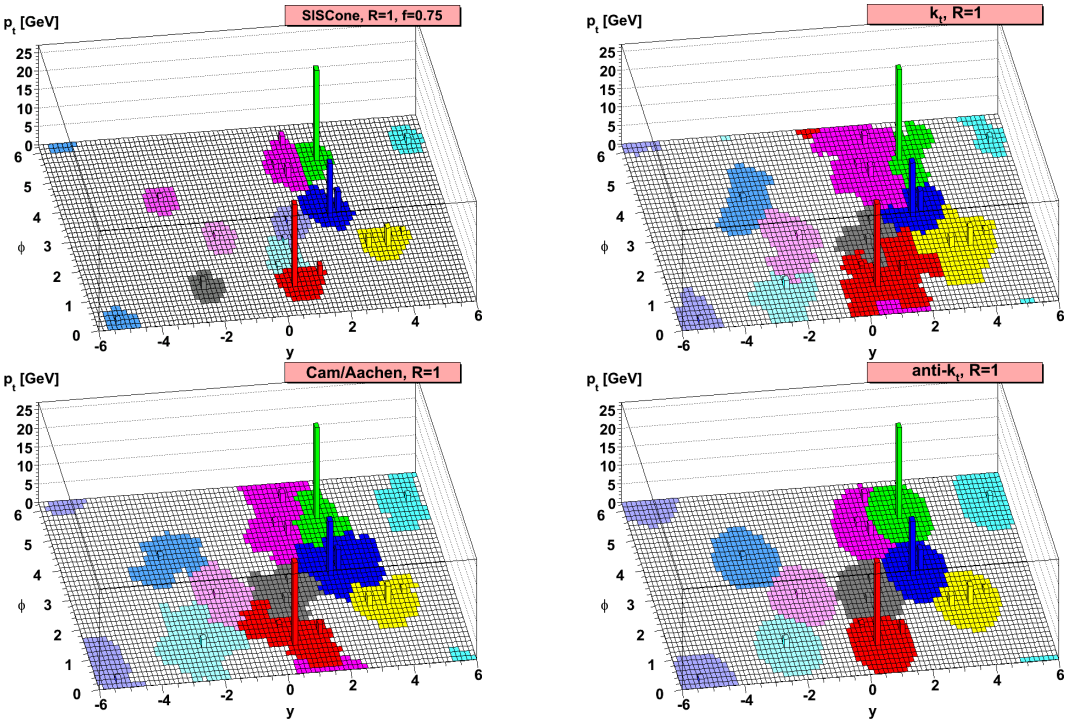


Figure 3.15: Jet areas for a selection of jet algorithms used within the CMS Collaboration [69].

The jet area A_j is defined [70] as the catchment area of the jet j in the η-ϕ space and depends on the used jet algorithm. While non-overlapping jets clustered with cone based algorithms have fixed jet areas defined by the R-parameter, the jet area for sequential clustering algorithms has to be measured for each jet. The

Table 3.2: File sizes and numbers of events for different analysis levels. This is an example for all data of the 2012 run in the dimuon trigger stream.

Format	no. of events	file size
EDM format	1.1×10^8	132.9 TB
Kappa skim	3×10^7	821 GB
flat n-tuple	1×10^6	500 MB

measurement [71] exploits the infrared-safety of the involved jet algorithms to estimate the area covered by a jet. To this end, a large number n_{all} of randomly and evenly distributed particles with negligible momenta are added to the event prior to the jet clustering. These so-called ghost particles are evenly distributed over the whole area within the detector acceptance $A_{\text{acceptance}}$. The area of a particular jet A_j is then assumed to be proportional to the number n_j of those ghost particles clustered into this jet.

$$A_j \equiv \frac{n_j}{n_{\text{all}}} \cdot A_{\text{acceptance}}$$

The acceptance extends up to $|\eta_{\text{max}}| = 5$, hence $A_{\text{acceptance}} = 20\pi$. The pile-up and underlying event p_{T}-density ρ is defined as

$$\rho \equiv \underset{j \in jets}{\text{median}} \frac{p_{\text{T}}^{\text{j}}}{A_j}$$

where all jets up to $|\eta_{\text{max}}| = 4.4$ are considered. For the ρ calculation, the k_{T} algorithm with a distance parameter of $R = 0.6$ is used for the jet clustering. Figure 3.15 illustrates the shape of jet areas for the cone-type SIS-Cone algorithm and the three described variants of the sequential clustering k_{T} algorithm.

3.4 Analysis workflow

The analysis is divided into three steps: Skimming, the analysis core and plotting as sketched in Figure 3.16. These steps have well-defined interfaces represented by data files in the ROOT [72, 73] file format. Each step significantly reduces the file sizes and the run time to further process the data as listed in Table 3.2 for the $Z(\rightarrow \mu\mu) + \text{jet}$ analysis in Chapter 4. Only basic changes to the analysis strategy require the time consuming first steps whereas most studies can be accomplished within seconds in the final plotting step.

Skimming The first step in each analysis is to retrieve the recorded information about CMS collisions and about the events simulated under the same conditions.

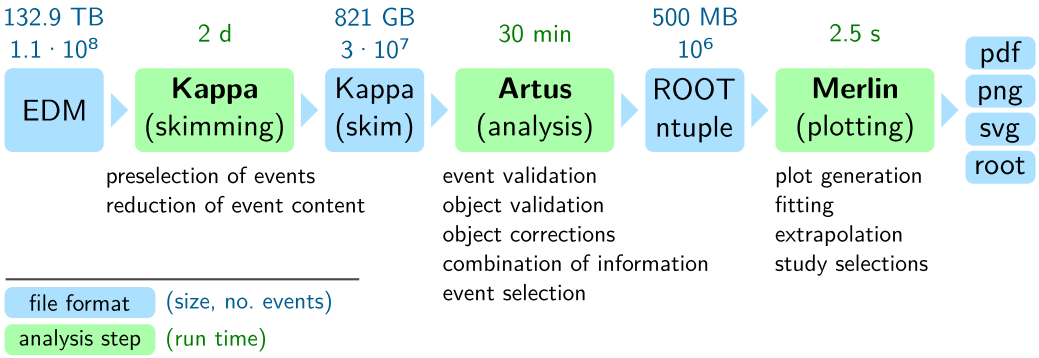


Figure 3.16: Analysis chain for analyses performed with the Artus analysis framework.

Reconstructed event data from CMS collisions as well as from simulated Monte-Carlo samples is stored in the Grid. This data is stored in the Event Data Model (EDM) format defined by CMS.

The typical size of all data recorded in 2012 in the double muon trigger stream amounts to 110 million events. Each of the several copies of this dataset occupy 132.9 TB of storage space on their respective Grid site. The CMS Software Framework (CMSSW) provides the definition of the EDM objects and classes to analyse them. The KAPPA framework, developed at the institute [74], provides a slim but comprehensive data format in which only the relevant event content is stored. The job submission tool GRID-CONTROL [75] was used to handle the GLITE [76] middleware and local batch systems for parallel processing of the skimming and analysis jobs on the Grid and local computing clusters. GRID-CONTROL is a feature-rich and user-friendly tool with a large set of configuration options.

The skimming job includes a rough preselection of events, a pair of reconstructed muons is required. This reduces the number of accepted events by a factor of 3.7. In addition, the information stored for each event is reduced to physically relevant part (factor 44). In combination, the size of the dataset written on disk is reduced by a factor of 160. This procedure increases the runtime of the following analysis step dramatically as the bandwidth for reading data is the limiting factor.

Analysis The main analysis logic is implemented in the *Artus* framework which was developed as part of this thesis for the Z+jet based jet energy calibration as presented in Chapter 4.2. One of the main features is that different settings can be processed for each event at once without rerunning the analysis. The object identification, the events selection, corrections and the calculation of derived quantities are done by a flexible set of producers, filters and consumers. All these processors can be individually enabled and configured by the options given in the configuration file,

written in the json file format.

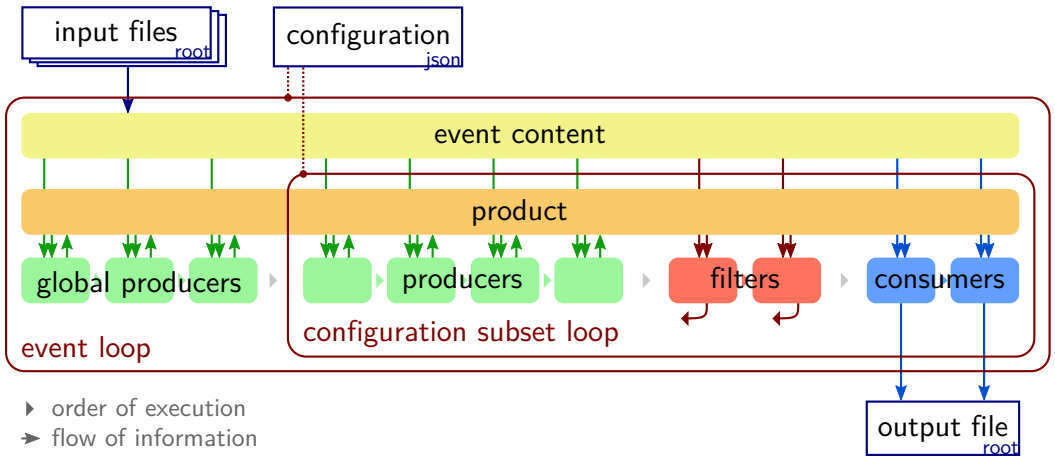


Figure 3.17: The workflow of an analysis performed with Artus.

Different *producers* provide all methods to validate, correct and sort the physical input objects such as muons or jets. Further producers calculate new quantities from the variables in the skim or produced by preceding producers, e. g. the Z boson from two muon candidates.

A subsequent set of configurable *filters* select the events that pass criteria given in the configuration file. Typical filters applied in this analysis are trigger decisions, the validation of events according to CMS data validation and the event selection based on physical quantities.

The last class of operators that act on each event are *consumers*, which retrieve the information from the event and the product as specified in the configuration. This information is then written to the output file in the form of n-tuples or histograms separately for each configuration subset.

Present CPUs are capable of running the analysis on all 2012 data within 30 min. The read rate of input data is the limiting factor and can further increase the run time. The turnaround time is reduced as GRID-CONTROL is used for data parallelisation. Table 3.3 summarizes the run time spent for different parts of the analysis.

Post-analysis and plot creation In a third step, the reduced sample of selected events is the basis to a versatile plotting framework. Additional cuts can be introduced while creating final plots and their effects can be studied with turnaround times of a few seconds. Generic plotting functions are based on python's plotting library MATPLOTLIB. Default parameters for all settings ensure that the plotting of arbitrary combinations of new variables are well formatted and most details can be overwritten

Table 3.3: Performance of the main analysis step. The time consumption is given with respect to a typical analysis set-up, run on the local batch cluster of the institute with 40 jobs running in parallel.

Analysis		Time (20 min)
read event	41.7%	00:08:20
correct objects	26.1%	00:05:13
sort objects	5.6%	00:01:08
cuts, calculations, ...	9.3%	00:01:51
write output tree	7.3%	00:01:27

by command line options. In this step, several datasets, processed in parallel so far, are combined into a defined set of plots. Filling histograms, fitting curves to data, extrapolation (cf. Section 4.8.3) falls into this step. The time consumption of the post-analysis calculations is small and most of the time is spent for opening the input file and for the output file creation by MATPLOTLIB. These benchmark numbers are summarised in Table 3.4.

Considering the performance, flexibility and feature set of this workflow, this framework became now more widely used within the institute and beyond. It will serve as basis for a large variety of particle physics analyses, ranging from QCD studies over Higgs analyses to new physics searches. The next chapter presents the original purpose for this framework to be created which served as the realistic benchmark scenario for the figures above.

Table 3.4: Performance of the post-analysis. Time consumption in different parts of the post-analysis for a single plot and for a set of thousand plots. The table shows that this is limited by the read rate of input data.

Plotting		**Time/s** (1 plot)	**Time [min:s]** (1000 plots)
read from tree	21.6%	0.555	09:15
open file	10.0%	0.257	00:01
load tree	0.9%	0.023	00:23
init histo	1.0%	0.026	00:26
project tree	9.7%	0.249	04:09
matplotlib conversion	1.2%	0.031	00:31
create new plot	2.4%	0.062	01:02
add curves	2.9%	0.074	01:14
labels, calculations, etc.	0.4%	0.010	00:10
matplotlib save plot	71.5%	1.836	30:36
sum		2.568	42:48

Chapter 4

Jet Energy Calibration with Z+Jet Events

At a high-energy hadron collider like the LHC, jets are ubiquitous. Thus, nearly every measurement in CMS involves the treatment of jets as part of its signal or background processes. The jet energy scale is defined as the energy of the measured jet with respect to the true jet's energy. The purpose of jet energy corrections is to determine correction factors to scale the energy of the measured jet back to the energy of the final state particle jet which is inaccessible in an experiment. The particle jet is clustered from final state Monte-Carlo particles by the given jet algorithm.

The jet energy scale is a major source of systematic uncertainty for many analyses. This is particularly true for studies of the VBF Higgs boson production topology as will be shown in Chapter 5.

The entire recorded data for the respective run period of the LHC was used in order to maximise the statistical precision and to provide corrections for all recorded data, whereas the figures in this chapter only show data from collisions at a centre-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$.

Figure 4.1 illustrates the evolution of a jet from a parton of the hard interaction to the final jet measured in the detector. Final state radiation can lead to further partons in addition to the initial parton in a process that is described by the parton shower in Monte-Carlo generators. Due to confinement, colour-neutral hadrons form the hadronic final state of the event. These particles enter the detector where the tracks of charge particles are measured in the silicon tracker and the energy of all particles is measured in the electromagnetic and hadronic calorimeters (ECAL and HCAL). From these deposits the particle flow candidates are reconstructed which are subsequently clustered into jets. The jet energy corrections relate the reconstructed jet energy to the true jet energy. Electronic noise and out-of-cone effects are indicated as examples for sources of miscalibration.

Additional parton-particle corrections connect the energy of the initiating parton with the energy of the true particle jet. These corrections account for out-of-cone effects and contributions from the underlying event and can only be derived from Monte-Carlo simulation.

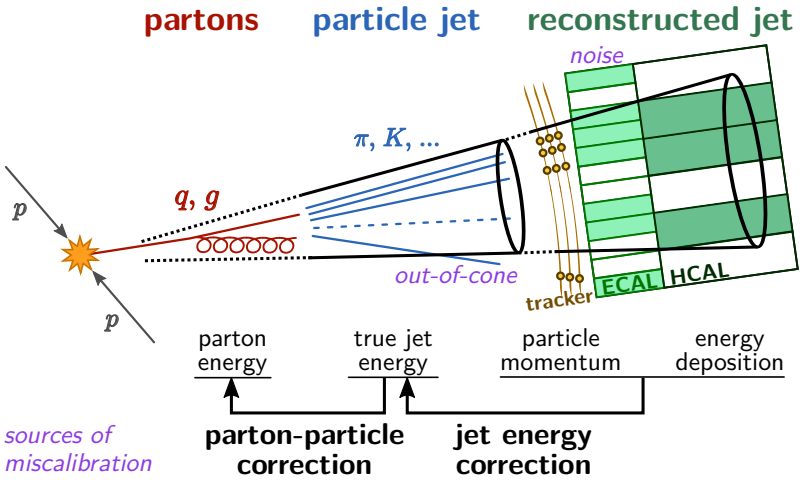


Figure 4.1: Schematic overview of jet measurement and correction.

4.1 Jet energy corrections in CMS

Within the CMS Collaboration, jets are corrected in a factorised approach. Three mandatory and several optional correction levels are applied to the raw jet momentum as provided by the object reconstruction and jet clustering. Each level is dedicated to the correction of a particular systematic bias on the jet energy scale.

The correction procedure starts with simulation-based corrections and is completed by the application of data-driven correction factors. This approach allows to study the underlying reasons for systematic deviations in the jet energy measurement in great detail using all the information only present in Monte-Carlo simulation. At the same time, it is ensured that the final correction factor is determined with the precision and robustness of data-driven methods and does not depend on assumptions about the parton shower and hadronization models.

Since the start-up of the LHC, the definitions of the correction levels has changed. The current scheme, used for recent publications is illustrated in Figure 4.2. The transverse momentum of the corrected jet $p_{\mathrm{T}}^{\mathrm{corr}}$ is calculated as

$$p_{\mathrm{T}}^{\mathrm{corr}} = c_3(\eta, p_{\mathrm{T}}^2) \cdot c_2(\eta, p_{\mathrm{T}}^1) \cdot c_1(\eta, \rho, A_j, p_{\mathrm{T}}^0) \cdot p_{\mathrm{T}}^0 \tag{4.1}$$

where $c_{1,2,3}$ are the correction factors of the three mandatory correction levels. Further optional levels can be added accordingly. p_{T}^0 is the transverse momentum of the uncorrected jet while $p_{\mathrm{T}}^{1,2}$ denote the transverse momentum after the corresponding correction level The correction factors are derived separately for 7 TeV and 8 TeV data.

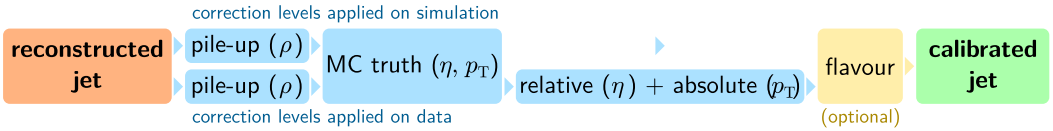


Figure 4.2: Factorised approach to jet energy corrections in CMS. Corrections for pile-up activity are followed by a Monte-Carlo based level. Data-driven corrections correct for the observed differences between data and Monte-Carlo. The flavour correction is shown as an example for optional correction levels.

Pile-up corrections

In a first step, the transverse momentum of the jet is corrected for the effects of pile-up. Particles emerging from additional pp interactions in the same event are clustered on the same footing as the hadronization products of the original parton from the hard interaction. The amount of energy added by this effect has to be subtracted by the pile-up corrections. Another source of additional event activity is out-of-time pile-up. Using two time slices (cf. Section 3.1.2) for the event reconstruction in 2012 instead of four reduces the overlap with the immediately preceding and following events. This reduces the amount of out-of-time pile-up to a negligible level, late out-of-time pile-up is even non-existent for all practical purposes. Furthermore, electronic noise can bias the jet energy measurement. However, the noise level is found to be negligible and does not require corrections. There are two methods for pile-up corrections: The *Average Offset* method and the *Jet Area* method.

The *Average Offset* method corrects for the pile-up contamination on an event-by-event basis. In a very good approximation, the average amount of energy added to the jet resulting from pile-up interactions is proportional to the number of primary vertices n_{PV} in the event. This is reflected by the correction factor $c_{1,\mathrm{AO}}$ of the first correction level in the Average Offset method in Eq. (4.2).

$$c_{1,\mathrm{AO}}(\eta, n_{\mathrm{PV}}; p_{\mathrm{T}}^{0}) = 1 - c_{\mathrm{AO}}(\eta) \cdot \frac{n_{\mathrm{PV}} - 1}{p_{\mathrm{T}}^{0}} \tag{4.2}$$

$$p_{\mathrm{T}}^{1,\mathrm{AO}} = c_{1,\mathrm{AO}}(\eta, n_{\mathrm{PV}}; p_{\mathrm{T}}^{0}) \cdot p_{\mathrm{T}}^{0} = p_{\mathrm{T}}^{0} - c_{\mathrm{AO}}(\eta) \cdot (n_{\mathrm{PV}} - 1) \tag{4.3}$$

The slope of this linear correction is determined separately for different η ranges. It also differs for data and simulation due to inconsistencies in the pile-up modelling and vertex reconstruction efficiencies. The p_{T} dependence is only a matter of notation as a multiplicative factor. Differences in the catchment area of individual jets are neglected in this approach.

The *Jet Area* method [77] is a more recent alternative approach that is able to determine correction factors on a jet-by-jet basis by taking into account the pile-up density ρ in the event and the jet area A_j.

Using the Jet Area method, the raw jet is corrected by a factor proportional to the pile-up density in the event and the area of the jet.

$$c_{1,\mathrm{JA}}(\eta, \rho, A_j; p_\mathrm{T}^0) = 1 - c_\mathrm{JA}(\eta) \cdot \frac{(\rho - \langle\rho_\mathrm{UE}\rangle) \cdot A}{p_\mathrm{T}^0} \tag{4.4}$$

$$p_\mathrm{T}^{1,\mathrm{JA}} = c_{1,\mathrm{JA}}(\eta, \rho, A_j; p_\mathrm{T}^0) \cdot p_\mathrm{T}^0 = p_\mathrm{T}^0 - c_\mathrm{JA}(\eta) \cdot (\rho - \langle\rho_\mathrm{UE}\rangle) \cdot A \tag{4.5}$$

The η dependence of the Jet Area method is taken from the η dependence of the Average Offset method using a global conversion factor. To avoid drastic over-corrections, the correction factor $c_{1,\mathrm{JA}}$ is not allowed to get negative. During the development phase of the Jet Area method, the correction factors were continuously cross-checked using $Z(\rightarrow \mu\mu)$ + jet events in the scope of this thesis.

Monte-Carlo based corrections

All other known effects on the jet energy scale are corrected for by a simulation-based correction level. The momentum of a reconstructed jet can differ from the one of the generated particle jet for several reasons: Additional out-of-cone effects can reduce the jet momentum. Detector effects, like a non-linear calorimeter response, reconstruction efficiencies which are particularly low for low energetic particles as well as known and simulated transparency losses in calorimeter cells, affect the jet momentum measurement. The jet response in simulated events is determined as

$$R = \frac{p_\mathrm{T}^{\mathrm{j,\ reco}}}{p_\mathrm{T}^{\mathrm{j,\ gen}}}$$

in bins of p_T^1 and η where the former is the transverse momentum already corrected for pile-up effects. The inverse of the determined jet response is applied as correction factor.

$$c_2(\eta, p_\mathrm{T}^1) = \frac{1}{\langle R\rangle_{p_\mathrm{T}^1, \eta}} \tag{4.6}$$

$$p_\mathrm{T}^2 = c_2(\eta, p_\mathrm{T}^1) \cdot p_\mathrm{T}^1 \tag{4.7}$$

A flavour mixture of typical LHC QCD events is assumed for this correction level.

This level completes the corrections on simulated events and the closure test in Figure 4.4 shows that both the pile-up and the p_T dependence are removed, leading to a corrected response close to unity. On data, all of the above effects can have unpredicted components. Therefore, additional data-driven corrections are needed.

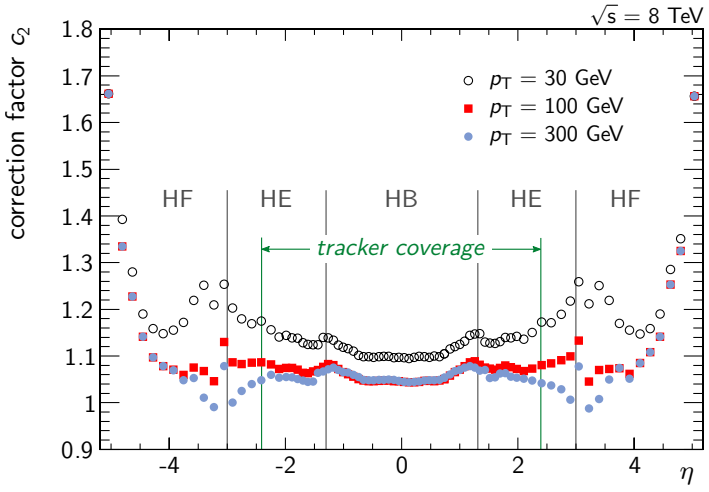


Figure 4.3: Correction factors of the Monte-Carlo based corrections over η for three different transverse jet momenta. Low energetic jets tend to be wider and require a larger correction factor. The correction factors are flat in the homogeneous central region of the detector. The transitions between detector components show discontinuities. Larger corrections factors are needed in the forward detector (derived from [78]).

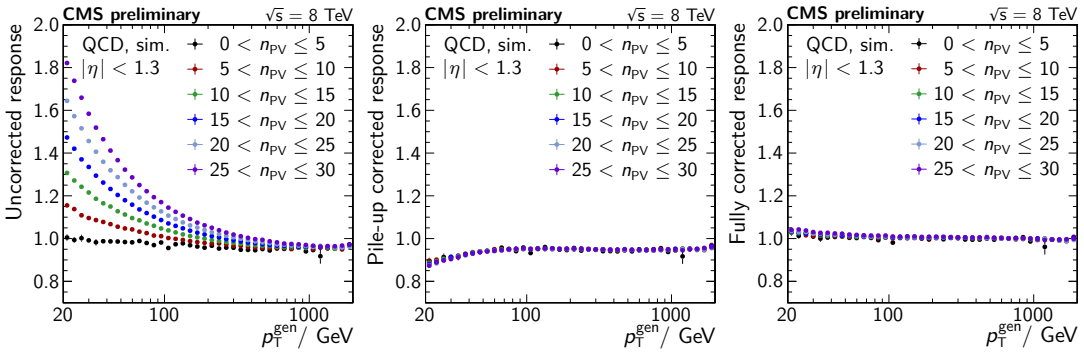


Figure 4.4: Closure test of the MC based corrections. Uncorrected jets show a strong dependence on the number of reconstructed vertices (*left*). The pile-up corrections successfully correct for this (*centre*) and the MC based corrections yield a jet response at unity over the whole p_T range [79].

Relative residual corrections

Particularly, the inhomogeneous parts of the detector outside the central barrel region are prone to the previously listed effects with an a priori unknown component. Therefore, jets in the reach of the endcaps or the forward calorimeters are calibrated relatively to jets in the central detector, before the calibration of the absolute scale. Dijet events with one jet in the barrel (tag) and the balancing jet in a higher pseudorapidity range (probe) are used. This method profits from the high cross section of dijet events at the LHC. Considering the comparably low resolution of both the tag and the probe jet, the direct p_{T} balance between the two suffers from a significant resolution bias. Calculating the jet response using the asymmetry $\mathcal{A}$ and averaging over a large number of events removes this bias.

$$R_{\mathrm{rel}}(\eta_{\mathrm{probe}}, p_{\mathrm{T}}^{\mathrm{ave}}) = \frac{1 + \langle \mathcal{A} \rangle}{1 - \langle \mathcal{A} \rangle} \tag{4.8}$$

with

$$\mathcal{A} = \frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}}} \tag{4.9}$$

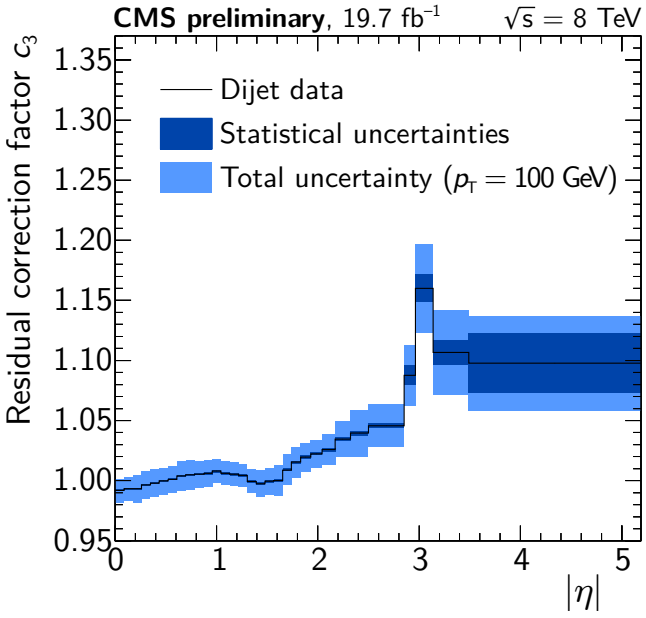


Figure 4.5: Correction factors for the relative residual corrections [78].

Absolute residual corrections

While so far, the jet energy scale is only corrected relatively to jets in the homogeneous centre of the detector ($|\eta| < 1.3$), the calibration of the absolute scale is missing. This is done by the last mandatory correction level, the absolute residual corrections which are described in greater detail in Section 4.2.

Flavour corrections

Jets emerging from partons of different flavour differ in their jet response by up to 3 %. Particularly gluon jets have a reduced jet response because they are wider and more prone to out-of-cone effects. Correction factors are derived for each flavour and for common flavour mixtures in QCD or $Z/\gamma + jets$ events. This simulation based correction level can be completed by a data-driven counterpart. The experimental techniques are explained in Section 4.10.

4.2 Correction of the absolute jet energy scale using Z($\rightarrow \mu\mu$)+jet events

As the true set of individual particles leading to the jet measurement in the detector is not accessible directly, a reference is needed that is physically closely related. As the proton bunches are located on the z-axis, the initial total transverse momentum in a collision is zero. It follows from the energy and momentum conservation that if one boson and one parton are produced in the hard interaction, the sum of the two momentum vectors vanishes in the transverse plane. In the commonly used coordinates, this is equivalent to the statement that the absolute value of both transverse momenta is equal and that the separation in the azimuthal angle is π, called a back-to-back topology. Hence, by measuring the energy and momentum of this boson, the transverse energy of the parton can be measured in collision data. Three event topologies are exploited by the CMS Collaboration in which a single jet is balanced by a reference object: $\gamma + \text{jet}$, $Z(\rightarrow ee) + \text{jet}$ and $Z(\rightarrow \mu\mu) + \text{jet}$. In all cases, the p_{T} of the reference object can be measured with significantly higher precision than the p_{T} of the jet.

While the Z boson has been subject of fundamental research beginning with its discovery in 1983 and leading to the precision measurements at the LEP collider, it can now serve as a reference object for probing jet physics experimentally. The clear signature of $Z(\rightarrow \mu\mu) + \text{jet}$ events is particularly suitable for a nearly background free study of those events and for the use as the reference object for the data-driven jet energy calibration.

Because $\gamma + \text{jet}$ production has a higher cross section at the LHC, it is possible to reach higher transverse momentum ranges with balancing jets and photons. However,

the precision of the ECAL momentum measurement is limited compared to muon based Z measurements (0.2%). The latter are based on the combined measurements of the tracking detector and the muon system, do not depend on calorimetric energy measurements and are only statistically limited. With the increased total luminosity of 2012 data, the calibration using the $Z(\rightarrow \mu\mu) + \text{jet}$ topology became the most reliable one.

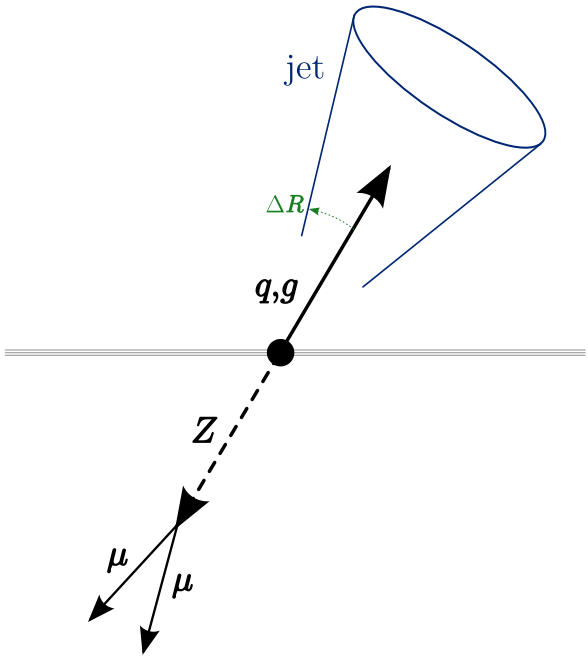


Figure 4.6: Sketch of the $Z(\rightarrow \mu\mu) + \text{jet}$ topology. A Z boson is balanced with a parton in the hard process of the interaction. While the Z boson decays into a pair of oppositely charged muons, the parton hadronises into a jet which can be clustered from the particles measured in the detector. ΔR illustrates the role of the jet size parameter.

The basic method for data-driven jet energy corrections using $Z(\rightarrow \mu\mu) + \text{jet}$ events has been laid out during the start-up phase of the LHC [80]. Very first comparisons [57, 81] to early 2010 data (50 events) showed the general feasibility of the method in a relatively low pile-up environment. Considering this low number of balanced $Z(\rightarrow \mu\mu) + \text{jet}$ events, large statistical uncertainties dominated the measurement and left a multitude of smaller systematic effects on the jet energy scale hidden. The performance of jet energy calibration in CMS and the first collaboration-wide set of correction factors were then published in [82]. The continuous studies of our group in this field have also been documented in [83, 84] and will be presented in this chapter.

During the 2011 and 2012 data-taking periods, the average number of pile-up

vertices in an event increased from 5 to 20. Methods to cope with the increased pile-up have been developed and lead to an improved understanding of pile-up influences and the detector in general. Further examples of previously unobserved effects are radiation damages and the subsequent loss of transparency in the calorimeters. In addition, the method was extended to cover all pseudorapidity ranges and the flavour dependence was measured.

4.3 Jet response measurement

The reason to use $Z(\rightarrow \mu\mu) + \text{jet}$ events in a data-driven method is based on the assumption that the transverse momentum of the jet's initial parton is equal to the transverse momentum of the Z boson as a consequence of the energy and momentum conservation. Therefore, the jet response is defined as the measured transverse momentum with respect to the transverse momentum of the reference object.

There are two complementary methods to determine the jet response called the p_T balance and the MPF method. The p_T balance response is defined as the direct transverse momentum ratio of the leading jet and the reference object

$$R_\mathrm{bal} = \frac{p_\mathrm{T}^\mathrm{jet1}}{p_\mathrm{T}^\mathrm{Z}} \,.$$

A possible bias is expected from final state radiation (FSR). In this case, the parton momentum is shared between the leading jet and the jet emerging from that radiation. This results in a lower measured jet response. The extrapolation of the jet response to vanishing secondary jet transverse momentum in Section 4.8.3 is designed to remove this bias.

The Missing E_T Projection Fraction (MPF) method is an alternative approach, first used at the Tevatron collider [85]. Exploiting the fact that $Z(\rightarrow \mu\mu) + \text{jet}$ events have no intrinsic missing transverse energy, the MPF method assumes a perfect balancing between the true transverse momenta of the Z boson and the recoiled jet. In measured events, each object has its response R and every mismeasurement of the transverse momentum finds its expression in the missing transverse energy. This is described by

$$\vec{p}_\mathrm{T}^{\,Z,\,\mathrm{gen}} + \vec{p}_\mathrm{T}^{\,\mathrm{jet1},\,\mathrm{gen}} = \vec{0} \tag{4.10}$$

$$R_Z \cdot \vec{p}_\mathrm{T}^{\,Z} + R_\mathrm{jet} \cdot \vec{p}_\mathrm{T}^{\,\mathrm{jet1}} = \vec{E}_\mathrm{T}^\mathrm{miss} \,. \tag{4.11}$$

This system of equations can be solved for the response of the hadronic recoil R_jet which is then identified with the jet response in a $\mathrm{Z} + 1$ jet event. Assuming the response of the Z boson to be one, this yields the definition of the MPF response:

$$R_\mathrm{MPF} = 1 + \frac{\vec{E}_\mathrm{T}^\mathrm{miss} \cdot \vec{p}_\mathrm{T}^\mathrm{Z}}{|p_\mathrm{T}^\mathrm{Z}|^2}$$

As it considers the whole recoil, the MPF method is less prone to the radiation bias or out-of-cone effects in general. However, it involves all particles in the event and is therefore more sensitive to any kind of mismodelling and acceptance effects also outside the homogeneous central detector region. While jets from residual pile-up bias the MPF response in individual events, these contributions cancel on average, making the MPF method a suitable complementary approach for the measurement of the jet energy scale. The methods for pile-up mitigation, explained in Section 4.4, prove to work and both methods provide comparably small uncertainties.

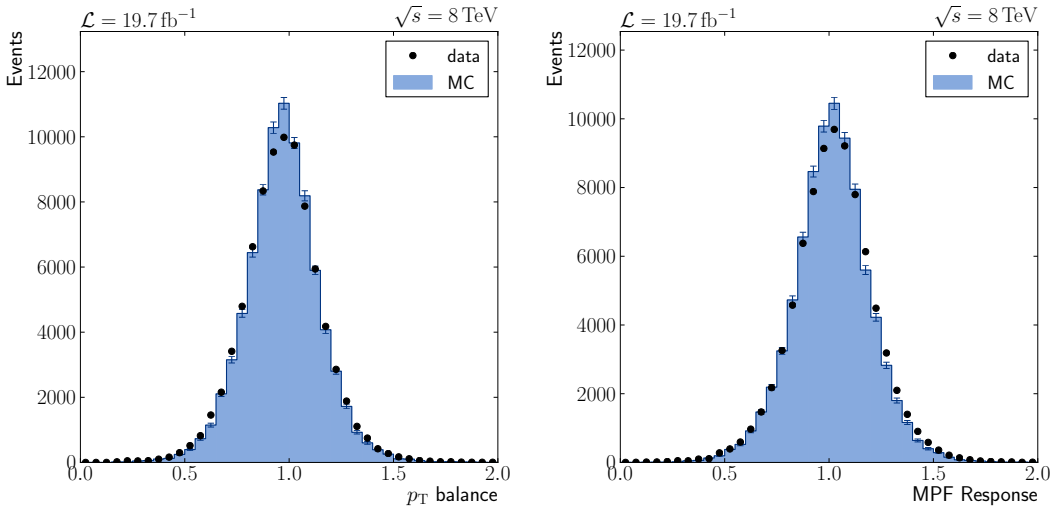


Figure 4.7: The two complementary response definitions p_T balance (*left*) and MPF (*right*) result in comparable distributions. In both cases, the resolution in data is slightly worse than in simulated events. The mean of these distributions is taken to estimate the jet response.

4.4 Pile-up mitigation

The high number of pile-up interactions in 2012 data reduced the number of usable events dramatically. While the intrinsic balance between jet and Z boson is not affected by pile-up interactions, the experimental measurement is complicated for several reasons: In rare cases, the transverse momentum of a jet emerging from pile-up can exceed the p_T of the jet from the hard interaction, leading to arbitrary but mostly low values for the measured jet response. As pile-up jets carry a transverse momentum of typically a few GeV, this is only problematic in the low p_T range. Much more frequent and much more difficult to handle are pile-up jets being the second

leading jet in an event. Events containing additional jets have to be rejected as this can be a result of final state radiation. In this case, they carry an unknown fraction of the initial parton energy. Further jets from pile-up interactions do not affect the jet response measurement.

Furthermore, increasing pile-up leads to a decreased muon isolation efficiency and to large tails in the missing energy distribution. All those effects require a thorough understanding of pile-up effects and techniques for their mitigation to allow for a jet energy calibration at the sub-percent level.

There are three complementary methods for pile-up mitigation employed in this analysis, intervening at different levels of the event reconstruction. Each method can be used on its own and in combination.

4.4.1 Charged hadron subtraction

The first countermeasure against pile-up intervenes at the level of Particle Flow candidates. All candidates, of which the track is compatible with a pile-up vertex, are removed from the set of input objects for the jet clustering. Tracks associated with the main vertex are preserved as well as tracks that can not be associated to any vertex. The latter was shown to be important to avoid over-subtraction of pile-up particles. The track association is illustrated in Figure 4.8 where charged hadrons drawn as solid red arrows are removed from the event prior to the jet clustering. This method results in Charged Hadron Subtraction (CHS) jets which can be used as an alternative to regular jets. Being a track-based method, this only works for charged particles and exclusively within the tracker acceptance. The analysis presented in this thesis was one of the first using this method, confirming its suitability for a wide range of pile-up affected analysis and paving the way for its use in the $H \to \tau\tau$ analysis in Chapter 5.

In combination with the jet identification criteria this not only removes the pile-up energy contribution to jets, but it can also remove entire jets. If all charged hadrons clustered into a jet are removed by CHS and only neutral hadrons are remaining, the jet does not fulfil the identification requirement of containing at least one charged component. This method alone removes about 70 % of the pile-up particles from jets and 30 % of the pile-up jets, thus reducing the pile-up biases to be removed by the following methods which are applied after jet clustering.

