Measurement of the coupling strength of the Higgs boson to τ leptons with the CMS experiment at the LHC

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

CERN-THESIS-2019-262
//2019



vorgelegt von

Valeria Botta

aus Perugia, Italien

Hamburg
2019

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, 02. Mai 2019

Valeria Botta

Gutachter/innen der Dissertation:	Dr. Alexei Raspereza Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	Prof. Dr. Dieter Horns Prof. Dr. Elisabetta Gallo Dr. Rainer Mankel Dr. Roman Kogler Prof. Dr. Gudrid Moortgat-Pick
Vorsitzende der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	19. Juni 2019
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Michael Potthoff
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang Hansen
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

I have a friend who's an artist and he's sometimes taken a view which I don't agree with very well. He'll hold up a flower and say, "Look how beautiful it is!" and I'll agree, I think. And he says: "You see, I as an artist can see how beautiful this is, but you as a scientist, oh, take this all apart and it becomes a dull thing." [...]

But I can appreciate the beauty of a flower. I can imagine the cells in there, the complicated actions inside which also have a beauty. I mean, it's not just beauty at this dimension of one centimeter, there is also beauty at a smaller dimension, the inner structure. Also the processes, the fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting - it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic?



All kinds of interesting questions, which shows that a science knowledge only adds to the excitement and mystery and awe of a flower. It only adds, I don't understands how it subtracts.

R. Feynman,
The pleasure of finding things out

Abstract

This thesis presents the measurement of the coupling strength of the 125 GeV Higgs boson to τ leptons. The measurement is performed on proton-proton collision data at a centre-of-mass energy of 13 TeV, collected by the CMS experiment at the LHC in the year 2016, and corresponding to an integrated luminosity of 35.9 fb^{-1}. The main focus of this work lies on the $\mu\tau_h$ and $e\tau_h$ final states - where τ_h indicates a hadronically decaying τ lepton - amounting to about 45 % of possible $\tau\tau$ decays. For each final state, selected events are classified into exclusive signal categories, targeting Higgs boson production through gluon-gluon fusion and vector boson fusion. A pair of kinematic variables is chosen for every category to disentangle the Higgs signal from the main background processes, and the corresponding two-dimensional distributions are used to extract the signal by means of a maximum-likelihood fit. The final result of this analysis combines the measurements in the $\mu\tau_h$ and $e\tau_h$ decay channels with the ones in the $\tau_h\tau_h$ and $e\mu$ final states. From the statistical combination of these four final states, the decay of the Higgs boson to a pair of τ leptons is observed with a local significance of 4.9 standard deviations, with the expectation being 4.7 σ. The signal strength is determined to be $\mu = \sigma/\sigma_{SM} = 1.09^{+0.27}_{-0.26} = 1.09 \pm 0.15(\text{stat.})^{+0.16}_{-0.15}(\text{syst.})^{+0.10}_{-0.08}(\text{theo.})^{+0.13}_{-0.12}(\text{bbb.})$, which is fully compatible with the Standard Model expectation within the uncertainty.

Furthermore, this thesis also contributed to the calibration of the CMS detector in the context of tracker alignment. At the end of 2016, the CMS pixel tracker has been replaced with a new upgraded version of it. Several studies on simulated events have been performed, with the aim of elaborating a strategy for the first track-based alignment of the new pixel detector after its installation. Based on this strategy, the pixel detector has been successfully aligned using cosmic ray tracks recorded in early 2017. Large misalignments, especially in the forward region of the detector, have been corrected, as confirmed by the improvements observed in several track-related distributions.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit der Messung der Kopplungstärke des 125 GeV Higgs-Bosons in zwei Tau-Leptonen. Die Analyse wird mit einem Datensatz aus Proton-Proton-Kollisionen mit einer Schwerpunktsenergie von 13 TeV durchgeführt, der im Jahr 2016 vom CMS Experiment am LHC aufgezeichnet worden ist und einer integrierten Luminosität von 35.9 fb^{-1} entspricht. Es werden die Zerfallskanäle $\mu\tau_h$ and $e\tau_h$ der beiden Tau-Leptonen betrachtet, die zusammen etwa 45 % aller $\tau\tau$ Zerfälle ausmachen. Hierbei steht τ_h für ein hadronisch zerfallendes Tau-Lepton. Für jeden Endzustand werden die ausgewählten Ereignisse in exklusive Signalkategorien einsortiert, um sie einem der Higgs-Boson Produktionskanäle (Gluon-Gluon-Fusion oder Vektorboson-Fusion) zuzuordnen. Um Ereignisse des Signalprozesses von den Hauptuntergrundprozessen zu trennen, wird für jede Signalkategorie ein Paar kinematischer Variablen ausgewählt. Diese werden in zweidimensionalen Verteilungen angeordnet und mittels eines Maximum Likelihood Fit angepasst, um das Signal zu extrahieren. Für das Endergebnis dieser Analyse werden die Messungen aus den $\mu\tau_h$ und $e\tau_h$ Zerfallskanälen mit den Ergebnissen aus $\tau_h\tau_h$ und $e\mu$ Endzuständen kombiniert. Die statistische Kombination dieser vier Endzustände ergibt die Beobachtung des Zerfalls des Higgs-Bosons in ein Tau-Leptonpaar mit einer lokalen Signifikanz von 4,9 Standardabweichungen, während $4.7\,\sigma$ erwartet werden. Die Signalstärke wird zu $\mu = \sigma/\sigma_{SM} = 1.09^{+0.27}_{-0.26} = 1.09 \pm 0.15(\text{stat.})^{+0.16}_{-0.15}(\text{syst.})^{+0.10}_{-0.08}(\text{theo.})^{+0.13}_{-0.12}(\text{bbb.})$ bestimmt. Diese steht vollständig in Einklang mit der Erwartung im Rahmen des Standardmodell innerhalb der angegebenen Unsicherheiten.

Weiterhin wird im Rahmen der vorliegenden Dissertation ein Beitrag zum Alignment des Spurdetektors beschrieben, als eine Komponente der Kalibration des CMS Detektors. Am Ende des Jahres 2016 ist der CMS Pixeldetektors einem Upgrade unterworfen und ausgetauscht worden. Als erster Schritt ist daher eine Vielzahl von Studien mit simulierten Daten durchgeführt worden. Diese hatten zum Ziel, eine erste spurbasierte Strategie zum Alignment des neuen Pixeldetektors auszuarbeiten. Basierend auf dieser Strategie ist im Jahr 2017 dann das Alignment des Pixeldetektor erfolgreich durchgeführt worden. Dazu sind Spuren kosmischer Strahlung benutzt worden, die vom Pixeldetektor aufgezeichnet worden sind. Dabei konnten größere Misalignment - vor allem in der Vorwärtsrichtung des Pixelspurdetektors - korrigiert werden. Dies kann durch Verbesserungen mehrerer spurbasierter Verteilungen nachgewiesen werden.

Contents

Introduction **1**

1 The Standard Model of particle physics **5**
1.1 The elementary particles . . . 6
1.2 Quantum electrodynamics . . . 7
1.3 Quantum chromodynamics . . . 9
1.4 The electroweak theory . . . 10
1.5 The Brout-Englert-Higgs mechanism . . . 14
1.6 Higgs boson phenomenology . . . 19
1.6.1 Mass constraints . . . 19
1.6.2 Production processes . . . 22
1.6.3 Decay modes . . . 23
1.7 The Higgs boson discovery . . . 25
1.8 Open questions . . . 27

2 The CMS experiment at the LHC **31**
2.1 The Large Hadron Collider . . . 31
2.2 The CMS detector . . . 37
2.2.1 Solenoid . . . 37
2.2.2 Tracker . . . 39
2.2.3 Electromagnetic calorimeter . . . 42
2.2.4 Hadron calorimeter . . . 44
2.2.5 Muon system . . . 45
2.2.6 Trigger system . . . 47
2.3 Event simulation in CMS . . . 49

3 Event reconstruction in CMS **53**
3.1 Particle flow algorithm . . . 53
3.1.1 Fundamental elements . . . 55
3.1.2 Link algorithm . . . 57
3.1.3 Particle reconstruction and identification . . . 57
3.2 Physics objects reconstruction . . . 58
3.2.1 Primary vertices . . . 58
3.2.2 Muons . . . 59
3.2.3 Electrons and photons . . . 61
3.2.4 Charged and neutral hadrons . . . 62
3.2.5 Jets . . . 62

3.2.6 Tau leptons . . . 65
3.2.7 Missing transverse energy . . . 70

4 Alignment of the CMS tracker 73
4.1 Introduction to track-based alignment . . . 74
4.1.1 Reference systems . . . 75
4.1.2 Alignment algorithms . . . 75
4.1.3 Alignment strategies . . . 78
4.1.4 Alignment validation . . . 80
4.2 Phase 1 upgrade of the pixel detector . . . 88
4.3 Alignment of the upgraded pixel tracker . . . 89
4.3.1 Expected misalignment after the installation . . . 89
4.3.2 Alignment strategy with simulated minimum bias events . . . 91
4.3.3 Realignment studies with cosmic ray events . . . 99
4.3.4 Conclusions on the studies with simulated events . . . 100
4.3.5 Alignment with cosmic ray data . . . 101
4.4 Conclusions . . . 106

5 Measurement of the Higgs boson decay to a τ lepton pair - Analysis strategy 107
5.1 Motivation and previous measurements . . . 108
5.2 Measurement at the LHC Run 2 . . . 109
5.3 Data samples . . . 111
5.4 Simulation samples . . . 112
5.5 The $\mu\tau_h$ and $e\tau_h$ channels . . . 117
5.5.1 Event selection for the $\mu\tau_h$ channel . . . 117
5.5.2 Event selection for the $e\tau_h$ channel . . . 124
5.5.3 Higgs boson reconstruction . . . 125
5.5.4 Estimation of background processes . . . 131
5.5.5 Corrections for simulated events . . . 135
5.5.6 Definition of signal categories . . . 149
5.6 The $\tau_h\tau_h$ channel . . . 150
5.7 The $e\mu$ channel . . . 151

6 Measurement of the Higgs boson decay to a τ lepton pair - Results 153
6.1 Signal extraction . . . 153
6.2 Uncertainty model . . . 158
6.2.1 Object reconstruction and identification uncertainties . . . 158
6.2.2 Background estimation uncertainties . . . 160
6.2.3 Theory uncertainties on the signal . . . 161
6.2.4 Other uncertainties . . . 162
6.3 Results . . . 164

7 Conclusion 175

A Reference System and Kinematic Variables 179

B Statistical Methods **181**
B.1 The likelihood function . . . 181
B.1.1 Treatment of systematic uncertainties . . . 182
B.2 Maximum likelihood fit . . . 183
B.3 Hypothesis testing . . . 183
B.3.1 Definition of the test statistics . . . 184
B.3.2 Significance calculation . . . 185
B.3.3 Limits setting . . . 185

Bibliography **187**

List of Figures **198**

List of Tables **203**

Acknowledegments **205**

Introduction

The Standard Model of particle physics is a quantum field theory describing the properties and interactions of elementary particles, including the electromagnetic, weak and strong forces, based on the principle of gauge invariance. Its predictions have been extensively tested on a large range of energy scales, and the experimental results are in agreement with the expectations within the uncertainties, thus making the Standard Model very solid and widely accepted.

In the Standard Model, elementary particles acquire their mass through the mechanism of electroweak spontaneous symmetry breaking. This is achieved in the interaction with the Higgs scalar field, of which the associated particle is the Higgs boson. Despite the existence of such a boson being postulated already in 1964, it took almost fifty years of experimental searches before it was actually discovered. In 2012 the analysis of proton-proton collisions data produced at the Large Hadron Collider (LHC) led to the discovery of a new resonance with a mass of 125 GeV being compatible with the Higgs boson. The announcement was received with great excitement by the scientific community and since then the newly discovered boson is being studied with increasing precision from LHC data. Within their uncertainties, all results obtained so far are in agreement with the predictions, indicating that the discovered Higgs boson interacts as expected from the Standard Model, and can indeed explain the origin of masses of elementary particles.

Despite its great success, several experimental evidences can not be explained by the Standard Model alone, among which, for example, the nature of dark matter and the asymmetric amount of matter and antimatter in our universe. Moreover, further open questions from a more theoretical perspective remain unanswered. To mention some of them, it is not clear which mechanism protects the Higgs mass from physics at higher scales and where the Higgs potential comes from. Several extensions of the Standard Model have thus been formulated, which could provide answers to at least some of the open questions. At the moment, however, such theories beyond the Standard Model lack of experimental proof, as any prediction in terms of new particles or processes has not yet been confirmed in reality.

In order to guide the theoretical efforts, not only direct searches for new physics but also precision measurements of known processes play a major role. The careful investigation of the recently discovered Higgs sector is particularly interesting, offering us

the possibility to explore the mechanism of electroweak symmetry breaking. Assessing the properties of the Higgs boson with increasing precision will guide us answering questions about the compositeness of the Higgs boson, its connection to dark matter, the contribution to CP violation, its relation to inflation in the early universe, and many others, potentially revealing important clues of physics beyond the Standard Model.

This doctoral thesis develops in this context, and presents the study of the Higgs boson decay to τ lepton pairs. At the LHC, this is the most sensitive process to probe the direct coupling of the Higgs boson to fermions. Therefore, measuring it with increasing precision is of crucial importance to confirm that mass generation proceeds as postulated by the Brout-Englert-Higgs mechanism of spontaneous symmetry breaking. Moreover, the $H \to \tau^+\tau^-$ process constitutes a background for direct searches of additional heavier Higgs bosons predicted by different theories beyond the Standard Model. The results presented in this thesis are based on data from proton-proton collisions at a centre-of-mass energy of 13 TeV, representing the very first published study of the Higgs to τ decay at this energy scale. Data have been collected in 2016 with the Compact Muon Solenoid (CMS) detector at the LHC and correspond to an integrated luminosity of $36\,\text{fb}^{-1}$. Particles produced in proton collisions are detected in the different systems of the CMS apparatus. The latter includes a fully silicon-based inner tracker, a homogeneous electromagnetic calorimeter and a hadronic calorimeter placed inside the bore of a superconductive magnet coil, providing a solenoidal magnetic field of 3.8 T. Such a high magnetic field provides large bending power to precisely measure the momentum of charged particles. Chambers for the detection of muons are embedded in the return yoke.

The $H \to \tau^+\tau^-$ process features different final states, depending on the decay mode of each τ lepton. The latter can decay to lighter leptons ($\ell = e, \mu$) accompanied by neutrinos, as well as to hadrons (τ_h), and a τ-neutrino. The work presented in this thesis focuses on the $\mu\tau_h$ and $e\tau_h$ decay channels, representing a fraction of 45% of the possible final states. Three exclusive event categories are defined for each final state, targeting different Higgs boson production modes. The signal is extracted by means of a maximum-likelihood fit of two-dimensional kinematic distributions built for each category, with the variables being chosen to disentangle the Higgs signal from background processes. Results are given in terms of local p-value and signal strength modifier μ, representing the ratio of the measured cross section with respect to the Standard Model expectation, separately for each final state and for their combination including the $\tau_h\tau_h$ and $e\mu$ final states as well.

This thesis is structured as follows. The first chapter presents a concise review of the Standard Model and of Higgs boson phenomenology at the LHC. In Chapter 2, the experimental apparatus used for collecting data analysed in this thesis, the CMS detector, is described. The algorithms adopted by the CMS collaboration for the reconstruction and identification of particles produced in the collisions are presented in Chapter 3. Chapter 4 is dedicated to a particular detector calibration, namely tracker alignment. In fact, part of the work presented in this thesis is dedicated to the alignment of the

CMS tracker. As the CMS pixel tracker was completely replaced at the end of 2016 with a new upgraded version of it, several studies were made for assessing the impact of a possible initial large misalignment and developing a strategy to correct for it with the first recorded tracks. The performance of such an alignment evaluated using cosmic ray data recorded in 2017 closes this chapter. Chapter 5 moves into the main topic of this dissertation, and presents the strategy adopted for the $H \to \tau^+\tau^-$ analysis. The event selection and background estimation applied for the $\mu\tau_h$ and $e\tau_h$ final states are discussed in greater detail. The signal extraction method, the uncertainty model and the obtained results are covered in Chapter 6. Chapter 7 concludes this dissertation by highlighting the most relevant aspects of this work and the main results.

Chapter 1

The Standard Model of particle physics

1.1 The elementary particles . . . 6
1.2 Quantum electrodynamics . . . 7
1.3 Quantum chromodynamics . . . 9
1.4 The electroweak theory . . . 10
1.5 The Brout-Englert-Higgs mechanism . . . 14
1.6 Higgs boson phenomenology . . . 19
1.6.1 Mass constraints . . . 19
1.6.2 Production processes . . . 22
1.6.3 Decay modes . . . 23
1.7 The Higgs boson discovery . . . 25
1.8 Open questions . . . 27

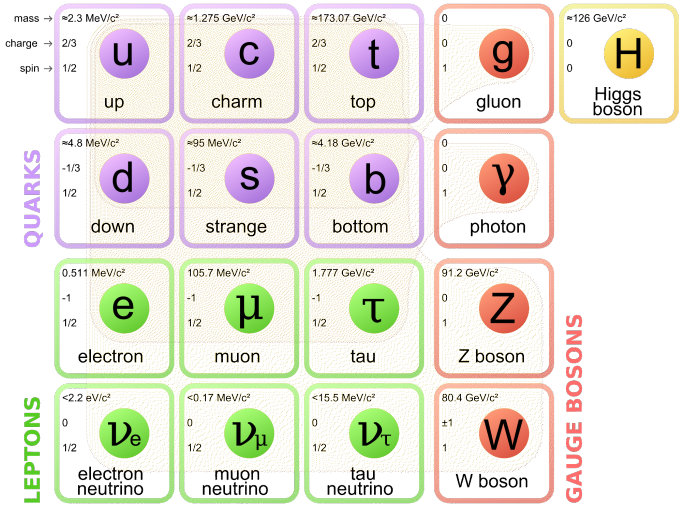
The Standard Model of particle physics is a quantum field theory based on gauge invariance principles. It describes the properties and interactions of elementary particles, including the electromagnetic, weak and strong forces, based on the $SU(3)\times SU(2)\times U(1)$ gauge group. Predictions of the Standard Model have been extensively tested on a large range of energy scales, and the experimental results are in agreement with the expectations within the uncertainties, making the Standard Model very solid and widely accepted. The first sections of this chapter introduce the elementary particles and their interactions as formulated by the Standard Model. The introduction of particles masses in the theory is achieved by spontaneous symmetry breaking through the Brout-Englert-Higgs mechanism, which is widely discussed in Sec. 1.5. The latter mechanism implies the existence of a scalar particle, the Higgs boson. The production and decay processes of this particle, as well as the theoretical constraints on its mass, are described in Sec. 1.6. Finally, the discovery of the Higgs boson in proton-proton collisions in 2012 and first measurements of its main properties are summarised in Sec. 1.7. The last section Sec. 1.8

completes the theoretical overview by mentioning the main questions which do not find an answer within the Standard Model.

1.1 The elementary particles

In the Standard Model, the elementary constituents of matter are fermions and their interactions are mediated by vector bosons. Fermions and vector bosons acquire their mass in the interaction with the scalar Higgs field, for which the Higgs boson is the associated particle. An overview of the Standard Model particle spectrum, including particles mass, charge and spin is shown in Fig. 1.1. As illustrated in Fig. 1.1, fermions are classified into two groups, namely quarks and leptons, depending on whether they do interact strongly or not. The strong interaction is responsible for quark confinement, so that quarks are not observed as free particles but only in bound states called *hadrons*.

Figure 1.1 Particle content of the Standard Model: quarks, leptons and bosons with their mass, charge and spin values. Picture from Wikipedia.



Leptons are organized into three isospin doublets, also called *families*, each including a massive lepton, namely the electron e, the muon μ and the τ lepton, and the corresponding neutrino:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}. \tag{1.1}$$

Neutrinos have very small masses and in the classical formulation of the Standard Model they are indeed considered massless. The heaviest lepton is the τ lepton, with a mass of 1.777 GeV [1]. The lower components of each lepton doublet carry electromagnetic charge $Q = -e$ while neutrinos are neutral particles and hence do not interact electromagnetically.

Similarly to the leptons, there are six quark *flavors*, namely the up (u), down (d), charm (c), strange (s), bottom (b) and top (t) quarks, combined into three families, as follows:

$$\left(\begin{array}{c} u \\ d \end{array} \right) \left(\begin{array}{c} c \\ s \end{array} \right) \left(\begin{array}{c} t \\ b \end{array} \right). \tag{1.2}$$

Quark masses span a large range, from a few MeV for the u and d quarks to 173 GeV for the top quark [1]. Quarks have fractional electromagnetic charge, $+2/3e$ and $-1/3e$ for the quarks in the upper and lower rows, respectively. In addition to the electromagnetic charge, quarks also carry *color* charge, as they take part in the strong interaction, and each quark flavor is present in three different color states.

Beside fermions, there are twelve vector bosons in the Standard Model, which are interpreted as force mediators: the photon γ for the electromagnetic force, the $W^{\pm}$ and Z bosons for the weak force, and eight gluons g for the strong force. While the photon and the gluons are massless, the $W^{\pm}$ and Z bosons have masses of 80 and 91 GeV [1], respectively, which they acquire through spontaneous symmetry breaking. The role of the boson fields and their number is understood in the formulation of the Standard Model as a quantum field theory based on gauge symmetries, which is presented in Secs. 1.3 and 1.4. As an introduction to that, the concept of gauge invariance and its application to build a quantum field theory is briefly discussed in the next section in the context of quantum electrodynamics.

1.2 Quantum electrodynamics

The importance of gauge invariance was first understood in the context of electromagnetism, and the first quantum field theory formulated based on gauge invariance was quantum electrodynamics (QED) [2], describing the electromagnetic interaction in the quantum regime.

Gauge invariance is a feature of the classical theory of electromagnetism [3]. In fact, Maxwell's equations are invariant under the transformation of the vector potential A^{μ}:

$$A^{\mu} \rightarrow A'^{\mu} = A^{\mu} - \partial^{\mu}\theta, \tag{1.3}$$

where $\partial^{\mu} = (\partial/\partial t, -\vec{\nabla})$, $\theta = \theta(x)$ is an arbitrary scalar function, and the vector potential $A^{\mu} = (V, \vec{A})$ is related to the magnetic and electric fields $\vec{B}$ and $\vec{E}$ as follows:

$$\begin{aligned} \vec{B} &= \vec{\nabla} \times \vec{A} \\ \vec{E} &= -\vec{\nabla} V - \frac{\partial \vec{A}}{\partial t}. \end{aligned} \tag{1.4}$$

Explicitly, this means the electric and magnetic fields $\vec{E}$ and $\vec{B}$ are preserved under the transformation in Eqn. 1.3. In quantum mechanics, however, it can be shown that the wavefunction ψ solving Schrödinger's equation for a particle in an electromagnetic

field is not invariant under the gauge transformation of Eqn. 1.3. In order to ensure the consistency of quantum mechanics with Maxwell's equation, the change in the potential A^μ must be accompanied by a change in the wavefunction, as follows:

$$\begin{aligned} A^\mu &\rightarrow A'^\mu = A^\mu - \partial^\mu \theta \\ \psi &\rightarrow \psi' = \psi e^{i\theta}. \end{aligned} \tag{1.5}$$

In fact it can be shown that the wavefunctions ψ and ψ' do describe the same physics [3]. As a consequence of Noether's theorem [4], the invariance under a local transformation implies the existence of a conserved quantity. In the case of the electromagnetic interaction, this leads to the conservation of the electric charge.

After explaining why gauge invariance should hold in QED, it is crucial to observe that the theory itself can be constructed starting from a gauge invariance principle. More precisely, QED can be fully derived starting from the Lagrangian density[1] for a free electron, and demanding its invariance under local $U(1)$ transformations, which is experimentally justified by the conservation of the electric charge [3]. The Lagrangian for a free massive Dirac field ψ of mass m is given by

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{1.6}$$

where ψ is a Dirac spinor, γ^μ are the Pauli matrices and $\overline{\psi} = \psi^+ \gamma^0$. Such a Lagrangian is not invariant under local $U(1)$ transformations of the spinor field

$$\psi \rightarrow \psi' = \psi e^{i\theta}. \tag{1.7}$$

The invariance can be imposed by substituting the partial derivative ∂^μ with the covariant derivative D^μ, defined by

$$D^\mu \equiv \partial^\mu + ieA^\mu, \tag{1.8}$$

which introduces an interaction of the electron with the vector field A^μ. Physically, the gauge field A^μ can be identified with a particle mediating the interaction between fermions with a coupling constant e, the photon. The transformation rules of the covariant derivative and of the vector field follow from the imposed gauge invariance for the Lagrangian.

The free fermion Lagrangian is completed by a term describing the dynamics of the gauge field:

$$\mathcal{L}_{kin} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad \text{with} \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{1.9}$$

which is Lorentz and gauge invariant. In the absence of matter, i.e. of fermion fields, the Lagrangian is reduced to the kinetic term $\mathcal{L}_{kin}$ of the gauge field, and it can be shown that its canonical variation yields the Maxwell's equations in vacuum [3].

[1] In the following, a Lagrangian density will be called Lagrangian for shortness reasons.

The formulation of QED starting from gauge invariance principles provides the needed guidance for the development of the Standard Model as a gauge theory. In the next section, the formulation of quantum chromodynamics is summarised.

1.3 Quantum chromodynamics

Quantum chromodynamics (QCD) [5] is the theory describing the strong interaction, which binds quarks together to form hadrons. From experimental evidences about the composition of hadrons, it was inferred that quarks carry an additional quantum number with respect to leptons, the color, and that the strong interaction can be described by a non-abelian $SU(3)$ color symmetry group [6]. Following a similar approach as for QED, the QCD Lagrangian can be derived [6] starting from the free Lagrangian for the quark spinor fields $q = q(x)$, and requiring it to be invariant under local $SU(3)$ transformations of the quark fields:

$$q \rightarrow q' = \exp[i\alpha_a T_a]\, q,\ a = 1, ..., 8 \tag{1.10}$$

where $T_a = \lambda_a/2$ are the eight generators of the $SU(3)$ group, expressed in terms of the Gell-Mann matrices λ_a. They obey the following commutation rules:

$$[T_a, T_b] = i f_{abc} T_c, \tag{1.11}$$

where f_{abc} are the structure constants of the group.

In order to obtain the invariance of the free quark Lagrangian under local $SU(3)$ gauge transformations, the covariant derivative must be defined as:

$$D_\mu \equiv \partial_\mu + i g_s T_a G^a_\mu, \tag{1.12}$$

where eight gauge vector fields G^a_μ have been introduced, transforming according to:

$$G^a_\mu \rightarrow G'^a_\mu = G^a_\mu - \frac{1}{g_s}\partial_\mu \alpha_a - f_{abc}\alpha^b G^c_\mu. \tag{1.13}$$

The full gauge invariant QCD Lagrangian [1] is given by:

$$\mathcal{L}_{QCD} = \sum_{f=1}^{N_f} \bar{q}_{f,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu T^c_{ab} G^c_\mu - m_q \delta_{ab}) q_{f,b} - \sum_{a=1}^{8} \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}, \tag{1.14}$$

where N_f is the number of quark flavors, and the gauge invariant strength field tensor $G^a_{\mu\nu}$ is defined as:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu. \tag{1.15}$$

Similarly to the case of QED, the vector fields G^a_μ represent eight gauge bosons mediating the strong interaction, the *gluons*. The constant g_s that appears in the expression of the covariant derivative in Eqn. 1.12 is the coupling constant of the strong interaction.

The complexity of the QCD Lagrangian describes the rich phenomenology of strong interactions. In particular, the strength field tensor of Eqn. 1.15 contains a term for the interaction of gluon fields. Conversely to the QED case, where the photon is neutral, gluons carry color charge themselves and interact with each other. Interaction vertices with three or four gluons are possible in QCD, while they have no analogue in QED. The self-interacting nature of QCD is a consequence of the choice of a non-abelian symmetry group [7]. Another remarkable consequence is asymptotic freedom, meaning that the effective strength of the interaction decreases with increasing transferred momentum Q^2 among the participants. More precisely, the strong interaction constant α_s can be written as

$$\alpha_s = \frac{g_s^2}{4\pi} \propto \frac{1}{\log(Q^2/\Lambda^2)}, \tag{1.16}$$

where Λ is an energy scale of the order of 200 MeV [5]. This "running" behaviour of α_s has the effect that partons are almost free at high energy scales, i.e. *asymptotically* free. Conversely, at low energies the coupling becomes so strong to confine quarks into color-neutral hadrons, explaining the experimental evidence that free color-charged particles are not observed.

Besides quantum chromodynamics, the Standard Model treats the electromagnetic and the weak interaction in a unified fashion. The basic principles and features of the electroweak theory are described in the next section.

1.4 The electroweak theory

The weak interaction retains its name because its strength is several orders of magnitude smaller than the strong and electromagnetic forces. The first evidences of the weak force date back to the study of nuclear β-decay. To explain the observed continuous energy spectrum in such decays, Pauli postulated the existence of neutrinos in 1930 [8], and the first model of the weak interaction was formulated by Fermi in 1933 in terms of a contact interaction involving four fermions [9]. Since then, the theoretical description of the weak interaction developed over many years. The current formulation is due to Glashow [10], Weinberg [11] and Salam [12], who proposed a gauge theory unifying the weak and the electromagnetic interactions, called *electroweak* theory.

The electroweak theory is based on a local $SU(2)\times U(1)$ gauge symmetry, the generators of the $SU(2)$ group are the weak isospin operators:

$$\vec{T} = \frac{1}{2}\vec{\sigma}, \tag{1.17}$$

with $\vec{\sigma}$ being the Pauli matrices, and obey the commutation relation:

$$[T^a, T^b] = i\epsilon_{abc}T^c. \tag{1.18}$$

The $U(1)$ symmetry is generated by the weak hypercharge operator, in a similar fashion as in the case of QED. The weak hypercharge Y can be related to the electrical charge Q and to the third component of the isospin I_3 by the Gell-Mann-Nishijima formula [13, 14]:

$$\mathrm{Y} = 2(Q - I_3). \tag{1.19}$$

The electroweak theory is a chiral theory, distinguishing between left and right handed spinor components. This is suggested by the experimental experience that charged weak interactions only couple to left-handed states. The experimental evidence that parity is violated in weak interactions [15] is reflected in the fact that the left and right handed components of the fermion spinors transform differently under gauge transformations. More precisely, the left-handed components of isospin doublets for leptons and quarks can be indicated, respectively, with:

$$\begin{aligned} \chi_L &= \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \\ \chi_L &= \begin{pmatrix} q_{up} \\ q_{down} \end{pmatrix}_L , \end{aligned} \tag{1.20}$$

while the right-handed components form isospin singlets:

$$\begin{aligned} \psi_R &= \ell_R \\ \psi_R &= \left(q_{up}\right)_R \, , \, \left(q_{down}\right)_R . \end{aligned} \tag{1.21}$$

Using the introduced notation, a gauge transformation of the $SU(2) \times U(1)$ group acts as follows:

$$\begin{aligned} \chi_L \;\rightarrow\; \chi'_L &= \exp(i\vec{\alpha}\cdot\vec{T} + i\beta\mathrm{Y})\chi_L \\ \psi_R \;\rightarrow\; \psi'_R &= \exp(i\beta\mathrm{Y})\psi_R. \end{aligned} \tag{1.22}$$

From the transformation rules, it is evident how the $SU(2)$ symmetry is imposed only for the left-handed components. For this reason, the electroweak symmetry group is often indicated with $SU(2)_L \times U(1)_Y$.

Following a similar approach as for QED and QCD, the electroweak Lagrangian can be derived starting from the free Lagrangian for the fermion fields and requiring it to be invariant under the transformations of Eqn. 1.22. For simplicity, the free Lagrangian for one lepton isospin doublet is considered, neglecting interaction and mass terms and assuming that neutrinos are fully left-handed:

$$\mathcal{L}_0 = \overline{\chi}_L\gamma^\mu\partial_\mu\chi_L + \overline{\psi}_R\gamma^\mu\partial_\mu\psi_R, \tag{1.23}$$

where mixed chirality terms do not contribute, as left and right handed components are orthogonal. The subscript "0" indicates that, at this stage, particles are assumed to be massless.

The gauge invariance under $SU(2)_L \times U(1)$ transformations can be imposed on the Lagrangian of Eqn. 1.23 by substituting the partial derivative with the following covariant derivative [7]:

$$D_\mu = \partial_\mu + ig\vec{T} \cdot \vec{W}_\mu(x) + ig' \frac{\mathrm{Y}}{2} B_\mu(x) \tag{1.24}$$

where a total of $3 + 1 = 4$ vector fields, indicated with $\vec{W}_\mu = (W^1_\mu, W^2_\mu, W^3_\mu)$ and B_μ, represent the gauge fields. The electroweak Lagrangian including kinetic terms for the gauge bosons can be written as:

$$\begin{aligned} \mathcal{L}_{EWK,0} &= \overline{\chi}_L \gamma^\mu \left[i\partial_\mu - g\vec{T} \cdot \vec{W}_\mu - g'(\tfrac{\mathrm{Y}_L}{2}) B_\mu \right] \chi_L \\ &+ \overline{\psi}_R \gamma^\mu \left[i\partial_\mu - g'(\tfrac{\mathrm{Y}_R}{2}) B_\mu \right] \psi_R \\ &- \tfrac{1}{4} \vec{W}^{\mu\nu} \cdot \vec{W}_{\mu\nu} - \tfrac{1}{4} B^{\mu\nu} B_{\mu\nu}, \end{aligned} \tag{1.25}$$

in which the field strength tensor $B_{\mu\nu}$ has a similar form as for the QED case (see Eqn. 1.9), while for the $SU(2)$ group:

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g\epsilon^{abc} W^b_\mu W^c_\nu, \tag{1.26}$$

where ϵ^{abc} is the totally antisymmetric Levi-Civita tensor. The hypercharge eigenvalues $\mathrm{Y}_L = -1$ and $\mathrm{Y}_R = -2$ in Eqn. 1.25 for the left and right-handed components have been derived from Eqn. 1.19.

The interpretation of the vector fields as the propagators of bosons mediating the electromagnetic and weak interactions is not as evident as in QED or QCD because of the "composite" nature of the symmetry, which is built from the $SU(2)_L$ and $U(1)_Y$ groups, the $U(1)_Y$ symmetry being imposed on the isospin doublet as a whole. However, the careful inspection of all the terms of the Lagrangian in Eqn. 1.25 and their interpretation in terms of charged and neutral currents allows to identify the vector fields corresponding to the mediators of the interaction, namely the $W^\pm$ bosons responsible for the charged weak current, the neutral Z boson for the weak neutral current, and the photon γ for the electromagnetic interaction. In the physical basis, the boson fields are given by [7]:

$$\begin{aligned} W^+_\mu &= \tfrac{1}{\sqrt{2}} \left(W^1_\mu - iW^2_\mu \right) \\ W^-_\mu &= \tfrac{1}{\sqrt{2}} \left(W^1_\mu + iW^2_\mu \right) \\ Z_\mu &= \tfrac{1}{\sqrt{g^2+g'^2}} \left(gW^3_\mu - g'B_\mu \right) \\ A_\mu &= \tfrac{1}{\sqrt{g^2+g'^2}} \left(g'W^3_\mu + gB_\mu \right). \end{aligned} \tag{1.27}$$

It can be observed that the Z and the photon field are obtained by rotating the gauge fields of an angle θ_W, called *Weinberg* angle or weak-mixing angle:

$$\begin{aligned} Z_\mu &= \cos\theta_W W^3_\mu - \sin\theta_W B_\mu \\ A_\mu &= \sin\theta_W W^3_\mu + \cos\theta_W B_\mu . \end{aligned} \tag{1.28}$$

Comparing the expressions for the fields Z_μ and A_μ in Eqns. 1.27 and 1.28, the Weinberg angle is related to the weak couplings g and g' as follows:

$$\sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} \quad , \quad \cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}. \tag{1.29}$$

The Weinberg angle is not predicted within the Standard Model, but is determined experimentally and measures $\sin^2\theta_W(Q^2 = m_Z) = 0.23122 \pm 0.00004$ [1].

Finally, the experimental observation of transitions between the upper and lower components of a quark flavor doublet across different families, e.g. from a u to a s quark, as well as the CP violation in weak interactions, are explained by the introduction of *quark mixing* [16]. This means that quark flavor eigenstates do not coincide with the weak isospin eigenstates, but the lower components of the doublets are rotated with respect to each other. More explicitly, the lower components of the weak isospin eigenstates d', s' and b' are related to the lower components of the flavor eigenstates d, s and b as follows:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.30}$$

The rotation matrix describing the transition probabilities between flavors is named after Cabibbo, Kobayashi and Maskawa, and is assumed to be unitary. Beside that, the single matrix elements are free parameters of the theory and are determined experimentally [1]. As a consequence of quark mixing, flavor-changing neutral currents are not allowed at tree level [17], which explains the fact they have not yet been observed.

Despite its elegance, the electroweak theory as presented up to here is affected by a big limitation: apparently, it can not describe massive particles. In fact, mass terms of the form $M^2 W^\mu W_\mu$ for vector bosons are not gauge invariant, and would introduce unrenormalizable divergences, making the theory unable to make predictions. Similarly, mass terms of the form $-m^2\overline{\psi}\psi$ for fermions would contain mixed chirality terms that do not comply to the gauge transformation rules, and can not be simply added to the Lagrangian because they are manifestly violating gauge invariance, the cornerstone of the theory itself. The mechanism for introducing massive particles in the theory without loosing gauge invariance is described in the next section.

1.5 The Brout-Englert-Higgs mechanism

The Standard Model is built on a $SU(3) \times SU(2) \times U(1)$ non-abelian gauge symmetry, corresponding to the symmetry groups of the strong and electroweak interactions discussed in the previous sections. The Brout-Englert-Higgs mechanism, shortly called Higgs mechanism, provides a way to generate mass terms in the Standard Model Lagrangian, preserving gauge invariance. The core idea of the Higgs mechanism is spontaneous symmetry breaking, a concept originally elaborated in condensed matter physics to describe, for example, superfluidity and superconductivity. The introduction of this concept for global symmetries is due to Nambu [18, 19], Goldstone [20] and Anderson [21]. Subsequently in 1964 Brout and Englert [22] and Higgs [23, 24] provided a formulation of spontaneous symmetry breaking for local symmetries, elaborated further by Guralnik, Hagen and Kibble [25].

The Goldstone model

The way to generate the mass of a particle by spontaneous symmetry breaking can be illustrated in a simple theory containing one complex scalar field ϕ with a quartic potential $V(\phi^*\phi)$ of the kind:

$$V(\phi^*\phi) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2. \tag{1.31}$$

The corresponding Lagrangian

$$\mathcal{L}_{Goldstone} = (\partial_\mu\phi)^*(\partial^\mu\phi) - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2, \tag{1.32}$$

with μ^2 and $\lambda > 0$ real constants, is manifestly invariant under global $U(1)$ transformations of the scalar field ϕ. The shape of the potential in Eqn. 1.32 depends on the sign of μ^2:

- if $\mu^2 > 0$, the potential has a "U" shape with a single minimum for $\phi = 0$;
- if $\mu^2 < 0$, the potential has a "mexican hat" shape, as illustrated in Fig. 1.2.

The interesting case to consider is $\mu^2 < 0$: in this situation, the vacuum is degenerate on the circumference with radius $v = \sqrt{-\mu^2/\lambda}$. The realisation of one particular state spontaneously breaks the rotational symmetry of the problem.

The theory corresponding to the Lagrangian of Eqn. 1.32, with $\mu^2 < 0$ is called *Goldstone* model. Its particle content is studied by expanding the Lagrangian around the minimum. Among all ground states lying on the circle of radius v, the state $(\phi_1, \phi_2) = (v, 0)$ is chosen, realizing spontaneous symmetry breaking. Small perturbations around the latter vacuum state can be parametrized as $(\phi_1, \phi_2) = (v + \eta, \xi)$, where η and ξ are

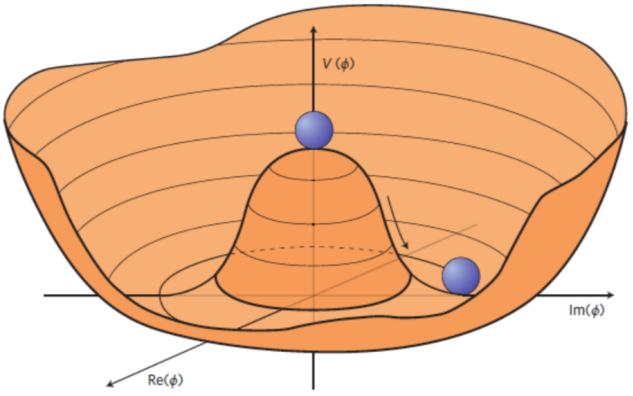
real fields. After performing the expansion around the chosen vacuum state up to the second order in the fields η and ξ, the Lagrangian of Eqn. 1.32 becomes:

$$\mathcal{L}(\eta, \xi) = \frac{1}{2}(\partial^\mu \xi)(\partial_\mu \xi) + \frac{1}{2}(\partial^\mu \eta)(\partial_\mu \eta) + \mu^2 \eta^2 + \text{const.} \; . \tag{1.33}$$

The presence of the term $\mu^2\eta^2$ shows that the η field has a mass $m_\eta^2 = -2\mu^2$, while the ξ field is massless. Visually, this can be understood from the shape of the potential drawn in Fig. 1.2, which is flat in the tangent ξ direction.

To summarise, the model contains one massive and one massless boson. This simple example shows how massive bosons can be accomodated in a gauge invariant theory. The presence of the massless field is a realisation of the Goldstone theorem [26], stating that massless scalars appear whenever a continuous symmetry of a physical system is spontaneously broken, i.e. the symmetry of the Lagrangian is not apparent in the ground state.

Figure 1.2 The "mexican hat" shape of the quartic potential allowing for spontaneous symmetry breaking. The minimum of the potential is described by the set of points along a circumference. This rotational symmetry of the ground state is spontaneously broken by the realisation of one specific vacuum state [27].



After discussing the mechanism of spontaneous symmetry breaking in the Goldstone model, the focus is moved to the specific case of the Standard Model, where the symmetry breaking needs to be introduced for $SU(2) \times U(1)$ gauge transformations.

Breaking a local $U(1)$ symmetry

The characteristics of a local $U(1)$ symmetry in terms of gauge transformations and covariant derivative have been discussed above in the context of QED in Sec. 1.2. The gauge invariant Lagrangian for a scalar field ϕ with a quartic potential term $V(\phi^*\phi)$ as

in Eqn. 1.31 has the form:

$$\mathcal{L} = D_\mu \phi^* D^\mu \phi - V(\phi^* \phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{1.34}$$

The symmetry of the Lagrangian is spontaneously broken by the choice of one particular ground state. Repeating the procedure followed in the previous section, the expansion of the Lagrangian around the vacuum state is performed by substituting $\phi(x) = \sqrt{\frac{1}{2}}\left(v + \eta(x) + i\xi(x)\right)$ to obtain [27]:

$$\begin{aligned} \mathcal{L}(\eta, \xi) \quad &= \quad \tfrac{1}{2}(\partial^\mu \xi)(\partial_\mu \xi) + \tfrac{1}{2}(\partial^\mu \eta)(\partial_\mu \eta) - v^2 \lambda \eta^2 \\ &\quad + \tfrac{1}{2} e^2 v^2 A^\mu A_\mu - e v A_\mu \partial^\mu \xi - \tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} + \text{ interaction terms.} \end{aligned} \tag{1.35}$$

In the expression above, a mass term for the gauge boson field A_μ, with value $m_A = ev$, has been generated as a consequence of the symmetry breaking, along with the massless scalar ξ and the massive scalar η, with $m_\eta^2 = 2\lambda v^2$. For a correct interpretation of the particle content, however, the Lagrangian in Eqn. 1.35 needs to be written in a base where the bilinear term $ev A_\mu \partial^\mu \xi$ cancels out. This can be achieved by fixing the gauge for A_μ performing the transformation [7]:

$$\begin{aligned} A_\mu \quad &\rightarrow \quad A'_\mu \quad = \quad A_\mu + \tfrac{1}{ev} \partial_\mu \xi \\ \phi \quad &\rightarrow \quad \phi' \quad = \quad \exp(-i\xi/v)\phi. \end{aligned} \tag{1.36}$$

With this choice, the Lagrangian of Eqn. 1.35 takes the form:

$$\mathcal{L}(\eta) = \frac{1}{2}\partial_\mu \eta \partial^\mu \eta - \frac{1}{2}\mu^2 \eta^2 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} e^2 v^2 A'^\mu A'_\mu + \text{tri- and quadri-linear interactions.} \tag{1.37}$$

The massless field ξ introduced by the choice of the ground state used to break the symmetry does not appear in the Lagrangian anymore. It has been absorbed by the gauge field A_μ, which acquired a mass, and hence an additional (longitudinal) polarization state. This is the Higgs mechanism. The massive scalar η is a distinguishing feature of spontaneous symmetry breaking, and is called *Higgs boson*. It is massive because the potential $V(\phi, \phi^*)$ is not flat in the radial direction.

After describing the Higgs mechanism for a theory with a local $U(1)$ gauge symmetry, the case of the Standard Model is discussed in the next paragraph.

Breaking the electroweak symmetry

In the Standard Model, the Higgs mechanism allows the $W^\pm$ and Z gauge bosons to acquire their mass without violating gauge invariance. Similarly to the case presented above, the procedure is to start from the gauge invariant electroweak Lagrangian $\mathcal{L}_{EWK,0}$

of Eqn. 1.25, and add a scalar field ϕ with a mexican-hat quartic potential $V(\phi, \phi^\dagger)$:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{EWK,0} + (D^\mu\phi)(D_\mu\phi)^\dagger - V(\phi,\phi^\dagger) \\ V(\phi,\phi^\dagger) &= \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \text{ with } \mu^2 < 0, \lambda > 0. \end{aligned} \tag{1.38}$$

In the Weinberg-Salam model [11, 12], the scalar field ϕ is chosen to be a complex doublet of $SU(2)$ with hypercharge $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.39}$$

The ground state ϕ_0 is chosen to be:

$$\phi_0 \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.40}$$

with isospin $I = 1/2$ and its third component $I_3 = -1/2$. From the Gell-Mann-Nishijima relation given in Eqn. 1.19, it follows that ϕ_0 is neutral ($Q = 0$), and therefore always invariant under local phase transformations of the form $\phi \to \phi' = \exp(i\theta q)\phi$, independently of the phase θ. As a consequence, the electromagnetic $U(1)$ symmetry is not broken and the photon remains massless [7]. The perturbative expansion of the Lagrangian of Eqn. 1.38 around the chosen ground state ϕ_0 is performed by representing the scalar field as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \eta \end{pmatrix}, \tag{1.41}$$

where the gauge has already been fixed to absorb the Goldstone bosons. Mass terms of the form $\frac{1}{2}M^2 W_\mu W^\mu$ appear for the gauge fields $W^1_\mu, W^2_\mu, W^3_\mu$, and for the scalar η field, i.e. the Higgs field. Remembering the relations between the electroweak gauge fields and the fields of the weak interaction bosons given in Eqn. 1.27, the masses of the $W^\pm$ and Z bosons can be derived:

$$m_W = \frac{1}{2}gv \ , \ m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}. \tag{1.42}$$

The vacuum expectation value v of the Higgs field $v = 246.22\,\text{GeV}$ can be calculated from the Fermi constant G_F [28], measured from the muon lifetime, as follows:

$$v^2 = \frac{1}{\sqrt{2}G_F}, \quad \text{where } G_F = \frac{\sqrt{2}}{8}\frac{g^2}{m_W^2} \text{ and } m_W = \frac{1}{2}gv. \tag{1.43}$$

The mass of the Higgs boson is given by

$$m_H^2 = -2\mu^2 \tag{1.44}$$

and is a free parameter in the Standard Model. Constraints on the Higgs boson mass coming from theoretical arguments are discussed in Sec. 1.6.1.

Fermion masses

As explained above, the Higgs mechanism allows to incorporate interacting massive vector bosons in the Standard Model preserving gauge invariance. In addition to that, the same mechanism is also able to explain the origin of masses for the fermions. Similarly to the case of vector bosons, also mass terms for the leptons can not directly be added to the Lagrangian. In fact, they would mix left and right handed components, and therefore manifestly violate $SU(2)$ gauge invariance. The way to generate mass terms for the lower component of the lepton doublets, i.e. for the electron, muon and τ lepton, through the Higgs mechanism proceeds as follows [7]. First, a gauge invariant term representing the Yukawa interaction of the Higgs field ϕ with the fermion fields ψ needs to be added to the Lagrangian [7]:

$$\mathcal{L}_{Yukawa} = -G_f[\overline{\psi}_R\phi^\dagger\psi_L + \overline{\psi}_L\phi\psi_R]. \tag{1.45}$$

Taking the electron as an example, the Yukawa term can be explicitly written as:

$$\mathcal{L}_{Yukawa,e} = -G_e\left[\bar{e}_R\phi^\dagger\begin{pmatrix}\nu_L\\ e_L\end{pmatrix} + (\bar{\nu}_L\bar{e}_L)\phi e_R\right]. \tag{1.46}$$

After performing the expansion of the Lagrangian around the vacuum of the Higgs field ϕ_0, i.e. after breaking the symmetry, the Yukawa term for the electron becomes:

$$\mathcal{L}_{Yukawa}(\eta) = -G_e\frac{v+\eta}{\sqrt{2}}(\bar{e}_R e_L + \bar{e}_L e_R). \tag{1.47}$$

Defining $\bar{e} \equiv (\bar{e}_R, \bar{e}_L)$ and $e = (e_L, e_R)^T$ and substituting in the expression above leads to:

$$\mathcal{L}_{Yukawa}(\eta) = -\frac{G_e v}{\sqrt{2}}\bar{e}e - \frac{G_e}{\sqrt{2}}\eta\bar{e}e, \tag{1.48}$$

from which the electron mass m_e can be read off to be:

$$m_e = \frac{G_e}{\sqrt{2}}v. \tag{1.49}$$

The second term of Eqn. 1.48 represents the interaction between the Higgs field and the electron, that is proportional to the electron mass. The same procedure can be repeated to give mass to fermions occupying the lower component of the isospin doublets, namely the muon, the τ lepton and the c, s and b quarks. The upper components of the doublets do not contribute to the Yukawa term of Eqn. 1.45, because the chosen ground state ϕ_0 of the Higgs potential has null upper component (see Eqn. 1.40).

In order to account for the masses of the up-type quarks, an additional Higgs doublet field $\tilde{\phi}$ has to be built from the first doublet ϕ [7]. In the Standard Model, this is chosen

as the charge conjugate of ϕ:

$$\tilde{\phi} = \phi_c = -i\tau_2\phi^* = \begin{pmatrix} -\overline{\phi^0} \\ \phi^- \end{pmatrix} \tag{1.50}$$

having hypercharge $Y = -1$ and same gauge transformation rules as ϕ. The Lagrangian is completed by the following Yukawa term:

$$\mathcal{L}^{up}_{Yukawa} = -G_u(\bar{u}, \bar{d})_L \phi_c u_R, \tag{1.51}$$

and the perturbative expansion around the ground state of ϕ_c is performed using:

$$\phi_c(\eta) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + \eta \\ 0 \end{pmatrix}. \tag{1.52}$$

Reminding that the quark flavor eigenstates (u, d) are rotated with respect to the weak interaction eigenstates (u, d'), the correct expression for the Yukawa Lagrangian of the quark sector is given by:

$$\mathcal{L}^{quark}_{Yukawa} = -G_d^{ij}(\bar{u}\bar{d'})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_{jR} - G_u^{ij}(\bar{u}\bar{d'})_L \begin{pmatrix} -\overline{\phi^0} \\ \phi^- \end{pmatrix} u_{jR} + \text{h.c.}, \tag{1.53}$$

where u and d stand for the upper and lower components of the quark doublets, and $i, j = 1, ..., N$ run over the three quark families.

1.6 Higgs boson phenomenology

1.6.1 Mass constraints

The Higgs boson mass is a free parameter of the Standard Model, and needs to be determined experimentally. Nevertheless, there are arguments coming from the theory itself which can be invoked to guide the experimental search. In particular, the Higgs boson mass is mainly bounded by three conditions: the unitarity, triviality and stability bounds, which are briefly discussed in the following paragraphs.

The unitarity bound arises from considerations about the scattering amplitudes of the longitudinal components of massive gauge bosons, which increase with their momenta. Hence, the cross section of processes like W^+W^- elastic scattering grows with the energy, and can eventually violate unitarity. Such processes involve the exchange of a Higgs boson, and an upper limit on the Higgs mass can be obtained by developing the scattering amplitude in partial waves and applying the optical theorem for the s-wave partial amplitude [29]. At high energies, in the limit $s \gg m_H^2$, the unitarity condition for the W^+W^- scattering brings to the upper bound $m_H \lesssim 870\,\text{GeV}$ [30]. This limit can be tightened by including other channels like ZZ, ZH, HH, WH and WZ and becomes $m_H \lesssim 710\,\text{GeV}$ [31, 32].

The triviality bound originates from the fact that ϕ^4 theories can not be perturbative at all scales unless they are trivial, i.e. the couplings are zero [33]. The quartic Higgs coupling is a function of the energy scale, $\lambda = \lambda(Q^2)$. At one-loop level, considering only contributions from the Higgs boson itself, the dependence of λ from the energy scale is given by the following relation [29]:

$$\lambda(Q^2) = \lambda(v^2)\left(1 - \frac{3}{4\pi^2}\lambda(v^2)\log\frac{Q^2}{v^2}\right)^{-1}, \tag{1.54}$$

where the reference energy point has been chosen to be the scale of electroweak symmetry breaking. From the latter expression, it is clear that the coupling λ grows with the energy scale and becomes infinite at the cut-off energy

$$\Lambda_C = v\exp\left(\frac{4\pi^2 v^2}{m_H^2}\right). \tag{1.55}$$

Therefore, the energy range in which the Higgs self-coupling is finite, hence the Standard Model is valid, depends on the Higgs mass. Higher values of the cut-off energy require lighter Higgs bosons. For example, for $\Lambda_C \sim 10^{16}\,\text{GeV}$, the Higgs boson mass $m_H \lesssim 200\,\text{GeV}$ [29].

The last constraint on the Higgs mass comes from the stability of the Higgs potential. When including other particles in the one-loop calculation of the quartic coupling $\lambda(Q)$, the top quark brings a negative contribution, which, if $\lambda(v)$ is too small, could make the coupling itself negative. In this case, the vacuum would have no minimum and be instable. The stability argument of the scalar potential provides a lower bound for the Higgs boson mass, which again depends on the cut-off energy. For $\Lambda_C \sim 10^{16}\,\text{GeV}$, $m_H \gtrsim 130\,\text{GeV}$ [29]. The plot in Fig. 1.3 shows the Higgs boson mass as a function of the cut-off energy scale, and the upper and lower bounds coming from the triviality and stability bounds, based on two-loop calculations of the Higgs quartic coupling. The allowed region for the Higgs boson mass lies between the two bands [34]. The stability bound can be relaxed if the vacuum is allowed to be metastable: in this case, the scalar potential has another minimum lower than the standard electroweak one, but the transition time to the lower state is longer than the universe lifetime. The stability of the Higgs potential as a function of the Higgs boson and of the top quark masses is shown in Fig. 1.4.

The latest searches for the Higgs boson, including the one presented in this thesis, are being conducted in proton-proton collisions at the Large Hadron Collider (LHC). In 2012, the ATLAS and CMS collaborations announced the discovery of a new resonance compatible with the Higgs boson hypothesis. The mass of the newly discovered particle is measured to be $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\,\text{GeV}$ [36]. Therefore, the review of Higgs boson production and decay modes in the following refers to the measured mass value. More details on the Higgs boson discovery are given in Sec. 1.7.

Figure 1.3 Triviality and stability bounds on the Higgs boson mass M_H as a function of the cut-off energy Λ_C. The values of the top mass and of the strong coupling constant α_S used are indicated in the plot. The thickness of the bands represents the effect of the uncertainties [34].

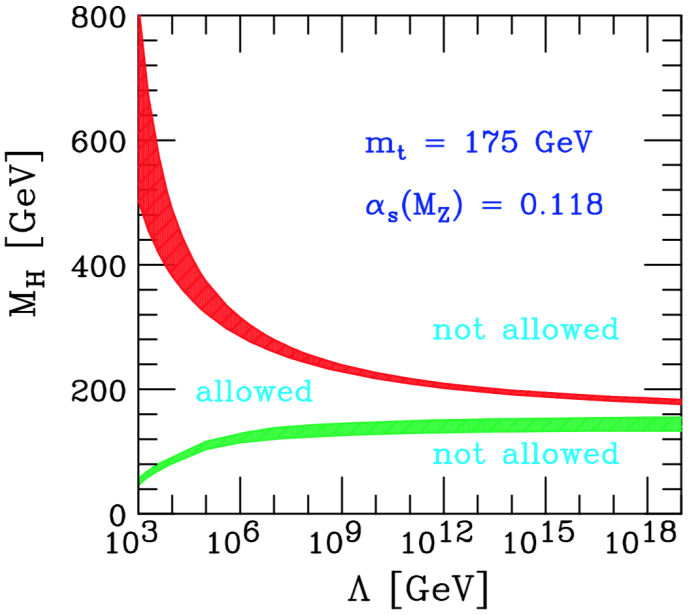
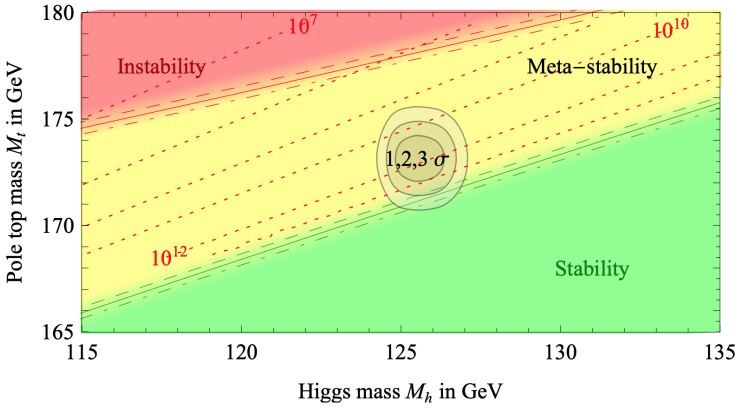


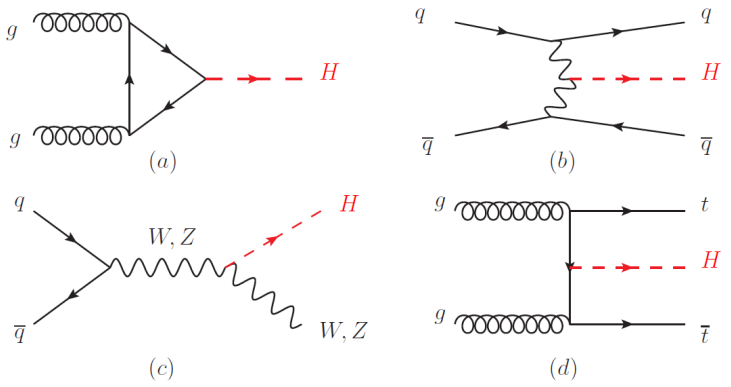
Figure 1.4 Stability of the Higgs potential as a function of the Higgs boson and of the top quark masses [35]. The current measurements of the top quark and of the Higgs boson masses indeed suggest that the electroweak vacuum is metastable. However, for a definite conclusion more accurate measurements especially of the top quark mass are needed.



1.6.2 Production processes

At hadron colliders, a Higgs boson can be produced through several processes [37], represented in Fig. 1.5.

Figure 1.5 Diagrams for Higgs boson production through gluon-gluon fusion (a), vector boson fusion (b), associated production with a vector boson (c) and with a top quark pair (d) [37].



The dominant process is the gluon-gluon fusion (ggH), with a cross section of almost 49 pb at $\sqrt{s}$ = 13 TeV for a Higgs mass m_H = 125 GeV. As the gluon is massless, it can not couple directly to the Higgs boson, and the process occurs via a quark loop, as shown in Fig. 1.5 (a). The dominant loop contribution arises from the top quark, because of its large mass. Despite its large production cross section, this production mode can be experimentally studied only when the Higgs boson decays to final states allowing sufficient background rejection.

The second Higgs production process is vector boson fusion (VBF or qqH), with a cross section of 3.8 pb at $\sqrt{s}$ = 13 TeV for a Higgs mass of m_H = 125 GeV. In this case, the Higgs boson is produced by the fusion of two W or Z vector bosons radiated by initial state quarks. The diagrams for this production mode are drawn in Fig. 1.5(b). The two additional quarks in the final state are detected as two hard jets, with a tendency to be in the forward and backward regions of the detector. Such peculiar topology is helpful to reject background events.

The third Higgs production mode in terms of cross section is the production in association with a vector boson (VH). This occurs when a virtual vector boson decays to its on-shell state radiating a Higgs boson, as shown in the diagram in Fig. 1.5(c) for the leading order contribution. Both the W and Z bosons contribute to this process, usually referred to as the VH production mode. At a center of mass energy $\sqrt{s}$ = 13 TeV, the VH cross section is $\approx$ 2.3 pb, of which about $\approx$ 0.9 pb due to the ZH process. Final

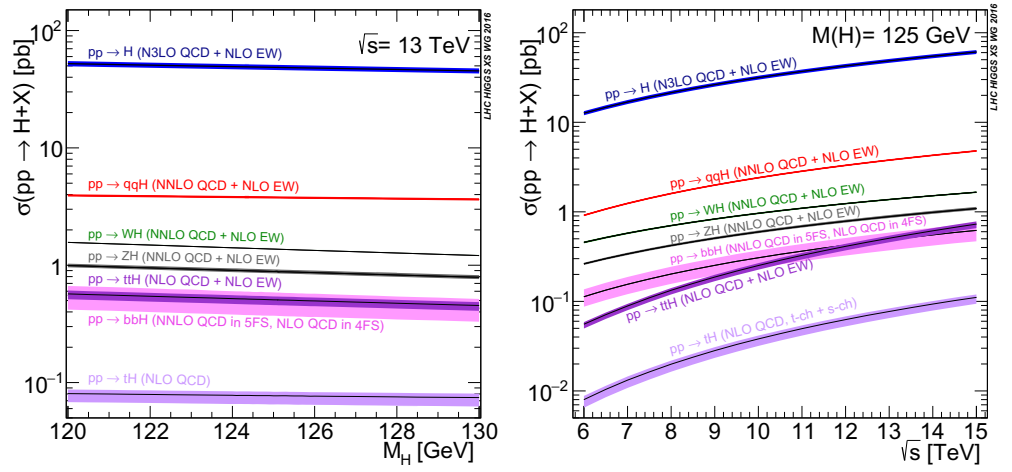
states with leptons from electroweak decays of vector bosons are particularly helpful for triggering and for background rejection in a hadronic environment.