4.4.2 Pile-up jet identification

A multivariate method is used to identify and reject jets emerging from pile-up vertices [86]. It combines the discriminating information of these variables:

As done by the Charged Hadron Subtraction, all particle tracks are associated with the primary vertex they originate from. With respect to the current jet, they can be

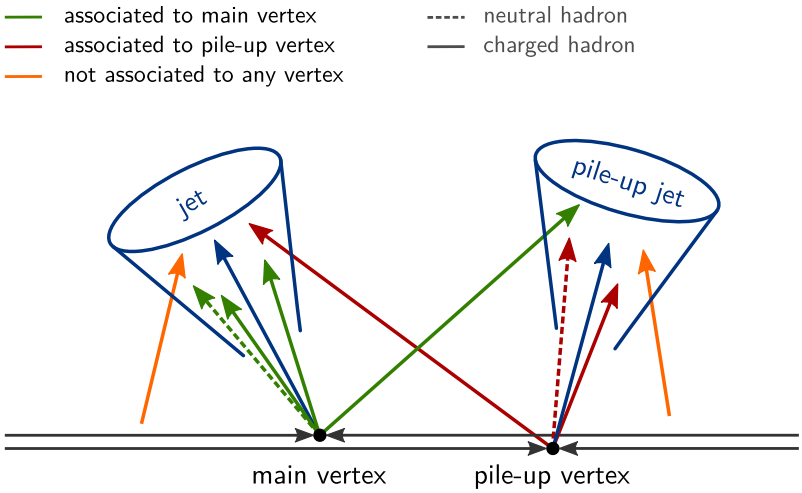


Figure 4.8: Association of particle tracks with primary vertices. Particles clustered into a jet from the hard interaction (jet) can emerge from the main vertex or from pile-up vertices. The Charged Hadron Subtraction (CHS) removes hadrons from pile-up vertices (red, dashed) prior to jet clustering. The pile-up jet identification removes an entire jet if it has a dominating contribution from pile-up hadrons.

sorted into three categories: the set of tracks coming from the same vertex as the jet M, tracks from other vertices P and tracks that can not unambiguously associated to one of the vertices N. This is illustrated in Figure 4.8. β-variables are calculated as the p_{T}-weighted fraction of tracks in this category, such that $\beta_M + \beta_P + \beta_N = 1$.

$$\beta = \beta_M = \frac{\sum_{i \in M} p_{\mathrm{T},i}}{\sum_i p_{\mathrm{T},i}} \qquad \beta^* = \beta_P = \frac{\sum_{i \in P} p_{\mathrm{T},i}}{\sum_i p_{\mathrm{T},i}} \tag{4.12}$$

d_z is defined as the distance between the track and the jet along the z-axis.

$$d_z = |z_{\mathrm{track}} - z_{\mathrm{jet}}|$$

These variables β, β^*, d_z and the number of primary vertices n_{PV} are able to discriminate between pile-up jets and jets from the hard interaction within the reach of the tracking detector.

Outside the tracker coverage, the identification of pile-up jets has to be based on shape variables. The mean width of a jet can be defined as

$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 p_{\mathrm{T},i}^2}{\sum_i p_{\mathrm{T},i}^2}$$

with $\Delta R_i = \Delta R_{\mathrm{jet},i}$ for all jet constituents i. In addition, the momentum sum in

rings around the central axis of the jet can be defined as

$$R_k = \frac{1}{p_{\mathrm{T}}^{\mathrm{jet}}} \sum_{i \in A_k} p_{\mathrm{T},i} \quad \text{with} \quad A_k = \{i \in \text{tracks} \mid k < \Delta R_i < k + \Delta k\}$$

with $k \in \{0.1, 0.2, 0.3, 0.4\}$ and $\Delta k = 0.1$. These values are appropriate for Anti-k_{T} 0.5 jets as the radius is limited to $R \leq 0.5$. It provides a discrete profile of the radial jet momentum distribution. While gluon jets also tend to be wide, they nonetheless have a harder centre on average. Further variables are the charged and neutral constituent multiplicities $n_{\pm}$ and n_0 and the fragmentation function

$$p_{\mathrm{T}}^{\mathrm{D}} = \frac{\sqrt{\sum_i p_{\mathrm{T},i}^2}}{\sum_i p_{\mathrm{T},i}} \,. \tag{4.13}$$

The discriminator between pile-up and non-pile-up jets is trained and evaluated separately for different η and p_{T} bins. The residual pile-up energy which is still clustered into a jet, is then subtracted by the pile-up jet corrections as explained in Section 4.1.

4.5 Data samples

The data collected in 2011 and 2012 comprises $L = 4.9\,\mathrm{fb}^{-1}$ and $L = 19.7\,\mathrm{fb}^{-1}$, respectively. The events have been preselected by a double muon trigger that requires one muon with $p_{\mathrm{T}}{}^{\mu_1} > 17\,\mathrm{GeV}$ and a second muon with $p_{\mathrm{T}}^{\mu_2} > 8\,\mathrm{GeV}$. Due to the large muon production cross section, the muon triggers with lower thresholds have to be prescaled, i. e. only a fixed fraction of those triggered events are recorded. For 2011 data, triggers with lower thresholds are chosen if an unprescaled trigger is available for this event (Appendix A.2.1) to allow for looser cuts on the muon momentum. The trigger turn-on significantly reduces the number of triggered dimuon events at the threshold but a constant efficiency of 96.5 % [87] is reached for a single muon above the analysis cut of 20 GeV.

The data results are compared to simulated events produced by the following Monte-Carlo generators. The NLO description of the Drell-Yan process by POWHEG [88] serves as a cross check for the default sample generated by MADGRAPH [89] which generates the Drell-Yan with any number of jets at NLO. In both cases, PYTHIA [90] is used for the parton shower and hadronization. For the study of fragmentation and hadronization effects as well as jet flavour studies, HERWIG++ [91] is used.

All events are passed to the CMS event simulation and event reconstruction. This includes the admixture of recorded pile-up events from the minimum bias data, following a fixed distribution of primary vertices.

The NNLO cross-sections [92] of Drell-Yan Z boson production accompanied by any number of additional jets are estimated using FEWZ [93, 94]. The total luminosity of recorded CMS data is determined from the integrated luminosity of all certified LHC runs and luminosity-sections [52] (Appendix A.1.1).

The efficiencies due to the detector acceptance, generator internal cuts, muon and jet isolation and validation have been taken into account as well as trigger efficiencies and event filter efficiencies,

$$\varepsilon = \varepsilon_{\text{acceptance}} \cdot \varepsilon_{\text{generator}} \cdot \varepsilon_{\text{isolation}} \cdot \varepsilon_{\text{validation}} \cdot \varepsilon_{\text{trigger}} \cdot \varepsilon_{\text{filters}} \,.$$

The trigger efficiency

$$\begin{aligned}\varepsilon_{\text{trigger}}(\eta, p_{\text{T}}) = {} & \varepsilon_8(\eta^{\mu^+}, p_{\text{T}}^{\mu^+}) \cdot \varepsilon_{17}(\eta^{\mu^-}, p_{\text{T}}^{\mu^-}) + \varepsilon_{17}(\eta^{\mu^+}, p_{\text{T}}^{\mu^+}) \cdot \varepsilon_8(\eta^{\mu^-}, p_{\text{T}}^{\mu^-}) \\ & - \varepsilon_{17}(\eta^{\mu^+}, p_{\text{T}}^{\mu^+}) \cdot \varepsilon_{17}(\eta^{\mu^-}, p_{\text{T}}^{\mu^-})\end{aligned}$$

is calculated from the single muon efficiencies ε_8 for the 8 GeV leg and ε_{17} for the 17 GeV leg of the double muon trigger. The trigger efficiencies are between 94 % and 98 % for muons with $p_{\text{T}} > 20\,\text{GeV}$. As the final result does not depend on absolute event counts but only on ratios of quantities within an event, the overall normalisation is fixed to the number of measured events in data.

4.5.1 Pile-up reweighting

The pile-up distribution in simulated samples is only a rough estimate and forward projection of the distribution during data taking. To get a realistic pile-up admixture the simulated events have to be weighted according to their mean number of pile-up interactions

$$n_{\text{PU}}^{\text{mean}} = \langle n_{\text{PU}} \rangle = L \cdot \sigma_{\text{minbias}}$$

independently for each run and lumi-section using the CMS luminosity measurement. The minimum bias cross section is estimated to be $\sigma_{\text{minbias}} = (69.3 \pm 1.8)\,\text{mb}$. In the $Z(\to \mu\mu) + \text{jet}$ sample, the value $\sigma_{\text{minbias}} = 68.5\,\text{mb}$ was used which yields the best agreement in the distribution of reconstructed primary vertices. Inconsistencies between the luminosity measurement in the HF detector during data-taking and the later pixel-luminosity measurement only lead to negligible uncertainties for reweighting purposes (cf. Section 3.1.1). However, unreliable measurements in events with more than 30 reconstructed primary vertices have to be discarded.

Figure 4.9a shows the distribution of the mean number of pile-up vertices in simulation after reweighting to the one estimated in data which agrees with the estimated data distribution by construction. In Figure 4.9b, the distribution of the number of reconstructed primary vertices n_{PV} is presented after the reweighting and compared to the measurement in data.

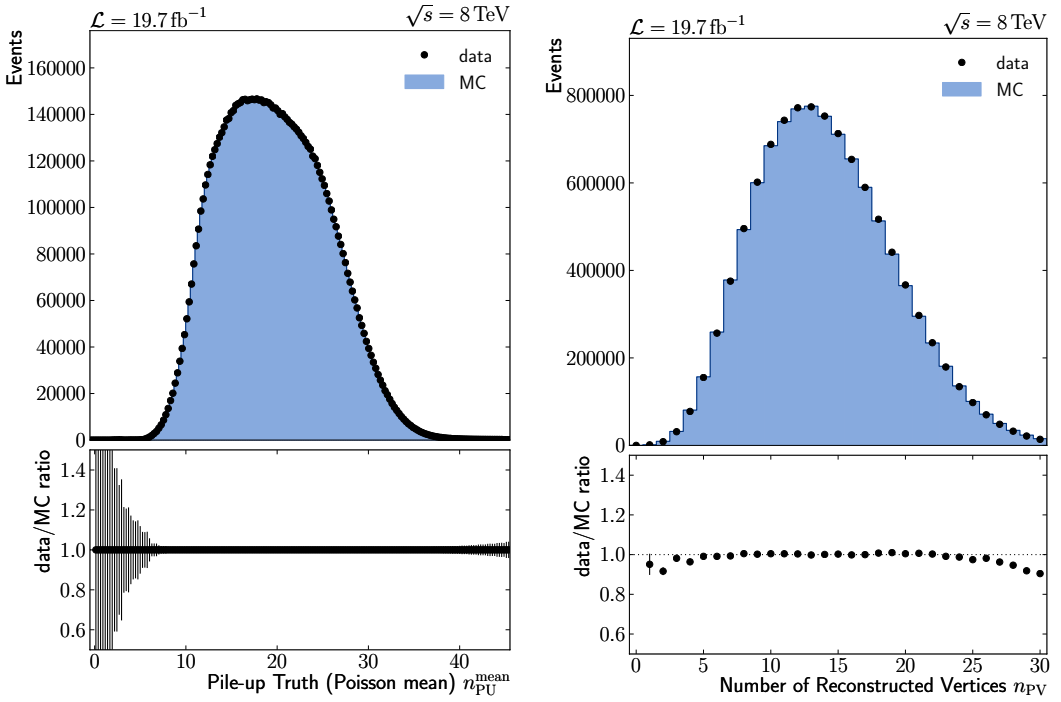


Figure 4.9: The distribution of the mean number of pile-up vertices (*left*) per lumi-section as mixed into the simulated samples after reweighting. After reweighting this distribution matches the estimated distribution in data corresponding to a minimum bias cross section of $\sigma_{\mathrm{minbias}} = 68.5\,\mathrm{mb}$. The distribution of reconstructed primary vertices (*right*) after reweighting agrees with the observation in data events.

The uncertainty on the measured jet response arising from the uncertainty on the lumi measurement, the minimum bias cross section, out-of-time pile-up and the reconstruction efficiency for primary vertices was found to be negligible.

4.6 Correction of reconstructed objects

Jets in the event are corrected by applying all correction levels described in Section 4.1, successively. All other involved physical objects require dedicated corrections which are applied as described in the following sections.

4.6.1 Missing transverse energy corrections

The different methods for jet corrections have their counterparts in the field of corrections of the missing transverse energy, as any mismeasurement of the hadronic activity automatically entails a bias in the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the same amount. Therefore, the missing energy has to be recalculated from the corrected objects.

Type-$\emptyset$ corrections

When the CHS removes charged hadrons from the event, that originate from pile-up vertices, this also requires to remove those charged hadrons from the missing energy calculation. By doing so, the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is improved and the dependence on pile-up is reduced. This correction step is called Type-$\emptyset$ correction.

Type-I corrections

After the application of jet energy corrections, the momentum of the jet is not equal to the four-vector sum of its constituents. Consequently, the momentum conservation is violated if the whole event is considered. To restore the event consistency, the jet energy correction has to be propagated to the missing transverse energy. To this end, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated after applying the jet energy corrections to all jets in the event that have a minimal transverse momentum of 10 GeV. This recalibration of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is called Type-I correction.

MET-ϕ corrections

Misalignment of detector components relative to each other and relative to the beam spot results in a modulation of the distribution of the azimuthal angle of the missing energy ϕ_{miss}. This modulation is different in collision data and simulated events. This misalignment was never fully corrected by re-reconstructions of the event content. For the missing transverse energy, correction terms have been determined to account for this effect [95].

$$E_x^{\mathrm{miss}\prime} = E_x^{\mathrm{miss}} - c_x^0 - c_x^1 \cdot n_{\mathrm{PV}} \tag{4.14}$$

$$E_y^{\mathrm{miss}\prime} = E_y^{\mathrm{miss}} - c_y^0 - c_y^1 \cdot n_{\mathrm{PV}} \tag{4.15}$$

As shown in Figure 4.10, this correction is able to significantly reduce the modulation. A complete removal method would have to be applied during the event reconstruction. This correction is not applied by default as there is no equivalent counterpart for all other objects in the event. Hence, this correction violates the event consistency, required for the two methods of measuring the jet response to be comparable.

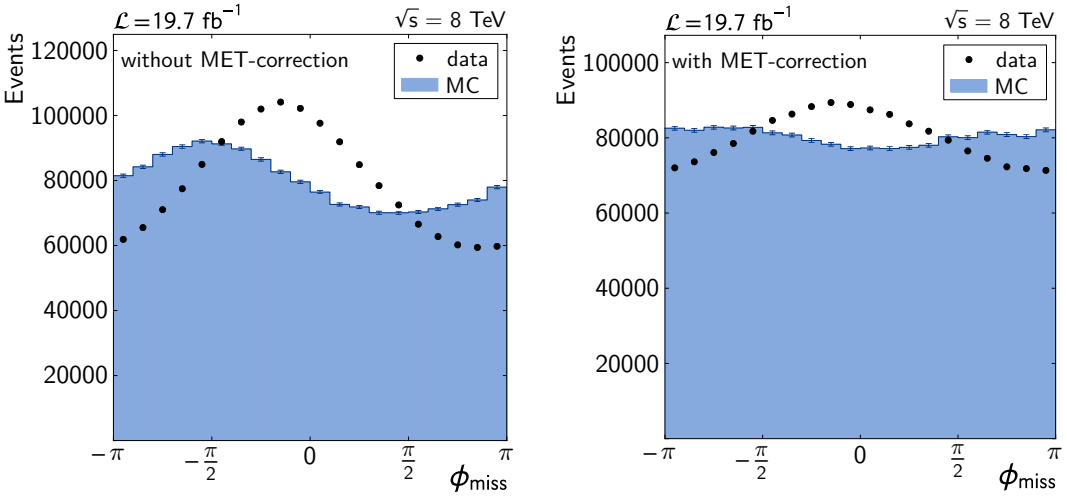


Figure 4.10: Correction of the ϕ-modulation of the missing transverse energy due to a misalignment of detector components. The strong modulation of $E_{\text{T}}^{\text{miss}}$ before the correction (*left*) is partially removed by the MET-ϕ corrections (*right*). However, the recommended corrections can not fully compensate for this effect – in particular the modulation in data.

4.6.2 Muon scale corrections

While muons are reconstructed very precisely by the combination of tracking information from the inner tracker and the muon chambers, this accuracy can be further improved by data-driven corrections. CMS provides muon scale corrections derived from fits around the Z boson mass peak to improve the scale of the muon momentum measurement and to smear the resolution in simulated dimuon events to the resolution found in data. This correction is applied separately for data and simulated events.

4.7 Event selection

The selection of $Z(\rightarrow \mu\mu) + \text{jet}$ events is based on the recommendations of the CMS Vector Boson Task Force [82]. The selection criteria have been slightly adapted to calibration purposes and are synchronized between the different channels of data-driven corrections.

4.7.1 Event criteria

The events are preselected using runs and lumi-sections containing physical interactions as recommended by the CMS Physics Validation Group (Appendix A.1.1).

MET event filters Events with clear signatures of mismeasurement of any kind have to be rejected. This is particular true if the missing energy is evaluated in the analysis because errors in the reconstruction of any object in the event has an impact on the reconstructed missing transverse energy, which in turn is used for the calculation of the MPF response.

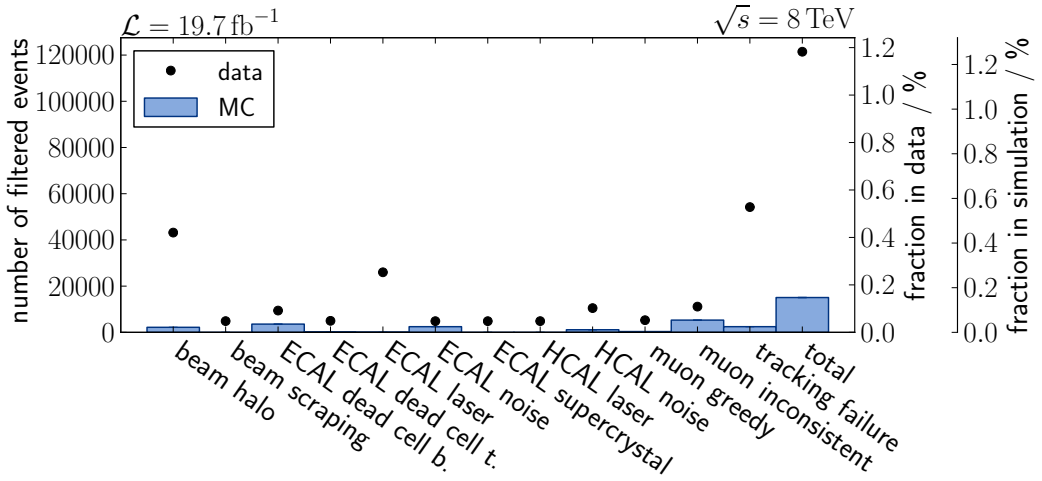


Figure 4.11: Event filters removing events in case of severe mismeasurements of the reconstructed missing transverse energy. Mismeasurements can arise from all components of the detector.

To this end, a sequence of event filters are applied (Fig. 4.11): The beam halo and beam scraping filters veto events with additional unwanted beam interactions besides the proton collisions. These interactions happen in almost every 200th event and only in recorded data events as such events are not simulated. Events in simulation rejected by these filters show a high scalar sum of transverse momenta in the event and are thus errors of the first kind. However, those cases are rare. Furthermore, filters are applied for events involving dead cells, defect super-crystals, noise and laser measurements in the electromagnetic calorimeter, as well as for events with noise or signals from laser calibration in the hadronic calorimeter. Dead cells in the ECAL are partly simulated but not to the full extend. The use of the laser calibration device of the ECAL during data-taking renders almost 26000 events useless. In addition, events in which the reconstructed calorimeter energy of a muon is larger than the muon momentum are rejected by the so-called muon greedy filter. Events containing an inconsistency in measurement of the muon transverse momentum between the tracker and the muon chambers of more than 10 % are discarded as well. The same holds for events with failures during track reconstruction. All these filters remove events with extremely large tails in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution which were found in dijet

events. $Z(\to \mu\mu) + \text{jet}$ events are less affected. The fraction of filtered events in total is 1.18 % for data events whereas only 0.16 % of the events in simulation are rejected by the $E_{\mathrm{T}}^{\mathrm{miss}}$ filters. The impact on the final result has been studied and found to be negligible.

4.7.2 Muon selection and Z boson reconstruction

The Z boson selection follows the recommendations of the CMS Vector Boson Task Force for Z bosons decaying into a pair of muons. The minimum requirement on the transverse muon momentum lies significantly above the trigger thresholds of 17 GeV and 8 GeV to avoid effects from the trigger turn on. The muon system acceptance restricts the reach in pseudorapidity to $|\eta| < 2.4$. The selection requirement is slightly tighter to avoid border effects. The following muon selection is applied in addition to the default CMS muon identification (A.2.2):

$$\begin{aligned} |\eta_\mu| &< 2.3\,, \\ p_{\mathrm{T}}^{\mu} &> 20\,\mathrm{GeV}\,. \end{aligned}$$

Furthermore, the muons have to be isolated to avoid non-prompt muons from jets

$$\sum_{\text{tracks}} p_{\mathrm{T}} < 3\,\mathrm{GeV} \qquad \text{within } \Delta R < 0.3\,.$$

The Z boson is reconstructed from the leading two opposite charged muons in the event. The invariant mass of the muon pair is required to be closer than 20 GeV to the Particle Data Group (PDG) value [1] of the Z boson mass and there is a minimum requirement on the transverse momentum as jets emerging from softer partons can not be calibrated using $Z(\to \mu\mu) + \text{jet}$ events because the relative influence of low energetic particles is too large.

$$|m_{\mu\mu} - m_{\mathrm{Z}}^{\mathrm{PDG}}| < 20\,\mathrm{GeV} \tag{4.16}$$

$$p_{\mathrm{T}}^{\mathrm{Z}} > 30\,\mathrm{GeV} \tag{4.17}$$

4.7.3 Jet selection

Jets are selected according to the CMS jet identification criteria. On average, 20 jets per event are rejected by this requirement when using CHS jets. For calibration purposes, the leading jet j_1 has to be reconstructed within the homogeneous central part of the detector and a minimal transverse momentum is required to minimize the influence of residual noise and pile-up. All other jets j have a lower but well-defined

requirement on their transverse momentum:

$$|\eta_{\mathrm{j1}}| < 1.3 \tag{4.18}$$

$$p_{\mathrm{T}}^{\mathrm{j1}} > 12\,\mathrm{GeV} \tag{4.19}$$

$$p_{\mathrm{T}}^{\mathrm{j}} > 5\,\mathrm{GeV}\,. \tag{4.20}$$

4.7.4 Z+jet topology cuts

The balanced $Z(\rightarrow \mu\mu)$ + jet topology is ensured by two additional cuts: The reconstructed jet has to be back-to-back to the Z boson and the contribution of additional jets is required to be smaller than 20 % of the Z boson p_{T} as illustrated in Figure 4.12.

$$|\Delta\phi(\mathrm{Z}, \mathrm{jet}_1)| > 2.8 \tag{4.21}$$

$$\alpha \equiv \frac{p_{\mathrm{T}}^{\mathrm{jet}_2}}{p_{\mathrm{T}}^{\mathrm{Z}}} < 0.2 \tag{4.22}$$

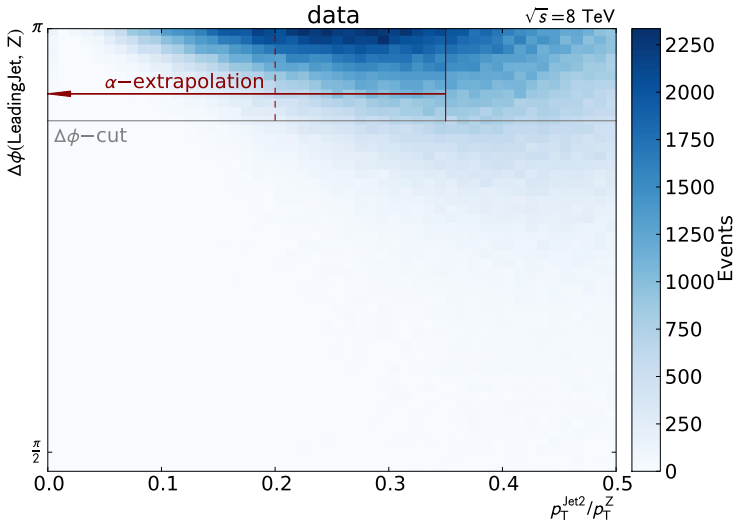


Figure 4.12: Event selection for the $Z(\rightarrow \mu\mu)$ + jet topology. The $\Delta\phi$-cut selects events close to a perfect balancing ($\Delta\phi = \pi$). For the distributions in this chapter, the second jet cut $\alpha < 0.2$ is applied. The final result is extrapolated to $\alpha = 0$ using all events with $\alpha < 0.35$ and thus most of the available event statistics.

Ideally, the Z boson should be balanced by exactly one jet and no other activity should be present in the event. However, pile-up and the underlying event activity

lead to further jets as well as initial and final state radiation. In Drell-Yan events measured from collision data, the average number of valid jets is 40. Therefore, the final results are extrapolated to the perfect topology by extrapolation of the second jet cut to zero additional jet activity. This procedure is further described in Section 4.8.3. The distributions of kinematic variables and event properties are shown with the fixed cut on α in Eq. (4.22).

4.7.5 Background estimation

This event selection guarantees a very clean sample of $Z(\rightarrow \mu\mu) + \text{jet}$ events. Apart from the Drell-Yan process, muons can be produced in $t\bar{t}$ events where both W bosons decay leptonically. W + jet events are another source of muons in pp collisions. Rarely, muons emerge from QCD interactions. While only 4.7 % of the muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are produced in the Drell-Yan process [63], the isolation criteria and the Z boson selection suppress these backgrounds drastically. The contribution of other processes is negligible as Table 4.1 shows. These estimations are based on the NNLO cross sections of the corresponding processes [92]. The QCD sample is produced with a filter for muons on generator level that enriches the fraction of events containing muons by a factor of 3.7×10^{-4}. However, none of those events passes the selection criteria. For this background as well as for the $t\bar{t}$ background, limits at 95 % CL are given instead of the expected number of events in the table. The Drell-Yan sample as well as the W+jet background are normalised with the cross-section from [93]. For $t\bar{t}$ events, the cross section is taken from [96].

Table 4.1: Background estimation for the $Z(\rightarrow \mu\mu)$+jet event selection based on Monte Carlo simulation for data corresponding to an integrated luminosity of $L = 19.7\,\mathrm{fb}^{-1}$.

Dataset	**Data**	**DY+jets**	**QCD**	$t\bar{t}$**+jets**	**W+jets**
σ / pb (NNLO)		3 531.9	3.64×10^8	245.8	36 703.2
Filter efficiency		1	3.7×10^{-4}	1	1
Events in sample		30 459 503	21 484 602	5 186 494	57 709 905
Z selection	1 615 904	1 574 974	< 328	7866	240
		98 %	< 0.02 %	0.49 %	0.01 %
Final selection	90 431	94 642	< 328	51	< 35
		105 %	< 0.36 %	0.06 %	< 0.04 %

4.8 Calibration of the absolute scale

4.8.1 Study of the reference object

For the calibration of the balanced jet using $Z(\rightarrow \mu\mu)+\text{jet}$ events, the understanding of the Z boson as reference object is of utmost importance. The precise muon measurement is demonstrated in Figure 4.13. The kinematics of both muons are well described by the simulation and precisely reconstructed. The muon scale corrections (Sec. 4.6.2) reduce the uncertainty on the muon momentum to 0.2 %. More kinematic quantities of the muons are shown in the Appendix A.3 where the high quality of the estimation of trigger efficiencies in simulated events is also demonstrated.

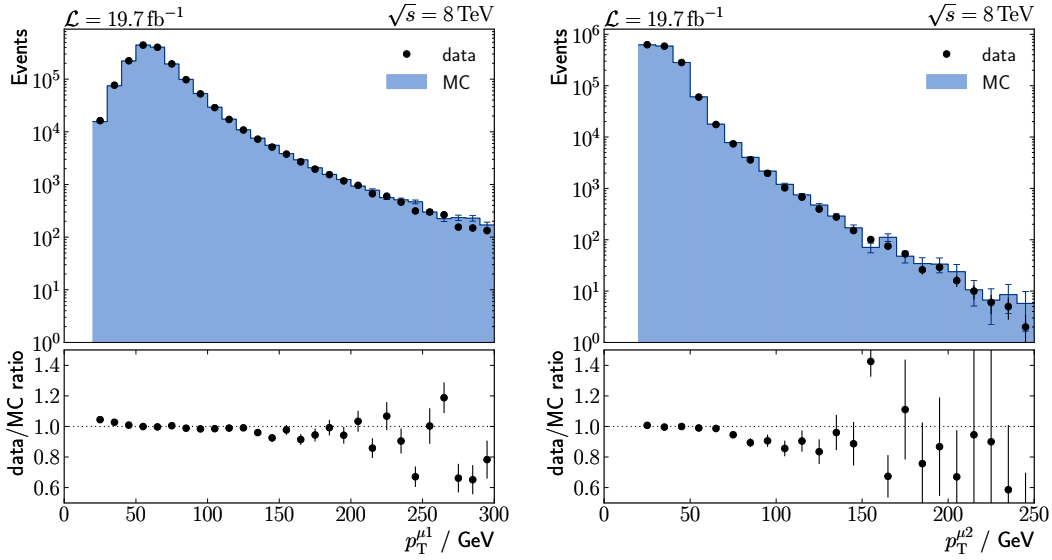


Figure 4.13: The ransverse momentum of the leading and second leading muon in data and simulation show a very good agreement over several orders of magnitude.

The Z boson is reconstructed from its decay products: a muon-antimuon pair. It thus inherits the high accuracy of the kinematic properties from the muons. The agreement with simulation is equally good as for the muons as Figure 4.14 shows. The same holds true for the pseudorapidity distribution.

The Z boson mass is another quantity that probes and demonstrates the precision of the reference object measurement. The distribution is shown in Figure 4.15. Being measured independently by precursor experiments, the location of the Z boson mass peak is suitable as a test of the Z boson reconstruction and muon momentum measurement. After the muon scale corrections, the measured value agrees with the

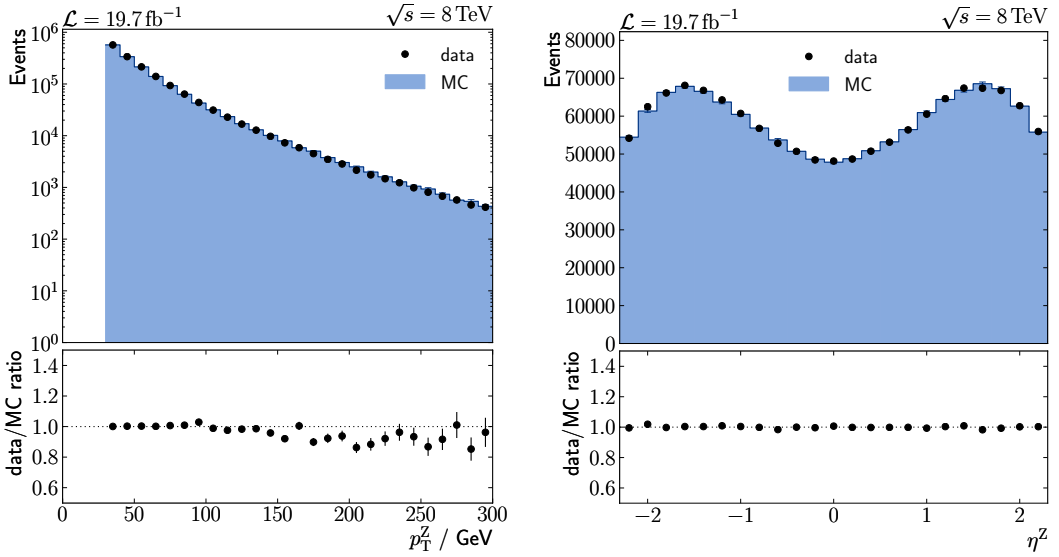


Figure 4.14: The transverse momentum $p_{\mathrm{T}}^{\mathrm{Z}}$ and the pseudorapidity η_Z of the Z boson as the reference object are well described by Monte-Carlo simulation. The double-peak structure in the pseudorapidity distribution results from the change in the binning with respect to the rapidity distribution. Both differ due to the non-vanishing mass of the Z boson. The latter has a smooth and monotone curvature and its maximum at $y = 0$.

expectation [1]. The measured width of the Z boson mass distribution in data is slightly wider than expected from simulated events. The behaviour of the Z boson mass as a function of the transverse momentum of the Z boson is another test of possible dependencies of the Z boson reconstruction on the transverse momentum. The slight decrease of the mean Z boson mass towards the low mass region is expected due to the constrained phase space and perfectly described by simulation.

The uncertainty on the Z boson momentum measurement is estimated by propagating the uncertainties on the muon momenta to the Z measurement. In addition, the scale and width of all components of the Z boson kinematics are compared for those shifts. This is done separately in case both muons are reconstructed in the barrel of the muon system, in case both muons are found in the muon endcaps and for mixed configurations. While the differences are mostly negligible, minor variations in the muon scale between the muon barrel and the endcap and in the Z mass reconstruction are measurable. Combining all uncertainty estimations, the precision of the reference

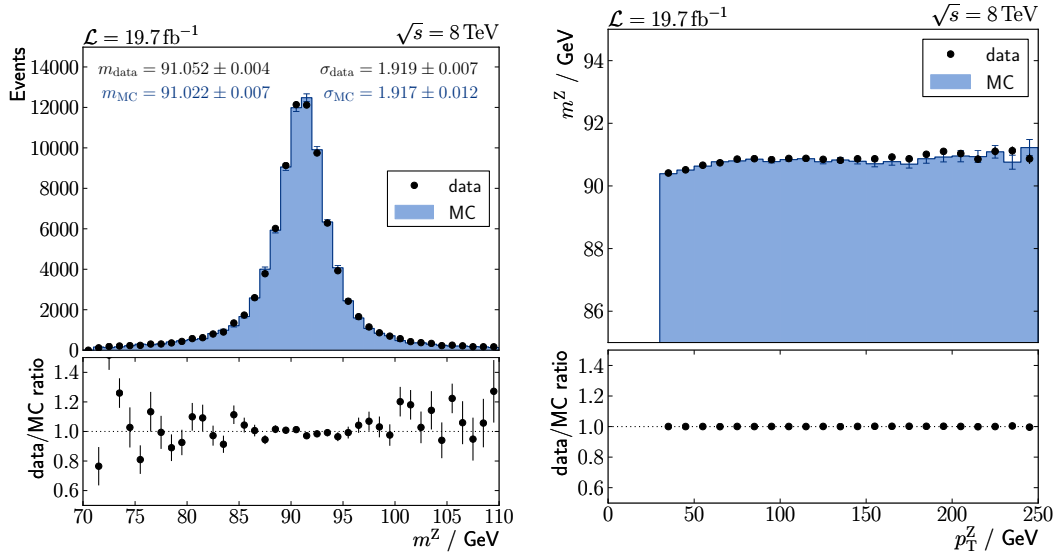


Figure 4.15: Z boson kinematics

scale is found to be

$$\frac{\Delta p_{\mathrm{T}}^{\mathrm{Z}}}{p_{\mathrm{T}}^{\mathrm{Z}}} = 0.2\,\%\,.$$

No other reference object can provide such a precision.

4.8.2 Measurement of the leading and second leading jet

The Particle Flow reconstruction of events allows to study the energy contribution of different particle classes to the total jet energy. The composition of the leading jet is depicted in Figure 4.16. On average, 60 % of the jet energy is carried by charged hadrons (CHad) such as charged pions and kaons. Photons, which mostly arise from neutral pion decays, make up about 30 % of the jet energy. The energy fraction of other particles detected as neutral hadrons (NHad) is 8.2 %. Only minor fractions of the jet energy are carried by electrons and positrons (e). Due to the isolation criteria for jets, the energy contribution of muons is negligible. The composition of the leading jet shows almost no dependence on the transverse momentum. High energetic jets contain less electrons while the photon fraction increases by 3 %. The differences in the jet composition between recorded data and simulation are small. They amount to less than about 1 % in some p_{T} regions while being smaller than 0.4 % on average.

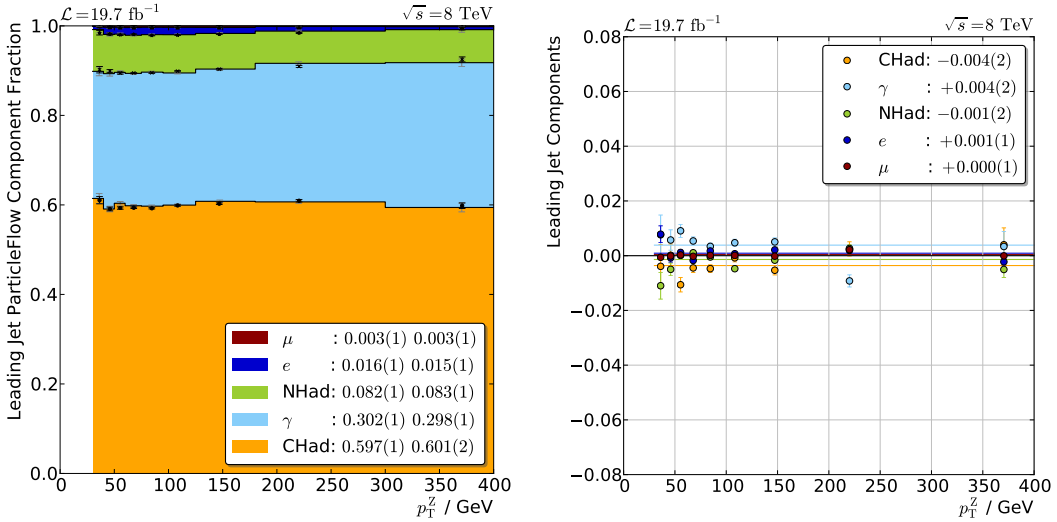


Figure 4.16: Energy deposits of the leading jet (*left*) in different detector components as a function of the transverse momentum of the Z boson. While the electron/-positron fraction decreases towards high p_{T}, the photon fraction increases slightly. The difference from the Monte-Carlo description (*right*) is below 0.5 % for all components.

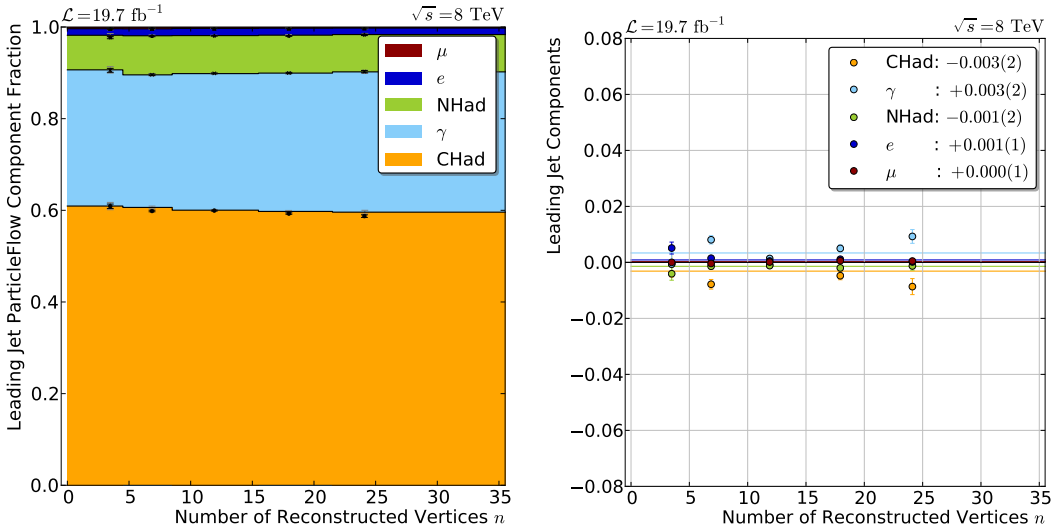


Figure 4.17: Energy deposits of the leading jet (*left*) in different detector components as a function of the number of reconstructed primary vertices n_{PV}. The stability with respect to the n_{PV} shows how well pile-up is under control. The data/MC differences (*right*) are $\leq 0.3\,\%$.

The energy fractions are also stable with respect to the number of reconstructed primary vertices in the event as Figure 4.17 shows. Since pile-up jets differ in their composition, this is a sign for the success of the pile-up mitigation techniques.

After the application of all preceding levels of jet corrections, the shape of the jet p_{T} spectrum is already reasonably described by Monte-Carlo simulation (Fig. 4.18). A small shift is nonetheless visible that requires the data-driven corrections described in this chapter. The distribution of the pseudorapidity is perfectly described considering the statistical uncertainties of simulated events.