Finally, a Higgs boson can be produced in association with heavy flavour quarks. The diagram for the production of a Higgs boson in association with a top quark pair ($t\bar{t}H$) is represented in Fig. 1.5(d). Despite the smaller cross section of ≈ 0.5 pb (at $\sqrt{s} = 13$ TeV and $m_H = 125$ GeV) with respect to other production modes, the measurement of the $t\bar{t}H$ production rate is extremely interesting to study the top-Higgs Yukawa coupling. Prompt leptons from top quark decays provide a handle to select signal events. A further Higgs production mode is the one in association with a $b\bar{b}$ quark pair. It proceeds through a similar diagram as for the $t\bar{t}H$ process and has a similar cross section, but it is experimentally very complicated to disentangle from the background.

The Higgs production cross section at the LHC is shown in Fig. 1.6, separately for each production mode, as a function of the Higgs mass (left) and as a function of the centre-of-mass energy (right).

The search for a Higgs boson decaying to a τ lepton pair presented in this thesis targets the two main production modes, namely gluon-gluon fusion and vector boson fusion. The VH and $t\bar{t}H$ production modes in the $\tau\tau$ final state are covered by dedicated analyses [38] [39].

Figure 1.6 Inclusive cross section for the production of a Standard Model Higgs boson from pp collisions, separately for each production mode, as a function of the Higgs boson mass (left) and of the center of mass energy (right). Calculations are performed up to the order indicated in the plot, color bands represent the uncertainty [40].



1.6.3 Decay modes

In the Standard Model, the total and partial decay widths of the Higgs boson can be predicted once the Higgs boson mass is known. The Higgs boson couples to fermions and vector bosons proportionally to their masses, and the preferred decay modes are the ones into the heaviest particles allowed by the kinematics.

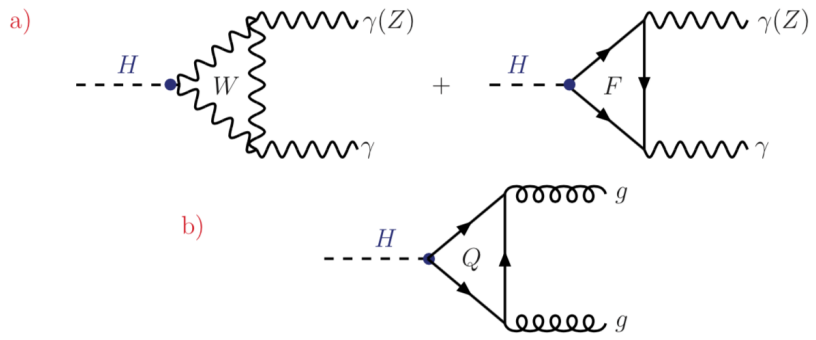
The coupling of the Higgs boson with the massive gauge bosons $V = W, Z$ is described in the Lagrangian density by the term $\mathcal{L}_{HVV} = (\sqrt{2}G_F)^{1/2} m_V^2 H V^\mu V_\mu$. Depending on the Higgs mass, the vector bosons in the final state may be both virtual, both real, or one virtual and one real. The decay width $H \to VV$ at tree level can be written as [41]:

$$\Gamma(H \to VV) = \frac{G_F}{\sqrt{2}} \frac{1}{16\pi} m_H^3 \delta_V F(m_1^2, m_2^2, m_H), \tag{1.56}$$

where $\delta_V = 2(1)$ for decays to W (Z) bosons, and the function $F(m_1^2, m_2^2, m_H^2)$ is a kinematic factor, depending on the squared invariant masses of the gauge bosons and of the Higgs boson.

Decays of the Higgs bosons to $\gamma\gamma$, $Z\gamma$ and gg are also possible through quantum loops involving massive particles, as shown in Fig. 1.7. The $\gamma\gamma$ and $Z\gamma$ couplings are mediated by W bosons and charged fermions, while the $H \to gg$ decay proceeds through quark loops. The dominant contribution to fermion loops comes from the top quark, while the b-quark starts playing a role for $m_H \gtrsim 100\,\text{GeV}$. These loop-induced decays are important for low Higgs masses $m_H \lesssim 130\,\text{GeV}$, where the total Higgs decay width is rather small. They are interesting as they could provide indirect evidence for the existence of new charged and/or colored heavy particles, which may be too heavy to be directly produced.

Figure 1.7 Loop-induced Higgs boson decays to $\gamma\gamma$ and $Z\gamma$ (top row) and to gg (bottom row) [29]



The Higgs boson coupling to fermion pairs of mass m_f is $g_{Hf\bar{f}} \propto m_f/v$, and the partial decay width, at tree level, is given by [29]:

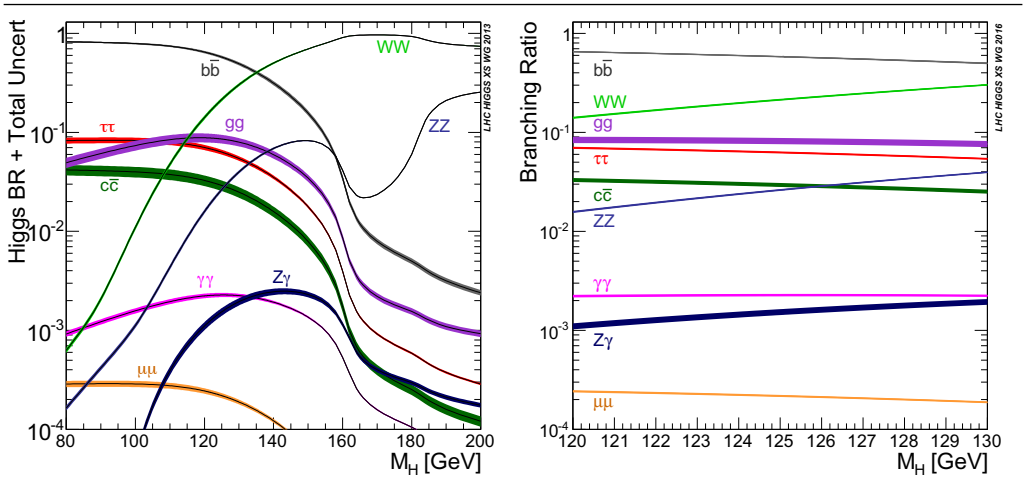
$$\Gamma(H \to f\bar{f}) = N_c \frac{G_F}{\sqrt{2}} \frac{1}{4\pi} m_H m_f^2 \beta_f^3, \tag{1.57}$$

where the factor N_c takes into account the color multiplicity of the final state, i.e. $N_c = 3$ for quarks and $N_c = 1$ for leptons, and the kinematic factor $\beta_f = (1 - 4m_f^2/m_H^2)^{1/2} \to 1$ for $m_H \gg m_f$.

The left plot of Fig. 1.8 shows the branching ratios for Higgs boson decays as a function of its mass, in the range from 80 to 200 GeV, while the right plot on the same figure

refers to the mass range between 120 and 130 GeV. For low masses, $m_H \lesssim 130\,\text{GeV}$, the dominant decay mode is to a $b\bar{b}$ quark pair, with a branching ratio of 57.7 % for $m_H = 125\,\text{GeV}$. The second fermionic decay mode is $H \to \tau^+\tau^-$, occurring 6.3 % of the times for a 125 GeV Higgs boson. Despite the smaller branching fraction, $H \to \tau^+\tau^-$ events are easier to identify at a hadron collider with respect to $H \to b\bar{b}$ decays, featuring a fully hadronic final state. Decays to $c\bar{c}$ quark pairs have a branching ratio of a few percent, but are extremely challenging to disentangle from the QCD background, while decays to a muon pair have a branching fraction at the per-mille level. At low masses, bosonic decays $H \to W^+W^-$ and $H \to ZZ$ have moderate branching ratios of a few percent, increasing quickly with the Higgs mass. For a 125 GeV Higgs boson, the decay fractions to W^+W^- and ZZ are 21 % and 2.6 %, respectively. Loop-induced decays to massless gauge bosons, namely $H \to gg$ and $H \to \gamma\gamma$, have branching ratios at the percent and per-mille levels, respectively. However, the clear signature of $H \to \gamma\gamma$ events, featuring two highly energetic photons emitted back-to-back, makes this one of the golden channels for Higgs boson searches at the LHC. Decays to vector bosons dominate for higher Higgs boson masses ($m_H \gtrsim 130\,\text{GeV}$).The total decay width as a function of the Higgs boson mass is shown in Fig. 1.9. For $m_H = 125\,\text{GeV}$, the total width measures $\Gamma_H = 4.07\,\text{MeV}$.

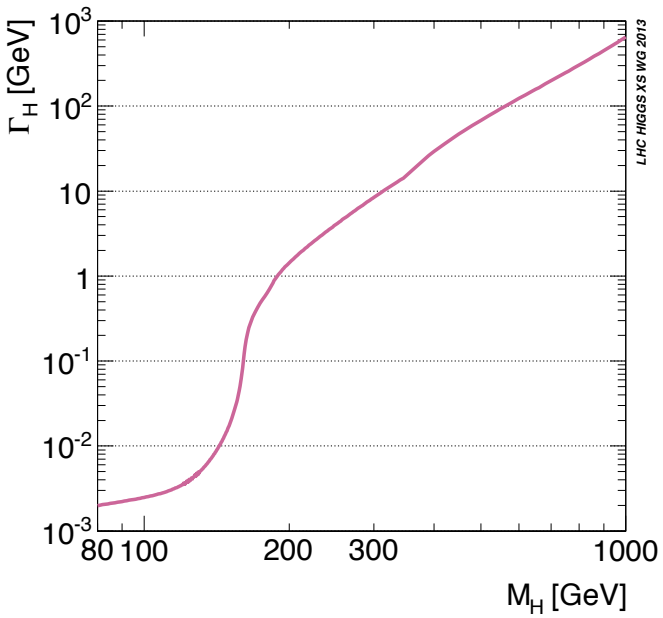
Figure 1.8 Branching ratios of Standard Model Higgs boson decays, as a function of its mass, in the range $80 < m_H < 200\,\text{GeV}$ (left), and for $120 < m_H < 130\,\text{GeV}$ (right) [40, 42].



1.7 The Higgs boson discovery

The first direct searches for the Higgs boson were conducted at the Large Electron-Positron Collider (LEP), operating from 1989 to 2000 at centre-of-mass energies from 91 GeV up to 209 GeV. As no direct evidence for Higgs boson production was found, the lower limit on the Higgs mass $m_H \geq 114.4\,\text{GeV}$ was set [43]. Direct searches at Tevatron excluded the range of $m_H \geq 150$ to 180 GeV [44]. Other experimental constraints came from precision measurements of electroweak observables, i.e. the W and Z

Figure 1.9 Total decay width of the Standard Model Higgs boson, as a function of its mass [40].



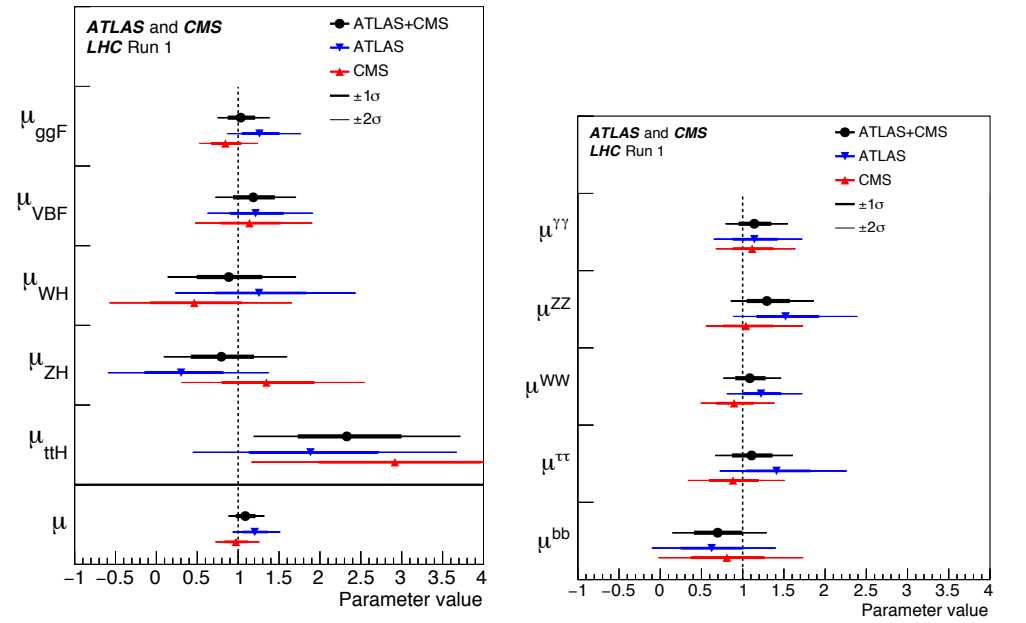
masses, which are sensitive to the Higgs boson mass through quantum corrections. The combination of the direct searches at LEP and Tevatron with electroweak precision measurements in 2011 led to the prediction $m_H = 125 \pm 10\,\text{GeV}$ [45].

In 2012, the ATLAS and CMS experiments at the LHC, by analysing proton-proton collision events at centre-of-mass energies of 7 and 8 TeV, discovered a new particle [46, 47, 48] with a mass of 125 GeV, compatible with the boson predicted by the Brout-Englert-Higgs mechanism of electroweak symmetry breaking. The discovery was driven by the bosonic decay channels, namely the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$, which are the most sensitive ones at the LHC.

At the end of the LHC Run 1, a combination of the ATLAS and CMS measurements of the Higgs boson properties was performed. The mass of the Higgs boson was measured from the decay channels with the best resolution, the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$, to be $m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})$ GeV [36]. From the combination of ATLAS and CMS analyses [49] targeting five production processes (ggF, VBF, WH, ZH and $t\bar{t}H$) and six decay modes ($H \to ZZ^*, W^+W^-, \gamma\gamma, \tau^+\tau^-, b\bar{b}$ and $\mu^+\mu^-$), the combined signal yield relative to the SM expectation was measured to be $1.09 \pm 0.07(\text{stat.}) \pm 0.08(\text{syst.})$. Figure 1.10 (left) shows the production signal strengths for the different production processes and their combination. The VBF production process was observed for the first time with a significance of $5.4\,\sigma$ ($4.6\,\sigma$ expected). The best fit results for the decay signal strengths for several analysed decay modes are shown in Fig. 1.10 (right), and are all compatible with the Standard Model expectation within the uncertainties. Moreover, couplings of the Higgs boson to Standard Model particles were also measured. Coupling modifiers κ_V and κ_f are introduced to allow the coupling of the Higgs to bosons

and fermions respectively to vary, the Standard Model predicting $\kappa_V = \kappa_f = 1$. New parameters can be derived from the coupling modifiers, to make explicit the dependence on the particle masses: linear for the Yukawa couplings to the fermions and quadratic for the gauge couplings of the Higgs boson to the weak vector bosons. The best fit results for such parameters as a function of the particle mass are shown in Fig. 1.11, together with the Standard Model prediction. All measured couplings are in good agreement with the predictions within the uncertainties.

Figure 1.10 Best fit results for the production (left) and decay (right) signal strengths for the combination of ATLAS and CMS Run 1 data, also shown separately for each experiment. The error bars indicate the 1 σ (thick lines) and 2 σ (thin lines) intervals [49].



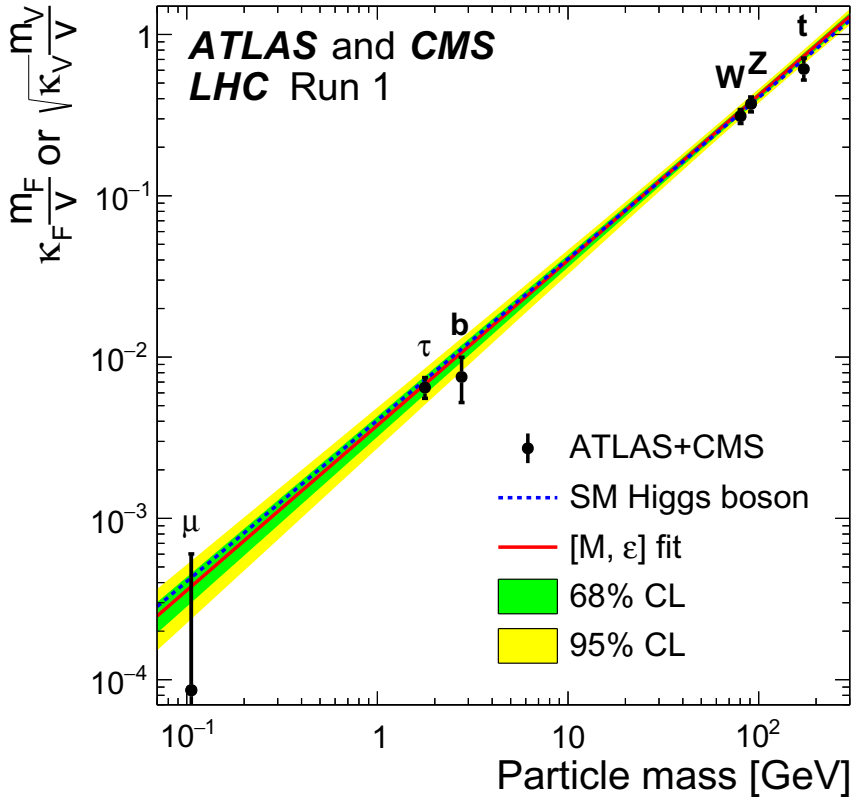
1.8 Open questions

The discovery of the Higgs boson and its characterisation being performed by several LHC measurements with increasing precision strengthen the solidity of the Standard Model. On the other hand, there are several questions to which our current understanding of nature based on the Standard Model can not give an answer. From the experimental point of view, the following evidences can not be explained:

- neutrinos have non-zero masses, as proven by neutrino oscillations;
- the asymmetry in the amounts of matter and antimatter is much larger than expected from *CP* violation in the Standard Model;
- there is additional non-interacting matter, called *dark matter*, for which no candidates are provided.

From a more theoretical perspective, other important points remain open:

Figure 1.11 Best fit values for coupling modifiers of the Higgs boson to Standard Model particles. For fermions, the parameter used is $\kappa_F m_F/v$, while for bosons it is $\sqrt{\kappa_V} m_V/v$, where v = 246 GeV is the vacuum expectation value of the Higgs field. The solid red line indicates the best fit results to the phenomenological model suggested in [50], while the dashed blue line indicates the predicted dependence on the particle mass for the Standard Model Higgs boson [49].



- what is the quantum description of the gravitational interaction;
- why forces have different coupling constants, and whether it is possible to unify them;
- which mechanism protects the Higgs mass from large quantum corrections which would push it towards the Planck scale $\Lambda_P \sim 1.2 \times 10^{19}$GeV;
- why there are so many, i.e. eighteen, free parameters in the Standard Model.

Plenty of theories have been proposed attempting to overcome the limits of the Standard Model, and they are being tested thoroughly. At the LHC, a large number of searches for new physics phenomena are being performed mainly by the ATLAS and CMS experiments, which thanks to their main purpose detectors can efficiently and precisely reconstruct several particles signatures.

Besides direct searches for new particles, hints for physics beyond the Standard Model could be found indirectly, measuring deviations of observables from their Standard

Model expectation. In particular, due to the recent discovery of the Higgs boson, the Higgs sector is still relatively unexplored. The aim of this thesis is to characterise the τ lepton Yukawa coupling in $H \to \tau^+\tau^-$ decays using data collected by the CMS experiment at the LHC. Therefore, the next chapter is dedicated to describe the main characteristics of the LHC accelerator and of the CMS detector.

Chapter 2

The CMS experiment at the LHC

2.1 The Large Hadron Collider 31
2.2 The CMS detector 37
2.2.1 Solenoid 37
2.2.2 Tracker 39
2.2.3 Electromagnetic calorimeter 42
2.2.4 Hadron calorimeter 44
2.2.5 Muon system 45
2.2.6 Trigger system 47
2.3 Event simulation in CMS 49

The data analysed in the context of this thesis has been collected by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). In this Chapter, the main features of the LHC accelerator and of the CMS detector are described. Finally, the techniques adopted to produce an accurate event simulation are summarised, as simulated events have been used in this work for the analysis optimisation and interpretation of the results.

2.1 The Large Hadron Collider

The Large Hadron Collider is the most powerful particle accelerator built so far. It is located at CERN near Geneva in Switzerland and installed in the 26.7 km long tunnel which formerly hosted the Large Electron Positron collider (LEP). The LHC has been built for fundamental research in high energy physics, especially aiming at the discovery of the Higgs boson and at the search for new physics beyond the Standard Model.

The LHC machine [51] is a two ring accelerator designed to collide proton bunches at a centre-of-mass energy of 14 TeV and with an instantaneous peak luminosity $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}. The latter parameter is defined as the proportionality factor between the rate of events

produced in the collisions for a given process, and its cross section at the machine operating energy:

$$R = \mathcal{L}\,\sigma \tag{2.1}$$

The instantaneous luminosity exclusively depends on the beam parameters; for identical bunches with a Gaussian profile and colliding head-on, it can be written as:

$$\mathcal{L} = f\frac{N^2 n_b}{4\pi\sigma_x\sigma_y} \tag{2.2}$$

where f is the revolution frequency, N the number of protons in each bunch, n_b the number of bunches per beam, σ_x and σ_y the transverse dimensions of the bunch. At the LHC, the nominal revolution frequency is 11.3 kHz, with collisions occurring every 25 ns. The beam can be filled with a maximum of 2808 bunches, each containing nominally 1.15×10^{11} protons and with average transverse dimensions $\sigma_x = 16\,\mu\text{m}$ and $\sigma_y = 100\,\mu\text{m}$ at the interaction points. The maximum particle density per bunch is limited by the nonlinear beam-beam interaction experienced by the particles in the interaction region. It results in a tune shift that must be kept reasonably small in order to avoid beam losses due to resonances. If the crossing angle at the interaction point is not null, the luminosity is reduced. At the LHC, the beam crossing angle is about 300 µrad, therefore a 15 % decrease in the luminosity calculated according to Eqn. 2.2 must be taken into account [51]. The force needed to bend 7 TeV proton beams along the desired trajectories is obtained with 1232 superconducting dipole magnets, each of them being 14.3 m long. They are cooled down to 1.9 K and produce a magnetic field of 8.3 T. The basic layout of the LHC lattice is shown in Fig. 2.1. It consists of eight arcs spaced by eight 500 meters long straight sections, which serve as experimental or utility insertion. The two beams circulate in two separated pipe lines, hosted in the dipole magnets, as can be seen in Fig. 2.2. Figure 2.3 shows the two beam pipes and the field lines of the LHC dipole magnets. When approaching the collision points, the beams are squeezed by dedicated quadrupole magnets to increase the luminosity and are then collided[1].

Currently, four main experiments are located on the LHC, their detectors being placed in correspondence of the interaction points (IP), as illustrated in Fig. 2.1. The ATLAS [55] and CMS [56] detectors are located at IP 1 and 5 respectively. They are multipurpose detectors designed to record data in high luminosity regime. The LHCb experiment [57] focuses on B-physics and is located at IP 8. Moreover, during special runs the LHC is also operated with lead ion beams and the dedicated experiment called ALICE [58] is located at IP 2. In addition to the main experiments, other three experiments are intended for more specific studies: TOTEM [59] and LHCf [60] focus on forward production, while MoEDAL [61] aims at searching for magnetic monopoles.

[1]Sextupole and octupole magnets are included in the LHC lattice to reduce tune shifts.

Figure 2.1 Schematic layout of the LHC. Beam 1 circulates clockwise and Beam 2 counterclockwise. The regions where the beams are squeezed and collided are indicated with a star sign. The positions of the four experiments on the interaction points are indicated [52].

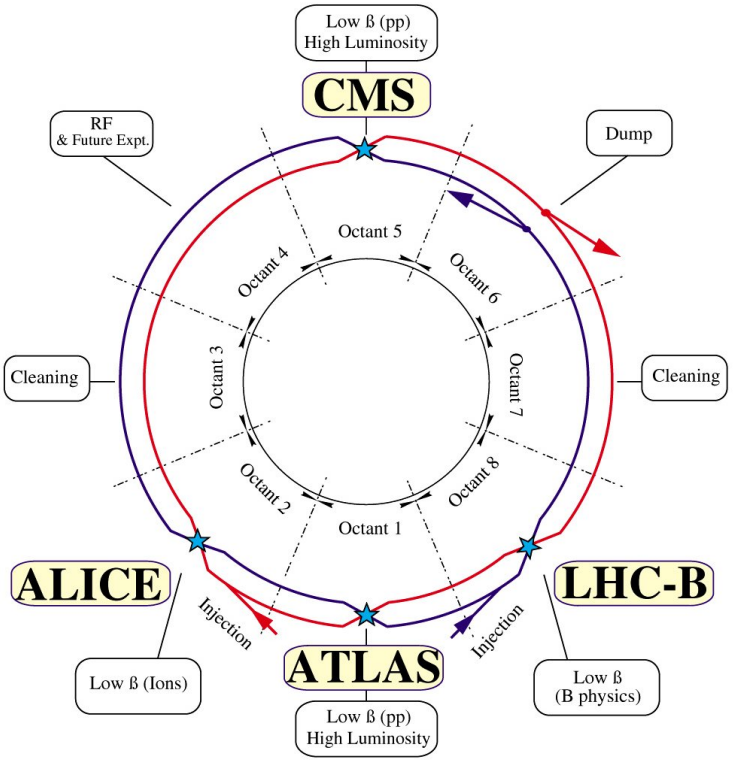


Figure 2.2 Cross section of a LHC dipole magnet showing its structure and features [53].

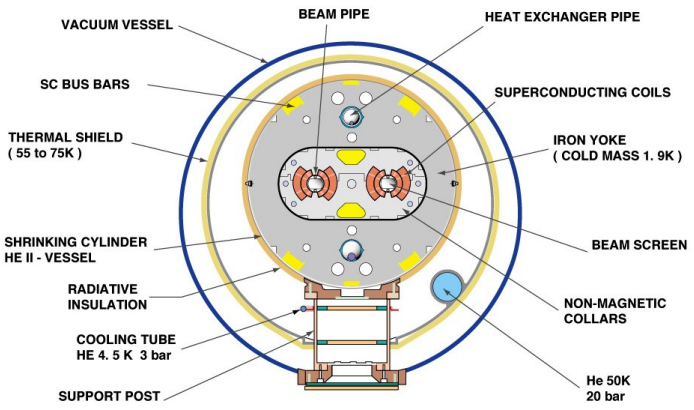
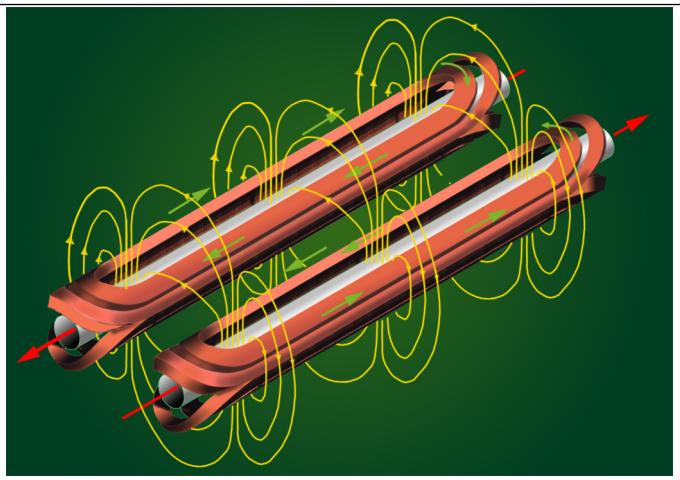


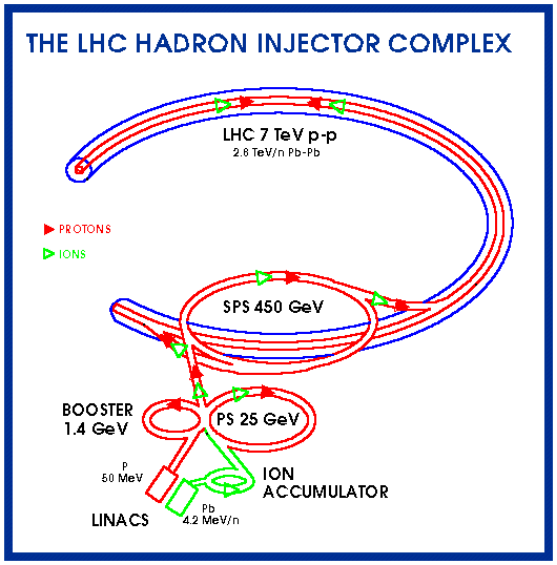
Figure 2.3 Beam pipes and magnetic field lines of the LHC dipole magnets [54].



LHC Machine Operation

Protons injected in the LHC are obtained by a source of hydrogen gas, stripping hydrogen atoms of their electrons using an electric field. They are accelerated in the Linac 2 linear accelerator bringing their energy up to 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the Proton Synchrotron (PS) pushing the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Then, they are finally transferred to the two beam pipes of the LHC and accelerated to the desired energy thanks to radio-frequency cavities. The minimum time required for ramping up the beam energy from 450 GeV to 7 TeV is approximately 20 minutes. Under normal operating conditions, beams circulate for many hours inside the LHC beam pipes. Figure 2.4 shows the different accelerating machines of the CERN complex.

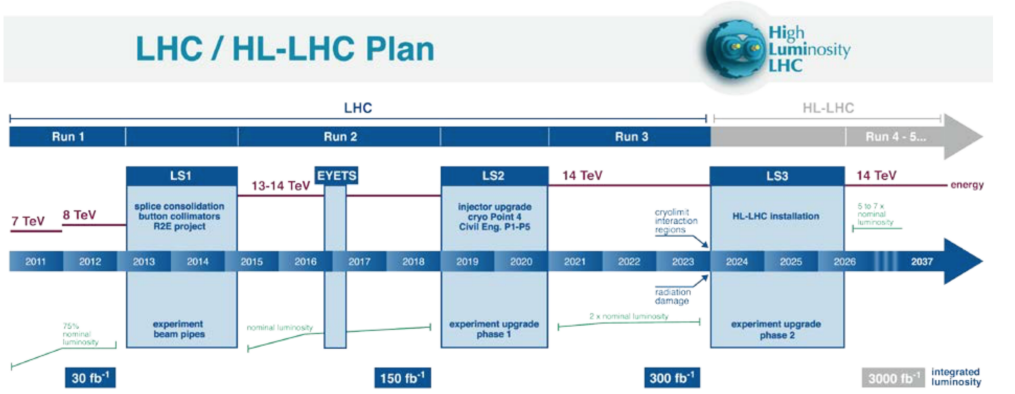
Figure 2.4 The LHC injection chain [51].



Schedule of the LHC Runs

Figure 2.5 summarises the plan of the whole LHC project from 2011 to approximately 2035, showing the different run times and the corresponding goals in terms of instantaneous and integrated luminosity, as well as the planned technical stops.

Figure 2.5 Timeline of the LHC project, from the beginning of its operation in 2010 to the end of the High Luminosity Phase in 2035. Technical stops and long shutdowns (LS) are needed to consolidate and upgrade both the LHC machine and the experiments [62].



The operation of the LHC [63] started in March 2010 with the first collisions of 3.5 TeV proton bunches, for a total centre-of-mass energy of 7 TeV. In 2012 the centre-of-mass energy was increased to 8 TeV, with around 1380 bunches per beam spaced in time by 50 ns.

The first phase of LHC operation, usually referred to as Run 1, lasted until December 2012. The total integrated proton-proton luminosity delivered to both the ATLAS and CMS experiments reached nearly 30 fb^{-1} and enabled the discovery of a Higgs boson. One of the main features of operation in 2011 and 2012 was the high bunch intensity (up to 150 % of nominal with 50 ns bunch spacing) and lower-than-nominal emittance thanks to the excellent performance of the injector chain of Booster, Proton Synchrotron and Super Proton Synchrotron. The cost to the experiments was high pileup, i.e. the number of interactions per bunch crossing, with an average of 21 in 2012. The first long shutdown (LS1) started in February 2013 and lasted about two years, during which both the accelerator complex and the detectors of the different experiments could be repaired and consolidated in preparation of the next run.

The LHC Run 2 has started in summer 2015 with the first proton collisions at an energy of 13 TeV and bunches spaced by 25 ns, and is supposed to continue until the end of 2018. Short shutdown periods are foreseen at the end of every year, allowing the consolidation of the accelerator complex and of the experiments. Similarly as in Run 1, the performance of the LHC accelerator exceeded the expectations, with about 4 and 40 fb^{-1} delivered to the CMS experiment in 2015 and 2016, respectively, as it can be observed in Fig. 2.6. The data analysed in this thesis for the search of the

Higgs boson decaying to τ leptons corresponds to the full 2016 dataset. The main LHC operational parameters in that year are summarised in Table 2.1. Most noticeably, the instantaneous luminosity reached up to $1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, resulting in an average pileup of 27 interactions per bunch crossing, as shown in Fig. 2.7.

Figure 2.6 Integrated luminosity delivered to CMS during stable beams and for proton collisions in 2015 (left) and 2016 (right) [64].

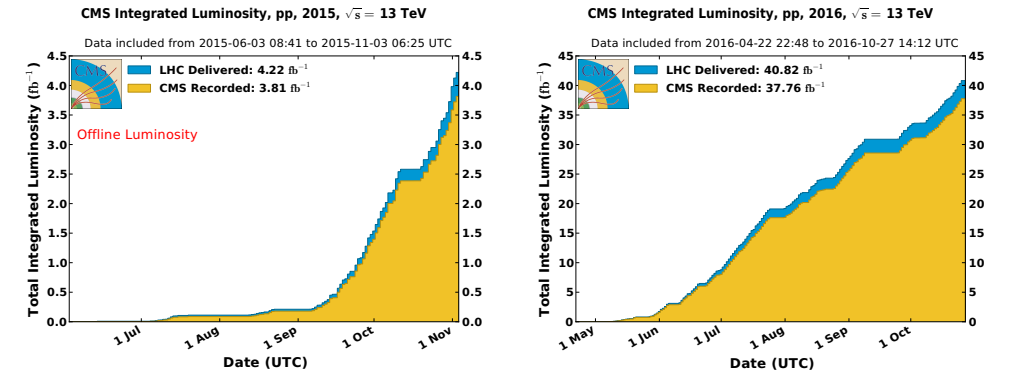
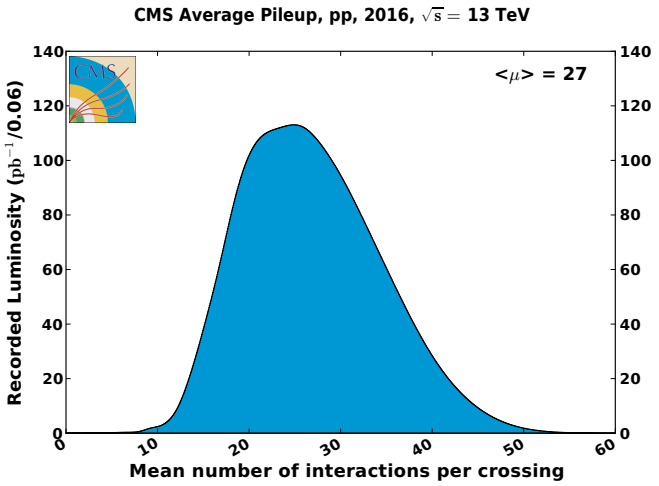


Table 2.1 Main operational parameters of the LHC accelerator in 2016.

Parameter	Value
Beam energy	6.5 TeV
Bunch spacing	25 ns
Max. number of bunches	2220
Protons per bunch	1.12×10^{11}
Peak luminosity at IP1 and 5	$1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$

Figure 2.7 Distribution of the mean number of interaction per bunch crossing (average pileup) in the data recorded with the CMS detector in 2016 [64].



At the end of Run 2 in 2018 a total integrated luminosity of $150\,\mathrm{fb}^{-1}$ is expected to be delivered by the LHC. After that, a second long shutdown of two years (LS2) is

foreseen and the LHC Run 3 should start in 2021 with collisions at 14 TeV and higher instantaneous luminosity. The goal for Run 3 is to reach a total integrated luminosity of 300 fb^{-1}. At the end of Run 3, another long shutdown time is foreseen to prepare for the high luminosity phase of the LHC, where the instantaneous luminosity is planned to be increased to 5-7 times the nominal one, allowing to collect integrated luminosity of 250 fb^{-1} per year with the goal to reach at least 3000 fb^{-1} around 2035.

2.2 The CMS detector

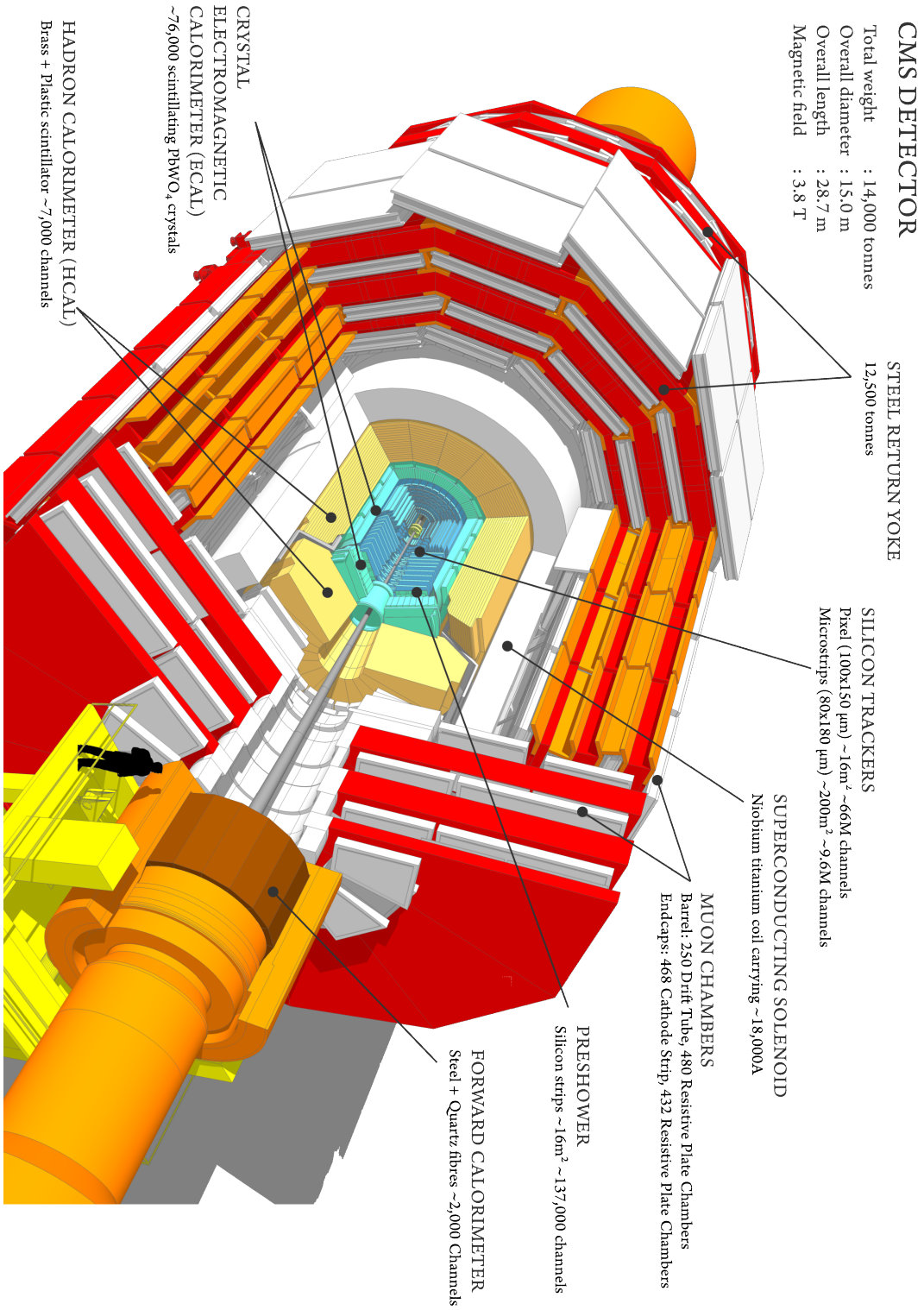
The Compact Muon Solenoid (CMS) [56] is one of the two general purpose detectors at the LHC, and it is located in correspondence of the IP 5 of the LHC. It has been designed to address a wide physics program, including tests of the Standard Model at the TeV scale, the search for the Higgs boson, and extensive searches for new particles predicted by supersymmetry and exotic models. Besides, precision measurements in the B-physics sector are of primary importance, as some rare decays might be enhanced due to the contributions of processes involving new particles. In order to meet the above mentioned goals at the LHC environment, the detector must provide good particle identification and momentum resolution for muons and other charged particles, with high efficiency and over a wide geometrical range. Moreover, good electromagnetic and hadronic energy resolution and almost hermetic coverage are required. Due to the large flux of particles coming from the interaction region, both the detectors and the front end electronics must be able to work in a highly irradiated environment. The adopted detector structure is cylindrically symmetric around the beam direction, with a big barrel region consisting of cylindrical detecting layers coaxial with the beam itself, closed at both ends with detecting disks, forming two endcaps to close the structure on both sides. The main distinguishing feature of the CMS detector is its compact design. The high solenoidal magnetic field provides large bending power to precisely measure the momentum of charged particles, and the return field is confined in an iron yoke in which the muon detectors are embedded. The full silicon-based inner tracker, the homogeneous electromagnetic calorimeter and the hadronic calorimeter are placed inside the bore of the magnet coil. The overall dimensions of the CMS detector are a length of 21.6 m, a diameter of 14.6 m and a total weight of 12 500 tons. A schematic view of the whole detector showing the different components is given in Fig. 2.8.

In the following, an overview of the most important components of the CMS detector is given.

2.2.1 Solenoid

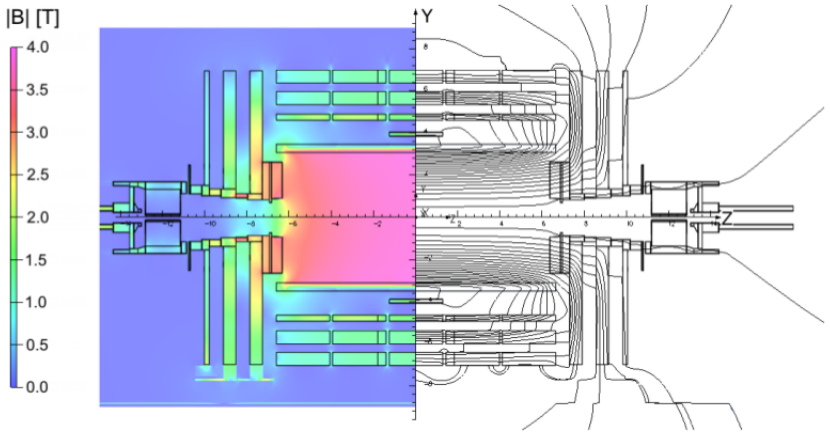
The CMS magnet is a superconducting solenoid providing a magnetic field of 3.8 T. Magnetic field strength and lines on a longitudinal section of the CMS detector are shown in Fig. 2.9. The coil has a inner radius of 5.9 m and a length of 12.9 m. It is

Figure 2.8 Sectional view of the CMS detector with all its subsystems. The LHC beams travel in opposite directions along the central axis of the CMS cylinder colliding in the middle of the detector [65].



cooled down to 4.5 K by a helium refrigeration plant and the whole structure is kept in insulation by two pumping stations providing vacuum on the cryostat volume. The field is closed by a 10 000 tons heavy iron return yoke made by five barrel wheels and two endcaps of three layers each. Muon stations are embedded between the yoke layers, exploiting the return field and enabling a very compact layout.

Figure 2.9 The magnetic field strength (left) and lines (right) of the CMS superconducting solenoid. Muon stations are embedded between the yoke layers, where the return field is exploited to bend muons [66].



2.2.2 Tracker

The tracker is the innermost part of the CMS detector, and has the purpose of providing high precision three-dimensional measurements of the hits generated by charged particles. Reconstructed tracks are used to identify the primary and secondary vertices of the event and to measure particle momentum p_T, which can be obtained as follows, once the radius of curvature R of the track is measured and the magnetic field strength B is known:

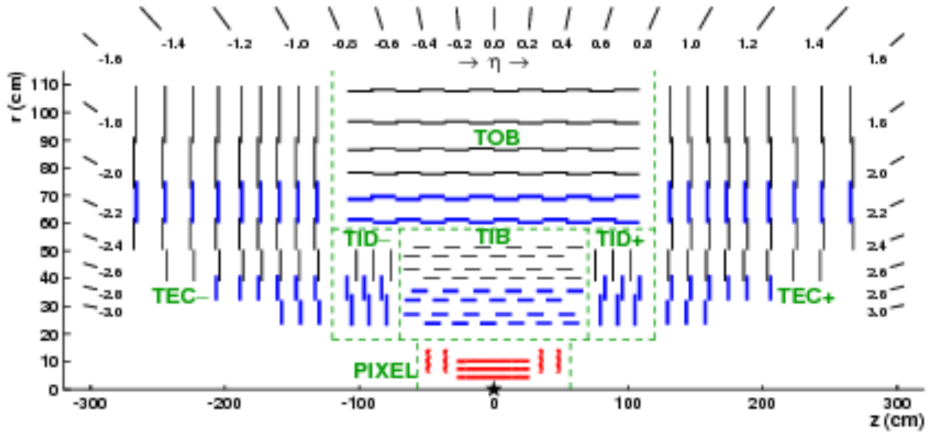
$$p_T\,[\text{GeV}] = 0.3 \cdot B\,[\text{T}] \cdot R\,[\text{m}] \tag{2.3}$$

The layout of the CMS tracker is shown in Fig. 2.10. It extends radially from 4.4 cm to nearly 110 cm away from the interaction point, and its total length is approximately 540 cm. The tracker is composed by silicon modules, with a total active area of more than 200 m^2. The use of silicon has the main following advantages:

- small charge collection time, thanks to the high mobility of charge carriers;
- high density, resulting in a high number of electron-hole pairs created by a ionising particle traversing it ($\sim$ 396 eV/µm corresponding to 110 electron-hole pairs/µm for a minimum ionising particle);
- good intrinsic energy resolution provided by a low ionisation energy ($\sim$ 3.6 eV).

In the innermost region, modules are segmented in pixels. Their small size allows keeping low occupancy despite the high flux of particles and achieving a better hit resolution, which is necessary for the reconstruction of the interaction vertices. Starting from a radius of 20 cm instead, silicon microstrip modules are used. The tracker is operated at low temperature to reduce radiation damage.

Figure 2.10 Schematic view of the CMS tracker in the $r - z$ plane. The pixel detector (red) has three layers in the barrel and four disks in the endcap region. Strip tracker modules are in black and blue (stereo modules) [67].

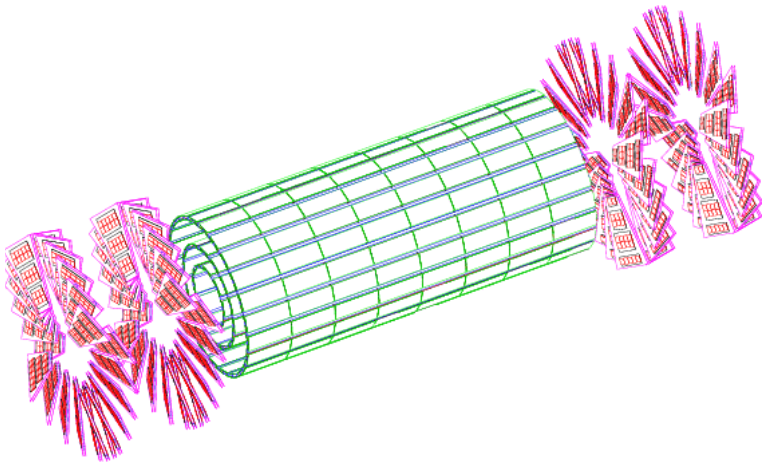


Pixel Tracker

The pixel detector consists of three cylindrical barrel layers with two endcap disks on each side. Barrel layers are located at a mean distance of 4.4 cm, 7.3 cm and 10.2 cm from the interaction point, and have a length of 53 cm. The endcap disks, extending from 6 to 15 cm in radius, are placed on each side at $|z| = 34.5$ cm and 46.5 cm. In order to achieve the optimal vertex position resolution, a design with an almost square pixel shape of 100×150 μm^2 has been adopted. The barrel comprises 768 pixel modules arranged into half-ladders of 4 identical modules each. The endcap disks comprise 672 pixel modules arranged in blades, each of them composed by seven modules. The overall layout of the pixel detector is sketched in Fig. 2.11. The total area covered by the pixel modules is about 1 m^2. Each pixel cell is bump bonded to a read out chip (ROC), and there are approximately 16000 ROCs for a total of about 66 million readout channels. The coverage of the pixel tracker extends up to $|\eta| = 2.5$.

The spatial resolution is measured to be about 10 µm for the $r - \phi$ measurement and about 20 µm for the z measurement. Such resolutions are achieved not only thanks to the small pixel size but also exploiting the charge sharing among neighbouring pixels induced by the CMS magnetic field [69] in which the tracker is immersed: charge carriers produced in the active volume do not drift along the electric field lines but are deflected by the Lorentz force, resulting in a sharing of the collected charge among more close-by pixels. The hit position is obtained evaluating the centre of gravity of the

Figure 2.11 View of the CMS pixel tracker, with the three barrel layers (green) and two encap disks (pink) on each side [68].



collected charge over the pixels, therefore improving the resolution. The magnitude of the Lorentz angle depends on the bias electric field and on the magnetic field applied. For the CMS pixel tracker, it has been measured to be $\sim 25°$ in the barrel region [70]. In order to induce charge sharing also in the endcap disks, the pixel endcap modules are arranged in a turbine-like geometry, with petals tilted by 20°.

Strip Tracker

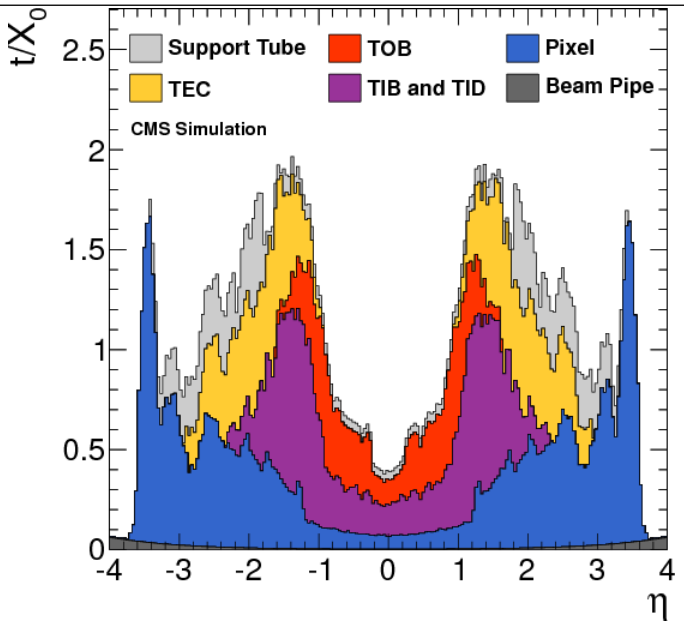
The silicon strip detector consists of almost 15400 modules, mounted on carbon-fiber structures and housed inside a temperature-controlled outer support tube. Silicon microstrip modules are placed from a radial distance of 20 cm up to 110 cm from the interaction point. The tracker barrel part is separated into an Inner (TIB) and an Outer Barrel (TOB). In order to avoid excessively shallow track crossing angles, the inner barrel is shorter than the outer barrel, and there are three additional inner disks on each side in the transition region between the barrel and the endcaps. The TIB is made of four layers and covers up to $|z| < 65$ cm, using silicon sensors with a strip pitch which varies from 80 to 120 µm. The first two layers are made with stereo modules, consisting of two silicon sensors mounted back-to-back with their strips aligned at a relative angle in order to provide a measurement in both $r - \phi$ and $r - z$ coordinates. An angle of 100 mrad has been chosen, resulting in a single-point resolution of between $23 - 34$ µm in the $r - \phi$ direction and 230 µm in z. The TOB comprises six layers with a half-length of $|z| < 110$ cm. The strip pitch varies from 120 to 180 µm. Also for the TOB the first two layers are rotated by the same stereo angle as for the TIB to achieve a single-hit $(r-\phi, z)$ resolution of $35-52$ µm and 520 µm respectively. The strip endcaps are divided into the TEC (Tracker End Cap) and TID (Tracker Inner Disks). Each TEC comprises 9 disks extending into the region $120 < |z| < 280$ cm, and each TID comprises 3 small disks filling the gap between the TIB and the TEC. Modules are arranged in rings with strips pointing towards the beam line and having a variable pitch ranging between

100 and 180 µm. Table 2.2 summarises the main characteristics of the various tracker subsystems. The material budget of the CMS tracker in units of radiation length X_0 as a function of η is provided in Fig. 2.12, showing the contributions from the different sub-components.

Table 2.2 Main characteristics of the different tracker subsystems, including their geometrical structure, coverage and segmentation. For the endcap region, the numbers refer to a single endcap [67].

Tracker subsystem	Layers	Pitch	Location
Pixel tracker barrel (BPIX)	3 cylindrical	$100 \times 150\,\mu m^2$	$4.4 < r < 10.2\,cm$
Strip tracker inner barrel (TIB)	4 cylindrical	$80 - 120\,\mu m$	$20 < r < 55\,cm$
Strip tracker outer barrel (TOB)	6 cylindrical	$122 - 183\,\mu m$	$55 < r < 116\,cm$
Pixel tracker endcap (FPIX)	2 disks	$100 \times 150\,\mu m^2$	$34.5 < \|z\| < 46.5\,cm$
Strip tracker inner disks (TID)	3 disks	$100 - 141\,\mu m$	$58 < \|z\| < 124\,cm$
Strip tracker endcap (TEC)	9 disks	$97 - 184\,\mu m$	$124 < \|z\| < 282\,cm$

Figure 2.12 Material budget of the CMS tracker in units of radiation length X_0, as a function of η [71].



2.2.3 Electromagnetic calorimeter

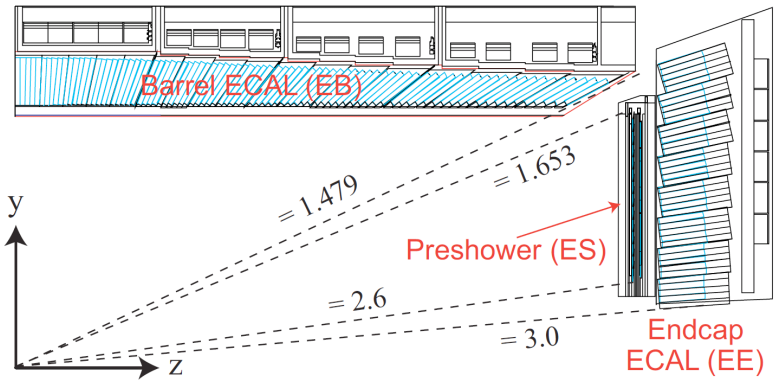
Electromagnetic calorimetry is based on the production of a shower initiated by a photon or an electron inside a scintillating material. An efficient collection of the emitted light allows to measure the initial particle energy. The Electromagnetic Calorimeter (ECAL) of the CMS detector is a hermetic homogeneous calorimeter composed by lead-tungstate ($PbWO_4$) scintillating crystals. This material has been chosen because of its short radiation length ($X_0 = 0.89\,cm$) and Molière radius (2.2 cm), together with fast light emission and radiation hardness. The light yield of about 30 photons/MeV is relatively

low and is collected by silicon-avalanche photodiodes APDs (vacuum phototriodes in the endcaps). Both the crystals and photodiodes responses are sensitive to temperature changes, requiring high temperature stability: a water cooling system is used to keep the ECAL temperature at 18 °C. The barrel ECAL section has an inner radius of 129 cm and is composed by 61200 crystals with a front cross section of about 22 × 22 mm^2, comparable to the typical transverse shower size in $PbWO_4$, and a length of 230 mm, corresponding to 25.8 X_0. The barrel covers the pseudorapidity region up to $|\eta| = 1.479$. The endcaps, placed at a distance of 314 cm from the interaction region and covering the pseudorapidity range $1.479 < |\eta| < 3.0$, are structured as two semi-circular aluminium plates containing units of 5 × 5 crystals, known as supercrystals. Each endcap hosts 7324 crystals with a front cross section of 28.6 × 28.6 mm^2 and a length of 220 mm (24.7 X_0). A preshower detector designed for a more efficient π_0 / γ separation is placed in front of the crystal calorimeter over much of the endcap η range. It is composed of two layers of lead absorber instrumented with orthogonal layers of silicon strip sensors. About 95 % of the incident photons are converted in a e^+e^- pair before reaching the second layer. Figure 2.13 shows the arrangement of the ECAL crystals belonging to one quarter of the whole calorimeter. The energy resolution of the CMS electromagnetic calorimeter can be decomposed in three different contributions:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{2.4}$$

where S and N are the stochastic and noise terms respectively, and C is a constant energy independent contribution. The values of these parameters have been obtained from a Gaussian fit to the reconstructed energy distributions measured during tests with electron beams [72]. The results and the resolution as a function of the particle energy are shown in Fig. 2.14. The optical transmission of the crystals is continuously measured by a laser system, allowing to correct for losses due to irradiation, which can reach up to 10 % in the endcaps at high luminosity.

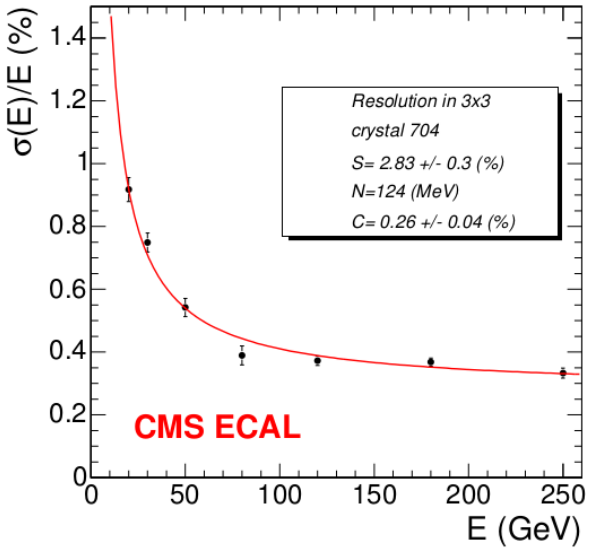
Figure 2.13 Geometric view of one quarter of the ECAL showing the arrangement of crystal modules in the barrel and endcaps region. The preshower detector is placed in front of the endcap [56].



2.2.4 Hadron calorimeter

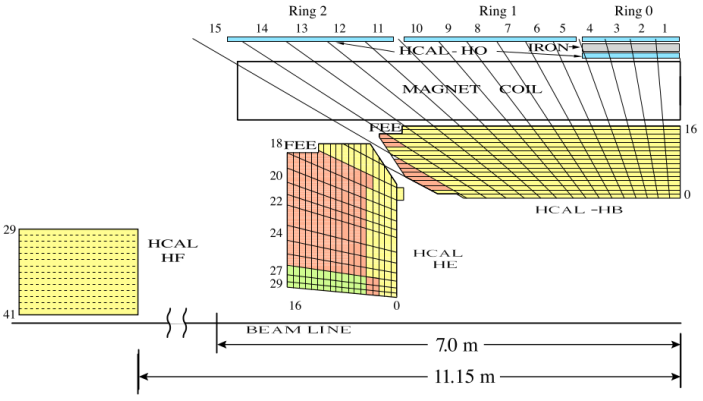
The CMS hadron calorimeter (HCAL) is a sampling calorimeter for hadron energy measurements. In a sampling calorimeter, layers of scintillating active material are spaced out by layers of absorber material. For the CMS HCAL, brass has been chosen as absorber because it has a reasonably short interaction length and is a non-magnetic material, while the active material consists of plastic scintillator tiles, read out by wavelength-shifting fibres. The HCAL should provide good containment and hermeticity, so that the measured missing energy in the plane orthogonal to the beam direction is actually due to non interacting particles (e.g. neutrinos), and not to mismeasured hadronic energy. The HCAL surrounds the ECAL and is designed to maximise the amount of material (quantified in interaction lengths) inside the magnet coil. The hadron calorimetry system is organised in an inner barrel, two endcaps, two forward calorimeters and an outer detector placed outside the magnet solenoid sampling the energy of penetrating hadron showers leaking through the back part of the barrel. The granularity of the three parts of the HCAL has been chosen to result in a similar transverse energy resolution. The hadron barrel (HB) part consists of wedges covering the region $|\eta| < 1.4$, with a final granularity of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The energy resolution in the barrel region is approximately $\sigma/E \sim 90\,\%/\sqrt{E}$. The hadron outer (HO) detector is placed in the wheels of the return yoke. It contains 10 mm thick scintillators with the same angular segmentation of the HB and covers the region $|\eta| < 1.26$. Thanks to the HO, the effective thickness of the HCAL in the central η region exceeds 10 interaction lengths, with the benefit of improving energy resolution. The two hadron endcaps consists of 14 η towers covering the region $1.3 < |\eta| < 3.0$. The granularity is higher in the outermost part, corresponding to smaller η, with $\Delta\phi$ from 5 to 10 degrees and $\Delta\eta$ from 0.087

Figure 2.14 Energy resolution of the electromagnetic calorimeter as a function of the electron energy, as measured in a test beam. Measured values of the three parameters in Eqn. 2.4 are listed in the box [72].



to 0.35 at the highest pseudorapidities. Finally, the hadron forward calorimeter (HF), located 11.15 m away from the interaction point, ensures coverage up to $|\eta| = 5$. As the forward region is characterised by a high flux of particles, a steel/quartz fibre structure has been chosen. The absorber has a depth of 1.65 m, the diameter of the quartz fibres is 0.6 mm and they are arranged in a square grid, 5 mm apart from each other. The signal originating from Cherenkov light emitted in the quartz fibres is channeled to photomultipliers. The overall assembly enables the HCAL to be built with essentially no dead areas in ϕ, while there is a gap between the barrel and the endcap HCAL, through which the services of the ECAL and the inner tracker find their way out of the detector. The affected η region is $1.48 < |\eta| < 1.53$. A schematic view of the HCAL structure is shown in Fig. 2.15.

Figure 2.15 Schematic quarter view of the CMS hadronic calorimeter (HCAL) in the barrel (HB and HO), endcap (HE) and forward (HF) regions [73].



2.2.5 Muon system

The muon system is responsible for muon identification, momentum measurement and triggering. It is located outside the solenoid and embedded in the iron return yoke and extends up to $|\eta| < 2.4$ without gaps. Due to the thickness of the CMS detector components located inside the magnet solenoid, dominated by the HCAL, almost all charged particles able to reach the muon detectors are actually muons. The muon system makes use of three types of gaseous detectors:

- drift tubes (DT) chambers in the barrel region ($|\eta| < 1.2$), where the expected particle rate is lower;
- cathode strip chambers (CSC) in the two endcaps (up to $|\eta| = 2.4$), where the flux of particle through the detector is higher;
- resistive plate chambers (RPC), used in addition to the above mentioned detectors, providing a complementary and independent trigger system up to $|\eta| = 1.6$.

Figure 2.16 Longitudinal layout of one quadrant of the CMS detector. The four DT stations in the barrel (MB1-4, green), the four CSC stations in the endcap (ME1-4, blue), and the RPC stations (red) are shown [74].

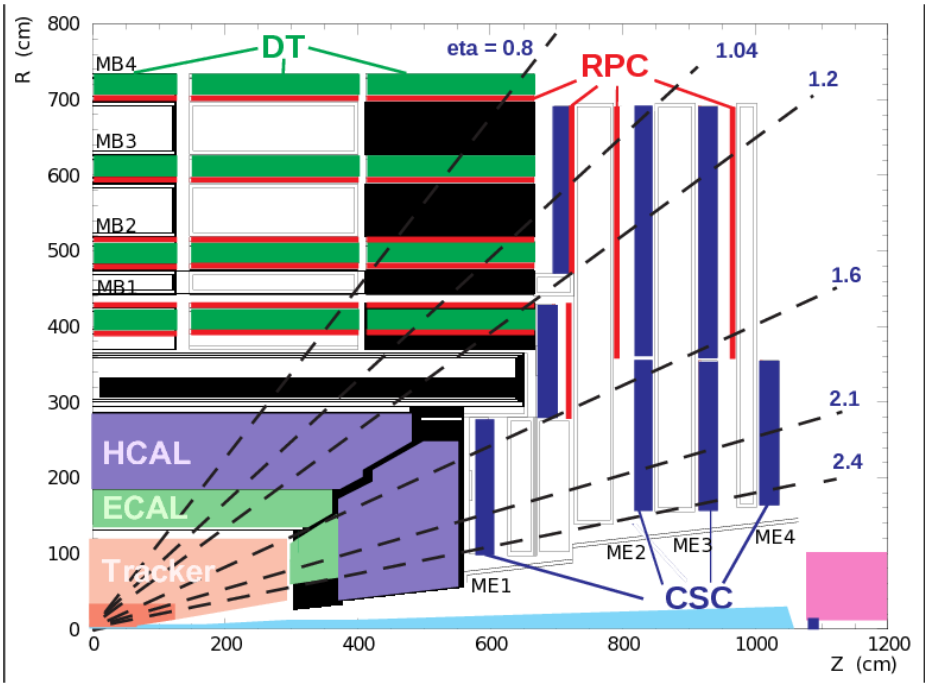


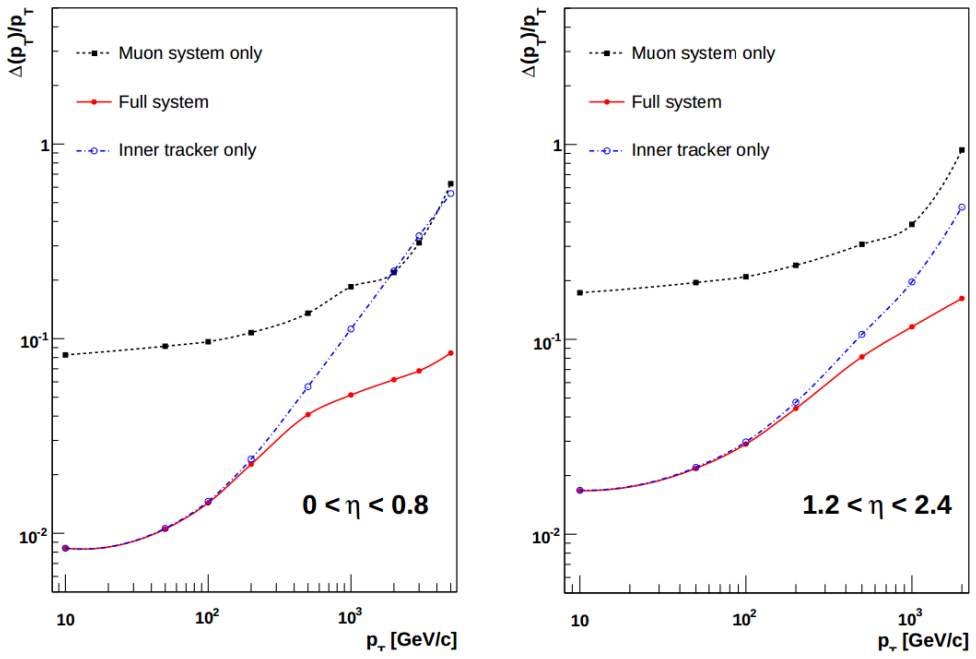
Figure 2.16 shows the different parts of the muon system and their arrangement.

In the barrel region, 250 drift tube chambers are grouped into 4 stations placed inside the magnet return yoke, extending radially from 4.0 to 7.0 m from the beam axis. Chambers in different stations are staggered to avoid dead spots and to allow stand-alone bunch crossing tagging. Each station is designed to measure the muon trajectory with a $r-\phi$ precision not worse than 100 µm in position and about 1 mrad in direction. The muon endcaps comprise a total of 468 CSCs chambers. Every CSC chamber has a trapezoidal shape and consists of six gas gaps, each gap having a plane of radial cathode strips and a plane of anode wires running almost perpendicularly to the strips. Charged particles traversing a plane of a chamber ionise the gas, and the primary ionisation charge is amplified close to the wire to produce an electron avalanche. The signal on the wires is fast and used for triggering, while the image charge induced on a group of cathode strips allows a precise position measurement by determining the centre-of-gravity of the charge distribution. Each CSC measures up to 6 space coordinates (r, ϕ, z). The spatial and angular $r-\phi$ resolution are about 200 µm and 10 mrad, respectively. The DT and CSC subsystems can efficiently trigger on the muon p_T independently of the rest of the detector. A further fast, independent and highly segmented trigger is provided by the RPC system. The RPC double-gap chambers cover the region up to $|\eta| < 1.6$ and provide fast response but coarser position resolution.

Muons produced in proton-proton collisions with sufficient transverse momentum are detected both in the inner silicon tracker and in the outer muon detectors. The muon

transverse momentum can then be measured independently taking into account only the tracker or the muon chambers hits, or combining the information. The p_T resolution as a function of the muon p_T itself is shown in Fig. 2.17, separately for the barrel and the endcap regions: for low p_T muons, the resolution is dominated by multiple scattering occurring in the traversed material, hence the tracker only measurement provides a better resolution. Muons with $p_T > 200\,\text{GeV}$ are much less sensitive to multiple scattering and benefit from the chamber high spatial resolution.

Figure 2.17 Muon transverse momentum resolution, using either the muon system only, the inner tracker only or their combination, in the barrel (left) and endcap (right) regions [56].



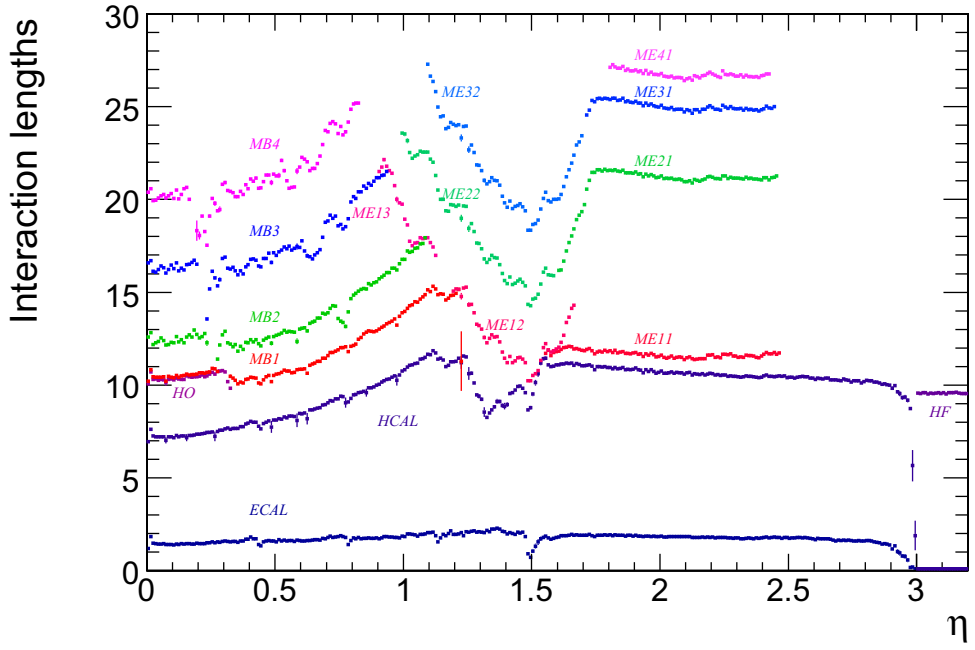
Finally, the amount of material of the CMS detector in terms of interaction lengths after the ECAL, HCAL and each muon station is provided in Fig. 2.18.

2.2.6 Trigger system

The trigger system is responsible for selecting online the small fraction of collision events which are interesting for the CMS physics program. At the LHC design luminosity, in fact, the proton-proton interaction rate is larger than 1 GHz, making it unbearable to store the full data for offline analysis. The CMS trigger system [75] consists of two successive decision levels: the level one (L1) trigger and the high-level trigger (HLT).

The L1 trigger is entirely hardware based, and is used to make a first and quick event selection based on simple quantities like signals from the muon system or energy clusters over a given threshold. A programmable menu selects the events that should be further processed by the high-level trigger. The maximum output rate of the L1 trigger is 100 kHz and the decision has to be taken within the fixed latency of the system, that is

Figure 2.18 Material thickness in interaction lengths after the ECAL, HCAL, and at the depth of each muon station as a function of pseudorapidity [68].



4 µs. First, object candidates like jets, electron and photons, τ leptons and muons are built from the energy deposits in the electromagnetic and hadronic calorimeters and from the hits in the muon chambers. From those candidates, an estimate of the missing transverse energy (E_T^{miss}) is also calculated. Then, the combined event information is evaluated in the so called global trigger. The latter implements a menu of triggers applying selection requirements on the final list of objects, i.e. isolation or topological conditions, close to the requirements applied at HLT level. The CMS global trigger was upgraded after Run 1 to allow an efficient data taking despite the higher collision rates expected in Run 2 [76]. The new global trigger system is based on MicroTCA technology and Field Programmable Gate Arrays (FPGA), making it possible to implement a higher number of L1 trigger paths including more sophisticated calculations, e.g. the invariant mass of two jets, within the trigger latency. The upgraded system took over the full trigger functionality at the beginning of 2016. The number of L1 trigger paths, which was limited to 128 before the upgrade, reached 267 at the end of 2016.

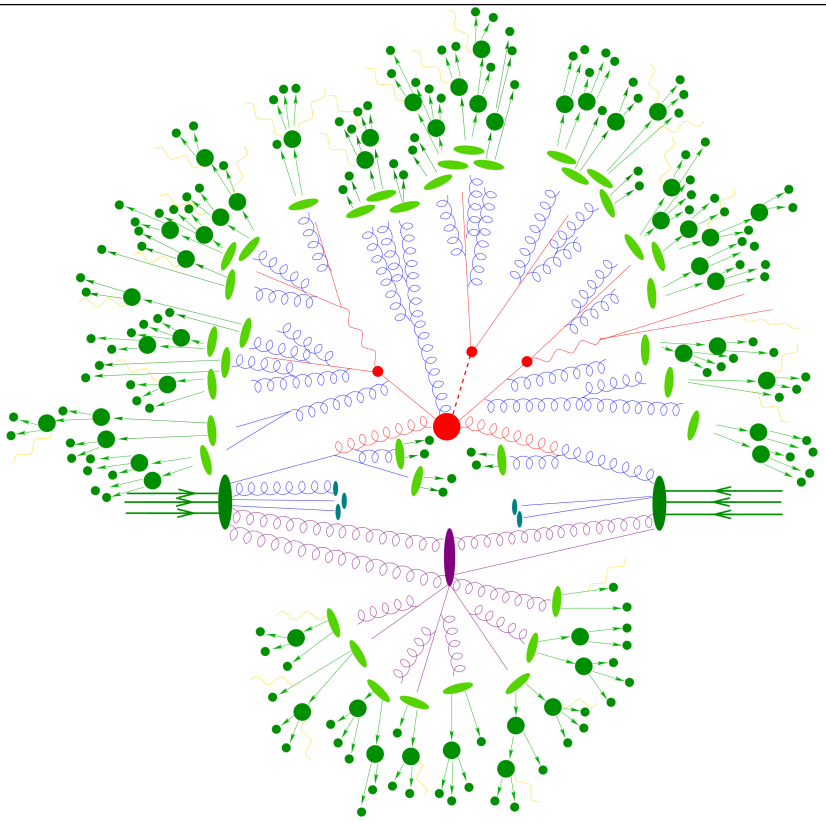
Events retained by the L1 trigger are processed further in the second level of the trigger system, the high-level trigger (HLT). The latter consists of an array of commercial computers and has the task of selecting events to be stored for offline analysis, keeping a maximum output rate of about 1 kHz. Events are processed through a chain of algorithmic steps run in a given order, in which the physics objects are reconstructed and selection criteria are applied. The reconstruction refinement and the selection complexity increase at each step, so that the most computationally intensive algorithms, e.g. track reconstruction, are run only on the few most interesting events. The reconstruction modules and selection filters of the HLT use the same software framework used for

offline reconstruction and analyses, including a lightweight version of the particle flow algorithm described in Chapter 3.

2.3 Event simulation in CMS

A realistic simulation of events for different physics processes is a crucial ingredient to calibrate the detector, design and tune the trigger selection, optimise the physics analyses and perform studies for future searches. As protons are composite objects, a proton-proton collision is a very complex phenomenon. The process of interest typically involves the interaction of two incoming partons occurring at high transverse momentum, called hard scattering. The remaining part of the collision is called underlying event, and includes the interaction of other incoming partons at lower transverse momenta with respect to the hard scattering, as well as the emission of additional particles by the hard scattering partons, either in the initial or in the final state[2]. Figure 2.19 shows a sketch of a $t\bar{t}H$ event, including both the hard scattering process and the underlying event.

Figure 2.19 Sketch of a $t\bar{t}H$ event produced from a pp collision. The hard scattering process is shown in red. The underlying event comprises additional multiparton interactions, shown in purple, and initial and final state radiation drawn in blue. The hadronisation products are shown in green [77].



[2]It is called initial or final state radiation, accordingly

The momentum of the interacting partons is described by their parton distribution functions (PDFs), which represent the probability for a parton of a given flavour to carry a certain longitudinal momentum, when the proton is probed at a given scale. They are non-perturbative functions which are not calculable a priori but extracted from fits to the data, mainly from electron-proton collisions. Thanks to the factorisation theorem [78, 79], any hadronic cross section can be expressed as a convolution of the two incoming partons' PDFs, evaluated at the factorisation scale, and a parton-parton cross section, which can be calculated in perturbation theory as an expansion in the strong coupling constant α_s. The factorisation scale is usually set to the scale of the hard scattering, which generally coincides with the renormalisation scale, i.e. the value at which the running constant α_s is evaluated.

As the vast majority of the calculations can not be performed analytically, results are obtained with numerical integration methods. Event simulation implies the use of Monte Carlo methods, because in quantum mechanics only the probability for a given final state can be calculated. For this reason, simulated events are often called Monte Carlo events. Several programs for event generation are available. Simulated events used for the physics analysis presented in this thesis have been produced by the following generators:

- MADGRAPH5_aMC@NLO [80] can compute the hard scattering process at LO or NLO, with up to four additional real partons included in the matrix element calculation. When used for the more intensive NLO computation, however, the number of additional partons is limited to two. In order to avoid divergences from soft gluon radiation, a p_T cutoff needs to be set for the simulated partons in the final state. For the simulation events used in this thesis, a p_T threshold equal to 20 GeV is used. This generator does not simulate the underlying event, which is instead produced by interfacing it with another software;
- POWHEG [81, 82, 83, 84] computes $2 \to 2$ processes at NLO, with at maximum one additional parton included in the matrix element calculation and is optimised for heavy quark simulations. Again, the underlying event is not simulated by POWHEG but delegated to another software;
- PYTHIA [85] can calculate only tree level $2 \to 2$ matrix elements, while higher order corrections are approximated with the parton shower algorithm. For this reason, it is generally not used as hard scattering event generator but only for the simulation of the parton shower and hadronisation. The emission of partons is ordered in p_T, from the hardest to the softest object.

The hard scattering and the parton shower steps are interconnected and need to be interfaced properly via dedicated jet matching and merging procedures to avoid double counting or uncovered regions of the phase space. The jet matching and merging is implemented differently for LO and for NLO matrix element calculations: for LO simulations, the MLM scheme is adopted [86], in which events with radiated partons with

p_T greater than the ones included in the matrix element calculation are discarded. For NLO simulations, instead, the jet matching and merging follows the prescriptions of the FxFx method [87], allowing to subtract the contribution of identical diagrams appearing at different orders. As a result of this subtraction, some events are assigned negative weights, with the drawback of effectively reducing the statistics of the sample. For this reason, NLO samples are not suitable for the analysis when large statistics is needed. In the specific case of NLO simulations performed with POWHEG instead, the matching procedure is simplified [88] as the latter generator always includes at maximum one radiated parton, and its p_T is used as starting point for the evolution of the parton shower. In this case, there is no overlap between emissions from the matrix element and from the parton shower. Beside the parton shower, the underlying event simulation includes additional partonic interactions other than the hard scattering process. The modelling of multiple partonic interactions in PYTHIA is based on phenomenological models whose parameters need to be tuned to match the observation. The sets of parameters used for the simulated samples of the $H \to \tau^+\tau^-$ analysis (see Chapter 5) is the CUETP8M1 tune [89], derived from combined fits to CMS pp collision data at $\sqrt{s}$ = 7 TeV and CDF $p\bar{p}$ data at lower energy.

In order to obtain a realistic simulation of a LHC proton-proton collision, pileup interactions need to be taken into account as well; they are simulated separately and then overlapped to the event of interest.

After the event simulation is complete, the response of the CMS detector is simulated using the Geant4 software [90], which takes care of propagating the particles through the detector, considering the interaction with the crossed material. Making use of the detector response, the signals of different particles are simulated and the same reconstruction algorithm as in data is applied, allowing for a consistent comparison between data and simulation. A detailed description of the detector geometry as well as of the response of the different detector subsystems are necessary to obtain a simulation being is as close as possible to the data.

Chapter 3

Event reconstruction in CMS

3.1 Particle flow algorithm 53
3.1.1 Fundamental elements 55
3.1.2 Link algorithm 57
3.1.3 Particle reconstruction and identification 57
3.2 Physics objects reconstruction 58
3.2.1 Primary vertices 58
3.2.2 Muons 59
3.2.3 Electrons and photons 61
3.2.4 Charged and neutral hadrons 62
3.2.5 Jets 62
3.2.6 Tau leptons 65
3.2.7 Missing transverse energy 70

This chapter summarises the event reconstruction in CMS, which is based on a global event description achieved thanks to the particle flow approach. The particle flow algorithm described in Sec. 3.1 aims at reconstructing and identifying all stable particles in the event by combining information from all subdetectors. Further reconstruction and identification of the different physics objects starting from the output of the particle flow algorithm is described in Sec. 3.2.

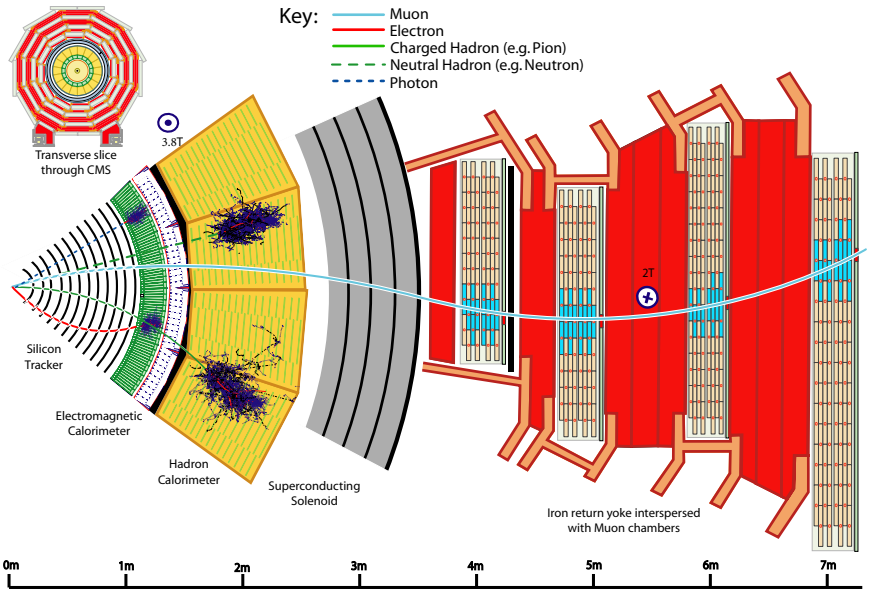
3.1 Particle flow algorithm

For the first time at a hadron collider, particle reconstruction and identification in CMS makes use of the Particle Flow (PF) approach [91], firstly developed and used by the ALEPH experiment at LEP. In the PF approach, a global event description is achieved by connecting the basic elements from the different subdetectors with the aim to identify each final-state particle. The basic idea can be understood thanks to the example of

Fig. 3.1, showing a slice of the CMS detector traversed by different particles: a muon (blue line), an electron (red line), a charged hadron (green line), a neutral hadron (green dotted line) and a photon (green dotted line). Muons are reconstructed by linking a track in the silicon tracker with hits in the muon chambers, while electrons and charged hadrons are characterised by the presence of a track matched to an energy deposit in the ECAL and HCAL respectively. Photons and neutral hadrons instead are identified looking for clusters in the ECAL and HCAL respectively, that are not matched to any track. The main advantage of the PF reconstruction is that particle properties are obtained by combining the measurements from several detector components, allowing to achieve better resolution with respect to the measurement where only one component is used. The CMS detector is well suited to perform a global event description based on the PF approach thanks to its fine segmented tracker, high granularity ECAL and hermetic HCAL, together with a strong magnetic field and an excellent muon system. A simplified version of the PF algorithm used for offline reconstruction is used at the HLT level, allowing to keep the offline and online object reconstructions as close as possible and limiting trigger inefficiencies. The particle-flow event reconstruction proceeds in three steps:

1. the fundamental elements, namely tracks and calorimetric clusters, are reconstructed;
2. the fundamental elements are linked into blocks possibly originated by the same particle;
3. particles are reconstructed and identified from the blocks.

Figure 3.1 A sketch of the specific particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector [91].



3.1.1 Fundamental elements

The fundamental elements of the particle-flow event reconstruction are tracks and calorimetric clusters. They are built using sophisticated algorithms achieving high efficiency and low fake-rate.

Tracks

In the quasi-uniform magnetic field in which the tracker is immersed, charged particles follow helical paths and therefore five parameters are needed to define a trajectory. The extraction of these five parameters requires at least three 3D hits, or two 3D hits and a constraint on the origin of the trajectory based on the assumption that the particle originated near the beam spot. The luminous region is related to the dimensions of proton bunches, which have a Gaussian profile in the transverse plane, with $\sigma_x = 16\ \mu$m and $\sigma_y = 100\ \mu$m, and are approximately 15 cm long. Reconstructing the trajectories of charged particles starting from the collection of hits in the tracker is a computationally challenging task. The tracking software of CMS is based on an iterative structure. Iteration 0, the source of most reconstructed tracks, is designed for tracks originating near the pp interaction point with $p_T > 0.8$ GeV and having three pixel hits. Iteration 1 is used to recover prompt tracks with two pixel hits only. Iteration 2 is configured to find low-p$_T$ prompt tracks. Iterations 3-5 are intended to find tracks originating outside the beam spot and to recover tracks not found in previous iterations. At the beginning of each iteration, hits associated with high-purity tracks found in previous iterations are excluded from consideration. The iterative tracking approach allows to extend the acceptance to tracks down to $p_T = 200$ MeV, well below the calorimeter thresholds. Each iteration proceeds in the four following steps, briefly described below: seed generation, track finding, track fitting and track selection [67].

Seed Generation Seeds define the starting trajectory parameters and associated uncertainties of potential tracks. They are constructed from hits in the pixel tracker and track candidates are built outwards. This approach benefits from the low channel occupancy and 3D spatial measurement in pixel sensors, providing more constraints and better estimates of trajectory parameters. Moreover, generating seeds in the inner tracker leads to a higher reconstruction efficiency, allowing to detect particles not reaching the outer tracker, e.g. electrons losing a significant part of their energy traversing the tracker, pions undergoing nuclear interactions or low-momentum tracks trapped by the strong magnetic field. Seeds are mainly triplets of pixel hits. Nevertheless, other trajectory seeds are needed to compensate for inefficiencies in the pixel detector (from gaps in coverage, non-functioning modules, and saturation of the readout), and partially to reconstruct particles not produced directly at the collision point.

Track Finding It is based on the Kalman filter [92] method. The track parameters are coarsely estimated from the trajectory seeds,and compatible hits from successive tracker layers are added to build track candidates. The track parameters are updated at each layer taking into account the position and uncertainty of the recorded hits and the amount of crossed material, which is used to estimate the effects of multiple Coulomb scattering and energy loss. The standard termination conditions are if a track reaches the end of the tracker or contains too many missing hits or if its momentum drops below a specified value. When the search for hits in the outward direction reveals a minimum number of valid hits, an inward search is initiated for additional hits. The track of a single charged particle can be reconstructed more than once, either starting from different seeds, or when a given seed develops into more than one track candidate. Therefore, after all the track candidates in a given iteration have been found, a cleaning step is performed to remove track duplicates: track pairs sharing more than a given fraction of hits are identified, and the worse reconstructed track, i.e. the one with the fewest hits or, when tracks have the same number of hits, with the largest χ^2, is discarded.

Track Fitting After the track finding step is completed a collection of hits and an estimate of the track parameters is available for each trajectory. In order to optimally exploit the full track information, which is only available at the end of the trajectory, the latter is refitted using a Kalman filter and smoother. This filtering and smoothing procedure takes into account not only the effect of material but also inhomogeneities of the magnetic field especially in the $|\eta| > 1$ region. Estimates of the track trajectory at any other points, such as the point of closest approach to the beam-line, are obtained by extrapolating the trajectory evaluated at the nearest hit to that very point using the same fitting technique. After filtering and smoothing, a search is made for hits incorrectly associated to the track, which might stem from another well-defined track, from nearby tracks, from delta rays or electronic noise.

Calorimetric clusters

Clustering algorithms in the calorimeters aims at detecting and measuring the energy of neutral hadrons and photons and separating them from the charged hadron energy deposits, as well as to reconstruct and identify electrons and accompanying bremsstrahlung photons. Whenever possible, the calorimetric energy measurement is used in combination with the momentum measurement of associated tracks, improving the energy resolution. The PF reconstruction relies on a dedicated clustering algorithm applied separately in the ECAL and HCAL barrel and endcaps and in the two preshower layers. As a first step, cluster seeds are found as cells with the maximum energy deposit with respect to the 8 (in the ECAL) or the 4 (in the HCAL) neighbouring cells. Then, topological clusters are built starting from the seeds and adding cells with at least one corner in common with the seed, and with energy above a given threshold. Finally, single clusters are identified within each topological cluster, with the algorithm described

in [91]. In order to minimise misreconstructed energy excesses and to efficiently identify neutral particles, an accurate calibration [91] of the calorimeters response to hadrons and photons is applied.

Muon tracks

Muon tracking relies on the muon chambers and on the silicon tracker. Three classes of muon candidates are reconstructed:

1. standalone muons, with tracks reconstructed from hits in the muon system only;
2. global muons, with a standalone muon matched to a track in the tracker (outside-in reconstruction). In this case, the two tracks segments are combined and fit, improving the momentum measurement for high-p_T muons;
3. tracker muons, seeded from a track in the tracker which can be linked to at least one muon segment (inside-out reconstruction). This class of muons includes mainly low p_T muons failing the global muon reconstruction due to scattering in the return yoke.

Thanks to the very high efficiency of both the inner tracker and the muon system, about 99 % of the muons produced within the geometrical acceptance of the muon system are reconstructed either as global or tracker muons.

3.1.2 Link algorithm

A given particle is, in general, expected to be related to several particle-flow elements connected to each other by a link algorithm. The linked elements represent the full reconstruction of a single particle through the detector. Starting from a pair of elements in the event, the link algorithm defines a distance between them and evaluates the quality of the link itself. Linked elements are then grouped in blocks. The link algorithm is tentatively performed on each pair of elements in the event and should avoid any possible double counting. Thanks to the high granularity of the CMS detectors, blocks typically contain a few elements only, and are used as input for the particle reconstruction and identification algorithm.

3.1.3 Particle reconstruction and identification

The reconstruction and identification of a set of particles from each block of elements is the core part of the particle-flow algorithm. It provides a list of reconstructed particles to be used in the subsequent physics analysis. For each block built by the link algorithm, particle reconstruction and identification is performed as follows:

- first, each global muon with a momentum compatible with the tracker-only measurement is identified as a particle-flow muon and the corresponding track is removed from the block.
- Then, electrons are identified, taking into account that their tracks tend to be short and that they lose energy by bremsstrahlung in the tracker layers. Tracks are followed all the way to the ECAL and matched with the corresponding clusters in the calorimeter. Once an electron has been identified, the associated track and ECAL clusters are excluded from further processing.
- Remaining tracks in the block are associated to a particle-flow charged hadron, whose momentum and energy are taken from the tracker, assuming the pion mass hypothesis.
- The detection of neutral particles in the block, e.g. photons and neutral hadrons, is done through a comparison between the momentum of the tracks and the corrected energy detected in the calorimeters: if the energies of the closest ECAL and HCAL clusters linked to the tracks are significantly larger than expected from the total momentum of the associated charged particles, the excess is interpreted as a photon and/or a neutral hadron.
- Finally, the remaining ECAL and HCAL clusters not linked to any track are interpreted as photons and neutral hadrons, respectively.

3.2 Physics objects reconstruction

3.2.1 Primary vertices

In a single bunch crossing multiple inelastic proton-proton interactions occur, giving rise to additional vertices and particles in the event. Primary vertices are reconstructed in a cylindrical region centred around the interaction point and developing along the beam axis, with a length of 24 cm and a radius of 2 cm. The primary vertex reconstruction aims at identifying and measuring the position of all proton-proton interaction vertices in the event, and proceeds in three steps:

1. first, prompt tracks originating from the interaction region are selected by imposing requirements on the transverse impact parameter significance with respect to the beam spot, the number of hits in the pixel and strip detectors and the normalised χ^2 of the track fit;
2. then, selected tracks are clustered based on the z-coordinate at the point of closest approach to the beam spot by using a deterministic annealing algorithm [67, 93];
3. finally, the candidate vertices containing at least two tracks are refitted using an adaptive vertex fitter [94] computing the best estimate of the vertex parameters.

In the adaptive fit, each track is assigned a weight between 0 and 1, reflecting the likelihood that it genuinely belongs to the considered vertex. The number of degrees of freedom in the vertex fit is defined as:

$$n_{dof} = -3 + 2 \sum^{tracks} w_i, \tag{3.1}$$

where w_i is the weight of the i-th track and the sum runs over all tracks associated to the vertex. The number of degrees of freedom in the vertex fit is strongly correlated with the number of tracks arising from the interaction region, and can therefore be used to select true proton-proton interactions. The primary vertices are ordered by the quadratic sum of the p_T of the associated tracks. The vertex with the highest value is identified as the hard-scatter vertex, whereas the other ones are considered as pileup vertices.

Pileup mitigation

For charged particles reconstructed within the tracker acceptance, it is possible to distinguish whether they originate from the primary vertex or not. If coming from pileup interactions, such particles are not considered when building physics objects. For photons, neutral hadrons and charged particles outside the tracker acceptance instead, it is not possible to identify the vertex from which they originate. The uniformity of the p_T density of pileup interactions in the (η, ϕ) plane, however, allows to derive corrections based on the average energy contribution per unit area ρ expected from pileup interactions.

3.2.2 Muons

Muons are reconstructed from muon tracks as described above in 3.1.1. In order to reject possible high energetic charged hadrons with showers leaking into the muon system, the information of energy deposits in ECAL, HCAL and HO matched to the muon track is included in the PF muon identification. The desired balance between identification efficiency and purity depending on the physics analysis is achieved by applying a selection based on muon identification variables, e.g. the number of hits in the tracker and in the muon chambers, the χ^2 of the muon track fit and the track impact parameters. A track-kink finder algorithm splits the muon track in the tracker into two tracks at several places along the trajectory, and checks if they are compatible with being a single track. The compatibility of a muon segment with a track is estimated by propagating the track to the muon system and checking how well the muon segments fit to the track in terms of position, direction and number of matched segments. This muon segment compatibility returns values between 0 and 1, with higher values indicating a higher probability of the track being a muon track.

The three most widely used muon identification algorithms in CMS are the following [95, 96]:

- the loose muon identification, designed to be highly efficient ($\epsilon \approx 99\,\%$) for prompt muons as well as from muons from heavy and light quark decays, considers all muons reconstructed by the particle flow algorithm, either as global or tracker muons;
- the medium muon identification, designed to be highly efficient ($\epsilon \approx 98\,\%$) for prompt muons and for muons from heavy quark decays, selects all loose muons having at least 80 % of valid tracker hits, and divides the muons into two classes: the good global muons and the other muons. Good global muons are the ones with normalised global-track $\chi^2 < 3$, position match between the tracker muon and standalone-muon with $\chi^2 < 12$, and track-kink finder $\chi^2 < 20$. The good global muons must have a track-segment compatibility > 0.303 in order to pass the medium muon identification, while the other preselected muons are required to pass a tighter cut on the track-segment compatibility > 0.451. The cuts on the segment compatibility are chosen to obtain an efficiency of 99.5 % for muons from W and Z bosons decays, measured in simulated events. The selection cuts to pass the medium muon identification working point are summarised in Table 3.1. This muon identification working point is used for the event selection of τ lepton pairs in the $\mu\tau_h$ final state of the $H \to \tau^+\tau^-$ analysis, described in Sec. 5.5.1;
- the tight muon identification, aiming at reducing the rate of muons from decays in flight, requires the PF muon to be reconstructed as a global muon, with the normalised track-fit $\chi^2 < 10$, at least one muon-chamber hit included in the global-muon track fit, muon segments in at least two muon stations, track hits in at least five layers of the tracker, including at least one pixel hit, and the track transverse and longitudinal impact parameters $d_{xy} < 2\,\text{mm}$ and $d_0 < 5\,\text{mm}$, respectively, with respect to the primary vertex.

Table 3.1 Selection criteria for the medium working point of the particle-flow muon identification. Muons satisfying preselection cuts and either the global muon criteria or the tight segment compatibility criteria pass the medium working point of the particle-flow muon identification.

Selection stage	Observable	Selection
Preselection	Particle-flow reconstruction	Global or tracker muon
	Fraction of valid tracker hits	> 0.8
Good global muon criteria	Particle-flow reconstruction	Global muon
	Normalised global-track χ^2	< 3
	Tracker-standalone position match	< 12
	Track kink finder	< 20
	Segment compatibility	> 0.303
Tight segment compatibility criteria	Segment compatibility	> 0.451

3.2.3 Electrons and photons

Electron reconstruction is based on the matching of a track to an ECAL supercluster. Electron track reconstruction accounts for the large probability of electrons to radiate while traversing the CMS tracker. For this purpose a Gaussian-sum filter (GSF) [97] is used, allowing for higher and more localised energy loss with respect to the KF. The ECAL supercluster may be composed of several clusters to collect the energy of possible brehmsstrahlung photons, typically covering a small area along η and being elongated in ϕ due to the bending of the conversion electrons in the magnetic field. In oder to form an electron seed, the track momentum and the ECAL cluster energy must be compatible. Electron seeding can be either initiated from the ECAL cluster or from the track. An electron candidate is found when an ECAL cluster and a track can be linked together, provided that the cluster is not linked to three or more tracks. Photon reconstruction is conducted at the same time as electron reconstruction and is seeded from an ECAL supercluster above a given threshold that has no associated GSF tracks. ECAL-initiated electron candidates and photons should not be matched to any HCAL cluster with energy larger than 10 % of the ECAL supercluster. Energy corrections as a function of E and η are applied to the ECAL clusters to account for the energy missed in the process, that is larger at low p_T and where the tracker thickness is largest. The electron energy is then derived from a combination of the GSF track momentum and of the corrected supercluster energy, while the electron direction is the one of the GSF track.

The identification of prompt electrons against those originating from photon conversion, jet misidentification or semileptonic decays of b and c quarks is based on several discriminating variables. The latter can be classified into three groups: observables comparing measurements obtained from the ECAL and the tracker, purely calorimetric observables based on the transverse shape of electromagnetic showers in the ECAL and the energy fractions deposited in the HCAL, and tracking observables exploiting the information obtained from the GSF-fitted track, and the difference between the information from the KF and GSF-fitted tracks.

To maximise the sensitivity of electron identification, several variables are combined using a Boosted Decision Tree (BDT) algorithm. Two types of BDT are defined [98] depending on whether the electron passes the HLT identification requirements ("triggering MVA") or does not ("not-triggering MVA"). Three working points corresponding to identification efficiencies of 90, 80 and 70 % are defined for the electron MVA-ID. The event selection of τ lepton pairs in the $e\tau_h$ final state, described in Sec. 5.5.2, requires the electron candidate to pass the 80 % efficiency working point of the non-triggering MVA-ID algorithm, with a fake rate corresponding to about 2 % in the barrel and 3 % in the endcap regions. Besides the MVA identification algorithms, a simpler cut-based identification approach is also applied, especially to define electron veto conditions [98].

An important source of background to prompt electrons arises from secondary electrons produced in conversions of photons in the tracker material. This background can be

suppressed exploiting the pattern of track hits: the trajectories of prompt electrons start from the beamline, and no missing hits are expected in the inner tracking layers. In addition to that, electrons from photon conversion are also rejected using a fit to the reconstructed electron tracks, performed under the hypothesis that two tracks, consistent with the topology expected from photon conversion, originate from the same vertex.

Photon candidates are retained if they are isolated from tracks and other calorimeter clusters in the event and if the ratio of energy in the ECAL and in the HCAL is compatible with the expectation from a photon shower. Non-isolated and soft photons are considered in the context of jet reconstruction.

3.2.4 Charged and neutral hadrons

The reconstruction of hadrons and non-isolated photons is performed after having identified muons, electrons and isolated photons. In the region of the tracker acceptance $|\eta| < 2.5$, ECAL and HCAL clusters that are not linked to any track are treated as photon and neutral hadron candidates, respectively. Beyond the tracker acceptance, instead, it is not possible to distinguish between charged and neutral hadrons. Therefore, ECAL and HCAL clusters which can be linked are supposed to belong to the same hadronic shower, while ECAL clusters without any match in the HCAL are classified as photons. The remaining HCAL clusters are linked to one or more tracks which can in turn be linked to ECAL clusters and such a topology is interpreted as the signature of charged hadrons. The sum of track momenta is compared to the calibrated calorimetric energy and if the latter is found to be in excess (larger then the expected calorimetric resolution) with respect to the sum of track momenta, such excess is interpreted with the presence of photons and/or neutral hadrons overlapping with the charged hadron.

The charged-hadron momenta are redetermined by combining the measurements in the tracker and in the calorimeter. In the simple case of a single track linked to a HCAL cluster, the combination reduces to a weighted average. Such combination allows to improve the energy resolution and to have a smooth transition between the low and the high energy regimes, where the resolution is driven, respectively, by the tracker and by the calorimeters.

In the rare cases when the calorimetric energy is significantly smaller than the sum of track momenta, a relaxed search for non-isolated muons depositing little energy in the calorimeters is performed. If the mismatch persists it is interpreted as arising from badly reconstructed tracks with large momentum uncertainty.

3.2.5 Jets

Jets are the experimental evidence of quarks and gluons, and are present in almost every event of interest at the LHC. Due to colour confinement, partons can not be observed

directly and their hadronisation originates a collimated shower of particles, the so called jet. Jets are reconstructed by clustering PF particles together and several algorithms can in principle be used. The aim of jet clustering algorithms is to provide an estimate of the jet energy and direction as close as possible to the ones of the particle initiating the shower. Therefore, a successful algorithm should fullfill two main requirements:

- collinear-safety, meaning that the outcome does not change if the energy carried by a single particle is instead distributed between two collinear particles
- infrared-safety, which is the stability of the result of the jet finding against the addition of soft particles.

The default technique for jet clustering in CMS is the anti-k_T algorithm [99], which belongs to a more general class of sequential recombination algorithms. In general, the latter introduce two parameters:

- d_{ij} between the two objects i, j to be clustered, i.e. the energy deposits in the calorimeter crystals;
- d_{iB} between the object i and the beam.

For each pair of objects, these quantities are evaluated. If d_{ij} is found to be smaller than d_{iB}, the objects are combined into a single object, while if $d_{ij} > d_{iB}$ the object i is identified as a jet and removed from the list of objects to be clustered. The procedure is repeated until no more entries are left. The two distance parameter mentioned above are defined as:

$$\begin{aligned} d_{ij} &= \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \\ d_{iB} &= k_{T,i}^{2p}, \end{aligned} \tag{3.2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and $k_{T,i}$, y_i and ϕ_i are, respectively, the transverse momentum, rapidity and azimuthal coordinate of object i. The parameter R defines the maximum (y, ϕ) radius of a jet cone. According to the value chosen for the parameter p, one can have:

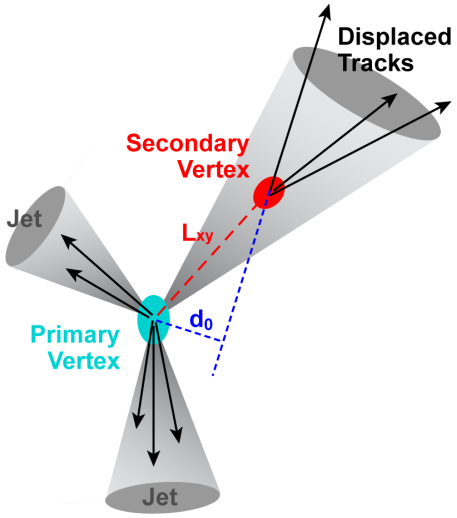
- the k_t algorithm, for p = 1
- the Cambridge/Aachen algorithm, for p = 0
- the anti-k_T algorithm, for p = -1.

The functionality of the anti-k_T algorithm lies in the fact that soft, low p_T objects tend to cluster with hard ones long before they cluster among each other. When a high p_T object has no neighbours in a radius R, then it starts to collect soft objects around it, and the result is a conical jet, provided that there are no overlaps among jets.

Identification of jets from heavy flavour quarks

Jets originating from the hadronisation of heavy flavour quarks are characterised by the presence of mesons and baryons with relatively long lifetimes ($c\tau \sim 500\,\mu m$) producing decay tracks with larger impact parameters and displaced secondary vertices within the jet, as illustrated in Fig. 3.2. B-tagging algorithms exploit such features in the attempt to provide a discriminant indicating of how likely it is for a given jet to originate from a b quark rather than from a quark of another flavour.

Figure 3.2 A simplified illustration of the structure of a b jet [100]. B hadron decays lead to the presence of displaced tracks with large impact parameters with respect to the primary vertex, and they possibly identify a secondary vertex in the jet due to the sizable lifetime of B hadrons.



Several algorithms have been developed in CMS:

- track based algorithms, discriminating using information from single tracks in the jet;
- vertex based algorithms, requiring the presence of a reconstructed secondary vertex in the jet cone and discriminating using its properties;
- combined algorithms combining the two previous approaches.

The most widely used b-tagging algorithm in the analysis of 2016 data is the Combined Secondary Vertex (CSV) algorithm, which includes information of several variables, like the multiplicity of tracks in the jet, the flight distance, the secondary vertex mass, the energy carried by tracks with respect to the total jet energy and the impact parameter significance of the first track with a mass higher than the c quark mass. More details on the CSV algorithm and on other b-tagging algorithms and their performance in 2016 are documented in [101] and [102], respectively.

Jet energy calibration

As the calorimeters response is non-uniform and non-linear, the reconstructed jet energy does not correspond to the total energy of the particles in the jet. Furthermore, electronic noise and pileup bring additional energy in the event. For these reasons, jet energy corrections (JEC) are applied to the reconstructed jet four-momentum, with the aim of relating, on average, the energy measured for the detector jet to the energy of the corresponding true-particle jet. The jet energy is corrected by applying a multiplicative factor to each of the four components of the raw momentum of the jet. The total correction is derived in a multi-stage procedure [103, 104] composed by several factors, each compensating for a specific effect. The following individual correction factors are applied in sequence and in a fixed order:

1. the additional energy coming from pileup interactions and electronics noise is removed. This energy correction is determined from the simulation of a sample of QCD dijet events processed with and without pileup overlay, and is parameterised as a function of the offset energy density, jet area, jet η and p_T;

2. the dependency of the jet response on the jet η and p_T is corrected. Corrections are determined on a QCD dijet sample, by comparing the reconstructed jet p_T to the particle-level one. The corrections are derived as a function of jet p_T and η and make the response uniform over these two variables;

3. residual corrections are applied to the data only, in order to improve the agreement between data and simulation. They are derived from Z+jets and γ+jets events, exploiting the p_T balance of the boson+jets system. The corrections act both on the relative and absolute jet energy scale, and are of the order of a few percent points.

The jet energy scale and resolution performances derived with early 13 TeV data collected in 2015 and 2016 can be found in [105].

3.2.6 Tau leptons

The τ lepton (m_τ =1.777 GeV) [106] is much heavier than the electron and the muon, and can therefore decay into hadrons h. Hadronic τ decays occur in about 65 % of the cases, typically producing one or three charged pions or kaons accompanied by up to two neutral pions π^0 - which decay almost exclusively into two photons - and one tau neutrino ν_τ. In the remaining 35 % of the cases, τ leptons decay to one of the lighter leptons ($\ell = e, \mu$) and two neutrinos. The most likely decay modes of a τ lepton are listed in Table 3.2 together with their branching ratios, and indicating the intermediate meson resonances involved in some of the hadronic decay modes.

Table 3.2 Branching ratios ($\mathcal{B}$) of τ lepton decay modes, to leptons (upper part) and to hadrons and ν_τ (lower part). For hadronic decays proceeding via an intermediate resonance, the resonance mass is also indicated [106, 107].

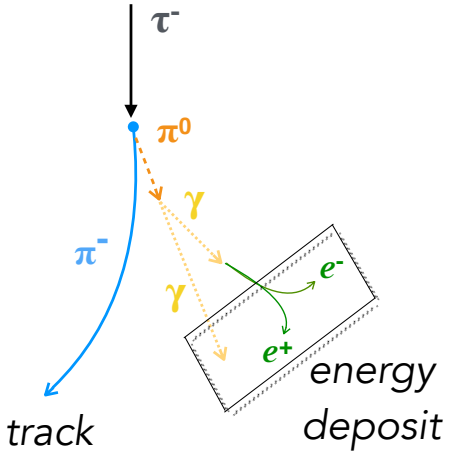
Decay mode	Meson resonance	$\mathcal{B}$ [%]
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other modes with hadrons		3.2
All modes containing hadrons		64.8

Electrons and muons originating from τ decays are difficult to distinguish from prompt electrons and muons produced in the primary proton-proton interaction, and they are reconstructed and identified using the standard PF approach.

The identification and reconstruction of hadronically decaying τ leptons (τ_h) is challenging due to the large multi-jet (QCD) event rate. However, there are handles to discriminate against such QCD events, as τ_h decays are characterised by lower particle multiplicity, narrower energy deposits, and a more isolated topology. At CMS, the dedicated Hadron Plus Strips (HPS) algorithm [107] is used to reconstruct τ_h candidates. The main idea of the HPS algorithm is to identify the charged hadrons - and possibly photons from π^0 decays - expected from τ_h decays, and link them together to build a τ_h candidate. Therefore, the HPS algorithm starts from PF jets and examine their constituents to find the charged hadrons and the photons. The latter have a high probability of converting to electron pairs within the CMS tracker, and such electron pairs are then bent by strong magnetic field, resulting in ECAL clusters with a typical strip shape, narrow along η and elongated along ϕ. A simplified illustration of a hadronic τ decay into one charged pion and one neutral pion is shown in Fig. 3.3.

The reconstruction of $\pi^0 \to \gamma\gamma$ decays consists in identifying such strips via an iterative procedure. It starts from the highest p_T electron or photon and searches for additional e or γ in a window around the previous one. If such an object is found, it is added to the strip and the strip position is updated with the energy-weighted average. The procedure is repeated until there are no more electrons or photons in the search window, and only strips with $p_T > 2.5\,\text{GeV}$ are considered as π^0 candidates. In the most recent version of the HPS algorithm [108] used as of 2015, the size of such window is not anymore fixed but is adjusted dynamically depending on the τ_h and on the strip p_T, avoiding that actual products of the τ_h decay going outside the fixed window boundaries contribute to the τ_h isolation. This can happen for example when a charged pion undergoes nuclear interaction with the tracker material and produces many low p_T particles, which are strongly deviated by the magnetic field, or when photons from π^0 decays undergo multiple conversion and bremsstrahlung. On the contrary, decay products of high p_T

Figure 3.3 A simplified illustration of a hadronic tau decay into one charged pion and one neutral pion. The latter decays to two photons, that may convert to an electron pair. Due to the strong magnetic field, the energy deposit of the electrons tend to be elongated along ϕ.



taus are expected to be collimated around the τ_h flight direction, allowing to use a smaller window size and to reduce background contributions. The size of the window along the η and ϕ directions is dynamically adjusted as a function of the p_T of the strip and of the next e or γ object to be merged. The functions defining the strip size are determined from a single τ gun simulation such that 95 % of all electrons and photons from the τ_h decay are contained inside a strip. The distance between the τ_h and the e/γ as a function of the e/γ p_T in the η (left) and ϕ (right) directions is shown in Fig. 3.4. The dashed lines represent the fit to the 95 % quantile defining the size of the strip.

Hadronic τ candidates are then formed by combining the strips with the charged-particle constituents of the jet with $p_T > 0.5\,\text{GeV}$. In order to mitigate the effect of pileup and of misreconstructed tracks, the distance of closest approach between the tracks of the charged constituents and the hypothetical production vertex of the τ candidate - taken to be the vertex closest to the charged particle of highest p_T within the jet - is required to be less than 0.4 cm in the z direction and less than 0.03 cm in the transverse plane. Multiple τ_h decay mode hypotheses are built, combining one or three charged particles and up to two strips, and the four-momentum of the τ_h candidate in each case is given by the four momentum of the charged particles and strips. The combinations of charged particles and strips considered by the HPS algorithm represent all τ_h decay modes in Table 3.2, except the one having three charged hadrons and one π^0 in the final state because of its contamination by jets. Each hypothesis is then tested for its compatibility with the different decay modes by applying criteria on the τ_h mass, charge and other quantities. As additional requirement, all τ_h constituents should lie inside a signal cone of radius $R = 3.0/p_T[\text{GeV}]$, with upper (lower) limit of 0.1 (0.05). If multiple combinations of charged hadrons and strips pass the mass window and the signal cone requirements, only the hypothesis for the candidate with largest p_T is

retained. The mass of the reconstructed τ_h leptons and the mass of the visible decay products of the $\mu\tau_h$ pairs from $Z \to \tau^+\tau^-$ events in 2016 data and in simulation are shown in Fig. 3.5.

Figure 3.4 Distance between the τ_h and the e/γ as a function of the e/γ p_T in the η (left) and ϕ (right) directions, from simulation. The dashed lines represent the fit to the 95 % quantile for each bin (dotted points), which defines the size of the strip [108].

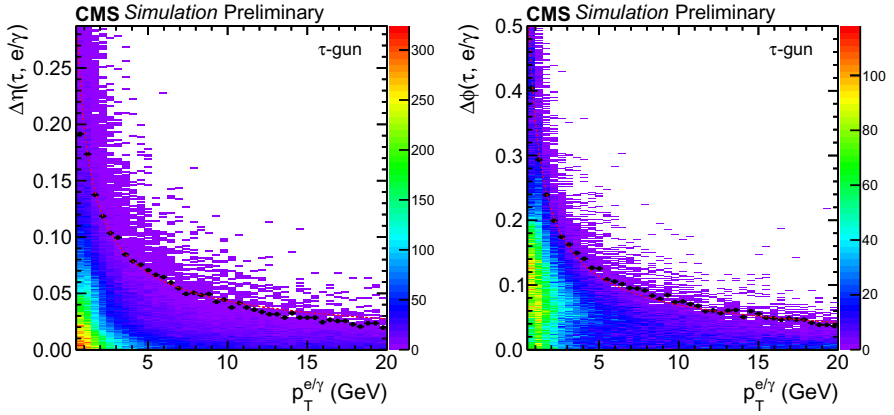
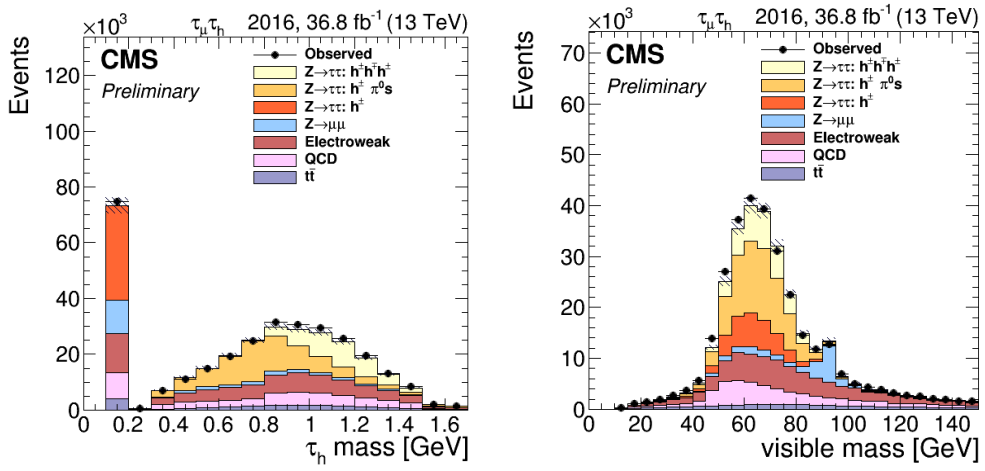


Figure 3.5 Reconstructed visible mass of the τ_h leptons (left) and of the $\mu\tau_h$ system (right) in $Z \to \tau^+\tau^-$ events, in 2016 data and in simulation. Simulated events with genuine τ_h candidates are divided accordingly to their reconstructed decay mode: 3-prong (pale yellow), 1-prong+π^0 (yellow) and 1-prong (orange) [109].

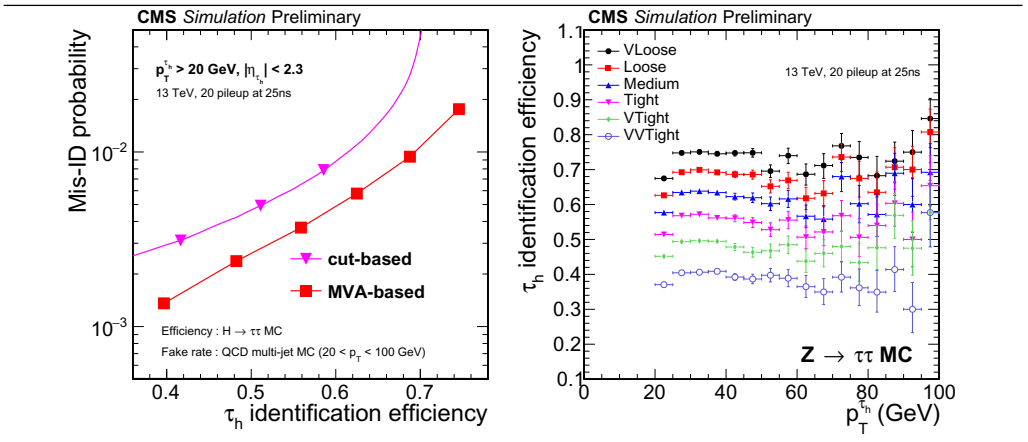


After having reconstructed τ_h candidates, rather strict isolation requirements need to be applied to reject fakes from multi-jet events. The isolation of a τ_h candidate is measured from the p_T sums of charged particles and photons - other than the constituents of the τ_h candidate - found within a cone of size $R = 0.5$ and $R = 0.3$, respectively, centred around the τ_h direction:

$$\begin{aligned} I_\tau &= I_\tau^{\text{charged}} + I_\tau^{\text{neutral}} \\ &= \sum p_T^{\text{charged}}(d_Z < 0.2\,\text{cm}) + \max\left(0, \sum p_T^{\gamma} - \Delta\beta \sum p_T^{\text{charged}}(d_Z > 0.2\,\text{cm})\right). \end{aligned} \tag{3.3}$$

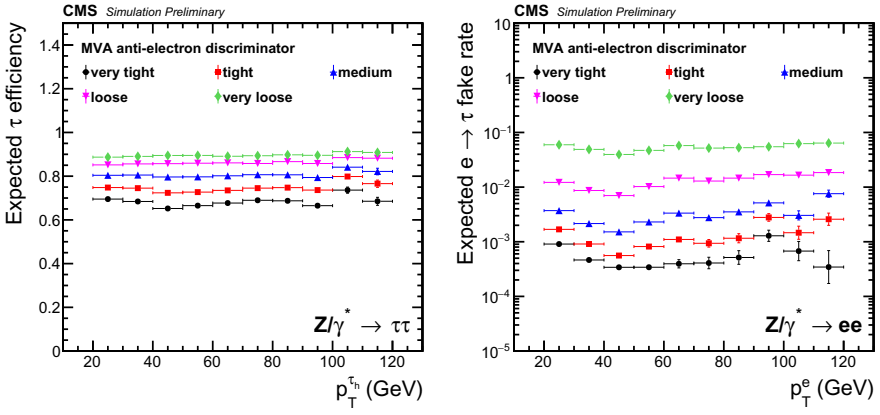
For the charged isolation sum $I_\tau^{charged}$, only tracks with $p_T > 0.5\,\text{GeV}$ and originating from the production vertex of the τ_h candidate, within $d_Z < 0.2\,\text{cm}$, are considered. For the neutral isolation, photons with $p_T > 0.5\,\text{GeV}$ are considered, and a correction for pileup is estimated by subtracting the p_T sum of charged particles with tracks not originating from the τ_h production vertex, i.e. with $d_Z > 0.2\,\text{cm}$, found within a cone of $R = 0.8$ around the τ_h candidate direction, multiplied by the factor $\Delta\beta = 0.2$. The isolation sum of Eqn. 3.3 is used for the cut-based τ_h identification. A more sophisticated approach, instead, consists in combining the τ_h isolation information with other discriminating variables in a multivariate analysis (MVA). In this case, a BDT is trained to distinguish genuine τ_h decays from quark and gluon jets using as input several variables, among which the charged- and neutral-particle isolation sums, the reconstructed τ_h decay mode, the impact parameters of the leading track of the τ_h candidate and their significances, as well as variables related to the tau shape, like the p_T-weighted ΔR, $\Delta\eta$ and $\Delta\phi$ of photons and electrons in strips inside or outside the signal cone with respect to the τ_h direction, and multiplicity (total number of photon and electron candidates in the signal and isolation cones). Different working points of the τ_h MVA isolation, corresponding to different τ_h identification efficiencies, can be defined by cuts on the BDT discriminant. Figure 3.6 (left) shows the misidentification probability - estimated from multi-jet simulated events - as a function of the τ_h identification efficiency - estimated from simulated $H \to \tau^+\tau^-$ events. The cut-based and MVA-based τ_h isolation are compared, showing the better performance of the MVA isolation with respect to the cut-based one for all working points. For a given working point, the threshold on the BDT discriminant is varied as a function of the τ_h candidate p_T to obtain a uniform efficiency as a function of p_T, as can be seen in Fig. 3.6 (right), showing the τ_h identification efficiency as a function of p_T from $Z \to \tau^+\tau^-$ simulated events. Due to its better performance, most of the physics analyses make use of the MVA isolation.

Figure 3.6 Left: misidentification probability jet$\to \tau_h$ as a function of the τ_h identification efficiency estimated from simulation, for different working points of the cut-based and MVA-based isolation discriminators. Right: τ_h identification efficiency for different working points of the MVA-based isolation, from $Z \to \tau^+\tau^-$ simulated events. The working points of the MVA-based discriminator are constructed to have a uniform efficiency as a function of p_T [108].



Not only jets but also isolated electrons and muons have a sizeable - though smaller - probability to be misidentified as τ_h in the 1-prong decay mode. Moreover, if an electron radiates a bremsstrahlung photon that converts, it can also be misidentified as a 1-prong+π^0 decaying τ_h. Two different anti-lepton MVA discriminators [107, 108], one against electrons and one against muons, are employed to suppress the $e \rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ fake rates, respectively. To reject electrons faking hadronic taus, a BDT discriminant is used, based on several variables like the electromagnetic energy fraction, the energy carried by photons, the change in the transverse momentum of the leading track due to bremsstrahlung, as well as mass of the τ_h candidate. Figure 3.7 shows the expected τ_h identification efficiency (left) and $e \rightarrow \tau_h$ fake rate (right) when applying different cuts on the anti-electron MVA discriminator, for τ_h candidates passing the loose working point of the MVA isolation.

Figure 3.7 Efficiency of the τ_h identification (left) and $e \rightarrow \tau_h$ fake rate (right) of the anti-electron discriminator from simulated Drell-Yan events, for τ_h candidates passing the loose MVA-isolation cut [108].



Similarly, another MVA based discriminator is trained to reject muons faking hadronic taus, mostly in the 1-prong decay mode. The variables used in the BDT are the calorimeter energy associated with the leading charged particle of the τ_h candidate, the PF calorimeter energy associated with charged particles or photons constituting the τ_h, the p_T fraction of the hardest charged particle, the number of track segments in the muon system and the number of muon stations with at least one hit within a cone of $\Delta R = 0.5$ around the τ_h direction.