The modelling of the second leading jet is crucial, too, as it affects the event selection via the α cut, Eq. (4.22), and the extrapolation to $\alpha \to 0$. Figure 4.19 shows the agreement of the second jet distribution in data with the simulation over many orders of magnitude.

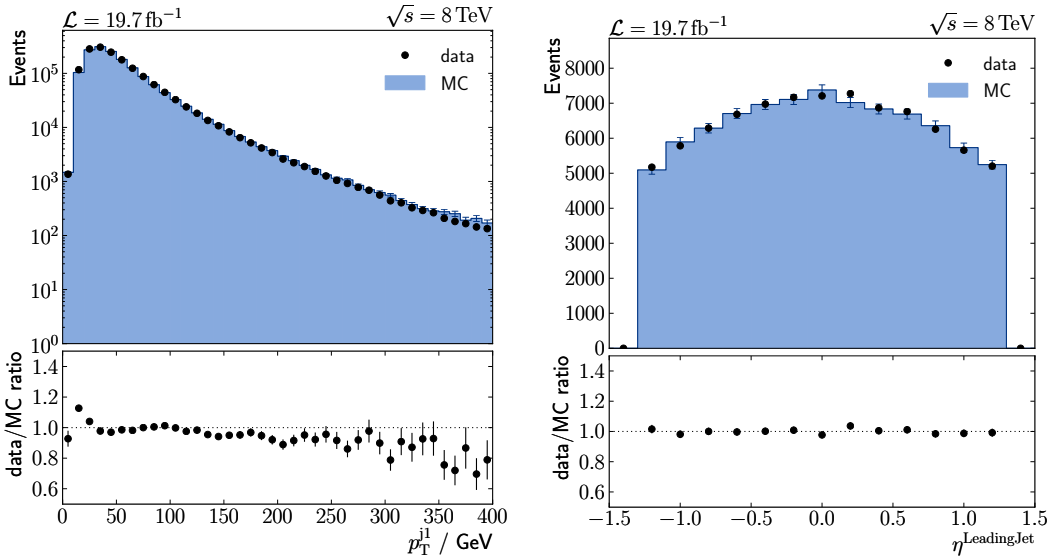


Figure 4.18: Kinematic distributions of the leading jet. The leading jet p_{T} (*left*) already shows a reasonable agreement between data and simulation. However, the disagreement at low p_{T} indicates a small shift in data. This requires the data-driven corrections derived in this chapter. The η distribution (*right*) is perfectly described by Monte-Carlo simulation.

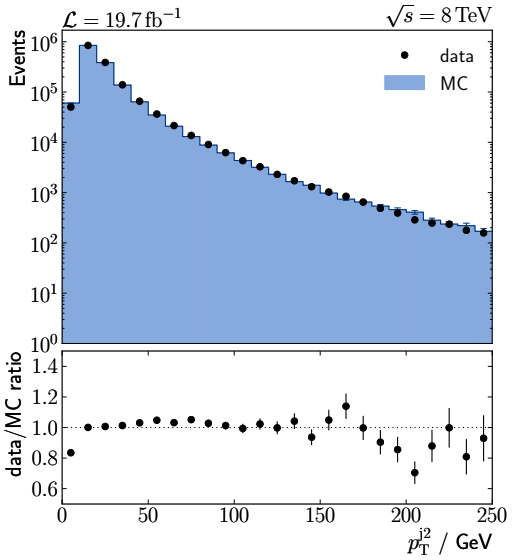


Figure 4.19: The transverse momentum distribution of the second leading jet in data agrees with simulation over many orders of magnitude due to the NLO description of the process by MADGRAPH.

4.8.3 Extrapolation to the ideal topology and determination of the absolute jet energy correction

The ideal topology of exactly one jet balancing the Z boson is extremely rare in real collisions. Final state radiation (FSR) leads to a splitting of the recoil jet and thus to one or more additional jets around the balanced jet. In this case, the leading shares the energy of the initial parton with the additional jet and is no longer balanced with the Z boson. This results in a reduced jet response with the balancing method. The MPF response is expected to be less biased by FSR as it considers the whole recoil unaffected by possible jet splitting. In addition, a large number of mostly low energetic jets from the underlying event and pile-up populate the event. To remove the radiation bias, the jet response is extrapolated towards the ideal event topology without additional jet activity. The response R is extrapolated as the second jet cut goes to zero,

$$\alpha \equiv \frac{p_{\mathrm{T}}^{\mathrm{jet_2}}}{p_{\mathrm{T}}^{\mathrm{Z}}} \tag{4.23}$$

$$R = \lim_{\alpha \to 0} R(\alpha) \tag{4.24}$$

In order to derive residual corrections for jets in data, all pile-up and simulation based jet energy correction levels are applied to jets in both simulated $Z(\to \mu\mu) + \text{jet}$ events and collision data. The extrapolation for such events is shown in Figure 4.20.

As expected, the p_{T} balance method exhibits a strong dependence on the momentum fraction α while the slope of the jet response extrapolation is small when measured using the MPF method. In both cases the slopes are well described by simulation. However, it is clearly visible that the jet energy scale differs by almost 2 % which shows the need for residual corrections for jets in data. The fact, that the jet response in simulated events lies above unity, is a consequence of the different flavour mixture in $Z(\to \mu\mu) + \text{jet}$ events compared to the average mixture in LHC events. The former are dominated by light quarks (65 %) as a result of the proton PDF. QCD production of jets has the highest cross section at the LHC and it is dominated by gluon jets. This results in a 1 % shift in the jet response as will be shown in Section 4.10. Both methods for measuring the jet response agree in the limit of the perfect event topology. The jet response ratios obtained using the p_{T} balance method and using the MPF method are

$$R^{\mathrm{bal}}_{\mathrm{data/MC}} = 0.981 \pm 0.002 \tag{4.25}$$

$$R^{\mathrm{MPF}}_{\mathrm{data/MC}} = 0.983 \pm 0.003 \tag{4.26}$$

where the response ratio is defined as $R_{\mathrm{data/MC}} = R_{\mathrm{data}}/R_{\mathrm{MC}}$ of the measured jet responses in data and Monte-Carlo simulation.

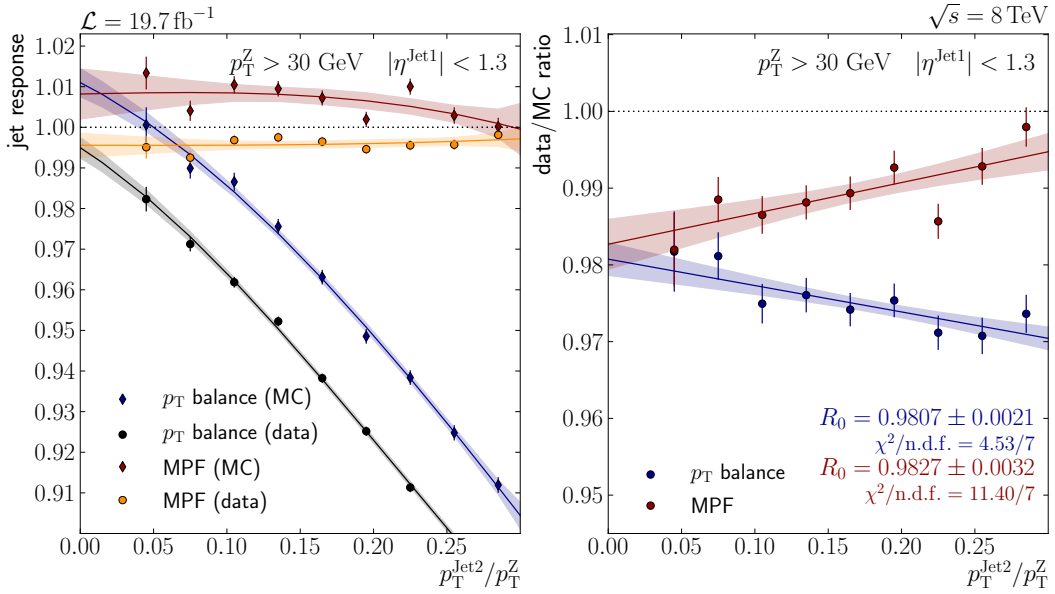


Figure 4.20: Extrapolation of the jet response to the ideal topology (*left*). To account for final state radiation, the measured jet response is extrapolated from non-negligible presence of second jets ($\alpha \approx 0.3$) to the ideal balancing against the only jet in the event. This affects the measurement using the p_T balance method while the MPF method is less influenced. The correction factors are derived from the extrapolated data/simulation ratio (*right*).

From this measurement, the final correction factor of the absolute scale residual corrections, c_abs, is calculated as the reciprocal of the measured and extrapolated jet response ratio. For the data recorded in 2012 it amounts to

$$c_\mathrm{abs} = 1.0189\,. \tag{4.27}$$

Applying the full set of jet energy corrections including the residual corrections gives the response measurements and ratios shown in Figure 4.21. The slope of the p_T balance is well modelled in simulation. The modelling of the whole event including the missing energy is slightly less accurate. This is because in data, additional pile-up activity in the endcaps affects the MPF response at higher values of α. All these effects are corrected for by the final state correction factor k_FSR derived from the α

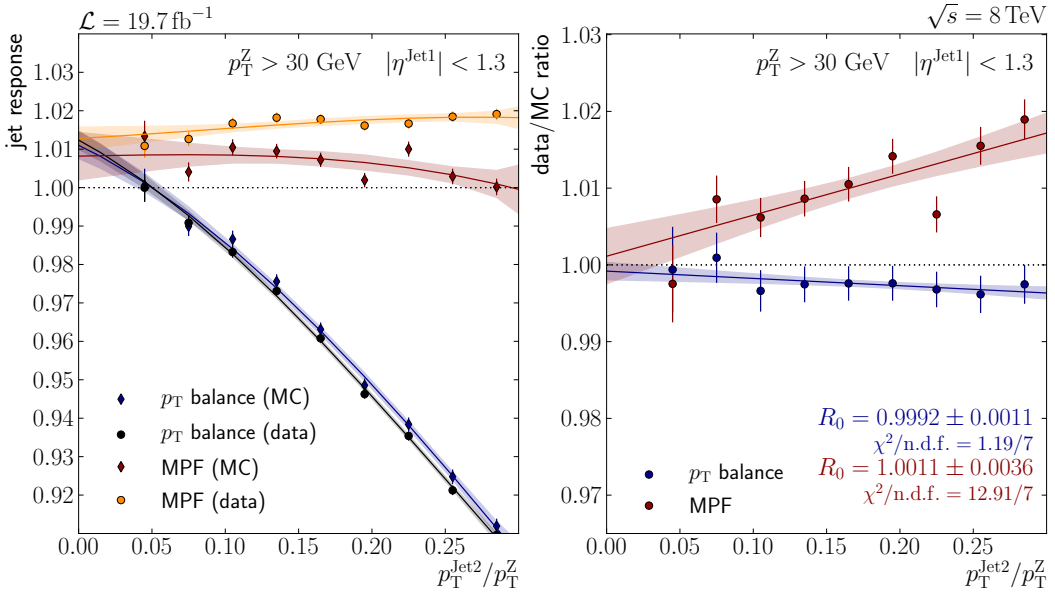


Figure 4.21: After the application of the derived residual corrections of the absolute scale, the measured jet responses agree in data and simulation (*left*). A different flavour mixture compared to QCD events results in a jet response above unity as will be discussed in Section 4.10. The data/MC ratio is brought to unity by the above-derived corrections (*right*).

extrapolation of the data over simulation ratio

$$k_\mathrm{FSR}^\mathrm{bal} \equiv \frac{R_\mathrm{data/MC}^\mathrm{bal}(\alpha = 0)}{R_\mathrm{data/MC}^\mathrm{bal}(\alpha < 0.2)} = 1.002 \tag{4.28}$$

$$k_\mathrm{FSR}^\mathrm{MPF} \equiv \frac{R_\mathrm{data/MC}^\mathrm{MPF}(\alpha = 0)}{R_\mathrm{data/MC}^\mathrm{MPF}(\alpha < 0.2)} = 0.994 \tag{4.29}$$

which is applied to response ratios in all following figures. Separate k_FSR factors are also applied to jet responses in data and simulation to remove the bias due to final state radiation from both methods. The ratio of the two factors for data and simulation is fixed to the value obtained in the extrapolation of the response ratio, Eq. (4.28) and (4.29). This way, the final correction factor of these data-driven method remains independent from the individual extrapolations, which serve illustrative purposes only. For the same reason, an additional global factor of 0.99 is applied to account for the difference in the jet response due to the different flavour mixture in $Z(\to \mu\mu) + \mathrm{jet}$ and QCD simulation which cancels in the response ratios. This factor is determined

in dedicated flavour studies (Sec. 4.10). As the flavour mixture varies for different jet p_T regions, this global factor is only a rough estimate and p_T dependencies of the jet response are expected.

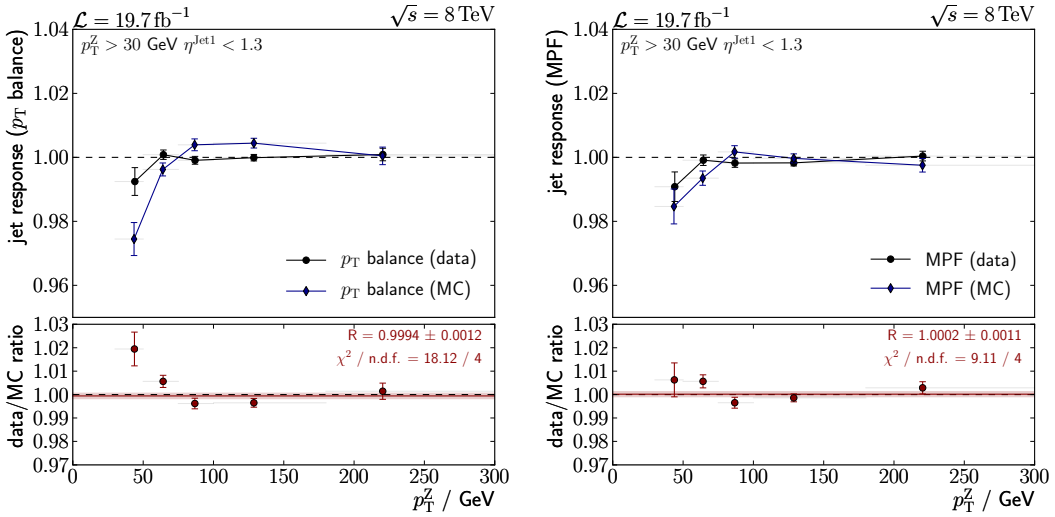


Figure 4.22: The jet response as a function of the transverse momentum of the reference object using both the p_T balance (*left*) and the MPF method (*right*). Both methods show a similar trend and a lower response at low p_T.

The jet response in data is well described by the corrected simulated jet response as can be seen in Figure 4.22 for both methods at a level of 0.5 %. The lower response for low energetic jets agrees with the aforementioned effect of changes in the flavour mixture. The p_T balance method shows this more clearly as out-of-cone effects of wider gluon jets have a larger impact. The average ratio serves as a closure test for the data-driven corrections and both methods are perfectly compatible with a response ratio at unity. However, a slightly falling and again rising trend over p_T can be seen in both methods. This deviation got significant only recently, as the statistical uncertainty decreased when the all data recorded in 2012 became available and the systematic uncertainty as a result of detailed studies of the jet energy scale. After a confirmation by the other channels of the data-driven jet energy calibration, this will be accounted for by a new set of then p_T-dependent absolute residual corrections.

The measurement of the jet response is stable with respect to pile-up influences as shown in Figure 4.23. This is a result of the advanced pile-up corrections and mitigation techniques. The only deviation is observed for $n_{PV} \leq 4$ which represents 1.2 % of all events, indicating an overcorrection of pile-up for low pile-up events.

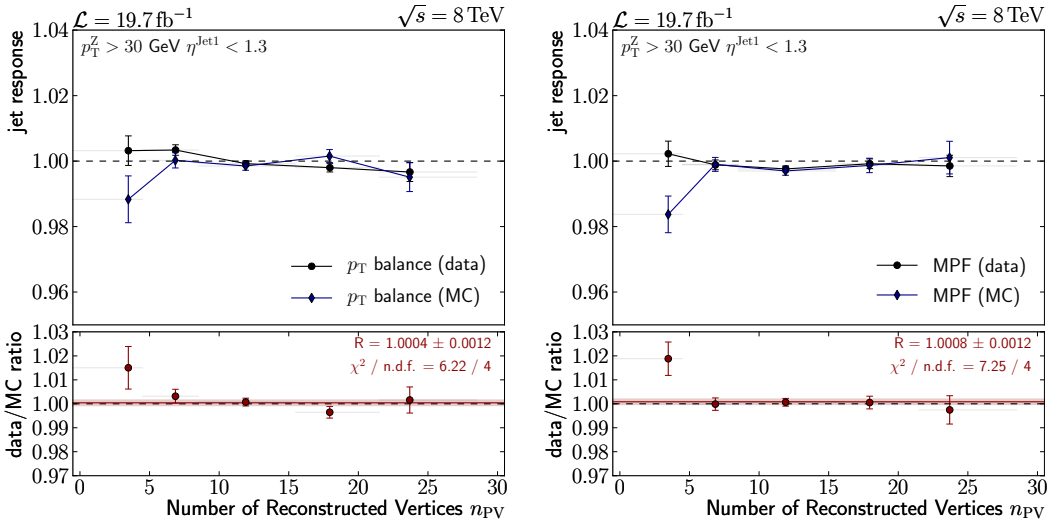


Figure 4.23: response using both the p_T balance (*left*) and the MPF method (*right*) as a function of the number of reconstructed primary vertices in the event. The measurement is stable and sufficiently independent from pile-up influences.

4.9 Extension to jets in the foward region

For data-driven corrections of the absolute scale, the homogeneous central part serves as an excellent reference region. However, for the calibration of jets with a higher pseudorapidity, the same techniques can be extended to regions covered by the endcap (HE) and the forward calorimeter (HF). Traditionally, jets in these regions are corrected relative to barrel jets using dijet events. Due to the increased number of events in 2012 data, the balancing in $Z(\rightarrow \mu\mu) + \text{jet}$ events became competitive and can measure the jet response with respect to the Z boson directly. At the same time, increased thresholds of the dijet triggers reduced the number of available events for the dijet balancing which is the traditional channel for data-driven calibrations of the endcaps and the forward region.

Loosening the restriction to jets in the central detector, Eq. (4.18), to $|\eta_{\text{j1}}| < 5.0$ extends the analysis to the forward region. The α-extrapolation has to be performed separately for different η regions. While the pile-up suppression using CHS is sufficient in the central detector and complementary methods contribute only little to the pile-up rejection, the pile-up jet identification (Section 4.4.2) becomes a crucial instrument in the endcap outside the tracker coverage and the forward calorimeters where the track-based approach of CHS is not possible. Inhomogeneities in the calorimeter layout, in particular the transition between the barrel and the endcap of the HCAL, represent major challenges for the jet reconstruction and calibration. Large uncertainties on the pile-up measurement in those regions are problematic as well as instabilities of the response measurement over time due to a decreasing calorimeter response.

The composition of the leading jet is depicted in Figure 4.24. As described before, charged hadrons provide the dominant energy contribution to jets in the barrel. The composition of the leading jet in the endcaps within the tracking region does not differ much from the one in the barrel. The end of the tracker coverage at $|\eta| = 2.4$ is clearly visible as an abrupt reduction of the charged particle contributions, namely the charged hadron fraction (CHad) and the contribution from electrons and positrons (e).

In the region outside the tracker acceptance, the reconstruction can solely distinguish between components from the electromagnetic and the hadronic calorimeter leaving less handles for component-specific studies. From this point until the transition from the endcaps to the forward region, all particles deposing their energy in the ECAL are treated as photons while all reconstructed hadrons are shown as neutral hadrons independent of their real charge. This twofold classification is also done for particles reconstructed from energy deposits in the forward calorimeter where the two components are called electromagnetic (HFem) and hadronic (HFhad) energy fraction in the hadron forward calorimeter range. The limited acceptance of the electromagnetic calorimeter reduces the electromagnetic fraction averaged in the

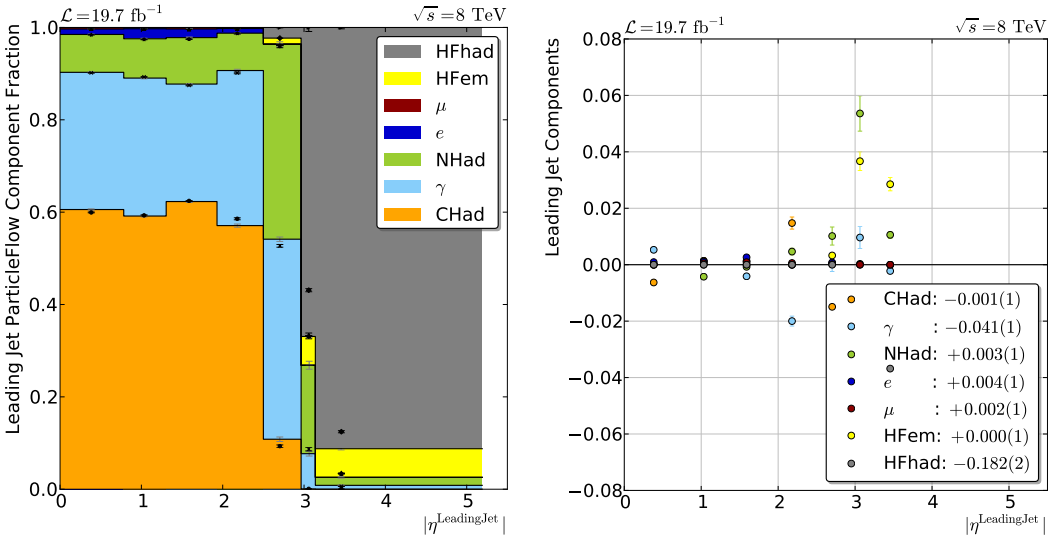


Figure 4.24: Composition of the leading jet over $|\eta|$. The fractions are relatively constant within the tracker coverage ($|\eta| < 2.4$). Outside the tracker acceptance, charged and neutral particles can not be distinguished any more and all particles are only divided in electromagnetic and hadronic calorimeter deposits. Significant differences between data and simulation are observed in the transition region between the endcap and the forward detector ($2.94 < |\eta| < 3.2$) and in the forward region.

forward region. The remaining small fraction of neutral hadrons and photons results from measurements at the border of the forward detector where some parts of the energy deposit are still measured in the endcaps. The composition in simulated events describes the data with an accuracy of $< 1\,\%$ in the whole region with tracker coverage. However, in the outer parts of the endcaps and in the forward region the accuracy is only of the order of $3\,\%$ to $5\,\%$.

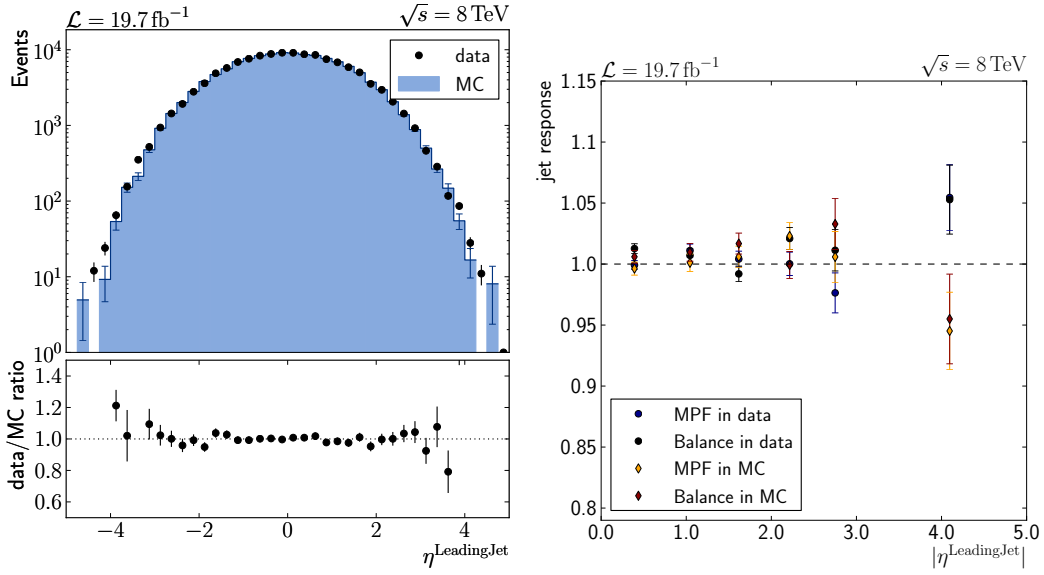


Figure 4.25: The η distribution of the leading jet (*left*) shows that the number of usable events for $Z(\to \mu\mu)+$jet calibrations is drastically reduced for high values of $|\eta|$. This is reflected in the increasing statistical uncertainties in the measured jet response (*right*). The p_{T} balance and the MPF method agree in the measured jet response. The correction factors from dijet balancing, which are applied as part of the common jet energy corrections can be confirmed by $Z(\to \mu\mu) +$ jet measurements.

The jet response is determined using both the MPF and the p_{T} balance method. The extrapolation $\alpha \to 0$ is again done for each method separately. However, the different environment in the detector and differences in the particle reconstruction requires the extrapolation to be performed separately in each η region. While the final state radiation was by far the main reason for the α dependence in the homogeneous barrel region, the amount of residual pile-up increases in the endcaps. Outside the tracker coverage, the in unavailability of the Charged Hadron Subtraction increases the pile-up contribution and hence the amount of additional jets in the event.

The average transverse momentum of jets is lower for higher values of η as a matter

of kinematic constraints. The fractional energy loss due to final state radiation is larger for low energetic jets. This entails a monotone increase of the slope of the extrapolation over η. Additional jets from the underlying event and hadronization products of the beam remnants lead to additional jet activity in the forward regions. Deriving individual extrapolation factors for each region accounts for those effects and allows to measure the jet response of an ideal configuration.

The η distribution of the leading jet in Figure 4.25a shows that the number of events decreases drastically for high values of $|\eta|$. Almost no leading jets from $Z(\rightarrow \mu\mu)+\text{jet}$ events have pseudorapidities of $|\eta| > 4$. The statistical uncertainties on the measured jet response outside the barrel region are therefore much larger. Leaving more freedom to the extrapolation fit, this in turn further increases extrapolation uncertainties. Thus, the measurement of the jet energy scale with uncertainties on the level of a few per mil, as demonstrated in the barrel, is not possible in the endcaps. However, a jet calibration comparable to the state of the art relative residual corrections is achievable up to $|\eta| = 3$ and cross checks with a reduced precision can be performed in the high rapidity region. This reach in $|\eta|$ is sufficient for most jets in typical analyses and even for the majority of tagging jets in VBF events (cf. Chapter 5).

The response measurements in Figure 4.25b confirm the dijet measurements and the calibration factors derived from them. The ratio of the jet response in data and simulation are compatible with unity over a wide η range. However, there is one exception which will be discussed in the following section.

4.9.1 Time dependence of the jet response

In the context of this thesis, a time dependence of the jet response was discovered. The jet energy scale is stable as expected in the central part of the detector and does not vary over time as shown in Figure 4.26. On the whole 2012 dataset, jets reconstructed in the parts of the endcap without tracker coverage show a significantly lower jet response in recorded data than expected. While the correction factors for the jet response was adequate in the first period of data-taking in 2012 (run 2012A), the reconstructed jet response fell by 15 % until the end of data-taking (run 2012D) in the first official reconstruction of collision data (prompt reco) which was used for the first six months after the end of the LHC run at $\sqrt{s} = 8\,\text{TeV}$. This time dependence of the jet response in the preliminary dataset is shown in Figure 4.27 as a grey band.

The reason for this time dependence was found to originate from a transparency loss in the hadronic calorimeter. Accumulating radiation damage, caused by the increased particle flux in the endcaps, leads to a loss of transparency of the hadronic calorimeter cells. As a consequence, the measured energy in those calorimeter cells is lower than the one of the incoming particle. As before the energy fractions in detector components are denoted by f_c. The time dependence of the energy fractions can be illustrated by the modified leading jet component fraction difference $\Delta f'_c$ which is

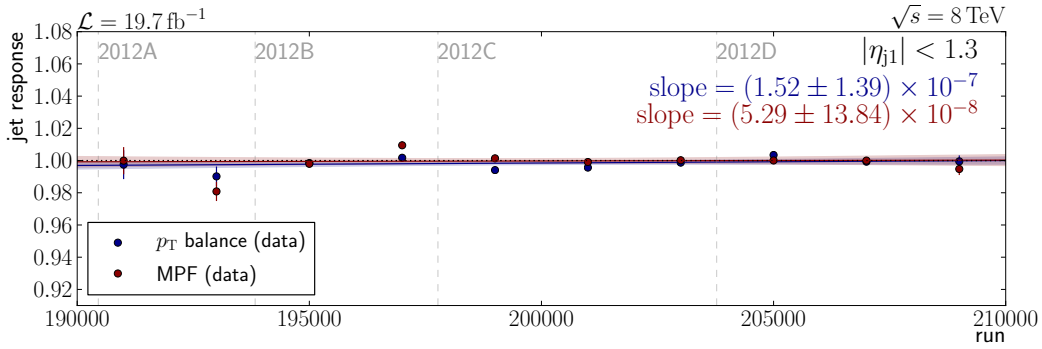


Figure 4.26: The time dependence of the jet response in the barrel ($|\eta| < 1.3$). In this region, the jet response is stable with respect to the number of the run in which the event was recorded.

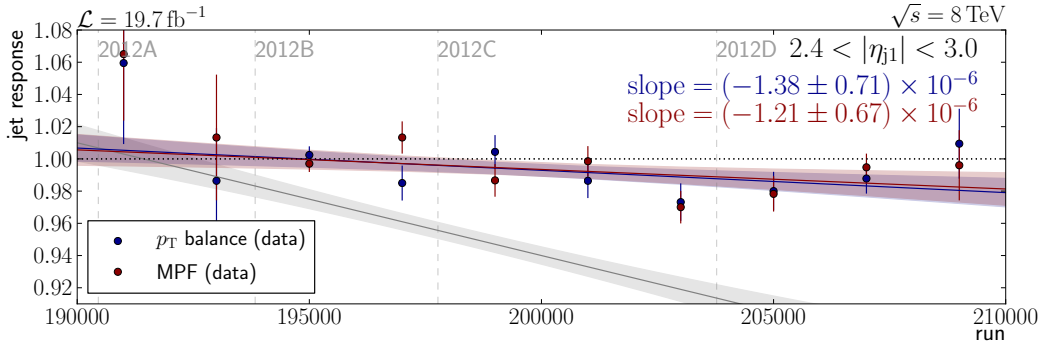


Figure 4.27: Time dependence of the jet response in the outer endcaps ($2.5 < |\eta| < 3.0$). In the prompt reconstruction of 2012 data, the jet response showed a dramatic dependence of the jet response on the number of the run in which the event was recorded as indicated by the grey band. The rereconstructed data was corrected for this time dependence. Both methods for measuring the jet response show a significant but now small reduction over time.

defined as

$$\Delta f'_c = R^{\text{data}}_{\text{bal}} \cdot f^{\text{data}}_c - R^{\text{MC}}_{\text{bal}} f^{\text{MC}}_c \qquad \text{with} \quad c \in [\text{CHad}, \text{NHad}, \gamma, e, \mu] \,.$$

By the multiplication with R_{bal}, this quantity $\Delta f'_c$ does not compensate the loss of response in one detector component with increased energy fractions of the other components. Therefore, it can be interpreted as the response of an individual detector component compared to the expectation in simulated events.

In Figure 4.28, the slope of the modified neutral hadron fraction f'_{NHad} as a function of the run in which the event was recorded is shown in grey for the first reconstruction of all data from the 2012 data-taking period. All data events were then re-reconstructed in 2013. In this procedure, the observed transparency loss was accounted for and the largest part of the time dependence was removed. The values of the modified fractions is now relatively stable also in the endcaps of the hadronic calorimeter. The new reconstruction is used throughout this thesis and Figure 4.27 shows that the stability of the jet energy scale has improved significantly with the updated energy calibration for the hadronic calorimeter.

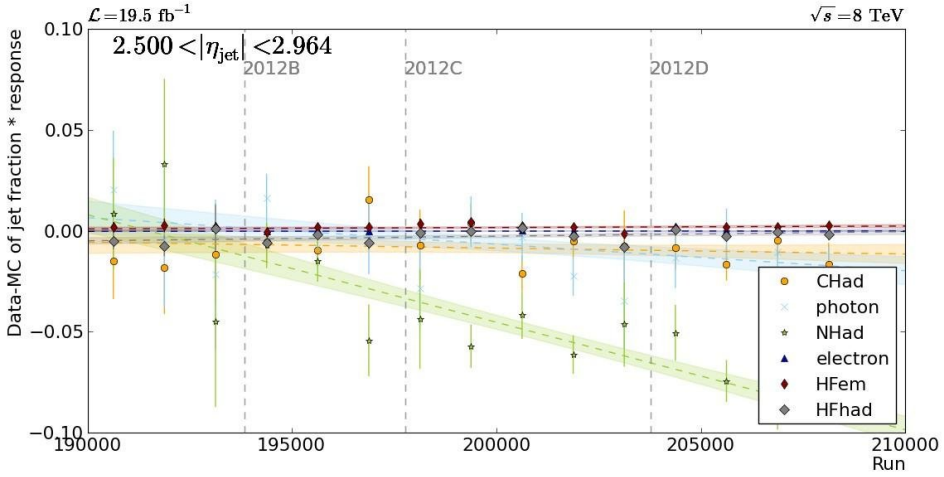


Figure 4.28: Time dependence of the leading jet composition. The green line shows the slope of the neutral hadron fraction in the first data reconstruction (prompt reco) resulting from a transparency loss of calorimeter cells. Correcting the energy of this jet component removes the downward trend of the jet response over time.

These findings show that the response measurements using the $Z(\to \mu\mu)+$jet event topology are now sensitive to small variations and biases over time. Furthermore, the Particle Flow reconstruction method enables the study of jet constituents to identify the source of such effects.

4.10 Flavour dependence of the jet response

The jet flavour is defined as the flavour of the initiating parton from which the jet emerges. Jets resulting from differently flavoured quarks and from gluons show differences in the momentum distributions and the measured jet response. This section presents the studies of the flavour dependence of the jet response in $Z(\rightarrow \mu\mu) + \text{jet}$ events.

Jet flavour definition based on generator information The purpose of a jet flavour definition is to relate the jet to the initial parton from which it emerged. While this seems straightforward using the generator information, it is often not unambiguously possible due to gluon splitting and initial state radiation. Therefore, three different definitions yield slightly different results:

- The *physics definition* matches the reconstructed jet to the initial parton from the matrix element process. The distance between the parton and the jet in the η-ϕ-space has to be within $\Delta R < 0.3$. In case there is no parton inside this cone, the jet flavour remains undefined. The same holds for the ambiguous case with more than one parton within the defined distance. The physics definition is used in the following section as it is the only definition supported by all used Monte-Carlo generators.
- The *algorithmic definition* addresses the sizeable fraction of unmatched jets in the physics definition. In contrast to the latter one, the final state partons after the parton shower are used for the matching procedure applying the same distance condition. However, if multiple partons are found within the cone, priority is given to c and b quarks. If none of those are matched, the hardest parton is returned as the jet flavour. While this is further away from the physically interesting hard process, the flavour identification efficiency of this algorithm is higher and it concentrates on the effects of hadronization.

About 70 % of the leading jets in $Z(\rightarrow \mu\mu) + \text{jet}$ events are light quark jets as can be seen in Figure 4.29. While the fraction of heavy quarks is low, the fraction of gluon jets amounts to 20 %. At low p_{T}, unmatched jets are mostly due to wide angle radiation without any parton inside the cone. In high p_{T} interactions, initial state radiation leads to additional partons in the matching cone of the physics definition.

As shown in Figure 4.30 for simulated events, the jet response strongly depends on the jet flavour. Compared to u- and d-quark jets, s-quark jets show a reduced jet response due to K^0_{L} production in the hadronization process, showing up as an increased fraction of neutral hadrons in Figure 4.30. These K^0_{L} mesons have a lower reconstruction efficiency than other hadrons. The electron and muon fractions are increased for c- and b-jets resulting from semi-leptonic decays also involving neutrinos.

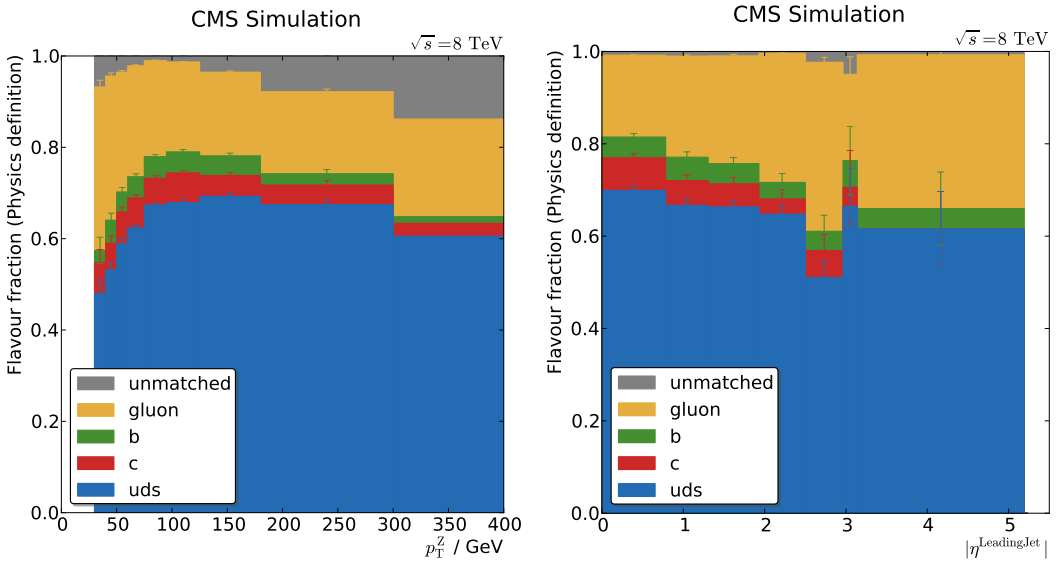


Figure 4.29: Fractions of jet flavours of the leading jet as described in simulation using the physics definition. The dependence on p_T (*left*) and on η (*right*) of the leading jet is shown. At low p_T, the fraction of unmatched jets is due to wide angle radiation where all partons are out of the jet cone. Initial state radiation leads to additional partons in the matching cone for high p_T jets and thus to a fraction of unmatched jets. The flavour composition over η is correlated with the p_T dependence as jets at high rapidities are softer on average.

As the latter cannot be measured in the detector, the jet response is drastically reduced for b-jets. Differences in the hadronization process of gluons lead to wider jets and thus more out-of-cone effects compared to light quark jets.

The out-of-cone effects can be taken into account by separate α-extrapolation for each jet flavour. The expected neutrino momenta can be taken from the Monte-Carlo truth information and can be added to the jet to account for this part of the missing energy. For this figure, no attempts were made to normalise the response to the true jet response. In particular the linear extrapolation of the final state radiation leads to a jet response above unity. By applying the aforementioned corrections for known flavour effects (Figure 4.30b), the jet response is agrees for all jet flavours showing the understanding of the underlying effects. Differences in the modelling of the parton-shower and hadronization between PYTHIA and HERWIG++ for all jet flavours can be seen in Figure 4.30. While the differences between both generators are significant for individual jet flavours, the average flavour mixture is comparable.