3.2.7 Missing transverse energy

The CMS detector can efficiently detect most particles produced in proton-proton collisions, with the exception of weakly interacting particles, such as neutrinos and hypothetical neutral weakly interacting particles which have not been discovered yet. Due to their low interaction cross section, these particles do not produce any signal in the detector but their presence can be inferred from the momentum imbalance in the transverse plane, i.e. the plane perpendicular to the beam direction. Assuming that the total

transverse momentum of the incoming protons is zero, momentum conservation in the transverse plane implies that the vectorial sum of transverse momenta of all final state particles must be zero as well. The missing transverse momentum $\vec{E}_T^{miss}$ is defined as the negative vectorial sum of all particles in the event, reconstructed by the PF algorithm. Many factors can lead to underestimated or overestimated $\vec{E}_T^{miss}$, e.g. calorimeter thresholds, tracking inefficiencies and non-linear detector responses. In order to reduce the bias on the $\vec{E}_T^{miss}$ measurement, the so-called "type-I" corrections are applied: the p_T of the jets is corrected to the particle level jet p_T using jet energy corrections and such correction is propagated to the $\vec{E}_T^{miss}$ calculation:

$$\vec{E}_T^{miss,corr} = \vec{E}_T^{miss} - \sum_{jets} (\vec{p}_{T,jet}^{\,corr} - \vec{p}_{T,jet}) \tag{3.4}$$

In the type-I corrected $\vec{E}_T^{miss}$, jet energy scale corrections are applied for all jets with $p_T > 15\,\text{GeV}$ and less than 90 % of their energy deposited in the ECAL.

Chapter 4

Alignment of the CMS tracker

4.1 Introduction to track-based alignment 74
4.1.1 Reference systems 75
4.1.2 Alignment algorithms 75
4.1.3 Alignment strategies 78
4.1.4 Alignment validation 80
4.2 Phase 1 upgrade of the pixel detector 88
4.3 Alignment of the upgraded pixel tracker 89
4.3.1 Expected misalignment after the installation 89
4.3.2 Alignment strategy with simulated minimum bias events 91
4.3.3 Realignment studies with cosmic ray events 99
4.3.4 Conclusions on the studies with simulated events 100
4.3.5 Alignment with cosmic ray data 101
4.4 Conclusions 106

In this chapter, the alignment of the CMS tracker based on reconstructed tracks is described in detail. The alignment of the CMS tracker is a crucial part of the detector calibration, and several contributions have been made in the context of this thesis, in the years 2016 and 2017. The installation of a new pixel detector at the beginning of 2017 posed the challenge of deriving the alignment constants on a short timescale. In order to achieve a good performance as soon as possible, several studies with simulated events have been performed, with the aim to develop a successful alignment strategy. The latter has then been applied for deriving the first alignment calibrations performed using cosmic ray data, allowing to correct for the relatively large tracker misalignment after the installation. The first part of this chapter introduces the alignment challenge and the methodologies used in CMS, together with the ultimate performance achieved with the full 2016 dataset. In the second part of the chapter, the original contribution of this thesis to the alignment of the upgraded CMS tracker is described in detail.

4.1 Introduction to track-based alignment

The CMS tracking system has been designed to precisely measure the trajectories and momentum of charged particles, with a design p_T resolution of 1.5 % for muons with $p_T = 100\,\text{GeV}$. This is made possible by the choice of a fully silicon-based tracker, immersed in a strong magnetic field. The CMS tracker is composed by two main subsystems:

- the pixel detector, which is the closest to the interaction region. It is formed by silicon modules segmented in pixels of the size of $100\,\mu\text{m} \times 150\,\mu\text{m}$, providing two-dimensional position information which allows to precisely reconstruct and distinguish the different interaction vertices;
- the strip detector, which surrounds the pixel tracker and is composed by silicon modules segmented in micro-strips. The pitch is $\approx 100\,\mu\text{m}$, depending on the position of the modules within the tracker.

A more detailed description of the CMS tracker is provided in Sec. 2.2.2. The trajectory of a charged particle interacting with the silicon sensors is reconstructed by finding and fitting its set of hits on the tracker modules. The hit position resolution is limited by the typical intrinsic hit resolution of $10 - 30\,\mu\text{m}$ [110], as well as by the knowledge of the tracker geometry, i.e. the positions and orientations of the tracker modules. The assumption of a wrong tracker geometry would introduce biases in the track reconstruction and affect the tracking performance. This has consequences on the physics performance, for example on b-jet identification [111] and on electroweak measurements [112]. In the presence of severe misalignment, hits might even be so far from their assumed position that they can not be associated to the track, causing inefficiencies that propagate also to the higher level physics objects built from tracks. To avoid the hit resolution being limited by the knowledge of the tracker geometry, the latter needs to be measured with an accuracy better than $10\,\mu\text{m}$, i.e. below the intrinsic silicon hit resolution. In order to achieve this, CMS employs track-based alignment techniques. They rely on the fact that tracks passing through the same modules correlate alignment parameters, and that the residuals between the measured and fitted track hits are a measure of the misalignment. Therefore, the tracker geometry can be determined by minimising the quadratic sum of the track-hit residuals. Due to the large number of modules composing the CMS tracker, $O(10^5)$ alignment parameters are necessary to describe the full tracker geometry, and several millions of reconstructed tracks are needed to measure them with the desired accuracy. This makes the tracker alignment task a computational challenge. The track-based alignment is suited to derive the positions and orientations of the tracker modules with respect to each other, but not the absolute position and orientation of the tracker. The latter is determined thanks to survey measurements of the TOB[1], which is the largest and most stable subcomponent of the strip tracker.

[1]Tracker Outer Barrel, see Sec. 2.2.2

4.1.1 Reference systems

Before describing the alignment algorithms employed in CMS, it is convenient to introduce the reference systems used in that context, namely the global and local coordinate systems. The global CMS reference system has the origin at the nominal collision point, the x axis pointing to the centre of the LHC, the y axis pointing up (perpendicular to the LHC plane), and the z axis along the anticlockwise beam direction. The local reference system, instead, is defined for each module, with the origin at the geometric centre of the active area of the module, the x axis defined along the more precisely measured coordinate of the module, the y axis orthogonal to the x axis and in the module plane, pointing away from the readout electronics, and the z axis normal to the module plane. The local coordinates are conventionally indicated with $x^{'}, y^{'}, z^{'}$ and the angles measured with respect to each of these axes are indicated with α, β, γ, respectively[2].

4.1.2 Alignment algorithms

Track-based alignment algorithms rely on the fact that a wrong assumption of the tracker geometry in the track reconstruction generally leads to an increase of the track-hit residuals, i.e. the spread of the measured and the fitted hit positions. Figure 4.1 shows a pictorial representation of this idea. The modules positions and orientations can thus be obtained by minimising the sum of squares of normalised residuals from a large sample of reconstructed tracks. Under the assumption that the hit measurements are independent, the quantity to minimise is:

$$\chi^2(\mathbf{p}, \mathbf{q}) = \sum_{j}^{tracks} \sum_{i}^{hits} \left(\frac{m_{ij} - f_{ij}(\mathbf{p}, \mathbf{q}_j)}{\sigma_{ij}} \right)^2 , \tag{4.1}$$

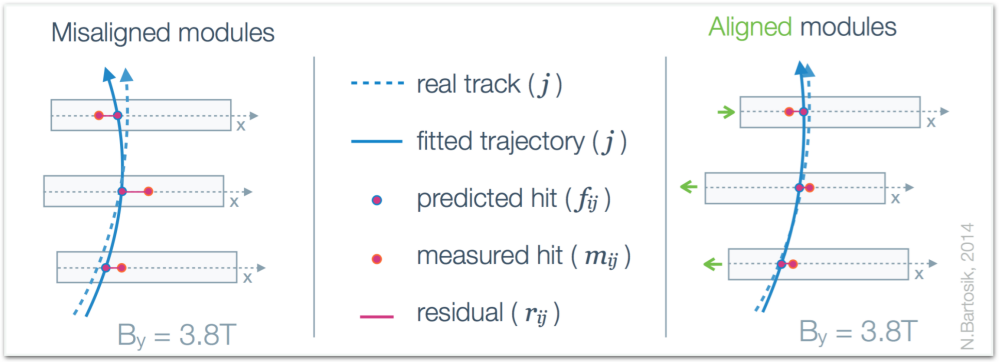
where m_{ij} are the measured hit positions with uncertainties σ_{ij}, and f_{ij} are the predicted hit positions from the track fit, which depend on the geometry $\mathbf{p}$ and on the track parameters $\mathbf{q_j}$. The latter are valid only for a given trajectory j and are therefore called local parameters. Conversely, the global parameters $\mathbf{p}$ describe the tracker geometry and contribute to all track fits. The aim of the alignment procedure is to determine the global parameters, thus also referred to as alignment parameters.

An initial description of the geometry $\mathbf{p}_0$, which is available from design drawings, survey measurements or previous alignment results, is used to determine a first estimate of the track parameters $\mathbf{q}_{j0}$. Assuming the corrections to the initial geometry being small, the term f_{ij} in Eqn. 4.1 can be approximated with a linear expansion around the initial values, and minimising the χ^2 is equivalent to solve a linear equation system of the form:

$$\mathbf{Ca} = \mathbf{b}, \tag{4.2}$$

[2]in some of the cited references, $(x^{'}, y^{'}, z^{'})$ are also called (u, v, w)

Figure 4.1 Schematic view of the fitted track prediction depending on the alignment of tracker modules. Locally measured hit positions are constant, but their global positions depend on the alignment of the corresponding modules. Misaligned modules pull the measured hit positions away from the real trajectory, leading to an increase of the track-hit residuals and to a distortion of the trajectory [113].



where $\mathbf{a}^T = (\Delta\mathbf{p}, \Delta\mathbf{q})$ represents the corrections to the geometry and to the track parameters of all used tracks, $\mathbf{C}$ is a symmetric n-by-n matrix, and $\mathbf{b}$ is a vector of size n, with $n = p + q$ being the number of local plus global parameters [114]. If the alignment corrections are not small, the linear approximation has a limited precision and the procedure needs to be iterated. The track parameters needed to describe a trajectory in a homogeneous magnetic field are five, in the absence of interactions with material. To include multiple scattering effects through thin scatterers, two additional parameters are needed for each traversed layer, leading to a large number of parameters ($n_{par} > 50$) especially in the case of cosmic ray tracks traversing the whole detector. The linear equation system of Eqn. 4.2 is solved with a global fit approach [114] implemented in the Millepede II software [115, 116]. The core idea is to take advantage from the special structure of the matrix $\mathbf{C}$: applying block matrix algebra, the large system in Eqn. 4.2 can be reduced to a smaller one:

$$\mathbf{C}'\Delta\mathbf{p} = \mathbf{b}', \tag{4.3}$$

which contains alignment parameters only, without losing information for the geometry determination. For a sample of 10^7 tracks with 20 parameters and 10^5 alignment parameters to be determined, the number of matrix elements is reduced by more than 4×10^6 [117]. The terms $\mathbf{C}'$ and $\mathbf{b}'$ sum contributions from each track. To solve the system of Eqn. 4.3, the solutions of the track fit equations $\mathbf{C}_j\Delta\mathbf{q}_j = \mathbf{b}_j$ are needed, as well as the full covariance matrix $\mathbf{C}_j^{-1}$. In the Millepede II software the latter is obtained using the General Broken Lines (GBL) [118] track refit, which starts from the direction and position of the first track hit, as reconstructed by the standard Kalman filter track, and propagates it through the silicon planes, taking into account magnetic field inhomogeneities, energy losses, and multiple scattering. The track parametrisation used in the GBL algorithm is such that most of the track parameters influence only a small part of the track, and this allows to significantly reduce the computation time.

As result of the χ^2 minimisation, the alignment parameters $\mathbf{p}$ are obtained. They

correspond to nine parameters per module, namely three shifts and three rotations in the local reference frame, as well as three parameters to describe the deviation of the sensor shape from a flat plane. This is particularly important for tracks with large angles of incidence relative to the module normal, as they are highly sensitive to variations in the local-z direction.

Weak modes

A major difficulty in the alignment algorithm arises from coherent changes of the alignment parameters $\Delta\mathbf{p}$ which can be compensated by changing the track parameters $\Delta\mathbf{q}$. In fact, when approximating the track model f_{ij} with a linear expansion around its initial values, linear combinations of alignment parameters can exist which leave the track-hit residuals, and hence the χ^2 of Eqn. 4.1 (almost) unchanged[3]. These linear combinations are called "weak modes". The kind of possible weak modes depends on the tracker geometry and granularity, the topology of the tracks used for alignment, and on the alignment and track parameterisation. As the highly segmented CMS tracker has a cylindrical symmetry and is immersed in a solenoidal magnetic field, weak modes of track-based alignments performed using only minimum bias tracks can be classified in cylindrical coordinates, i.e. by changes of the modules position Δr, Δz and $\Delta\phi$, as a function of r, z and ϕ [117]. The solution for reducing the presence of weak modes is to include track topologies which bring additional information in Eqn. 4.1. In particular, muon tracks from cosmic rays and from $Z \rightarrow \mu^+\mu^-$ decays are considered. Cosmic muon tracks originate outside the detector and can traverse two opposite halves of the tracker, allowing to correlate movements of distant modules. This feature is helpful to avoid the so called telescope weak mode, in which modules are shifted along the z axis proportionally to their distance from the beam line r, i.e. $\Delta z \propto r$. Figure 4.2 shows how cosmic ray tracks are sensitive to the telescope weak mode: such a deformation would introduce a kink in the track, increasing the distance between the predicted and the measured track hits, and thus the χ^2 of Eqn. 4.1. Conversely, collision tracks can still be fitted without increasing the track-hit residuals.

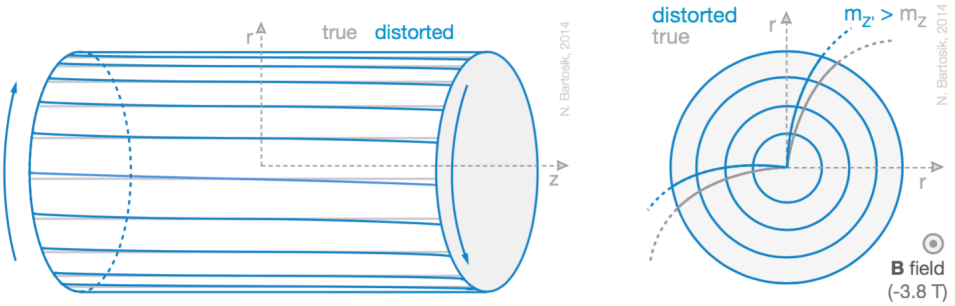
Tracks of muon pairs originating from Z bosons decays allow to introduce an additional constraint in the alignment fit, as the muon pair invariant mass must be compatible with the Z boson mass. For this reason, such tracks are sensitive to the so called twist weak mode, in which modules are rotated along the beam line direction ϕ depending on their longitudinal position z, i.e. $\Delta\phi \propto z$. Figure 4.3 illustrates that the twist deformation introduces a bias in the track momentum. This can be reduced requiring the two tracks being compatible with the Z boson decay, by replacing the two independent track parameterisations with a two-body-decay parameterisation and imposing a constraint on the reconstructed Z boson mass.

[3]From a mathematical point of view, it means that the matrix containing the alignment parameters is numerically (close to) singular.

Figure 4.2 Illustration of the telescope weak mode, in which $\Delta z \propto r$. The real tracker geometry (grey) is compared with the distorted one (in blue), in the rz plane. Collision tracks can be fitted in both geometries without increasing the track-hit residuals (left). This is not the case for cosmic muon tracks, for which the telescope distortion would introduce a kink in the trajectory (right) [113].



Figure 4.3 Illustration of the twist weak mode, in which $\Delta\phi \propto z$. The real tracker geometry (grey) is compared with the distorted one (in blue) (left). Tracks can be reconstructed in the distorted geometry without increasing the track-hit residuals by changing the track curvature, and hence the track momentum (right). Track pairs originating from Z boson decays are sensitive to such deformation, as their invariant mass should be compatible with the one of a Z boson [113].



4.1.3 Alignment strategies

The CMS tracker is structured in a hierarchical way, with a few large mechanical structures composed by smaller parts, the smallest unit being always the single module. A full overview of the structure hierarchies of the CMS tracker is shown in Fig. 4.4. The mechanical structure of the tracker, and with it the possible physical movements of its substructures, is taken into account by the alignment algorithm, which allows treating translations and rotations of the substructures by introducing six alignment parameters for each of them. This adds redundant degrees of freedom, since the movements of the large structures can be equivalently expressed either by their alignment parameters or by the parameters of their components. The redundancy is removed by linear equality constraints, imposed on the original equation system by using Lagrange multipliers. The possibility of taking into account larger structures in the alignment algorithm is very advantageous, especially when the track sample is not large enough to derive the alignment parameters independently for each module, i.e. to perform a module level alignment. In this case, a high level structures alignment can be performed to

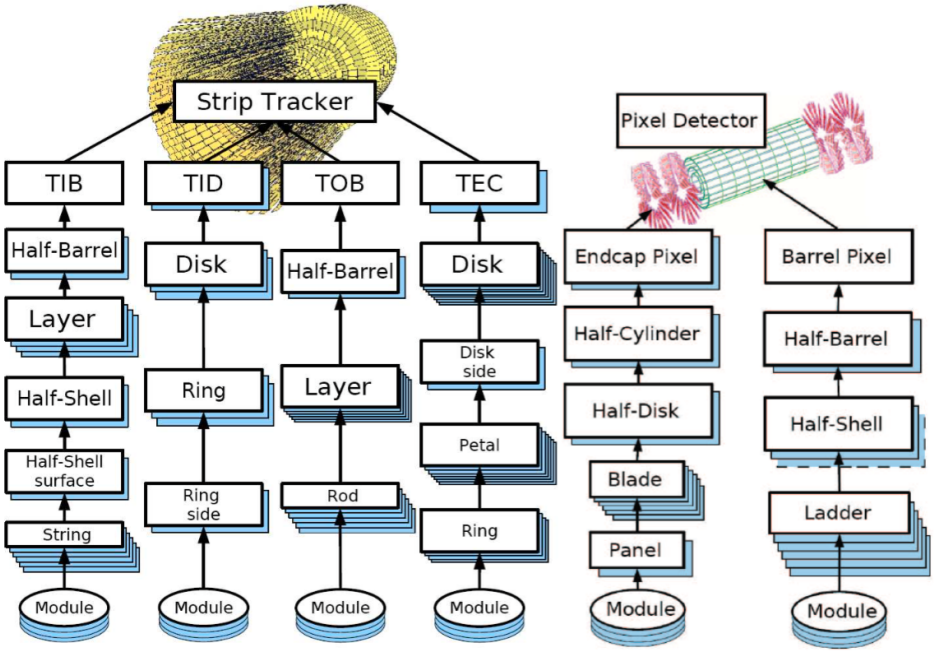
determine the positions and orientations of the selected large structures and propagate it to their subcomponents. The alignment workflow also allows to introduce time dependent and independent parameters simultaneously. The time-dependent parameters are valid only in a given "interval of validity" (IOV), and are determined using only tracks recorded in that time window, while the time independent parameters use the full track sample. This approach, called differential alignment, is often combined with the hierarchical alignment, such that the time dependency is taken into account for the high level structures, while the positions of the single modules with respect to their high level structure is derived using the full statistical power of the dataset.

Different track topologies can be used simultaneously for track-based alignment:

- cosmic ray tracks, which can be recorded either with or without magnetic field;
- collision tracks from minimum bias events;
- isolated muon tracks with high p_T, from W boson decays;
- tracks of muon pairs produced in the decay of a resonance, typically from $Z \rightarrow \mu^+\mu^-$ events.

All tracks used for alignment are required to be reconstructed from hits with a good signal-to-noise ratio, which should be larger than 12 for strip hits, and the angle between the module surface and the track should not be smaller than 5°. Moreover, at least two of the track hits must provide a precise 2D measurement, i.e. they are coming either from

Figure 4.4 Hierarchical structure of the CMS tracker, for the strip (left) [119] and for the pixel detectors [120].



pixel or from strip stereo modules. The track transverse impact parameter is required to be less than 50 cm. In addition, other selection cuts are applied depending on the track topology, as summarised in Table 4.1. Muon tracks must be associated to global muon candidates and have at least one pixel hit. The muons must be isolated, with the sum of the p_T from charged tracks and of the electromagnetic and hadronic transverse energies measured in a cone of radius $R = 0.3$ not exceeding 15 % of the muon p_T. The selection of muon tracks from $Z \rightarrow \mu^+\mu^-$ events requires two oppositely charged muons, with an invariant mass between 85.8 and 95.8 GeV. Finally, tracks are not considered in the final fit if their χ^2 value exceeds the 99.87 % quantile (corresponding to three standard deviations) of the χ^2 distribution with a number of degrees of freedom equal to the one of the track.

Table 4.1 Selection criteria for tracks used in the alignment procedure. For isolated muon and di-muon events, additional requirements are applied, as detailed in the text.

Track topology	N(hits)	$\lvert\eta\rvert$	p	p_T	N(tracks)/event
Cosmic ray track	≥ 8	n.a.	> 4 GeV	n.a.	≤ 1
Collision track	≥ 10	< 2.6	> 8 GeV	> 1 GeV	n.a.
Isolated muon track	≥ 10	< 2.5	> 8 GeV	> 5 GeV	≤ 1
Di-muon event track	≥ 10	< 2.5	> 8 GeV	> 15 GeV	2

Alignment of the tracker based solely on cosmic ray tracks is usually performed before the start of a new pp collision run, in order to derive the first alignment corrections to be applied. As soon as collision events are recorded, more track topologies are included in the alignment fit. Tracks from minimum bias events can be collected at high rates, and provide a large dataset to derive alignment parameters at higher granularity, e.g. at the level of single modules. Isolated muons from W boson decays provide well measured high-p_T tracks, while muon tracks from cosmic rays and from $Z \rightarrow \mu^+\mu^-$ decays are important to reduce the presence of weak modes, as explained above.

In summary, several aspects need to be determined to define an alignment strategy, namely which tracker substructures are to be aligned and at which granularity, both in terms of alignment parameters and of time dependency, as well as the track topologies to use. In most of the cases it is not possible to determine a priori what is the best strategy to follow, but different alignments are investigated in parallel, and their performance is compared with the means of the alignment validation, described in the following section, to identify the most successful one.

4.1.4 Alignment validation

The validation of the alignment result consists of evaluating the achieved accuracy and investigating the presence of systematic misalignment effects. Several observables are used for these purposes, as described in the following. The description is complemented

with concrete examples, showing the results of the alignment performed using the full 2016 dataset. This was the largest alignment campaign ever performed at CMS: corrections were derived at the module level for the whole tracker, using a sample of about 100 million reconstructed tracks, both from cosmic rays and from collision events. The alignment parameters for the high level structures were treated as time-dependent, and derived separately for 36 time intervals. Several iterations of the alignment algorithm were performed to ensure the accuracy of the result.

Comparison of tracker geometries

The comparison of tracker geometries is the first way to understand the alignment results. The geometry comparison is performed by measuring the difference of the module positions and orientations according to two different sets of the alignment parameters. By comparing the modules positions before and after the alignment corrections, for example, it is possible to identify the movements in the different detector substructures and to evaluate their magnitude. This information is very important to understand features that might appear in other validation methods. Figure 4.5 shows the comparison of the tracker geometries before and after performing the 2016 end-of-year alignment. Every dot represents a module, and the colour indicates the tracker subsystem to which it belongs. "Dead" modules which do not yield any signal, and thus can not be aligned, are represented with empty circles. Every plot of Fig. 4.5 represents the modules positions differences projected along a given direction, e.g. $\Delta r, \Delta z, r\Delta\phi$, as a function of the modules positions along another direction, e.g. r, z, ϕ.

Median of track-hit residuals

The statistical accuracy of the alignment result, as well as the presence of systematic biases, can be evaluated from the distribution of the median of the track-hit residuals [117]. Such a distribution is obtained as follows: each track is refitted using the alignment parameters under study, and the hit prediction for each module is obtained from all of the other track hits, not including the hit under consideration. Then, the median of the distribution of track-hit residuals is taken for each module and used to fill a histogram. The median is chosen because it is relatively robust against stochastic effects from multiple scattering. The width of this distribution of the medians of residuals (DMR) is therefore taken as a measure of the local precision of the alignment results; if the mean of the distribution deviates from zero it indicates possible biases. The DMR distributions are produced separately for modules belonging to different tracker subsystems. For pixel modules providing a 2D hit measurement, the track-hit residuals are projected along the x' and y' directions, while for strip modules only the DMR distribution along the x' direction is considered. In principle, any track topology could be used to produce the DMR distributions. In practice, however, minimum

Figure 4.5 Difference of the modules positions after the dedicated alignment at the end of 2016, with respect to the modules positions in data taking. Each point represents a module. Dead modules are indicated with empty circles, they are not aligned at module level but can change position following the alignment of the higher-level mechanical structure they belong to. The colours correspond to the different tracker sub-detectors (see Fig. 2.10). The different plots show on the x-axis the coordinate r, ϕ and z and on the y-axis the module position differences Δr, $r\Delta\phi$ and Δz [121].

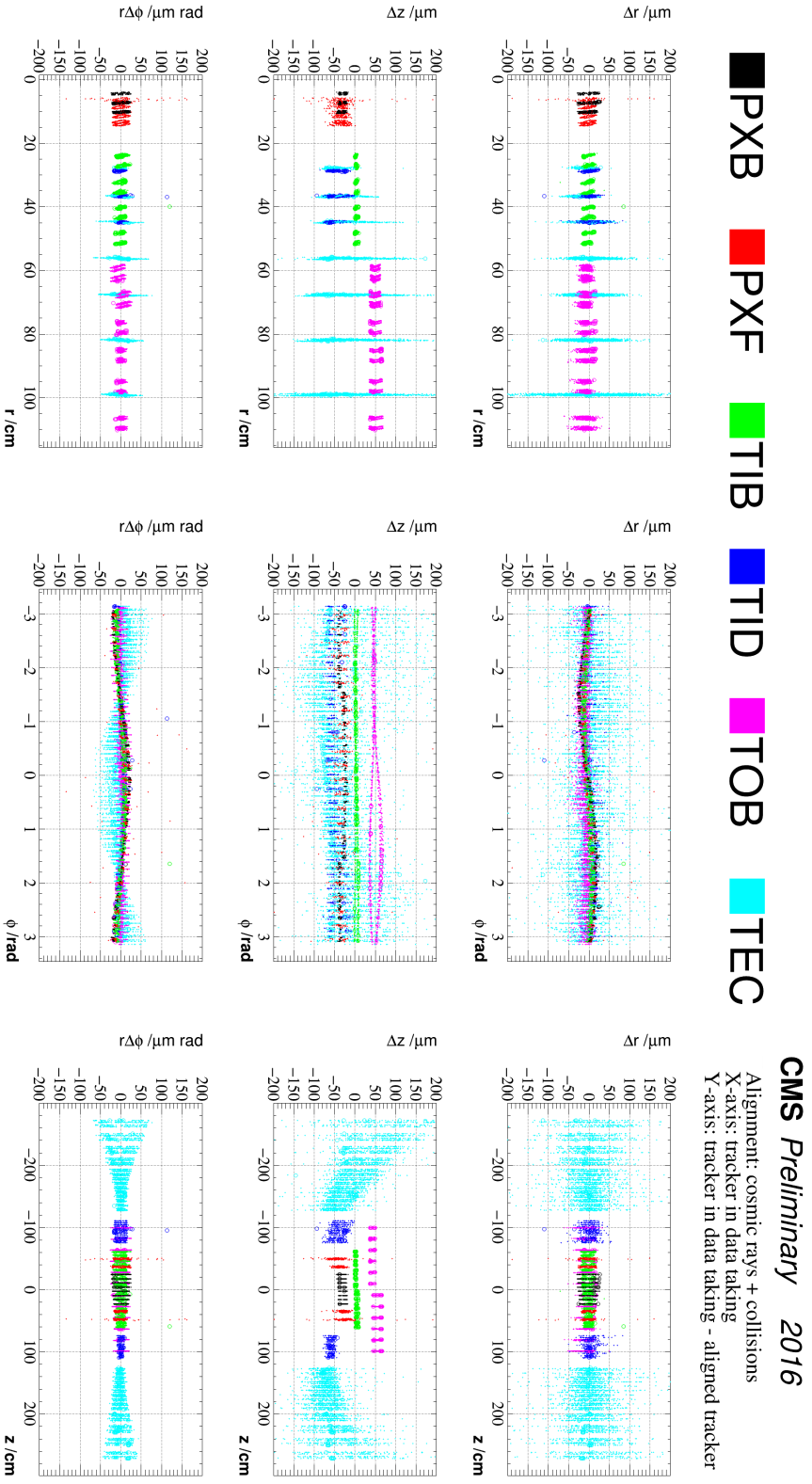
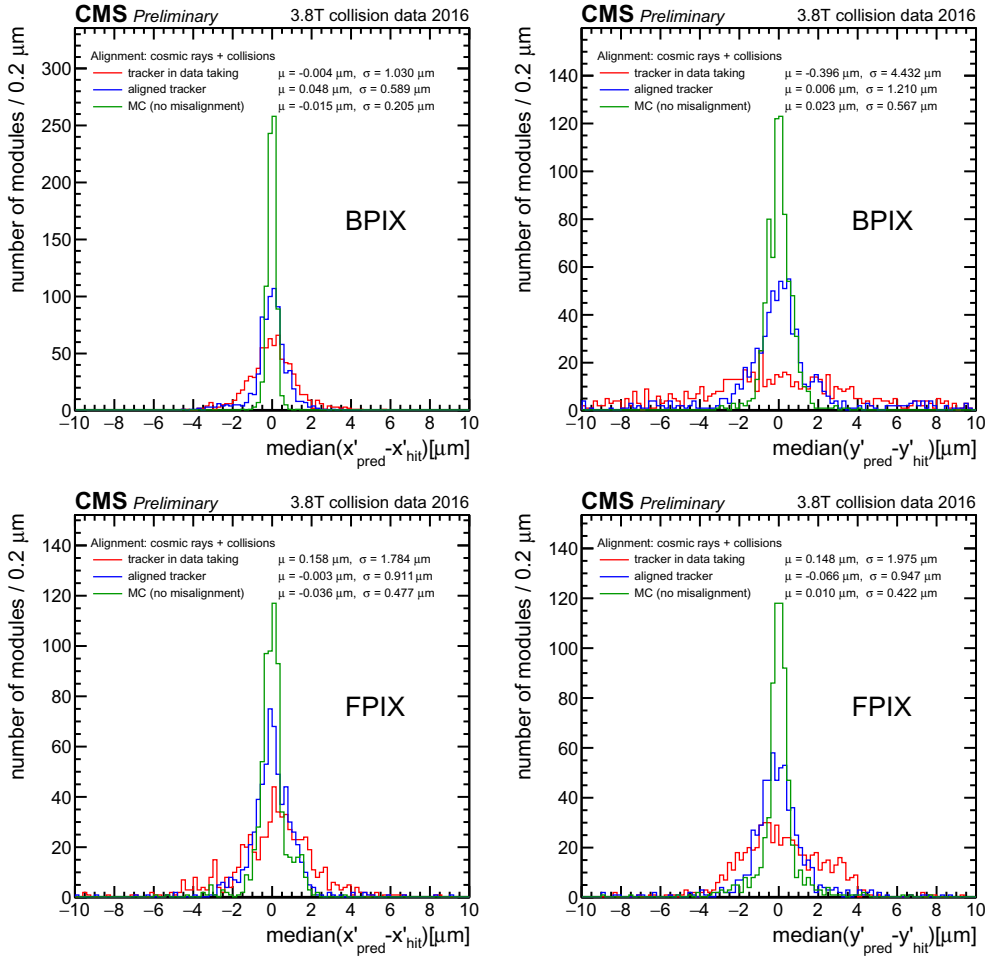
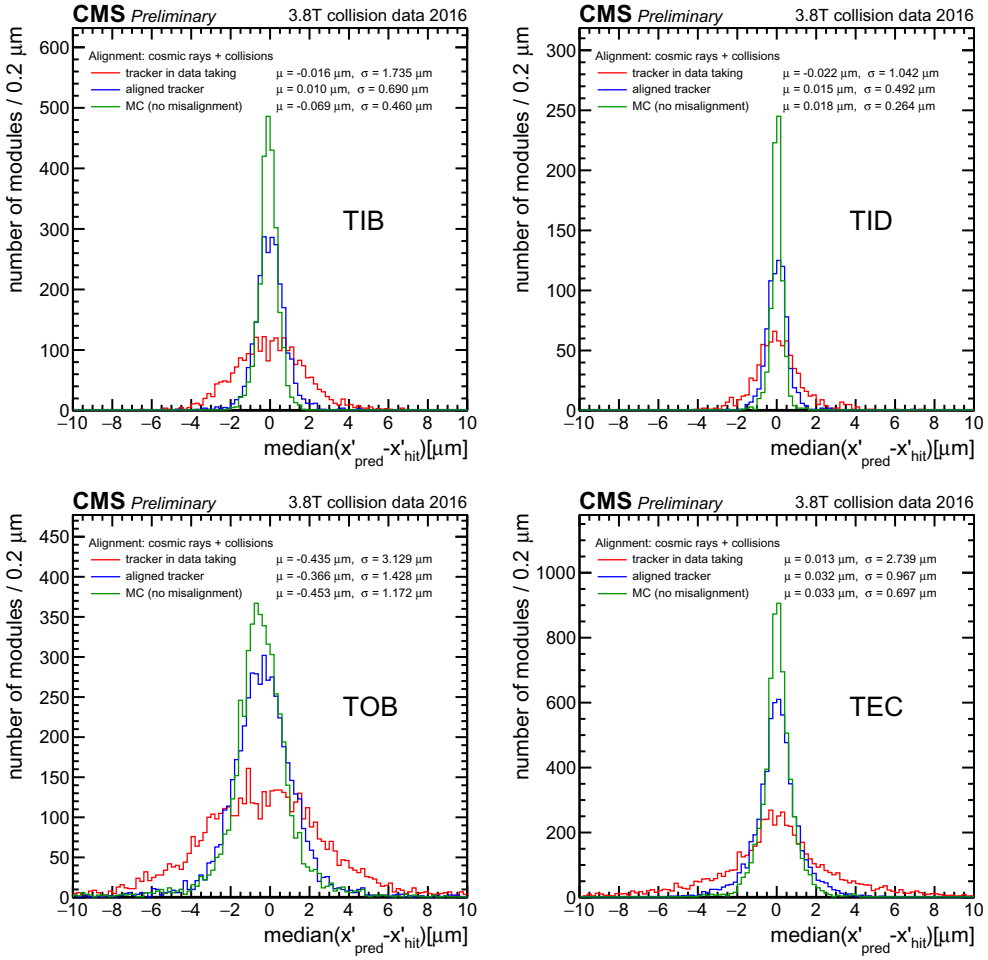


Figure 4.6 The distribution of median residuals along the x' and y' directions for the pixel detector (BPIX and FPIX). Distributions are produced using 1 million collision events collected with the magnetic field at 3.8 T. The red line shows the geometry used in data taking during 2016, while the alignment shown in blue is obtained after the end of the 2016 data taking using 3.8 T cosmic ray and collision data from the whole 2016 pp run. As a reference, the green line shows the distribution obtained from a Monte Carlo simulation, where no misalignment is applied [121].



bias events are preferred because of the large number of events available. As the median of the track-hit residuals is strongly correlated with the quantity minimised by the alignment algorithm (in Eqn. 4.1), alignment and validation should preferably be performed using independent sets of tracks. Finally, when comparing the performance of different alignment results, it is important to know that the DMR distribution has an intrinsic width component due to the limited number of tracks used to produce it, and therefore distributions can only be compared if they are produced with the same number of tracks. Figures 4.6 and 4.7 represent the DMR distributions for the different tracker subsystems, showing the improvement of the dedicated alignment run at the end of 2016 with respect to the alignment used in data taking. The comparison with simulated events where no misalignment is applied (ideal case) quantifies the purely statistical width of the DMR distributions. The precision achieved by the end-of-year

Figure 4.7 The distribution of median residuals along the x' direction for the strip detector subcomponents (TIB, TID, TOB and TEC). Distributions are produced using 1 million collision events collected with the magnetic field at 3.8 T. The red line shows the geometry used in data taking during 2016, while the alignment shown in blue is obtained after the end of the 2016 data taking using 3.8 T cosmic ray and collision data from the whole 2016 pp run. As a reference, the green line shows the distribution obtained from a Monte Carlo simulation, where no misalignment is applied [121].



2016 alignment is of the order of 1 μm for all subdetectors, which is very close to the ideal case.

Track-vertex residuals

The primary vertex residual validation [117] is based on the study of the distance between the track and the vertex, when the latter is reconstructed without taking into account the track under consideration, to avoid biases. Collision events used for the primary vertex validation are recorded using minimum bias triggers and the following selection criteria are applied:

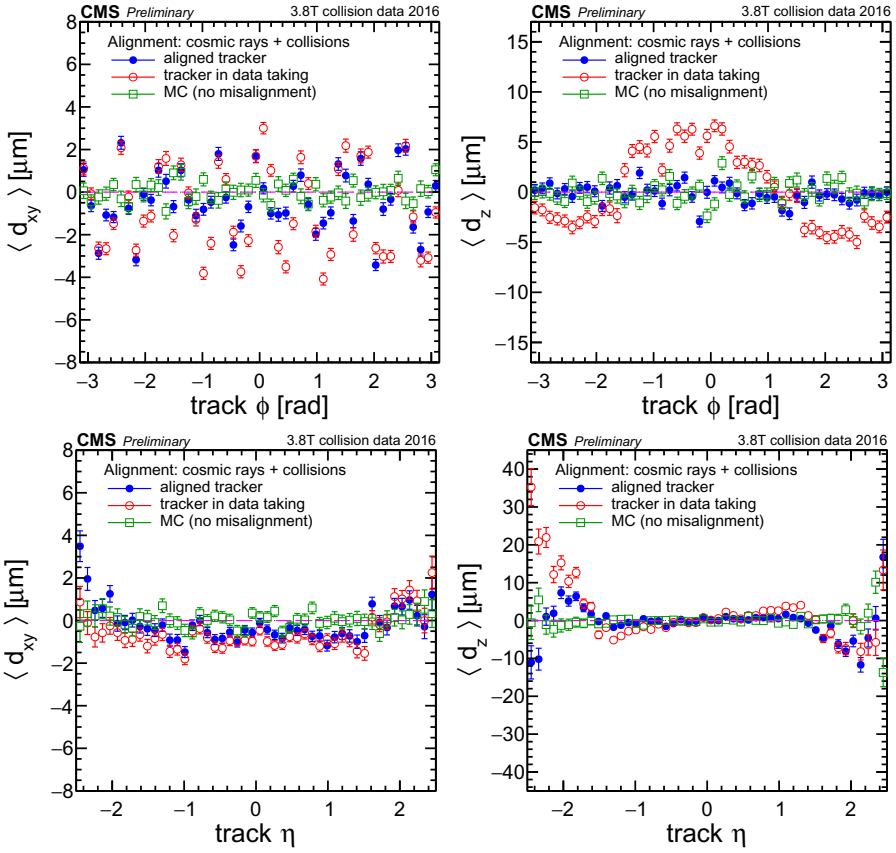
- the vertex fit must have at least 4 degrees of freedom, i.e.the vertex must be reconstructed from at least four tracks (see Sec. 3.2.1);
- for each of the vertices, the impact parameters are measured for tracks with more than six hits in the tracker, of which at least two are in the pixel detector and at least one in the first layer of the barrel pixel or the first disk of the forward pixel tracker;
- the normalised χ^2 for the track fit smaller than 5.

The track-vertex residuals, projected on the transverse plane (d_{xy}) or along the longitudinal direction (d_z), can be studied as a function of the track pseudorapidity η or azimuthal angle ϕ. More precisely, the mean of the track-vertex residuals in bins of the track η or ϕ is considered to reveal systematic misalignments in specific regions of the detector, which pull the mean value away from zero, i.e. the expected value in the ideal case. This kind of validation, usually referred to as primary vertex validation, is particularly sensitive to misalignments in the pixel detector, which drives the resolution of the reconstructed vertex position, as it is the tracker component closest to the beam line and with the best hit resolution. Figure 4.8 shows the result of the primary vertex validation for the 2016 end-of-year alignment, compared to the alignment used during the 2016 data taking and to the ideal case (from simulation). In general, the alignment performance is very close to ideal, with mean values of the track-vertex residuals of the order of few μm. The plot of the mean longitudinal residual $< d_z >$ as a function of the track ϕ shows an interesting feature: for the alignment in data taking, the distribution exhibits a step pattern in correspondence of $\phi \in [-\pi/2; +\pi/2]$ and $\phi \in [+\pi/2; \pi] \cup [-\pi; -\pi/2]$, corresponding to the group of modules on the two half shells of the barrel pixel. This pattern is corrected thanks to a shift of the two half barrels along the global z direction, which is introduced as a result of the final 2016 alignment.

Z boson mass from dimuon events

Muonic decays of the Z boson can be used as a standard candle to evaluate the alignment results [117]. This validation, usually referred to as the $Z(\mu\mu)$ validation, is based on the fact that several distortions of the tracker geometry can cause a bias on the reconstructed track curvature, thus leading to a bias in the reconstructed transverse momentum of the track. The selection of well-reconstructed $Z \to \mu^+\mu^-$ candidates requires the presence of two global muons with opposite sign charge, $p_T > 20\,\text{GeV}$, $|\eta| < 2.4$ and invariant mass $80 < M_{\mu\mu} < 120\,\text{GeV}$. The distribution of the reconstructed invariant mass is fitted to a Voigtian function to model the reconstructed Z-boson shape and an exponential function to model the background. The mass of the Z boson is then estimated from the mean of the Voigtian as a function of relevant observables, e.g. η or ϕ of the positively or negatively charged muons. In the ideal case, no dependence of the Z boson invariant mass on the muon kinematics is expected, any other behaviour is a hint for systematic misalignment effects. Figure 4.9 shows the distribution of the reconstructed Z boson

Figure 4.8 The distributions of the unbiased track-vertex residuals in the transverse plane, d_{xy} (top) and in the longitudinal direction, d_z (bottom) as a function of the track ϕ (left) and η (right), for a sample of 250 thousand collision events collected through minimum bias triggers. The performance of a dedicated alignment performed using the full 2016 dataset is compared to the one used during the 2016 data taking, as well as to the ideal case from simulation, where no misalignment is introduced [121].



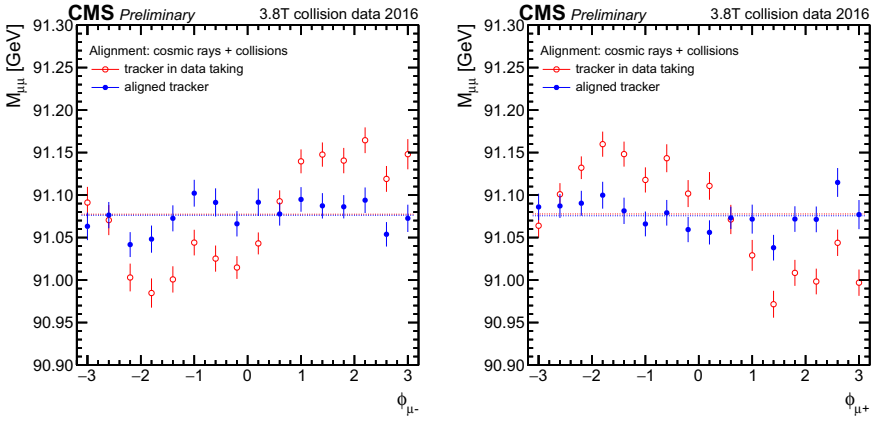
mass as a function of the azimuthal angle of the negatively and positively charged muons, for the alignment used during the 2016 data taking and for the one performed after the 2016 pp run. The updated alignment calibration visibly reduces modulation on the Z boson mass, which can be seen for the alignment used during 2016.

Track-split validation

The peculiar topology of cosmic ray tracks, which traverse the detector from the top to the bottom, can be used to perform the "track-split" validation of the alignment results [110]. In this validation, cosmic ray tracks are split at the point closest to the origin and the two parts[4] are refitted independently with the alignment constants under consideration. With ideal alignment, the track parameters of both halves should be identical. The differences in various track parameters between the two half-tracks

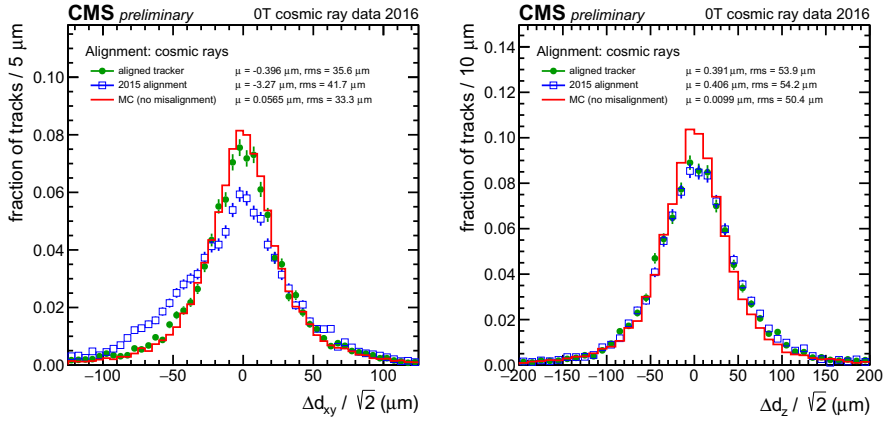
[4]These are called half-tracks in the following, meaning that each track is split into two, but not that they are exactly half of the entire track

Figure 4.9 Reconstructed $Z \to \mu\mu$ mass as a function of ϕ of the negatively (left) and of the positively (right) charged muons, for the initial alignment used during the 2016 data taking and for the one performed after the 2016 pp run (aligned tracker). A constant function is fitted to illustrate the deviation from the ideal behaviour [121].



are studied: the width of the distribution depends on the achieved alignment precision, while a mean value away from zero indicates possible biases. Figure 4.10 shows the normalised distributions of the difference in the transverse and longitudinal track impact parameters for two different alignment results, further compared to the ideal case when no misalignment is considered. Other variables which can be studied are for example the track polar and azimuthal angles θ and ϕ, the track pseudorapidity η and, if data have been recorded with the magnet turned on, the track transverse momentum.

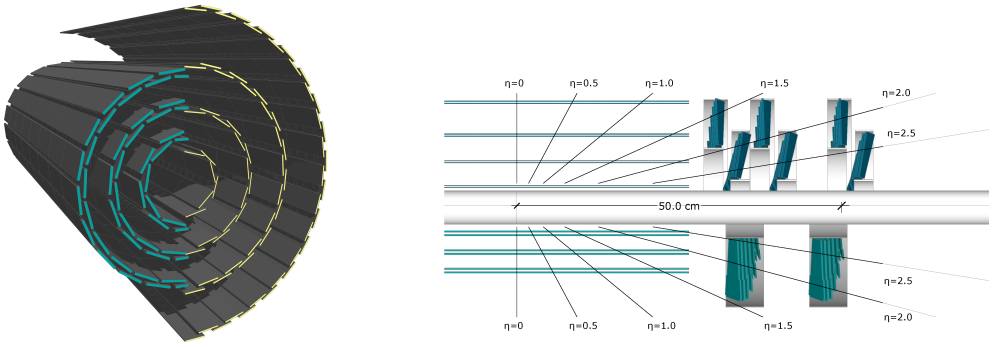
Figure 4.10 The normalised differences between two halves of a cosmic track, split at the point of closest approach to the interaction region, in the track transverse (left) and longitudinal (right) impact parameters, d_{xy} and d_z. The first alignment of the CMS tracker performed in 2016 with cosmic ray data recorded at $B = 0\,\mathrm{T}$ is compared to the tracker geometry of 2015, which is no longer valid mostly because of temperature and magnetic field changes. The mean and rms values show that the bias is reduced compared to the 2015 geometry, and the precision comes close to the ideal case (from simulation) where no misalignment is included [122].



4.2 Phase 1 upgrade of the pixel detector

The increase of the instantaneous luminosity, foreseen according to the LHC project plan shown in Fig. 2.5, requires upgrades of the experiments in order to retain the high design performances in a more challenging environment. The first phase of upgrades of the CMS experiment is taking place in the shutdown periods from 2012 to 2018, with interventions on different detector systems [123]. The most important change concerns the pixel detector, being completely replaced during the technical stop at the end of 2016 [124]. When keeping the original pixel system, readout inefficiencies would have decreased the ability of reconstructing tracks, and the higher bias voltage needed to compensate for radiation damages would have spoiled the hit resolution. The new pixel detector is in operation since the beginning of 2017. It consists of four barrel layers, instead of the previous three, and of three endcap disks on each side. The innermost layer is placed at a radial distance of 3.0 cm, instead of 4.4 cm of the current tracker from the beam, requiring the installation of a new beam pipe with a smaller diameter. The geometries of the new and of the previous pixel detector are compared in Fig. 4.11.

Figure 4.11 The pixel detector geometry before and after the Phase 1 upgrade, in the barrel and in the endcap regions. Left: the original barrel geometry (blue) is place aside the upgraded four-layer geometry (yellow). Right: the previous tracker design (bottom) compared to the upgraded geometry (top) [124].



The addition of the fourth layer reduces the distance between the last pixel hit and the strip system, improving track extrapolation towards the silicon strip tracker. Moreover, the additional layers provide four hit coverage over the full $|\eta| < 2.5$ range making it easier to detect triplet track seeds used for track reconstruction and therefore reducing track inefficiency. The pixel size is unchanged at $100 \times 150\ \mu m^2$, while the number of pixels to be read out increases to 79 million. The new size and arrangement of the modules in the endcap blades allows an easier production process and more flexibility in replacing damaged modules with spare ones. Before, each side of a blade was hosting several small modules with different sizes, while in the new design there is only one larger module per side with the same dimensions of a barrel pixel module. As a consequence, the total number of FPIX modules remained unchanged to 672. Despite the addition of one barrel layer and two endcap disks, the amount of material in the central η region has not increased. This is possible thanks to a lighter mechanical support, a new cooling system (from liquid C_6F_14 to dual phase CO_2), and by moving

services further out in z from the interaction point. Further important improvements have been introduced in the readout chip, including a faster readout of the pixel charge and increased number of buffer cells, to avoid data losses at high luminosity.

4.3 Alignment of the upgraded pixel tracker

The alignment of the upgraded pixel tracker represented a big challenge as a completely new pixel detector needed to be aligned in a short time scale,due to the early restart of the proton-proton collisions. In order to provide the best possible alignment under such challenging conditions, extended studies have been performed well ahead of the installation of the new detector, by using samples of simulated tracks. Several alignment configurations have been tested, aligning the pixel detector at different levels and with different number of tracks. Initially, the studies have been performed assuming that no cosmic ray data would have been available for alignment purposes before the pp collision run. Then, as the schedule for the LHC restart became more defined, the possibility for a short cosmic ray data taking prior to the proton-proton run emerged, and further alignment scenarios with simulated cosmic ray tracks have been tested. Such studies are described in Secs 4.3.2 and 4.3.3, and their most relevant findings are summarised in Sec. 4.3.4. In Sec. 4.3.5 the alignment of the new pixel detector with the first amount of cosmic ray data is described, including its different iterations performed and the validation results. Both the simulation studies and the first alignment of the pixel detector with cosmic ray data were performed in the context of this thesis.

4.3.1 Expected misalignment after the installation

As a starting point for the realignment studies, an assumption for the expected misalignment of the pixel detector after the installation is made, based on the suggestions of experts and their experience in the installation of the current CMS pixel detector at the beginning of the LHC Run 1, as well as its re-insertion before the start of the LHC Run 2. The misaligned geometry is built by shifting the different pixel detector components around their ideal position. The magnitude of the shifts in each direction are chosen randomly, either from a uniform or from a Gaussian distribution, its width reflecting the expected installation precision. The positions of the two pixel endcap half cylinders, instead, have been intentionally modified by fixed amounts, such that the two halves of the same cylinder are shifted by similar amounts in the longitudinal direction. The largest applied misalignment is a shift of $\approx 4\,\text{mm}$ for one pixel endcap in the z direction. The strip detector is assumed to remain well aligned, as found to be the case after the start of the LHC Run 2 in 2015, and thus only a small misalignment at module level of the order of 20 μm is considered. The misalignments applied to the different structures of the tracker are summarised in Table 4.2. The comparison of the ideal and misaligned geometries is shown in Fig. 4.12.

Figure 4.12 Comparison of the design and misaligned tracker geometries, for the misalignment scenario assumed after the pixel installation. Each point represents a module. The position differences (design - realigned) are plotted as a function of the design positions. The projection of the module position differences along several directions are shown. Most notably, the two pixel endcaps are observed to be shifted by $\approx 4\,\text{mm}$ and $\approx 0.7\,\text{mm}$, respectively, in the z direction, and the pixel barrel by $\approx 0.5\,\text{mm}$ in the x, y and z directions with respect to their design positions, as can be seen in the second, third and fourth plots of the last column.

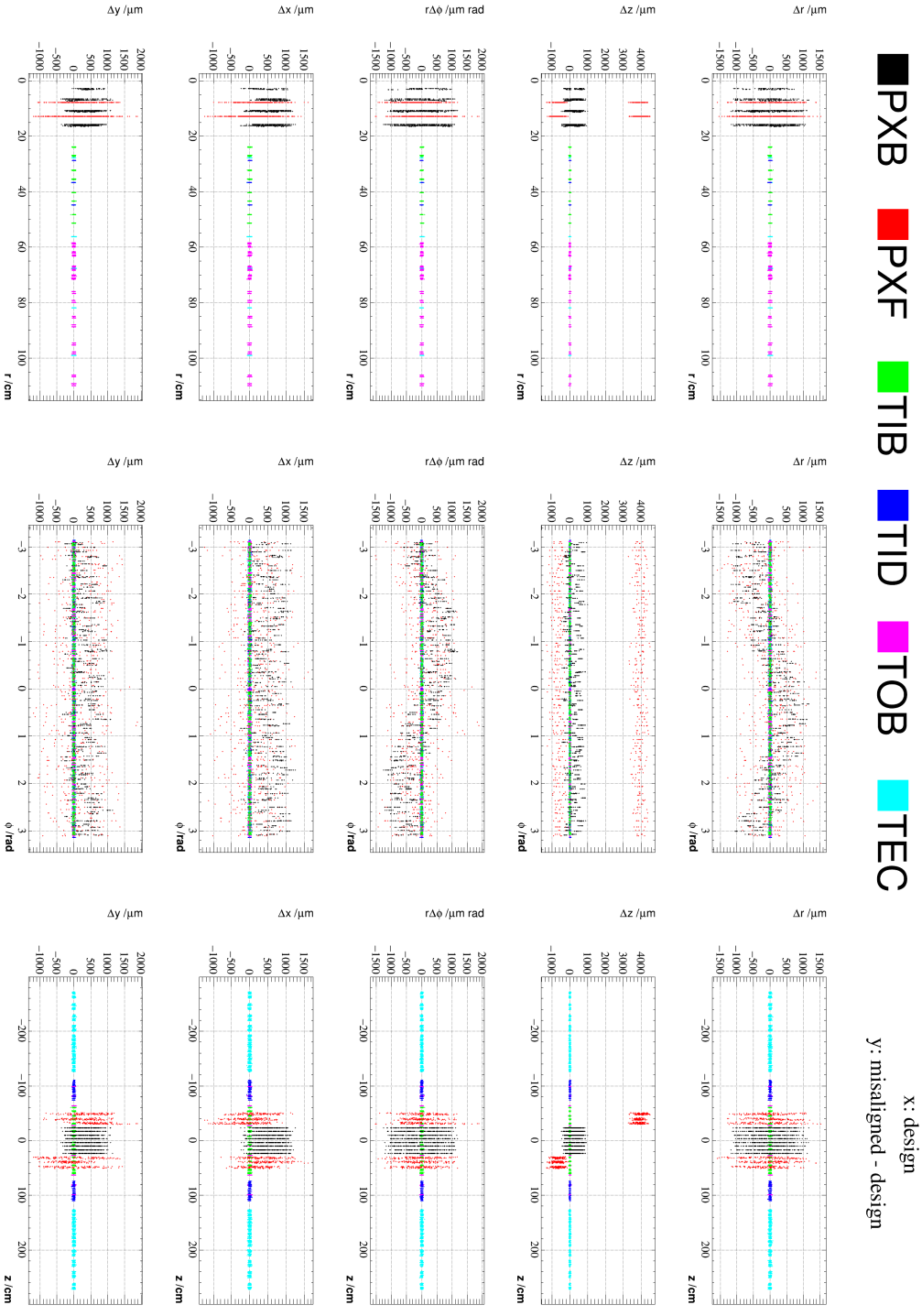


Table 4.2 Assumed misalignment scenario after the pixel detector installation. The shifts applied at the different structure levels of the tracker with respect to the design geometry, are reported. For the high level structures, shifts are applied in the x, y and z directions, while for the single modules they are applied in the local coordinate system (x', y', z'). Uniform shifts are extracted randomly from a uniform distribution between $-\Delta s$ and $+\Delta s$, while Gaussian shifts are sampled from a Gaussian distribution with a mean value of zero and a standard deviation of Δs, with s being the value of the shift applied in a given direction. For fixed shifts, the substructures are misaligned by the indicated amount.

Tracker Structure	$\Delta x[\mu m]$	$\Delta y[\mu m]$	$\Delta z[\mu m]$	Distribution
Pixel Barrel				
Barrel	500	500	1000	uniform
Half Barrels	100	100	100	uniform
Ladders	500	500	500	uniform
Modules	100	100	100	gaussian
Pixel Endcaps				
Half Cylinder 1 (+z side)	250	250	-680	fixed
Half Cylinder 2 (+z side)	250	-250	-720	fixed
Half Cylinder 1 (-z side)	250	250	3940	fixed
Half Cylinder 2 (-z side)	-250	250	4020	fixed
Half Disks	250	500	100	uniform
Blades	300	300	200	gaussian
Modules	200	200	50	gaussian
Strip Detector				
Modules	20	20	20	gaussian

4.3.2 Alignment strategy with simulated minimum bias events

Several strategies to recover from misalignment after the installation process have been tested using simulated samples. Initially, the development of the realignment strategy assumes that no cosmic ray data will be available, and only tracks from minimum bias events are used. Events are first simulated and reconstructed assuming the design detector geometry. In the studies presented here, tracks are then refitted using the alignment constants under consideration, without performing the track finding step once more, as it is computationally very intensive. However, it has been checked on a smaller subset that results do not change significantly when the full track reconstruction is run, instead of the track refit only.

Due to large initial misalignment expected after the installation, the best strategy is to follow an iterative approach starting with the alignment of the pixel high-level structures, and then performing additional alignment iterations with finer granularity, for example at the level of blades and ladders and ultimately at module level. In the initial situation, in fact, many tracks are rejected in the alignment fit because of a high χ^2 value. However, the position of large detector structures can be derived with relatively small

set of tracks, as only a few alignment parameters have to be determined. After every alignment iteration, the number of rejected tracks decreases and makes it possible to derive more parameters. Moreover, following such a staged approach during the first days of data taking with the new pixel detector has the big advantage that alignment parameters can be derived quickly and then be used for the track reconstruction of the data collected afterwards.

As the initial misalignment of the pixel detector is large, it is necessary to inflate the uncertainties, which is done by increasing the alignment position errors (APE) and including them in the alignment fit. Conversely, many hits can not be matched to a track as they lay too far from their expected position. Initially, studies are performed assuming APE values of 500 μm in the pixel and 20 μm in the strip detectors. The choice of the APE values and its effect on the recovery from misalignment is discussed at the end of this Section. Due to the preferred topology of minimum bias tracks, which have usually a very low polar angle and transverse momentum, the alignment algorithm is hardly sensitive to the misalignment along the longitudinal direction in the forward pixel tracker. For this reason, the offsets in the z' direction are kept fixed, i.e. not aligned, in the forward pixel tracker, when performing module level alignment with minimum bias tracks only. The alignment procedures are performed using a sample of up to 25 million minimum bias tracks.

In the first alignment iteration, the pixel detector high level structures, namely the two pixel half barrels and the four pixel endcap half cylinders, are aligned. Then, a second iteration is performed to align the pixel detector at the level of single modules. An intermediate alignment iteration before performing the module level alignment has also been attempted, to align the pixel detector at the level of ladders in the barrel pixel tracker and blades in the forward pixel tracker. However, this additional iteration is found not to significantly improve the recovery from misalignment and the following results thus refer to the realignment strategy with two iterations. Table 4.3 shows the mean position and the spread of the pixel modules around their ideal positions, which quantify the recovery from misalignment. The remaining differences in the module positions with respect to the design geometry after the first and the second alignment iteration are plotted in Figs. 4.13 and 4.14, respectively. In summary, the most relevant conclusions that can be drawn from this study are the following:

- the systematic shifts of the large structures can be well recovered already after the first alignment iteration, as it can be observed from the comparison of geometries before and after the latter alignment in Fig. 4.13, as well as from Table 4.3;
- the recovery from misalignment is comparable if the intermediate alignment iteration (aligning the pixel detector at the level of blades and ladders) is omitted, and the module level alignment is performed as second step, for which the results are given in Table 4.3;

- the random misalignment of the forward pixel modules along the z' direction can not be recovered with minimum bias tracks alone, but requires the use of cosmic ray tracks, as shown later in Sec. 4.3.3.

Table 4.3 Mean (μ) and RMS (σ) of the differences of pixel modules positions before and after each step of the realignment strategy, with respect to their design positions, along the x, y and z directions. Values are shown separately for the barrel and forward pixel modules.

Module positions differences		Alignment		
		Misaligned	High level structures	Module level
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	-517 ± 300	-12 ± 299	-14 ± 50
	$\mu(y) \pm \sigma(y)$	-331 ± 301	6 ± 301	-3 ± 55
	$\mu(z) \pm \sigma(z)$	-239 ± 308	-9 ± 307	-7 ± 103
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	-178 ± 469	60 ± 385	-18 ± 73
	$\mu(y) \pm \sigma(y)$	-205 ± 568	-18 ± 381	2 ± 85
	$\mu(z) \pm \sigma(z)$	1630 ± 2312	-174 ± 319	-48 ± 239

The impact of the misalignment on physics quantities has been evaluated by performing the primary vertex validation (see Sec. 4.1.4), as it is the validation most sensitive to misalignments in the pixel detector. Figure 4.15 shows the mean unbiased track-vertex residuals, projected into the transverse plane and on the longitudinal direction, as a function of the track η and ϕ. Due to the large initial misalignment, the vertex reconstruction is found to be severely biased for the misaligned geometry (empty circles). After the first iteration of the realignment strategy, i.e. the pixel high level structures alignment (full circles), the biases are slightly reduced, and the largest improvement can be seen for the $d_z(\eta)$ distribution at high negative η values: this is driven by the large misalignment of the $z-$ pixel endcap, which is strongly reduced after the first alignment iteration. After the module level alignment (full squares), the behaviour of the track-vertex residuals distributions resembles much more the one without misalignment (empty squares). Residual biases of the order of $50 - 100\,\mu$m are observed especially as a function of the track ϕ and in the $d_z(\eta)$ distribution for $\eta < -1.5$, due to the larger initial misalignment of the FPIX minus endcap.

Pileup dependence

One important aspect to study for the alignment of the pixel detector at start-up is whether the recovery from misalignment has a dependency on the number of pileup interactions in the event, as the number of proton-proton interactions per bunch crossing is expected to increase quite rapidly in the first days after the LHC restart. The

Figure 4.13 Comparison of the realigned geometry after the pixel high level structures alignment with the design geometry. Each point represents a module. The position differences (design - realigned) are plotted as a function of the design positions. The projection of the module position differences along several directions are shown. The large systematic shifts if the pixel barrel (black points) and of the pixel endcaps (red points) visible in the misaligned geometry (Fig. 4.12) can be cured already after this first alignment iteration.

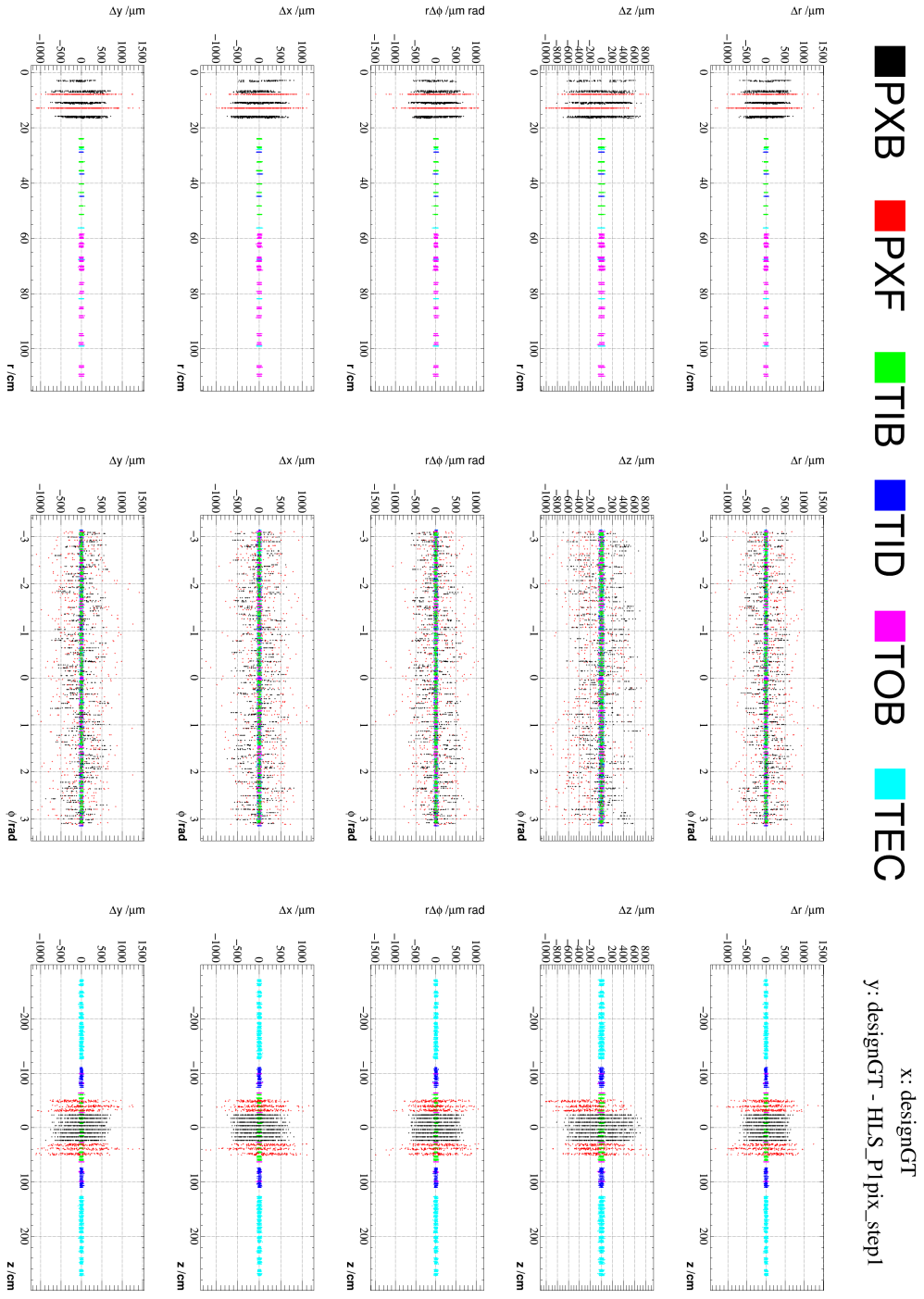


Figure 4.14 Comparison of the realigned geometry after the pixel module level alignment with the design geometry. Each point represents a module. The position differences (design - realigned) are plotted as a function of the design positions. The projection of the module position differences along several directions are shown. As explained in the text, the forward pixel modules are only allowed to be aligned in the x' and y' directions, while the z' direction is kept fixed. After this second alignment iteration, the spread of the module positions is considerably reduced.

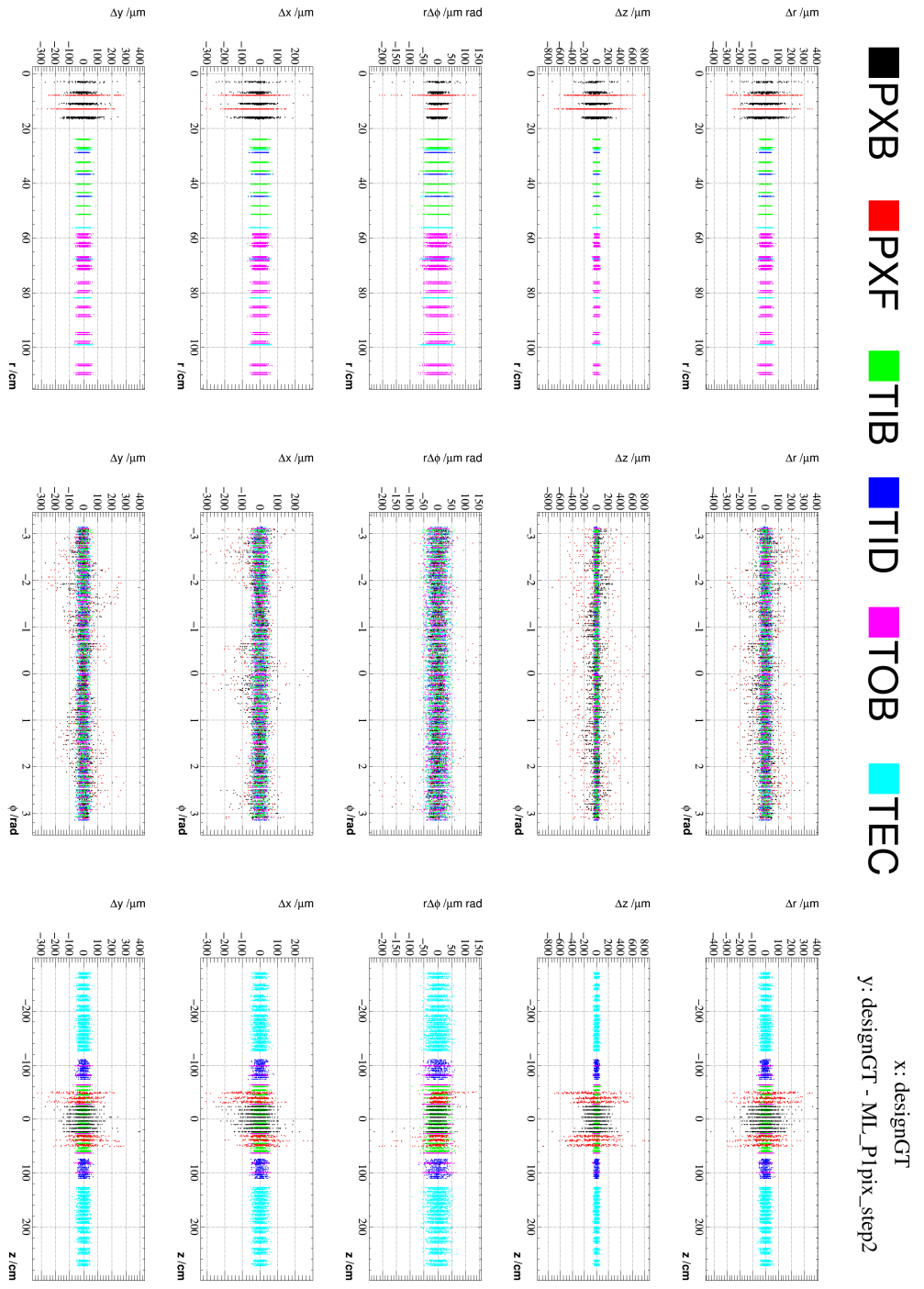
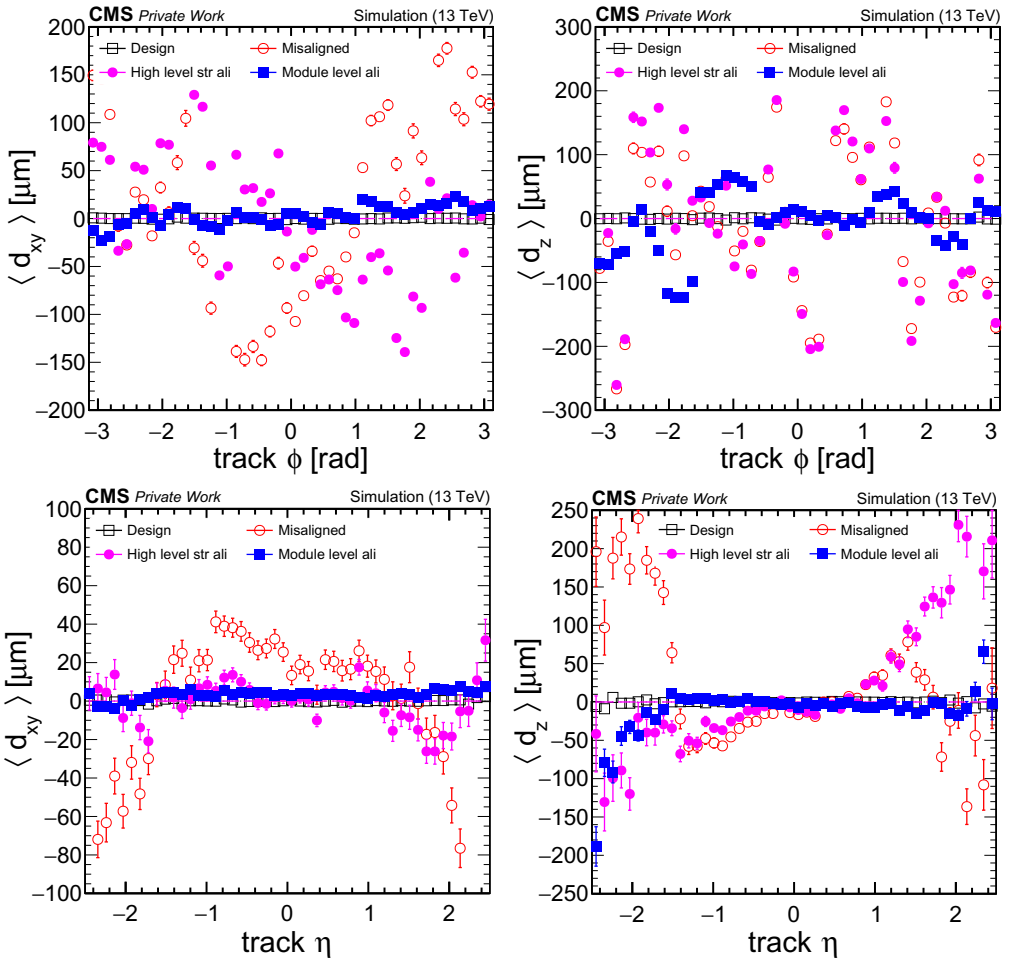


Figure 4.15 Mean values of the transverse d_{xy} and longitudinal d_z track-vertex unbiased residuals as a function of the track ϕ and η, for the initial misalignment scenario (empty circles), the design geometry (full circles), and after the first and second realignment steps, namely the pixel high level structures alignment (full circles) and the pixel module level alignment (full squares). The strong biases observed for the initial misaligned geometry are gradually reduced after each alignment iteration.



effect of pileup on the recovery from misalignment has been studied by performing the realignment strategy using tracks from events with different numbers of simulated pileup interactions. The two realignment iterations, namely the pixel high level structures alignment and the pixel module level alignment, have been performed three times using tracks in events with pileup less than 30, between 30 and 40 and more than 40. The pileup profile of the simulated sample is such that the number of tracks used in the alignment correspond to about 9, 11 and 5 M in the three considered pileup bins, respectively. The recovery from misalignment is quantified by the mean and RMS of the modules positions differences after each alignment iteration, with respect to the ideal positions. Results are summarised in Table 4.4, and show that there is no evident trend as a function of pileup, the largest differences being of $\approx 20\,\mu$m in the modules positions RMS. This indicates that, within the tested pileup interval, the performance of the chosen alignment strategy should not strongly depend on the instantaneous luminosity at which the data are recorded.

Table 4.4 Recovery from misalignment as a function of pileup. The realignment strategy, consisting of a pixel high level structures alignment followed by the pixel module level alignment, has been performed using tracks in events with different numbers of simulated pileup (PU) interactions. The mean (μ) and RMS (σ) of the differences of pixel modules positions before and after the realignment with respect to their design positions, along the x, y and z directions, are shown separately for the barrel and forward pixel modules, after each alignment iteration and for each pileup bin.

Pixel high level structures alignment		PU < 30	30 < PU < 40	PU > 40
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	55 ± 302	59 ± 302	61 ± 303
	$\mu(y) \pm \sigma(y)$	−12 ± 302	−12 ± 302	−12 ± 303
	$\mu(z) \pm \sigma(z)$	−16 ± 314	−22 ± 316	−24 ± 318
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	−54 ± 385	−52 ± 385	−51 ± 385
	$\mu(y) \pm \sigma(y)$	20 ± 435	20 ± 435	20 ± 435
	$\mu(z) \pm \sigma(z)$	−150 ± 347	−145 ± 348	−142 ± 348
Pixel module level alignment		**PU < 30**	**30 < PU < 40**	**PU > 40**
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	22 ± 68	23 ± 70	23 ± 72
	$\mu(y) \pm \sigma(y)$	−7 ± 75	−9 ± 79	−10 ± 80
	$\mu(z) \pm \sigma(z)$	7 ± 154	7 ± 162	7 ± 165
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	20 ± 113	21 ± 116	19 ± 111
	$\mu(y) \pm \sigma(y)$	−19 ± 128	−19 ± 139	−22 ± 125
	$\mu(z) \pm \sigma(z)$	65 ± 461	61 ± 475	67 ± 470

Study of the APE choice

The Alignment Position Errors (APE) are the estimated uncertainty on the module position in the three global coordinates, and their value is added in quadrature to the hit errors in the pattern recognition and in the track fitting [110]. As the initial misalignment of the pixel detector is large, it is necessary to inflate such uncertainties and to include them in the alignment fit. This avoids that too many track hits can not be associated to the tracks during the refit, due to them being too distant from their expected positions, as well as that a large number of tracks is not considered by the alignment algorithm because of their poor quality. On the other hand, choosing too large APE can degrade the results, as also fake or poorly reconstructed tracks are taken into account. The most suitable values of the APE to include in the alignment fit at each step of the pixel realignment strategy have been identified by repeating each

alignment step using different APE for the pixel modules, and comparing the recovery from misalignment. Again, the latter is quantified by looking at the mean and RMS of the modules positions differences with respect to the design geometry, reported in Table 4.5. APEs of 500, 400, 300, 200 and 100 μm for the pixel modules have been tested. The APE for strip modules have not been varied and amount to 20 μm. As a result, it turns out that the optimal choice is to use large APE for the first alignment iteration (pixel high level structures alignment), namely APE of 500 μm. Using this APE value, the second alignment iteration is repeated for the different APE sets being tested: the best choice for the second iteration (pixel module level alignment) is to reduce the APE to 200 μm. For a further alignment iteration, e.g. a third iteration performing again a pixel module level alignment, the APE can be further reduced to 10 μm.

Table 4.5 Recovery from misalignment as a function of the APE choice for pixel modules. The realignment strategy, consisting of a pixel high level structures alignment followed by the pixel module level alignment, has been performed using different APE values for pixel modules. The mean (μ) and RMS (σ) of the differences of pixel modules positions before and after the realignment with respect to their design positions, along the x, y and z directions, are shown separately for the barrel and forward pixel modules, after each alignment iteration and for each set of APE.

Pixel high level structures alignment		APE 500	APE 400	APE 300	APE 200	APE 100
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	-12 ± 300	-6 ± 299	-1.5 ± 299	-39 ± 298	-167 ± 300
	$\mu(y) \pm \sigma(y)$	6 ± 301	7 ± 302	5 ± 303	-14 ± 304	-102 ± 308
	$\mu(z) \pm \sigma(z)$	9 ± 307	-5 ± 307	-1.6 ± 307	-13 ± 307	-16 ± 307
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	59 ± 385	77 ± 385	90 ± 384	65 ± 386	-36 ± 424
	$\mu(y) \pm \sigma(y)$	-18 ± 431	-16 ± 432	-10 ± 432	-8 ± 437	-55 ± 476
	$\mu(z) \pm \sigma(z)$	-174 ± 319	-281 ± 392	-390 ± 602	537 ± 935	-752 ± 1409
Pixel module level alignment		APE 500	APE 400	APE 300	APE 200	APE 100
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	-14 ± 50	-14 ± 44	-14 ± 37	-14 ± 34	-19 ± 65
	$\mu(y) \pm \sigma(y)$	-3 ± 60	-4 ± 47	-4 ± 40	-3 ± 38	-0.3 ± 70
	$\mu(z) \pm \sigma(z)$	-7 ± 103	-6 ± 88	-5 ± 73	-5 ± 60	-2 ± 99
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	-19 ± 73	-17 ± 68	-14 ± 62	-12 ± 60	-23 ± 79
	$\mu(y) \pm \sigma(y)$	1.7 ± 85	-1.3 ± 78	0.4 ± 69	5 ± 69	12 ± 94
	$\mu(z) \pm \sigma(z)$	-48 ± 247	-46 ± 245	-41 ± 245	-42 ± 246	-35 ± 283

4.3.3 Realignment studies with cosmic ray events

The alignment of the upgraded pixel detector at start-up has also been prepared for the scenario in which the first recorded data are not from proton-proton collisions, but from cosmic ray events, recorded first without magnetic field. The realignment strategy applied here is the one developed with minimum bias events (see Sec. 4.3.2), consisting of two iterations:

1. the alignment of the pixel high level structures, namely the two half barrels and the four half cylinders of the pixel endcaps;
2. the alignment of the pixel detector at the level of single modules. Differently from minimum bias tracks, cosmic ray tracks are sensitive to misalignment in the z' direction for forward pixel modules, and this degree of freedom is no longer excluded from the alignment fit.

The main difference between the cosmic ray and the minimum bias datasets is their size: the collection rate of minimum bias events is in fact much larger than for cosmic ray events. For this reason, when estimating the recovery from misalignment that can be achieved with cosmic ray tracks, it is important to study the dependence on the number of events available for the alignment. In the study presented here, the first two alignment iterations (pixel high level structures and module level alignment) are performed using cosmic ray data recorded at $B = 0\,\mathrm{T}$, with datasets of 2, 1, 0.5 or 0.2 M events. The recovery from misalignment achieved with cosmic ray events is again quantified through the mean and RMS of the modules position differences with respect to the ideal geometry, and is estimated in Table 4.6 for the four different dataset sizes considered. On average, the number of tracks selected and used in the alignment fit amounts to ≈ 0.5 tracks per reconstructed event. The results of this study show that, even with a relatively small amount of cosmic ray events (0.2 M), systematic shifts introduced in the misaligned geometry can be considerably reduced, from $O(100\ \mu\mathrm{m})$ to $O(10\ \mu\mathrm{m})$, and that increasing the statistics does not significantly improve the result for the high level structures alignment. After the second alignment iteration, the systematic shifts are further reduced to the order of a few μm, especially in the barrel pixel tracker. The performance of the module level alignment is strongly dependent on the number of tracks used to derive the alignment, as can be seen from the RMS of the modules positions differences, which decreases when increasing the number of events used for the alignment. For the forward pixel tracker, the recovery from misalignment is not so effective because of the lower number of cosmic ray tracks crossing the pixel endcaps modules.

Table 4.6 Recovery from misalignment when using different amounts of cosmic ray tracks. The realignment strategy, consisting of a pixel high level structures alignment followed by the pixel module level alignment, has been performed using cosmic ray events recorded without magnetic field. The mean (μ) and RMS (σ) of the differences of pixel modules positions before and after the realignment with respect to their design positions, along the x, y and z directions, are shown separately for the barrel and forward pixel tracker modules, after each alignment iteration and for the different dataset sizes. To facilitate the comparison, the first column shows the values for the initial misalignment scenario.

Pixel high level structures alignment		Misaligned	0.2 M events	0.5 M events	1 M events	2 M events
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	-517 ± 300	-24 ± 297	-25 ± 297	-25 ± 297	-25 ± 297
	$\mu(y) \pm \sigma(y)$	-331 ± 301	-11 ± 303	-18 ± 302	-22 ± 302	-22 ± 302
	$\mu(z) \pm \sigma(z)$	-239 ± 308	-4 ± 306	-2 ± 306	-2 ± 306	-1 ± 306
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	-178 ± 469	-7 ± 387	-6 ± 385	-8 ± 385	-10 ± 386
	$\mu(y) \pm \sigma(y)$	-205 ± 567	-6 ± 448	-5 ± 443	16 ± 443	-16 ± 442
	$\mu(z) \pm \sigma(z)$	1630 ± 2312	-120 ± 280	146 ± 298	-146 ± 299	-137 ± 293
Pixel module level alignment		Misaligned	0.2 M events	0.5 M events	1 M events	2 M events
Barrel Pixel [μm]	$\mu(x) \pm \sigma(x)$	-517 ± 300	-1.1 ± 252	-0.6 ± 155	-3 ± 75	-3 ± 21
	$\mu(y) \pm \sigma(y)$	-331 ± 301	-3 ± 260	-1.6 ± 154	-2 ± 59	-0.5 ± 16
	$\mu(z) \pm \sigma(z)$	-239 ± 308	-3 ± 268	-1.2 ± 161	-6 ± 61	-6 ± 29
Forward Pixel [μm]	$\mu(x) \pm \sigma(x)$	-178 ± 469	-31 ± 386	-20 ± 373	-11 ± 312	-7 ± 193
	$\mu(y) \pm \sigma(y)$	-205 ± 567	5 ± 449	-8 ± 432	-9 ± 350	-8 ± 208
	$\mu(z) \pm \sigma(z)$	1630 ± 2312	-16 ± 242	35 ± 235	-25 ± 186	-19 ± 135

4.3.4 Conclusions on the studies with simulated events

Extensive studies on simulated events have been performed to evaluate the recovery from the misalignment expected after the installation of the new pixel detector. Two scenarios have been considered:

- the first available dataset is from proton-proton collisions, recorded with minimum bias triggers;
- the first available dataset is from cosmic ray tracks, recorded without magnetic field.

A two-step realignment procedure, aligning first the high level structures and then the single modules of the pixel detector, has been tested. The studies show that large

shifts of the high level structures can be recovered successfully already after the first alignment iteration, even when it is performed with a small sample (0.2 M events) of cosmic ray events. Due to the initial large misalignment expected, it is important to include APE in the alignment fit. For the first iteration, the studies suggest to use APE of 500 μm for pixel modules, and to reduce them to 200 μm for the second iteration. In order to reach a good alignment precision at the level of single modules, a large sample of reconstructed tracks needs to be collected. This can be easily achieved recording minimum bias events, as the recovery from misalignment showed to be stable as a function of pileup. However, this track topology can not be used to align the forward pixel modules in the z' direction, as most of the tracks have a large η and small p_T and the alignment fit is not able to recover the misalignment in this direction. On the other hand, cosmic ray tracks are sensitive to the latter misalignment and can effectively reduce it. In summary, the studies show that it is very beneficial to collect cosmic ray data with the new pixel detector after its installation, before the start of the proton-proton collision run, such that the first important alignment corrections can be derived and applied already before recording the first collision events.

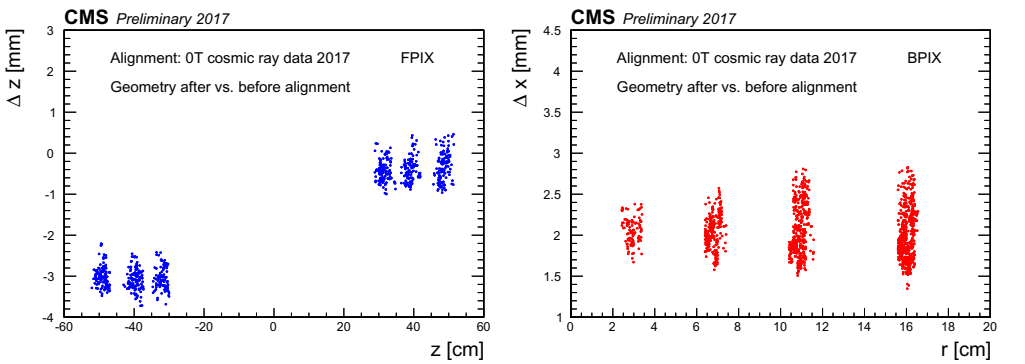
4.3.5 Alignment with cosmic ray data

The first data with the new pixel detector have been recorded in spring 2017, prior to the restart of the LHC proton-proton run. After an initial period of detector commissioning to derive the initial calibrations, cosmic ray data for alignment purposes have been recorded. Initially, events have been recorded without magnetic field, during the "Cosmic Run at Zero Tesla" (CRUZET). Then, the CMS magnet has been switched on and a "Cosmic Run At Four Tesla" (CRAFT) took place before the first collisions. The cosmic ray data was crucial to derive the first geometry corrections for the new pixel tracker, and as expected from the studies performed on Monte Carlo, large shifts of the high level structures could be identified and corrected successfully at an early stage. In the following, the alignment results of the upgraded pixel detector obtained with the cosmic ray data prior to the proton-proton collision run are presented, together with their validation.

Alignment with 0 T data

The first alignment of the pixel detector has been derived using cosmic ray tracks recorded at 0 T. The new pixel detector has been calibrated and integrated in the data readout in two steps, starting from the forward pixel tracker and continuing with the barrel pixel tracker. Already before performing any alignment, the observed asymmetric track rate between the two forward pixel endcaps provided a hint of large initial misalignment for this subdetector. The first alignment of the forward pixel high level structures, namely the four half cylinders, has been performed using 1.5 M reconstructed 0 T cosmic ray tracks. Also the high-level structures of the strip detector have

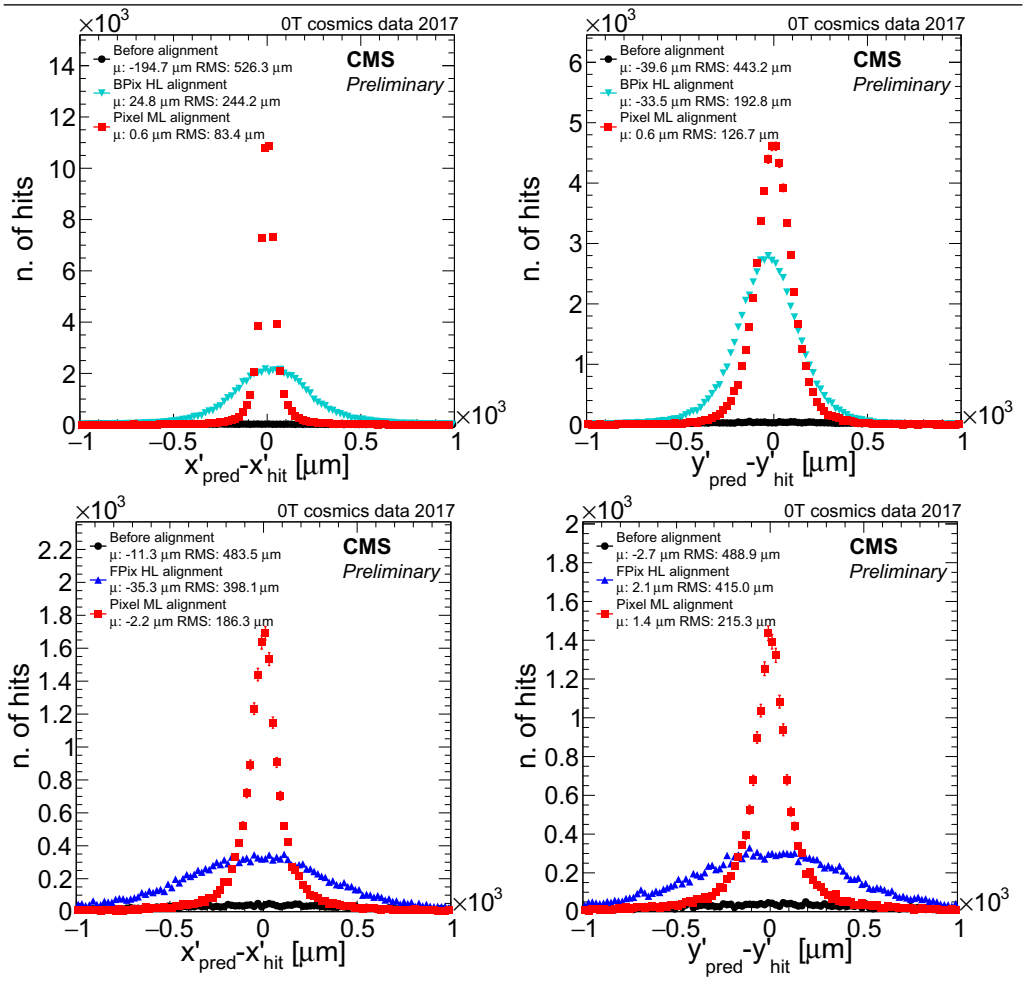
Figure 4.16 Largest corrections after the first pixel high-level structures alignment, for the forward pixel (left) and for the barrel pixel (right). The plots show the modules positions corrections after the alignment, with respect to the geometry before performing any alignment, as a function of the design positions. The alignments were performed with 0 T cosmic ray tracks. In the forward pixel tracker, the largest correction is a shift of ≈ -2.8 mm of the z- endcap in the longitudinal direction. For the barrel pixel tracker instead, both half cylinders appear to be systematically misaligned by ≈ 2 mm in the horizontal plane along the x direction. Since the detector is also rotated with respect to the geometry assumed before alignment, there is an increasing spread of Δx with the radius r [125].



been aligned to avoid introducing any bias in the pixel alignment due to misalignment in the strip tracker. APE of 500 μm and 20 μm for pixel and strip modules, respectively, have been used for the track reconstruction and fit, according to the outcome of the studies performed on simulated events (see Sec. 4.3.2). The largest measured correction to the assumed geometry is a shift of ≈ -2.8 mm of the z- endcap of the forward pixel in the longitudinal direction, as shown in Fig. 4.16 (left). Such a large misalignment explained the asymmetric track rate between the two forward pixel sides. In fact, after applying the alignment correction this behaviour has not been observed anymore.

Regarding the strip tracker, the magnitude of the alignment corrections is of $O(10\ \mu\text{m})$ for most of the substructures, with the exception of the tracker endcaps (TEC), which move by $\approx 100\ \mu$m in the longitudinal direction. Overall, the assumption that the strip tracker would not be substantially misaligned at the startup has been confirmed. Due to the large correction introduced in the forward pixel, data collected before the alignment update has been reprocessed before using it for other alignment iterations. After the barrel pixel detector was included in the data taking, the alignment correction for the positions and orientations of the two half barrels could be derived. The alignment uses ≈ 0.5 M cosmic ray tracks, recorded at 0 T. The largest geometry correction is a shift of ≈ 2 mm in the x direction of the whole barrel, as it can be seen in Fig. 4.16 (right). The alignment of the pixel detector at the level of single modules has been performed using the full statistics of the CRUZET run, with ≈ 3.2 M tracks used for the alignment. For the pixel modules, the APE have been reduced to 100 μm. The distributions of the unbiased track-hit residuals in the local x and y directions for the barrel and for the forward pixel are shown in Fig. 4.17, respectively, for the two alignment iterations and for the initial geometry assumed before performing any alignment. After each alignment iteration,

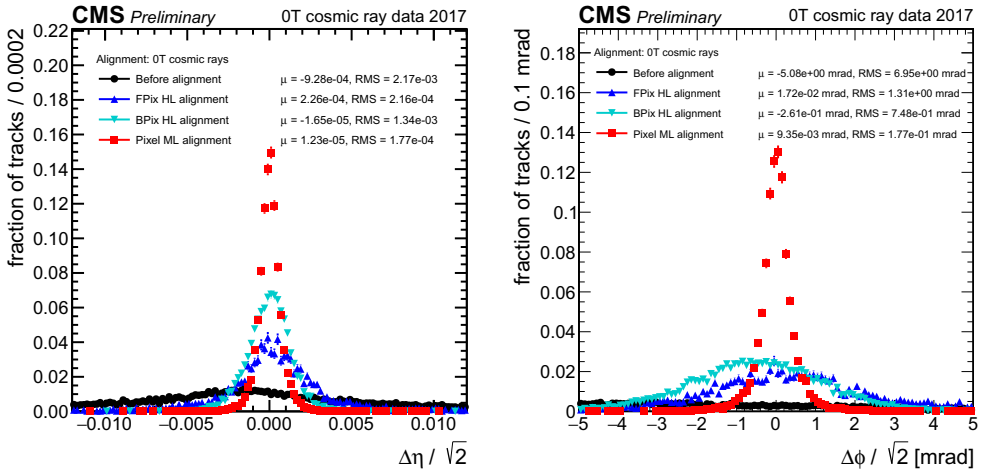
Figure 4.17 Unbiased track-hit residuals in the barrel (top) and forward (bottom) pixel, in the local x (left) and y (right) directions. The distributions are shown for the different alignment iterations performed: the black circles are relative to the geometry assumed before performing any alignment, the triangles are relative to the high level structures alignment of the barrel and forward pixel, and the red squares represent the pixel module level alignment structures alignment. After each alignment iteration, both the bias and the width of the distributions are strongly reduced [125].



the bias and the width of the distributions decrease significantly, proving the reduction of systematic misalignment effects and an improved local precision.

The improvement achieved thanks to the different alignment iterations can be seen in the track-split validation as well. As explained in Sec. 4.1.4, this kind of validation consists of splitting cosmic ray tracks into two halves at the point of closest approach to the origin, refitting them independently and looking at the differences in the track parameters for the two halves. Figure 4.18 shows the differences in the track's η and ϕ parameters at the different stages of the pixel realignment. The improvement achieved after each alignment iteration can be observed, as both the bias and the width of the distributions are considerably reduced.

Figure 4.18 Normalised difference between the two halves of a cosmic ray track at 0 T, split at the point of closest approach to the interaction region, in the track η (left) and ϕ (right). The distributions are shown for the different alignment iterations performed: the black circles are relative to the geometry assumed before performing any alignment, the triangles are relative to the high level structures alignment of the barrel and forward pixel, and the red squares represent the pixel module level alignment structures alignment. After each alignment iteration, both the bias and the width of the distributions are strongly reduced [125].



Alignment with 3.8 T **data**

A further alignment update was necessary after the CMS magnetic field was turned on, as the magnetic field change is known to induce sizeable movements of the pixel detector high level structures. As a consequence, the geometry derived at 0 T became inadequate, showing biases in the track-hit residuals distributions (Fig. 4.19, red squares) and in the track-split validation (Fig. 4.20, red squares). Therefore, a further alignment of the pixel detector has been performed using cosmic ray data collected at 3.8 T. The limited statistics of this dataset, corresponding to $\approx$ 300 k events, does not allow to perform a pixel module level alignment, and the corrections have been derived for the strip and pixel high level structures. The observed biases have been successfully cured, as can be observed from Figs. 4.19 and 4.20 (green crosses).

Figure 4.19 Unbiased track-hit residuals in the barrel (top) and forward (bottom) pixel, in the local x (left) and y (right) directions. The distributions are shown for the geometry assumed before performing any alignment (black circles), for the geometry derived with 0 T cosmic ray tracks (red squares) and for the alignment derived with 3.8 T cosmic ray tracks (green crosses) [125].

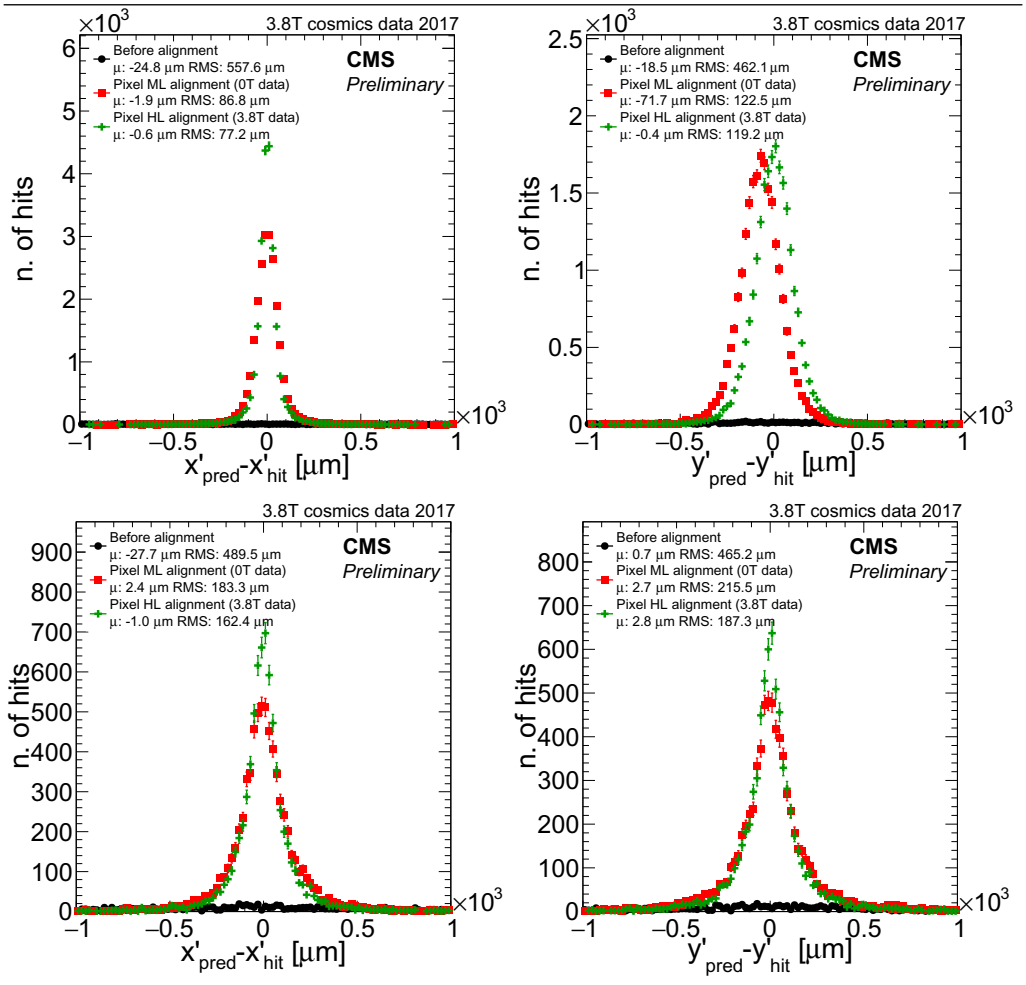
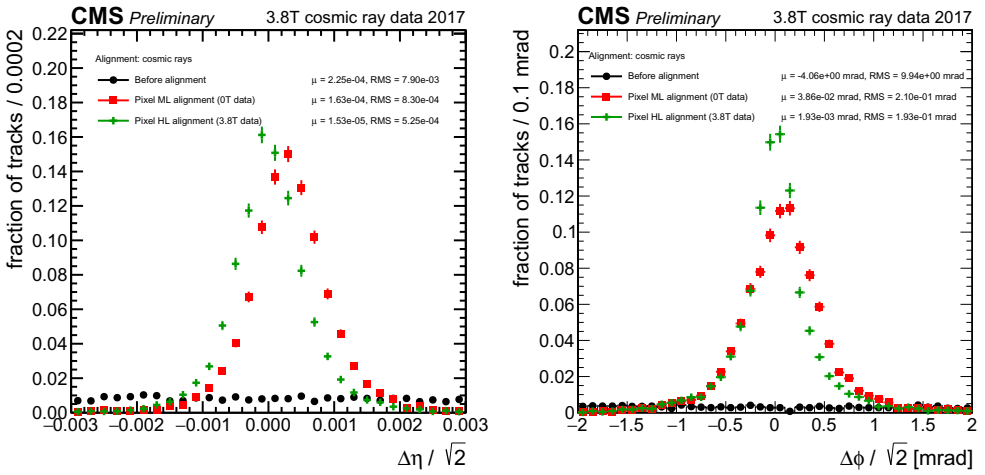


Figure 4.20 Normalised difference between the two halves of a cosmic ray track at 3.8 T, split at the point of closest approach to the interaction region, in the track η (left) and ϕ (right). After the ramp up of the magnetic field, the positions of the high level structures have been recomputed using 3.8 T cosmic ray tracks. The latter alignment (green crosses) shows improved performance over the geometry derived from 0 T cosmic ray data [125].



4.4 Conclusions

The alignment of the CMS tracker represents an important part of the detector calibration, and aims at measuring the positions of the tracker modules with a precision of a few μm. This can be achieved with the means of track-based alignment, its implementation and performance being described in the first part of this chapter. The tracker alignment calibration is a complex problem: from the computational point of view, because the number of parameters to determine is typically very large, and from the point of view of the configuration, because it is usually not clear a-priori what is the best strategy to follow in terms of alignables granularity, track topologies and statistics to use. A particularly challenging time in the CMS tracker alignment activities occurred at the beginning of 2017, when the pixel detector of the CMS experiment has been entirely replaced with a new upgraded one and the alignment corrections needed to be derived right after the detector installation, in a short timescale. The work done in the context of this thesis was fundamental in such a crucial time: thorough preparatory studies have been performed on simulated events, with the aim to compare different scenarios and identify the best realignment strategy. Then, the identified strategy has been applied for the first actual alignment measurements performed on cosmic ray data, collected prior to the proton-proton collision run: first, the large misalignments present after the installation have been corrected, then the position of the pixel tracker down to the level of single modules have been measured. As a result, the precision of the hit position measurement has been largely improved before the start of the proton-proton collision run, allowing to collect collision data with an already good tracker alignment.

Chapter 5

Measurement of the Higgs boson decay to a τ lepton pair - Analysis strategy

5.1 Motivation and previous measurements . . . 108
5.2 Measurement at the LHC Run 2 . . . 109
5.3 Data samples . . . 111
5.4 Simulation samples . . . 112
5.5 The $\mu\tau_h$ and $e\tau_h$ channels . . . 117
5.5.1 Event selection for the $\mu\tau_h$ channel . . . 117
5.5.2 Event selection for the $e\tau_h$ channel . . . 124
5.5.3 Higgs boson reconstruction . . . 125
5.5.4 Estimation of background processes . . . 131
5.5.5 Corrections for simulated events . . . 135
5.5.6 Definition of signal categories . . . 149
5.6 The $\tau_h\tau_h$ channel . . . 150
5.7 The $e\mu$ channel . . . 151

In this chapter, the strategy adopted to measure the coupling strength of the Higgs boson to τ leptons at CMS is described, including the event selection, the background estimation techniques and the corrections applied to simulated events. The analysis is performed on 35.9 fb^{-1} of LHC proton-proton collision data recorded with the CMS detector in 2016. The four main final states resulting from the decay of a τ lepton pair are addressed, namely the fully hadronic $\tau_h\tau_h$, the semileptonic $\mu\tau_h$ and $e\tau_h$ and the fully leptonic $e\mu$ final state. In particular, the original contribution of this thesis work to the analysis includes the measurement of the lepton identification and trigger efficiencies and the study of the $\mu\tau_h$ and $e\tau_h$ channels, which are therefore described in greater detail. In the last part of the chapter, the peculiarities of the analysis in the $\tau_h\tau_h$ and

in the $e\mu$ final states are briefly summarised, as the final results presented in the next chapter are obtained from the combination of all four analysed channels.

5.1 Motivation and previous measurements

Measuring the direct coupling of the Higgs boson to fermions is necessary to prove that the mass generation mechanism proceeds as postulated by the Brout-Englert-Higgs mechanism of spontaneous electroweak symmetry breaking, according to which the coupling strength of the scalar boson to fermions is proportional to the fermion mass. Among all decay modes of the Higgs boson to fermions, the most sensitive one at the LHC is $H \to \tau^+\tau^-$, which for the Standard Model Higgs boson of mass $m_H = 125.09\,\text{GeV}$ has a branching ratio of 6.3 %, and is the second preferred decay channel to fermions. The most likely Higgs decay to a b-quark pair has a much larger branching ratio ($\mathcal{B} = 57\,\%$), but at hadron colliders its sensitivity is limited by the very large multi-jet background.

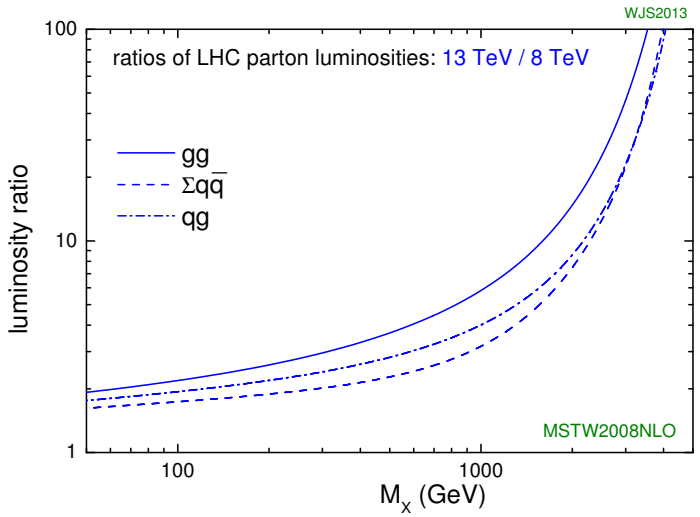
Before the start of the LHC, searches for the Higgs boson decaying to a pair of τ leptons were conducted by the experiments at the LEP and TEVATRON colliders. LEP was a e^+e^- circular collider operating at centre-of-mass energies up to 209 GeV, thus the main Higgs production process was the associated ZH production. The four LEP experiments looked for the Higgs boson in the decay channels with the highest expected branching ratio, namely $b\bar{b}$ and $\tau^+\tau^-$. The ALEPH experiment reported a $3\,\sigma$ excess in the four jet analysis, compatible with the signal expected from a Higgs boson with mass $m_H = 114\,\text{GeV}$ [126], while the DELPHI, L3 and OPAL experiments did not observe any significant excess and set lower limits on the Higgs boson mass from 110 to 114 GeV [127, 128, 129]. Searches for the Higgs boson decaying to τ leptons were then performed also at the TEVATRON, in $p\bar{p}$ collision events at a centre-of-mass energy of 1.96 TeV. The CDF and D0 collaborations, searching for $H \to \tau^+\tau^-$ events in final states with an electron or muon and a hadronically decaying τ, set upper limits on the cross section times branching ratio for Higgs masses from 100 to 150 GeV [130, 131].

At the LHC, both the ATLAS and CMS collaborations searched for the Higgs boson decays to τ leptons, and published their first results on the LHC Run 1 data, corresponding to an integrated luminosity of approximately $5\,\text{fb}^{-1}$ and $20\,\text{fb}^{-1}$ at 7 and 8 TeV, respectively. Both experiments reported an excess of events compatible with the $H \to \tau^+\tau^-$ signal. For ATLAS the observed significance of this excess was $4.5\,\sigma$ (expected $3.4\,\sigma$) and the signal strength was measured to be $\mu = 1.43^{+0.43}_{-0.37}$ [132]. CMS reported a significance of $3.2\,\sigma$ (expected $3.7\,\sigma$) and a signal strength $\mu = 0.78 \pm 0.27$ [133]. The combination of the two results lead to the observation of the Higgs boson decaying to τ leptons with a significance of $5.5\,\sigma$ (expected $5.0\,\sigma$) [49] and a signal strength $\mu = 1.11^{+0.24}_{-0.22}$.

5.2 Measurement at the LHC Run 2

The LHC Run 2 started in 2015 with the first proton-proton collisions at a centre-of-mass energy of 13 TeV, providing the opportunity to explore this energy regime for the first time. Not only the energy but also the instantaneous luminosity increased, allowing the experiments to collect more data. In 2015, CMS recorded data corresponding to an integrated luminosity of $3.8\,\text{fb}^{-1}$, out of which $3.2\,\text{fb}^{-1}$ were certified to be used for physics analyses. In 2016, the performance of the LHC exceeded the expectations and CMS recorded data for $37.8\,\text{fb}^{-1}$, out of which $35.9\,\text{fb}^{-1}$ could be used for physics analyses. Both the higher energy and the increased integrated luminosity of the Run 2 dataset make it absolutely interesting for performing precision measurements as well as for searching for new physics. The expected ratio of parton luminosities at 13 TeV with respect to 8 TeV as a function of the mass of the intermediate resonance M_X is shown in Fig. 5.1. For the 125 GeV Higgs boson, the increase of the centre-of-mass energy from 8 to 13 TeV results in an increase of the Higgs production cross section of a factor ≈ 2.5. Table 5.1 compares the cross sections at 8 TeV and 13 TeV for the main Higgs boson production modes as well as for the Standard Model processes constituting the main background sources for most of the Higgs boson measurements.

Figure 5.1 Expected ratio of LHC parton luminosities for a centre-of-mass energy of 13 TeV with respect to 8 TeV as a function of the mass of the intermediate resonance M_X [134].



Regarding the Higgs boson decaying to a τ lepton pair, it is particularly interesting to analyse the Run 2 dataset searching for such events. In fact, the larger dataset and the higher signal cross section increase the expected sensitivity of this analysis, allowing to improve the precision with respect to the Run 1 result. Measuring the $H \to \tau^+\tau^-$ process with higher precision is of crucial importance, as it is the most sensitive process to study the direct coupling of the Higgs boson to fermions. The latter is interesting

Table 5.1 Comparison of cross sections at 8 and 13 TeV for the different Higgs boson production modes (for m_H = 125 GeV) and for the main Standard Model background processes.

Process	Cross section [pb]		Cross Section Ratio
	$\sqrt{s}$ = 8 TeV	$\sqrt{s}$ = 13 TeV	13 TeV / 8 TeV
Higgs boson production			
ggH	19.24	48.58	2.6
qqH	1.579	3.782	2.4
VH	1.117	2.257	2.0
$t\bar{t}H$	0.129	0.507	3.9
Standard Model background			
W+jets, $W \to \ell\nu$	36703	61527	1.7
$Z/\gamma^* \to \ell\ell$, $m_{\ell\ell} > 50$ GeV	3531.9	5765.4	1.6
$t\bar{t}$	252.8	831.8	3.3

not only in the context of the Standard Model, but also to test new physics models predicting enhanced coupling of the Higgs to τ leptons. Finally, the $H \to \tau^+\tau^-$ process constitutes a background for new physics searches, e.g. the search for minimal supersymmetric (MSSM) partners of the Higgs boson. Along with its promising features, the analysis of the Run 2 dataset also poses some challenges. Due to the high instantaneous luminosity reached by the LHC, the 2016 dataset is characterised by an average of 27 collisions happening in the same bunch crossing. In such a dense event environment, the trigger thresholds need to be increased and the object reconstruction and identification becomes less performant in some cases. Moreover, the increase of the centre-of-mass energy causes some background processes to be enhanced with respect to Run 1, most noticeable is the $t\bar{t}$ pair production cross section increasing of a factor larger than three, and the contribution of the QCD background in the sideband used to constrain the W+jets background.

The CMS collaboration has performed the search for the Higgs boson decay to a τ lepton pair using the full 2016 dataset, corresponding to an integrated luminosity of 35.9 fb^{-1} [135]. The four most likely final states for the decay of τ lepton pairs have been considered, namely the fully hadronic channel, $\tau_h\tau_h$ ($\mathcal{B}$ = 42 %), the semi-hadronic channels $\ell\tau_h$ where $\ell = e, \mu$ ($\mathcal{B}$ = 45 % in total) and the fully leptonic channel $e\mu$ ($\mathcal{B}$ = 6 %). In each channel, events are further divided into three exclusive signal categories which target different Higgs production modes. The e^+e^- and $\mu^+\mu^-$ final states ($\mathcal{B}$ = 6 % in total) have not been included in the first place because of their low sensitivity due to the large Drell-Yan background. The branching fraction of the different final states for τ lepton pairs decays are reported in Table 5.2.

The work done in the context of this thesis focuses mainly on the $e\tau_h$ and $\mu\tau_h$ channels, which are therefore described in greater detail. As the final results are obtained from the statistical combination of all channels, the main features and analysis techniques

used for the $\tau_h\tau_h$ and $e\mu$ channels analysis are also shortly discussed at the end of this chapter.

Table 5.2 Branching fraction of the different final states for τ lepton pairs.

Final state	Branching fraction $\mathcal{B}$
$\tau_h\tau_h$	42 %
$\mu\tau_h$	22 %
$e\tau_h$	23 %
$e\mu$	6 %
$\ell\ell$, $\ell = e, \mu$	6 %

5.3 Data samples

The analysis is based on the proton-proton collision data at a centre-of-mass energy of 13 TeV recorded with the CMS detector in 2016. Only data taken in stable detector conditions and fully certified by the CMS collaboration are taken into account, corresponding to an integrated luminosity of 35.9 fb^{-1}. The datasets used for the analysis in the different channels are listed in Table 5.3, together with their integrated luminosity.

The name of the datasets is composed by three parts, separated by a slash, with the following meaning:

- the first part of the name indicates the group of triggers that have been used to record the events. For the $\mu\tau_h$ and $e\tau_h$ channels, events are selected online requiring the presence of a single isolated muon or electron, respectively. These triggers are part of the `SingleMuon` and `SingleElectron` datasets. The latter also include $\ell + \tau_h$ cross triggers, which are seeded by a lepton and require the presence of a hadronic tau candidate at trigger level. For the $e\mu$ channel, events are triggered by $e + \mu$ cross triggers included in the `MuonEG` dataset, while in the $\tau_h\tau_h$ channel triggers requiring the presence of hadronic taus included in the `Tau` dataset are used;
- the second part of the name identifies the time of the year in which the data have been taken and the detector calibration used for the reconstruction. The 2016 data taking is divided into 8 runs, identified with letters from A to H, depending on the LHC schedule and the CMS detector configuration used for data taking. Run 2016 A is dedicated to commissioning and therefore it is not used for physics analyses;
- the third part is the name of the data format.

Table 5.3 Datasets used for the analysis in each channel, with the corresponding integrated luminosity.

Channel	Dataset	Integrated Luminosity [fb^{-1}]
$\mu\tau_h$	/SingleMuon/Run2016B-03Feb2017-ver2-v2/MINIAOD	5.8
$\mu\tau_h$	/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD	2.6
$\mu\tau_h$	/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD	4.3
$\mu\tau_h$	/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD	4.0
$\mu\tau_h$	/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD	3.1
$\mu\tau_h$	/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD	7.5
$\mu\tau_h$	/SingleMuon/Run2016H-03Feb2017-ver2(3)-v1/MINIAOD	8.6
$e\tau_h$	/SingleElectron/Run2016B-03Feb2017-ver2-v2/MINIAOD	5.8
$e\tau_h$	/SingleElectron/Run2016C-03Feb2017-v1/MINIAOD	2.6
$e\tau_h$	/SingleElectron/Run2016D-03Feb2017-v1/MINIAOD	4.3
$e\tau_h$	/SingleElectron/Run2016E-03Feb2017-v1/MINIAOD	4.0
$e\tau_h$	/SingleElectron/Run2016F-03Feb2017-v1/MINIAOD	3.1
$e\tau_h$	/SingleElectron/Run2016G-03Feb2017-v1/MINIAOD	7.5
$e\tau_h$	/SingleElectron/Run2016H-03Feb2017-ver2(3)-v1/MINIAOD	8.6
$e\mu$	/MuonEG/Run2016B-03Feb2017-ver2-v2/MINIAOD	5.8
$e\mu$	/MuonEG/Run2016C-03Feb2017-v1/MINIAOD	2.6
$e\mu$	/MuonEG/Run2016D-03Feb2017-v1/MINIAOD	4.3
$e\mu$	/MuonEG/Run2016E-03Feb2017-v1/MINIAOD	4.0
$e\mu$	/MuonEG/Run2016F-03Feb2017-v1/MINIAOD	3.1
$e\mu$	/MuonEG/Run2016G-03Feb2017-v1/MINIAOD	7.5
$e\mu$	/MuonEG/Run2016H-03Feb2017-ver2(3)-v1/MINIAOD	8.6
$\tau_h\tau_h$	/Tau/Run2016B-03Feb2017-ver2-v2/MINIAOD	5.8
$\tau_h\tau_h$	/Tau/Run2016C-03Feb2017-v1/MINIAOD	2.6
$\tau_h\tau_h$	/Tau/Run2016D-03Feb2017-v1/MINIAOD	4.3
$\tau_h\tau_h$	/Tau/Run2016E-03Feb2017-v1/MINIAOD	4.0
$\tau_h\tau_h$	/Tau/Run2016F-03Feb2017-v1/MINIAOD	3.1
$\tau_h\tau_h$	/Tau/Run2016G-03Feb2017-v1/MINIAOD	7.5
$\tau_h\tau_h$	/Tau/Run2016H-03Feb2017-ver2(3)-v1/MINIAOD	8.6

5.4 Simulation samples

Simulated events are produced accordingly to the procedure described in Sec. 2.3 and used to model both background and signal production processes relevant for the analysis. The most relevant background processes for the $H \to \tau^+\tau^-$ analysis are the following:

- the Drell-Yan production of opposite sign lepton pairs $Z/\gamma^* \to \ell^+\ell^-$, with $\ell = e$, μ or τ, with leading-order Feynman diagram shown in Fig. 5.2. $Z \to \tau^+\tau^-$ events constitute the main background in all channels, as their topology is very hard to distinguish from the one of the signal, except for the invariant mass of the τ lepton pair, which can be reconstructed with a limited precision due to the presence of neutrinos from τ decays. In the $\mu\tau_h$ and $e\tau_h$ channels, $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events, respectively, contribute to the background when a lepton fakes the τ_h, that happens especially for τ_h candidates reconstructed in the 1-prong and 1-prong+π^0 decay modes;
- the production of W bosons in association with jets, occurring mainly through the process shown in Fig. 5.3. In the $\mu\tau_h$ and $e\tau_h$ channels, W+jets events can mimic

the signal topology when the W boson decays leptonically and the τ_h candidate is faked by a jet;

- the production of events with more than one jet in the final state from QCD processes. This background is the dominant one in the $\tau_h\tau_h$ channel, but due to its very large cross section it has a relevant contribution also in the other channels;
- the production of $t\bar{t}$ pairs, which occurs thorough gluon-gluon fusion and $q\bar{q}$ annihilation, as shown in the diagrams of Fig. 5.4. The decay of $t\bar{t}$ pairs produces two opposite-sign W bosons and two b-jets. This background affects mostly the $e\mu$ channel, but can also mimic the signal topology in the $\mu\tau_h$ and $e\tau_h$ channels, especially in the case of $t\bar{t}$ events in the lepton+jets final state, with a lepton from the W boson decay and a jet faking the τ_h;
- other processes, like single top and diboson production, contribute to a smaller extent to the background composition and are also taken into account.

Several event generators described in Sec. 2.3 are used for the simulation of the different background and signal events. The default set of parton distribution functions (PDFs) is NNPDF30_nlo_as_0118 [136], and the underlying event is always simulated with Pythia8, setup to use the CUETP8M1 tune [89]. The jet matching and merging is performed according to different schemes, depending on the order and software of the matrix element calculation.

Figure 5.2 Feynman diagram of the Drell-Yan process, at leading order.

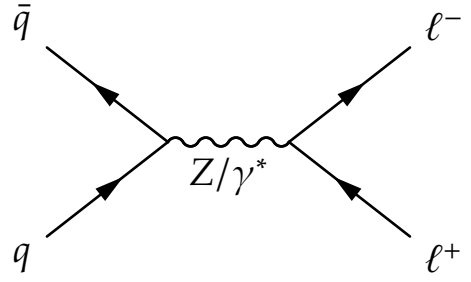


Figure 5.3 Feynman diagrams for the production of W bosons and their leptonic decay, at leading order.

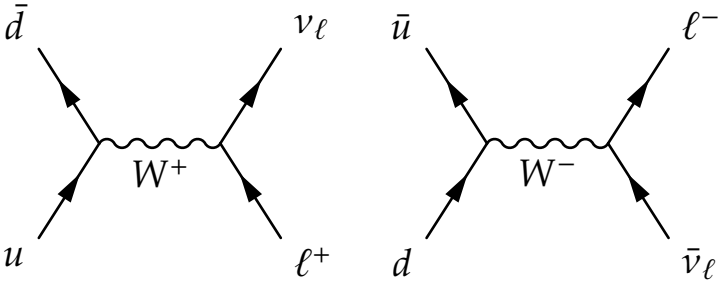
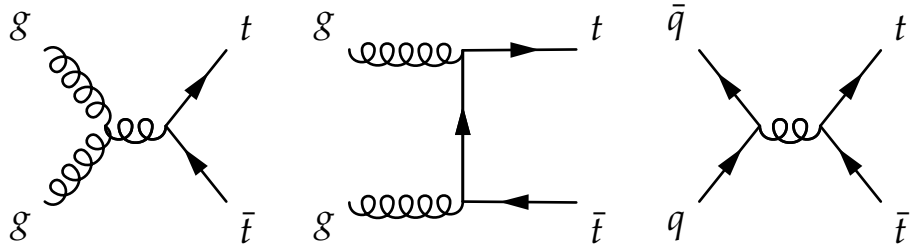


Figure 5.4 Feynman diagrams of $t\bar{t}$ production processes. At the LHC, the dominant production mode is the one initiated by gluon-gluon fusion.