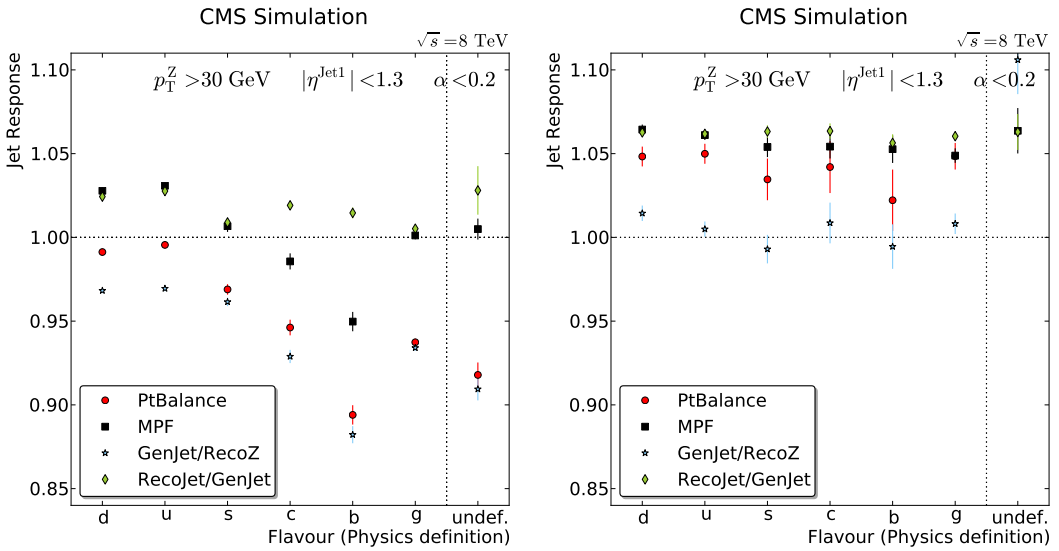


Figure 4.30: Flavour response for simulated jets using different response definitions. K_L, which can be produced in the hadronization process of s-quarks, have a lower reconstruction efficiency than other hadrons. Semi-leptonic decays in b-jets involve neutrinos which are not measured in the detector. Gluon jets are usually wider leading to more out-of-cone effects. By correcting for these effects, adding neutrinos and the correction of final state radiation via extrapolation, the flavour differences in the jet response can be removed.

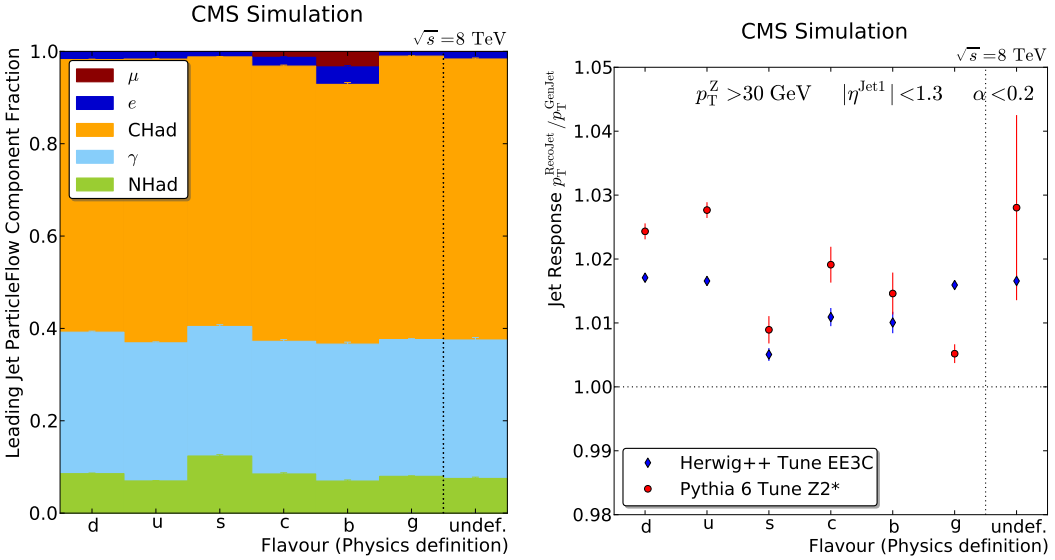


Figure 4.31: Jet composition of the leading jet for different jet flavours using the physics definition (*left*). *s*-jets contain a larger fraction of neutral hadrons (K^0) while *b*-jets contain muons from W decays. The different composition and width of the jets for different flavours lead to different jet responses. A comparison of the jet response for different jet flavours between PYTHIA and HERWIG++ is shown (*right*). While the true jet response for quark jets is lower in events simulated with HERWIG++ than in the PYTHIA sample, the opposite is observed for gluon jets.

Jet flavour tagging in reconstructed events In reconstructed events, the jet flavour can only be estimated indirectly from jet properties or the association to displaced vertices. While light jet flavours (u, d, s) are too similar to be distinguished, the identification of b-jets is an established technique and to some extent it is also applicable to c-jets. The second, more recently developed method is able to tell quark-jets from gluon-jets. These two methods are used to identify jet flavours in reconstructed event data.

- *b-tagging*: Jets coming from a b-quark can be identified by their displaced vertex. Hadrons containing b-quarks have a sufficient lifetime for a secondary vertex to be identified by the CMS tracker ($c\tau_{\mathrm{B}} \approx 0.5\,\mathrm{mm}$). In addition, b-jets are usually wider as light quark jets, have higher invariant masses and constituent multiplicities. All this information can be used for the experimental identification of b-jets (b-tagging) [97]. The combined secondary vertex b-tagging discriminator D_{b} allows to discriminate light quark jets ($D_{\mathrm{b}} = 0$) and b-jets ($D_{\mathrm{b}} = 1$). As the mass of c-quarks is in between light quarks and b-quarks, c-jet properties are frequently found to be in between the properties of the other corresponding jet flavours. Thus, even though the distinction is not very clear, the fraction of c-jets is enriched for discriminator values between 0.3 and 0.7.

- *quark-gluon discriminator*: A likelihood-product discriminator is defined including three variables that are able to distinguish between quark-jets and gluon-jets [98]. All variables are based on the particle flow candidates representing the constitutents of the jet with two restrictions: Neutral candidates have a minimum p_{T} requirement of $p_{\mathrm{T},i} > 1\,\mathrm{GeV}$ and the tracks of charged candidates have to be compatible with the main interaction vertex. The first variable is the particle multiplicity n_{c} as quark jets are expected to have a lower number of constituents n_{c} than gluon jets for a given p_{T} scale. Jet shapes can be evaluated considering the matrix

$$M = \frac{1}{\sum_i p_{\mathrm{T},i}} \cdot \sum_i p_{\mathrm{T},i}^2 \begin{pmatrix} \Delta\eta_i^2 & -\Delta\eta_i\Delta\phi_i \\ -\Delta\eta_i\Delta\phi_i & \Delta\phi_i^2 \end{pmatrix} \tag{4.30}$$

  representing the quadratic part of a conic section that describes the jet as an ellipse in the η-ϕ-space. The square-roots of the eigenvalues $\sigma_{1,2} = \sqrt{\lambda_{1,2}}$ of this matrix are the major and minor axes of the ellipse. The minor axis σ_2 is used as a second variable. As employed for the pile-up jet identification, the fragmentation function, Eq. (4.13), is also used to discriminate between quark and gluon jets. All three variables are evaluated for different η regions and in fine bins of p_{T} and the pile-up density ρ. The discriminator has a value between $D_{\mathrm{qg}} = 0$ for gluon jets and $D_{\mathrm{qg}} = 1$ for light quark jets.

In order to estimate the jet response for different flavours from the flavour mixture of $Z(\rightarrow \mu\mu)$ + jet events in data, different zones (1 – 4) are defined with respect to the tagger values. Each zone contains an enriched fraction of events of a particular jet flavour. As illustrated in Figure 4.32, zone 1 contains events with $D_{\mathrm{qg}} > 0.9$ and $D_{\mathrm{b}} < 0.3$ and thus mostly leading jets emerging from light quarks (uds), zone 2 is c-jet enriched ($0.3 < D_{\mathrm{b}} < 0.7$), b-jets are predominantly found in zone 3 ($D_{\mathrm{b}} > 0.9$) and gluon jets in zone 4 ($D_{\mathrm{qg}} < 0.1$ and $D_{\mathrm{b}} < 0.3$).

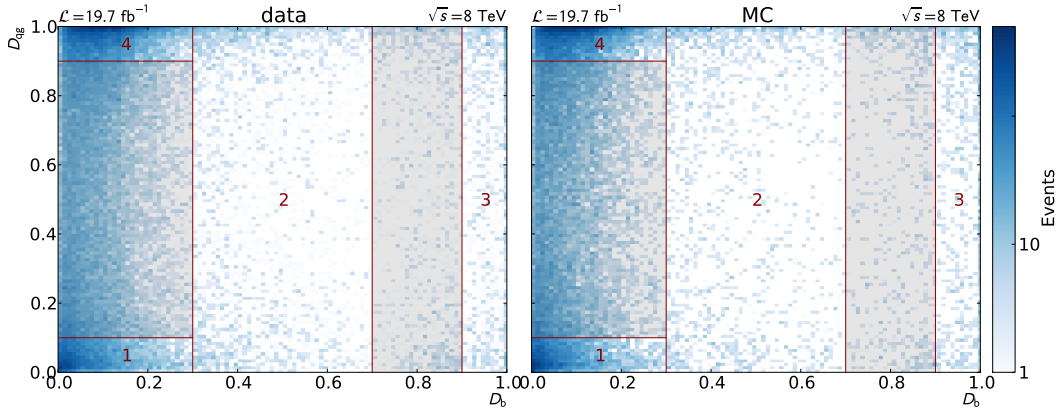


Figure 4.32: The four jet flavour tagging zones in data (*left*) and simulation (*right*). The number of events is shown in bins of the quark-gluon discriminator D_{qg} and CSV b-tagging discriminator D_{b}. Within the zones, different jet flavours are enriched: light quark jets in zone 1, c-jets in zone 2, b-jets in zone 3 and gluon-jets in zone 4. The shaded regions between 1 and 4 and between 2 and 3 contain events with ambiguous flavour mixtures that are not used for the flavour studies.

Within each of the four zones, the jet response can be determined and it shows strong differences between the zones. The most prominent deviation from the average is observed for b-enriched jets (zone 3) which is expected due to the semileptonic decay of B mesons involving neutrinos which do not contribute to the measured jet. These decays also involve leptons which is reflected by the higher electron and muon contributions to the jet as demonstrated in Figure 4.31. This can be seen in Figure 4.33 along with the corrections for the known effects due to jet flavours.

The fractions f_k^i of a true flavour k in zone i can be expressed in form of a matrix:

$$\begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{pmatrix} = \begin{pmatrix} f^1_{\mathrm{uds}} & f^1_{\mathrm{c}} & f^1_{\mathrm{b}} & f^1_{\mathrm{g}} \\ f^2_{\mathrm{uds}} & f^2_{\mathrm{c}} & f^2_{\mathrm{b}} & f^2_{\mathrm{g}} \\ f^3_{\mathrm{uds}} & f^3_{\mathrm{c}} & f^3_{\mathrm{b}} & f^3_{\mathrm{g}} \\ f^4_{\mathrm{uds}} & f^4_{\mathrm{c}} & f^4_{\mathrm{b}} & f^4_{\mathrm{g}} \end{pmatrix} \cdot \begin{pmatrix} R_{\mathrm{uds}} \\ R_{\mathrm{c}} \\ R_{\mathrm{b}} \\ R_{\mathrm{g}} \end{pmatrix} .$$

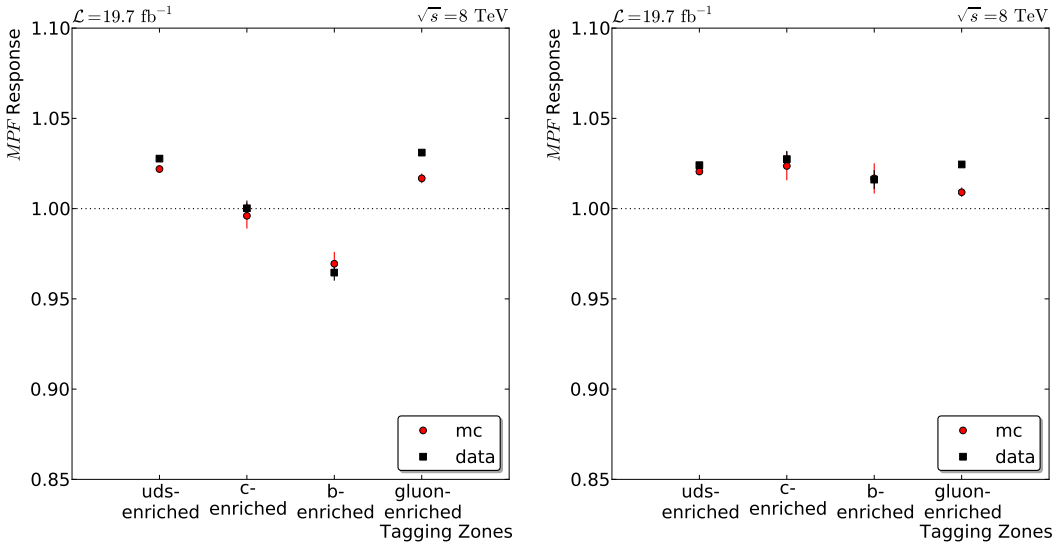


Figure 4.33: Jet response for different flavour-tag zones. On the left hand side, the unmodified MPF response is shown for each tagging zone, both for data and simulation. The jet response differs by about 5 %. On the right hand side, the MPF response is shown after applying flavour correction factors. The global correction factor for the flavour composition is not applied in these plots to show the response of each flavour. This results in jet responses above unity when corrected to the average in $Z(\rightarrow \mu\mu) + \text{jet}$ events.

The inversion of this matrix provides a measure for the jet flavour response on reconstructed events. The large contribution of other flavours in the c-enriched region (zone 2) results in larger uncertainties for the c-jet response. Flavour corrections can be applied as a function of the Particle Flow composition of a jet. Figure 4.34 shows that these effects are well understood as the flavour correction in combination with the extrapolation removes the differences which were observed in the uncorrected jet response.

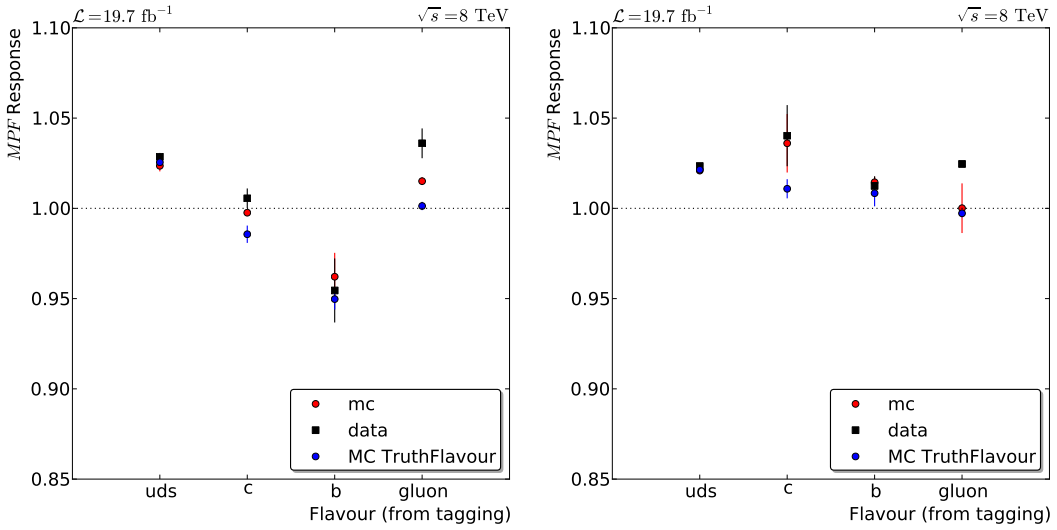


Figure 4.34: Jet response using the MPF method for different jet flavours using the flavour tag inversion method on data and Monte-Carlo samples. In addition, the flavour response is shown for the true flavour in Monte-Carlo simulation as given by the physics flavour definition. The uncorrected flavour response (*left*) shows differences of 7 % between light quark jets and b-jets. After Particle-Flow based flavour corrections (*right*), the responses for different flavours nearly agree.

4.11 Combination of channels for the absolute corrections

For the final jet energy correction level, the jet response measurement in the $Z(\rightarrow \mu\mu) + \text{jet}$ channel is combined with the other two channels of data-driven corrections of the absolute jet energy scale: $Z(\rightarrow ee) + \text{jet}$ and $\gamma + \text{jet}$. The former provides a direct comparison to the $Z(\rightarrow \mu\mu) + \text{jet}$ results due to the same reference object and lepton universality. Most differences between the two $Z + \text{jet}$ channels can be attributed to effects in the reconstruction of the electrons. Since the electron energy scale is sufficiently known only in the barrel of the electromagnetic calorimeter, the acceptance has to be reduced to $|\eta_e| < 1$. This introduces subtle differences between the two channels, resulting from uncertainties of the relative jet energy scale. Relying on calorimeter measurements, the precision of the electron momentum measurement is lower than for muons. The energy scale uncertainty of the ECAL is shared with the $\gamma + \text{jet}$ derived calibration. The latter has the huge advantage of a high statistical precision for events with a momentum transfer above about 200 GeV compared to the other channels. On the lower end of the probed p_T range, the photon triggers are highly prescaled and the number of usable events is very low.

The data over simulation ratio, $R_{\text{data/MC}}$, of the jet response measured in all three channels is shown in Figure 4.35. The response ratios show a remarkable agreement between the two methods in each channel. Furthermore, a similar value for the residual correction is obtained from all three channels and the remaining differences are subject of ongoing studies. The ECAL scale is suspected of introducing a shift in the reference scale for both $Z(\rightarrow ee) + \text{jet}$ and $\gamma + \text{jet}$. The downward trend observed previously in $Z(\rightarrow \mu\mu) + \text{jet}$ events is visible in all three channels. A p_T dependent residual correction is prepared to account for this slope. The permil level precision of the $Z(\rightarrow \mu\mu) + \text{jet}$ measurements result in a high weight for this channel in the combination.

In all three topologies, the leading jet p_T spectrum is falling exponentially, limiting the reach of data-driven jet corrections to several hundred GeV. The jet energy scale of high energetic jets is therefore extrapolated from the behaviour in simulated events. Multijet events in which the leading jet is balanced by multiple jets in ϕ offer a method for in-situ calibration of high energetic jets. This topology is is being studied and it is planned to complete the residual corrections in future.

4.12 Jet energy uncertainties

The uncertainties on the jet energy scale determination and correction are composed of a large number of sources listed below which are assumed to be uncorrelated. The sources are grouped by correction level. Each correction level has a set of uncertainty components that account for the uncertainties which arise from the derivation of the

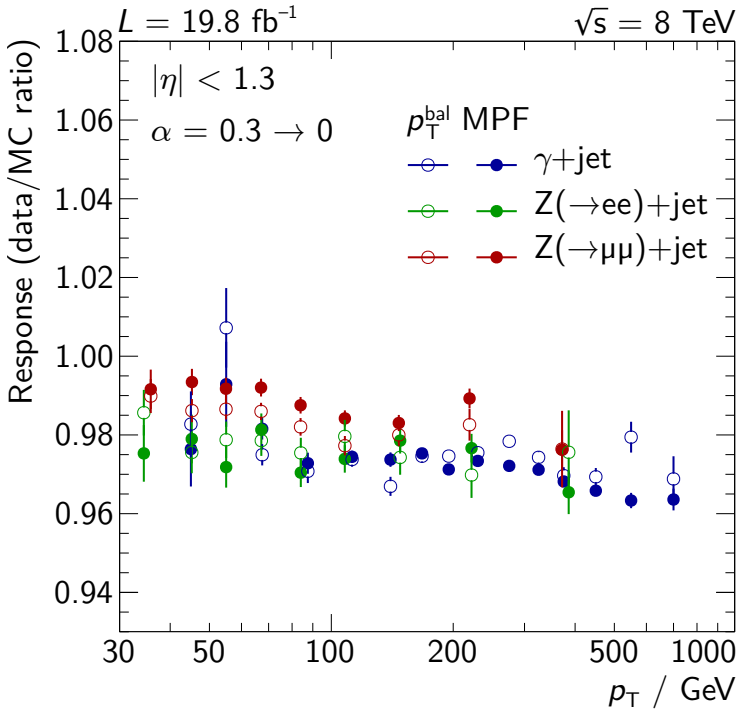


Figure 4.35: All channels of the data-driven absolute residual corrections. All three channels, $\gamma + \text{jet}$, $Z(\to ee) + \text{jet}$ and $Z(\to \mu\mu) + \text{jet}$, and both methods to measure the jet response, p_T balance and MPF, agree within a 1 % uncertainty over the full p_T range. While the $Z(\to \mu\mu) + \text{jet}$ result dominates the combination at low p_T due to the high precision, $\gamma + \text{jet}$ events are much more frequent above $p_\text{T} \approx 200\,\text{GeV}$.

correction factors of this level [99]. The total uncertainty on the jet energy scale is given by the quadratic sum

$$\Delta(p_{\mathrm{T}}, \eta) = \sqrt{\sum_i s_i(p_{\mathrm{T}}, \eta)^2}$$

of all sources s_i at a given p_{T} and η of the jet.

Pile-up (4 components) The difference between data-driven pile-up measurements in Zero Bias data and in Monte-Carlo simulation is not fully understood. While this effect is corrected for separately in data and simulation, 20 % of the difference is still used as an uncertainty on the pile-up measurement methods. In addition, three independent components are attributed to p_{T} dependence of the pile-up measurement in all of the HCAL subcomponents: the barrel, the endcap and the forward calorimeter.

Relative scale (10 components) The determination of the relative jet energy scale is affected by the resolution bias as the reference object (barrel jet) itself has a non-negligible resolution. This uncertainty is split into three regions: the endcap within the tracker acceptance, outside the tracker acceptance and in the forward detector (HF).

The p_{T} uncertainties with respect to the central fit in an η region requires four more uncertainty sources. These are constant values for the central detector region and the forward calorimeter to account for the inhomogeneities of the p_{T} measurement within these detector components. These values are not p_{T} dependent. In both parts of the endcaps however, non-linearities in the detector response occur possibly due to radiation damage. This requires an uncertainty modelled as a log-linear function of p_{T} in both parts of the endcap.

The correction factor for the final state radiation in dijet events comes with an additional uncertainty component. Its value is correlated between different regions of the detector but increases towards the forward region.

The limited number of suitable events for the relative residual corrections in the part of the HE outside tracking and in the HF leads to two further sources of statistical uncertainties.

Absolute scale (6 components) The uncertainty of the absolute scale measurement is based on the uncertainties of the reference object in the muon-system based $Z(\rightarrow \mu\mu)+$jet and the ECAL based $Z(\rightarrow ee)+$jet and $\gamma+$jet analysis methods. Furthermore, the uncertainty of the initial and final state radiation is taken into account. The flavour mixture in those three channels and a potential bias of the MPF method are two further sources of uncertainty.

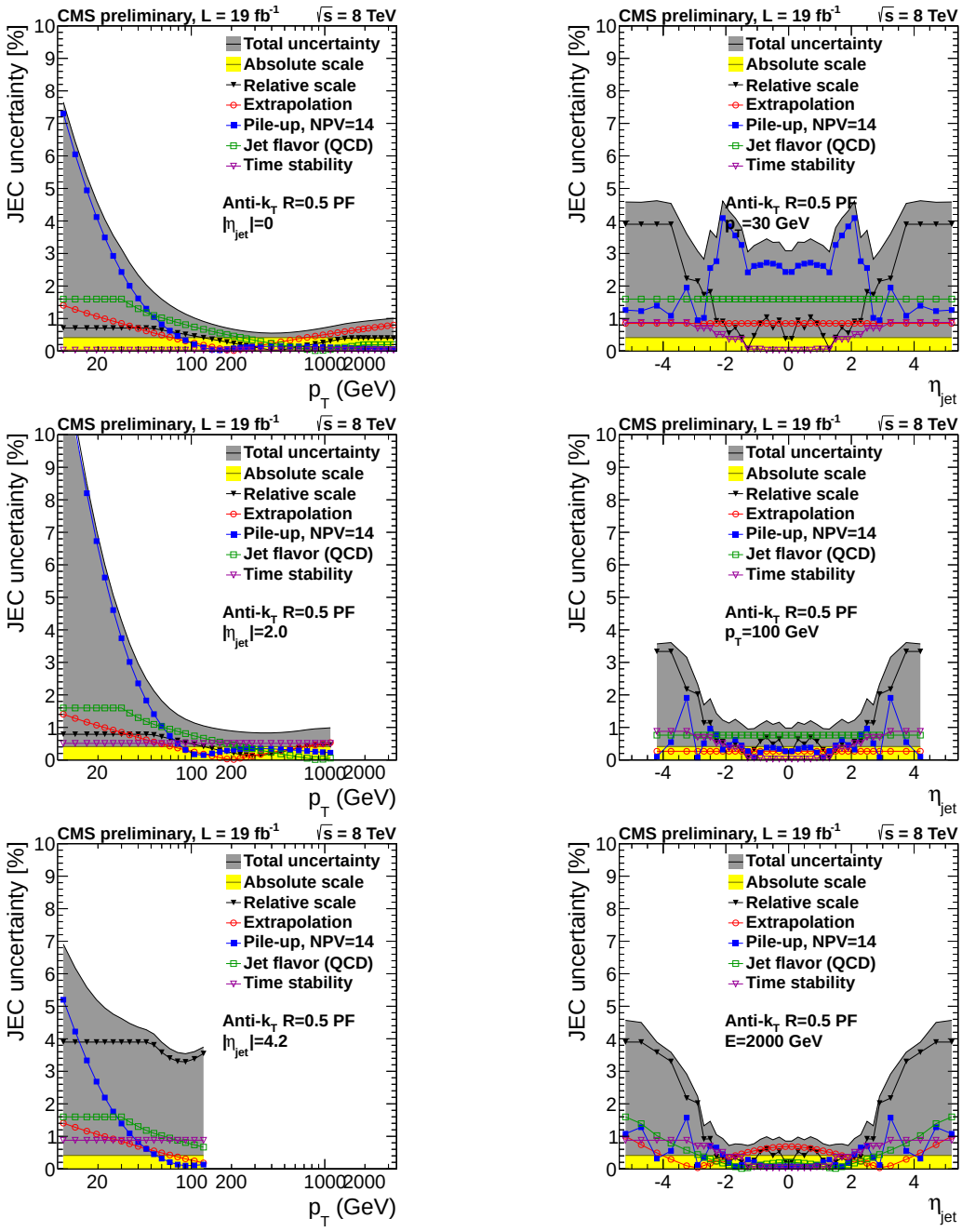


Figure 4.36: Sources of jet energy uncertainties and total uncertainties presented as cross sections through the p_{T}-η space. On the left, the uncertainties over the transverse jet momentum is shown in the barrel region ($\eta = 0$), in the endcaps ($\eta = 2$) and in the forward detector ($\eta = 4.2$). On the right, the uncertainties are shown as a function of the pseudorapidity for jets at $p_{\mathrm{T}} = 30\,\mathrm{GeV}$, $100\,\mathrm{GeV}$ and $2000\,\mathrm{GeV}$ [78].

The reach in p_T of the data-driven methods is limited statistically for high p_T values. The extrapolation uncertainty is estimated from the PYTHIA/HERWIG++ differences in the fragmentation of high energetic jets and the underlying event modelling using the full event simulation for both generators. In addition, an uncertainty of 3 % on the response of single pions in the ECAL and HCAL is propagated to reconstructed jets. The components of the absolute scale uncertainties are listed in Table 4.2. The total uncertainty on the absolute scale is only 0.42 %.

Table 4.2: Uncertainties on the absolute scale measurement.

Uncertainty source	2011	2012
Z/γ statistics	0.22 %	0.20 %
Reference scale	0.30 %	0.10 %
Method extrapolation	0.40 %	—
Flavour mixture	0.25 %	0.35 %

Flavour (7 components) Flavour uncertainties are determined from the differences between the Monte-Carlo generators PYTHIA with the Z2* tune and HERWIG++ 2.3. Separate uncertainties are estimated for each jet flavour (pure g-, uds-, c- and b-jets). In addition, the uncertainties on the flavour mixture in QCD events, γ+jet and Z+jet events are derived, the former being the default flavour uncertainty set. An alternative uncertainty can be chosen depending on the flavour mixture of a given analysis. The flavour uncertainties are depicted in Figure 4.37.

Time dependence (1 component) The observed decline in the jet response as discussed in Section 4.9.1 leads to a further source of uncertainty in the endcaps.

Correlations To account for correlations between the different methods and correction levels, the sources are grouped whenever they share methods or effects. These groups concern uncertainties on the MPF method that is shared by the data-driven correction levels and other inter-calibration correlations, flavour uncertainties which are common to all levels, especially uncertainties on the jet energy scale of b-jets.

Two previously considered sources are not used any longer: The sample uncertainties for the relative corrections are currently not used as the differences between dijets and Z/γ+jet samples are already covered by the uncertainty sources listed above. The same holds for the uncertainties of the pile-up estimation method in Monte-Carlo which is now covered by the p_T dependent pile-up uncertainties.

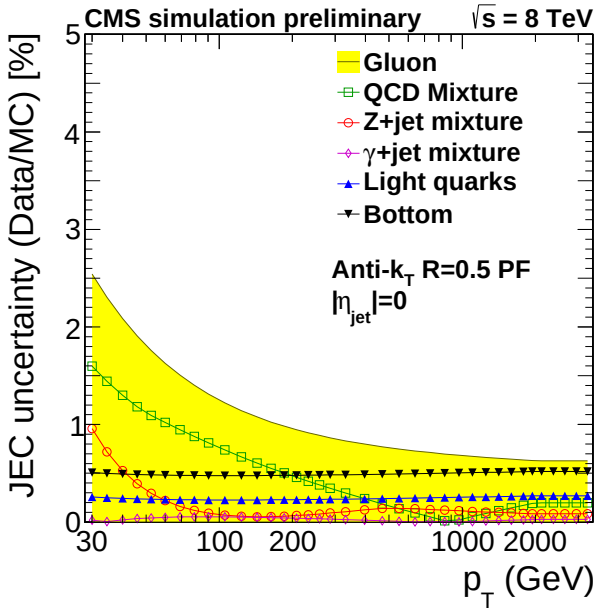


Figure 4.37: Flavour uncertainties on the jet energy scale as a function of the transverse momentum of the jet [78]. The uncertainty is given for individual flavours as well as for typical flavour mixtures of a selection of datasets.

4.13 Summary of the jet energy calibration

In CMS, the uncorrected jet energy differs from the true jet energy by up to 80 %. The combination of simulation based studies and robust data-driven corrections made it possible to calibrate the jet energy with unprecedented precision. While the jet energy scale could be determined with a precision of several percent when measured with the first collisions in 2010, it was quickly reduced to $1-2\,\%$ in the central barrel region of the detector in 2011. Such a precision of jet energy corrections was never achieved by other collider experiments on such a short time scale. The current uncertainties for a jet of $p_{\mathrm{T}} = 300\,\mathrm{GeV}$ in the barrel region are only 0.55 %.

This evolution of jet energy corrections during data taking is documented in Detector Performance Summaries [78, 79, 100, 101]. A full description of the jet energy calibration in the 8 TeV pp collision data is going to be published soon [102].

The data-driven residual corrections of the absolute jet energy scale exploiting the $Z(\rightarrow \mu\mu)+$jet topology was presented as part of the work of this thesis. These studies were the major contribution to the improvement of the jet energy scale measurements. The uncertainty of the corrections, derived from $Z(\rightarrow \mu\mu)+$ jet events, could be reduced by a factor of 4 from 2 % in 2010 to now 0.42 %. While the increase of statistical precision played a major role, this result would not have been possible with

out the application of several pile-up mitigation techniques. The original method was not prepared for the dramatic increase of the number of pile-up interactions as it rejects events with additional pile-up jets. However, with the suppression of and correction for pile-up effects, the measurement of the jet energy scale could was stable with respect to pile-up biases and was further improved. The detailed study of jet properties, performed for this thesis, was a key to reduce systematic uncertainties. The agreement of two complementary methods for measuring the jet response gives further confidence in the obtained result. The measurement of the flavour response in data, the time stability of the jet energy scale and the independent cross check of the calibration in the endcaps and the forward detector are further contributions of this work.

In the upcoming years, the calibration of jets in an environment, that is even more affected by pile-up interactions, will be the major challenge of this field. This involves selection criteria, isolation efficiencies, jet identification and most of all the pile-up contamination of jets. For this reason, the pile-up mitigation techniques are further developed and it is planned to combine the different methods to consistently remove the pile-up influence from all components of an event, including jets and missing energy.

Furthermore, multi-jet topologies are planned for the extension of the data-driven methods to the TeV scale. In addition, corrections of subjets are an emerging topic. Subjets can be seen as the hard parts of wider jets which typically result from hadronic decay products of a highly boosted object.

The detailed knowledge about jet physics, jet reconstruction and calibration is of utmost importance for a wide range of analysis, where the jet energy scale is often a major source of systematic uncertainty. In particular, the Higgs search in the ditau decay mode profited from the understanding of jets in the vector boson fusion production channel as described in the following chapter.

Chapter 5

The Search for the Higgs Boson in the H $\rightarrow \tau\tau$ Channel

The search for the Higgs boson was one of the main reasons for building the LHC. It was driving the design of the accelerator as well as the layout of the detectors. While the Higgs boson has been found in the bosonic decay modes $\gamma\gamma$, ZZ and WW [14, 15] in 2012, evidence for the fermionic decay of the Higgs boson was yet to be delivered. Due to the nature of Yukawa couplings, the search for Higgs bosons decaying into heavy fermions is particularly promising. While the top mass is above the threshold for a light Higgs boson, decays into a pair of b-quarks exhibit the largest fermionic branching ratio that is kinematically accessible. Huge irreducible backgrounds from heavy quark production prevent this channel from exploiting the main Higgs production channel: gluon fusion. Therefore, the search for $H \rightarrow b\bar{b}$ is mainly based on the vector boson fusion [103], the vector boson associated production [104] and top-pair associated production [105] modes. QCD, $t\bar{t}$ and vector boson + jet backgrounds make this search challenging and reduce the sensitivity compared to the $\tau\tau$ channel which is the channel of the next lighter fermion. The search for Higgs bosons decaying into τ pairs is dominated by backgrounds from electroweak processes. A large international collaboration was able to improve the understanding of the reconstructed physical objects, to refine the analysis techniques and to reduce the uncertainties in this channel such that we have now evidence for the fermionic Higgs decay [106] as will be presented in this chapter.

The first section of this chapter will introduce the $H \rightarrow \tau\tau$ search, whereas the second section concentrates on the vector boson fusion Higgs production process and the implication of jet measurements to it. The τ-lepton being an unstable particle, only its decay products can be analysed. Therefore, the search is split into analyses of the subchannels. The third section of this chapter describes the dedicated analysis of the $H \rightarrow \tau\tau \rightarrow \mu\mu$ decay mode to which this thesis made a major contribution. Finally, the combined results of the search for the Standard Model Higgs boson decaying into a pair of τ-leptons are presented.

5.1 Overview of the $H \to \tau\tau$ search

While the mass of the Higgs boson is not predicted by the Standard Model, the range of possible Higgs masses was restricted by theoretical constraints and previous measurements at the LEP and Tevatron colliders [40]. The width of the Higgs resonance is small with respect to its mass only for low Higgs masses ($m_{\mathrm{H}} \lesssim 400\,\mathrm{GeV}$). These constraints lead to a search window between 90 GeV and 1000 GeV with particular interest in the low mass region between 90 GeV and 145 GeV. Even after the discovery of a Higgs boson at 125 GeV, the analysis was pursued as a mass-independent search in the $H \to \tau\tau$ channel. In order not to bias the final result by drawing premature conclusions and by tuning the analysis for a maximal outcome given the observed data, the procedures of the analysis were fixed before looking at the final results in data.

5.1.1 Invariant mass of the ditau system

One of the most sensitive variables to identify $H \to \tau\tau$ events in the overwhelming backgrounds – $Z \to \tau\tau$ in particular – is the invariant mass of the ditau system $m_{\tau\tau}$. This mass can be approximated by the invariant mass of the visible decay products. This is only a rough approximation, since decays of τ-leptons always involve at least two neutrinos. In the semi-leptonic or fully leptonic channels, one or two additional neutrinos are produced in the decays. Compared to the visible mass, the scale and resolution of the reconstructed ditau mass can be improved by making assumptions on the neutrino kinematics and including the measured missing transverse energy in the reconstruction of $m_{\tau\tau}$.

One approach is the collinear approximation. It assumes that all neutrinos in tau decays point in the same spatial direction in the laboratory frame as the corresponding visible decay products. It is further assumed that all missing energy has its origin in the undetected neutrinos from tau decays. With these constraints one gets a system of equations that can be solved for the invariant ditau mass. However, this construction does not provide valid solutions for all configurations of the visible decay products and the $E_{\mathrm{T}}^{\mathrm{miss}}$.

SVFit mass The reconstruction mechanism used in this $H \to \tau\tau$ search is the SVFIT algorithm [106, 107] which is a likelihood-based approach identifying the most compatible mass given the kinematics of the decay products and the missing transverse energy. The four-momenta of the visible decay products can be measured as p_1^{vis} and p_2^{vis}. In addition, there are three degrees of freedom per τ decay left unconstrained (two in case of hadronic decays), which can be chosen to be: the energy fraction, x, carried by the visible decay products with respect to the energy of the τ-lepton, the azimuthal angle of the τ-lepton ϕ_τ and the invariant mass of the neutrino system $m_{\nu\nu}$

of that decay. The latter differs from zero only in leptonic decays. These parameters can be grouped into a vector $\vec{a} = (x, \phi_\tau, m_{\nu\nu})$ for each τ decay. Further information is provided by the two components of the missing transverse energy $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. The likelihood function f of the components of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is constructed given the unknown parameters $\vec{a}_1$ and $\vec{a}_2$ and the four-momenta of both visible decay products p_1^{vis} and p_2^{vis}. Using this model, the probability of the mass hypothesis $m'_{\tau\tau}$ can be calculated as

$$P(m'_{\tau\tau}) = \int \delta\big(m'_{\tau\tau} - m_{\tau\tau}(p_1^{\mathrm{vis}}, p_2^{\mathrm{vis}}, \vec{a}_1, \vec{a}_2)\big) \cdot f(p_1^{\mathrm{vis}}, p_2^{\mathrm{vis}}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}; \vec{a}_1, \vec{a}_2)\, \mathrm{d}\vec{a}_1\, \mathrm{d}\vec{a}_2 \tag{5.1}$$

The result of the SVFIT algorithm is the value of $m'_{\tau\tau}$ that maximises this probability.

The likelihood function $f(p_1^{\mathrm{vis}}, p_2^{\mathrm{vis}}, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}; \vec{a}_1, \vec{a}_2)$ is composed of three factors. One for each of the two τ decays and a third factor to quantify the compatibility of a certain τ decay configuration with the reconstructed missing transverse energy. Due to the different number of final state particles, leptonic and hadronic τ decays have a different likelihood function.

The leptonic τ decay is a three-body decay with two massless neutrinos. The differential distributions are modelled using matrix elements for unpolarized τ decays [108]. The likelihood for such a decay is thus given by

$$\mathcal{L}_{\tau\to\ell}(\vec{a}) = \frac{\mathrm{d}\Gamma}{\mathrm{d}x\mathrm{d}\phi_\tau\mathrm{d}m_{\nu\nu}} \propto \frac{m_{\nu\nu}}{4m_\tau^2}(m_\tau^2 + 2m_{\nu\nu}^2)(m_\tau^2 - m_{\nu\nu}^2) \tag{5.2}$$

within the physically allowed region $0 \leq x \leq 1$ and $0 \leq m_{\nu\nu} \leq m_\tau\sqrt{1-x}$. Besides the explicit dependence on $m_{\nu\nu}$, the integration boundaries lead to an implicit dependence on x.

For hadronic decays, the visible decay products are reconstructed as a single object called the hadronic tau τ_{h}. The differential width of this two-body decay is given by [1]

$$\mathcal{L}_{\tau\to\tau_{\mathrm{h}}}(\vec{a}) = \frac{\mathrm{d}\Gamma}{\mathrm{d}x\mathrm{d}\phi_\tau} \propto \frac{1}{1 - m_{\mathrm{vis}}^2/m_\tau^2} \tag{5.3}$$

which is defined within the physically allowed region $m_{\mathrm{vis}}^2/m_\tau^2 \leq x \leq 1$.

The neutrinos from the τ decays are assumed to be the only source of reconstructed missing energy in the event. The compatibility of a decay hypothesis with the measured missing energy is quantified by the likelihood function

$$\mathcal{L}_\nu(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}) = \frac{1}{2\pi\sqrt{\det V}} \cdot \exp\left[-\frac{1}{2}(\vec{E}_{\mathrm{T}}^{\mathrm{miss}} - \textstyle\sum \boldsymbol{p}_{\mathrm{T}}^{\nu})^{\top} V^{-1} (\vec{E}_{\mathrm{T}}^{\mathrm{miss}} - \textstyle\sum \boldsymbol{p}_{\mathrm{T}}^{\nu})\right] \tag{5.4}$$

where the covariance matrix V represents the expected $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution as estimated in [109].

5.1.2 Decay channels

The analysis is split into decay channels according to the final state particles of the Higgs decay reconstructed from the detector information. The three options for each τ (τ_h, μ, e) lead to six independent channels for the analysis: $\tau_h\tau_h$, $e\tau_h$, $\mu\tau_h$, $e\mu$, ee and $\mu\mu$, which all have their dedicated analysis strategy. As a measure of the sensitivity of each decay channel, the expected limits at 95 % CL are shown for different mass hypotheses in Figure 5.1. The ordering of the decay channels in sensitivity is nearly

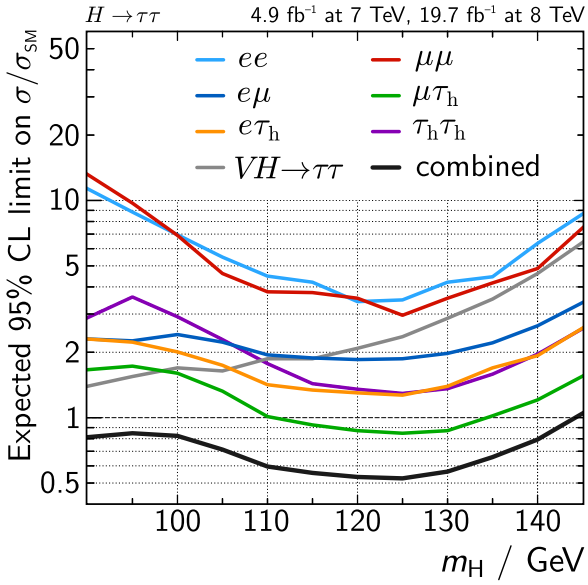


Figure 5.1: Expected limits of the decay channels in the $H \to \tau\tau$ search as a function of the probed mass of the Standard Model Higgs boson, m_H. This serves as a measure for the sensitivity in each channel. It shows the ordering from the most sensitive semi-leptonic channels to the same-flavour lepton channels.

kept over the whole investigated mass range. The expected limit on the Higgs boson production cross section for a Higgs mass of 125 GeV for each decay channel is also listed in Table 5.1 together with the branching ratios, $\mathcal{B}$, and the dominant source of background events. In addition, the vector boson associated Higgs production (VH) is addressed by a dedicated group of analyses as explained in [106]. It is not covered by the following description. A detailed explanation can be found in [110].

In case both τ-leptons decay hadronically ($\tau_h\tau_h$), QCD multijet events containing jets misidentified as τ decay products are the dominant background. The integrated luminosity used for the analysis of this final state is smaller than for the following channels due to prescaling of the available trigger menus. In addition, the successful rejection of the QCD background, in case one τ-lepton decays to a lighter lepton,

makes the sensitivity of the $e\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels surpass the full-hadronic decay mode despite the lower branching ratio. The dominant background in these two semi-leptonic channels arises from $Z \to \tau\tau$ events with the same final state. The same holds for the $e\mu$ channel which is the most sensitive dilepton channel. It has to cope with the reduced branching ratio of the dilepton final states. The $t\bar{t}$ background is reduced using a multivariate method exploiting the following variables: the kinematic quantities of the $e\mu$ system, the missing transverse energy, the distance of closest approach between the leptons and the primary vertex and the b-tagging discriminant for the leading jet with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. This channel vetoes same flavour lepton events and can therefore drastically suppress Drell-Yan production of light leptons. This is not possible for the same flavour lepton channels $\mu\mu$ and ee which makes the search in these channels particularly challenging in addition to the small branching ratio of about 3 %. The precise muon reconstruction and the differences in the acceptance lead to the slightly higher sensitivity of the $\mu\mu$ channel compared to the otherwise very similar treatment of the ee final state. The combined expected limit is a factor of two better than the most sensitive decay channel, $\mu\tau_{\mathrm{h}}$.

5.1.3 Event categories

After a preselection of events, that is unique for each channel, all accepted events are sorted into mutually disjoint categories. This categorisation maximises the sensitivity of the combined result by enhancing the signal-to-background ratio in some categories while preserving all signal events after the preselection. This is achieved by designing the categories such that they account for features of the Higgs production processes.

A common classification criterion across all channels is the jet multiplicity. The jet multiplicity of all valid reconstructed jets is 35 on average and shows large tails to a maximum at around 90 jets per event. Jets that enter the multiplicity criteria of

Table 5.1: The subchannels of the $H \to \tau\tau$ search ordered by branching ratio as given in [1]. The expected limit on the cross section in multiples of the expected Standard Model Higgs boson production cross section is given for $m_H = 125\,\mathrm{GeV}$. well as the dominant background for each final state.

Channel	Branching ratio	Expected limit	Dominant background
$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$	41.94 %	1.29	QCD
$e\tau_{\mathrm{h}}$	23.09 %	1.27	$Z \to \tau\tau$
$\mu\tau_{\mathrm{h}}$	22.54 %	0.848	$Z \to \tau\tau$
$e\mu$	6.21 %	1.87	$Z \to \tau\tau$
ee	3.18 %	3.48	$Z \to ee$
$\mu\mu$	3.03 %	2.96	$Z \to \mu\mu$

the $H \to \tau\tau$ search are required to be within the detector acceptance and above a transverse momentum threshold which is imposed against the vast amount of low energetic pile-up jets,

$$p_{\mathrm{T}}^{\mathrm{j}} > 30\,\mathrm{GeV}, \qquad |\eta_j| < 4.7\,.$$

These cuts reduce the number of jets to the range of 0 to 2 in most events.

2-jet The 2-jet category contains all events with at least two jets. This aims at the VBF production channel with its characteristic two jets at high rapidities which are also called *tagging jets*. A strict distinction of the production processes is not possible. Roughly half of the expected signal in this category is still produced by gluon fusion. In this category, dijet quantities are defined as

Invariant mass of the dijet system	$m_{jj} \equiv m_{j_1,j_2}$
Rapidity gap between the two jets	$\Delta\eta_{jj} \equiv \lvert\eta_{j_1} - \eta_{j_2}\rvert$

where j_1 and j_2 are the leading and second leading jets in the event.

1-jet Events in which the Higgs boson recoils to a jet are another advantageous configuration. The fraction of those events is enhanced by requiring at least one jet in the event. They are mostly produced via gluon fusion with an additional radiation of a high p_{T} gluon in the fermion loop or from initial or final state radiation. Such boosted events exhibit a better resolution of the mass variables and the missing transverse energy. This category can also contain events with more than one jet, in case of stricter VBF requirements for the 2-jet categories in addition to the jet multiplicity criterion of some channels.

0-jet All preselected events without any jet belong to the 0-jet category. It is dominated by large backgrounds and mostly used for calibration purposes and constraints on the backgrounds.

All but the full-hadronic channel follow this scheme and in addition split the 0- and 1-jet categories according to the transverse momentum of one of the leptons. In many channels, this basic set of categories is further subdivided: The mixed-flavour channels exploit the variables $\Delta\eta_{jj}$ and m_{jj} for loose and tight VBF categories for events containing two jets. Events that do not fulfil neither of those requirements are passed on to the 1-jet categories. Channels with at least one hadronically decaying τ-lepton introduce boosted categories by requiring a certain minimal transverse momentum of the reconstructed ditau system given by

$$p_{\mathrm{T}}^{\tau\tau} = |\vec{p}_{\mathrm{T}}^{\,\tau_{\mathrm{h}}} + \vec{p}_{\mathrm{T}}^{\,L} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}|\,.$$

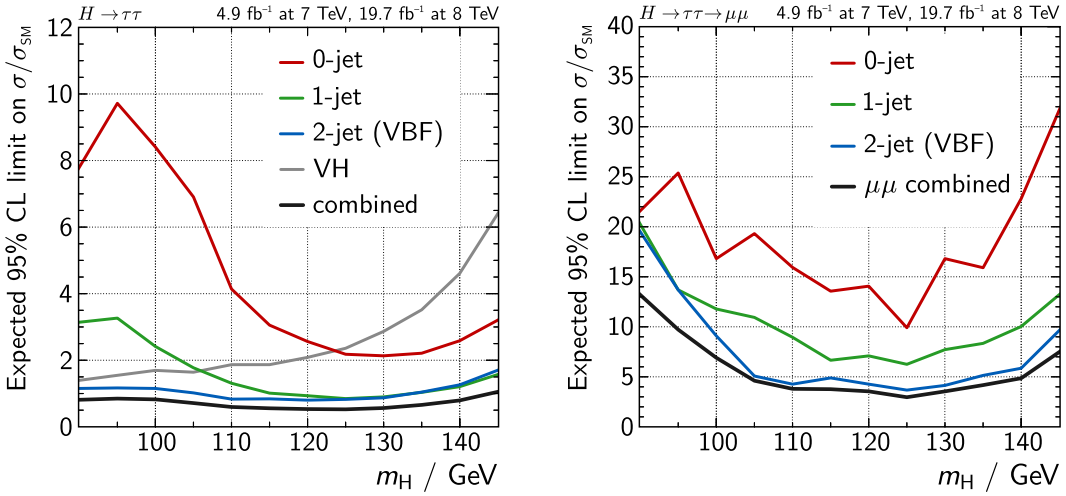


Figure 5.2: Expected limits of the event categories of the $H \to \tau\tau$ search. The left plot shows the expected limits for all categories combined across all τ decay channels. The curve of the 0-jet category shows the sensitivity of a pure counting experiment which is less sensitive for a Higgs boson mass around the Z boson mass. The 2-jet category is the most sensitive category over nearly the whole probed mass range. The right plot shows the expected limits for the categories of the $H \to \tau\tau \to \mu\mu$ channel where again the 2-jet category is the most sensitive.

A complete description of the categories containing further subtle differences can be found in [106].

To compare the sensitivity of the event categories, Figure 5.2 shows the expected limit for the categories of the $\mu\mu$ decay channel. The sensitivity of the 0-jet category is the one of a nearly pure counting experiment which looses sensitivity around the Z boson mass peak. Compared to this, the selection of the 1- and 2-jet categories improves the sensitivity drastically.

While in other Higgs decay modes the evidence for the VBF topology is yet to be established, the 2-jet category is the most sensitive in all channels of the $H \to \tau\tau$ search. The two tag jets are a distinctive feature of the signal events while the signal-background separation is much more difficult for Higgs events produced via gluon fusion, because the production process leaves no other detectable traces in the event but the decay products of the Higgs boson candidate.

5.2 Jets in vector boson fusion $H \to \tau\tau$ events

The high sensitivity of the VBF (2-jet) categories mainly results from the clear signature of the corresponding Higgs production topology. This emphasizes the particular interest in this production mode and its properties. While a clear phenomenological isolation of the VBF signal is not possible [111], increasing the signal over background ratio in these categories is crucial for improving the significance of the result. These relatively strict requirements for the 2-jet category typically lead to a smaller number of events compared to the other categories. As an example from the $\mu\mu$ channel, which is discussed in the next section, the 2-jet category contains 197 655 events in data compared to 6 839 183 events in the other categories. At the same time, the number of the classified events has to remain sufficiently large to keep the statistical precision. Furthermore, it must be guaranteed that the categorisation criteria can be determined with adequate precision. As the jet multiplicity is the main characteristic of the VBF selection, the jet energy scale has a sizeable impact on the event classification. The uncertainty on the jet energy scale can result in a change in acceptance of up to 20 % as measured by propagating the jet energy uncertainties through the full analysis. It is thus the second largest systematic uncertainty concerning all signal and background processes – comparable only to the τ energy scale. This requires a thorough understanding of jets at high rapidities and in-situ corrections of the jet energy if necessary.

5.2.1 Jet energy corrections in VBF events

The jet energy scale in VBF events can be evaluated using data-driven calibration techniques as presented in Chapter 4. The jet response was studied for all η regions within the acceptance of the detector using $Z(\to \mu\mu) + \text{jet}$ events in addition to the CMS wide determination of the jet energy scale explained in the previous chapter.

Time dependence of the jet energy scale in the endcaps

The systematic shift in the jet energy scale over data taking periods, shown in Section 4.9.1, reduces the acceptance in data significantly. While the acceptance in the first reconstruction of collision data was reduced by 5 %, this effect was partly accounted for by a recalibration of the hadronic calorimeter in the re-reconstructed events that were finally used for this analysis. However, the time dependence did not fully vanish. Therefore, dedicated corrections, based on number of the LHC run n_r in which the event was recorded, are applied that remove the acceptance loss during 2012 the data-taking period. The additional residual correction factor

$$c_{\text{time}} = 1 + c_\eta \cdot (n_r - n_r^0) \cdot f_{\text{NHad}}^{\text{jet}} \tag{5.5}$$

is applied to the jet p_{T} as a linear function of the run number n_r and the energy fraction of neutral hadrons $f_{\mathrm{NHad}}^{\mathrm{jet}}$ in the jet. The constant c_η is derived in η bins according to the towers of the hadronic calorimeter. It is positive for most η regions. n_r^0 denotes the run with a good a priori calibration and hence a vanishing time dependence correction.

The observed time dependence of the jet response was corrected in the same-flavour lepton channels of this analysis. Apart from this effect no significant deviations from the correction factors and uncertainties presented in Chapter 4.2 appeared. This also confirms the applicability of general jet energy correction to different event topologies.

Jet response in a two-jet balance

The calibration methods described in Chapter 4 were applied so far for in-situ measurements and cross checks of the jet energy scale in the η ranges of typical VBF tagging jets. As a further cross check, the jet energy scale can be examined in a dedicated event selection for the VBF Z boson production. This way, the jet measurement can be studied directly using the events that enter the 2-jet category.

These cuts follow the typical VBF criteria and are close to most 2-jet categories of the $H \to \tau\tau$ channels. Apart from the cuts on the two tagging jets, a central jet veto rejects events with a third jet within the rapidity gap between the tagging jets.

$$\begin{aligned} p_{\mathrm{T}}^{\mathrm{j1,j2}} &> 30\,\mathrm{GeV} \\ p_{\mathrm{T}}^{j3} &< 30\,\mathrm{GeV} \qquad |\eta_{\mathrm{j3}} - \frac{\eta_{\mathrm{j1}} - \eta_{\mathrm{j2}}}{2}| < \frac{1}{2}|\eta_{\mathrm{j1}} - \eta_{\mathrm{j2}}| \end{aligned}$$

The size of the rapidity gap η_{jj} and the invariant mass of the dijet system are also considered by the cuts

$$\begin{aligned} \Delta\eta_{jj} &> 4 \\ m_{jj} &> 500\,\mathrm{GeV}\,. \end{aligned}$$

The Z boson reconstruction and selection is identical to Section 4.7.2, except that there is no minimum requirement for the transverse momentum of the Z boson, in order to increase the statistical significance of the study.

The MPF method can be employed unchanged as the full recoil to the reference object is used for the measurement. On the other hand, the p_{T} balance method has to be modified. The balancing technique employed here is similar to the one in Chapter 4 but extended to two recoiling jets as an adaptation to the VBF topology. These two jets are the leading ones, identical to the tagging jets of the VBF selection. The p_{T} balance method for this study reads

$$R_{\mathrm{bal}} = \frac{p_{\mathrm{T}}^{\mathrm{jj}}}{p_{\mathrm{T}}^{\mathrm{Z}}} \qquad \text{with } p_{\mathrm{T}}^{\mathrm{jj}} = (p_{j1} + p_{j2})_{\mathrm{T}}\,.$$

The central jet veto replaces the cut on additional jets. An extrapolation to zero additional jet activity is not possible due to the comparably low number of events.

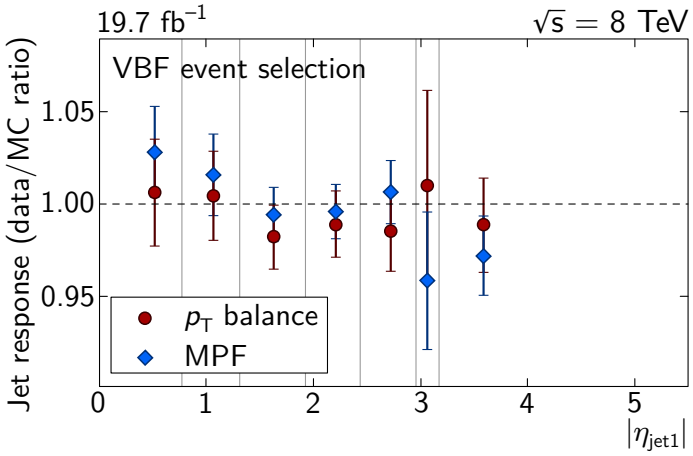


Figure 5.3: Data-driven measurement of the jet response in VBF-like events. The data/MC ratio is compatible with unity over the whole pseudorapidity range, when including the time dependence corrections for the region $2.5 < |\eta| < 3.0$, where the transparency loss in data would otherwise result in a lower response by 5 %.

Figure 5.3 shows the data/MC ratio of the jet response measured using these two methods over the pseudorapidity of the leading jet. The absolute value of η is shown as the response ratio is symmetric. The data/MC ratio is compatible with unity over almost the whole pseudorapidity range. In the region $2.5 < |\eta| < 3.0$, the aforementioned transparency loss in data results in a lower response ratio in this interval. The additional time dependence corrections solve this issue.

The low number of events in this selection results in large statistical uncertainties. In addition, no radiation corrections can be applied. The total uncertainty on this response measurement amounts to 1.5 % to 4.5 % which is large compared to the per mil level measurements in the previous chapter.

A correct description of the jet energy scale and jet physics in general should result in a correct description of jet variables by the simulation. In particular the different detector components, their response and their behaviour with respect to pile-up, are known to affect the jet selection as demonstrated before.

This can be seen in the pseudorapidity distributions of both tagging jets, shown for the 2-jet category of the $\mu\mu$ channel in Figure 5.4. Over a wide range, the description is very precise, especially in the barrel region ($|\eta| < 1.3$) in which the jets are calibrated using $Z(\to \mu\mu)$ + jet based residual corrections. On the other hand, it also reveals the suboptimal description of jets in the problematic part of the endcap outside the tracker coverage even after applying all aforementioned corrections. The

description of jets in the rapidity gap of 2-jet events are discussed in Section 5.3.6.

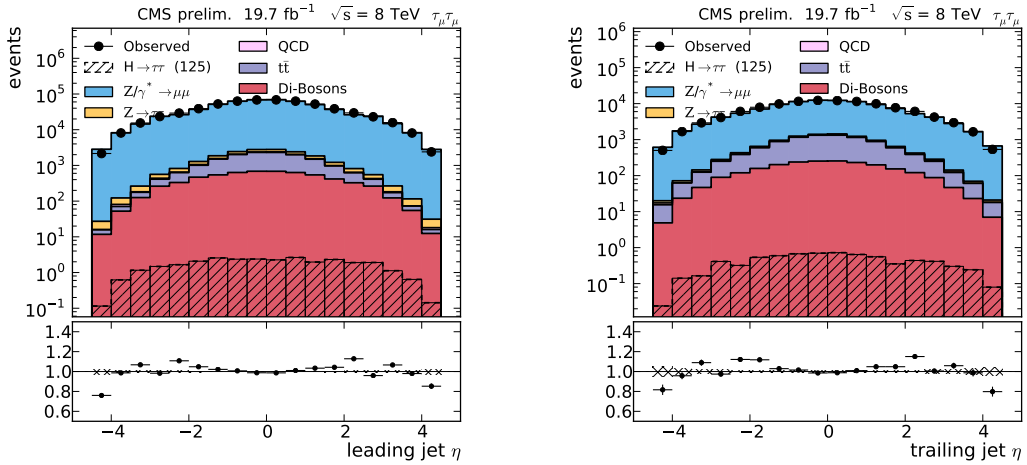


Figure 5.4: Pseudorapidity distributions of the leading and second leading jet in the 2-jet category.

5.3 The $H \to \tau\tau \to \mu\mu$ channel

This section describes the search for the Standard Model Higgs boson decaying into a pair of oppositely charged τ-leptons with a subsequent τ decay into muons [112]. The Higgs search in this channel has been prepared over many years and the analysis techniques have been optimised continuously [113–116].

The other same-flavour lepton channel ee is not discussed in this section. It uses equivalent analysis techniques for electrons in an otherwise identical procedure. In particular, the improvements on the reconstruction and identification of jets shown here are exploited in the ee channel, as well.

With respect to other channels, the $\mu\mu$ decay mode profits from the precise muon triggering and measurement and the absence of hadronic Higgs decay products, particularly in the rapidity gap of a VBF produced Higgs boson. However, the expected significance of the dilepton channels is lower than in others due to the low branching ratio (3 %) and large irreducible background as can be seen in Figure 5.1. The latter arises from the fact that Drell-Yan background can not be rejected by an opposite charge lepton veto. In addition, the four neutrinos in the final state reduce the resolution of the ditau mass and therefore complicate the distinction between Z and Higgs boson decays in addition.

5.3.1 Recorded CMS data and simulated events

As for all channels of the $H \to \tau\tau$ search, the full amount of data, recorded by the CMS detector in 2011 and 2012, is used. It corresponds to an integrated luminosity of $4.9\,\text{fb}^{-1}$ and $19.7\,\text{fb}^{-1}$, respectively, determined by the pixel-luminosity measurements [53]. The selection of valid events follows the CMS recommendations, the dataset names can be found in Appendix A.1.1. Since both data-taking periods differ in their centre-of-mass energy, they are analysed separately, followed by the statistical analysis of each individual channel and all channels combined. For the first part of 2011, the IsoMu17 trigger requiring an isolated muon with $p_\text{T} > 17\,\text{GeV}$ was unprescaled and could be used. With the increasing luminosity, this analysis switched to double muon triggers with a minimum p_T requirement of 13 GeV (8 GeV) for the leading (second leading) muon. The same double muon triggers as in Section 4.5 are used to identify pairs of muons in collision data at $\sqrt{s} = 8\,\text{TeV}$.

Events of all signal and background processes with a relevant contribution after the event selection are simulated using Monte-Carlo generators. The generated events are then passed on to the CMS detector simulation using GEANT4 [117, 118]. Simulated inelastic pp collisions are added to model the effect of pile-up in these events following a rough forward projection of the expected pile-up distribution.

Event weights Weights are applied to the events of all simulated samples for the following reasons: For a precise modelling of the pile-up distribution, all simulated events are weighted according to the a posteriori determined number of pile-up vertices to match the distribution expected in data as explained in Section 4.5.1. The total inelastic cross section of proton-proton scattering used for this analysis is $\sigma_\text{minbias} = 68.0\,\text{mb}$ at 7 TeV and 69.4 mb at 8 TeV.

Further weights are applied to normalize the simulated samples to the expected event yield. The latter is estimated from the cross section of the process σ, efficiencies of filters applied during the event generation ε_MC and the total integrated luminosity of the respective run period $\mathcal{L}$. In addition, weights are applied to account for trigger efficiencies in the same way as in Section 4.5. The product of these weights

$$w = w_\text{PU} \cdot w_\text{trigger} \cdot \frac{\mathcal{L} \cdot \sigma \cdot \varepsilon_\text{MC}}{n_\text{sample}}$$

is applied per event.

Signal modelling

According to the Standard Model, the Higgs boson can be produced in the four main production channels at the LHC as described in Section 2.7.1. The gluon fusion and the VBF production have the largest expected cross sections. These two processes are

generated by POWHEG [88, 119–121] followed by the parton shower and hadronization of PYTHIA [90], which also accounts for initial and final state radiation.

For completeness, the associated production with vector bosons or top pairs is simulated using PYTHIA. Theoretical uncertainties affecting the production of simulated signal events are studied in Section 5.3.9.

5.3.2 Background processes

Due to the small Higgs boson production cross section and the small branching ratio of the $H \to \tau\tau \to \mu\mu$ channel, the reduction of a long list of background processes constitutes a major challenge of this analysis. Except for the QCD background which is generated by PYTHIA, all background processes are generated using MADGRAPH.

Since the expected yield of signal events is small compared to the number of events from background processes, a precise modelling of all backgrounds is a key in the strive for maximal sensitivity. While the smaller backgrounds are sufficiently well described by the event generators, the modelling of all major backgrounds needs to be improved by data-driven methods. The main backgrounds of the $H \to \tau\tau \to \mu\mu$ analysis are:

Drell-Yan muons In contrast to other $H \to \tau\tau$ channels, $Z \to \ell\ell$ represents the largest background in the same-flavour lepton channels. As same-flavour leptons carrying opposite charges are also part of the signal signature, vetoes on the presence of leptons with the same flavour can not be applied. The huge fraction of events containing Drell-Yan muons requires a thorough data-driven calibration of the expected dimuon yield. This is achieved with the help of the finite lifetime of the τ-lepton as will be explained in Section 5.3.7 after the description of all variables and analysis techniques.

Drell-Yan τ-leptons The equivalent process to pairs of τ-leptons, is the most akin to the signal process and thus the only irreducible background. To make maximal use of data themselves when modelling this background, the embedding technique was developed [110, 113]: This method exploits $Z \to \mu\mu$ events recorded from collision data. The muon pair and its detector response in these events are replaced by simulated τ-leptons and their decay products, while keeping the rest of the event, including the underlying event and pile-up activity. TAUOLA [122, 123] is used for the simulation of the τ decay, followed by the CMS detector simulation. Two variants have been studied, the replacement on the level of reconstructed muon hits and on the level of reconstructed Particle Flow candidates. Exploiting lepton universality, the $Z \to \tau\tau$ yield can be measured from the yield of $Z \to \mu\mu$ in respective control regions. Weights are applied to correct for the differences.

Top pairs Top quarks decay into b-quarks under the emission of a W boson. In $t\bar{t}$ events, the subsequent decay of each of the W bosons into a muon and its neutrino results in the detection of a muon pair of opposite charges. While the shape of the final distributions can be taken from the simulation, the normalization is fitted in a $t\bar{t}$ enriched side-band region containing events with large missing transverse energy. Events containing b-jets are vetoed in the event selection, reducing the $t\bar{t}$ background significantly.

QCD Even though muons are extremely rare in QCD processes, the sheer amount of strong interactions at the LHC results in a non-negligible QCD background. In addition it is difficult to model. In the simulated samples, the fraction of events containing muons is enriched by a factor 2.7×10^3 by imposing a muon cut already during Monte-Carlo event generation. But as even this sample does not contain a sufficient number of events to monitor this background correctly, the distribution of this background has to be taken from data. To this end, events with two muons carrying the same charge are selected to determine the shape of QCD events. The isolation criteria have to be loosened to increase the statistical precision. A scale factor, taken from Monte-Carlo simulation, accounts for the ratio of the number of same-charge with respect to opposite-charge events.

Electroweak A small background arises from electroweak diboson production. WW events in which each W boson decays to a muon and the corresponding neutrino can pass the preselection, as well as semi-leptonic ZZ decays. Being a significant background in channels with one hadronically decaying τ-lepton, the background of W + jet events is small in the $H \to \tau\tau \to \mu\mu$ channel. Both the shapes of the distributions and the normalization of this background are taken from the simulation at NLO.

5.3.3 Reconstruction of physical objects

The definition of muons and jets in this analysis is identical to the one used in Chapter 4. This includes all corrections and improvements developed there. The muon scale is corrected according to Section 4.6.2.

The advanced understanding of jet physics, gained from $Z(\to \mu\mu)$ + jet events, has been exploited in this analysis. If this analysis used standard jets, pile-up jets would fill the rapidity gap of VBF events. This would lead to a misclassification of about 60 % of genuine VBF Higgs events as non-VBF by the use of a central jet veto. The pile-up treatment using Charged Hadron Subtraction (CHS) removes most of the pile-up particles in the central detector region which represents the largest part of a typical VBF rapidity gap. Rejecting jets identified as emerging from pile-up interactions (Section 4.4.2) reaches even outside the tracker acceptance. These

methods for pile-up mitigation allow to use a central jet veto for the 2-jet category The reconstruction methods and the validation criteria are documented in [124].

In contrast to the corrected $E_{\mathrm{T}}^{\mathrm{miss}}$ used for jet calibration purposes, the missing transverse energy is corrected using a multivariate regression technique. To this end, five variables are calculated from Particle Flow candidates that allow to identify sources of mismeasurement. These variables are defined as the negative vectorial sum of different sets of particles: all Particle Flow candidates (this corresponds to the raw $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$), all tracks associated with the main vertex, charged Particle Flow candidates from the main vertex as well as neutral candidates that are clustered into a jet which is not identified as pile-up jet, charged candidates that are associated with the main vertex as well as neutral candidates from pile-up jets and finally tracks associated with the main vertex as well as which are not clustered into a pile-up jet. This method implicitly corrects for the influences by jet mismeasurement that were explicitly addressed by the Type-I corrections. This correction improves the resolution of the missing transverse energy at high pile-up by 40 % [124].

Subsequently, the description of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is further improved by an in-situ recoil calibration in the analysis [112]. This calibration leads to an agreement of the measured missing energy in data and simulation. It follows the method of the data-driven CMS $E_{\mathrm{T}}^{\mathrm{miss}}$ corrections [109] using $Z \to \mu\mu$ events.

5.3.4 Preselection of events

The muon selection follows the cuts given in Section 4.7, except for the looser cut on the second muon: $p_{\mathrm{T}}^{\mu_2} > 10\,\mathrm{GeV}$. The muons in signal events originate from a secondary vertex close to the main primary vertex. Therefore, the muon is required to be close to the main vertex. The transverse impact parameter has to be $|d_{xy}| < 0.2\,\mathrm{mm}$ and the longitudinal one $|d_z| < 2\,\mathrm{mm}$. Instead of being based on the tracks of charged particles, the muon isolation requirement is based on the particle flow candidates. The isolation variable

$$i_{\mu}^{\mathrm{PF}} = \sum_{\Delta R<0.4} p_{\mathrm{T}}^{\mathrm{CHad}} + \max\left[\sum_{\Delta R<0.4}\left(p_{\mathrm{T}}^{\gamma} + p_{\mathrm{T}}^{\mathrm{NHad}}\right) - \Delta\beta, 0\right]$$

is calculated and required to be $i_{\mu}^{\mathrm{PF}}/p_{\mathrm{T}}^{\mu} < 0.1$ for muons in the barrel and < 0.15 for muons in the endcap. The additional term, $\Delta\beta$, corrects for the pile-up contribution to neutral particles.

5.3.5 Event categories

Events passing the preselection are classified by jet multiplicity and the transverse momentum of the leading muon. A schematic overview of the five mutually disjoint categories of the $\mu\mu$ channel is given in Figure 5.5.

In addition to the common requirement of all channels for the 2-jet category to contain at least 2 jets, the same-flavour lepton channels impose a central jet veto. Apart from the two tagging jets, no further jet with a transverse momentum $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is allowed in between the two tagging jets. This is possible because the pile-up mitigation techniques successfully remove 98 % of the pile-up jets within the rapidity gap. This enhances the fraction of VBF produced Higgs boson events in this category by about 10 %. Vetoed events are passed to the 1-jet category. However, the fraction of multijet events in the 1-jet category remains low. As for the other channels, all preselected events without any jet belong to the 0-jet category.

The latter two categories are split into subcategories according to the transverse momentum of the leading muon at $p_{\mathrm{T}}^{\mu_1} = 35\,\mathrm{GeV}$. This distinction enhances the signal fraction in the high p_{T} categories, as at least the leading muon from a resonant $H \to \tau\tau \to \mu\mu$ decay carry a higher transverse momentum on average, compared to muons from $Z \to \mu\mu$. The sensitivity of the categories increases with the jet multiplicity, as shown in Figure 5.2, and with the leading muon p_{T} for the same jet multiplicity.

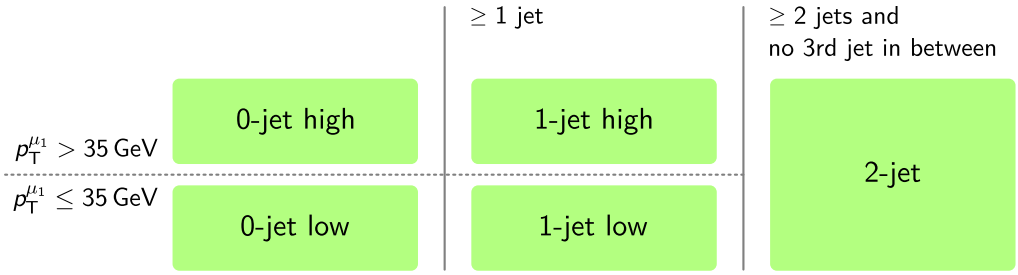


Figure 5.5: Schematic overview of the five event categories of the $H \to \tau\tau \to \mu\mu$ analysis. The events are classified by jet multiplicity. The 0-jet and 1-jet categories are further split according to transverse momentum of the leading muon, $p_{\mathrm{T}}^{\mu_1}$.

5.3.6 Multivariate analysis

Within the above-defined categories multivariate analysis methods MVA are used to improve the sensitivity beyond the reach of cut based analyses. The method of choice is the method of Boosted Decision Trees (BDTs) using the ROOT TMVA [125] implementation. Two separate BDT analyses are performed for the 0- and 1-jet categories and for the 2-jet category, respectively. These BDTs use different sets of variables, each dedicated to the corresponding Higgs boson production process.

For both MVA analyses, a set of variables is chosen that is suited to discriminate events containing Higgs bosons from the background. These background events mainly contain Z bosons with similar properties as the expected Higgs boson. The

variables comprise mass variables, angular and kinematic variables. Most quantities are derived from muon properties as they can be precisely reconstructed and are very accurately modelled in the simulation. This avoids pile-up influences and the uncertainties on parton showers and hadronization models. The distributions of these variables are shown in Figure 5.6 and 5.7. While the same steps were performed with 7 TeV data accordingly, the distributions in this chapter all show 8 TeV data only.

The following list contains a description of the variables that are common in both MVA analyses. It starts with the two mass definitions which provide the most sensitive variables for the $H \to \tau\tau$ search. The problem of having nearly identical positions of the peak in both signal and background events, illustrates the challenges of the $\mu\mu$ channel.

- The invariant mass of the dimuon system $m_{\mu\mu}$ is the first approximation to the mass of the resonant decay. Being reconstructed only from the visible decay products in case of signal events, the mass peak is shifted to lower values due to the missing contribution of the four neutrinos. The same holds for the $Z \to \tau\tau$ background where the mass peak of the Z boson is shifted. In the largest background, $Z \to \mu\mu$, the dimuon mass closely follows the peak of the Z boson. The same can be seen in the diboson background (electroweak) which is about $2 - 3$ orders of magnitude smaller.

- The invariant mass of the reconstructed ditau system $m_{\tau\tau}$ is calculated using the SVFIT algorithm (Section 5.1.1). In signal events, the shape of the ditau mass distribution nearly follows a Gaussian distribution around the given Higgs boson mass with a width of about 20 GeV.

- The DCA significance $\log_{10}$ DCAsig(μ, μ) is defined as the decimal logarithm of the significance of the distance of closest approach (DCA) which is the spatial distance between the two measured muon tracks. It is typically small for prompt muons from a Z decay, where both particles originate from the same vertex. In this case, the distance depends only on the track resolution. In the case of muons from τ decays, the fact that τ-leptons have a mean decay length of $c\tau = 87.18\,\mu\text{m}$ typically results in a measurable distance between the two muon tracks.

- The angular variable $\cos\Omega^*$ is defined as the cosine of the angle between the anti-muon three-momentum and the dimuon decay plane in the dimuon rest frame. The dimuon decay plane is given by the three-momentum of the dimuon system and the z-axis. This angle is typically small for direct decays of a Z boson into a pair of muons whereas muons from τ decays tend to be produced with a larger value of Ω^*.

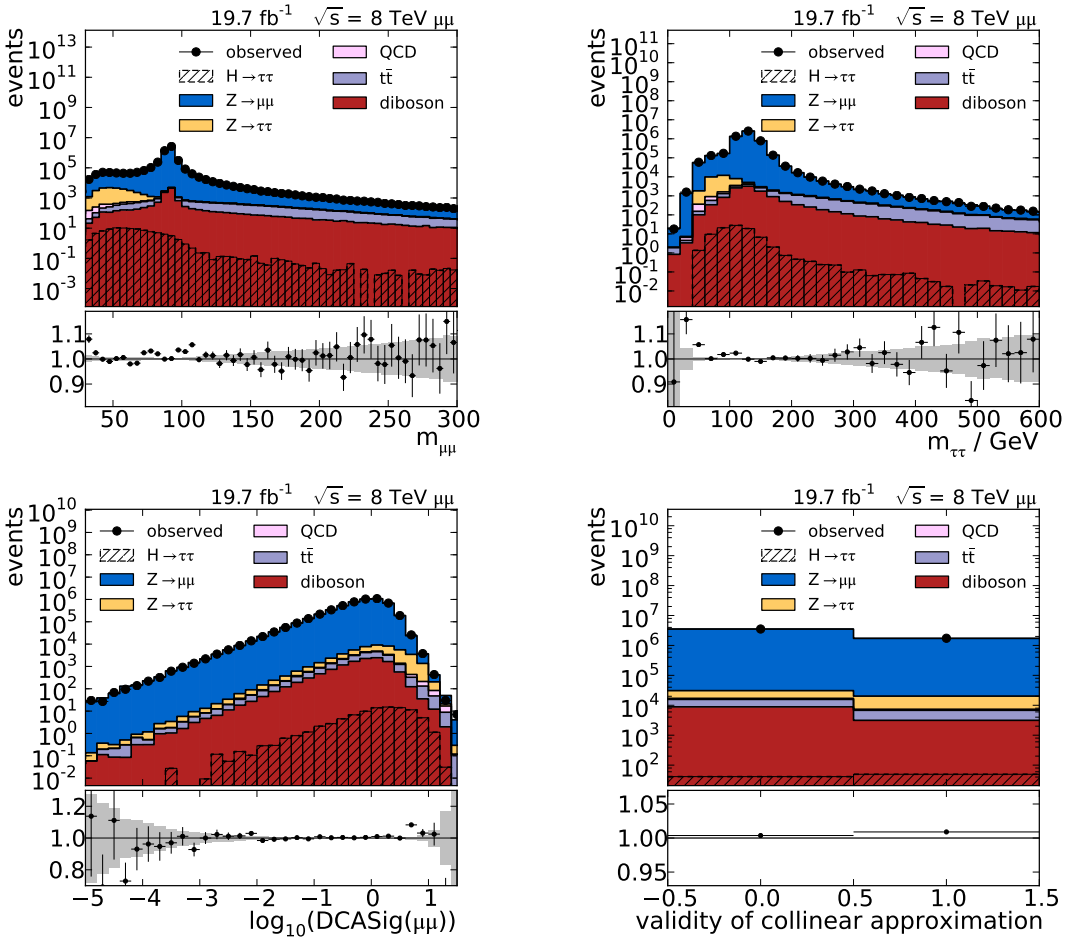


Figure 5.6: Common variables in all BDT trainings of the $H \to \tau\tau \to \mu\mu$ analysis (part I). The distributions show the level of agreement between data and simulation of the invariant mass of the dimuon pair, the reconstructed ditau mass using the SVFIT algorithm, the significance of the DCA variable and the validity of the collinear approximation.