The hard scattering of the Z+jets and W+jets processes are simulated at leading order (LO) with the MADGRAPH5_aMC@NLO generator [80], and the jet matching and merging is performed according to the MLM scheme [86]. The same event generator is used for the simulation of diboson processes at next-to-leading order (NLO), with the FxFx jet matching and merging [87], while single top and $t\bar{t}$ production are simulated with POWHEG [81, 82, 83, 84], with the dedicated jet matching and merging technique [88].

For the signal processes, samples with different Higgs boson masses hypotheses of 110, 120, 125, 130 and 140 GeV are used. The $H \to W^+W^-$ process is considered as a background for the $H \to \tau^+\tau^-$ search. Higgs boson production via gluon-gluon fusion (ggH), vector boson fusion (VBF or equivalently qqH) and associated production with a Z, W^+ or W^- boson are considered, while the $t\bar{t}H$ production mode has a negligible contribution due to its smaller cross section and the selection cuts applied in the analysis. The signal processes are simulated at NLO with the POWHEG generator.

The sets of simulated samples for background and signal processes used in the analysis are given in Tables 5.4 and 5.5, respectively.

Table 5.4 Simulation samples of background processes considered for the analysis, together with their cross sections. For the Drell-Yan and W+jets processes, a k-factor, i.e. the ratio between the inclusive NLO and LO cross section, of 1.21 and 1.16, respectively, is used to correct the LO cross section value.

Process	Generator	Cross section [pb]
Drell-Yan Z/γ^* + jets, $10 < m_{\ell\ell} < 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	18610 (LO)
Drell-Yan Z/γ^* + jets, $m_{\ell\ell} > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	5765.4
Drell-Yan Z/γ^* + 1 jet, $m_{\ell\ell} > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	1012.5 (LO)
Drell-Yan Z/γ^* + 2 jets, $m_{\ell\ell} > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	332.8 (LO)
Drell-Yan Z/γ^* + 3 jets, $m_{\ell\ell} > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	101.8 (LO)
Drell-Yan Z/γ^* + 4 jets, $m_{\ell\ell} > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	54.8 (LO)
Drell-Yan Z/γ^* + jets, $m_{\ell\ell} > 150\,\text{GeV}$	MADGRAPH5_aMC@NLO	6.657 (LO)
$t\bar{t}$	POWHEG	831.76
W+ jets	MADGRAPH5_aMC@NLO	61526.7
W+ 1 jet	MADGRAPH5_aMC@NLO	9644.5 (LO)
W+ 2 jets	MADGRAPH5_aMC@NLO	3144.5 (LO)
W+ 3 jets	MADGRAPH5_aMC@NLO	954.8 (LO)
W+ 4 jets	MADGRAPH5_aMC@NLO	485.6 (LO)
single t-quark, tW channel	POWHEG	35.6
single $\bar{t}$-quark, tW channel	POWHEG	35.6
single t-quark, t channel	POWHEG	136.02
single $\bar{t}$-quark, t channel	POWHEG	80.95
$WZ \to 1\ell 3\nu$	MADGRAPH5_aMC@NLO	3.05
$WZ \to 1\ell 1\nu 2q$	MADGRAPH5_aMC@NLO	10.71
$WZ \to 2\ell 2q$	MADGRAPH5_aMC@NLO	5.595
$WW \to 1\ell 1\nu 2q$	MADGRAPH5_aMC@NLO	1.212
$ZZ \to 2\ell 2q$	MADGRAPH5_aMC@NLO	3.22
$VV \to 2\ell 2\nu$	MADGRAPH5_aMC@NLO	11.95
electroweak W^-+2 jets, $W \to \ell\nu$, $m_W > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	20.25
electroweak W^++2 jets, $W \to \ell\nu$, $m_W > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	25.62
electroweak Z+2 jets, $Z \to \ell\ell$, $m_Z > 50\,\text{GeV}$	MADGRAPH5_aMC@NLO	3.987
electroweak Z+2 jets, $Z \to \nu\nu$	MADGRAPH5_aMC@NLO	10.01
$W\gamma$, $W \to \ell\nu$	MADGRAPH5_aMC@NLO	489.0
$W\gamma^* \to \ell\nu\mu\mu$	MADGRAPH5_aMC@NLO	2.793
$W\gamma^* \to \ell\nu ee$	MADGRAPH5_aMC@NLO	3.526
$gg \to H \to WW \to 2\ell 2\nu$, $m_H = 125\,\text{GeV}$	POWHEG	1.001
$qq \to H \to WW \to 2\ell 2\nu$, $m_H = 125\,\text{GeV}$	POWHEG	0.0858

Table 5.5 Simulation samples of Higgs boson signal production processes included in the analysis, together with their cross sections and branching ratio for the $H \to \tau^+\tau^-$ decay mode. Events have been simulated using the Powheg generator.

Process	Cross section [pb]	$\mathcal{B}(H \to \tau^+\tau^-)$
$ggH, H \to \tau^+\tau^-, m_H = 110\,\text{GeV}$	57.90	0.0791
$ggH, H \to \tau^+\tau^-, m_H = 120\,\text{GeV}$	52.22	0.0698
$ggH, H \to \tau^+\tau^-, m_H = 125\,\text{GeV}$	48.58	0.0627
$ggH, H \to \tau^+\tau^-, m_H = 130\,\text{GeV}$	45.31	0.0541
$ggH, H \to \tau^+\tau^-, m_H = 140\,\text{GeV}$	36.00	0.0360
$qqH, H \to \tau^+\tau^-, m_H = 110\,\text{GeV}$	4.434	0.0791
$qqH, H \to \tau^+\tau^-, m_H = 120\,\text{GeV}$	3.935	0.0698
$qqH, H \to \tau^+\tau^-, m_H = 125\,\text{GeV}$	3.782	0.0627
$qqH, H \to \tau^+\tau^-, m_H = 130\,\text{GeV}$	3.637	0.0541
$qqH, H \to \tau^+\tau^-, m_H = 140\,\text{GeV}$	3.492	0.0360
$ZH, H \to \tau^+\tau^-, m_H = 110\,\text{GeV}$	1.309	0.0791
$ZH, H \to \tau^+\tau^-, m_H = 120\,\text{GeV}$	0.994	0.0698
$ZH, H \to \tau^+\tau^-, m_H = 125\,\text{GeV}$	0.884	0.0627
$ZH, H \to \tau^+\tau^-, m_H = 130\,\text{GeV}$	0.790	0.0541
$ZH, H \to \tau^+\tau^-, m_H = 140\,\text{GeV}$	0.6514	0.0360
$W^+H, H \to \tau^+\tau^-, m_H = 110\,\text{GeV}$	1.335	0.0791
$W^+H, H \to \tau^+\tau^-, m_H = 120\,\text{GeV}$	0.9558	0.0698
$W^+H, H \to \tau^+\tau^-, m_H = 125\,\text{GeV}$	0.8400	0.0627
$W^+H, H \to \tau^+\tau^-, m_H = 130\,\text{GeV}$	0.7414	0.0541
$W^+H, H \to \tau^+\tau^-, m_H = 140\,\text{GeV}$	0.6308	0.0360
$W^-H, H \to \tau^+\tau^-, m_H = 110\,\text{GeV}$	0.8587	0.0791
$W^-H, H \to \tau^+\tau^-, m_H = 120\,\text{GeV}$	0.6092	0.0698
$W^-H, H \to \tau^+\tau^-, m_H = 125\,\text{GeV}$	0.5328	0.0627
$W^-H, H \to \tau^+\tau^-, m_H = 130\,\text{GeV}$	0.4676	0.0541
$W^-H, H \to \tau^+\tau^-, m_H = 140\,\text{GeV}$	0.3940	0.0360

5.5 The $\mu\tau_h$ and $e\tau_h$ channels

5.5.1 Event selection for the $\mu\tau_h$ channel

The $\mu\tau_h$ final state is characterised by the presence of one high-p_T, well-identified, and isolated muon from the leptonic τ decay, and one isolated hadronically decaying τ, with opposite sign charge. No additional isolated leptons are expected to be found in the event. Interesting events are selected in data by single muon triggers, complemented by a $\mu + \tau_h$ cross trigger. The trigger paths with the lowest p_T thresholds being active and unprescaled during the whole 2016 data taking period have been chosen. The single muon trigger selects events with an isolated muon with $p_T > 22\,\text{GeV}$ in the region $|\eta| < 2.1$, while the $\mu + \tau_h$ cross trigger selects events with one muon of $p_T > 19\,\text{GeV}$ and one τ_h with $p_T > 20\,\text{GeV}$. In the offline selection, muons with $p_T > 23\,\text{GeV}$ are required to have fired the single muon trigger, while events with muons of p_T between 20 and 23 GeV should be matched to the muon leg and to the tau leg of the $\mu + \tau_h$ cross trigger. The trigger matching is performed requiring the offline muon (or τ_h in case of the tau leg) to be found within a cone of $\Delta R = 0.5$ from an online object firing the trigger of interest.

Muon isolation

The muon isolation is the amount of energy deposited in a cone of size $\Delta R = 0.4$ around the muon direction. Contributions from charged particles, as well as from neutral hadrons and photons, all reconstructed through the particle flow algorithm, are considered. While for the charged component only particles with tracks originating from the primary vertex are included, this distinction is not possible for neutral particles, which have no track and can not be associated to the production vertex. The contribution from pileup neutral particles is instead estimated as one half of the charged isolation, measured from particle-flow charged particles not originating from the primary vertex. This is motivated by the fact that the fraction of neutral to charged energy in jets from proton-proton collisions, measured from simulation, is about one half. The neutral pileup isolation is subtracted from the neutral isolation component, and this correction is called $\Delta\beta$ correction. If the neutral isolation after the correction is negative, it is not considered in the total isolation. The relative $\Delta\beta$ corrected muon isolation, or simply muon isolation, is therefore defined as follows:

$$R^{\mu}_{\text{Iso}} = \left(\sum p_T^{\text{charged}} + \max\left[0, \sum p_T^{\text{neutral}} + \sum p_T^{\gamma} - p_T^{\text{PU}}\,(\mu)\right]\right) / p_T^{\mu}, \tag{5.1}$$

and it is used to distinguish prompt from non-prompt or fake muons. The same definition is used also for electrons, with the difference that the quantities are measured in a cone of size $\Delta R = 0.3$.

Selection of muons

For the $\mu\tau_h$ channel, muons reconstructed offline with the PF algorithm are required to pass the following quality criteria:

- the muon is required to pass the medium working point of the standard CMS muon identification (described in Sec. 3.2.2);
- the muon should have $p_T > 20\,\text{GeV}$ if it is matched to the cross trigger, or $p_T > 23\,\text{GeV}$ if it is matched to the single muon trigger, and in both cases it should be within the trigger acceptance region $|\eta| < 2.1$;
- the impact parameter of the muon track with respect to the primary vertex should be $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ in the longitudinal direction and in the transverse plane, respectively;
- the relative $\Delta\beta$-corrected muon isolation, calculated according to Eqn. 5.1, should be less than 0.15. This cut corresponds to the tight isolation working point defined by the CMS Muon Object Group.

Selection of hadronic taus

Hadronic tau candidates (τ_h) are reconstructed with the HPS algorithm, as described in Sec. 3.2.6, and they are considered if they satisfy the following conditions:

- the τ_h decay mode is 1-prong, 1-prong+π^0 or 3-prong;
- the τ_h passes the tight working point of the MVA-based tau isolation;
- the τ_h has $p_T > 30\,\text{GeV}$ and $|\eta| < 2.3$;
- the τ_h should pass the tight working point of the anti-muon and the very loose working point of the anti-electron discriminators, to reject the Drell-Yan background;
- for events selected by the $\mu + \tau_h$ cross trigger, the τ_h should be matched to the respective trigger object within a cone with $\Delta R = 0.5$.

Selection of $\mu\tau_h$ pairs

The muon and τ_h candidates must be well separated, with $\Delta R > 0.5$. When more than one muon or τ_h are found that meet the above mentioned selection, all possible pairs are built and only the pair with the most isolated objects is kept. If there is still more than one pair found in the event, only the one made with the highest p_T objects is selected. In this way, exactly one $\mu\tau_h$ pair is considered for each selected event.

Additional leptons veto

Besides the quality cuts on the single objects, the event is rejected if additional leptons are found. In order to reduce the Drell-Yan background contribution, a second-lepton veto is applied. In the $\mu\tau_h$ channel, events containing a second muon with opposite sign charge and separated by at least $\Delta R = 0.15$ with respect to the signal muon, with $p_T > 15\,\text{GeV}$, $|\eta| < 2.4$, $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$, and a relative $\Delta\beta$-corrected isolation smaller than 0.3 are vetoed. Moreover, events are also vetoed if an extra lepton - either a muon or an electron - passing a loose selection is found in the event. This allows to reduce contributions from backgrounds with multi-lepton final states and ensures no overlap between events selected in different channels. The loose muon and loose electron selections for the lepton veto are defined as follows:

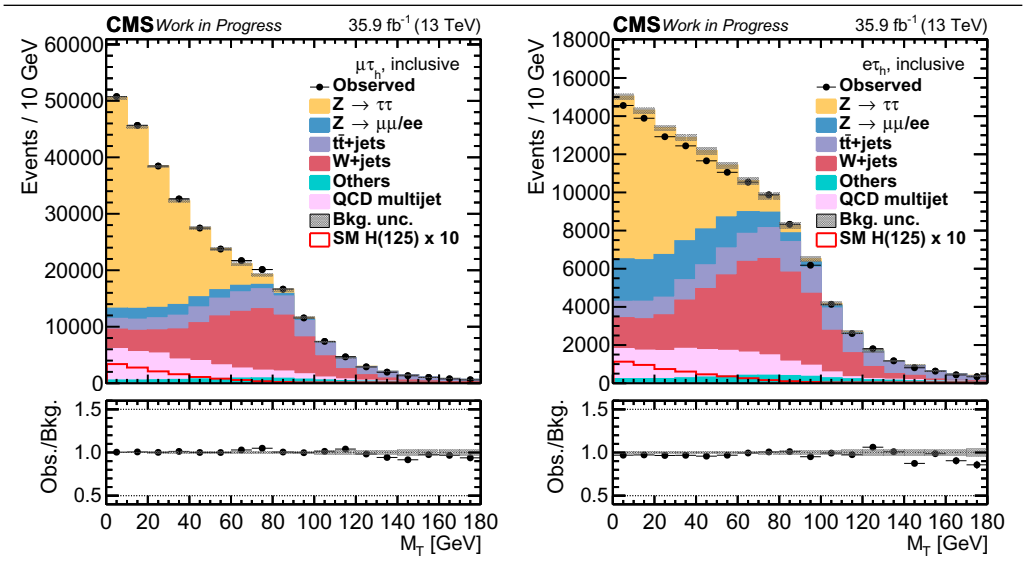
- the muon passes the medium muon identification working point;
- the muon has $p_T > 10\,\text{GeV}$ and $|\eta| < 2.4$;
- the muon track impact parameters, calculated with respect to the primary vertex, satisfy the conditions $d_0 < 0.045\,\text{cm}$ and $d_Z < 0.2\,\text{cm}$;
- the muon is loosely isolated, with $\Delta\beta$-corrected relative isolation $R^{\mu}_{\text{Iso}} < 0.3$.

- The electron passes the 90 % efficiency working point of the MVA-based identification algorithm;
- the electron passes the conversion veto;
- the electron track has ≤ 1 missing inner track hit;
- the electron has $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$;
- the electron track impact parameters, measured with respect to the primary vertex, satisfy the conditions $d_0 < 0.045\,\text{cm}$ and $d_Z < 0.2\,\text{cm}$;
- the electron is loosely isolated, with $\Delta\beta$-corrected relative isolation $R^{e}_{\text{Iso}} < 0.3$.

Topological cuts

The muon and the τ_h from the $H \to \tau^+\tau^-$ signal are expected to be of opposite electric charge (OS), and the transverse invariant mass of the muon+E_T^{miss} system is expected to be smaller than for other processes, like W+jets and $t\bar{t}$ production. In fact, neutrinos produced in the τ decay are nearly collinear with the visible decay products, as the τ energy is much larger than its mass for τ decay products passing the p_T threshold. Conversely, in W+jets events, the high mass of the W boson results in a neutrino

Figure 5.5 Observed and expected distributions of the transverse mass between the muon and the transverse missing energy, calculated according to Eqn. 5.2, in the $\mu\tau_h$ (left) and $e\tau_h$ (right) channels. The uncertainty is statistical only.



approximately opposite to the lepton in the transverse plane, while a jet is misidentified as τ_h. The transverse mass M_T is calculated according to the definition in Eqn. 5.2:

$$M_T = \sqrt{2 p_T^{\mu} E_T^{miss} \left(1 - \cos(\Delta\phi)\right)}, \tag{5.2}$$

where p_T^{μ} and E_T^{miss} are the transverse momentum of the muon and the missing transverse energy (from PF algorithm with type-I corrections applied, see Sec. 3.2.7), and $\Delta\phi$ is the azimuthal angle between the muon and the $\vec{E}_T^{miss}$. The distribution of the M_T variable for events in the $\mu\tau_h$ channel, both for data and simulation, is shown in Fig. 5.5 (left). Events entering the final selection are required to have an opposite sign $\mu\tau_h$ pair and a transverse mass $M_T < 50\,\text{GeV}$. The cut on the transverse mass has been chosen to maximise the analysis sensitivity.

Control distributions

The offline selection applied to select events in the $\mu\tau_h$ channel is summarised in Table 5.6. The distributions of the p_T, η and ϕ variables for the muon and the τ_h, as well as the τ_h mass and the missing transverse energy, after applying the offline selection, including the topological cuts, are shown in Fig. 5.6. The distributions measured in data (black dots) are compared to the ones from simulated events, which are plotted in a stacked histogram including the following contributions:

- the Drell-Yan production of a pair of τ leptons, indicated with the name $Z \to \tau\tau$ and shown in yellow;

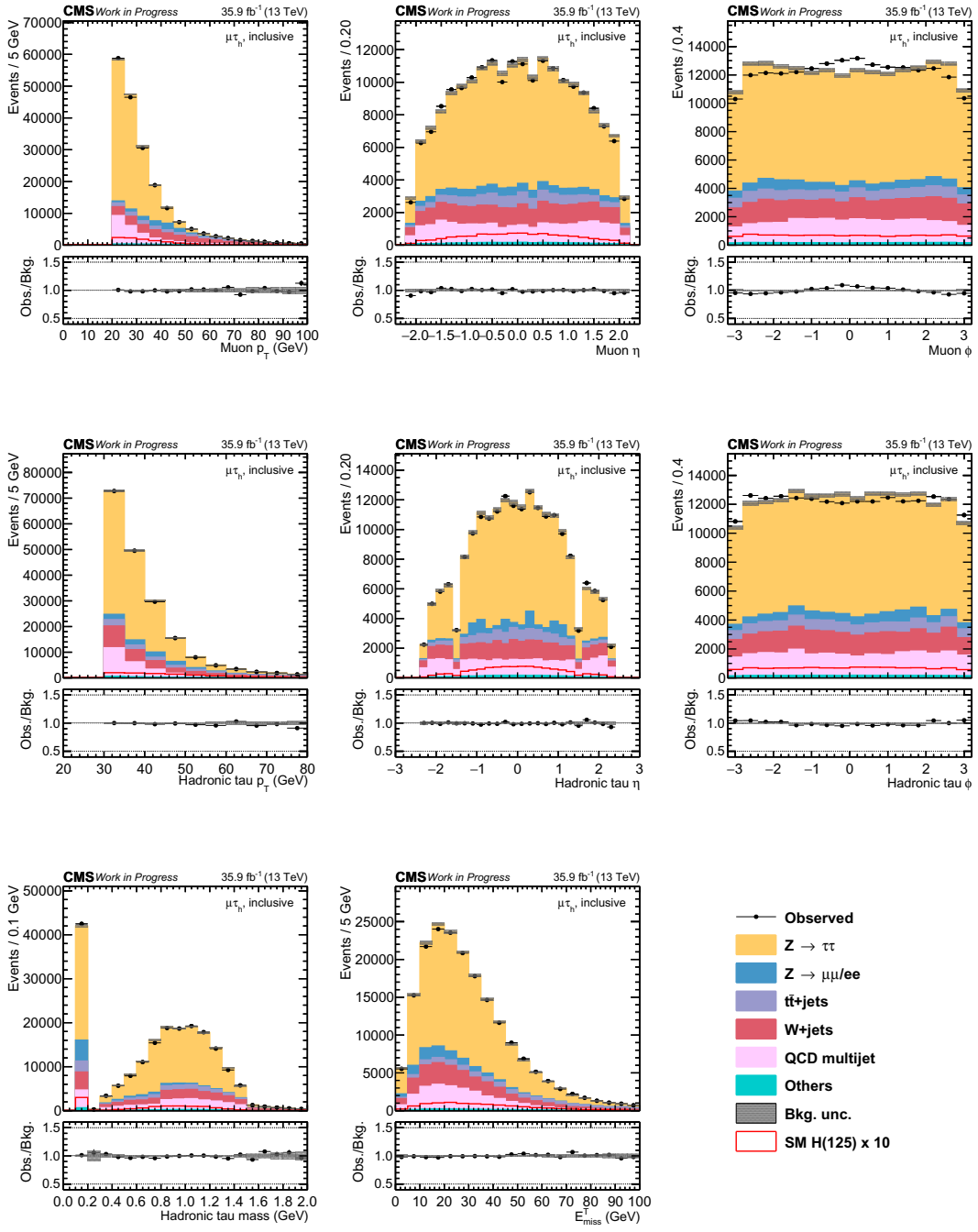
- the Drell-Yan production of a pair of electrons or muons, indicated with the name $Z \to \mu\mu/ee$ and shown in blue;
- the production of $t\bar{t}$ pairs in association with jets, labelled as $t\bar{t}$+jets and shown in violet;
- the production of W bosons in association with jets, labelled as W+jets and shown in red;
- the QCD multi-jet background, estimated from data and shown in pink;
- other minor backgrounds, like the diboson and single top production, are shown in cyan.

The estimation of the different background sources is performed as explained in Sec. 5.5.4, and the corrections described in Sec. 5.5.5 are applied to simulated events. The uncertainty on the simulated events is shown as a grey band and corresponds to the pure statistical uncertainty. The expected Higgs boson signal, in the hypotheses of a Standard Model Higgs boson of mass $m_H = 125\,\text{GeV}$, multiplied by a factor 10, is shown as a red line. The plots shown in Fig. 5.6 are "pre-fit", meaning that the constraints on the backgrounds coming from the final fit are not taken into account. Overall, good agreement between data and simulation is observed.

Table 5.6 Offline event selection applied in the $\mu\tau_h$ channel.

Muon selection	
Trigger	$\mu(19) + \tau_h(20)$ cross trigger or single muon trigger $\mu(22)$ Matching to trigger object within $\Delta R < 0.5$
Kinematics	$p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.1$
Identification	Medium identification working point Muon track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex
Isolation	$\Delta\beta$-corrected relative isolation < 0.15 (see Eqn. 5.1)
Tau selection	
Trigger	Matching to trigger object within $\Delta R < 0.5$ for $\mu + \tau_h$ cross trigger
Kinematics	$p_T > 30\,\text{GeV}, \lvert\eta\rvert < 2.3$
Identification and isolation	HPS reconstruction, 1-prong, 1-prong+π^0 or 3-prong Tight working point of the MVA-based tau isolation Tight working point of the anti-muon discriminator Very loose working point of the anti-electron discriminator
$\mu\tau_h$ pair selection	
	$\Delta R(\mu, \tau_h) > 0.5$ Select the most isolated pair Select the highest p_T pair
Additional leptons veto	
No second muon with	Opposite sign charge, $\Delta R(\mu, \mu) > 0.15$ PF global muon $p_T > 15\,\text{GeV}, \lvert\eta\rvert < 2.4$ Track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex $\Delta\beta$-corrected relative isolation < 0.3
No other muon with	Medium identification working point $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.4$ Track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex $\Delta\beta$-corrected relative isolation < 0.3
No electron with	MVA-based identification (90 % efficiency working point) Passed conversion veto Track with ≤ 1 missing inner track hit $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex $\Delta\beta$-corrected relative isolation < 0.3
Topological cuts	
	μ and τ_h with opposite sign charge $M_T(\mu, E^T_{miss}) < 50\,\text{GeV}$ (see Eqn. 5.2)

Figure 5.6 Pre-fit control distributions for events passing the $\mu\tau_h$ channel inclusive selection, after applying all simulation corrections. Shown are, from left to right, the muon p_T, η and ϕ, the τ_h p_T, η, ϕ and mass, and the missing transverse energy. The uncertainty band corresponds to the pure statistical uncertainty. A common plot legend is shown in the bottom right corner.



5.5.2 Event selection for the $e\tau_h$ channel

The $e\tau_h$ final state is characterised by the presence of one high-p_T well-identified and isolated electron from the leptonic τ decay, and one isolated hadronically decaying τ, having opposite sign charges. No additional isolated leptons are expected to be found in the event. Events are selected online by a single electron trigger with threshold of $p_T > 25\,\text{GeV}$ and coverage up to $|\eta| = 2.1$. The $e + \tau_h$ cross trigger has not been used, as in 2016 the p_T threshold of the electron leg was of 24 GeV, i.e. only 1 GeV lower than the single electron trigger threshold, and the sensitivity gain obtained by including it in the analysis would have been very small. Similarly as for the $\mu\tau_h$ channel, only events with at least one offline electron matched to the a trigger object are selected.

Electron isolation

The isolation of electrons is measured relative to their transverse momentum p_T^e, by summing over the transverse momenta of charged and neutral PF particles in a cone with $\Delta R = 0.3$ around the electron. The contribution to the neutral isolation expected from pileup events is subtracted. The calculation is performed as already explained in the case of muons in Eqn. 5.1.

Selection of electrons

For the $e\tau_h$ channel, electrons reconstructed offline with the PF algorithm must meet the following requirements:

- the electron passes the 80 % efficiency working point of the non-triggering MVA-ID algorithm (described in Sec. 3.2.3);
- the electron track has at most 1 missing inner hit and passes the conversion veto, to reject electrons originating from photon conversion;
- the electron has $p_T > 26\,\text{GeV}$ and $|\eta| < 2.1$, corresponding to the trigger acceptance;
- the relative $\Delta\beta$-corrected isolation (defined in Eqn. 5.1) should be smaller than 0.1.

Selection of hadronic taus

The selection of hadronic tau candidates applied in the $e\tau_h$ channel is the same as the one for the $\mu\tau_h$ channel, and is summarised in Table 5.7. The only difference is given by the cuts on the anti-lepton discriminators: in the $e\tau_h$ channel, the τ_h candidate should pass the loose working point of the anti-muon and the tight working point of the anti-electron discriminators.

Selection of $e\tau_h$ pairs

The logic to select the $e\tau_h$ pairs is the same as for the $\mu\tau_h$ pairs for the $\mu\tau_h$ channel and is summarised in Table 5.7. Again, each selected event contains exactly one $e\tau_h$ pair.

Additional leptons veto

Similarly as for the $\mu\tau_h$ channel, the event is rejected if additional leptons are found. In order to reduce the Drell-Yan background contribution, a second electron veto is applied to events with an additional electron identified with the cut-based identification algorithm, having $p_T > 15\,\text{GeV}$, $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ and a relative $\Delta\beta$-corrected isolation smaller than 0.3. Moreover, events are also vetoed if an extra lepton - either a muon or an electron - which passes a loose selection is found in the event. The criteria applied for the extra lepton veto are the same as for the $\mu\tau_h$ channel.

Topological cuts

As the event topology in the $e\tau_h$ channel is very similar to the one in the $\mu\tau_h$ channel, the same topological cuts are applied, namely the electron and the τ_h should have opposite sign charge and the transverse invariant mass of the electron+E_T^{miss} system, M_T, calculated in analogy to Eqn. 5.2, should be less than 50 GeV. The distribution of the M_T variable for events passing the $e\tau_h$ channel selection is shown in Fig. 5.5 (right).

Control distributions

The offline selection applied to select events in the $e\tau_h$ channel is summarised in Table 5.7. The distributions of the p_T, η and ϕ variables for the electron and the τ_h, as well as the tau mass and the missing transverse energy, after applying the offline selection, including the topological cuts, are shown in Fig. 5.7. The distributions measured in data (black dots) are compared to the ones from simulated events. The grouping and colour scheme of the different background contributions follow the same convention as for the $\mu\tau_h$ channel.

5.5.3 Higgs boson reconstruction

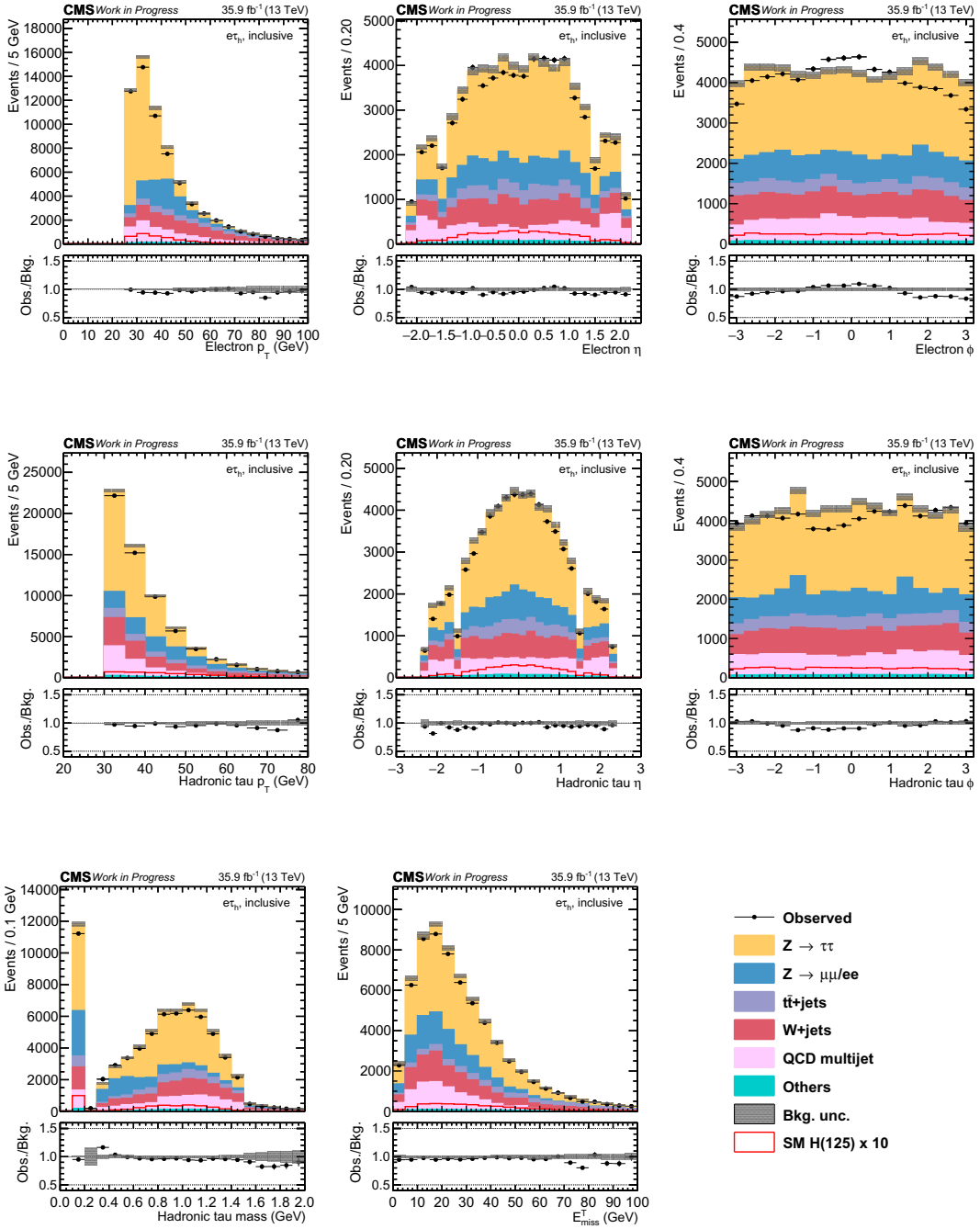
Invariant mass

Due to the presence of neutrinos produced in the decay of a τ lepton - one in the case of hadronic and two in the case of leptonic decay modes - it is not possible to measure the full four-momentum of the $\tau\tau$ system: neutrinos do not interact with the detector and only the sum of the missing energy in the transverse plane E_T^{miss} can be measured.

Table 5.7 Offline event selection applied in the $e\tau_h$ channel.

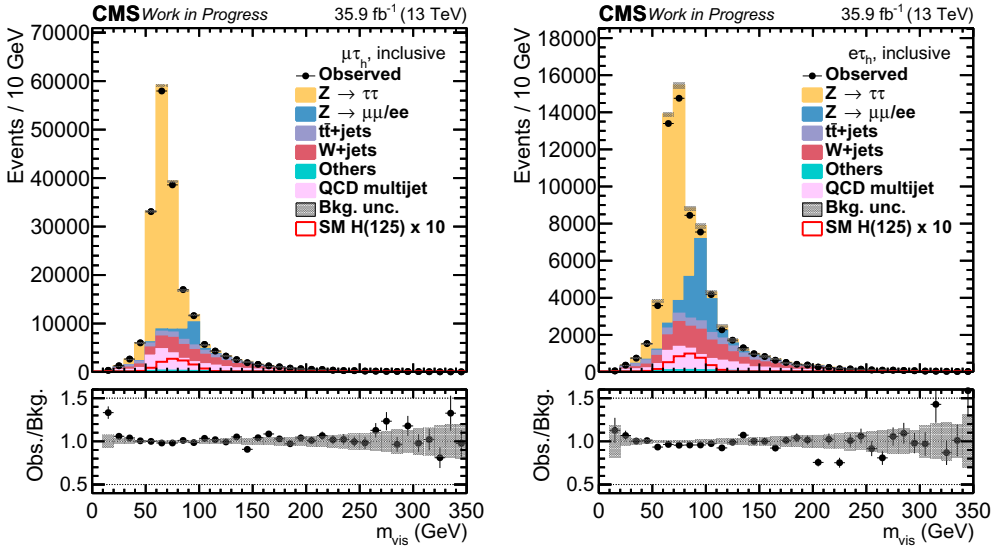
Electron selection	
Trigger	Single electron trigger $e(25)$
	Matching to trigger object within $\Delta R < 0.5$
Kinematics	$p_T > 26\,\text{GeV}, \lvert\eta\rvert < 2.1$
Identification	Non-triggering MVA algorithm (80 % efficiency working point)
	Passing conversion veto
	Track missing inner hit ≤ 1
	Track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex
Isolation	$\Delta\beta$-corrected relative isolation < 0.1 (see Eqn. 5.1)
Tau selection	
Kinematics	$p_T > 30\,\text{GeV}, \lvert\eta\rvert < 2.3$
Identification and isolation	HPS reconstruction, 1-prong, 1-prong+π^0 or 3-prong
	Tight working point of the MVA-based tau isolation
	Tight working point of the anti-electron discriminator
	Loose working point of the anti-muon discriminator
$e\tau_h$ pair selection	
	$\Delta R(e, \tau_h) > 0.5$
	Select the most isolated pair
	Select the highest p_T pair
Additional leptons veto	
No second electron with	Opposite sign charge, $\Delta R(e, e) > 0.15$
	Loose cut-based identification
	$p_T > 15\,\text{GeV}, \lvert\eta\rvert < 2.5$
	Track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex
	$\Delta\beta$-corrected relative isolation < 0.3
No muon with	Medium identification working point
	$p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.4$
	track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex
	$\Delta\beta$-corrected relative isolation < 0.3
No other electron with	MVA-based identification (90 % efficiency working point)
	Passed conversion veto
	Track with ≤ 1 missing inner track hit
	$p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$
	Track $d_Z < 0.2\,\text{cm}$ and $d_0 < 0.045\,\text{cm}$ wrt. primary vertex
	$\Delta\beta$-corrected relative isolation < 0.3
Topological cuts	
	e and τ_h with opposite sign charge
	$M_T(e, E^T_{miss}) < 50\,\text{GeV}$ (see Eqn. 5.2)

Figure 5.7 Pre-fit control distributions for events passing the $e\tau_h$ channel inclusive selection, after applying all simulation corrections. Shown are, from left to right, the electron p_T, η and ϕ, the τ_h p_T, η, ϕ and mass, and the missing transverse energy. The uncertainty band corresponds to the pure statistical uncertainty. A common plot legend is shown in the bottom right corner.



Therefore, the invariant mass of the $\tau\tau$ system, $m_{\tau\tau}$, corresponding to the Higgs boson mass, can not be fully reconstructed. One approximation of the $m_{\tau\tau}$ mass can be obtained by considering only the visible decay products of the τ leptons and computing their invariant mass, the so called visible (invariant) mass m_{vis}. The distribution of the visible mass for events passing the $\mu\tau_h$ (left) and $e\tau_h$ (right) channel selection is shown in Fig. 5.8.

Figure 5.8 Pre-fit distribution of the visible mass for events passing the $\mu\tau_h$ (left) and the $e\tau_h$ (right) channel selection, after applying all simulation corrections. The uncertainty band corresponds to the pure statistical uncertainty.



The visible mass, however, does not provide a good separation between the $Z \to \tau^+\tau^-$ background and the $H \to \tau^+\tau^-$ signal, as the H and Z bosons masses are close to each other. In order to have a more precise estimate of the $\tau\tau$ mass, a dedicated algorithm has been developed in CMS, allowing to derive the most likely invariant mass $m_{\tau\tau}$ on an event-by-event basis, based on a maximum likelihood approach. The algorithm is called SVFit, and it is briefly described in the following, while a detailed explanation can be found in [133, 137].

The kinematic of a hadronic τ decay (τ_h) can be described by six parameters, i.e. the θ and ϕ angles of the visible τ decay products in the τ rest frame, three boost parameters to the laboratory frame, and the visible mass m_{vis} of the decay products. As leptonic τ decays feature two neutrinos in the final state, one additional parameter is needed, which can be chosen to be the invariant mass $m_{\nu\nu}$ of the two-neutrino system. The measured observable, however, is only the four-momentum of the visible decay products in the lab frame, and is not sufficient to fully reconstruct the τ lepton decay. Additional two (three) parameters are needed to fully describe the kinematics of a τ decay to hadrons (leptons). One possible choice of such parameters is the following:

- the fraction x of the τ energy carried by the visible decay products, measured in the laboratory frame;

- the azimuthal angle ϕ of the τ direction in the laboratory frame;
- in case of leptonic τ decays, the invariant mass $m_{\nu\nu}$ of the two-neutrino system. In case of hadronic τ decays, where only one neutrino is produced, this parameter is fixed to be $m_{\nu\nu} \equiv 0$.

The x and y components of the $\vec{E}_T^{miss}$ are also reconstructed and provide a constrain on the momenta of the neutrinos. Based on the available information from the event reconstruction, the di-tau invariant mass $m_{\tau\tau}$ is estimated by combining the measured values of the missing transverse energy, $E_T^{miss,x}$ and $E_T^{miss,y}$, with a likelihood model including the τ decay kinematics and the E_T^{miss} resolution. Matrix elements for τ decays available in the literature for leptonic and hadronic decays are used to define the probability density functions of the unknown parameters x, ϕ and $m_{\nu\nu}$. The possible difference between measured $\vec{E}_T^{miss}$ and neutrino momenta is taken into account assuming Gaussian resolution, and a likelihood term is built which includes the E_T^{miss} covariance matrix. Then, the probability of a certain value of $m_{\tau\tau}$ mass given the measured decay observables is calculated by marginalization of the unknown parameters within their physical boundaries. The procedure is repeated for all values of the mass between 5 and 2000 GeV in steps of 2.5 GeV, and the $m_{\tau\tau}$ mass that maximises the probability is chosen. Figure 5.9 compares the distributions of the visible mass m_{vis} and the $m_{\tau\tau}$ mass obtained with the SVFit algorithm, for simulated $H \to \tau^+\tau^-$ events with $m_H = 125\,\text{GeV}$, showing the achieved improvement in the $m_{\tau\tau}$ mass estimation with respect to m_{vis}. In the Run 1 $H \to \tau^+\tau^-$ search, using the $m_{\tau\tau}$ mass reconstructed by SVFit, instead of the visible mass m_{vis}, improved the analysis sensitivity by 40 %. In the following, the $m_{\tau\tau}$ mass is always reconstructed with the SVFit algorithm.

Transverse momentum

The Higgs boson candidate transverse momentum $p_T(H)$ is obtained from the vectorial sum of the visible τ decay products and of the $\vec{E}_T^{miss}$, as follows:

$$\vec{p}_T(H) = \vec{p}_T(\tau_{1,vis}) + \vec{p}_T(\tau_{2,vis}) + \vec{E}_T^{miss}. \tag{5.3}$$

The distribution of the reconstructed Higgs boson transverse momentum in $\mu\tau_h$ and $e\tau_h$ events is shown in Fig. 5.10, both for data and simulated events.

Figure 5.9 Normalised distributions of the visible mass m_{vis}, and of the SVFit mass, $m_{\tau\tau}$, for simulated events with a Higgs boson of mass $m_H = 125\,\text{GeV}$ decaying to a pair of τ leptons in the $\mu\tau_h$ final state [138].

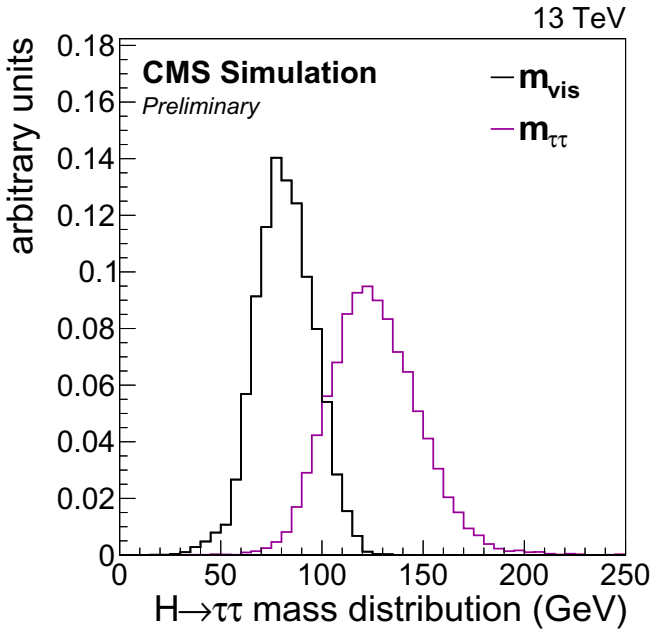
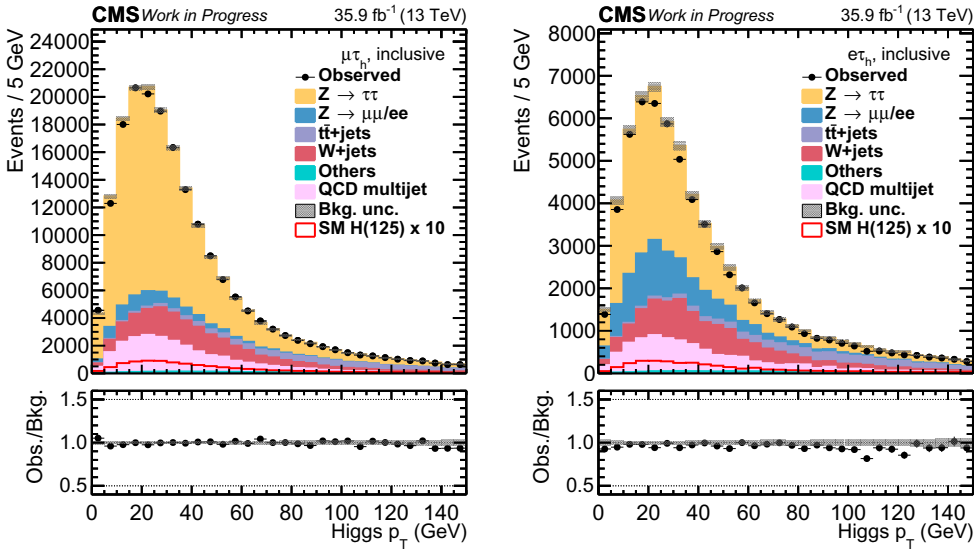


Figure 5.10 Pre-fit distribution of the reconstructed Higgs boson transverse momentum for events passing the $\mu\tau_h$ (left) and the $e\tau_h$ (right) channel selection, after applying all simulation corrections. The uncertainty band corresponds to the pure statistical uncertainty.



5.5.4 Estimation of background processes

The main background processes mimicking the Higgs signal signature in the $\mu\tau_h$ and $e\tau_h$ channels are the following:

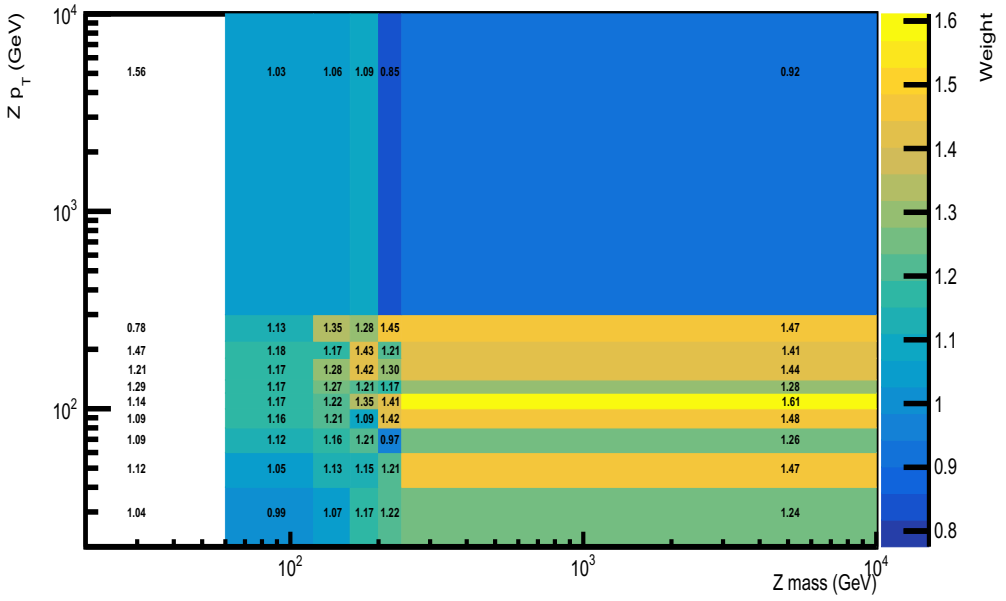
- Drell-Yan $Z/\gamma^* \to \tau^+\tau^-$ is the main background, with a final state identical to the Higgs signal, the main difference being the invariant mass of the di-tau system. Not only $Z/\gamma^* \to \tau^+\tau^- \to \ell\tau_h$ events but also $Z/\gamma^* \to \ell^+\ell^-$ events contribute to the background in the $\ell\tau_h$ channels, when an isolated lepton fakes an hadronic tau - mostly reconstructed in the 1-prong decay mode;
- W+jets production constitutes a background for the $\ell\tau_h$ channels when the leptonic decay of the W boson produces an isolated prompt lepton, and a jet is misidentified as the τ_h candidate;
- QCD multi-jet events, when a jet is misreconstructed as an isolated lepton and another one as a τ_h. Even though the probability for a jet to be reconstructed as an isolated lepton is very small, the contribution of this background is not negligible due to its extremely large cross section at the LHC;
- $t\bar{t}$ pair production, mostly in the semi-leptonic top pair decays, with an isolated lepton from a leptonic W decay and a jet faking the τ_h;
- other backgrounds having a minor contribution are di-boson and single top production, with one ore more leptons in the final state. Also the small contribution from Higgs boson decays to W^+W^- pairs is included in the background for the $H \to \tau^+\tau^-$ search.

In the following, the techniques used to model each background contribution in the $\mu\tau_h$ and $e\tau_h$ channels are explained in detail. In the next section, the corrections applied to the simulated events to achieve a better description of the data are described.

Drell-Yan

The Drell-Yan production of $Z/\gamma^* \to$ decaying to $\tau^+\tau^-$ or $\ell^+\ell^-$ is the largest source of irreducible background for the $H \to \tau^+\tau^-$ search. This background is modelled from simulation with dedicated corrections applied to achieve a better description of the data. Simulated events are split into three groups, depending on the generator-level information about the origin of the reconstructed τ_h, which can be a genuine τ_h, a fake τ_h matched to a prompt lepton (e or μ), or a fake τ_h from a jet. Reconstructed τ_h which can not be matched to any generator-level object are considered as jet fakes. The matching with generator-level objects is established within a cone of $\Delta R < 0.2$. Splitting this background in different contributions allows applying dedicated simulation corrections for each group of events, along with the related uncertainties, as it will be explained in Secs. 5.5.5 and 6.2.

Figure 5.11 Weights applied to the Drell-Yan LO simulation, derived from a $Z \to \mu^+\mu^-$ control region in data, as a function of Z p_T and Z mass $m_{\mu\mu}$. These weights only correct the shape of the distribution and do not affect the yield.



It has been observed in $Z \to \mu^+\mu^-$ events that the LO simulation of the Drell-Yan events does not describe the data well at high Z mass and p_T. Therefore, weights are derived as a function of the generated Z p_T and Z mass $m_{\mu\mu}$ and applied to the simulated Drell-Yan events. The correction is measured in a $Z \to \mu^+\mu^-$ control region defined by selecting events with two opposite-sign, well-identified muons passing the medium identification working point, with $p_T > 25\,\text{GeV}$, and with a relative isolation less than 0.15. The $Z \to \mu^+\mu^-$ distribution in data is obtained by subtracting the W+jets, $t\bar{t}$, and QCD background events. The weights are then derived by comparing the normalised $m_{\mu\mu}$ mass and Z p_T distributions in data and in simulation, and they are applied as a function of the generated Z mass and p_T, correcting the shape of the Drell-Yan simulation, but not its yield. The actual values of the correction are shown in Fig. 5.11. In the VBF category, defined in Sec. 5.5.6, additional weights as a function of the di-jet invariant mass $m(jj)$ are applied to obtain good agreement between data and simulation.

W+jets

The yield of the W+jets background is obtained with a data-driven technique, while the shape is estimated from simulation. To measure the yield of the W+jets background from data, events are selected with the nominal selection but inverting the cut on the transverse mass, by requiring $M_T > 80\,\text{GeV}$. This high-M_T sideband is dominated by W+jets events, but also other processes like QCD and $t\bar{t}$ events contribute (see Fig. 5.5). In order to obtain the yield of the W+jets background in the signal region, characterised

by opposite sign $\ell\tau_h$ pairs with $M_T < 50\,\text{GeV}$, the measured yield in the high-M_T region is extrapolated to the low-M_T region with a factor obtained from simulation:

$$N(W_{low-M_T,data}) = \frac{N(W_{low-M_T,MC})}{N(W_{high-M_T,MC})} \cdot N(W_{high-M_T,data}), \tag{5.4}$$

where $N(W_{low-M_T,data})$ is the determined yield of W+jets events in the signal region, $N(W_{low-M_T,MC})$ and $N(W_{high-M_T,MC})$ are the yields of W+jets events in the low- and high-M_T regions, respectively, obtained from simulation, and $N(W_{high-M_T,data})$ is the yield of W+jets events in the high$-M_T$ sideband in data. The latter is obtained subtracting all expected backgrounds from data as follows:

$$N(W_{high-M_T,data}) = N(data_{high-M_T}) - \sum_{i=DY,t\bar{t},QCD,other} N(i_{high-M_T}). \tag{5.5}$$

In Eqn. 5.5, all the terms entering the sum are taken from simulation, except for the QCD yield, which is measured from data in the same-sign (SS) sideband, subtracting all expected backgrounds:

$$N(QCD_{high-M_T,SS}) = N(data_{high-M_T,SS}) - \sum_{i=DY,t\bar{t},W,other} N(i_{high-M_T,SS}), \tag{5.6}$$

and then extrapolated back to the opposite-sign (OS) region using the factor $\frac{OS}{SS}$ (see QCD estimation in the next paragraph):

$$N(QCD_{high-M_T}) = \frac{OS}{SS} \cdot N(QCD_{high-M_T,SS}). \tag{5.7}$$

The W+jets yield in the same-sign high-M_T sideband, which appears in Eqn. 5.6, is taken purely from simulation for simplicity reasons, and because the effect of neglecting data-driven corrections for this term is expected to be small: the QCD contribution in the high-M_T region is roughly 20 %, and the data-driven correction of the QCD contribution in the opposite sign high-M_T is also of the order of 20 %, meaning that the overall W+jets normalisation would change by only 4 %. The entire procedure is performed separately for each signal category defined in Sec. 5.5.6. The shape of the W+jets background is estimated from simulation. Due to the limited statistical precision of the simulated sample, a relaxed selection is used, where the cut on the lepton relative isolation is relaxed from 0.15 to 0.3 and the working point of the τ_h identification is loosened from tight to medium. The relaxed selection is verified not to bias the shape of the W+jets distributions by performing Kolmogorov-Smirnov tests [139].

QCD

The QCD multi-jet events constitute another important source of reducible background in the $\mu\tau_h$ and $e\tau_h$ channels. This source of background is entirely estimated from data using sideband regions, orthogonal to the signal region and enriched in QCD events.

The nominal yield of QCD events in the signal region, with opposite-sign $\ell\tau_h$ pairs, $N_{QCD,OS}$ is derived from same-sign events, i.e. with the ℓ and τ_h having the same charge, as explained below. First, the number of same-sign QCD events is determined by subtracting the contribution of all other background processes from data:

$$N(QCD_{SS}) = N(data_{SS}) - \sum_{i=DY,W,t\bar{t},others} N(i_{SS}), \tag{5.8}$$

where the Drell-Yan, $t\bar{t}$ and other backgrounds are taken from simulation. As the W+jets background is estimated with a data driven technique as well, the W+jets and QCD background estimations are connected to each other. In particular, in Eqn. 5.8, the yield of same-sign W+jets events in the low-M_T region $N(W_{SS})$ is estimated from simulation and corrected with the following scale factor:

$$W_{SF} = \frac{N(W_{high-M_T,data})}{N(W_{high-M_T,MC})}, \tag{5.9}$$

to account for the difference between the data-driven and the estimation from simulation. Then, the yield of QCD events in the signal region $N(QCD)$ is obtained extrapolating the yield of same-sign QCD events $N(QCD_{SS})$ to the opposite-sign region through the $\frac{OS}{SS}$ factor:

$$N(QCD_{OS}) = N(QCD_{SS}) \cdot \frac{OS}{SS}, \tag{5.10}$$

to take into account differences between the background composition in the same-sign and opposite-sign regions. This $\frac{OS}{SS}$ factor is obtained by comparing the yield of the QCD background in same-sign and opposite-sign events, selected inverting the lepton isolation requirement. For electrons, it is required that $0.1 < R^e_{iso} < 0.3$, while for muons $0.15 < R^\mu_{iso} < 0.3$. The $\frac{OS}{SS}$ extrapolation factor is derived separately for each of the signal categories defined in Sec. 5.5.6 and for each final state, and the measured values are shown in Table 5.8. The shape of the QCD background is taken from Eqn. 5.10 as well, but a relaxed selection as for the W+jets estimation is used to reduce the statistical uncertainty.

Table 5.8 Extrapolation factors from same-sign (SS) to opposite-sign (OS) events for the QCD estimation, for each signal category (see Section 5.5.6) of the $\mu\tau_h$ and $e\tau_h$ channels. The uncertainty includes both the statistical and systematic components.

Channel	Category	OS/SS factor	Uncertainty
$\mu\tau_h$	0-jet	1.07	4 %
$\mu\tau_h$	Boosted	1.06	6 %
$\mu\tau_h$	VBF	1.00	20 %
$e\tau_h$	0-jet	1.00	7 %
$e\tau_h$	Boosted	1.28	10 %
$e\tau_h$	VBF	1.00	20 %

$t\bar{t}$

The production of $t\bar{t}$ pairs is modelled from simulation, and is split into two components, one with genuine τ_h candidates, and the other one with fake τ_h, mostly from jets. In the context of several analyses involving top quark pairs in the final state [140, 141], it has been noticed that the top p_T spectrum is softer in data than in simulation. Corrections have been derived in the context of those analyses to achieve a better description of data by simulated events, and are applied in the form of event-by-event weights. The latter are calculated as a function of the generator level p_T of the t and $\bar{t}$, as follows:

$$w(t\bar{t}) = \sqrt{w(t_1) \cdot w(t_2)}, \tag{5.11}$$

where $w(t_i)$ depends on the p_T of the t and $\bar{t}$:

$$w(t_i) = a + b \cdot e^{b \cdot p_T(t_i)}, \tag{5.12}$$

and $a = 0.0615$ and $b = -0.0005$ have been derived from a fit of the simulation to the data [140, 141].

Other backgrounds

The diboson, single top and electroweak Z boson production processes represent a minor fraction of the background events for the $e\tau_h$ and $\mu\tau_h$ channels and are estimated from simulation.

5.5.5 Corrections for simulated events

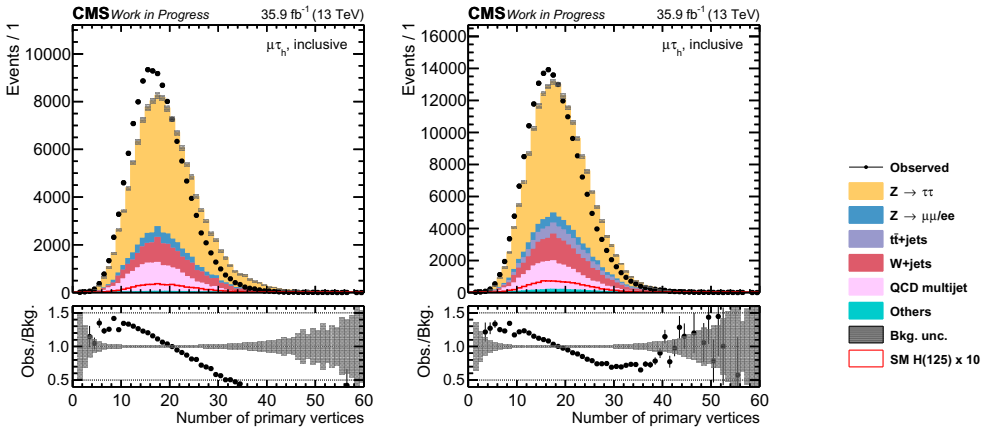
Several data-driven corrections are applied to simulated events in order to obtain a better modelling of the data by the simulation. All the corrections applied in the analysis are described in the following paragraphs.

Pileup reweighing

The effect of pileup is included in the Monte Carlo simulation by generating minimum bias collision events overlapping with the process of interest. However, the distribution of such pileup interactions can not be precisely predicted at the time when the simulated events are produced, as it depends on the operational parameters of the LHC, and the pileup profiles in data and simulation end up to be different from each other. To reduce this difference, the pileup distribution used in the simulation is reweighted to match the one measured in data, and a correction is applied to all simulated events, on an event-by-event basis. The pileup distribution in data is obtained by computing the

instantaneous luminosity for each bunch crossing, and multiplying it by the total proton-proton inelastic cross section. The pileup correction is a pure shape correction, i.e. it does not affect the overall event yield. Figure 5.12 shows the number of reconstructed primary vertices for $\mu\tau_h$ events, before and after applying the pileup correction. After applying the reweighing, the agreement between data and simulation is improved.

Figure 5.12 Number of primary vertices distributions in data and in simulation, before (left) and after (right) applying the pileup reweighing, for events passing the selection summarised in Table 5.6.



Lepton identification and isolation

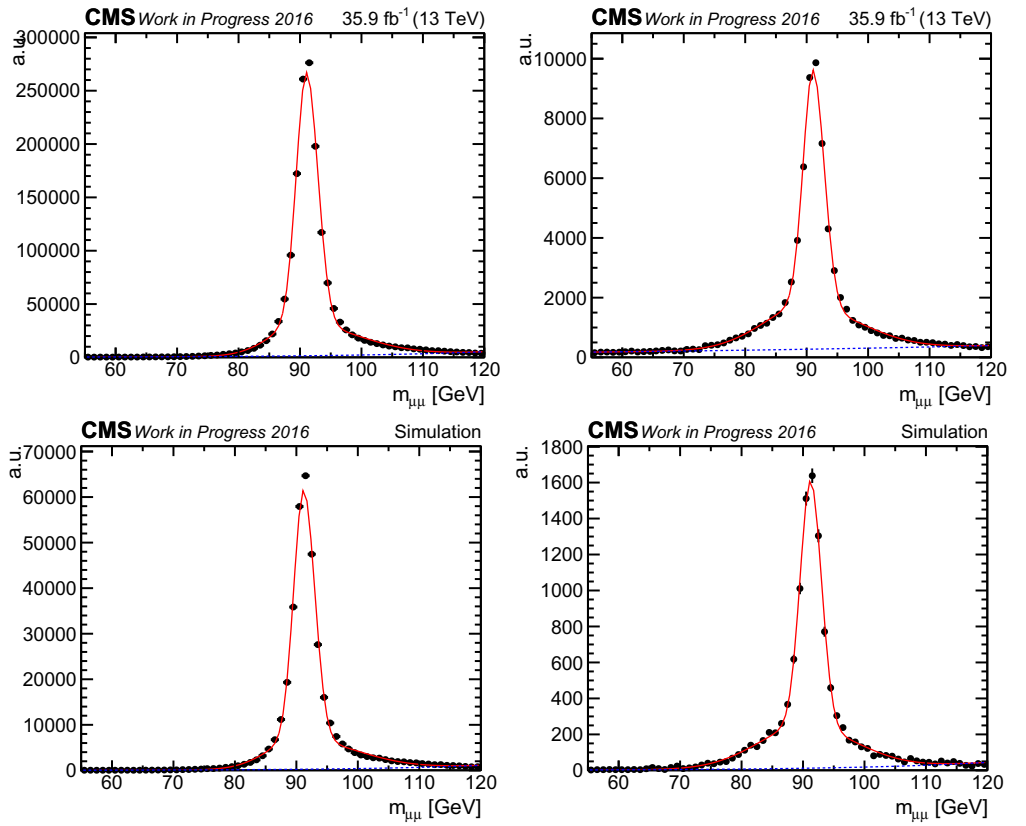
Lepton identification and isolation efficiencies are measured using a tag-and-probe technique [142] on $Z \to \ell^+\ell^-$ events. The main idea of the tag-and-probe method is to use resonances decaying into a pair of identical particles, e.g. muons or electrons, and apply tight cuts on one of them (the "tag"), in order to tag the resonance, and loose cuts on the second one (the "probe") on which the selection criteria are going to be tested. The set of events where a tag-and-probe pair is found is split into two exclusive categories, depending on wether the probe passed or failed the selection criteria. The selection efficiency ϵ is then given by the ratio between the number of passing probes and the number of all probes:

$$\epsilon = \frac{N_{pass}}{N_{pass} + N_{fail}}. \tag{5.13}$$

The tag-and-probe measurement is performed separately on data and on Drell-Yan simulated events, and the efficiency ratio $\epsilon_{data}/\epsilon_{MC}$ is applied as a weight on an event-by-event basis to correct the simulation. Efficiencies are measured in bins of lepton p_T and η.