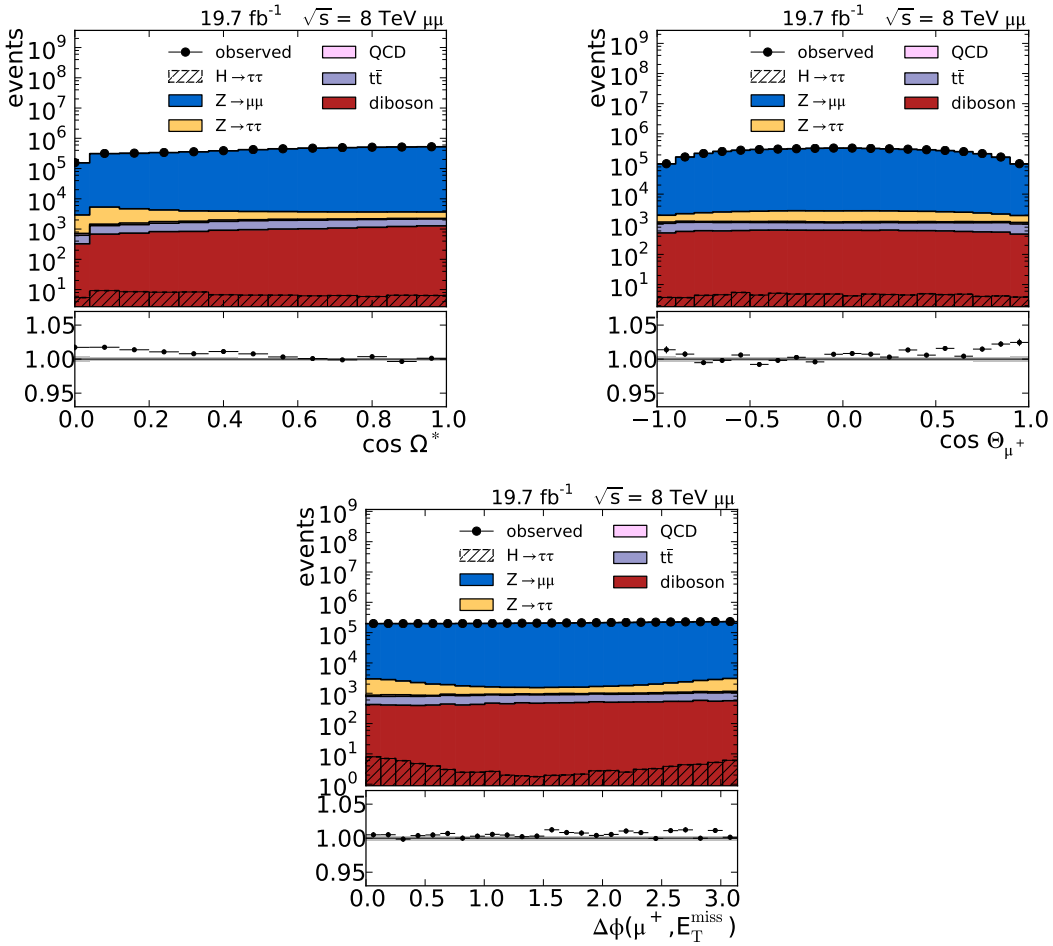


Figure 5.7: Common variables in all BDT trainings of the $H \to \tau\tau \to \mu\mu$ analysis (part II). The angular variables $\cos\Omega^*$ and $\cos\Theta^*_{\mu^+}$ and the azimuthal distance between the anti-muon and the missing transverse energy.

- The variable $\cos\Theta^*_{\mu^+}$ is defined as the cosine of the angle between the anti-muon and the dimuon system in the dimuon rest frame. This variable is sensitive to the spin of the decaying objects. The distribution in signal events is relatively flat compared to the one of the main background which prefers wide angle decays.

- The difference $\Delta\phi(\mu^+, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})$ between the azimuthal angle of the anti-muon and the direction of the missing transverse energy is exploited as muons from tau decays are accompanied by missing energy in a similar direction. In $Z \to \mu\mu$ events, the missing energy is almost evenly distributed in ϕ with a slight preference towards π – the direction of a recoiling jet.

- The last variable shared between both MVA analyses is the validity of the collinear approximation. Even though the collinear approximation as estimate of the ditau mass is superseded by the better performing SVFIT algorithm, one characteristic property is still used: The collinear approximation is not able to find valid solutions for all muon configurations. This logical value is sensitive to signal events as a failure of the collinear approximation is more likely in background events than for muons from a $H \to \tau\tau$ decay.

Two more variables, sensitive to different characteristics of the Drell-Yan process and gluon fusion Higgs production, are shown in Figure 5.8. They are used only in the analysis of the 0- and 1-jet categories.

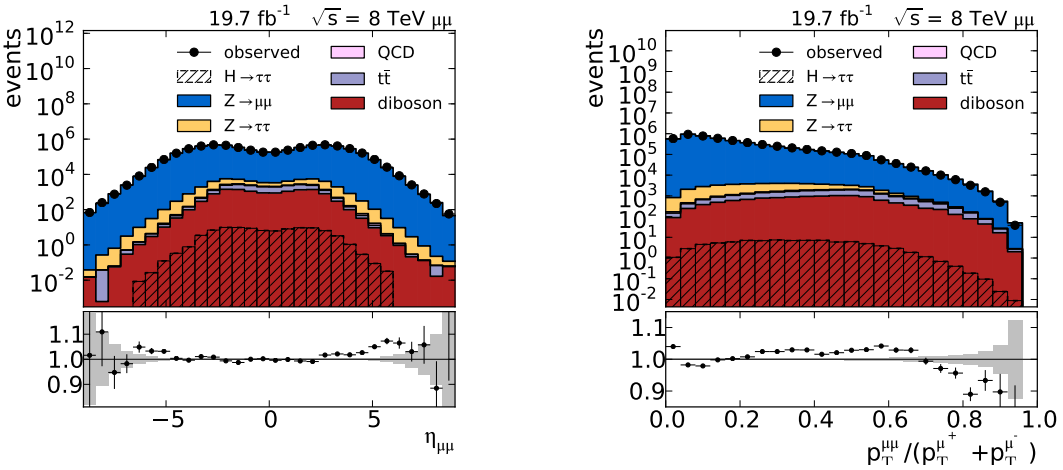


Figure 5.8: The pseudorapidity of the dimuon system and the relative dimuon momentum are additionally used in the 0-jet and 1-jet categories of the $H \to \tau\tau \to \mu\mu$ analysis.

- The pseudorapidity of the dimuon system $\eta_{\mu\mu}$ is concentrated on the central part of the detector in case the muons come from a Higgs boson produced via gluon fusion as the longitudinal momentum distribution is the same for both initial gluons. In the Drell-Yan process, the incoming quark, which is in most cases a valence quark, carries a higher momentum fraction on average than the anti-quark, which is a sea quark.

- The relative dimuon momentum is defined as $p_{\mathrm{T}}^{\mu\mu}/(p_{\mathrm{T}}^{\mu^+} + p_{\mathrm{T}}^{\mu^-})$. It is correlated with the angle between the two muons and thus sensitive to the boost of the dimuon system.

The production mechanism of the VBF topology is best exploited using the characteristic properties of the tagging jets and the missing transverse energy. The distributions of the following variables are shown in Figure 5.9.

- The invariant dijet mass m_{jj} of the two tagging jets in a 2-jet event is an extra factor in the matrix element of the VBF Higgs boson production. This leads to a relatively flat distribution in signal events of this production channel while all backgrounds show a steep exponentially falling distribution.

- A large pseudorapidity difference $\Delta\eta_{jj}$ between the tagging jets (rapidity gap) is a prominent signature of events produced in VBF. The momentum vectors of the incoming quarks only change their directions minimally by the radiation of a vector boson. Therefore, the two tagging jets are most often reconstructed in the forward region in opposite sides. In Drell-Yan events including additional radiation of jets as well as in dijet events and all other backgrounds, the two leading jets are close to each other in most cases.

- The missing energy can be exploited in the 2-jet category to suppress the $Z \to \mu\mu$ background which has no intrinsic $E_{\mathrm{T}}^{\mathrm{miss}}$. The characteristic of large tails of the missing energy distribution is only shared by the $t\bar{t}$ background which also contains neutrinos from the leptonic W decays.

Additional study of the Zeppenfeld variable

In addition to the variables shown before, further variables have been considered to discriminate between signal and background. One example is the "Zeppenfeld variable" [126] z and the reduced counterpart z^* which are defined as

$$z = \eta_{\mathrm{j}_3} - \frac{\eta_{\mathrm{j}_1} + \eta_{\mathrm{j}_2}}{2} \qquad z^* = \frac{z}{|\eta_{\mathrm{j}_2} - \eta_{\mathrm{j}_1}|}\,. \tag{5.6}$$

All involved jets are required to have a transverse momentum of more than 30 GeV to avoid pile-up influences at low p_{T}. Lower thresholds have been considered but

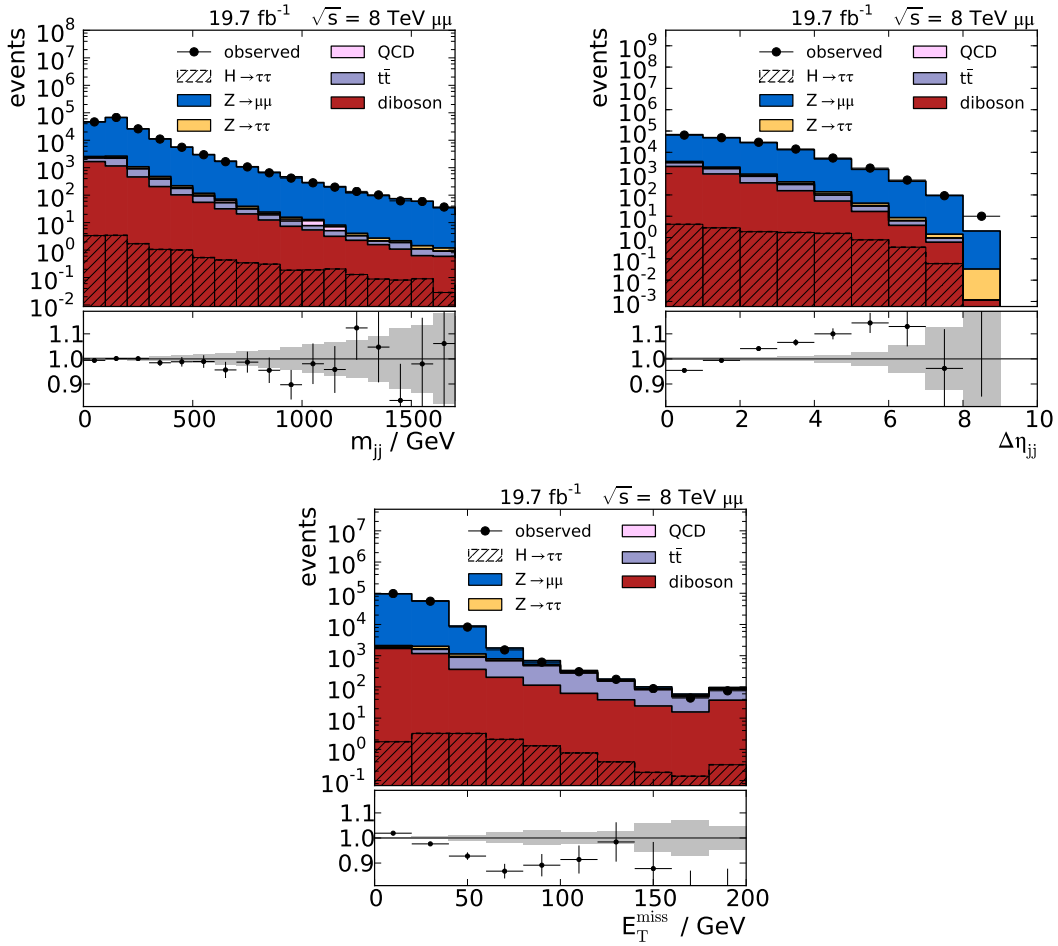


Figure 5.9: Additional variables for the analysis of events in the 2-jet category of the $H \to \tau\tau \to \mu\mu$ analysis. The invariant mass of the two tagging jets, the distance in rapidity between the two and the missing transverse energy are only used in this category to identify VBF events.

they show similar distributions. This variable is studied with a definition of the 2-jet category that does not involve the central jet veto. A third jet inside the rapidity gap is equivalent to $z^* \in [-\frac{1}{2}, \frac{1}{2}]$, values of $|z^*| \approx \frac{1}{2}$ occur for a third jet close to one of the tagging jets.

A characteristic of the VBF topology is the absence of hadronic activity within the rapidity gap while additional jets can still occur as initial or final state radiation around the tagging jets in the forward region. The theoretical uncertainties studied in comparisons of the description of the events by Monte-Carlo generators are small [127] which allows to use this variable in the event selection.

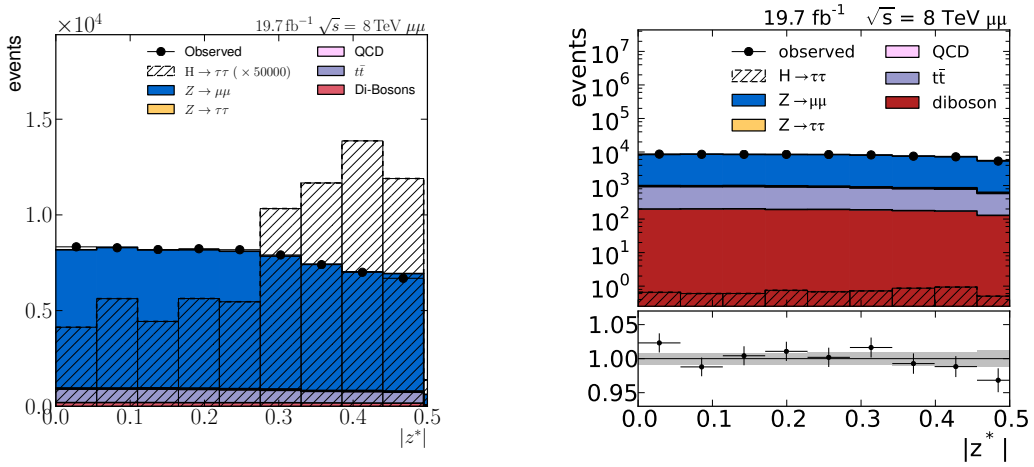


Figure 5.10: The distribution of the absolute value of the z^* variable in 2-jet events on a linear scale (*left*) and on a logarithmic scale (*right*) shows a promising separation between signal and background.

In reconstructed events, jets emerging from pile-up interactions fill the rapidity gap unless they are rejected by the Figure 5.10 shows the distribution of the absolute value of the z^* variable in the 2-jet category for signal events both on a linear and a logarithmic scale. As a result of the detailed modelling of pile-up in CMS and the advanced mitigation techniques presented in this work, the distribution of z^* is well described in simulated events. This has to be emphasized as z^* is not only affected by residual pile-up but it also combines properties of three jets requiring both higher-order QCD calculations and a correct modelling of the parton shower and hadronization.

An almost flat distribution with a slightly falling slope is observed for all backgrounds. In signal events, the maximum of third jets is found in the vicinity of one of the tagging jets. A lower number of events is observed directly around $|z^*| \approx \frac{1}{2}$ as hadronization products that would otherwise result in a third jet can not be clustered

separately in case they cover the same η-ϕ space.

The contributions from residual pile-up jets and jets from the underlying event in the centre of the rapidity gap are well controlled to allow the use of the z^* variable for the selection of VBF events. For the $H \to \tau\tau \to \mu\mu$ analysis is was decided not to use it in the MVA analysis but in the event categorisation. As described in Section 5.1.3, for this subchannel a central jet veto is imposed which is roughly equivalent to the requirement

$$|z^*| \geq \frac{1}{2}\,.$$

5.3.7 Calibration of the dimuon background

Differences in the shapes of this variable can be exploited to evaluate the contribution from the dimuon Drell-Yan background and thus to calibrate the event yield of this background based on collision data. A fit is performed that fits the templates of the Drell-Yan backgrounds to the observed data distribution of the DCA significance to obtain scale factors for the normalization of the simulated samples. All but the Drell-Yan backgrounds are subtracted from the data distributions prior to the fit.

The correction factors are derived separately in different regions of the variable space which are defined as follows. To perform this calibration separately for different compositions between the two Drell-Yan backgrounds, dedicated BDTs are trained to separate between the two. The mass variables $m_{\mu\mu}$ and $m_{\tau\tau}$ are removed from the list of BDT variables as well as the DCA significance itself which is used for template fit. This reduces correlations to the standard BDTs in order not to bias the final result. The discriminator of this reduced BDT is called D_{red}.

20 rectangular regions are defined in the $m_{\mu\mu}$-D_{red} space for events in the 0- and 1-jet categories and 12 regions for the 2-jet category. Five intervals in the dimuon mass $m_{\mu\mu}$ with the borders at $70, 91, 110$ and $150\,\text{GeV}$ account for different ranges of the Z mass distribution. The last two intervals are joined in the 2-jet category for statistics reasons. The borders along the D_{red} axis are $-1.0, -0.7, -0.4, -0.2, 1.0$ for the 0- and 1-jet categories and $-1.0, -0.4, -0.25, 1.0$ in the 2-jet category. In these 4×5 (3×4) regions, the DCA template fits are performed resulting in scale factors for the $Z \to \mu\mu$ background.

5.3.8 BDT trainings and combination of the discriminators

The Boosted Decision Trees are trained separately for the 0- and 1-jet categories and for the 2-jet category. In each of these two parts of the analysis, two BDTs are trained to discriminate against different backgrounds. The first of these two BDTs is dedicated to separate the $\tau\tau$ final states against all other background contributions of which $Z \to \mu\mu$ is the dominant one. The rejection of the $Z \to \mu\mu$ background is the major challenge of this analysis. The second BDT is trained to separate the $H \to \tau\tau$

signal from the $Z \to \tau\tau$ background. Half of the events, the BDT is trained on, are signal events, the other half is background. For the signal events, a superposition of all considered Higgs boson masses is used to avoid preferences for a particular mass hypothesis.

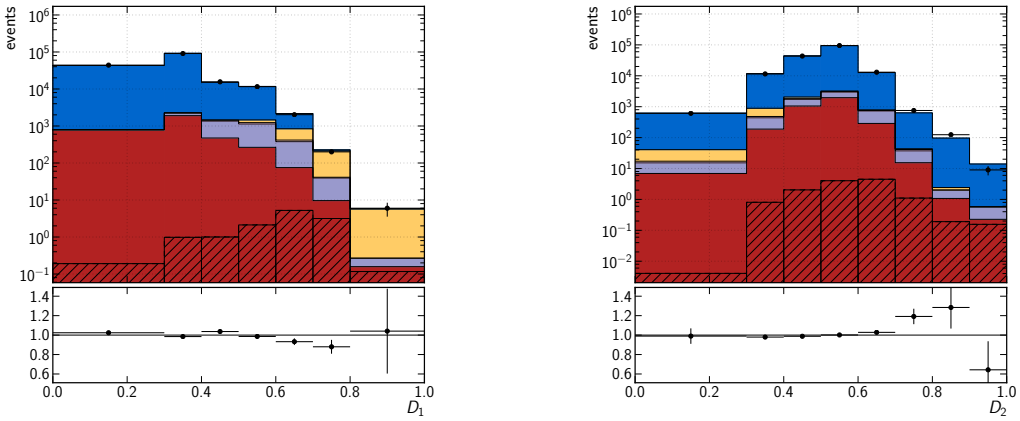


Figure 5.11: The two discriminators D_1 and D_2 in the 2-jet category.

Each of these two BDTs is evaluated separately in each of the categories. The output of both BDTs in the 2-jet category is shown in Figure 5.11. The discriminator D_1 of the BDT, trained to separate between $Z \to \mu\mu$ and $H \to \tau\tau$, (Fig. 5.11a), shows a clear separation between the two classes of events. This output is mainly based on the dimuon mass $m_{\mu\mu}$ and the DCA significance. Backgrounds containing τ decays or other decays involving neutrinos get high values of D_1 like the signal events. The disciminator D_2 of the second BDT distinguishes τ decays from Z bosons from those of the Higgs boson (Fig. 5.11b).

For the signal extraction, the discriminators D_1 and D_2 of the two BDTs are combined using a PDF integration method. The codomain of each of the multivariate methods is $[0, 1]$. For each event category the two-dimensional probability density function $f_{\text{sig}}(D_1, D_2)$ of the two discriminators is constructed for the signal hypothesis. It is constructed from the signal events passing the respective category selection. For every event, f_{sig} of the corresponding category is then integrated to the point (D_1, D_2) to get the final discriminator D.

$$D = \int_0^{D_1} \int_0^{D_2} f_{\text{sig}}(D_1', D_2')\, \mathrm{d}D_1' \mathrm{d}D_2' \tag{5.7}$$

Figure 5.12 shows the cumulative signal PDF as a function of D_1 and D_2 and the

final discriminator D in the 2-jet category. The observed distribution of data events is close to the distribution in simulation as it has been the case for the input variables and the discriminators of the individual BDTs. The most sensitive bin $D \in (0.8, 1.0]$ shows a downward fluctuation which is compatible with both the signal and the background hypothesis. The complete set of discriminators for all categories and data-taking periods that enter the combination can be found in [112, 128].

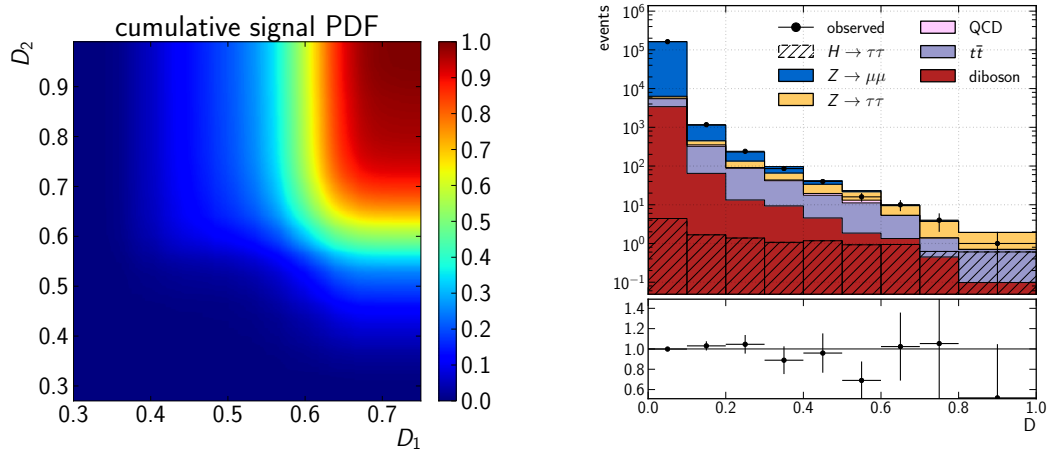


Figure 5.12: The cumulative signal PDF $f_{\text{sig}}(D_1, D_2)$ of the two discrimiators D_1 and D_2 in the 2-jet category (*left*). It is used to calculate the final discriminator D (*right*) using Eq. (5.7).

5.3.9 Theoretical uncertainties

The simulation of the gluon fusion and VBF signal processes, which have the largest production cross section at the LHC, is done using the Monte-Carlo generator POWHEG. The theoretical uncertainties on the modelling of jets in the event are estimated, following the methods that were consistently used for all $H \to \tau\tau$ channels [129]. These uncertainties affect the properties of jets in the event and thus the classification of events into categories as well as the shapes of the distributions of jet variables inside the 2-jet category.

The acceptance for each category is given by $a_k \equiv \frac{n_k}{n_{\text{all}}}$, the ratio of the number of accepted events n_k in category k with respect to the number of all events in the preselection. In addition, shape uncertainties are estimated for the final discriminator in the 2-jet category. All other categories do not exploit jet properties. The two variables $\Delta\eta_{jj}$ and m_{jj} in the 2-jet category are affected by the theoretical uncertainties on the simulation of jets in the VBF topology and accepted gluon fusion events.

This study includes the uncertainties on the PDFs, α_S, the renormalization and factorization scale, the matching scale and the parton shower and hadronization modelling of the event generator. For each of these uncertainties, several samples are produced with different Monte-Carlo generators and varying parameters. All uncertainties are studied separately for the two main production channels VBF and gluon fusion.VBF events are generated using the POWHEG-BOX with VBF@NLO [24]. For the gluon fusion events that are accepted by the 2-jet selection, MADGRAPH [89] and POWHEG+MINLO [130] are compared to the default POWHEG sample [24].

To avoid the expensive detector simulation and reconstruction step for each of the variations, the varied samples are produced by Monte-Carlo generators only. From these samples the distributions of $\Delta\eta_{jj}$ and m_{jj} are studied and compared to the reference sample. Weights are calculated according to

$$w(\Delta\eta_{jj}, m_{jj}) = \left[\frac{n_{\text{varied}}}{n_{\text{reference}}}\right]_{\Delta\eta_{jj}, m_{jj}}$$

in the two dimensional $\Delta\eta_{jj}$-m_{jj} plane. These weights are then applied to the fully simulated reference sample according to the values of these two variables calculated from the particle jets of the event on generator level. The full analysis is then performed with all weighted variants. The significance of the theory uncertainties is determined from the variance of subsamples.

Renormalization and factorization scale uncertainty The choice of the renormalization and factorization scales, μ_{R} and μ_{F}, is arbitrary and should not affect the results. As NLO generators are used, the dependence on these scales is expected to be small. The values chosen in the reference sample are

$$\mu_{\text{R}} = \mu_{\text{F}} = \frac{1}{2} m_H \qquad \text{for VBF and} \tag{5.8}$$

$$\mu_{\text{R}} = \mu_{\text{F}} = m_H \qquad \text{for gluon fusion.} \tag{5.9}$$

Both parameters are varied up and down by a factor of two under the condition that $\frac{1}{2} \leq \mu_{\text{R}}/\mu_{\text{F}} \leq 2$, resulting six variations of the scales. The uncertainty on the acceptance is defined as

$$\Delta a_k^{\text{R,F}} = \frac{1}{2} \cdot \frac{a_k^{\max} - a_k^{\min}}{a_k^{\text{ref}}}$$

where a_k^{ref} is the acceptance in the reference sample. An uncertainty of 0.2 % (7.8 %) is added linearly for VBF (gluon fusion) events to account for the uncertainty on the inclusive cross section, following the recommendations in [17]. The shape uncertainties are only statistically significant in the range $D \in [0.5, 1.0]$ a corresponding shape uncertainty is added for the 2-jet category.

Matching scale uncertainty As the NLO matrix element has to be interfaced with the parton shower modelled with PYTHIA, double counting of events with higher jet multiplicities has to be avoided. The matching scale $p^2_{\mathrm{T,min}}$ separates the low p_{T} range in which jets are produced by the parton shower from the high p_{T} range where the jets are produced by the matrix element calculation. The matching scale of the reference sample is $p^2_{\mathrm{T,min}} = 1\,\mathrm{GeV}^2$. This value is varied within the range from 0.5 GeV to 30 GeV. The uncertainty on the category acceptance

$$\Delta a_k^{\mathrm{M}} = \frac{1}{2} \cdot \frac{a_k^{\mathrm{max}} - a_k^{\mathrm{min}}}{a_k^{\mathrm{ref}}}$$

is 1 %. No significant impact on the shape of the final discriminator can be observed.

Generator comparison The modelling of the available generators for the gluon fusion process, POWHEG, MADGRAPH and POWHEG+MINLO and MADGRAPH, shows differences in the description of jet multiplicities and jet kinematics. The envelope of these differences is treated as another component of the theory uncertainties. The category acceptance is affected by up to 18 % and it also has the largest contribution to the shape uncertainties on the final discriminator. It is quadratically added to the scale uncertainties of the gluon fusion sample as it is an independent source of uncertainty.

Parton shower The uncertainties on the modelling of the parton shower are studied by comparing different tunes in PYTHIA. The default tune in CMS (Z2*) is compared to the AUET2 used by ATLAS. The full difference between the two is taken as the parton shower uncertainty

$$\Delta a_k^{\mathrm{PS}} = \frac{a_k^{\mathrm{AUET2}} - a_k^{\mathrm{Z2^*}}}{a_k^{\mathrm{Z2^*}}} \,.$$

This uncertainty is up to 9 % depending on the category.

The aforementioned uncertainties on the acceptance of the different categories of the $H \to \tau\tau \to \mu\mu$ analysis are summarized in Table 5.2. The impact on the final discriminator in the 2-jet category of the $H \to \tau\tau \to \mu\mu$ analysis are shown in Figure 5.13 for the gluon fusion and the VBF produced events.

PDF uncertainty To estimate the uncertainty due to the limited knowledge about the proton PDFs, the PDF4LHC working group [33] recommends to evalute the PDF sets of the NNPDF [36], MSTW [35] and CT10 [131] Collaborations. The evaluation of these PDF sets is done using LHAPDF [37]. The CT10 and MSTW PDF sets contain 25 and 20 pairs of eigenvectors, respectively. The NNPDF uncertainties are

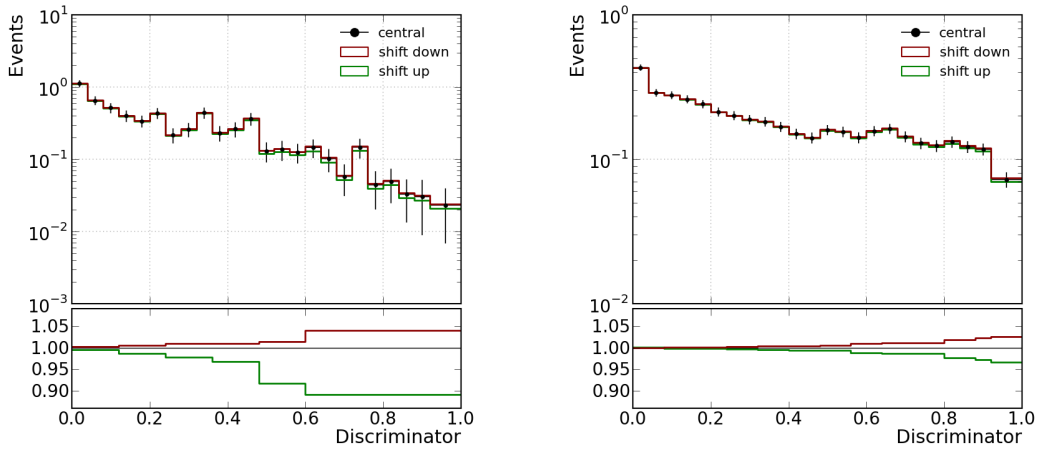


Figure 5.13: Theoretical uncertainty on the shape of the final discriminator distributions in the 2-jet category of the $H \to \tau\tau \to \mu\mu$ analysis resulting from scale uncertainties and differences in the generator modelling for gluon fusion (*left*) and VBF events (*right*).

evaluated as the variance of its 100 member PDFs. Weights are calculated for each member PDF

$$w_i^{\mathrm{PDF}} = \frac{x_1 f_i(x_1, Q^2, q_1)}{x_1 f_0(x_1, Q^2, q_1)} \cdot \frac{x_2 f_i(x_2, Q^2, q_2)}{x_2 f_0(x_2, Q^2, q_2)} \tag{5.10}$$

where x_j is the momentum fraction of the parton j, f_i is the member PDF i of the PDF set, Q^2 is the resolution scale of the process and q_j is the flavour of the parton for $j \in 1, 2$. Subsequently, the acceptances of all categories are determined and the shapes of $\Delta\eta_{jj}$ and m_{jj} for each of the 190 member PDFs. The uncertainty on the acceptance for the variation of the CT10 and MSTW PDF sets are determined as

$$\Delta a_k^+ = \sqrt{\sum_{i=1}^{n} \max(a_{k,i}^+ - a_{k,0}, a_{k,i}^- - a_{k,0}, 0)^2} \tag{5.11}$$

$$\Delta a_k^- = \sqrt{\sum_{i=1}^{n} \min(a_{k,i}^+ - a_{k,0}, a_{k,i}^- - a_{k,0}, 0)^2} \tag{5.12}$$

whereas the acceptance uncertainty for the NNPDF variations is calculated as

$$\Delta a_k^\pm = \sqrt{\frac{1}{n_\pm} \sum_{i=1}^{n_\pm} (a_{k,i} - a_{k,0})^2}\,. \tag{5.13}$$

where n_+ (n_-) are the members of the PDF set with $a_i > a_0$ ($a_i < a_0$). The uncertainties derived from the CT10 PDF set are scaled by a factor of 0.608 as they provide 90 % CL uncertainties. The envelope of these uncertainties from the three PDF sets is calculated and symmetrized as

$$\Delta a_{\mathrm{PDF}} = \frac{1}{2}\big(\max_i(a_i + \Delta a_i^+) - \min_i(a_i - \Delta a_i^-)\big) \tag{5.14}$$

The PDF uncertainties are found to be small. They amount to 1 % for VBF and 2 % for gluon fusion events. An uncertainty of 2.8 % and 7.5 % is added linearly for VBF and gluon fusion events, respectively, to account for the uncertainty on the inclusive cross section [17].

In the same way, the two-dimensional set of weights $w(\Delta\eta_{jj}, m_{jj})$ are calculated for each member of the PDF sets using equivalent formulae as Eq. (5.12) and (5.13). For each PDF set, the resulting weights are applied to the fully simulated datasets for both VBF and gluon fusion events in order to propagate the uncertainties to the shape of the final discriminator. The symmetrized envelope of all PDF sets is shown in Figure 5.14 for both production processes. All theory uncertainties on the categorisation are summarized in Table 5.2.

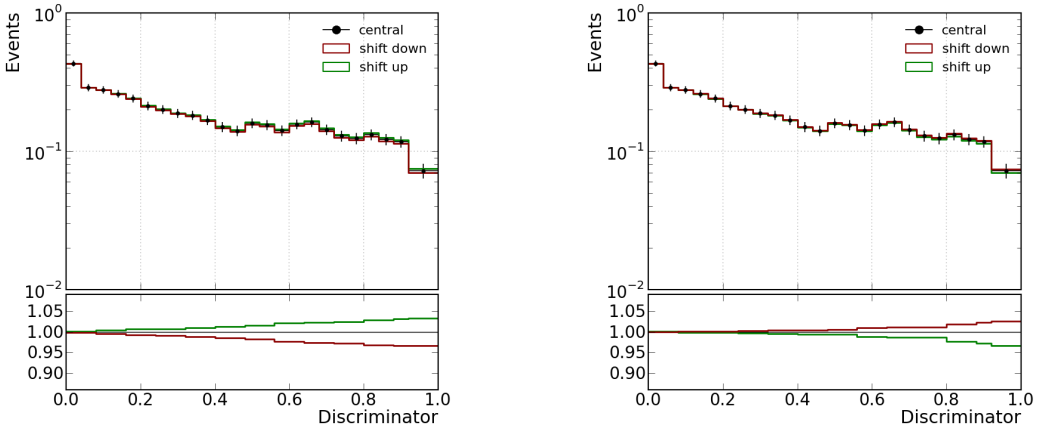


Figure 5.14: Theoretical uncertainty on the shape of the final discriminator distributions in the 2-jet category of the $H \to \tau\tau \to \mu\mu$ analysis resulting from the PDF uncertainties for gluon fusion (*left*) and VBF events (*right*).

Table 5.2: Uncertainties on the acceptance of categories [129]. Three classes of uncertainties are treated separately: The PDF uncertainties, the underlying event and parton shower uncertainties (UE, PS) and the QCD scale uncertainties. For each of the two production modes, the number of events n_{exp} expected in the respective category is given.

Category	**vector boson fusion**				**gluon fusion**			
	$n_{\text{exp}}^{\text{VBF}}$	PDF	UE, PS	QCD scale	$n_{\text{exp}}^{\text{gf}}$	PDF	UE, PS	QCD scale
0-jet high	0.40	3.6 %	6.3 %	2.8 %	33.3	9.7 %	4.2 %	10.0 %
0-jet low	0.50	3.6 %	8.9 %	8.9 %	36.0	9.7 %	3.5 %	10.3 %
1-jet high	1.41	3.6 %	0.4 %	1.3 %	9.4	9.7 %	−2.2 %	10.7 %
1-jet low	3.60	3.6 %	0.0 %	0.8 %	18.8	9.7 %	−1.6 %	10.9 %
2-jet	4.15	3.6 %	−1.4 %	0.9 %	7.7	9.7 %	−7.4 %	−18.2 %

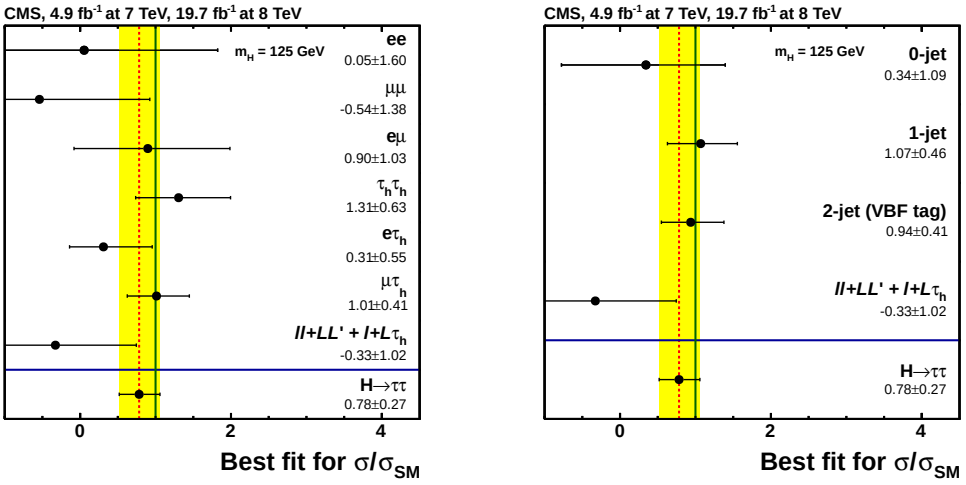
Table 5.3: Overview of the results of the $H \to \tau\tau \to \mu\mu$ decay channel. The upper part of this table shows the results for the study at 7 TeV (2011) and at 8 TeV (2012) in the lower part. For each category, the event numbers are given for discriminator values between $D = 0.6$ and $D = 1$, as this is the most sensitive region. The number of observed events (obs.) in collision data is compared to the number of events expected from the sum of all background processes. The uncertainties from the normalization of each background contribution (norm.), the luminosity measurements (lumi.), the muon identification efficiencies (ε_μ), the b-tagging efficiencies ($\varepsilon_{\text{b-tag}}$), the jet energy scale (JES) and the scale of the missing transverse energy (MET). The right part shows the expected number of signal events from all four considered production modes together with the theory uncertainties comprising the proton structure (PDF), QCD scales (scale) and the underlying event and parton shower model (UE, PS) uncertainties, while other uncertainties that are shared with the background are summarized in the last column. This table can not fully reflect the treatment of correlated uncertainties as it is the case in the statistical combination of the results.

Category	**obs.**	**expected background**							**signal**				
			norm.	lumi.	ε_μ	$\varepsilon_{\text{b-tag}}$	JES	MET		PDF	scale	UE,PS	other
Collisions at 7 TeV recorded in 2011													
0-jet low	407	442.90	±9.46	±0.04	±12.51	∓0.02	∓0.04	∓0.07	1.06	±0.10	±0.10	±0.04	±0.23
0-jet high	49	46.61	±1.44	±0.11	±1.01	∓0.13	∓0.12	±0.11	0.92	±0.09	±0.09	±0.03	±0.22
1-jet low	25	26.75	±0.72	±0.02	±0.90	∓0.04	±0.01	±0.02	0.45	±0.04	±0.04	∓0.01	±0.16
1-jet high	28	27.39	±1.00	±0.08	±0.63	∓0.16	∓0.10	±0.02	0.58	±0.04	±0.04	∓0.00	±0.18
2-jet	1	2.66	±0.18	±0.01	±0.11	∓0.04	∓0.05	±0.10	0.27	±0.17	∓0.03	∓0.01	±0.13
Collisions at 8 TeV recorded in 2012													
0-jet low	1665	1676.34	±34.91	±0.10	±45.68	∓0.06	∓0.37	∓0.02	3.40	±0.32	±0.33	±0.14	±0.42
0-jet high	314	306.12	±6.34	±0.45	±5.42	∓0.57	∓0.54	±0.09	2.81	±0.26	±0.28	±0.10	±0.39
1-jet low	63	68.93	±1.94	±0.07	±2.33	∓0.19	∓0.15	∓0.00	1.72	±0.15	±0.16	∓0.03	±0.31
1-jet high	95	72.10	±2.46	±0.18	±1.77	∓0.47	∓0.28	±0.14	2.17	±0.18	±0.19	∓0.02	±0.34
2-jet	15	15.50	±1.00	±0.05	±0.61	∓0.32	±0.10	∓0.04	2.17	±0.43	∓0.10	∓0.08	±0.36

5.4 Results in the search for $H \to \tau\tau$

The statistical combination of all channels and categories is based on a maximum likelihood fit and the CL_s method as described in Section 2.7.3.