The selection of the tag-and-probe pairs is defined as follows. The tag lepton must satisfy the kinematic, identification and isolation requirements used in the analysis and it should also be matched to a single lepton trigger object within $\Delta R = 0.5$. Probe leptons are only required to pass the kinematic cuts. Pairs of tag-and-probe leptons are considered when the leptons are same flavour and opposite sign charge, are well

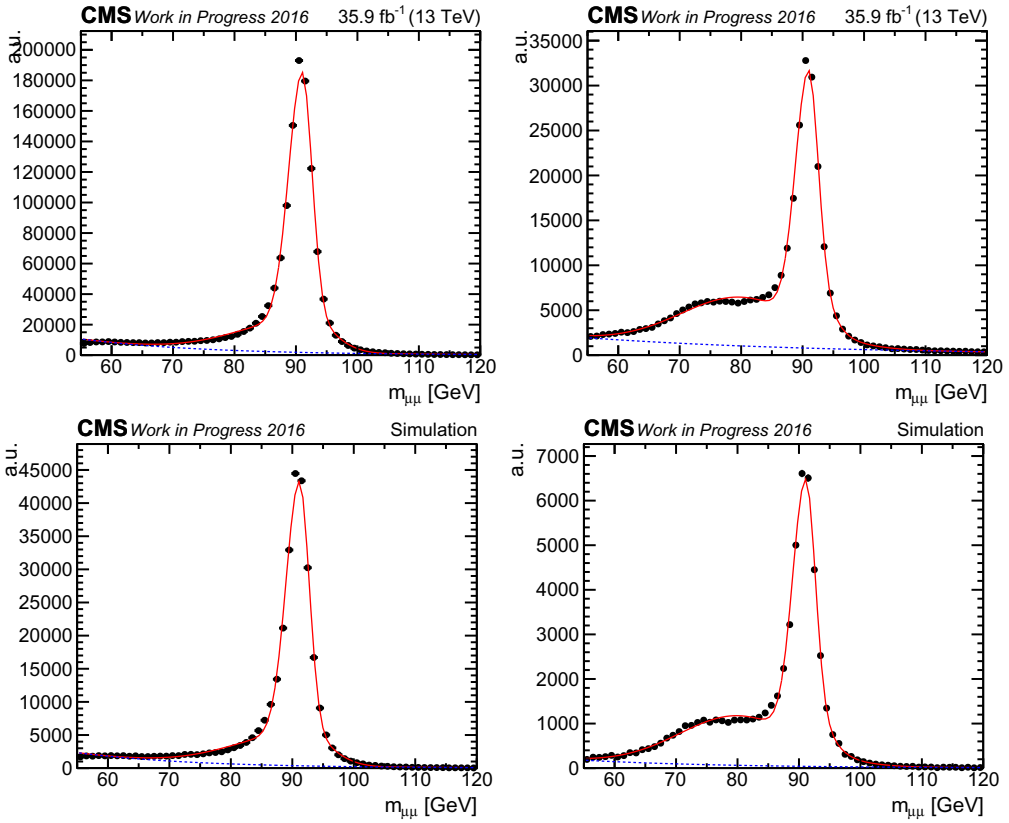
Figure 5.13 Example of fit of the dimuon invariant mass distributions, for muons with $50 < p_T < 60$ GeV and $|\eta| < 0.9$, for data (top) and Drell-Yan simulation (bottom). The distribution of passing probes (left) and of failing probes (right) are shown. The red line indicates the signal+background fit, while the blue dotted line represents the fitted background.



separated ($\Delta R > 0.5$), and their invariant mass $m_{\ell\ell} > 50$ GeV. All combinations of tag-and-probe pairs passing the above mentioned selection are considered. The numbers of passing and failing probes are extracted from a fit to the dilepton invariant mass in a mass window around the Z mass, $75 < m_{\ell\ell} < 105$ GeV, using an exponential function to model the background and two asymmetric Gaussians to model the signal. An example of fitted invariant mass distributions for passing and failing probes, in data and simulation, is shown is Fig. 5.13. A final state radiation component ($Z \to \ell^+\ell^- + \gamma$), modelled as a Gaussian function with its central value shifted to the left of the Z peak, is added to better model the failing probes distribution in the p_T ranges $25 < p_T < 40$ GeV, as it can be seen in Fig. 5.14.

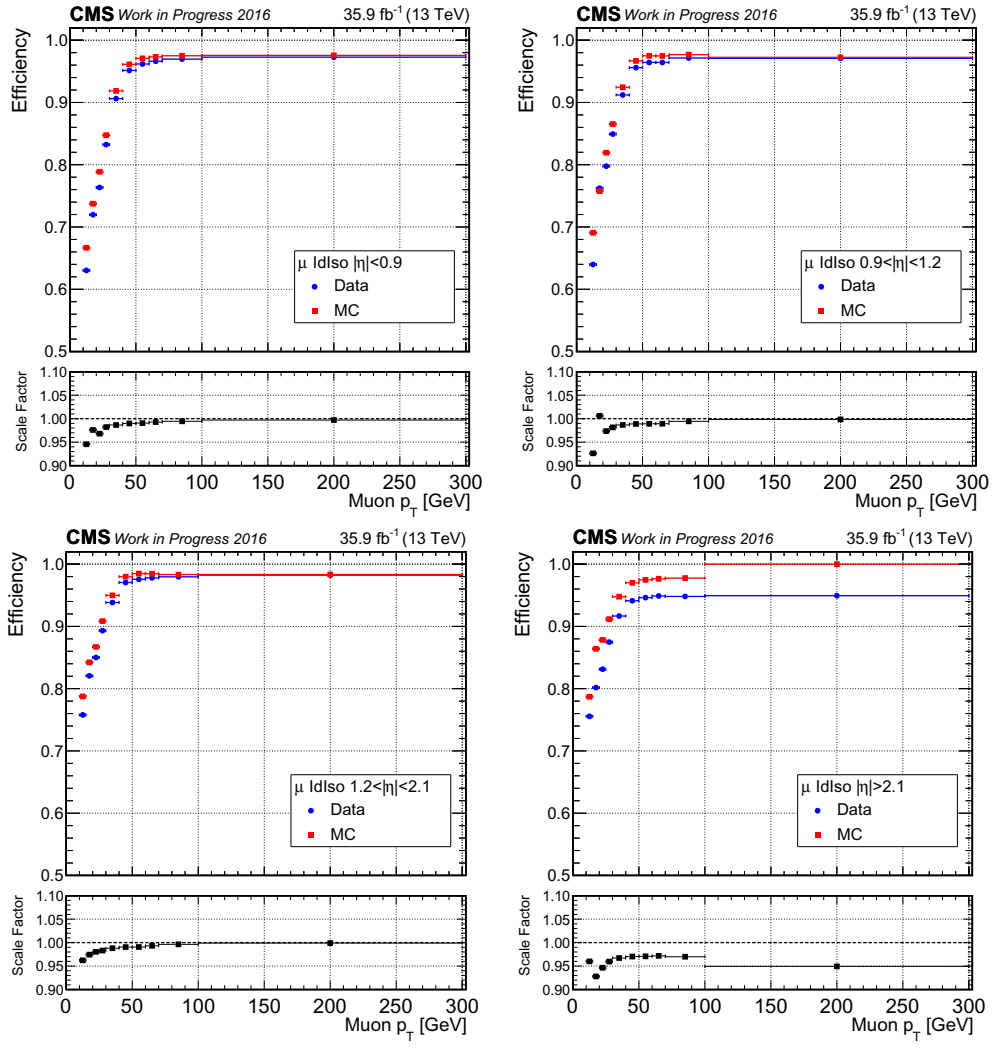
The identification and isolation efficiency measurement is performed in bins of p_T and $|\eta|$ of the probe lepton.
The p_T bin boundaries have been chosen as p_T = [10, 15, 20, 25, 30, 40, 60, > 60] GeV. The binning in $|\eta|$ is different for muons and electrons, as it reflects the structure of the muon system and of the electromagnetic calorimeter. For muons, the binning is $|\eta|$ = [0.0, 0.9, 1.2, 2.1, 2.4], while for electrons $|\eta|$ = [0.0, 1.48, 2.1, 2.5]. The scale factor applied to simulated events is given by the ratio $\epsilon_{data}/\epsilon_{MC}$ for the (p_T, η) bin of the

Figure 5.14 Example of fit of the dimuon invariant mass distributions, for muons with $25 < p_T < 30$ GeV and $|\eta| < 0.9$, for data (top) and Drell-Yan simulation (bottom). The distribution of passing probes (left) and of failing probes (right) are shown. The red line indicates the signal+background fit, while the blue dotted line represents the fitted background. To better fit the distribution of failing probes in this p_T regime, a Gaussian function with its central value shifted to the left of the Z peak is added to the signal template to account for final state radiation events $Z \to \ell\ell + \gamma$.



selected lepton. Figures 5.15 and 5.16 show the efficiencies in data and simulation and their ratio, as a function of the lepton p_T and in the different $|\eta|$ bins, for muons and electrons respectively. Even though the scale factor measurement extends down to lepton p_T of 10 GeV, only the values for muon $p_T > 20$ GeV and electron $p_T > 26$ GeV are actually used in the $\mu\tau_h$ and $e\tau_h$ channels, respectively. In the case of muons, the identification and isolation efficiencies are very similar in data and in simulation, and the corrections are of the order of a few percent depending on the p_T and η bin. A plateau efficiency of 95 % or higher is reached for muon with p_T of about 40 GeV.
In the highest η region, $2.1 < |\eta| < 2.4$ the corrections are generally larger than in the other η regions, over the whole p_T range. In the case of electrons, the identification and isolation efficiencies differ at maximum by 20 % between data and simulation, the largest effects being observed at low p_T and in the highest η bin. Electron identification and isolation efficiencies measure approximately 80 % at the plateau, which is reached for electrons with $p_T > 60$ GeV.

Figure 5.15 Efficiency of the muon identification and isolation criteria used in the $\mu\tau_h$ channel, measured with the tag-and-probe method in the full 2016 dataset (blue circles) and in Drell-Yan simulated events (red squares), as a function of the muon p_T in four different η regions: $|\eta| < 0.9$ (top left), $0.9 < |\eta| < 1.2$ (top right), $1.2 < |\eta| < 2.1$ (bottom left) and $2.1 < |\eta| < 2.4$ (bottom right). The bottom panel in each plot shows the ratio of the efficiency in data and in simulation, which is applied in the analysis as a scale factor to correct simulated events. The last p_T bin includes muons with $p_T > 100\,\text{GeV}$.



Trigger efficiency

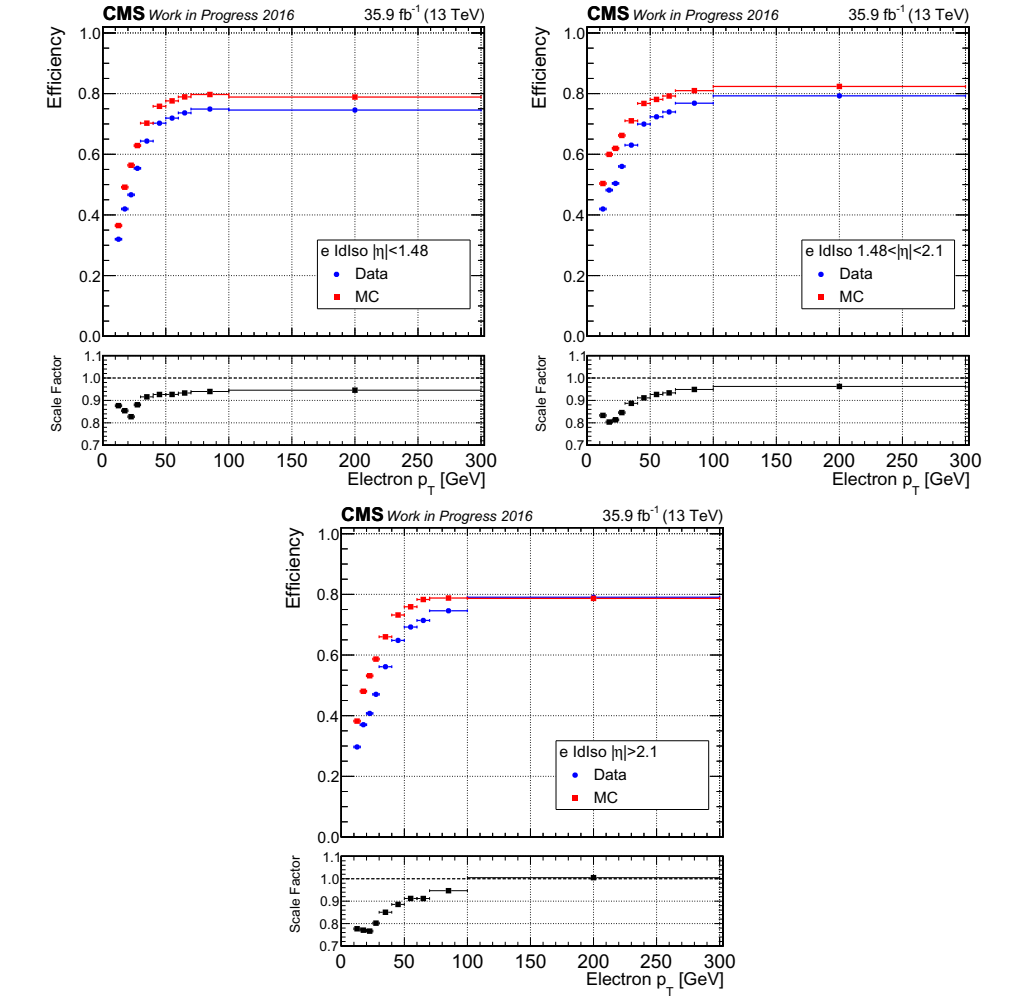
Similarly as lepton identification and isolation efficiencies, trigger efficiencies are measured with the tag-and-probe method in data and in simulation, and an event-by-event weight is obtained and applied to simulated events.

The efficiencies of the single lepton triggers and of the lepton leg of the $\mu + \tau_h$ cross trigger used in the $\mu\tau_h$ and $e\tau_h$ channels are measured from $Z \to \ell^+\ell^-$ events. The requirements for the tag lepton are the same as for the lepton identification and isolation efficiency measurement, while probe leptons must additionally be matched to a trigger object within $\Delta R = 0.5$. The sample of passing probes includes the ones firing the

trigger under consideration, while the failing probes are those not firing that trigger. As all tag-and-probe pairs entering the trigger efficiency measurement are required to pass the identification and isolation criteria, the trigger efficiency scale factor can be applied as an additional independent event weight. For the same reason, the sample of tag-and-probe pairs is very pure in $Z \to \ell^+\ell^-$ events, allowing to simplify the way of counting the numbers of passing and failing probes: the latter are obtained from the histogram integral around the Z mass peak in the range $75 < m_{\ell\ell} < 105\,\text{GeV}$. The small background contribution is estimated from the left and right sides of the Z peak, i.e. $60 < m_{\ell\ell} < 75\,\text{GeV}$ and $105 < m_{\ell\ell} < 120\,\text{GeV}$ respectively.

In the $\mu\tau_h$ channel, a combination of several single muon triggers with a p_T threshold of 22 GeV is used to select events with a muon with $p_T > 23\,\text{GeV}$. The trigger efficiencies

Figure 5.16 Efficiency of the electron identification and isolation criteria used in the $e\tau_h$ channel, measured with the tag-and-probe method in the full 2016 dataset (blue circles) and in Drell-Yan simulated events (red squares), as a function of the electron p_T in three different η regions: $|\eta| < 1.48$ (left), $1.48 < |\eta| < 2.1$ (middle), $2.1 < |\eta| < 2.5$ (right). The bottom panel in each plot shows the ratio of the efficiency in data and in simulation, which is applied in the analysis as a scale factor to correct simulated events. The last p_T bin includes electrons with $p_T > 100\,\text{GeV}$.



shown in the top row of Fig. 5.17 are measured for such combination of triggers. The scale factor applied to simulated events is given by the ratio of the efficiency in data and simulation. Events with muon p_T between 20 and 23 GeV are selected using a $\mu + \tau_h$ trigger, seeded by muons, and with thresholds of $p_T > 19$ GeV for the muon and $p_T > 20$ GeV for the τ_h.The trigger efficiency of the muon leg of the $\mu + \tau_h$ trigger is measured following the same procedure as for the other muon triggers, and the results are shown in the bottom row of Fig. 5.17. The efficiency of the τ_h leg of the $\mu + \tau_h$ trigger is also measured with a tag-and-probe technique on $\mu\tau_h$ events, where the muon is used as a tag and the τ_h as a probe [143]. For events being selected with the $\mu + \tau_h$ cross trigger, the overall trigger scale factor is calculated as the product of the scale factors for each trigger leg. The muon trigger efficiency curves show a steep turn-on curve, with a plateau efficiency in data of $\approx 90\,\%$. The scale factors for the combination of single muon triggers differ from unity of a maximum of 8 %, and the largest differences between data and simulation are seen in the intermediate η region $1.2 < |\eta| < 2.1$.

In the $e\tau_h$ channel, events are selected using a single electron trigger with a p_T threshold of 25 GeV, the measured efficiencies are shown in Fig. 5.18. The turn-on of this trigger is quite slow because the level-one trigger seeds with the lowest thresholds were prescaled in the runs with higher instantaneous luminosity, in order to limit the trigger rate. The electron trigger scale factors range from 0.8 to 1.0, the largest differences between data and simulation are observed in the turn-on region, with a similar behaviour in both η bins. The plateau efficiency is reached for electron $p_T > 60$ GeV and is $\approx 90\,\%$ and $\approx 85\,\%$ in the lower and higher η bins, respectively.

After applying the scale factors for the lepton identification, isolation and trigger, the agreement between data and simulation is improved, as can be seen in Figs. 5.19 and 5.20, comparing the muon and electron p_T and η distributions with and without applying the scale factors.

The lepton identification and trigger efficiencies derived in the context of this thesis were also applied in other CMS analyses performed in the context of the Higgs to tau tau group, which used the same identification and/or trigger requirements for the leptonic channels.

Figure 5.17 Efficiency of the muon triggers used in the $\mu\tau_h$ channel, measured with the tag-and-probe method in the full 2016 dataset (blue circles) and in Drell-Yan simulated events (red squares), as a function of the muon p_T in three different η regions: $|\eta| < 0.9$ (left), $0.9 < |\eta| < 1.2$ (middle), $1.2 < |\eta| < 2.1$ (right). The efficiency curves of the combination of single muon triggers with p_T threshold of 22 GeV are shown on the top row, while the bottom row shows the efficiency curves of the muon leg of the $\mu + \tau_h$ trigger. The bottom panel in each plot shows the ratio of the efficiency in data and in simulation. A fine binning in p_T is used especially around the trigger threshold to better model the turn on of each trigger. A vertical dashed line indicates the nominal trigger threshold. The last p_T bin includes muons with $p_T > 200$ GeV.

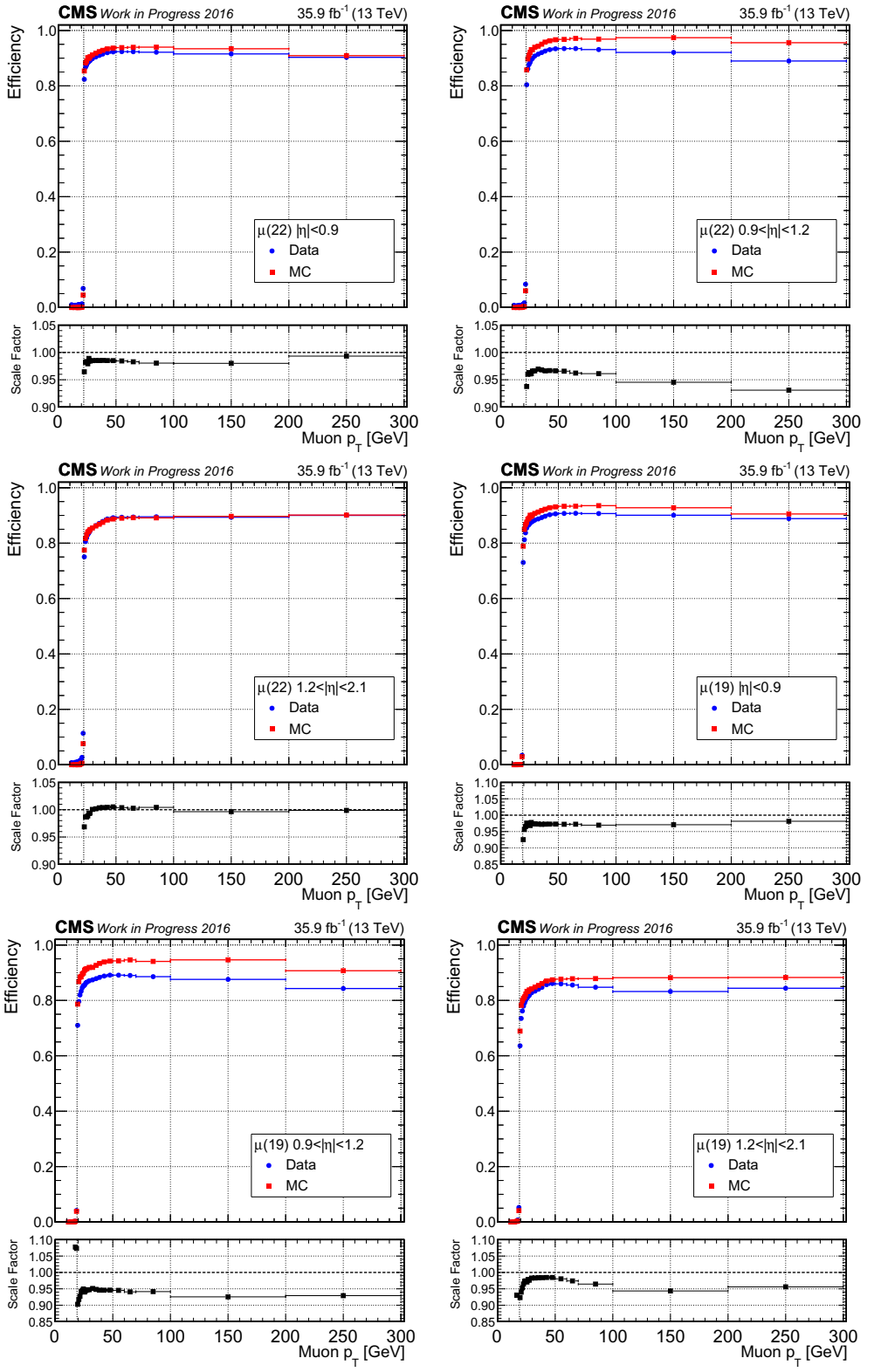


Figure 5.18 Efficiency of the single electron trigger used in the $e\tau_h$ channel, measured with the tag-and-probe method in the full 2016 dataset (blue circles) and in Drell-Yan simulated events (red squares), as a function of the electron p_T in two different η regions: $|\eta| < 1.48$ (left) and $1.48 < |\eta| < 2.1$ (right). This trigger has a p_T threshold of 25 GeV, that is indicated by the vertical dashed line. The bottom panel in each plot shows the ratio of the efficiency in data and in simulation. A fine binning in p_T is used especially around the trigger threshold to better model the trigger turn on. The last p_T bin includes electrons with $p_T > 200$ GeV.

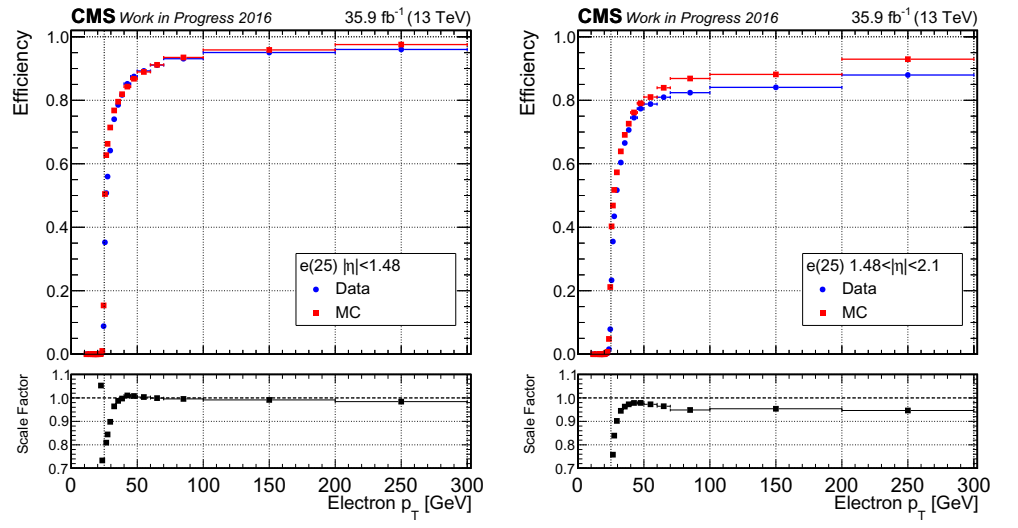


Figure 5.19 Distributions of the muon p_T (top row) and η (bottom row), before (left) and after (right) applying the muon identification, isolation and trigger scale factors. Applying the corrections improve the agreement between data and simulation.

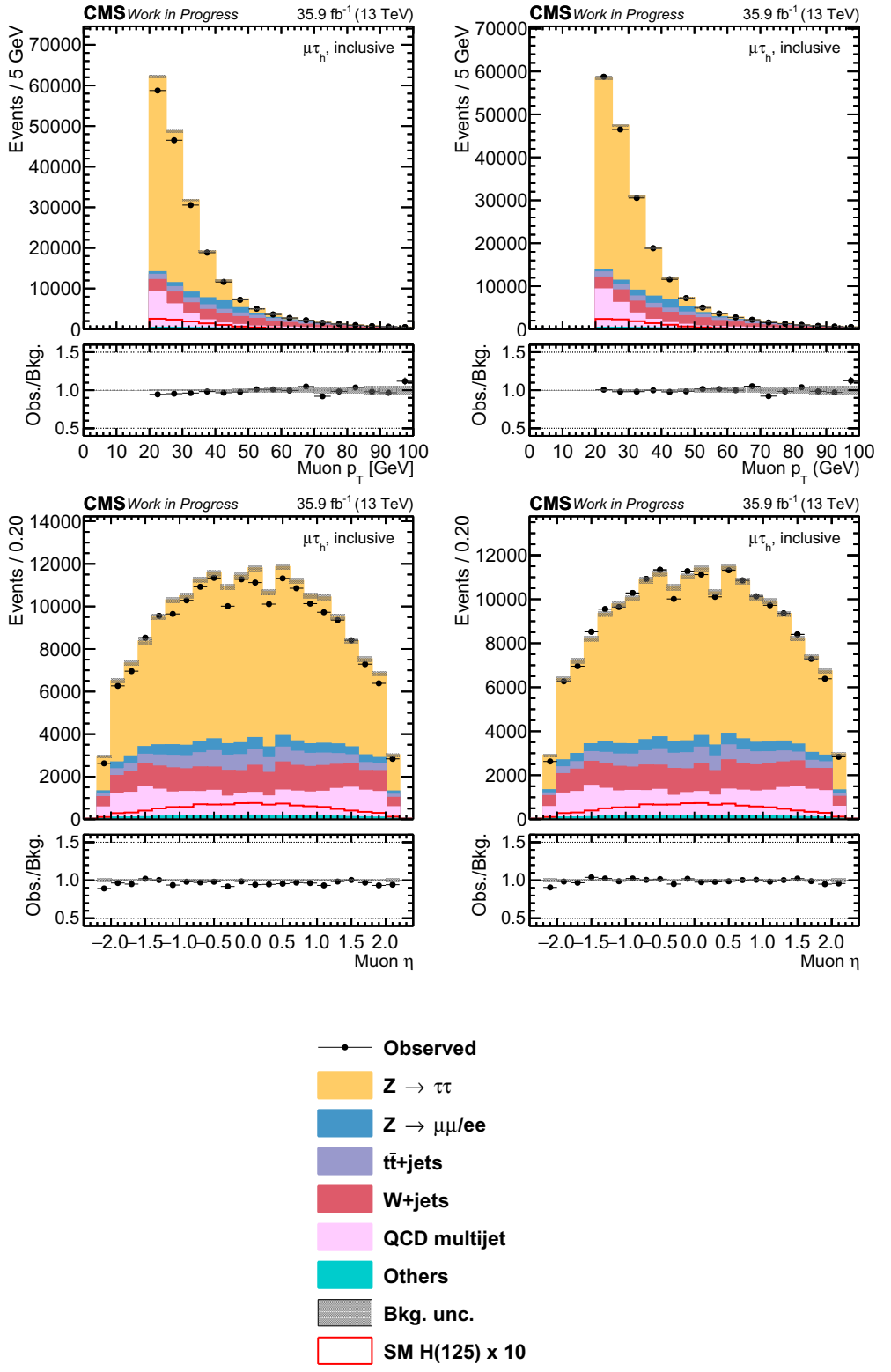
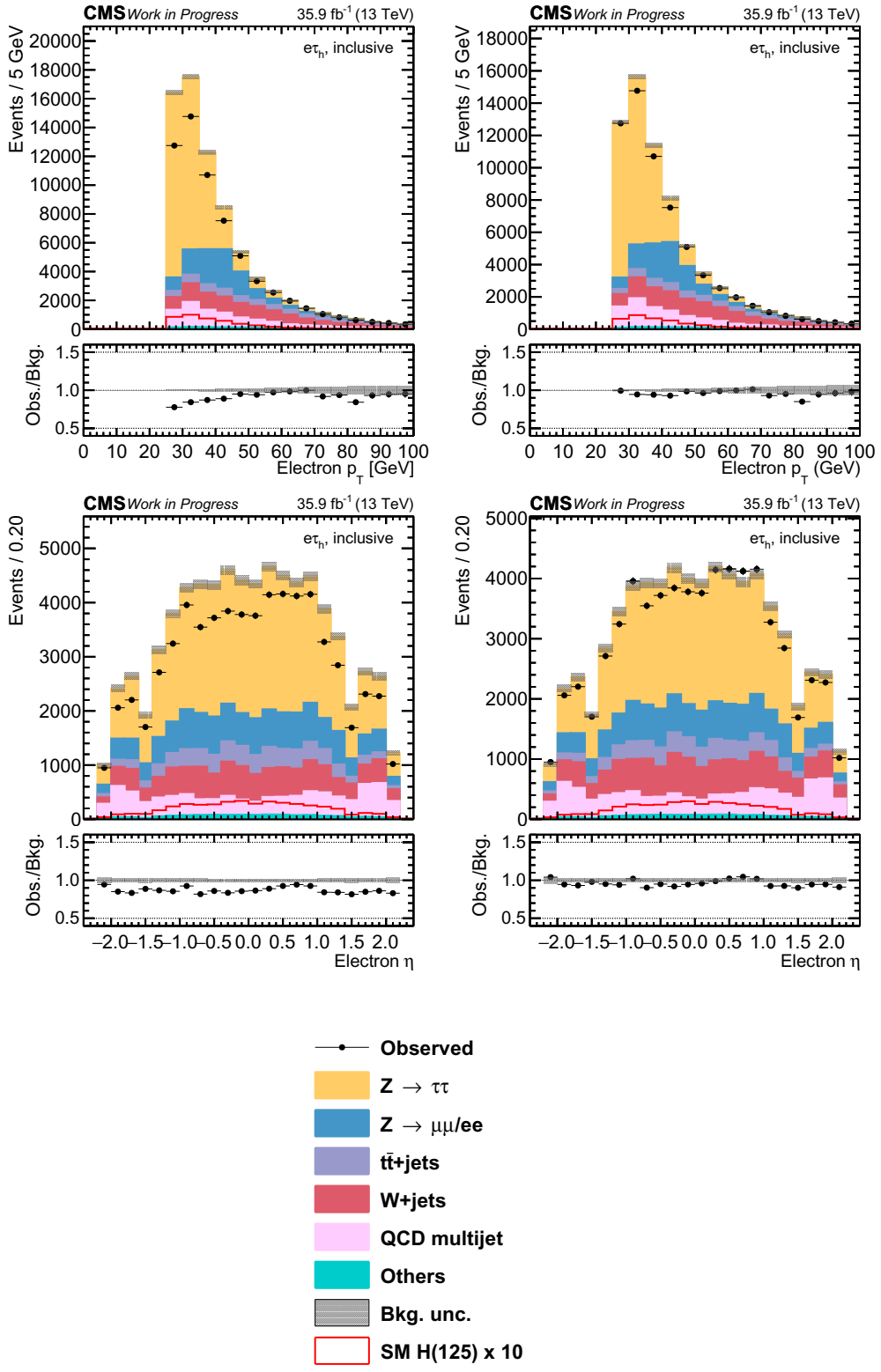


Figure 5.20 Distributions of the electron p_T (top row) and η (bottom row), before (left) and after (right) applying the electron identification, isolation and trigger scale factors. Applying the corrections improve the agreement between data and simulation.



Recoil corrections for $\vec{E}_T^{miss}$

Simulated events where a vector boson recoils against jets are corrected for the different vector boson recoil between data and simulation. The $\vec{E}_T^{miss}$ distribution in $Z \to \ell^+\ell^-$ events, where no genuine missing energy is expected, exhibits some disagreement between data and simulation. The latter arises from a poor modelling of the vector boson recoil $\vec{U}$, defined as the vectorial difference of the measured missing transverse momentum and total transverse momentum of neutrinos originating from the decay of Z, W or H bosons:

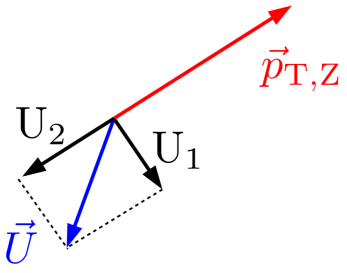
$$\vec{U} = \vec{E}_T^{miss} - \vec{p}_{T,\nu}. \tag{5.14}$$

The vector boson recoil $\vec{U}$ can be precisely measured in $Z \to \mu^+\mu^-$ events, as they do not contain neutrinos and the four-vector of the Z boson can be fully reconstructed. For such events, the recoil vector $\vec{U}$ defined in Eqn. 5.14 can be expressed as

$$\vec{U} = \vec{E}_T^{miss} = -\vec{p}_{T,\mu\mu} - \vec{H}_T, \tag{5.15}$$

where $\vec{H}_T$ is the transverse momentum of hadronic recoil and $\vec{p}_{T,\mu\mu}$ is the transverse momentum of the Z boson decaying to a muon pair (leptonic recoil). In other words, Eqn. 5.15 means that in events without neutrinos, the $\vec{E}_T^{miss}$ comes from unbalanced hadronic and leptonic recoil, due to experimental effects. When such effects are not well simulated, the measured and simulated E_T^{miss} distributions can show discrepancies. In order to correct for that, the vector boson recoil $\vec{U}$ is projected onto the axes parallel ($\vec{U}_1$) and orthogonal ($\vec{U}_2$) to the boson p_T, as represented in Fig. 5.21. The two components $\vec{U}_1$ and $\vec{U}_2$ are measured in data and in simulation from $Z \to \mu^+\mu^-$ events, and corrections are applied such that the mean and width of the U_1 and U_2 distributions in simulated events match the ones observed in data.

Figure 5.21 Representation of the vector boson recoil $\vec{U}$, projected onto the axes parallel ($\vec{U}_1$) and orthogonal ($\vec{U}_2$) to the Z boson p_T.



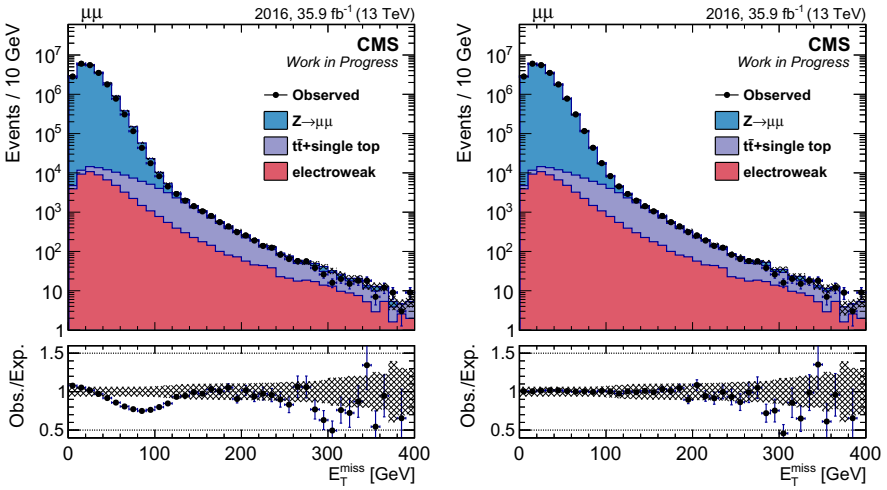
The recoil corrections are derived in bins of the Z boson p_T = 0-10, 10-20, 20-30, 30-50, and > 50 GeV and of the jet multiplicity ($N_{\text{jets}} = 0$, 1 or at least 2). The mean values of the U_1 projection, $\langle U_1 \rangle$, and the resolutions of the $U_{1,2}$ projections, $\sigma(U_{1,2})$, are determined from Gaussian fits. Finally, the corrections to the $U_{1,2}$ variables are computed on a

event-by-event basis, as follows:

$$U_1' = \langle U_1 \rangle_{data} + (U_1 - \langle U_1 \rangle_{MC}) \frac{\sigma(U_1)_{data}}{\sigma(U_1)_{MC}}, \qquad U_2' = U_2 \frac{\sigma(U_2)_{data}}{\sigma(U_2)_{MC}}, \tag{5.16}$$

and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is corrected accordingly using Eqn. 5.14. Recoil corrections are applied to simulated events of Drell-Yan, W+jets and Higgs boson production processes. The improvement of the modelling of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in on the $Z \to \mu^+\mu^-$ events, obtained by applying the recoil corrections, is shown in Fig. 5.22.

Figure 5.22 Distribution of the E_T^{miss} variable in $Z \to \mu\mu$ events, before (left) and after (right) applying recoil corrections. The recoil corrections improve the agreement between data and simulation.



Hadronic tau corrections

Several corrections are applied to the reconstructed τ_h candidates in simulation, to account for differences between data and simulation. In particular, the identification efficiency and the energy scale corrections are derived separately for genuine and lepton-fake τ_h candidates and applied in the analysis. The measurement of the hadronic tau corrections is performed by the CMS group dedicated to the τ_h reconstruction and performance. The measurement of the τ_h performance and data/simulation scale factors performed on the full 2016 dataset are documented in [143]. In the following, the measurements relevant for the $H \to \tau^+\tau^-$ analysis are summarised.

The efficiency of the reconstruction and identification of genuine τ_h candidates has been measured in data and in simulation with a tag-and-probe technique using $Z/\gamma^* \to \tau^+\tau^-$ events in the $\mu\tau_h$ final state. For the tight identification working point, used in the $\mu\tau_h$ and $e\tau_h$ channels of the $H \to \tau^+\tau^-$ analysis, the scale factor is found to be 0.95 ± 0.05 and is applied to Drell-Yan events with a genuine τ_h candidate.

The probability for a lepton (e or μ) to be reconstructed as a τ_h is reduced thanks to the application of dedicated anti-lepton discriminators, and it is typically at the percent level for electrons and at the per-mille level for muons. However, the $\ell \to \tau_h$ misidentification probability is observed not to be well described in simulation. The discrepancies between data and simulation can be reduced by measuring the $\ell \to \tau_h$ misidentification rate in data and in simulation and applying the ratio as a scale factor to correct the simulation. The measurement of the $\ell \to \tau_h$ fake rate is performed in $Z/\gamma* \to \ell^+\ell^-$ events using a tag-and-probe technique. The scale factors are measured as a function of the lepton η and range between 1.1 and 2.4 for muons and between 1.2 and 2 for electrons.

The jet$\to \tau_h$ misidentification probability has been measured in $W \to \mu\nu$+jets events, as a function of the jet η and p_T, and it is found to be compatible between data and simulation within the uncertainties.

The τ_h energy scale is defined as the average reconstructed τ_h energy, with respect to the energy of the visible τ_h decay products at generator level. For genuine τ_h candidates, the simulation correction is derived from a fit to the distribution of sensitive variables, like the reconstructed tau mass m_{τ_h} and m_{vis}, separately for the 1-prong, 1-prong+π^0 and 3-prong τ_h decay modes. Drell-Yan events in the $\mu\tau_h$ final state are used, selected following the baseline of the $H \to \tau^+\tau^-$ analysis, with the difference that the τ_h identification and the m_T cut are tightened to further reject fake τ_h candidates. The shapes of the mass distributions for events with genuine τ_h are shifted between $-6\,\%$ and $+6\,\%$ in steps of $0.1\,\%$ to produce several templates to which the data are compared. The size of the correction amounts to $-1.8\,\%$ for 1-prong, $1.0\,\%$ for 1-prong+π_0 and $0.4\,\%$ for 3-prong τ_h candidates. The overall uncertainty is estimated to be $1.2\,\%$ and takes into account the limited statistics of simulated events with respect to data, the effect of the residual contamination from misidentified τ_h, as well as the choice of the binning and of the fit range.

In the $H \to \tau^+\tau^-$ analysis, additional corrections to the $\ell \to \tau_h$ background are derived separately per τ_h decay mode, from events of the 0-jet category (defined in Sec. 5.5.6) in the $\mu\tau_h$ and $e\tau_h$ channels, with the aim of correcting the yield and energy scale of $\ell \to \tau_h$ simulated events. Such additional corrections are obtained by subtracting all background contributions from the data, except for the $\ell \to \tau_h$ events, and performing Gaussian fits to the visible mass distributions in data and simulation. The procedure is repeated separately per τ_h decay mode and per channel. In the $\mu\tau_h$ channel, the yield corrections are 0.74 ± 0.25 and 1.0 ± 0.25 while in the $e\tau_h$ channel they measure 0.98 ± 0.12 and 1.20 ± 0.12 for the 1-prong and 1-prong+π^0 decay modes, respectively. The energy scale corrections for τ_h candidates originating from muons are measured to be $(-0.2 \pm 1.5)\%$ and $(1.5 \pm 1.5)\%$, while for electrons they are $(0 \pm 3)\%$ and $(9.5 \pm 3)\%$ for the 1-prong and 1-prong+π^0 decay modes. The 3-prong region is essentially free from $\ell \to \tau_h$ fakes and no corrections are applied there. The energy scale corrections for genuine τ_h candidates and for $\ell \to \tau_h$ fakes applied in the $\ell\tau_h$ channels are summarised in Table 5.9.

Table 5.9 Energy scale corrections applied in the analysis for genuine and lepton-fake τ_h candidates, depending on the reconstructed τ_h decay mode.

Genuine τ_h energy scale corrections

τ_h decay mode	1-prong	1-prong+π^0	3-prong	Uncertainty (prefit)
	−1.8 %	1.0 %	0.4 %	1.2 %

$\mu \to \tau_h$ energy scale corrections

τ_h decay mode	1-prong	1-prong+π^0	3-prong	Uncertainty (prefit)
	−0.2 %	1.5 %	n.a.	1.5 %

$e \to \tau_h$ energy scale corrections

τ_h decay mode	1-prong	1-prong+π^0	3-prong	Uncertainty (prefit)
	0.0 %	9.5 %	n.a.	3 %

5.5.6 Definition of signal categories

In order to increase the analysis sensitivity, events reconstructed in each final state are further divided into three mutually exclusive signal categories, targeting different Higgs production mechanisms. The definition of the categories for the $\mu\tau_h$ and $e\tau_h$ channel is the following:

1. **0-jet category**, targeting gluon fusion production, includes events with no loosely identified PF jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 4.7$. Even though the gluon fusion production is the Higgs production mode with the highest cross section at the LHC, this category is the least sensitive due to the large contribution of background events, especially from the irreducible Drell-Yan background, but it is useful to constrain the different background processes because of its high statistical precision;

2. **VBF category**, targeting events where a Higgs boson is produced through vector boson fusion. The main signature of VBF production is the presence of two jets with large invariant mass and a large angular separation. Events with two loosely identified PF jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 4.7$ with invariant mass $m(jj) > 300\,\text{GeV}$ belong to this category. No cut is applied on the pseudorapidity difference between the two jets, because it is strongly correlated with $m(jj)$, and an additional cut on this variable is found not improve the analysis sensitivity;

3. **boosted category**, containing all other events that did not enter one of the previous categories. They come from boosted gluon-gluon fusion Higgs production in association with jets, VBF production where one of the jets is not detected or is too soft to satisfy the $m(jj)$ cut, and Higgs bosons produced in association with a vector boson, when the latter decays hadronically (the leptonic decays are rejected by the additional lepton vetoes).

The repartition of signal events in the three categories depending on their production mode has been estimated from simulation. In the 0-jet category, 98 % of the signal events come from from gluon-gluon fusion. In the VBF category, the fraction of VBF events with respect to the total number of signal events is 57 % when integrating over the whole m_{jj} range, with larger values in high m_{jj} bins. The composition of signal events in the boosted category differs depending on the final state, with 78 to 80 % of signal events coming from gluon-gluon fusion Higgs production in association with jets, about 12 % from VBF production and the remaining 4 to 8 % from VH associated production.

5.6 The $\tau_h\tau_h$ channel

Events in the $\tau_h\tau_h$ final state are selected online using a double-tau trigger with symmetric thresholds of 35 GeV. In the offline event selection, τ_h candidates are required to pass the tight working point of the MVA-based tau identification, and to satisfy the following requirements:

- $p_T > 50$ (40) GeV for the leading (subleading) τ_h;
- $|\eta| < 2.1$;
- pass the very-loose (loose) working point of the anti-electron (muon) discriminator.

At least one $\tau_h\tau_h$ pair with opposite-sign charge τ_h candidates and separated by $\Delta R > 0.5$ must be found to consider the event. If more than one $\tau_h\tau_h$ pair satisfies the requirements, the most isolated one is chosen. Events with loosely identified electrons or muons are rejected.

The estimation of the QCD background is performed in a different way with respect to the $\mu\tau_h$ and $e\tau_h$ channels. For the $\tau_h\tau_h$ channel, in fact, the shape of the QCD background is found to be significantly different in the same-sign and opposite-sign regions, and an alternative method using an opposite-sign sideband with inverted isolation has been developed. The opposite-sign non-isolated region is defined by inverting the isolation requirement on one τ_h and relaxing the isolation cut for the other one, and it is used to derive the QCD shape. The estimated QCD yield in this region, obtained subtracting from data the contribution of other background processes, is scaled by a loose-to-tight isolation extrapolation factor, derived from the same-sign sideband. The uncertainties related to the shape estimation, to the yield in the sideband regions, and to the method itself are taken into account. The W+jets background contribution is rather small in the $\tau_h\tau_h$ final state, and is estimated from simulation.

The definition of the boosted and VBF categories is slightly different for the $\tau_h\tau_h$ channel, due to the background composition being dominated by QCD events: the VBF category

contains events with at least two jets, separated by $|\Delta\eta| > 2.5$ and with Higgs boson candidate $p_T > 100\,\text{GeV}$. The definition of the boosted category is modified accordingly to include events not entering any of the other two categories.

5.7 The $e\mu$ channel

Events in the $e\mu$ channel are selected online thanks to a combination of $e + \mu$ cross triggers with asymmetric thresholds. Namely, two cross triggers are used, which have p_T thresholds of 8 GeV and 23 GeV or 23 GeV and 12 GeV for the muon and electron respectively. The lepton selection differs with respect to the one applied in the $\ell\tau_h$ channels in the following requirements:

- the electron must have $p_T > 13\,\text{GeV}$ and $|\eta| < 2.5$;
- the relative electron isolation calculated inside a cone with radius $R = 0.3$ should be $R^e_{Iso} < 0.15$;
- the muon must have $p_T > 15\,\text{GeV}$ and $|\eta| < 2.4$;
- the relative muon isolation calculated inside a cone with radius $R = 0.4$ should be $R^\mu_{Iso} < 0.2$.

Moreover, each lepton is required to match the respective trigger object of the fired trigger, and the transverse momentum of the lepton firing the highest p_T leg of the trigger must be greater than 24 GeV. Opposite-sign $e\mu$ pairs are selected if $\Delta R(e, \mu) > 0.3$, and only the most-isolated pair in the event is considered.

In the $e\mu$ channel, the second most important background after the $Z/\gamma^* \to \tau^+\tau^-$ production is the $t\bar{t}$ process. The latter can be reduced exploiting the fact that in signal events the direction of neutrinos and of the visible τ decay products are closer to each other than in $t\bar{t}$ or W+jets events. This topological property is quantified by the p_ζ observable, which has been used for the first time by the CDF collaboration [144]. The $t\bar{t}$ background is further suppressed by discarding events with a b-tagged jet.

Similarly as for the $\tau_h\tau_h$ channel, the W+jets background contribution is small and is estimated from simulation.

Chapter 6

Measurement of the Higgs boson decay to a τ lepton pair - Results

6.1 Signal extraction 153
6.2 Uncertainty model 158
6.2.1 Object reconstruction and identification uncertainties 158
6.2.2 Background estimation uncertainties 160
6.2.3 Theory uncertainties on the signal 161
6.2.4 Other uncertainties 162
6.3 Results 164

In this chapter, the results of the measurement of the Higgs boson decay to a τ lepton pair are presented. First, the signal extraction method and the systematic uncertainty model are discussed. Then, the final fitted distributions in every channel and the combined results in terms of the $H \to \tau^+\tau^-$ significance and signal strength are shown.

6.1 Signal extraction

The final results of the analysis are obtained taking into account all four analysed channels, namely $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, and $e\mu$. In each channel, events have been classified into three exclusive signal categories, the 0-jet, boosted and VBF categories, as explained in Sec. 5.5.6. The Higgs boson signal is extracted by performing a binned maximum-likelihood fit to two dimensional distributions, one per channel and per category, for a total of 12 distributions. The variables used to build such distributions have been chosen because of their power to disentangle the signal from the main background processes to maximise the analysis sensitivity. One variable is always the invariant mass of the $\tau\tau$ system, calculated with the SVFIT algorithm (see Sec. 5.5.3), with the exception of the 0-jet category, where the invariant mass of the visible decay products m_{vis} is chosen.

In fact, the use of the SVFit algorithm, which takes O(1 s) to process each event, is not envisaged for the 0-jet category due to the large number of events.

The second variable of the final distributions is chosen as follows:

- for the 0-jet category, the reconstructed τ_h decay mode (1-prong, 1-prong+π^0 or 3-prong) is used in the $\ell\tau_h$ channels, as it allows to separate the $Z \to \ell^+\ell^-$ background, which is more important for the 1-prong and 1-prong+π^0 decay modes, while it is negligible for 3-prong τ_h decays. In the $e\mu$ channel, the muon p_T is used. In the $\tau_h\tau_h$ channel, the one dimensional SVFit mass distribution is used for the final fit, and is not further binned, e.g. as a function of the $\tau_h(p_T)$, because of the quickly falling p_T spectrum of the QCD background;
- for the VBF category, the invariant mass of the two leading jets $m(jj)$ is chosen, as the signal purity increases with $m(jj)$;
- for the boosted category, the reconstructed Higgs boson p_T is used, as the signal purity increases at high Higgs p_T: in the boosted regime, in fact, several background processes, such as QCD, are reduced.

Figure 6.1 shows the two-dimensional distributions of the most important Higgs signal and background processes for the VBF and boosted categories in the $\mu\tau_h$ channel, obtained from simulated events, demonstrating how the signal and background can be disentangled with the chosen variables. Table 6.1 summarises the selection defining the categories and the variables used in the signal extraction in each channel. The observed and expected event yields for the $\mu\tau_h$ and $e\tau_h$ channels, divided per category and per process, are given in Table 6.2.

Table 6.1 Definition of signal categories and observables used to build the 2D distributions for the final fit, for each channel.

Channel	Category	Selection	Observables
$\mu\tau_h$	0-jet	No jet with $p_T > 30\,\text{GeV}$ and $\lvert\eta\rvert < 4.7$	τ_h decay mode, m_{vis}
	VBF	≥ 2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T(H) > 50\,\text{GeV}$, $p_T(\tau_h) > 40\,\text{GeV}$	$m_{jj}, m_{\tau\tau}$
	Boosted	Others	$p_T(H), m_{\tau\tau}$
$e\tau_h$	0-jet	No jet with $p_T > 30\,\text{GeV}$ and $\lvert\eta\rvert < 4.7$	τ_h decay mode, m_{vis}
	VBF	≥ 2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T(H) > 50\,\text{GeV}$	$m_{jj}, m_{\tau\tau}$
	Boosted	Others	$p_T(H), m_{\tau\tau}$
$\tau_h\tau_h$	0-jet	No jet with $p_T > 30\,\text{GeV}$ and $\lvert\eta\rvert < 4.7$	$m_{\tau\tau}$
	VBF	≥ 2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T(H) > 100\,\text{GeV}$, $\Delta\eta_{jj} > 2.5$	$m_{jj}, m_{\tau\tau}$
	Boosted	Others	$p_T(H), m_{\tau\tau}$
$e\mu$	0-jet	No jet with $p_T > 30\,\text{GeV}$ and $\lvert\eta\rvert < 4.7$	$p_T(\mu), m_{vis}$
	VBF	≥ 2 jets, $m_{jj} > 300\,\text{GeV}$	$m_{jj}, m_{\tau\tau}$
	Boosted	Others	$p_T(H), m_{\tau\tau}$

Figure 6.1 Distributions for the signal (left) and for the dominant background processes (right) of the two observables chosen in the 0-jet (top), VBF (centre), and boosted (bottom) categories in the $\mu\tau_h$ channel, demonstrating how the chosen variables help in disentangling the signal from the background. Distributions are obtained from simulated events only, and their binning is not the one used for the final distributions [135].

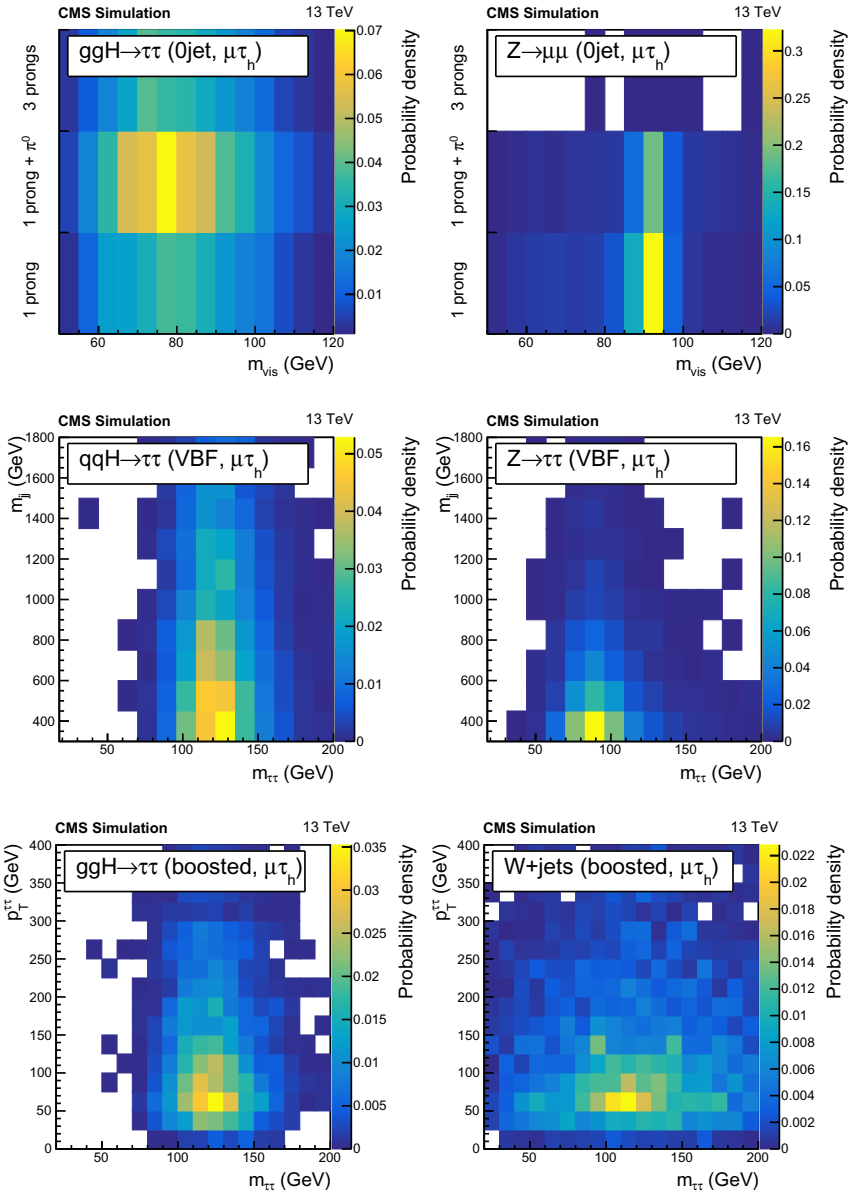


Table 6.2 Expected and observed event yields in the $\mu\tau_h$ and $e\tau_h$ channels, for each category. The expected number of events are given before the final fit. The uncertainty is statistical only.

Event yields in the $\mu\tau_h$ channel

Process	0-jet	Boosted	VBF
$Z \to \tau\tau$	91700±300	32200±140	1250±30
$Z \to \mu\mu/ee$	7100±100	2100±40	37±4
$t\bar{t}$+jets	194±9	9980±60	1100±20
W+jets	15000±300	8200±100	218±9
QCD multijet	16000±200	8250±130	140±20
Other backgrounds	690±10	1960±20	165±5
$ggH, H \to \tau\tau$	570±6	360±5	31.1±1.5
$qqH, H \to \tau\tau$	5.8±0.2	58.4±0.6	29.5±0.4
$WH, H \to \tau\tau$	3.4±0.1	20.0±0.3	1.22±0.07
$ZH, H \to \tau\tau$	1.55±0.08	11.3±0.2	0.79±0.05
Total background (exp.)	130800±500	62700±200	2910±40
Total signal (exp.)	581±6	449±5	62.6±1.6
Observed	129686	62215	3036

Event yields in the $e\tau_h$ channel

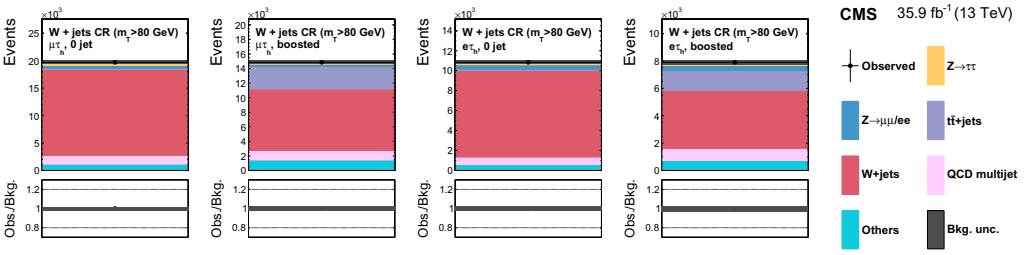
Process	0-jet	Boosted	VBF
$Z \to \tau\tau$	22510±160	9340±70	720±20
$Z \to \mu\mu/ee$	7840±120	2630±50	102±8
$t\bar{t}$+jets	79±6	4160±40	782±17
W+jets	6770±200	3220±50	100±4
QCD multijet	5000±100	3660±100	200±20
Other backgrounds	281±6	815±12	127±4
$ggH, H \to \tau\tau$	187±4	131±3	18.2±1.1
$qqH, H \to \tau\tau$	1.8±0.1	18.9±0.3	16.9±0.3
$WH, H \to \tau\tau$	1.4±0.1	8.2±0.2	0.63±0.05
$ZH, H \to \tau\tau$	0.54±0.04	4.44±0.12	0.36±0.03
Total background (exp.)	42470±300	23820±150	2040±40
Total signal (exp.)	191±4	163±3	36.1±1.2
Observed	41189	22189	2088

Control regions

Several control regions, each enriched in one of the main background processes, are defined and included in the final fit, to further constrain the background normalisation from data.

For the W+jets background, four dedicated control regions are defined, applying the signal region selection for the 0-jet and boosted categories of the $e\tau_h$ and $\mu\tau_h$ channels, but reverting the M_T cut to $M_T > 80\,\text{GeV}$. Figure 6.2 shows the control regions in the 0-jet and boosted categories of the $\mu\tau_h$ and $e\tau_h$ channels. The distributions in the control regions are constituted by a single bin, because they are used uniquely to constrain the normalisation of the W+jets process. In each category, the W+jets background yield in the control region is normalised to the observed yield, in such a way as to obtain perfect agreement between the yields of data and predicted backgrounds in the high-M_T sideband. No dedicated control region is defined for the VBF category, as it is statistically limited. Instead, the constraints obtained in the boosted category are extrapolated to the VBF category of the corresponding final state because the topology of boosted and VBF events is similar.

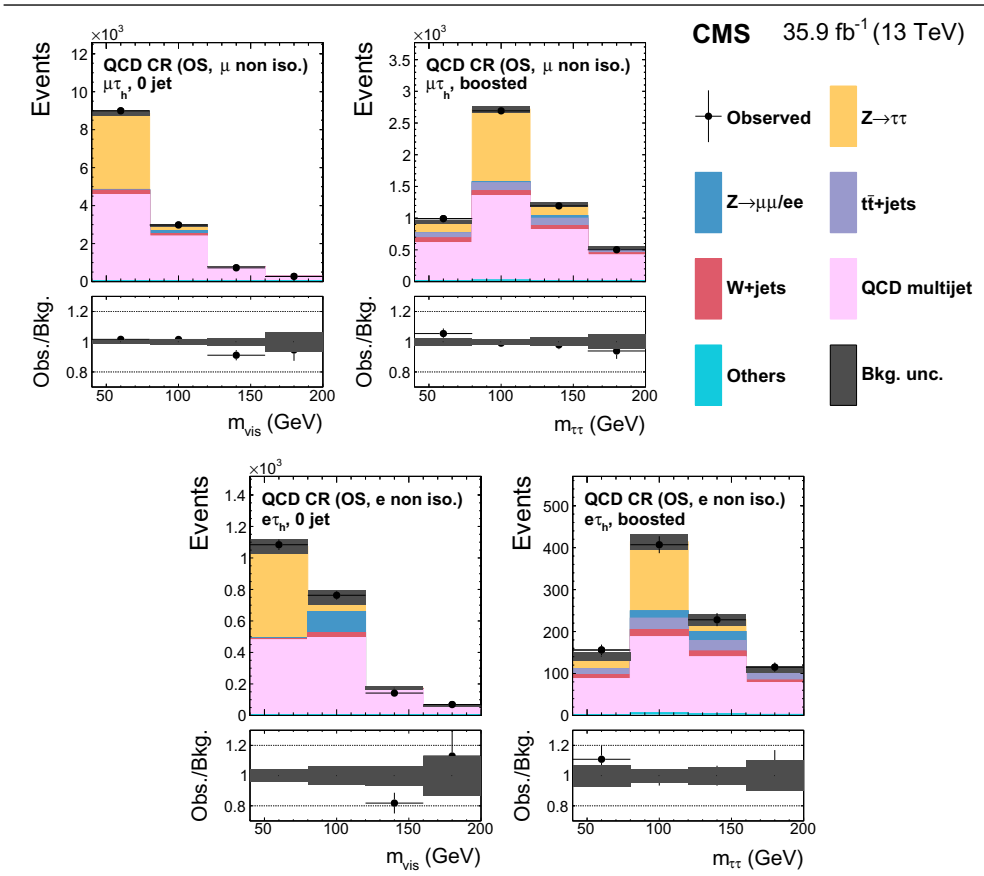
Figure 6.2 Control regions used to derive the normalisation of the W+jets background, for the 0-jet and boosted categories of the $\mu\tau_h$ and $e\tau_h$ channels. During the final fit, the W+jets yield is fitted such to have a perfect agreement between the observed and the expected yields in the high-M_T sideband [135].