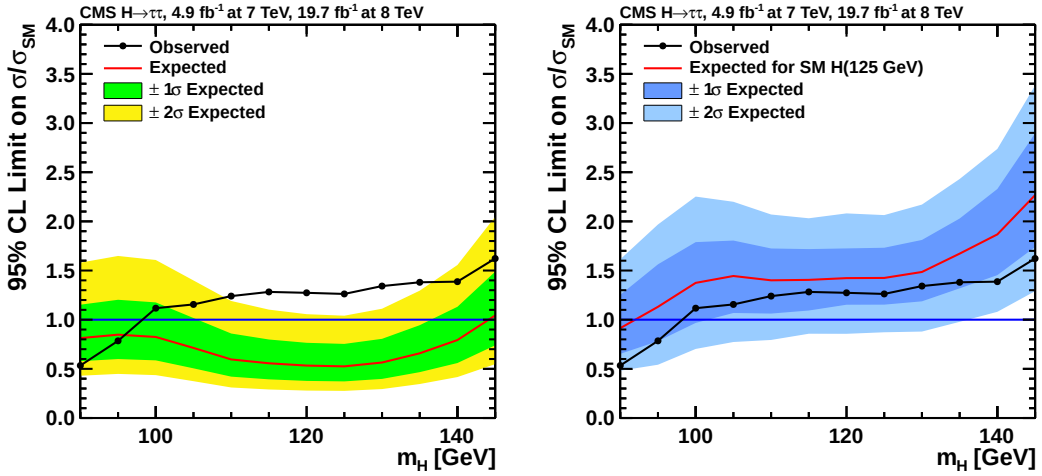
In most channels, the final discriminating variable, which is used for the global fit, is either $m_{\tau\tau}$ or the visible mass m_{vis}. In the ee and $\mu\mu$ channels, the final discriminator D, combined from the discriminators D_1 and D_2 of the two BDTs, enters the global fit. These distributions are combined using a binned likelihood [106]. This involves the expected numbers of signal events, of background events and the observed number of events in that bin. The expected number of signal events is scaled by the signal strength modifier μ which is the single parameter of interest in the global fit. The systematic uncertainties are represented by nuisance parameters which can be correlated across channels, categories and bins. The probability density function of these parameters is assumed to follow a log-normal distribution. The statistical uncertainty in the observed event yields is the dominant uncertainty.



(a) (b)

Figure 5.15: Best fit values for $\mu = \sigma/\sigma_{\text{SM}}$ (a) for individual channels and (b) separately for each category, combined across the channels for the Standard Model Higgs boson with $m_H = 125\,\text{GeV}$. All channels and categories show a consistent result [106].

Figure 5.15 shows the best fit values for $\mu = \sigma/\sigma_{\text{SM}}$ for all channels and categories combined across the decay channels for the Standard Model Higgs boson with $m_H = 125\,\text{GeV}$ These plots reflect the ordering in sensitivity of the decay channels and event categories as discussed in Sections 5.1.2 and 5.1.3, respectively. The



(a) (b)

Figure 5.16: 95 % CL limits on the Higgs cross section relative to the Standard Model cross section (a) for the background-only hypothesis and (b) the Standard Model including the Higgs boson at $m_H = 125\,\text{GeV}$ While the combined analysis is sensitive over the whole mass range as shown in (a), the Higgs boson can not be excluded due to a broad excess in the observed limit. This excess is compatible (b) with the Standard Model Higgs boson [106].

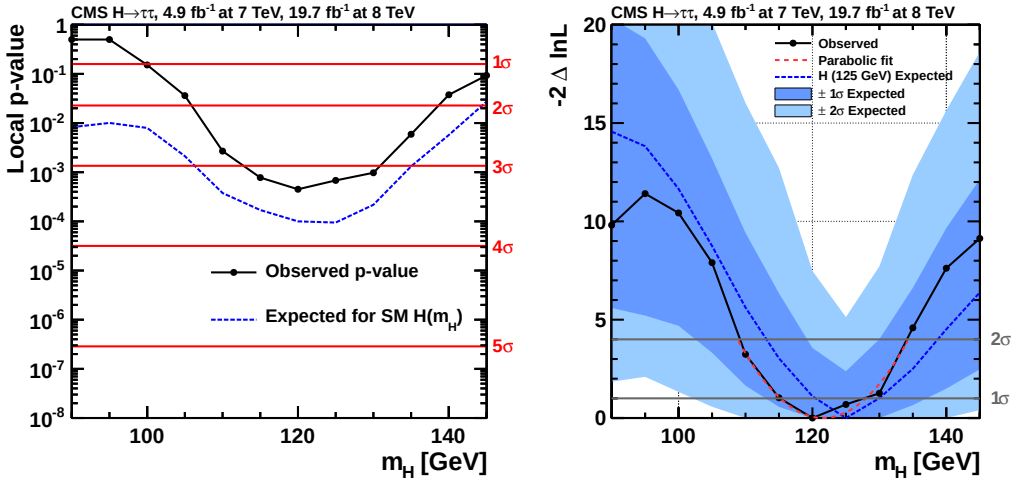
combined best fit value of the common signal strength modifier is

$$\hat{\mu} = 0.78 \pm 0.27$$

and thus compatible with the Standard Model expectation.

The 95 % CL limits on the Higgs cross section relative to the Standard Model cross section is shown in Figure 5.16 for different mass hypotheses. The probed mass hypotheses are equally spaced by 5 GeV and linearly interpolated. The combined $H \to \tau\tau$ search is sensitive over the whole investigated mass range up to Higgs boson masses of 145 GeV. Only a small part of this mass range can be excluded, $m_H < 98\,\text{GeV}$, which was already excluded by the LEP experiments. Over the rest of the mass range, a broad excess is observed. For Higgs boson masses between 109 GeV and 137 GeV, the deviation from the Standard Model expectation is more than 2σ. In the same mass range, the results are compatible with the Standard Model expectation including the Higgs boson with a mass of $m_H = 125\,\text{GeV}$ as shown in Figure 5.16b. The contribution from Higgs decay modes other than $H \to \tau\tau$, in particular $H \to WW$, has been treated as part of the background.

The probability for the background-only hypothesis to describe the data is given by the local p-value. It is shown in Figure 5.17a as a function of the probed Higgs



(a) (b)

Figure 5.17: The local p-value (a) quantifies the significance of the excess to be 3.2σ, providing evidence for the Higgs decay to τ-leptons. The mass scan (b) shows that the best compatible mass for the $H \to \tau\tau$ data is $122 \pm 7\,\text{GeV}$ [106].

boson mass hypotheses. The contribution from $H \to WW$ is again considered as part of the background. The look-elsewhere effect did not have to be accounted for as the mass range in which the signal was expected was known at least since the discovery of the Higgs boson in 2012 in other channels. The p-value reaches 3.2σ at a mass of 120 GeV, thus providing evidence for Higgs decays to τ-leptons. Combining the $H \to \tau\tau$ and $H \to b\bar{b}$ searches of the CMS Collaboration the p-value increases to 3.8σ [132]. While the discovery of the Higgs boson in 2012 cannot be explained in terms of the Standard Model without the coupling to fermions in the gluon fusion production process, $H \to \tau\tau$ is the first channel providing evidence for Higgs couplings to down-type fermions and for fermionic Higgs decays in general.

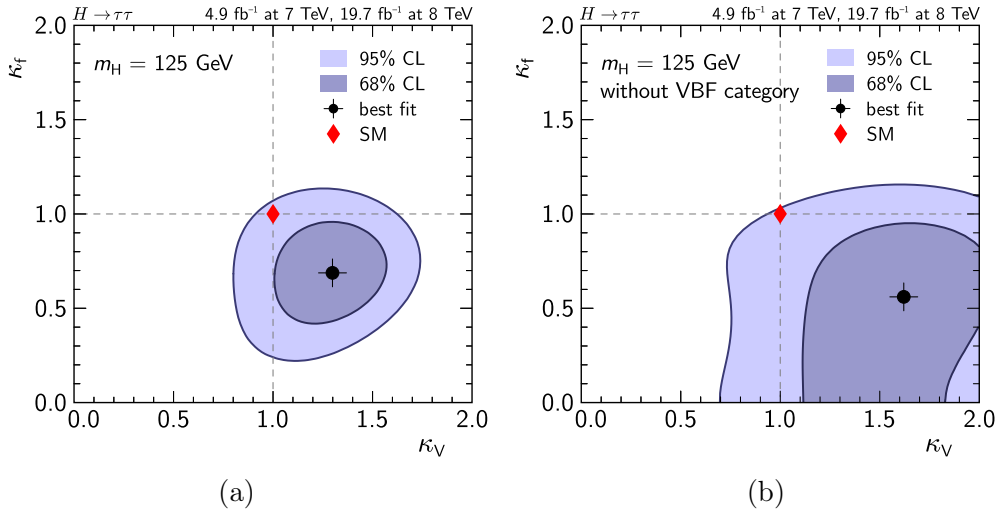
The mass resolution in the $H \to \tau\tau$ channel is limited due to the two to four neutrinos in the Higgs boson decay escaping detection in the experiment. A scan over different Higgs boson mass hypotheses in the range from 90 GeV to 145 GeV is shown in Figure 5.17b. The best fit value for the Higgs mass in the combined $H \to \tau\tau$ search is

$$m_{\text{H}} = 122 \pm 7\,\text{GeV}$$

as indicated by a parabolic fit to the discrete $-2\Delta \ln L$ values below four of the tested mass hypotheses. The measurement lies within $\pm 1\sigma$ of the expectation for a Higgs boson at 125 GeV. The Higgs mass estimate in this channel is thus compatible with the measurement in the high resolution channels $\gamma\gamma$ and ZZ [133] in which it has

been measured to be $m_{\mathrm{H}}^{\mathrm{CMS}} = 125.6 \pm 0.2\,(\mathrm{syst.}) \pm 0.4\,(\mathrm{stat.})\,\mathrm{GeV}$.

The relative couplings to fermions κ_{f} and to vector bosons κ_{V} with respect to the Standard Model prediction can be fitted independently [134]. This fit includes the fraction of $H \to WW$ events in the analysis as signal. Figure 5.18a shows that these



(a) (b)

Figure 5.18: Relative couplings to vector bosons κ_{V} and to fermions κ_{f} with respect to the Standard Model prediction. The full $H \to \tau\tau$ search (a) is able to constrain both couplings to values within 2σ from the Standard Model value. Neither the fermionic nor the bosonic couplings are compatible with the background-only value $\kappa_{\mathrm{V}} = \kappa_{\mathrm{f}} = 0$. Without the VBF category (b), this analysis would not have been able to constrain the fermionic couplings.

relative couplings are compatible with the Standard Model expectations within 2σ and far from the background-only value at $(0, 0)$. The slight deficit in the μ-value corresponds to a value below unity for the relative fermionic coupling.

Figure 5.18b shows the same study without exploiting the most sensitive VBF category in all decay channels. This drastically relaxes the constraints on κ_{V} and removes the ability to constrain κ_{f}, emphasizing the importance of the VBF category for this analysis.

These results are consistent with the Standard Model predictions in all measurable properties of the recently discovered Higgs boson. They also agree with the results in other decay channels and the results from the ATLAS Collaboration [15].

Chapter 6

Conclusion

The measurements performed with the CMS experiment at the LHC rely on a precise determination of the energy of all involved physical objects. As jets are ubiquitous at a hadron collider, a precise measurement of the jet energy plays an important role for most analyses.

In this thesis, $Z(\rightarrow \mu\mu) + \text{jet}$ events are used to determine the jet energy scale from collision data and to derive jet energy corrections. This method requires exactly one balancing jet and was designed for interactions without additional activity in the event. The dramatic increase of the instantaneous luminosity, and hence of multiple pp collisions during the same bunch crossing (pile-up), required dedicated methods to remove the impact of pile-up interactions from reconstructed events. These pile-up mitigation techniques comprise the subtraction of charged particles from pile-up vertices, the rejection of entire jets emerging from pile-up interactions and corrections for contributions to the jet energy from the residual pile-up activity. These and further improvements to the Z-jet balancing technique resulted in a successful jet energy calibration for all data-taking periods in which collisions were recorded by the CMS experiment. Detailed systematic studies helped to reduce the uncertainty of absolute jet energy corrections by a factor of four. This was a major contribution to the jet energy calibration in CMS which reached a precision of up to half a percent in the central detector region. The measurement of the flavour response in data, the monitoring and correction of the time stability of the reconstructed jet energy as well as the extension of the calibration method to the endcaps and the forward detector are further results of this work.

The understanding of jets was exploited in the search for Higgs bosons decaying to pairs of τ-leptons. The success of the $H \rightarrow \tau\tau$ search is predominantly based on the analysis of Higgs bosons produced via vector boson fusion (VBF) – in contrast to other Higgs decay channels in which the results are by far dominated by gluon fusion produced Higgs bosons. Two jets in the forward region provide a clear signature for this topology. This is exploited in a dedicated VBF event selection that defines the most sensitive category throughout all subchannels.

Studies of these characteristic jets are presented and lead to a reduction of the systematic uncertainties on jets and indirectly on the missing energy. Furthermore, they revealed a systematic trend of the jet energy over time which required a new reconstruction of all CMS data and corresponding jet energy corrections.

The $H \to \tau\tau \to \mu\mu$ subchannel has been described in detail and is part of the combined result. The background – in particular the large Drell-Yan contribution – is reduced by multivariate methods and requires a comprehensive background estimation. In this channel, a central jet veto rejects events with a third jet in the rapidity gap between the two jets of the VBF selection. This criterion leads to a 10 % improvement in the signal to background ratio in the 2-jet category. However, it is particularly sensitive to a possible miscalibration of the jet energy in the low energetic range and additional jets from pile-up interactions. It is therefore only possible if jets emerging from pile-up interactions are successfully removed from the central detector region. This is accomplished by the use of the aforementioned pile-up mitigation techniques. Furthermore, the theoretical uncertainties on Higgs production in the main production modes have been studied for the $H \to \tau\tau \to \mu\mu$ decay channel.

This work is part of the official CMS results on the $H \to \tau\tau$ search. The statistical combination of all $\tau\tau$ decay modes shows an excess over the background-only hypothesis with a local significance of 3.2σ for a Higgs boson with $m_H = 125\,\text{GeV}$. In combination with the results from the search for $H \to b\bar{b}$, the significance further increases to 3.8σ as it will be published in Nature Physics [132]. This provides evidence for the Higgs boson decaying into a pair of τ-leptons and thus direct evidence for couplings to down-type fermions as predicted by the Standard Model.

Appendix A

Appendix

A.1 Datasets

A.1.1 CMS collision data

Only certified runs and luminosity sections were analysed, following the recommendation condensed in the json file released by the CMS Physics Validation group:

`Cert_160404-180252_7TeV_ReRecoNov08_Collisions11_JSON_v2`
`Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON`

Table A.1: Datasets of the 2011A and B and 2012A, B, C and D runs. The luminosities are calculated with the pixel luminosity method.

Run	Dataset	Events
2011A	`/DoubleMu/Run2011A-May10ReReco-v1/AOD`	
	`/DoubleMu/Run2011A-PromptReco-v4/AOD`	
	`/DoubleMu/Run2011A-Aug05ReReco-v1/AOD`	
	`/DoubleMu/Run2011A-Oct03ReReco-v1/AOD`	
2011B	`/DoubleMu/Run2011B-PromptReco-v1/AOD`	
2012A	`/DoubleMuParked/Run2012A-22Jan2013-v1/AOD`	876 /pb
2012B	`/DoubleMuParked/Run2012B-22Jan2013-v1/AOD`	4412 /pb
2012C	`/DoubleMuParked/Run2012C-22Jan2013-v1/AOD`	6766 /pb
2012D	`/DoubleMuParked/Run2012D-22Jan2013-v1/AOD`	7369 /pb
for prompt reco comparisons:		
2012A	`/DoubleMu/Run2012A-13Jul2012-v1/AOD`	4 913 546
	`/DoubleMu/Run2012A-recover-06Aug2012-v1/AOD`	722 728
2012B	`/DoubleMu/Run2012B-13Jul2012-v4/AOD`	22 786 354
2012C v1	`/DoubleMu/Run2012C-24Aug2012-v1/AOD`	2 181 586
2012C v2	`/DoubleMu/Run2012C-PromptReco-v2/AOD`	26 816 721
2012D	`/DoubleMu/Run2012D-PromptReco-v1/AOD`	30 748 995

A.1.2 Monte-Carlo datasets

The Monte Carlo studies are performed with the Drell-Yan to leptons and jets sample of the Summer12 production, using MADGRAPH for the matrix element generation and PYTHIA with the tune Z2* for the generation of the underlying event (see Table A.2).

Table A.2: Monte-Carlo datasets used in the $Z(\rightarrow \mu\mu) + \text{jet}$ anaysis.

Generator	Dataset	Events
MADGRAPH	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30 459 503
POWHEG	/DYToMuMu_M-20_CT10_TuneZ2star_v2_8TeV-powheg-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	48 819 386
QCD	/QCD_Pt_20_MuEnrichedPt_15_TuneZ2star_8TeV_pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v3/AODSIM	
W + jets	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/ Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM	
$t\bar{t}$ + jets	/TTJets_mass175_5_TuneZ2star_8TeV-madgraph-tauola/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	

A.2 Object and event selection criteria

A.2.1 Trigger paths

The HLT_Mu17_Mu8 trigger is used exclusively for 2012 data. The complete list of unprescaled single and double muon triggers used for the analysis of 2011 data: HLT_DoubleMu7, HLT_Mu8, HLT_Mu13_Mu8, HLT_Mu17_Mu8.

A.2.2 Muon identification

Muon identification criteria for 7 TeV (8 TeV) data.

- identified as muon by the Particle Flow algorithm
- $\chi^2/n_{\text{dof}} < 10$
- $n_{\text{valid } \mu \text{ hits}} \geq 1$, $n_{\text{matched stations}} > 1$, $n_{\text{tracker layers}} > 8$ (5)
- $d_{xy}(\text{track}, \text{primary vertex}) < 2\,\text{mm}$, $(d_z(\text{track}, \text{vertex}) < 5\,\text{mm})$
- $n_{\text{pixel hit}} \geq 1$

A.3 Additional plots

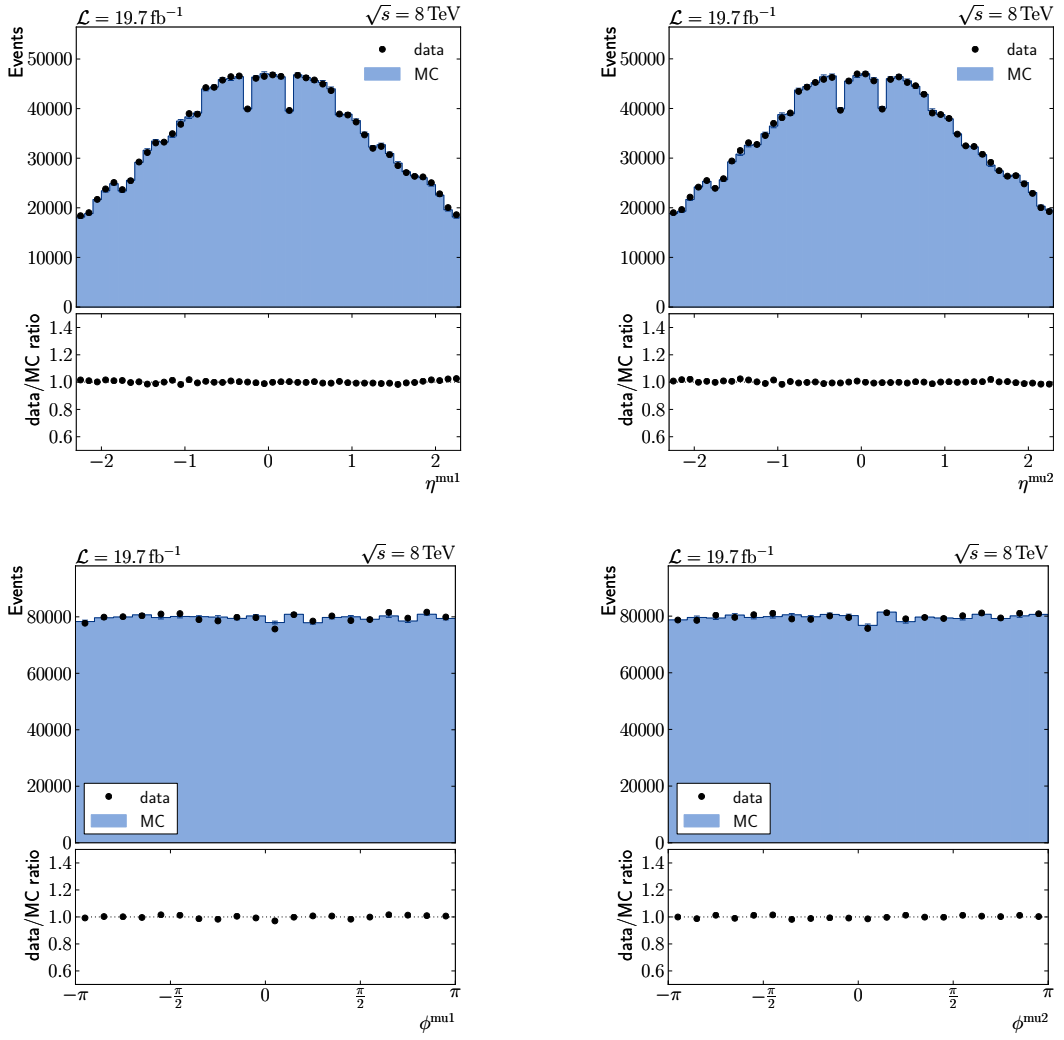


Figure A.1: Muon kinematics: η and ϕ distributions.

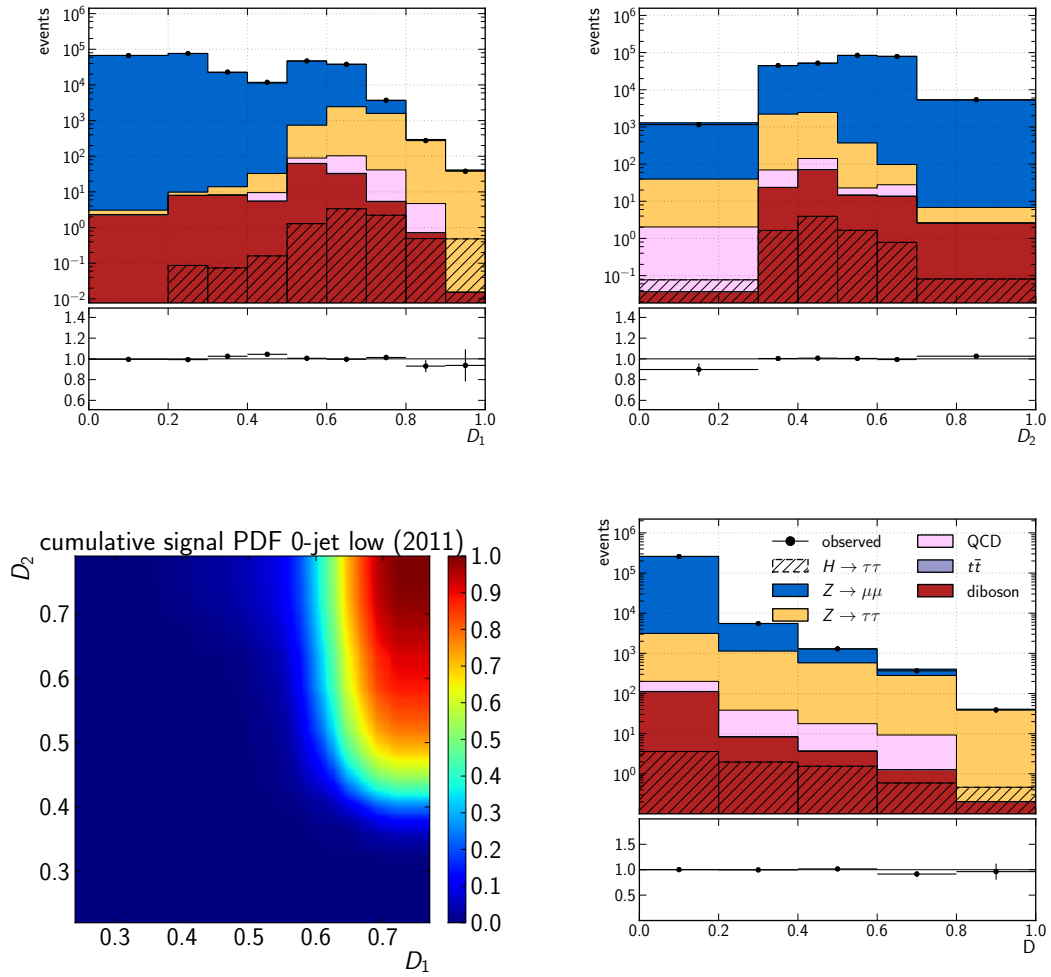


Figure A.2: Discriminators of the two BDTs D_1 and D_2, cumulative signal PDF and the combined discriminator D in the 0-jet low category of the 2011 analysis at 7 TeV using $\mathcal{L} = 4.9\,\text{fb}^{-1}$ of data.

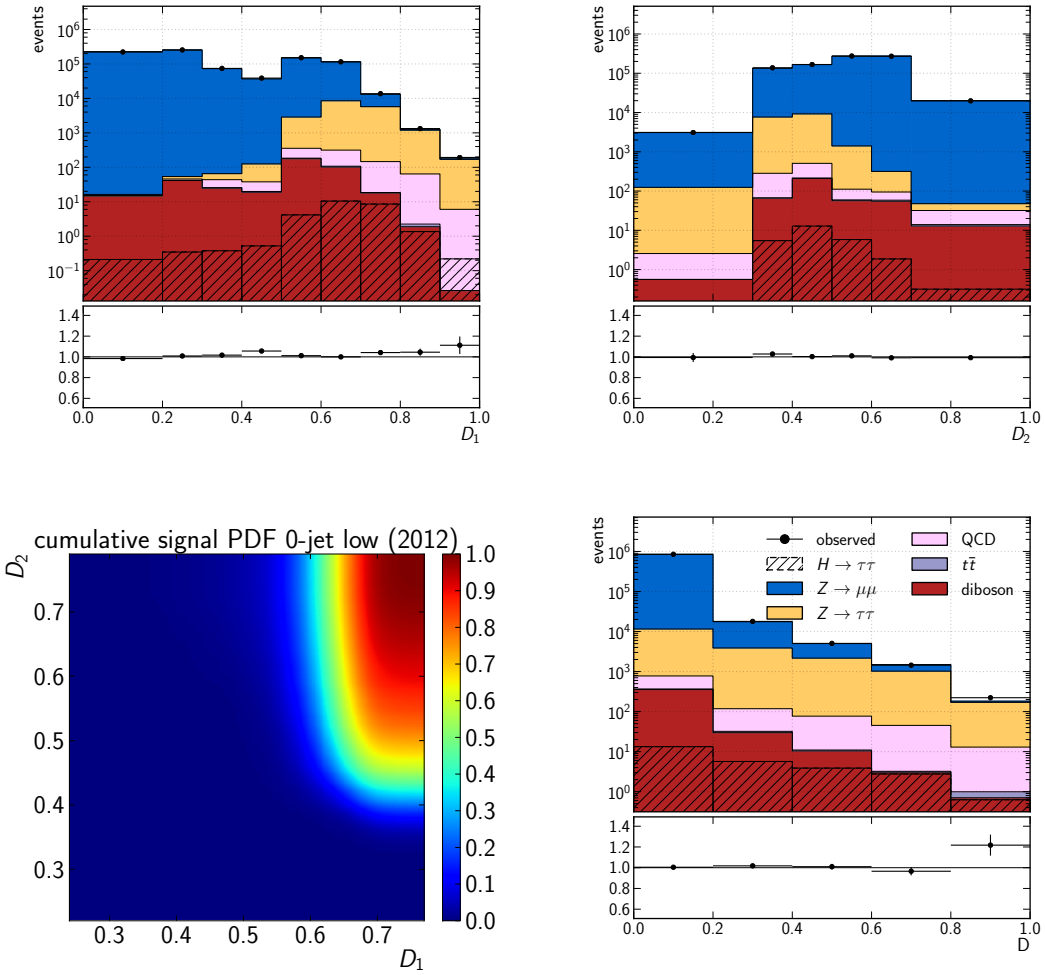


Figure A.3: Discriminators of the two BDTs D_1 and D_2, cumulative signal PDF and the combined discriminator D in the 0-jet low category of the 2012 analysis at 8 TeV using $\mathcal{L} = 19.7\,\mathrm{fb}^{-1}$ of data.

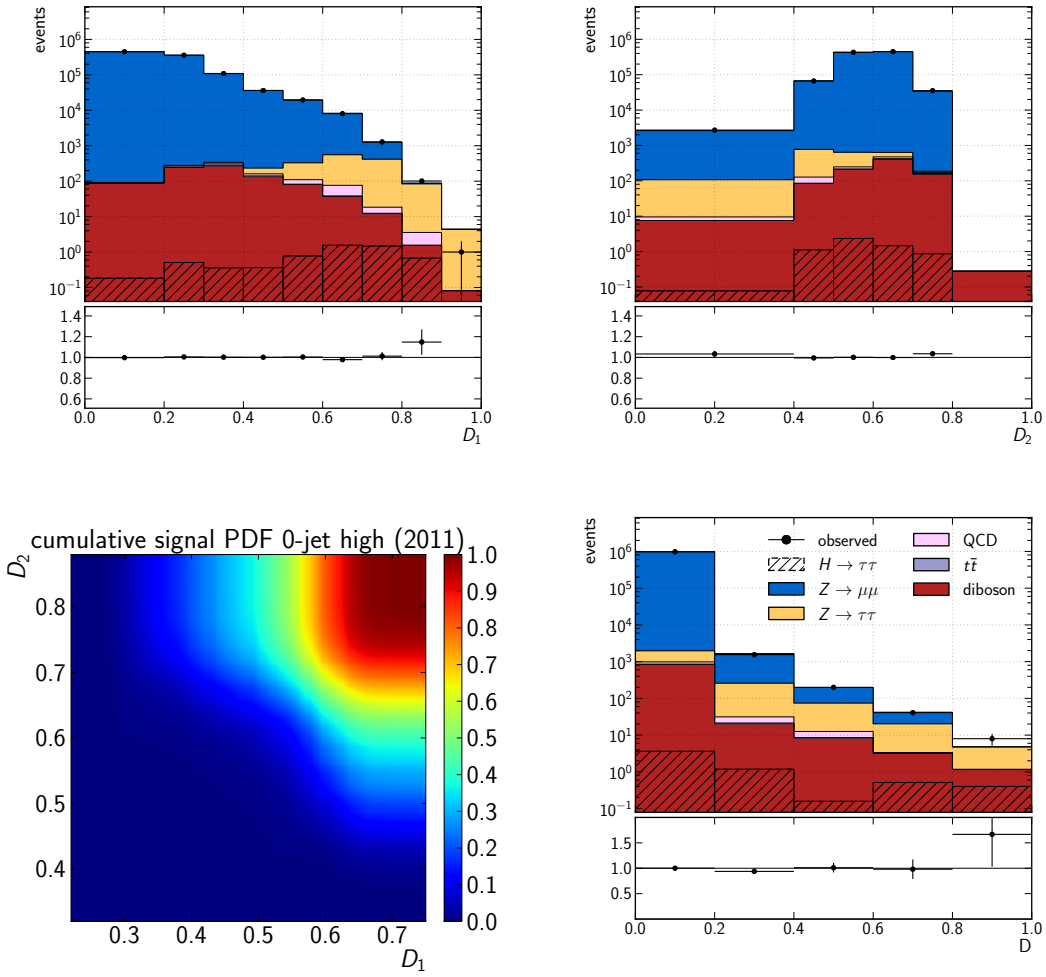


Figure A.4: Same as Figure A.2 for the 0-jet high category.

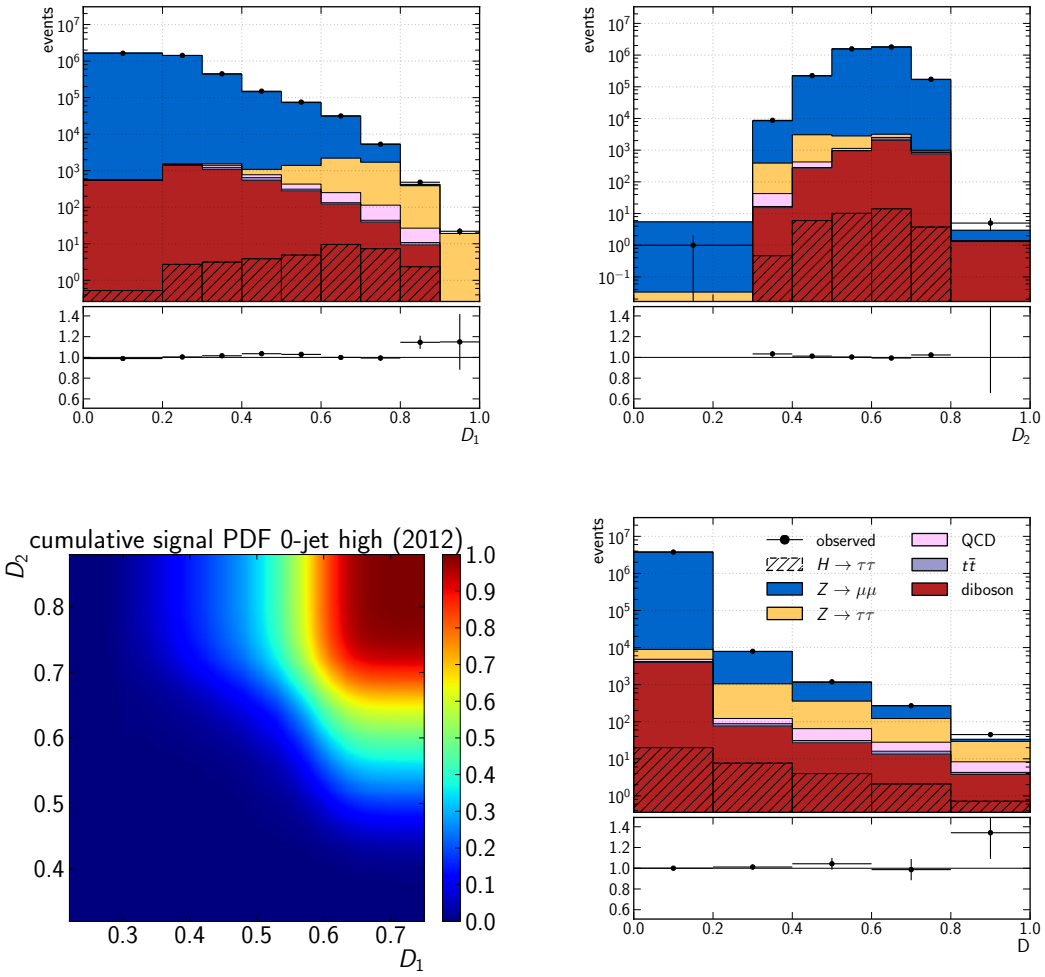


Figure A.5: Same as Figure A.3 for the 0-jet high category.

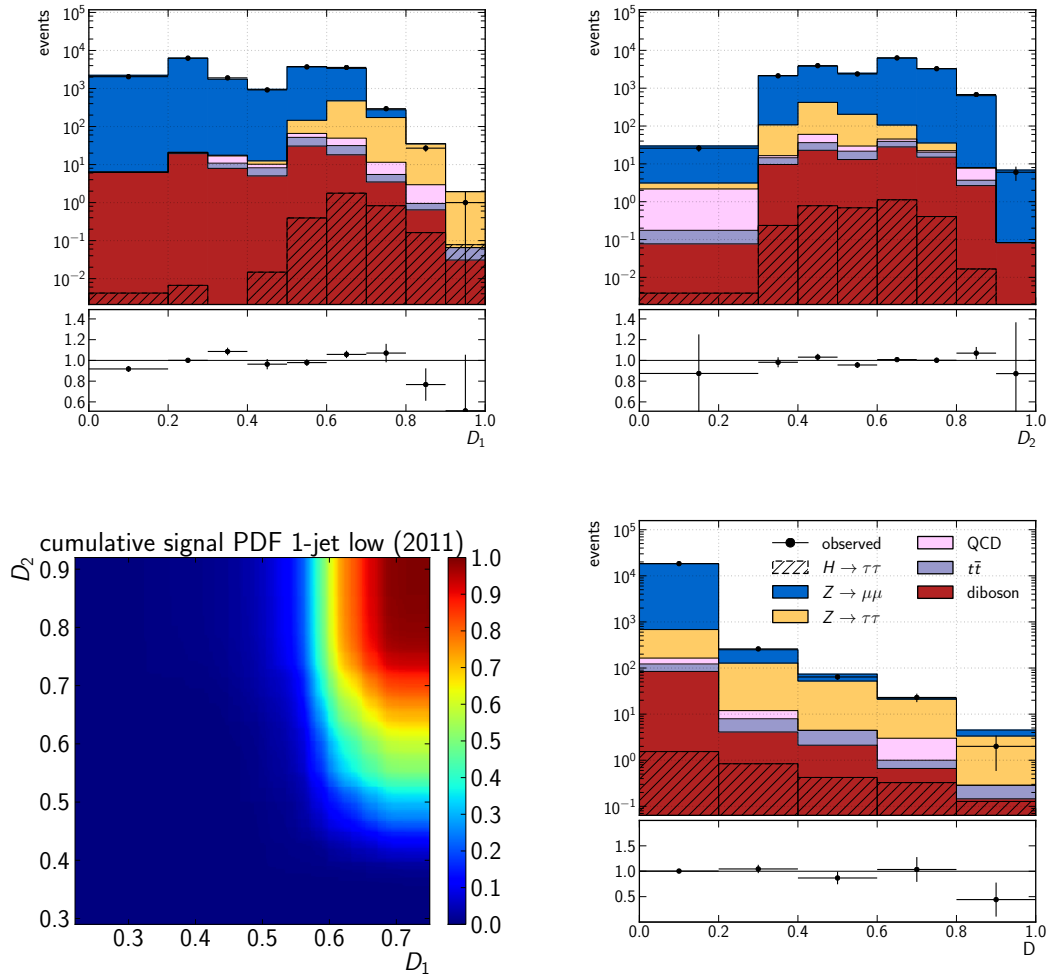


Figure A.6: Same as Figure A.2 for the 1-jet low category.

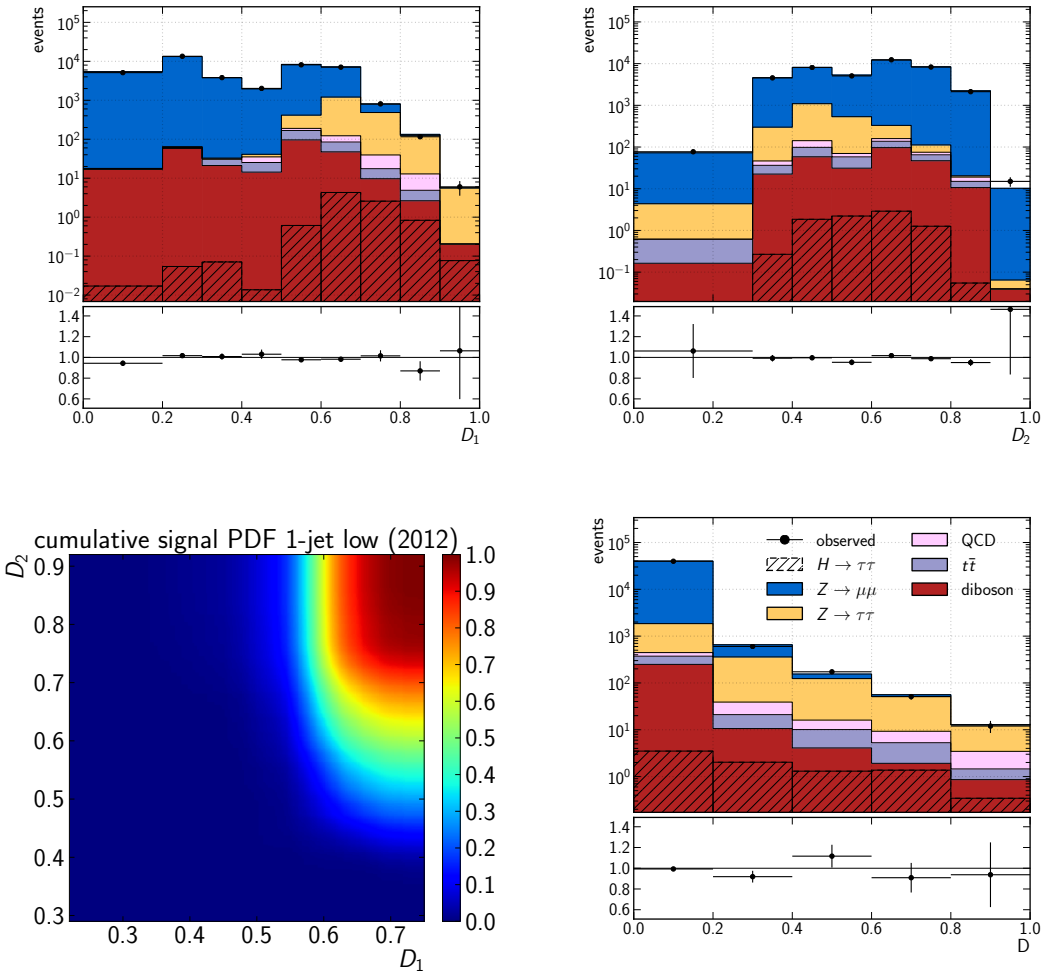


Figure A.7: Same as Figure A.3 for the 1-jet low category.

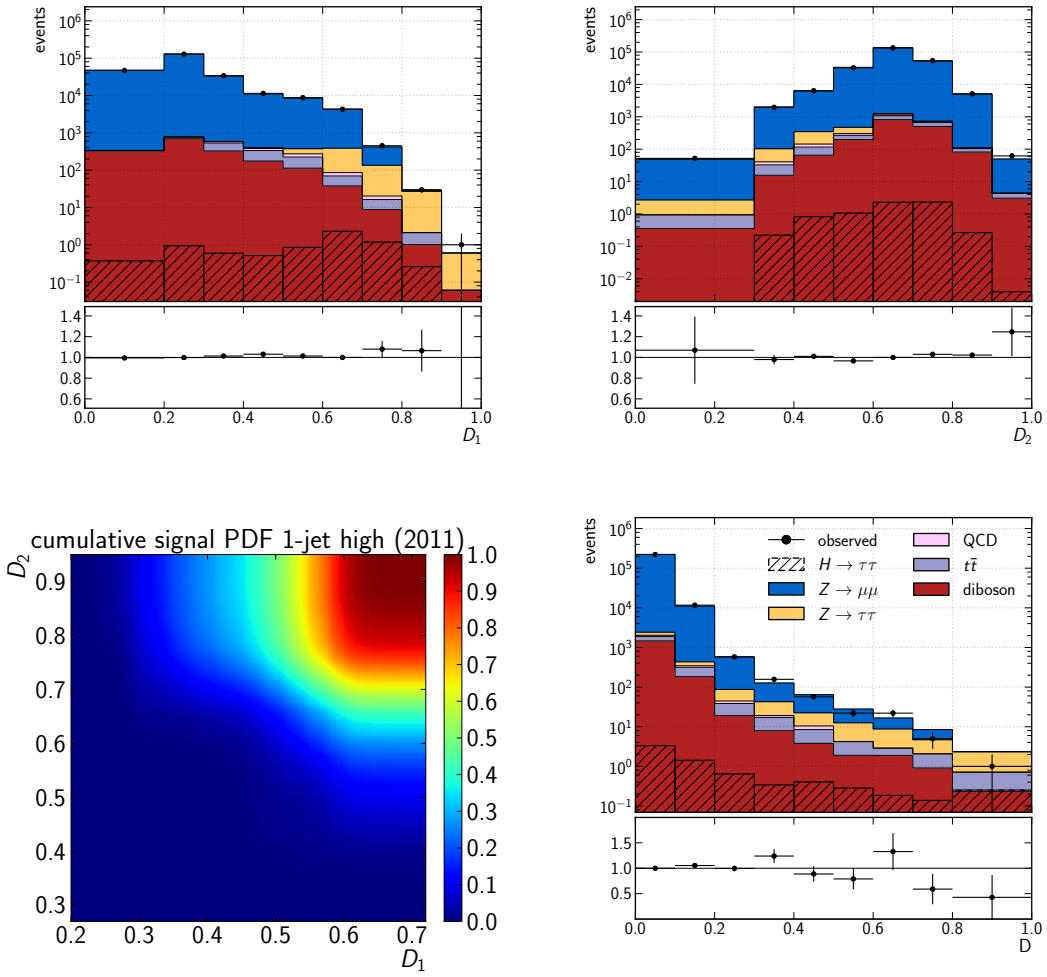


Figure A.8: Same as Figure A.2 for the 1-jet high category.

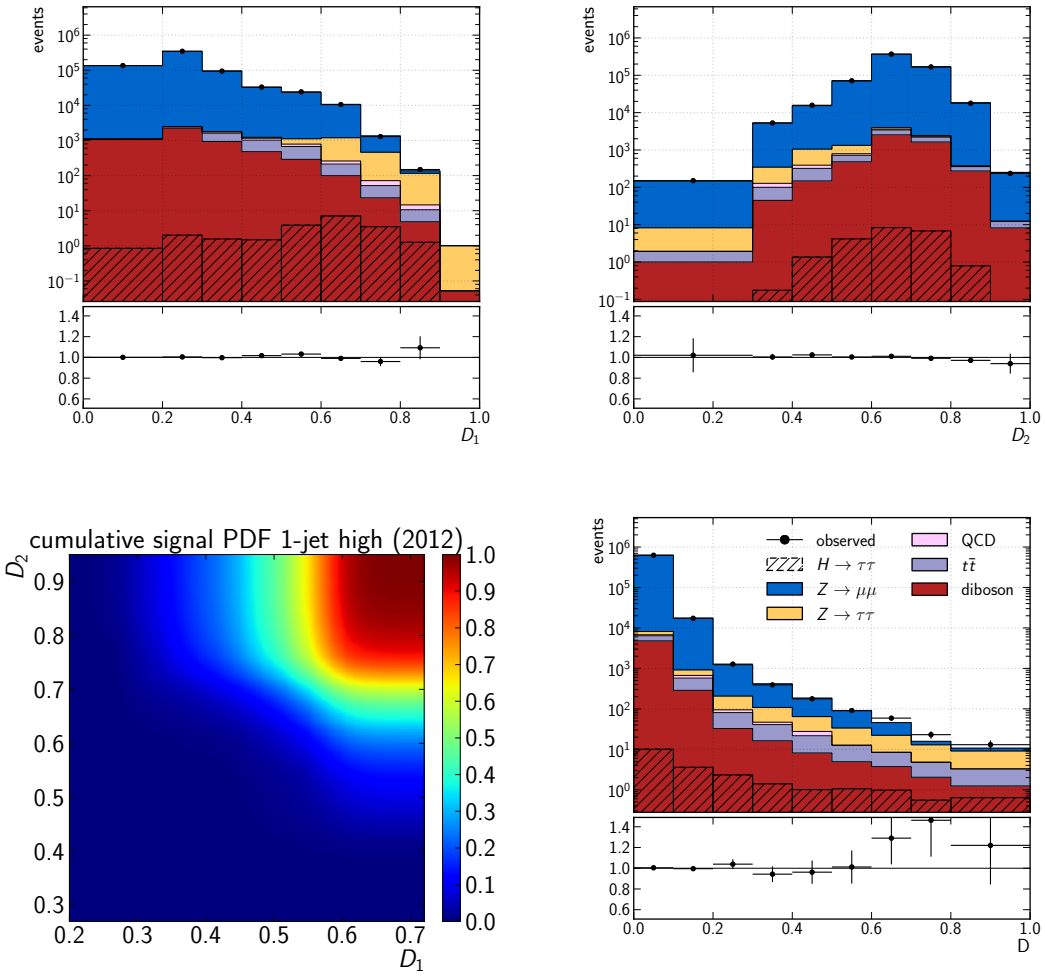


Figure A.9: Same as Figure A.3 for the 1-jet high category.

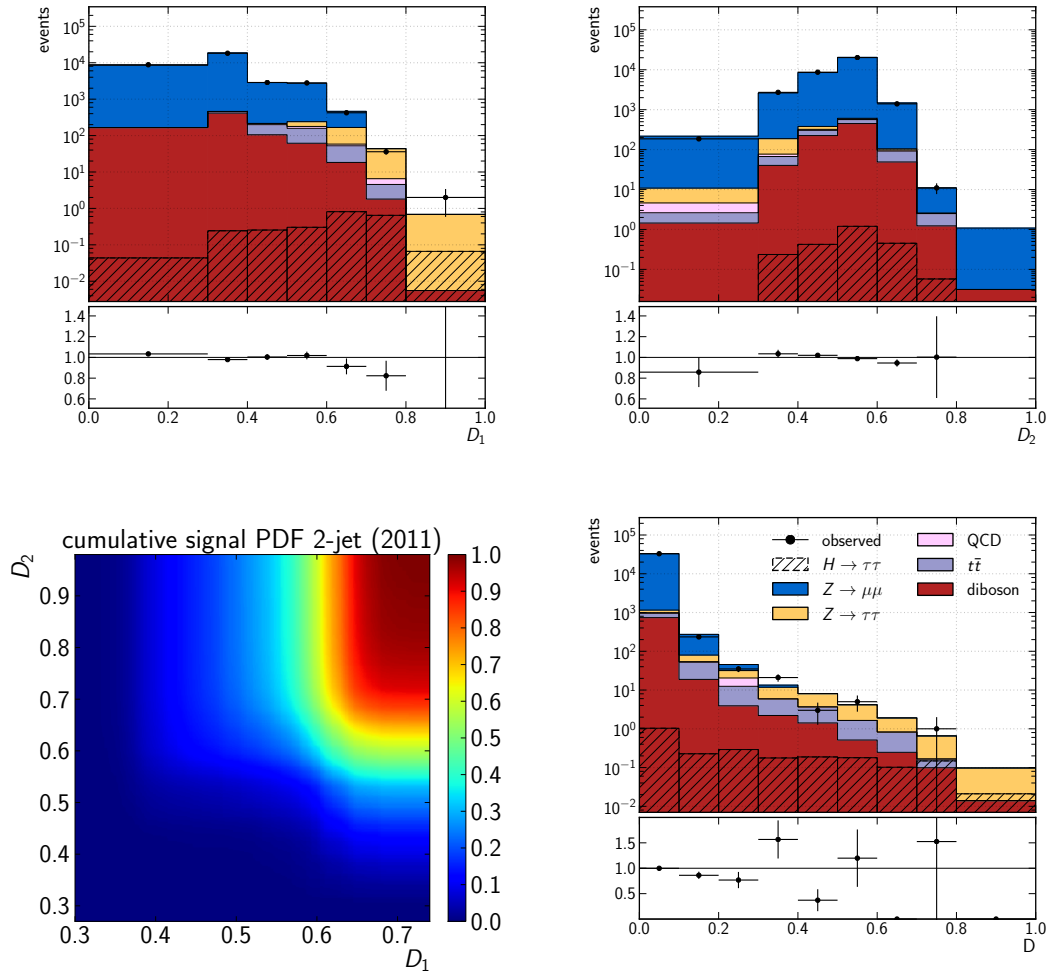


Figure A.10: Same as Figure A.2 for the 2-jet category.

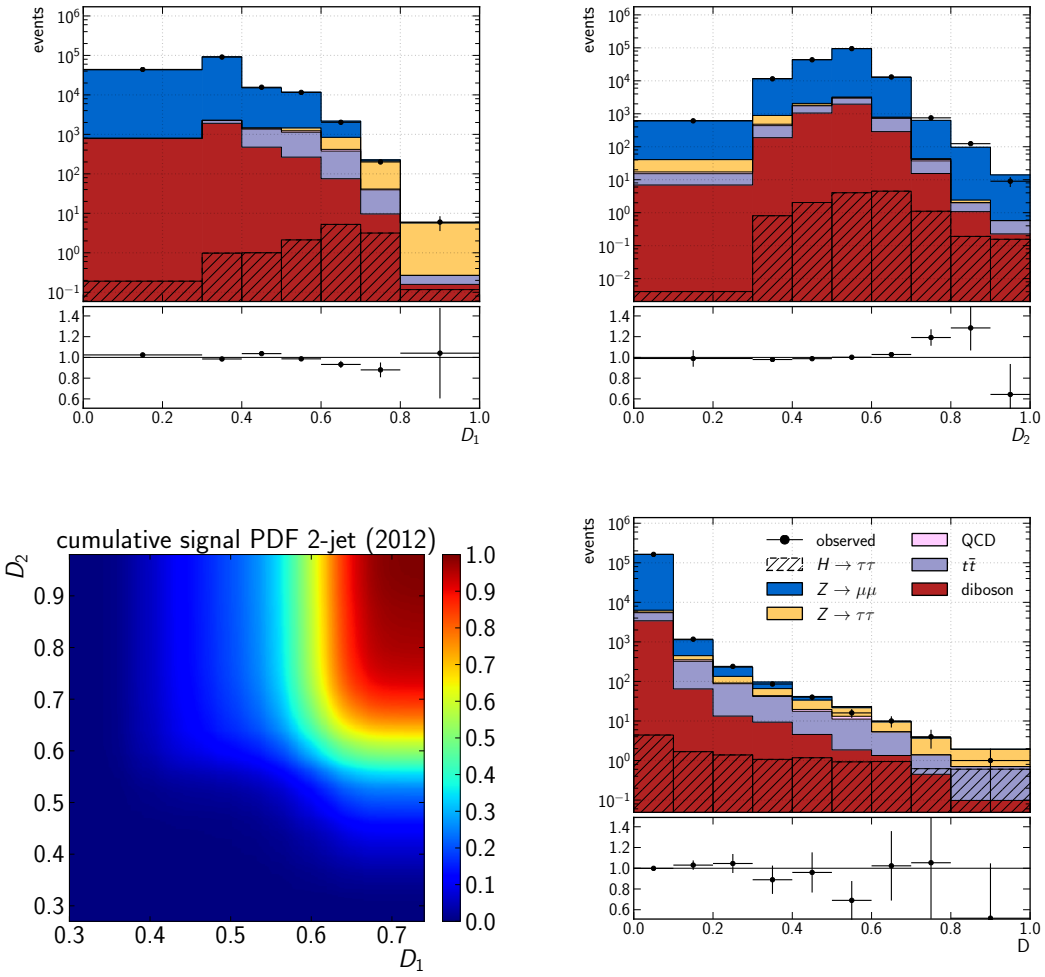


Figure A.11: Same as Figure A.3 for the 2-jet category.

A.4 Software

This is a list of software packages and their versions used for this work.

Software	Version
ROOT	5.32
CMSSW	5.3.9
LHAPDF	5.9.1
POWHEG-Box	1.0
MADGRAPH	5
HERWIG++	2.5.0
ThePEG	1.7.0
PYTHIA	6.426
FASTJET	3.0.1
VALGRIND	3.6.1
BOOST	1.47.0
TAUOLA	27.121.5

List of Acronyms

ALICE	A Large Ion Collider Experiment ... 30
ATLAS	A Toroidal LHC ApparatuS
BDT	Boosted Decision Tree ... 124
CERN	European Organisation for Nuclear Research ... 29
CHS	Charged Hadron Subtraction ... 47
CMS	Compact Muon Solenoid ... 36
CMSSW	CMS Software Framework ... 51
ECAL	electromagnetic calorimeter ... 40
EDM	Event Data Model ... 51
FSR	final state radiation ... 63
HCAL	hadronic calorimeter ... 41
LEP	Large Electron-Positron Collider ... 29
LHC	Large Hadron Collider ... 19
LHCb	Large Hadron Collider Beauty ... 30
MPF	Missing E_{T} Projection Fraction ... 63
MSTW	Martin-Stirling-Thorne-Watt parton distribution functions ... 26
MVA	multivariate analysis ... 124
NLO	next-to-leading order ... 19
NNLO	next-to-next-to-leading order ... 19
NNPDF	Next-to-next-to-leading order PDF ... 26
PDF	parton distribution function ... 26
PDG	Particle Data Group ... 73
PU	pile-up ... 34
QCD	quantum chromodynamics ... 20
SI	International System of Units ... 10
TOTEM	Total elastic and diffractive cross section measurement ... 30
VBF	vector boson fusion ... 20

List of Figures

List of Tables

Bibliography

All references to CERN publications can be found on the CERN Document Server: cds.cern.ch. References to theses published at the Institut für Experimentelle Kernphysik are available at ekp.kit.edu/invenio. For arXiv preprints, please refer to arxiv.org. The digital object identifier (DOI) is used to uniquely identify referenced papers and can be resolved on dx.doi.org.

[1] J. Beringer et al., "Review of particle physics, 2012-2013. Review of particle properties", *Phys. Rev. D* **86** (2012) 010001, DOI: 10.1103/PhysRevD.86.010001.

[2] Nicola Cabibbo, "Unitary symmetry and leptonic decays", *Phys. Rev. Lett.* **10** (1963) 531–533, DOI: 10.1103/PhysRevLett.10.531.

[3] Makoto Kobayashi and Toshihide Maskawa, "CP violation in the renormalizable theory of weak interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, DOI: 10.1143/PTP.49.652.

[4] "Search for nucleon decay into charged anti-lepton plus meson in Super-Kamiokande I and II", *Phys. Rev. D* **85** (2012) 112001, DOI: 10.1103/PhysRevD.85.112001 [arXiv:1203.4030].

[5] Bureau International des Poids et Mesures, "Le Systéme International d'Unités (SI), The International System of Units (SI), 8th edition", *SI brochure* **8** (2006) 1–180, URL: http://www.bipm.org/utils/common/pdf/si_brochure_8.pdf.

[6] Peter J. Mohr, Barry N. Taylor and David B. Newell, "CODATA recommended values of the fundamental physical constants: 2010", *Rev. Mod. Phys.* **84** (2012) 1527–1605, DOI: 10.1103/RevModPhys.80.633 [arXiv:1203.5425].

[7] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, DOI: 10.1103/PhysRevLett.13.321.

[8] Peter W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133, DOI: 10.1016/0031-9163(64)91136-9.

[9] Peter W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, DOI: 10.1103/PhysRevLett.13.508.

[10] G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (1964) 585–587, DOI: 10.1103/PhysRevLett.13.585.

[11] T. W. B. Kibble, "Symmetry breaking in non-abelian gauge theories", *Phys. Rev.* **155** (1967) 1554–1561, DOI: 10.1103/PhysRev.155.1554.

[12] Peter W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156–1163, DOI: 10.1103/PhysRev.145.1156.

[13] Abdelhak Djouadi, "The anatomy of electro-weak symmetry breaking. I: The Higgs boson in the Standard Model", *Phys. Rept.* **457** (2008) 1–216, DOI: 10.1016/j.physrep.2007.10.004 [arXiv:hep-ph/0503172].

[14] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, DOI: 10.1016/j.physletb.2012.08.021 [arXiv:1207.7235].

[15] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, DOI: 10.1016/j.physletb.2012.08.020 [arXiv:1207.7214].

[16] Christoph Hackstein, "Searches for the Higgs boson at the LHC based on its couplings to vector bosons", PhD thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2011, IEKP-KA/2011-19.

[17] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs cross sections: 3. Higgs properties*, 2013, DOI: 10.5170/CERN-2013-004 [arXiv:1307.1347].

[18] D. Graudenz, M. Spira and P. M. Zerwas, "QCD corrections to Higgs boson production at proton-proton colliders", *Phys. Rev. Lett.* **70** (1993) 1372, DOI: 10.1103/PhysRevLett.70.1372.

[19] M. Spira et al., "Higgs boson production at the LHC", *Nucl. Phys. B* **453** (1995) 17–82, DOI: 10.1016/0550-3213(95)00379-7 [arXiv:hep-ph/9504378].

[20] Robert V. Harlander and William B. Kilgore, "Next-to-next-to-leading order Higgs production at hadron colliders", *Phys. Rev. Lett.* **88** (2002) 201801, DOI: 10.1103/PhysRevLett.88.201801 [arXiv:hep-ph/0201206].

[21] Charalampos Anastasiou and Kirill Melnikov, "Higgs boson production at hadron colliders in NNLO QCD", *Nucl. Phys. B* **646** (2002) 220–256, DOI: 10.1016/S0550-3213(02)00837-4 [arXiv:hep-ph/0207004].

[22] V. Ravindran, J. Smith and W. L. van Neerven, "NNLO corrections to the total cross-section for Higgs boson production in hadron-hadron collisions", *Nucl. Phys. B* **665** (2003) 325–366, DOI: 10.1016/S0550-3213(03)00457-7 [arXiv:hep-ph/0302135].

[23] Alexey Pak, Mikhail Rogal and Matthias Steinhauser, "Virtual three-loop corrections to Higgs boson production in gluon fusion for finite top quark mass", *Phys. Lett. B* **679** (2009) 473–477, DOI: 10.1016/j.physletb.2009.08.016 [arXiv:0907.2998].

[24] Paolo Nason and Carlo Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", *JHEP* **02** (2010) 037, DOI: 10.1007/JHEP02(2010)037 [arXiv:0911.5299v2].

[25] N. Kauer et al., "$H \to WW$ as the discovery mode for a light Higgs boson", *Phys. Lett. B* **503** (2001) 113–120, DOI: 10.1016/S0370-2693(01)00211-8 [arXiv:hep-ph/0012351].

[26] T. Han, G. Valencia and S. Willenbrock, "Structure-function approach to vector-boson scattering in pp collisions", *Phys. Rev. Lett.* **69** (1992) 3274, DOI: 10.1103/PhysRevLett.69.3274 [arXiv:hep-ph/9206246].

[27] M. Ciccolini, A. Denner and S. Dittmaier, "Strong and electroweak corrections to the production of a Higgs boson + 2 jets via weak interactions at the Large Hadron Collider", *Phys. Rev. Lett.* **99** (2007) 161803, DOI: 10.1103/PhysRevLett.99.161803 [arXiv:0707.0381].

[28] M. Ciccolini, A. Denner and S. Dittmaier, "Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC", *Phys. Rev. D* **77** (2008) 013002, DOI: 10.1103/PhysRevD.77.013002 [arXiv:0710.4749].

[29] M. Dührssen et al., "Determination of Higgs-boson couplings at the LHC" (2004) [arXiv:hep-ph/0407190].

[30] A. L. Read, "Presentation of search results: the CL_s technique", *J. Phys. G: Nucl. Part. Phys.* **28** (2002) 2693, DOI: 10.1088/0954-3899/28/10/313.

[31] Bernhard Mistlberger and Falko Dulat, "Limit setting procedures and theoretical uncertainties in Higgs boson searches" (2012) [arXiv:1204.3851].

[32] CMS Collaboration, *Combination of CMS searches for a Standard Model Higgs boson*, CMS Physics Analysis Summary, 2011, CMS-PAS-HIG-11-032.

[33] Michiel Botje et al., "The PDF4LHC Working Group interim recommendations" (2011) [arXiv:1101.0538].

[34] Jun Gao et al., "CT10 Next-to-next-to-leading order global analysis of QCD", *Phys. Rev. D* **89** (2014) 033009, DOI: 10.1103/PhysRevD.89.033009 [arXiv:1302.6246].

[35] A. D. Martin et al., "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009) 189–285, DOI: 10.1140/epjc/s10052-009-1072-5, [arXiv:0901.0002v3].

[36] Richard D. Ball et al., "A first unbiased global NLO determination of parton distributions and their uncertainties", *Nucl. Phys. B* **838** (2010) 136–206, DOI: 10.1016/j.nuclphysb.2010.05.008, [arXiv:1002.4407].

[37] M. R. Whalley, D. Bourilkov and R. C. Group, "The Les Houches Accord PDFs (LHAPDF) and Lhaglue" (2005) [arXiv:0508110].

[38] Nobel Media AB 2013, *The Nobel Prize in Physics 1984*, 2010, URL: http://nobelprize.org/nobel_prizes/physics/laureates/1984 (visited on 05/05/2014).

[39] Nobel Media AB 2013, *The Nobel Prize in Physics 1992*, 2014, URL: http://nobelprize.org/nobel_prizes/physics/laureates/1992/ (visited on 05/05/2014).

[40] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups, "Precision electroweak measurements on the Z resonance", *Phys. Rep.* **427** (2006) 257–454, DOI: 10.1016/j.physrep.2005.12.006.

[41] Tim Berners-Lee, *Information management: A proposal*, 1989, URL: http://www.w3.org/History/1989/proposal.html (visited on 05/05/2014).

[42] Lyndon Evans and Philip Bryant, "LHC machine", *JINST* **3** (2008) S08001, DOI: 10.1088/1748-0221/3/08/S08001.

[43] G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003.

[44] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, DOI: 10.1088/1748-0221/3/08/S08004.

[45] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, DOI: 10.1088/1748-0221/3/08/S08005.

[46] ALICE Collaboration, "The ALICE Experiment at the CERN LHC", *JINST* **3** (2008) S08002, DOI: 10.1088/1748-0221/3/08/S08002.

[47] TOTEM, "The TOTEM Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007, DOI: 10.1088/1748-0221/3/08/S08007.

[48] LHCf Collaboration, "The LHCf Detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006, DOI: 10.1088/1748-0221/3/08/S08006.

[49] Christiane Lefevre, *CERN LHC: The guide*, Geneva, 2009, p. 60, CERN-Brochure-2009-003-Eng.

[50] D. Acosta et al., *CMS physics: technical design report, Volume 1: Detector performance and software*, Geneva, 2006, CERN-LHCC-2006-001.

[51] Mike Lamont, *J. Phys.: Conf. Ser.* **455** (2013) 012001, DOI: 10.1088/1742-6596/455/1/012001.

[52] CMS, *Luminosity – public results*, 2012, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 05/05/2014).

[53] CMS Collaboration, *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*, CERN, 2013, CMS-PAS-LUM-13-001.

[54] Alexx Perloff, "Pileup measurement and mitigation techniques in CMS", *J. Phys.: Conf. Ser.* **404** (2012) 012045, DOI: 10.1088/1742-6596/404/1/012045.

[55] *CMS members*, URL: http://greybook.cern.ch/programmes/experiments/CMS.html (visited on 20/05/2014).

[56] CMS Collaboration, "CMS physics: technical design report, Volume 2: Physics performance", *J. Phys. G* **34** (2007) 995, DOI: 10.1088/0954-3899/34/6/S01.

[57] Joram Berger, "Analysis of Z($\rightarrow \mu\mu$)+Jet events at the LHC", diploma thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2010, IEKP-KA/2010-23.

[58] Michael Hoch and Maximilien Brice, *CMS Experiment*, Geneva, 2008, CMS-PHO-GEN-2008-028.

[59] Katja Klein, "Lessons learned during CMS tracker end cap construction", *Nucl. Instr. Meth. Phys. Res. A* **579** (2007) 731–735, DOI: 10.1016/j.nima.2007.05.283.

[60] Maximilien Brice, "First half of CMS inner tracker barrel" (2006), CERN-EX-0610026.

[61] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, CMS Physics Analysis Summary, 2009, CMS-PAS-PFT-09-001.

[62] CMS Collaboration, *Commissioning of the particle-flow event reconstruction with leptons from J/Ψ and W decays at 7 TeV*, CMS Physics Analysis Summary, 2010, CMS-PAS-PFT-10-003.

[63] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, DOI: 10.1088/1748-0221/7/10/P10002 [arXiv:1206.4071].

[64] Matteo Cacciari and Gavin P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006-08) 57–61, DOI: 10.1016/j.physletb.2006.08.037 [arXiv:hep-ph/0512210v2].

[65] Matteo Cacciari, Gavin P. Salam and Gregory Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, DOI: 10.1140/epjc/s10052-012-1896-2 [arXiv:1111.6097].

[66] Gavin P. Salam and Gregory Soyez, "A practical seedless infrared-safe cone jet algorithm", *JHEP* **05** (2007) 086, DOI: 10.1088/1126-6708/2007/05/086 [arXiv:0704.0292v2].

[67] J. M. Butterworth et al., "KtJet: A C++ implementation of the $K_\perp$ clustering algorithm", *Comput. Phys. Commun.* **153** (2003) 85–96, DOI: 10.1016/S0010-4655(03)00156-5 [arXiv:hep-ph/0210022].

[68] S. Catani, Yu.L. Dokshitzer and B.R. Webber, "The $k_\perp$-clustering algorithm for jets in deep inelastic scattering and hadron collisions", *Phys. Lett. B* **285** (1992) 291–299, DOI: 10.1016/0370-2693(92)91467-N.

[69] Matteo Cacciari, Gavin P. Salam and Gregory Soyez, "The Anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, DOI: 10.1088/1126-6708/2008/04/063.

[70] Matteo Cacciari, Gavin P Salam and Gregory Soyez, "The catchment area of jets", *JHEP* **04** (2008) 005, DOI: 10.1088/1126-6708/2008/04/005 [arXiv:0802.1188].

[71] CMS Collaboration, *Measurement of the Underlying Event Activity with the Jet Area/Median Approach at* 0.9 *TeV*, CMS Physics Analysis Summary, 2010, CMS-PAS-QCD-10-005.

[72] Rene Brun and Fons Rademakers, "ROOT - An Object Oriented Data Analysis Framework", *Nucl. Instrum. Meth. A* **389** (1997) 81–86, DOI: 10.1016/S0168-9002(97)00048-X.

[73] ROOT team, *ROOT*, 5.26, CERN, Geneva, 2009, URL: http://root.cern.ch (visited on 05/05/2014).

[74] Fred Stober, "Measurement of the three-jet mass cross-section at $\sqrt{s} = 7$ TeV", PhD thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2012, IEKP-KA/2012-24.

[75] The grid-control team, *Grid-control: A job submission tool*. KIT, Karlsruhe, 2010, URL: https://ekptrac.physik.uni-karlsruhe.de/trac/grid-control.

[76] Stephen Burke et al., *gLite 3.2 user guide*, Geneva, 2011, CERN-LCG-GDEIS-722398.

[77] Matteo Cacciari and Gavin P Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126, DOI: 10.1016/j.physletb.2007.09.077 [arXiv:0707.1378].

[78] CMS Collaboration, *8 TeV jet energy corrections and uncertainties based on 19.8 fb^{-1} of data in CMS*, CMS Detector Performance Summary, 2013, CMS-DP-2013-033.

[79] CMS Collaboration, *Jet energy corrections and uncertainties. Detector performance plots for 2012*, CMS Detector Performance Summary, 2012, CMS-DP-2012-012.

[80] Volker Büge, "Virtualisation of grid resources and prospects of the mesurement of Z boson production in association with jets at the LHC", PhD thesis, Universität Karlsruhe (TH), 2008, IEKP-KA/2008-18.

[81] Danilo Piparo, "Statistical combination of Higgs decay channels and determination of the jet-energy scale of the CMS experiment at the LHC", PhD thesis, Karlsruhe: Karlsruhe Institut of Technology, 2010, IEKP-KA/2010-22.

[82] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, DOI: 10.1088/1748-0221/6/11/P11002.

[83] Thomas Hauth, "Dynamic extension of local batch systems with cloud resources and measurement of the jet energy scale with the CMS detector", diploma thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2011, IEKP-KA/2011-07.

[84] Dominik Haitz, "Analyse von Z$\rightarrow \mu^+\mu^-$+Jet-Ereignissen und Kalibration der Jet-Energieskala des CMS-Detektors", diploma thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2013, IEKP-KA/2013-12.

[85] DØ Collaboration, "Determination of the absolute jet energy scale in the DØ calorimeter", *Nucl. Instrum. Meth. A* **424** (1999) 352–394, DOI: 10.1016/S0168-9002(98)01368-0.

[86] CMS Collaboration, *Pileup Jet Identification*, CMS Physics Analysis Summary, 2013, CMS-PAS-JME-13-005.

[87] CMS Collaboration, *Single muon efficiencies in 2012 data*, CMS Detector Performance Summary, 2013, CMS-DP-2013-009.

[88] Carlo Oleari, "The POWHEG-BOX", *Nucl. Phys. B* **205-206** (2010) 36–41, DOI: 10.1016/j.nuclphysbps.2010.08.016 [arXiv:1007.3893v1].

[89] Johan Alwall et al., "MadGraph 5: going beyond", *JHEP* **06** (2011) 128, DOI: 10.1007/JHEP06(2011)128 [arXiv:1106.0522v1].

[90] Torbjörn Sjöstrand, Stephen Mrenna and Peter Skands, "PYTHIA 6.4 – physics and manual", *JHEP* **05** (2006) 026, DOI: 10.1088/1126-6708/2006/05/026 [arXiv:hep-ph/0603175].

[91] Manuel Bähr et al., "Herwig++ physics and manual", *Eur. Phys. J. C* **58** (2008) 639–707, DOI: 10.1140/epjc/s10052-008-0798-9 [arXiv:0803.0883].

[92] CMS, *Standard Model cross sections for CMS at 8 TeV*, 2013, URL: https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat8TeV (visited on 05/05/2014).

[93] Ryan Gavin et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388–2403, DOI: 10.1016/j.cpc.2011.06.008.

[94] S. Alekhin, "The NNLO predictions for the rates of the W/Z production in pp and $\bar{p}p$ collisions" (2003) [arXiv:hep-ph/0307219].

[95] M. Marionneau and C. Veelken, *Study of E_{T} performance in the full 2012 dataset using $Z \to \mu\mu$ events*, 2013, CMS-AN-2013-233.

[96] Michał Czakon, Paul Fiedler and Alexander Mitov, "The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, DOI: 10.1103/PhysRevLett.110.252004 [arXiv:1303.6254].

[97] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, DOI: 10.1088/1748-0221/8/04/P04013, [arXiv:1211.4462v2].

[98] CMS Collaboration, *Performance of Quark/Gluon Discrimination in 8 TeV pp Data*, CMS Physics Analysis Summary, 2013, CMS-PAS-JME-13-002.

[99] CMS Collaboration, *Jet energy scale uncertainty sources*, 2014, URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources (visited on 05/05/2014).

[100] CMS Collaboration, *Jet energy scale performance in 2011*, CMS Detector Performance Summary, 2012, CMS-DP-2012-006.

[101] CMS Collaboration, *Status of the 8 TeV jet energy corrections and uncertainties based on 11 fb^{-1} of data in CMS*, CMS Detector Performance Summary, 2013, CMS-DP-2013-011.

[102] CMS Collaboration, *Jet energy calibration in the 8 TeV pp data*, to be published in JINST, 2014, CMS-PAS-JME-13-004.

[103] CMS Collaboration, *Search for the standard model Higgs boson produced in vector boson fusion, and decaying to bottom quarks*, CMS Physics Analysis Summary, 2013, CMS-PAS-HIG-13-011.

[104] CMS Collaboration, *Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks (LHCP 2013)*, CMS Physics Analysis Summary, 2013, CMS-PAS-HIG-13-012.

[105] CMS Collaboration, *Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons*, CMS Physics Analysis Summary, 2013, CMS-PAS-HIG-13-019.

[106] CMS Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *JHEP* **05** (2014) 104, DOI: 10.1007/JHEP05(2014)104 [arXiv:1401.5041].

[107] CMS Collaboration, *Search for MSSM Neutral Higgs Bosons Decaying to a Pair of Tau Leptons in pp Collisions*, CMS Physics Analysis Summary, 2014, CMS-PAS-HIG-13-021.

[108] B. K. Bullock, K. Hagiwara and A. D. Martin, "τ polarization and its correlations as a probe of new physics", *Nucl. Phys. B* **395** (1993) 499, DOI: 10.1016/0550-3213(93)90045-Q.

[109] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, DOI: 10.1088/1748-0221/6/09/P09001, [arXiv:1106.5048].

[110] Armin Burgmeier, "Position resolution and upgrade of the CMS pixel detector and search for the Higgs boson in the $\tau^+\tau^-$ final state", PhD thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2014, IEKP-KA/2014-09.

[111] Pietro Govoni and Chiara Mariotti, "Z production via vector boson fusion at LHC" (2010) [arXiv:1001.4357].

[112] Joram Berger et al., *Search for Higgs boson decays to τ pairs in the di-muon and di-electron channels*, 2013, CMS-AN-2013-192.

[113] Manuel Zeise, "Study of Z boson decays into pairs of muon and τ-leptons with the CMS detector at the LHC", PhD thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2011, IEKP-KA/2011-20.

[114] Thomas Müller, "Unterdrückung des Z-Untergrundes zur Higgs-Suche im Kanal $H \to \tau\tau$ durch Multivariate Analyse von $\tau\tau$-Endzuständen in pp-Kollisionen am LHC", master's thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2012, IEKP-KA/2012-14.

[115] Raphael Friese, "Multivariate Methoden zur Identifikation von $H \to \tau\tau \to \mu\mu$-Zerfällen", master's thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2013, IEKP-KA/2013-14.

[116] Thomas Müller, "Search for the Higgs boson in the $H \to \tau\tau \to \mu\mu$ channel", PhD thesis in preparation, Karlsruhe: Karlsruhe Insitute of Technology, 2015.

[117] S. Agostinelli et al., "Geant4 simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303, DOI: 10.1016/S0168-9002(03)01368-8.

[118] J. Allison et al., "Geant4 developments and applications", *IEEE Transactions on Nuclear Science* **53** (2006) 270–278, DOI: 10.1109/TNS.2006.869826.

[119] Paolo Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, DOI: 10.1088/1126-6708/2004/11/040 [arXiv:hep-ph/0409146].

[120] Stefano Frixione, Paolo Nason and Carlo Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, DOI: 10.1088/1126-6708/2007/11/070 [arXiv:0709.2092].

[121] Simone Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, DOI: 10.1007/JHEP06(2010)043 [arXiv:1002.2581].

[122] Stanisław Jadach et al., "The τ decay library TAUOLA, version 2.4", *Comput. Phys. Commun.* **76** (1993) 361–380, DOI: 10.1016/0010-4655(93)90061-G.

[123] Zbigniew Wąs, "TAUOLA the library for τ lepton decay, and KKMC/KORALB/KORALZ/... status report", *Nucl. Phys. B – Proc. Suppl.* **98** (2001) 96–102, DOI: 10.1016/S0920-5632(01)01200-2 [arXiv:hep-ph/0011305].

[124] A. Apyan et al., *Physics objects in the Higgs to $\tau\tau$ analysis*, 2013, CMS-AN-2013-188.

[125] Andreas Hoecker et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS(ACAT2007)040* (2007) [arXiv:physics/0703039].

[126] D. Rainwater, R. Szalapski and D. Zeppenfeld, "Probing color-singlet exchange in Z+2-jet events at the CERN LHC", *Phys. Rev. D* **54** (1996) 6680.

[127] Max Fischer, "Workflow-Management auf verteilten Computingsystemen für Analysen in der Teilchenphysik", diploma thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2013, IEKP-KA/2013-16.

[128] CMS Collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of τ-leptons*, 2014, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig13004PubTWiki (visited on 05/05/2014).

[129] Yuta Takahashi et al., *Theoretical uncertainty for the Higgs production via VBF and gluon fusion process*, 2013, CMS-AN-2013-262.

[130] Keith Hamilton, Paolo Nason and Giulia Zanderighi, "MINLO: Multi-scale improved NLO", *JHEP* **10** (2012) 155, DOI: 10.1007/JHEP10(2012)155 [arXiv:1206.3572].

[131] Pavel M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev. D* **78** (2008) 013004, DOI: 10.1103/PhysRevD.78.013004 [arXiv:0802.0007].

[132] CMS Collaboration, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions", *Nature Physics* (2014), DOI: doi:10.1038/nphys3005 [arXiv:1401.6527].

[133] CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state", *Phys. Rev. D* **89** (2014) 092007, DOI: 10.1103/PhysRevD.89.092007 [arXiv:1312.5353].

[134] René Caspart, "Determination of the coupling of the Higgs boson to τ-leptons by statistical combination of different production and decay channels with data from the CMS experiment at the LHC", master's thesis, Karlsruhe: Karlsruhe Insitute of Technology, 2014, IEKP-KA/2014-16.

Dank

Ich möchte mich bei Prof. Günter Quast für die ausgezeichnete Betreuung und die umfassende Unterstützung während der ganzen Zeit und auch für den Aufenthalt am CERN bedanken. Ebenso gilt mein Dank Prof. Wim de Boer für die Übernahme des Korreferats.

Desweiteren möchte ich mich bei allen bedanken, mit denen ich in den letzten Jahren zusammengearbeitet habe. Bei Volker Büge und Danilo Piparo für die Einarbeitung in die Thematik. Besonderer Dank gilt Thomas Hauth und Dominik Haitz für die erfolgreiche Zusammenarbeit in unterschiedlichen Zeitabschnitten im Bereich der Jetenergie-Kalibration. Außerdem möchte ich mich bei Mikko Voutilainen und allen Subgroup-Convenorn für die vielen guten Anregungen bedanken.

Ich möchte mich auch bei der Gruppe zur Higgs-Suche im $\mu\mu$-Endzustand unter der Leitung von Alexei Raspareza für die freundliche Zusammenarbeit bedanken, bei Thomas Müller, Raphael Friese und Fabio Colombo. Ebenso für die guten Diskussionen über die statistische Kombination mit René Caspart und für die vielen guten Anregungen von Roger Wolf.

Ebenso danke ich Max Fischer, Daniel Funke, Klaus Rabbertz, Georg Sieber, Fred Stober und Manuel Zeise für die gute Arbeitsatmosphäre, die aufschlussreichen Diskussionen und für viele gute Ideen.

Von Herzen möchte ich mich bei meiner geliebten Frau Krystyna bedanken, die mich in der ganzen Zeit so liebevoll unterstützt hat. Ich bedanke mich bei meinen Eltern für alle Unterstützung und bei allen die mich im Gebet begleitet haben. Vor allem gebührt der Dank Gott, der mich getragen hat.