The control regions for the QCD background are defined by inverting the lepton isolation requirements. For the $e\tau_h$ and $\mu\tau_h$ channels, a total of four control regions are defined, for the 0-jet and boosted category separately, and the constraints obtained from the boosted control region are extrapolated to the VBF category. The distributions in the QCD control regions after the final fit are shown in Fig. 6.3. The yield of the QCD background in the $\tau_h\tau_h$ channel is controlled by three control regions (one per category), formed by selecting events with opposite-sign τ_h candidates passing relaxed isolation requirements.

The normalisation of the $t\bar{t}$ background is derived from one control region, defined from the $e\mu$ channel selection with inverted p_ζ cut (see Sec. 5.7), and rejecting events with no jets. The constraint on the $t\bar{t}$ background coming from this control region fit is propagated to all channels.

Figure 6.3 Control regions for the QCD background, for the 0-jet and boosted categories of the $\mu\tau_h$ and $e\tau_h$ channels. They are obtained inverting the isolation requirement on the leptons, namely $0.1 < R^e_{iso} < 0.3$ for electrons and $0.15 < R^\mu_{iso} < 0.3$ for muons [135].



6.2 Uncertainty model

The sources of uncertainty taken into account are described below, and the experimental uncertainties are summarised in Table 6.3.

6.2.1 Object reconstruction and identification uncertainties

The largest contribution to the experimental uncertainties comes from uncertainties related to the τ_h reconstruction, which include:

- the τ_h identification uncertainty for genuine τ_h candidates, derived from the measurement of the τ_h identification scale factor, amounts to 5 %. This uncertainty is only partially correlated among final states (4 %) because the anti-lepton discriminator cuts applied are different;
- the τ_h trigger efficiency uncertainty, amounting to 5 % per τ_h candidate;

- the uncertainty on the visible energy scale measurement of genuine τ_h candidates. It amounts to 1.2 %, it is measured inclusively in τ_h p_T and η, and is uncorrelated among decay modes. After performing the final fit, this uncertainty is constrained to 0.3 % for all decay modes. This is explained by several reasons: first, this analysis includes not only the $\mu\tau_h$ final state, but also $\tau_h\tau_h$ and $e\tau_h$, and applies a finer categorisation that isolates regions very pure in $Z \to \tau^+\tau^-$ events. In addition to that, the large majority of τ_h candidates have $p_T < 100\,\text{GeV}$ and are in the barrel region of the detector, therefore their energy can be measured precisely from the track p_T. The uncertainty on the visible τ_h energy scale results in an uncertainty of 5 % on the signal strength;

- the migration of events between reconstructed decay modes. The relative contribution of τ_h candidates reconstructed in a given decay mode is allowed to vary by 3 %, while keeping the total yield unchanged, to account for possible differences in the τ_h reconstruction and identification efficiency depending on the decay mode. This uncertainty is relevant only in the 0-jet category of the $\ell\tau_h$ channels, where events are categorised as a function of the reconstructed τ_h decay mode;

- the τ_h identification uncertainty for $\ell \to \tau_h$ fakes, which amounts to 12 % for $e \to \tau_h$ and 25 % for $\mu \to \tau_h$ fakes, mainly in the $\ell\tau_h$ channels. In the 0-jet categories of the $\ell\tau_h$ channels the final fit is performed on the m_{vis} distribution separately for each τ_h decay mode, allowing to reduce this uncertainty to about 5 %;

- the τ_h energy scale uncertainty for $\ell \to \tau_h$ fakes, being estimated as 1.5 % for muons and 3 % for electrons. It is uncorrelated between τ_h decay modes and is constrained to about one third of its initial value after the final fit;

- the identification efficiency uncertainty for τ_h originating from jets. It is estimated from W+jets events with a highly virtual W boson to probe the τ_h high-p_T region. As the jet $\to \tau_h$ misidentification rate could be not well modelled in simulations, an uncertainty increasing linearly by 20 %/100 GeV, up to a maximum of 40 %, is applied to τ_h candidates matched to jets.

Other experimental uncertainties are related to the reconstruction and identification of leptons, jets and missing transverse energy:

- the uncertainty on the lepton identification, isolation and trigger efficiency is estimated to be 2 % for both muons and electrons;

- the uncertainty on the electron energy scale amounts to 1 % in the barrel and 2.5 % in the endcaps, and is applied in the $e\mu$ channel, where it is propagated to the shape of the mass distributions. In the $e\tau_h$ channel, this uncertainty is not taken into account because it is smaller than the τ_h energy scale uncertainty;

- the uncertainty on the jet energy scale. It affects the $m(jj)$ and the number of jets variables, and therefore the repartition of events among the categories and the shape of the distributions in the VBF category. The jet energy scale uncertainty is measured event-by-event as a function of the jet p_T and η, and it consists of 27 different components [145], which are treated as uncorrelated sources to avoid jets in a given region constraining other independent jets;

- the uncertainty on the b-jet veto efficiency, applied in the $e\mu$ channel, amounts to a maximum of 5 % for $t\bar{t}$ events;

- the scale uncertainties on the missing energy [146], which are computed on an event-by-event basis, and affect both the event yield (through the M_T cut) and the shape of the $m_{\tau\tau}$ mass distribution. The missing energy scale uncertainties include two different components: the clustered energy, coming from the jet energy corrections uncertainty, and the uncertainty on the unclustered energy deposits in the tracker, ECAL, HCAL and forward calorimeters. The latter is estimated by considering each object in the event (jets, electrons, muons, taus, photons and PF candidates not within jets) and by varying its momentum within the scale and resolution uncertainties. Overall, the missing energy scale uncertainties are responsible for an uncertainty on the signal strength of about 10 %.

6.2.2 Background estimation uncertainties

The following background components are affected by uncertainties arising from the background estimation methods described in Sec. 5.5.4:

- the $Z \to \tau^+\tau^-$ background is estimated from simulation with data-driven corrections. Both the yield and the shape of this background are allowed to vary to account for the uncertainty on the applied simulation corrections. The uncertainty on the Z boson p_T-weights, derived as a function of the Z mass and p_T, is estimated to be 10 %. In the VBF category, additional weights are applied as a function of $m(jj)$. Such weights have been observed to vary when changing the muon p_T threshold used to select $Z \to \mu^+\mu^-$ events, and uncertainties up to 20 % are applied as a function of $m(jj)$ to cover these differences;

- the shape of the W+jets background in the $\ell\tau_h$ channels is estimated from simulation, while its yield is measured in data from dedicated control regions included in the fit. Uncertainties on the yield include the statistical uncertainty of data and simulation, and the systematic uncertainty due to the contamination of other background processes in the control regions. An additional uncertainty ranging between 5 and 10 % takes into account the extrapolation of the constraints from the high-M_T sideband to the low-M_T signal regions. The latter is measured by comparing the M_T distributions of simulated and observed $Z \to \mu^+\mu^-$ events where one of the muons has been removed and replaced by missing transverse energy,

to mimic W+jets events. In the $\tau_h\tau_h$ and $e\mu$ channels, the W+jets background is estimated fully from simulation, and the uncertainty on its yield is 4 and 20 %, respectively. Such numbers include the uncertainty on the misidentification rate of jets to leptons (for the $e\mu$ channel) or to τ_h (for the $\tau_h\tau_h$ channel);

- the QCD background is estimated from data in all channels, but with different methods. In the $\ell\tau_h$ and $\tau_h\tau_h$ channels, dedicated control regions are included in the final fit to constrain the QCD yield from the data. The uncertainty from the fit of such control regions, as well as an uncertainty for the extrapolation from the control regions to the signal regions, are applied, amounting to 20 % for the $\ell\tau_h$ channels and between 3 and 15 % for the $\tau_h\tau_h$ channel. In the $e\mu$ channel, the QCD yield uncertainty ranges between 10 and 20 % depending on the category;

- the $t\bar{t}$ background shape is corrected by applying event-by-event weights. The uncertainty on this reweighing procedure is taken into account by constructing alternative templates built applying no weight and twice the weight. The yield of the $t\bar{t}$ background is determined from data using a dedicated control region included in the final fit. The uncertainty from the fit in the control region is propagated to the signal regions, and amounts to 5 %;

- the single-top and diboson backgrounds are estimated from simulation and a 5 % uncertainty on the total yield is applied, based on recent CMS measurements of their cross sections [147, 148].

6.2.3 Theory uncertainties on the signal

The simulated Higgs boson p_T distribution is corrected to match the NNLO plus NNLL prediction by the HRES2.1 generator [149, 150]. The following sources of uncertainty affect the signal processes:

- uncertainties on the parton distribution functions (PDFs) and on the QCD renormalisation and refactorization scales affect the rate and the acceptance of signal events. The PDFs uncertainty amounts to 3.2, 2.1, 1.9 and 1.6 % for the ggH, VBF, WH and ZH Higgs production modes, respectively, and the corresponding renormalisation and factorisation scales uncertainties are 3.9, 0.4, 0.7 and 3.8 % [40], respectively. The acceptance uncertainties coming from the renormalisation and factorisation scales are treated as shape uncertainties for the ggH process, as they increase linearly as a function of the reconstructed Higgs boson p_T and $m(jj)$;

- the uncertainty on the parton shower modelling, estimated by varying the parton shower tune in Herwig++ 2.6 samples [151] change the acceptance up to 7 %;

- the theoretical uncertainty on the $H \to \tau^+\tau^-$ branching ratio, equal to 2.1 % [40].

The dependence of the theoretical uncertainties as a function of the jet multiplicity is taken into account as recommended in [152]. The overall contribution of the theory uncertainties on the signal strength amounts to 10 %.

6.2.4 Other uncertainties

The uncertainty on the integrated luminosity is 2.5 %, it is applied to background processes estimated from simulation, and it is fully correlated among channels and categories. Statistical uncertainties due to the limited size of simulated samples and of the event yield in the control regions are taken into account. They are considered separately for each bin of the distributions used in the final fit, if the uncertainty is larger than 5 %. Such bin-by-bin uncertainties are treated as uncorrelated and affect mostly the VBF category, where templates are less populated than in the other categories. The overall contribution of the bin-by-bin uncertainties to the signal strength uncertainty is 12 %.

Table 6.3 Summary table of the experimental uncertainties considered for the $H \to \tau\tau$ analysis. The postfit value of the uncertainty is indicated when it is significantly different from its prefit value.

Object reconstruction and identification

Source	Prefit	Postfit
τ_h energy scale	1.2%	0.2-0.3%
τ_h identification	5%	3.5%
τ_h trigger (per τ_h)	5%	3%
τ_h decay mode migration ($\ell\tau_h$)	3%	2%
e energy scale ($e\mu$)	1-2.5%	0.2-0.5%
$e \to \tau_h$ energy scale	3%	0.6-0.8%
$e \to \tau_h$ rate	12%	5%
e identification and trigger	2%	-
$\mu \to \tau_h$ energy scale	1.5%	0.3-1.0%
$\mu \to \tau_h$ rate	25%	3-8%
μ identification and trigger	2%	-
$\vec{E}_T^{miss}$ energy scale (unclustered)	Dep. on (p_T, η)	-
$\vec{E}_T^{miss}$ energy scale (clustered)	Dep. on (p_T, η)	-
Jet energy scale	Dep. on (p_T, η)	-
Jet $\to \tau_h$ rate	$0.2 \cdot p_T(\tau_h)/100\,\text{GeV}$	15%

Background estimation

Source	Prefit	Postfit
Drell-Yan reweighing	Dep. on Z mass, p_T, $m(jj)$	-
Drell-Yan normalisation	7-15%	3-15%
W+jets normalisation (from high M_T sideband, $\ell\tau_h$)	5-10%	-
W+jets normalisation (from control region, $\ell\tau_h$)	5-15%	-
W+jets normalisation (from simulation, $\tau_h\tau_h$ and $e\mu$)	4-20%	-
QCD multijet normalisation (from anti-iso. sideband, $\tau_h\tau_h$)	3-15%	3-10%
QCD multijet normalisation (from anti-iso. sideband, $\ell\tau_h$)	20%	7-10%
QCD multijet normalisation (from control region, $\ell\tau_h$ and $\tau_h\tau_h$)	5-15%	-
QCD multijet normalisation ($e\mu$)	10-20%	5-20%
$t\bar{t}$ reweighing	Dep. on top p_T	-
$t\bar{t}$ normalisation (from control region)	5%	-
b-jet veto ($e\mu$)	3.5-5.0%	-
Single top cross section	5%	-
Diboson cross section	5%	-
Integrated luminosity	2.5%	-

6.3 Results

The results are obtained from a global binned maximum likelihood fit of the 2D distributions in all channels and categories, together with the control regions for the QCD, W+jets, and $t\bar{t}$ backgrounds. The likelihood function involves the expected and observed numbers of events in each bin of the 2D distributions and of the control regions. The expected number of signal events is the one predicted for the Standard Model Higgs boson with mass $m_H = 125.09\,\text{GeV}$ decaying to a pair of τ leptons, multiplied by a signal strength modifier $\mu = (\sigma \cdot BR)/(\sigma \cdot BR)_{SM}$, which is treated as a free parameter in the fit. The systematic uncertainties are represented by nuisance parameters, being varied in the fit according to their probability density functions. For uncertainties affecting only the yield of the background processes, the nuisance parameters probability density function is assumed to be log-normal, while for uncertainties affecting the shape of the templates the nuisance parameters are assumed to have a Gaussian probability density function [153]. More details on the statistical methods applied are given in Appendix B and in the cited references.

The fitted distributions are shown in Figs. 6.4, 6.5, 6.6 and 6.7 for the $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ and $e\mu$ channels, respectively. The binning of such distributions is finer in the region where the Higgs signal is expected, and in general it is chosen depending on the statistical precision of the simulation and data templates, leading to wider bins in the less populated categories, e.g. VBF. The signal prediction for a Higgs boson with mass $m_H = 125.09\,\text{GeV}$ is shown as a red line and is normalised to the fitted signal strength. The constraints on the backgrounds coming from the control region fit are applied.

The most sensitive category is VBF, followed by the boosted and 0-jet categories. Among the channels, the most sensitive is the fully hadronic one, followed by $\mu\tau_h$, $e\tau_h$ and $e\mu$. An excess of events in data with respect to the Standard Model backgrounds is observed in Fig. 6.8, showing the distribution of the decimal logarithm of the ratio between the number of signal events S and the number of signal-plus-background events $S+B$ for all bins. The excess of events in data is well described by the Standard Model Higgs signal, as shown in the inset of Fig. 6.8. The expected background and signal contributions in each channel, as well as the observed number of events, are indicated in Table 6.4 for the most sensitive bins with $\log_{10}(S/(S+B)) > 0.9$. The excess of events can be seen also in the combined $m_{\tau\tau}$ invariant mass distributions shown in Fig. 6.9. They are obtained by summing the mass distributions for a constant range of the second dimension and weighting them by the $S/(S+B)$ ratio to enhance the contribution of the most sensitive categories. The first and last bins of the distributions are not included, as their signal yield is negligible. Due to the different binning used, it is not possible to combine all channels and categories into a single $m_{\tau\tau}$ distribution, and at least two distributions are needed to include all channels and categories which use the $m_{\tau\tau}$ variable as observable: the two panels of Fig. 6.9 group compatible bins of Figs. 6.4, 6.5, 6.6 and 6.7.

Figure 6.4 Observed and predicted 2D distributions in the $\mu\tau_h$ channel, after the final fit, for the 0-jet (top), boosted (middle) and VBF (bottom) categories. The signal distribution is normalised to its best fit signal strength. The uncertainty band includes systematic and statistical uncertainties, after the final fit. The signal is shown both as a stacked filled histogram and an open overlaid histogram. The bottom panel shows the difference between the data and the predicted backgrounds divided by the background uncertainty, and compares it to the fitted $H \to \tau\tau$ signal [135].

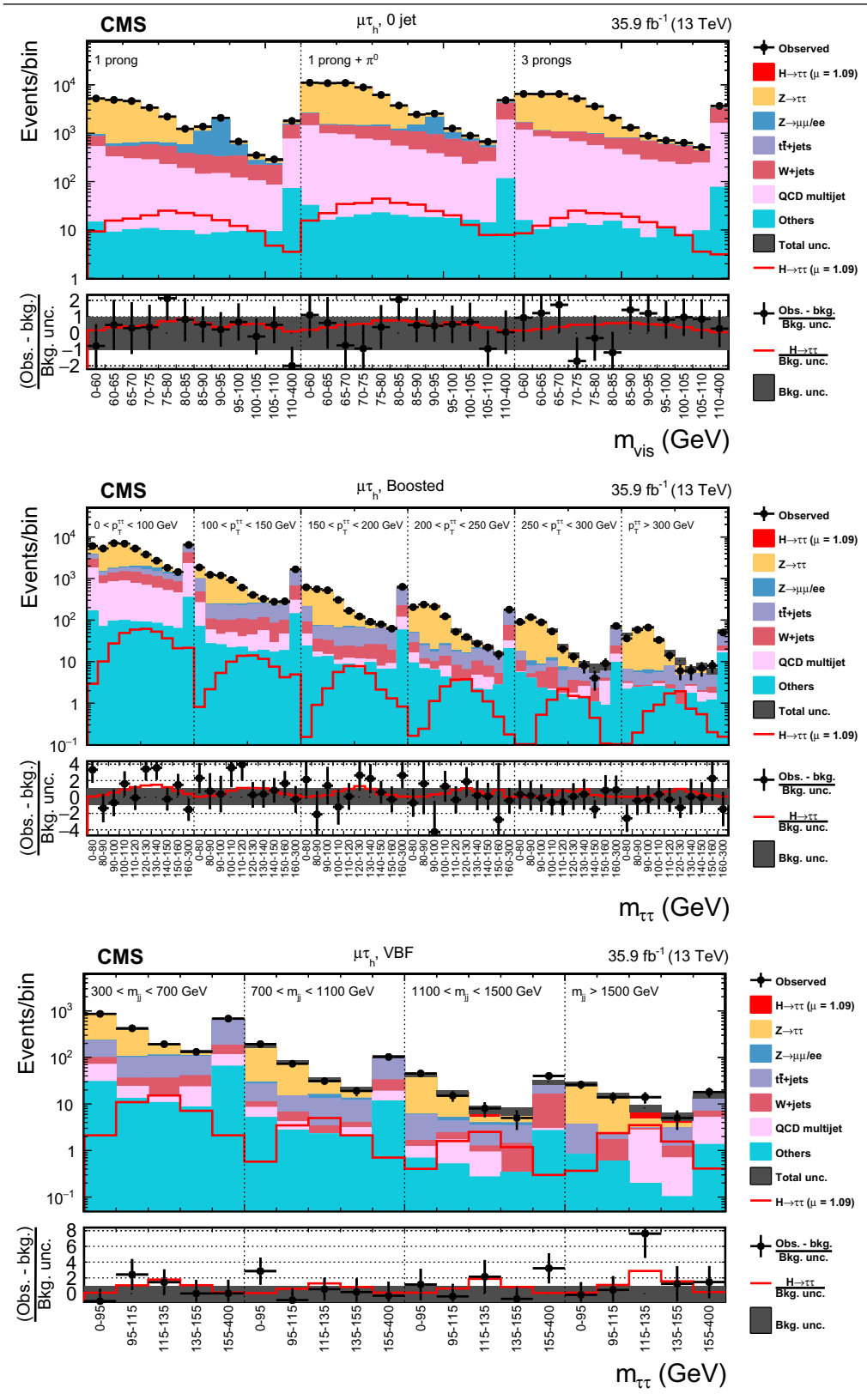


Figure 6.5 Observed and predicted 2D distributions in the $e\tau_h$ channel, after the final fit, for the 0-jet (top), boosted (middle) and VBF (bottom) categories. The signal distribution is normalised to its best fit signal strength. The uncertainty band includes systematic and statistical uncertainties, after the final fit. The signal is shown both as a stacked filled histogram and an open overlaid histogram. The bottom panel shows the difference between the data and the predicted backgrounds divided by the background uncertainty, and compares it to the fitted $H \to \tau\tau$ signal [135].

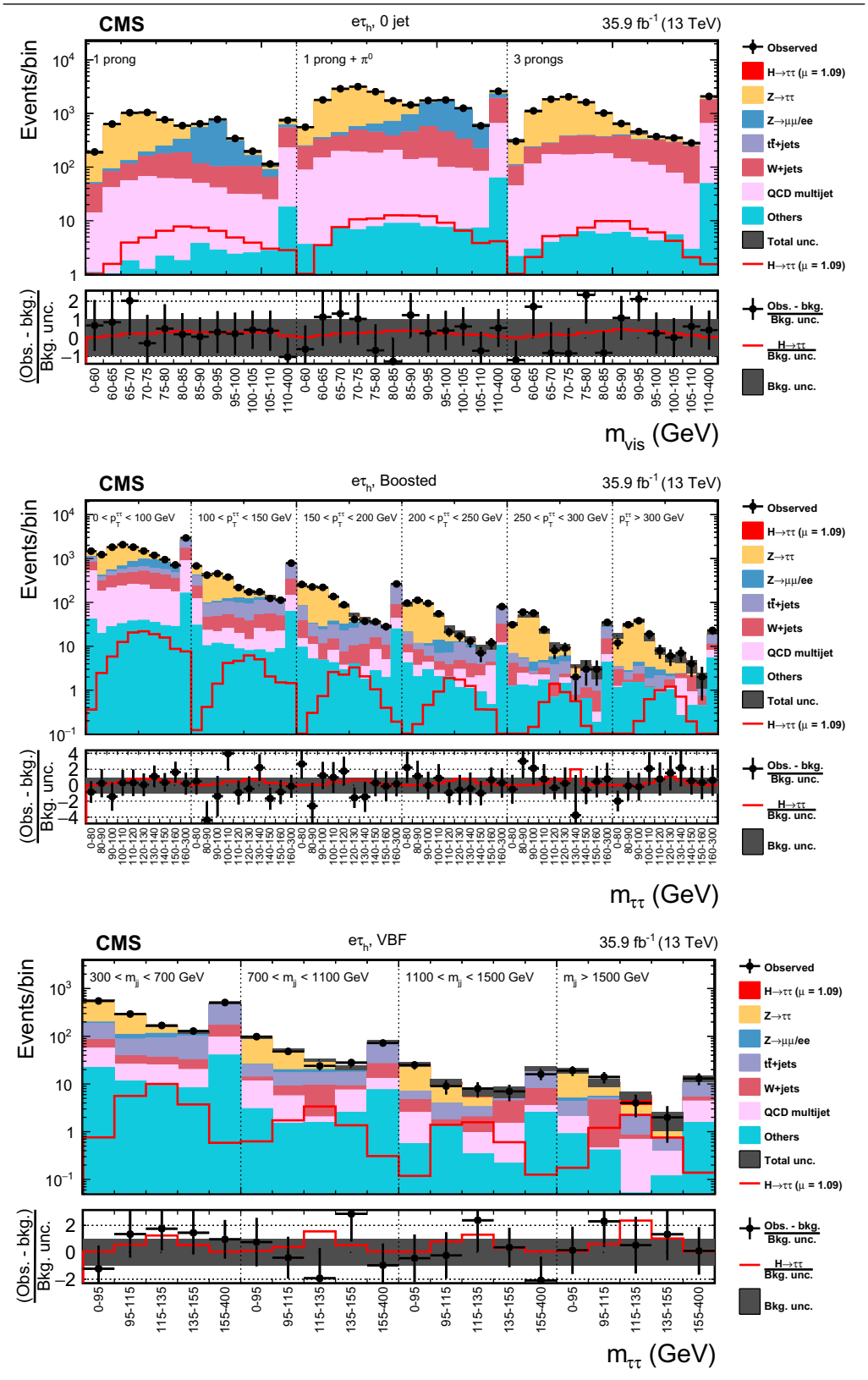


Figure 6.6 Observed and predicted 2D distributions in the $\tau_h\tau_h$ channel, after the final fit, for the 0-jet (top), boosted (middle) and VBF (bottom) categories. The signal distribution is normalised to its best fit signal strength. The uncertainty band includes systematic and statistical uncertainties, after the final fit. The signal is shown both as a stacked filled histogram and an open overlaid histogram. The bottom panel shows the difference between the data and the predicted backgrounds divided by the background uncertainty, and compares it to the fitted $H \to \tau\tau$ signal [135].

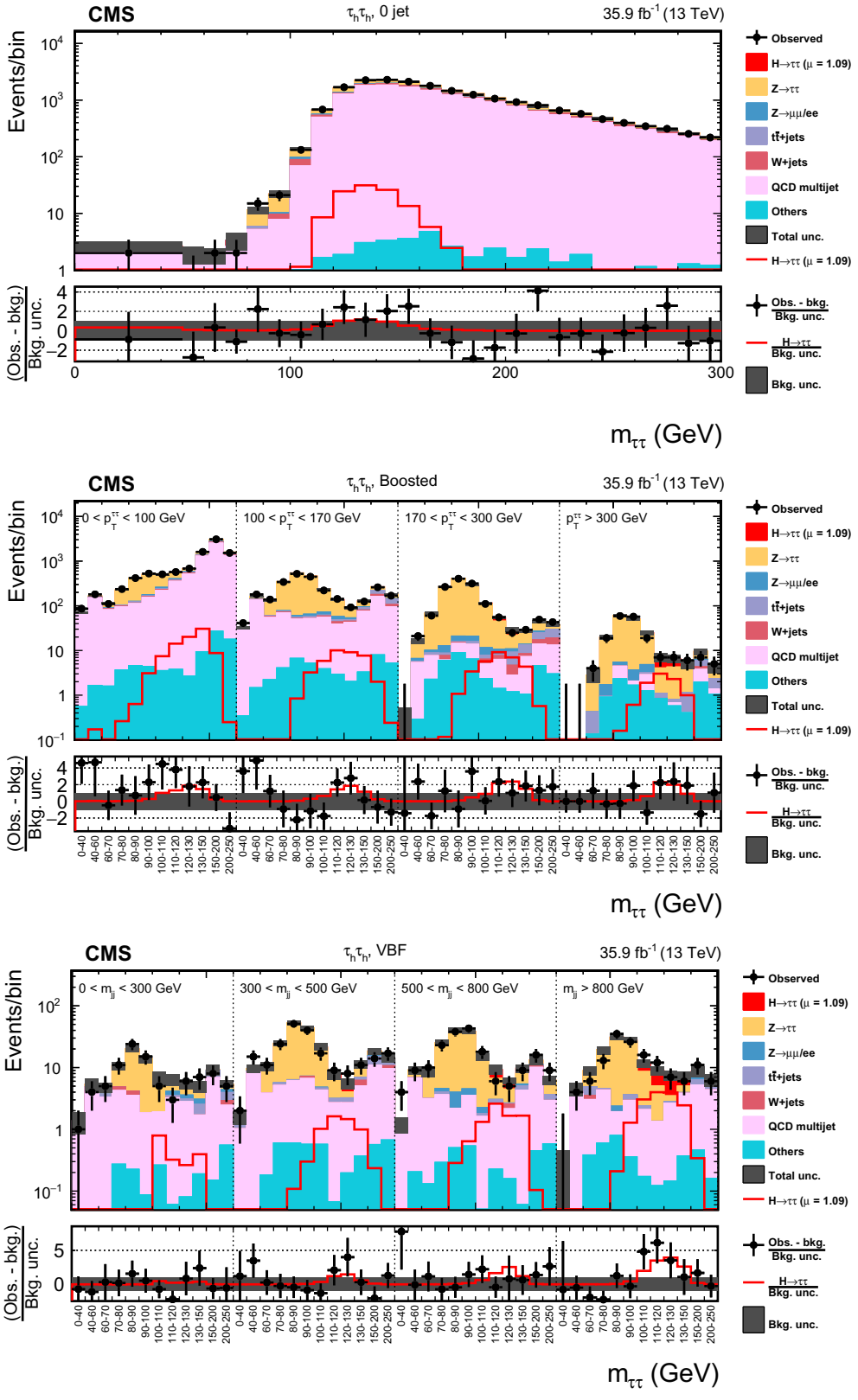


Figure 6.7 Observed and predicted 2D distributions in the $e\mu$ channel, after the final fit, for the 0-jet (top), boosted (middle) and VBF (bottom) categories. The signal distribution is normalised to its best fit signal strength. The uncertainty band includes systematic and statistical uncertainties, after the final fit. The signal is shown both as a stacked filled histogram and an open overlaid histogram. The bottom panel shows the difference between the data and the predicted backgrounds divided by the background uncertainty, and compares it to the fitted $H \to \tau\tau$ signal [135].

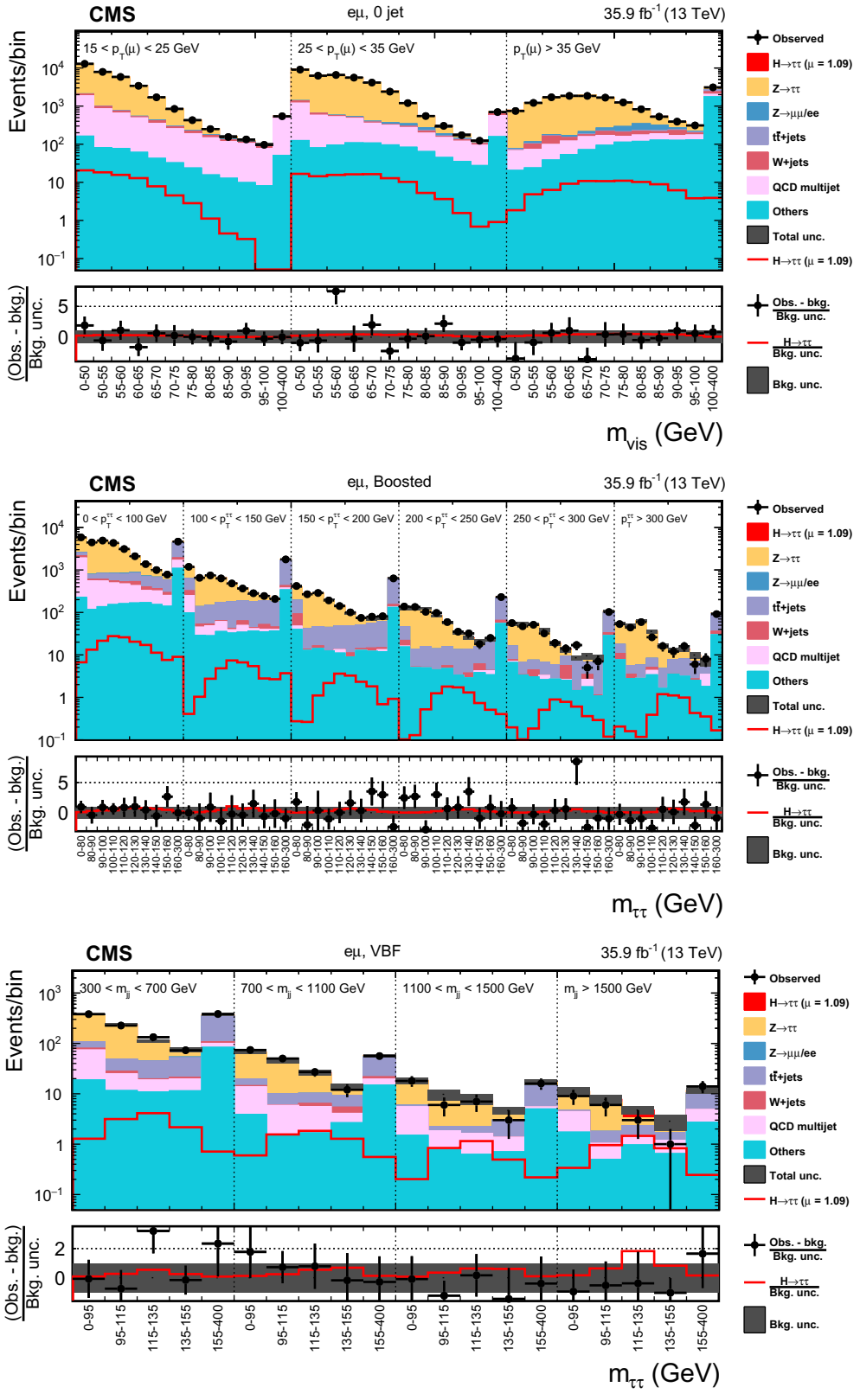


Figure 6.8 Distribution of the decimal logarithm of the S/(S+B) ratio, where S and B are the numbers of expected signal and background events in each bin of the mass distributions, respectively, for the different decay channels. The inset shows the corresponding difference between the observed data and expected background distributions divided by the background expectation, as well as the signal expectation divided by the background expectation [135].

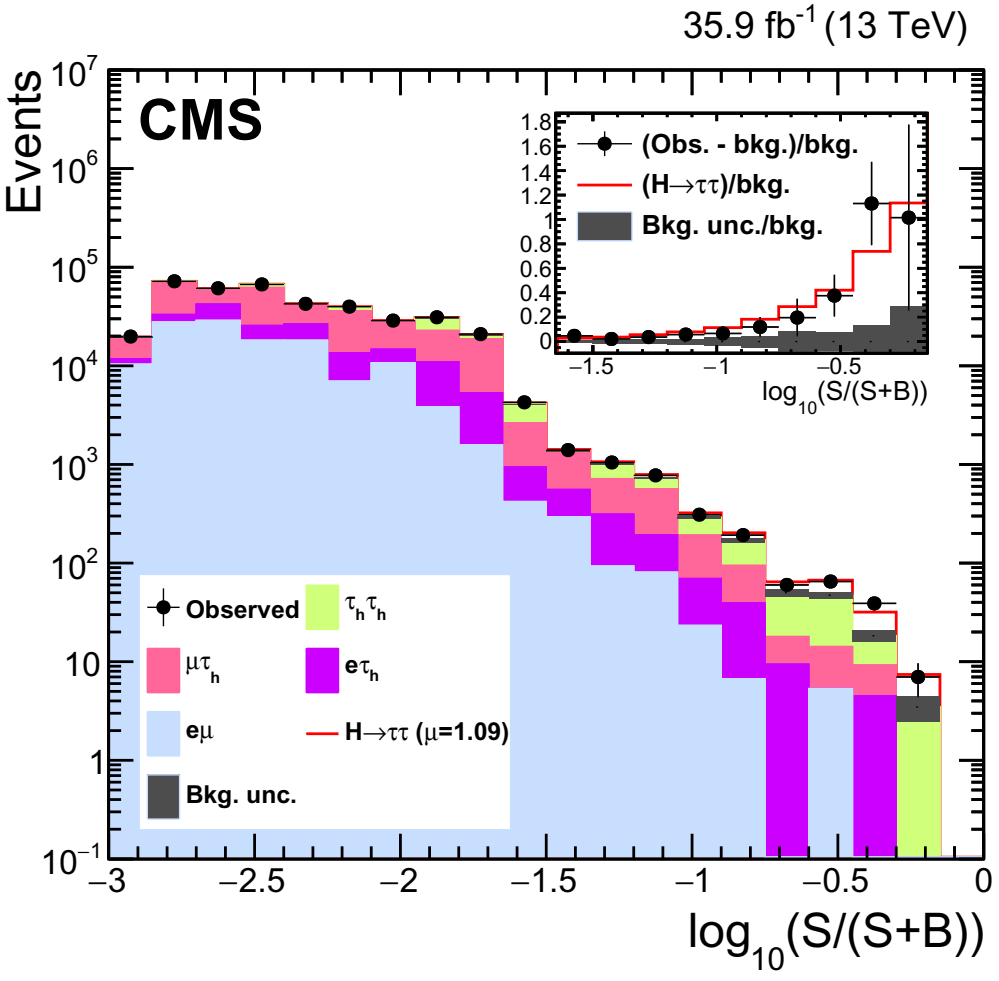


Table 6.4 Expected background and signal contributions in each channel, together with the observed number of events, for the most sensitive bins of Fig. 6.8, with $\log_{10}(S/(S+B)) > 0.9$ [135].

Process	$\tau_h\tau_h$	$\mu\tau_h$	$e\tau_h$	$e\mu$
$Z \to \tau\tau$	89.1 ± 6.9	34.6 ± 4.9	21.2 ± 3.3	0.8 ± 2.2
$Z \to ee/\mu\mu$	5.0 ± 0.2	3.7 ± 0.2	2.9 ± 0.2	0.0 ± 0.0
$t\bar{t}$+jets	13.9 ± 0.5	22.2 ± 1.8	10.4 ± 0.3	1.9 ± 0.1
W+jets	7.6 ± 0.8	6.6 ± 1.3	4.0 ± 0.3	0.8 ± 0.02
QCD multijet	35.5 ± 2.1	5.0 ± 1.3	3.3 ± 2.5	2.1 ± 0.3
Other backgrounds	7.3 ± 0.2	6.1 ± 0.2	5.2 ± 0.2	1.4 ± 0.1
$ggH, H \to \tau\tau$	27.4 ± 2.1	6.0 ± 0.6	5.0 ± 0.6	0.6 ± 0.1
VBF, $H \to \tau\tau$	17.5 ± 1.0	12.6 ± 1.0	5.1 ± 0.5	2.8 ± 0.3
$VH, H \to \tau\tau$	1.3 ± 0.1	0.2 ± 0.0	0.3 ± 0.0	0.0 ± 0.0
Total background	156.2 ± 7.3	77.7 ± 5.5	46.5 ± 4.1	12.1 ± 2.2
Total signal	48.3 ± 2.6	19.2 ± 1.4	10.9 ± 0.8	3.4 ± 0.4
Observed	207	91	54	11

To quantify the observed excess of events, a profile ratio test statistic is used to calculate the local p-value and the corresponding significance of the excess. Several Higgs boson mass hypotheses have been tested, namely $m_H = 110, 120, 125.09, 130$ and $140\,\text{GeV}$. Figure 6.10 shows the expected and observed local p-value as a function of the Higgs mass, which has its minimum for $m_H = 125.09\,\text{GeV}$. For this mass value, the observed significance is $4.9\,\sigma$, the expected value being $4.7\,\sigma$.

The best fit value for the signal strength is found to be $\mu = 1.09^{+0.27}_{-0.26}$ for $m_H = 125.09\,\text{GeV}$, and is compatible with the Standard Model expectation within the uncertainty. The latter can be decomposed into four main components, namely statistical uncertainty, experimental uncertainties, theoretical uncertainty and bin-by-bin statistical uncertainties on the simulation, to obtain $\mu = 1.09 \pm 0.15(\text{stat.})^{+0.16}_{-0.15}(\text{syst.})^{+0.10}_{-0.08}(\text{theo.})^{+0.13}_{-0.12}(\text{bbb})$. The best fit signal strength per channel and category are shown in Fig. 6.11. Results are consistent among channel and categories and with the expectation from the Standard Model Higgs boson.

A scan of the negative logarithm of the likelihood function is performed as a function of the κ_V and κ_f parameters, which measure the coupling of the Higgs boson to vector bosons V and to fermions f with respect to the Standard Model expectation. For this measurement, the $H \to W^+W^-$ process is included in the signal. The result of the (κ_V,κ_f) scan is shown in Fig. 6.12, where the best fit result, the Standard Model expectation and the 68 % and 95 % confidence level intervals are plotted. The measurement shows that the best fit result is in agreement with the Standard Model expectation $\kappa_V = \kappa_f = 1$.

Figure 6.9 Observed and predicted $m_{\tau\tau}$ distributions, combined for the categories with compatible binning, namely the VBF category of the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels (top) and the three categories of the $\tau_h\tau_h$ channels and the boosted categories of the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels (bottom). The normalisation of the backgrounds and of the signal and the uncertainty are adjusted to the values obtained from the final fit. The mass distributions for a constant range of the second dimension are weighted according to S/(S+B), where S and B are computed, respectively, as the signal and background contribution in the mass distribution excluding the first and last bins. The inset shows the corresponding difference between the observed data and expected background distributions, together with the signal expectation for the Standard Model Higgs boson [135].

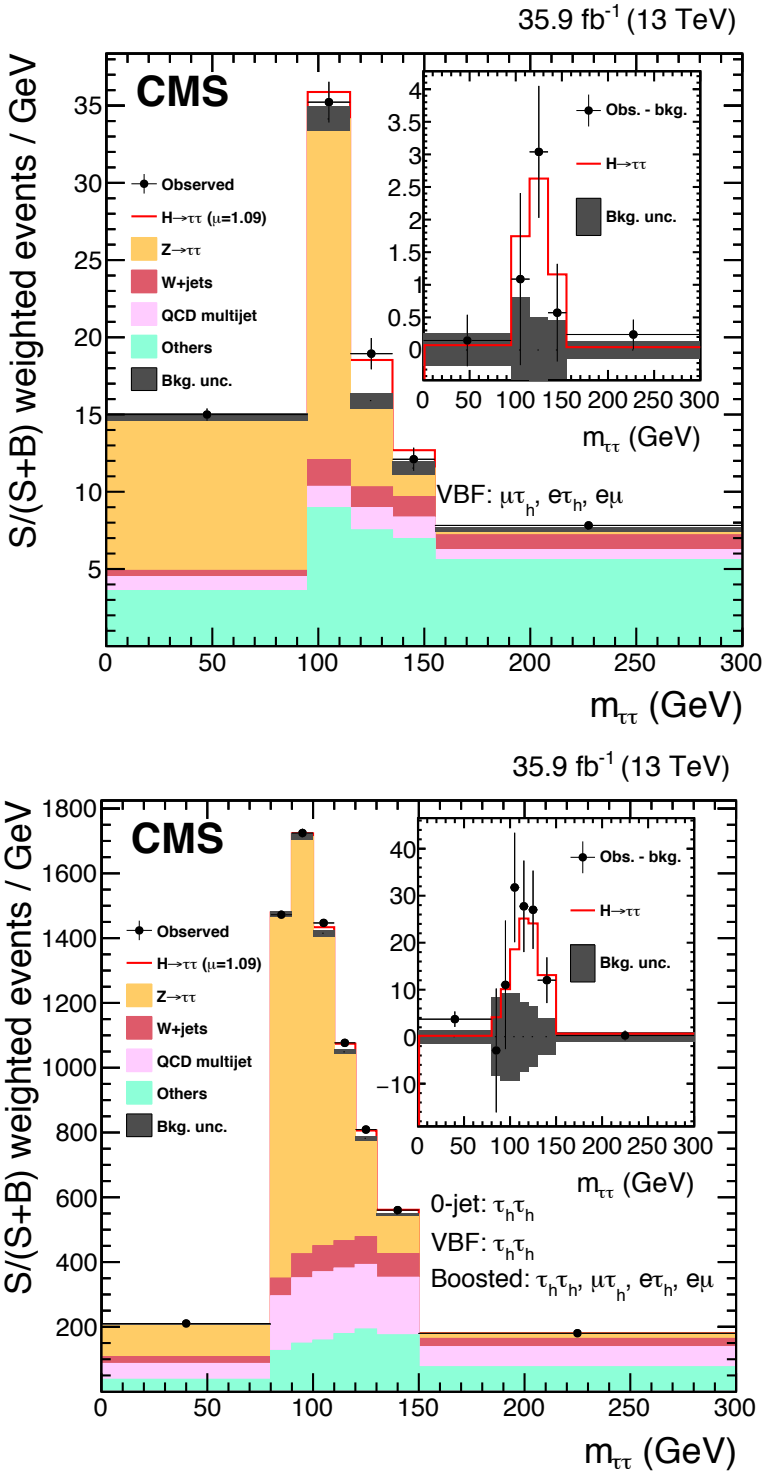


Figure 6.10 Local p-value and significance as a function of the Standard Model Higgs boson mass hypothesis. The observation (red, solid) is compared to the expectation (blue, dashed) for a Higgs boson with a mass $m_H = 125.09\,\text{GeV}$: the observed significance is $4.9\,\sigma$, expected $4.7\,\sigma$ [135].

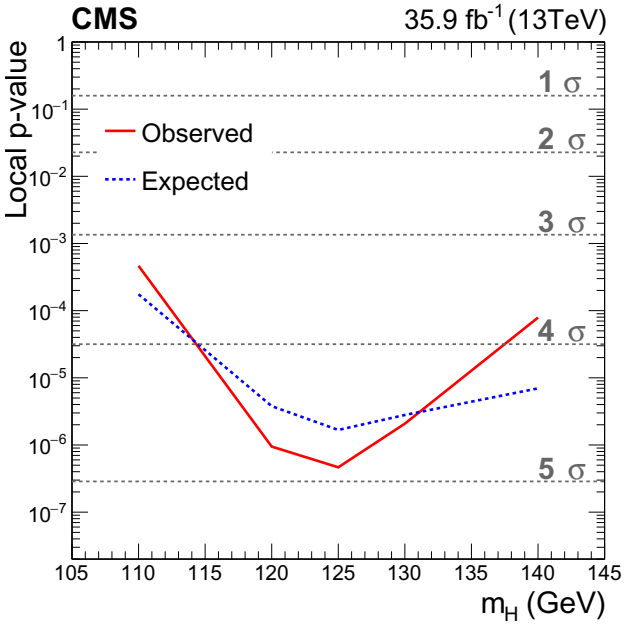


Figure 6.11 Best fit signal strength per category (left) and channel (right), for $m_H = 125.09\,\text{GeV}$. The constraints from the global fit are used to extract each of the individual best fit signal strengths. The combined best fit signal strength is $\mu = 1.09^{+0.27}_{-0.26}$, and it is compatible among categories and channels and with the Standard Model expectation within the uncertainties [135].

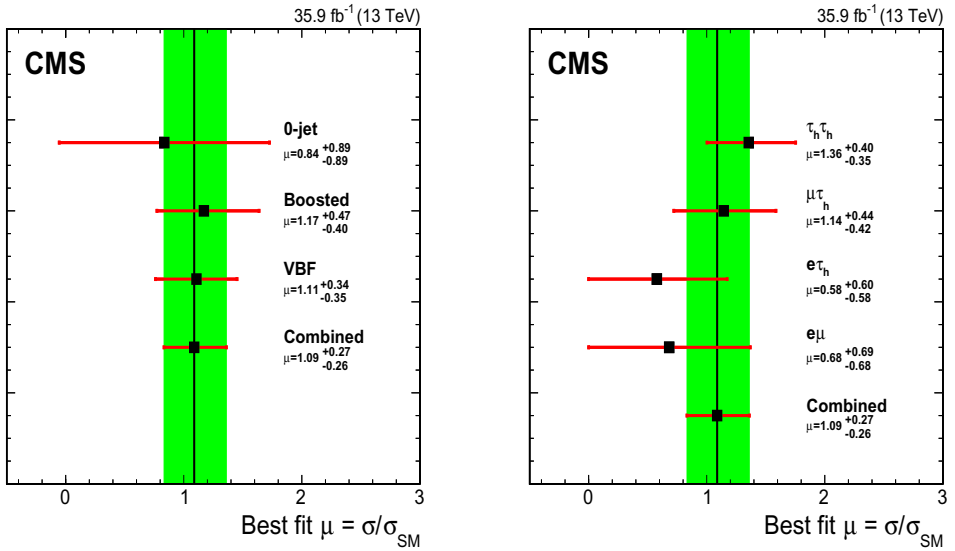
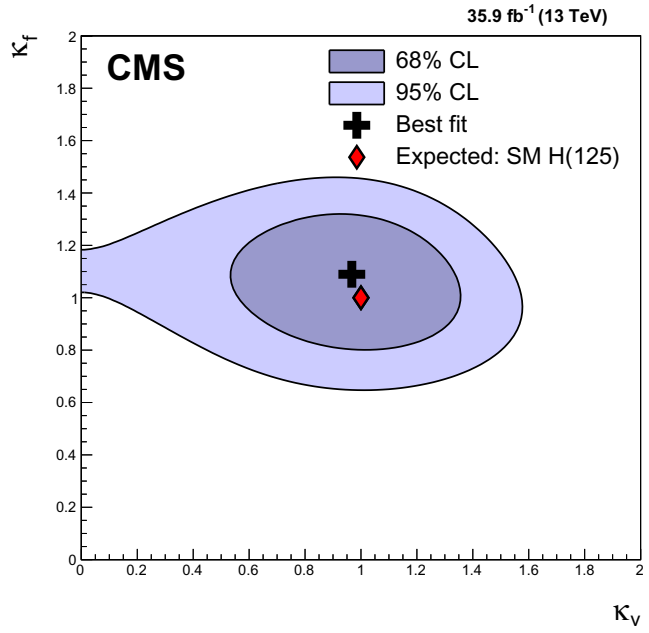


Figure 6.12 Scan of the negative log-likelihood as a function of κ_V and κ_f, for $m_H = 125.09\,\text{GeV}$. All nuisance parameters are profiled for each point. For this scan, the $H \to W^+W^-$ contribution is treated as a signal process [135].



Combination with the Run 1 result

Finally, a combination of this analysis with the previous one by CMS, which used the 7 and 8 TeV data [133], has been performed. Theoretical uncertainties have been treated as fully correlated, while other systematic uncertainties as fully uncorrelated. In the combination, the observed significance of the $H \to \tau^+\tau^-$ signal is 5.9 σ (expected 5.9 σ) and the best fit for the signal strength $\mu = 0.98 \pm 0.18$ for $m_H = 125.09\,\text{GeV}$.
This is the first observation of direct $H \to \tau^+\tau^-$ decays by a single experiment.

Chapter 7

Conclusion

The discovery of a new resonance with a mass of 125 GeV compatible with the Higgs boson hypothesis, announced on July, 4$^{\text{th}}$ 2012 by the ATLAS and CMS collaborations, has opened up the path for the exploration of the mechanism of electroweak spontaneous symmetry breaking, which is supposed to be responsible for the origin of masses of elementary particles. In particular, fermions acquire their mass in the Yukawa interaction with the Higgs field. Establishing the couplings of the Higgs boson with the different fermion flavors is one of the most important steps towards the characterisation of this newly discovered particle, which may open a portal to discover physics beyond the Standard Model.

According to the Standard Model, a Higgs boson with a mass of 125 GeV decays to a pair of tau leptons with a branching ratio of 6.3 %. Thanks to a good balance between its branching ratio and the amount of background events, the $H \rightarrow \tau^+\tau^-$ decay is the most sensitive fermionic channel at the LHC. In this thesis, the measurement of the coupling strength of the Higgs boson to τ leptons, performed using proton-proton collision events at a centre-of-mass energy of 13 TeV, has been presented. The analysed dataset has been recorded with the CMS detector in the year 2016 and corresponds to an integrated luminosity of 35.9 fb^{-1}. This measurement of the $H \rightarrow \tau^+\tau^-$ signal strength is the most sensitive at the time of its publication and it is the first one being performed at $\sqrt{s}$ = 13 TeV. The observed significance of the $H \rightarrow \tau^+\tau^-$ signal is 4.9 σ, to be compared to an expected significance of 4.7 σ. The measured signal strength is

$$\mu = \sigma/\sigma_{SM} = 1.09^{+0.27}_{-0.26} = 1.09 \pm 0.15(\text{stat.})^{+0.16}_{-0.15}(\text{syst.})^{+0.10}_{-0.08}(\text{theo.})^{+0.13}_{-0.12}(\text{bbb.}),$$

where the overall uncertainty has been split into four sources, namely the data statistical (stat.) uncertainty, the experimental (syst.) and theoretical (theo.) systematic uncertainties, and the bin-by-bin (bbb.) uncertainties, accounting for the limited statistics of simulated events. The result is fully compatible with the Standard Model expectation within the uncertainties. The combination with the same measurement performed with data collected by the CMS experiment at centre-of-mass energies of 7 and 8 TeV results in an observed significance of 5.9 σ, equal to the expected significance. This is

the first observation of Higgs boson decays to τ lepton pairs by a single experiment, and represents a milestone in the path towards the characterisation of the Higgs boson particle.

The work done in the context of this thesis contributes to the success of the analysis in the $\mu\tau_h$ and $e\tau_h$ final states, two of the four channels from which the 13 TeV results is obtained. In particular, the analysis selection and the background modelling have been extensively studied. Corrections for muon and electron identification, isolation and trigger efficiencies have been measured as a function of the lepton kinematics for the specific selection used in the analysis and applied to all simulated events, making the simulation closer to the observed data. Moreover, the full analysis of the $\mu\tau_h$ and $e\tau_h$ channels has been performed, up to the production of all the templates needed for the combined likelihood fit, with the aim to provide a thorough cross-check of all the steps of the analysis up to the final result. This played a very important role in increasing the confidence that such a complex analysis is performed correctly and ready for publication. The paper [135] has been published in Physics Letters B.

Obtaining such great results is possible only if a sophisticated data analysis is complemented by excellent data quality. This requires superb detector hardware and precise calibration of all subsystems. During this doctoral work important contributions have been made to the alignment of the CMS silicon tracker. In this context, the main result of this thesis is the elaboration of a strategy to align the upgraded CMS pixel detector soon after its installation at the beginning of 2017. Several strategies to align the new detector with the first set of recorded tracks, consisting of cosmic ray and collision tracks, have been studied and evaluated. The best strategy among those has then been identified and applied for the first alignment of the new pixel detector, performed using cosmic ray tracks. Large initial misalignments could be successfully identified and corrected, as confirmed by the significant improvements in the quality of reconstructed tracks.

The sensitivity of the $H \to \tau^+\tau^-$ analysis is currently limited by the available amount of data and by the experimental systematic uncertainties, contributing equally to the overall uncertainty. The statistical uncertainty on the simulated events is also relevant, due to the low population of templates, especially in the VBF category. In the near future, these uncertainties could be reduced thanks to the increase of the dataset size. In fact, already by the end of 2018, the amount of data recorded by CMS at a centre-of-mass energy of 13 TeV has reached 150 fb^{-1}. The availability of such a large dataset should allow reducing both the statistical and the experimental uncertainty, as the corrections to the simulation may be derived more precisely. Moreover, the increasing size of the dataset will encourage a more systematic development and use of data-driven background estimation techniques, with the consequent reduction of the use of simulated data and of the necessary corrections.

Finally, it is important to be aware that the price to pay for collecting a large quantity of data in a short time is high instantaneous luminosity, and dealing with a high pile-up environment will be one of the main challenges in the coming years. In the next phase

of the LHC program, which is planned to start in 2026, proton-proton collisions will happen at extremely high instantaneous luminosities, producing events with up to 200 pile-up interactions. In ten years of operation, a dataset corresponding to 3000 fb^{-1} is expected to be produced, allowing to perform an exciting variety of measurements [154]. To mention some, the $H \to \mu^+\mu^-$ and $Z\gamma$ decay channels should become visible for the first time, and the main Higgs boson couplings, including the one to τ leptons, should be measurable with uncertainties at the level of a few percent. The large size of the dataset will also allow to probe rare exotic decay modes of the Higgs boson, up to branching ratios of the order of 10^{-6}, as well as to set stringent upper limits at the percent level on the Higgs decay width to invisible final states. Even the namely the double Higgs production might become visible with a significance of about $4\,\sigma$.

To summarise, an exciting path through the unexplored Higgs territory is being traced, and there are reasons to believe that it will lead to a detailed view of the Higgs landscape in the next two decades.

Appendix A

Reference System and Kinematic Variables

The reference frame used to describe the CMS detector and the recorded events from collisions has its origin in the geometrical centre of the solenoid. Two different global coordinates are used:

- Cartesian coordinate system: the x axis unit vector $\hat{i}$ is oriented towards the centre of the LHC, the y axis unit vector $\hat{j}$ upwards, perpendicular to the LHC plane, and the z axis unit vector $\hat{k}$ point in counterclockwise direction along the beam line, completing a right handed frame.
- Polar coordinate system: the particle directions are defined with an azimuthal angle $\phi = \arctan\frac{y}{x}$ and a polar angle $\theta = \arctan\frac{r_\perp}{z}$, with $r_\perp$ being the radius in the xy plane.

The basic kinematic variables used in the analyses are

- Transverse momentum p_T

$$p_T = p \sin\theta \tag{A.1}$$

  It is the component of the particle momentum in the plane perpendicular to the beam direction.

- Rapidity y

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{A.2}$$

  with E being the energy and p_z the longitudinal momentum of a particle.

- Pseudorapidity η

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{A.3}$$

It is an approximation of the rapidity for particles with mass $m << E, p_z$. For massless particles, rapidity and pseudorapidity are identical.

Appendix B

Statistical Methods

In this Appendix, the statistical methods used to derive the results of the $H \to \tau^+\tau^-$, presented in Chapter 6, are summarised.

B.1 The likelihood function

In general, the likelihood function is the probability density function (pdf) of the observed data, as a function of a set of parameters. The analysis presented in this thesis is based on binned histograms, where the content of each bin is independent from the others and can be regarded as the outcome of a single counting experiment with average rate λ. The probability of such an experiment to observe n events follows the Poisson distribution and is a function of λ:

$$f(n) = \frac{e^{-\lambda}\lambda^n}{n!}. \tag{B.1}$$

Typically, the true value of the parameter λ is not known, and its best estimator is the expected number of events a in the time interval of the experiment. The likelihood function, that quantifies the agreement between the expectation and the outcome of the experiment, is

$$\mathcal{L}(n|a) = \frac{e^{-a}a^n}{n!}. \tag{B.2}$$

In the case of N independent counting experiments, each observing n_i and expecting a_i events, the likelihood is the product of the likelihoods for each experiment. Therefore, the likelihood function for a histogram with N bins, with n_i observed a_i expected events in each bin i, can be written as:

$$\mathcal{L}(\vec{n}|\vec{a}) = \prod_{i=1}^{N} \frac{e^{-a_i}a_i^{n_i}}{n_i!}, \tag{B.3}$$

with $\vec{n} = \{n_1, ...n_N\}$ and $\vec{a} = \{a_1, ..., a_N\}$ being the vectors of observed and expected number of events in each bin of the histogram.

B.1.1 Treatment of systematic uncertainties

Systematic uncertainties that influence the model, and therefore the expected outcome of the measurement, can be taken into account in the likelihood function [153] by introducing a set of parameters $\vec{\theta}$, often called "nuisance" parameters - as they have an impact on the result but they are not the main object of interest. Each source of systematic uncertainty is represented by a different nuisance θ_j, whose nominal value $\tilde{\theta}_j$ is usually determined by dedicated measurements, and its uncertainty is described by the pdf $p(\tilde{\theta}_j|\theta_j)$. The systematic uncertainties change the expected number of events $\vec{b}$, that becomes a function of the nuisance parameters $\vec{\theta} = \{\theta_1, ..., \theta_L\}$, and the likelihood function of Eqn. B.3 is modified as follows:

$$\mathcal{L}(\vec{n}|\vec{a}) = \prod_{i=1}^{N} \frac{e^{-a_i} a_i^{n_i}}{n_i!} \prod_{j=1}^{L} p(\tilde{\theta}_j|\theta_j). \tag{B.4}$$

The choice of the pdf depends on the nature of the uncertainty. In the analysis presented in this Thesis, the following approach is adopted:

- systematic uncertainties affecting only the event yield are modelled with a log-normal distribution $lnN(\theta)$ of parameter κ [155]:

$$lnN(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln(\kappa))^2}\right)\frac{1}{\theta}, \tag{B.5}$$

  choosing the width of the distribution κ to be equal to the measured uncertainty. The lnN distribution is preferred to the classical normal distribution because it goes to zero at $\theta = 0$, avoiding negative unphysical contributions to the observables, i.e. the expected number of events;

- systematic uncertainties changing the shape of the distribution are modelled with a linear extrapolation method: two alternative shapes are provided, obtained by varying the nuisance parameter by $\pm 1\sigma$ around its initial value, and a parameter θ is added to the likelihood to interpolate between the up and down shapes using a vertical morphing technique [156];

- uncertainties arising from the limited number of simulated events used to estimate the background in every bin of the distribution, called "bin-by-bin" uncertainties, are a special case of shape uncertainties, because, within a single process, the behaviour of each bin is independent from the others. Therefore, several nuisance parameters need to be introduced, one for every bin of each process [157], multiplying the number of expected events in a given bin for the given process. As

a consequence, the number of nuisances in the likelihood can increase rapidly. On the other hand, many of these parameters usually have a small effect on the results, and can be neglected.

B.2 Maximum likelihood fit

The scope of the analysis presented in this Thesis is to make a statement on the $H \to \tau\tau$ signal, based on the data. For this reason, it is convenient to explicitly write the number of expected events $\vec{a}(\vec{\theta})$ as the sum of two contributions, the background and the signal, the latter multiplied by a positive real number $\mu \geq 0$, that modifies the signal yield:

$$\vec{a}(\vec{\theta}) \to \mu \cdot \vec{s}(\vec{\theta}) + \vec{b}(\vec{\theta}). \tag{B.6}$$

The condition $\mu = 0$ represents the absence of the signal, while for $\mu = 1$ the yield corresponds to the prediction from the Higgs mechanism. Applying the notation of Eqn. B.6 to the likelihood function $\mathcal{L}(\vec{n}|\vec{a})$ of Eqn. B.4 explicitly introduces the signal strength μ as one of its parameters, in addition to the nuisances. The values of the parameters giving the best match between the expectation and the observation can be obtained by performing the maximum likelihood fit: the values of the nuisance parameters and of the signal strength are varied according to their pdf distributions in order to find the values that maximise the likelihood. Technically, the fitting procedure is often performed on the logarithm of the likelihood, instead than on the likelihood function itself, such that the maximum needs to be found for a sum of terms instead of for a product. The best-fit signal strength is the value of μ that maximises the likelihood - or equivalently, that minimises its negative logarithm. The bounds of the $\pm 1\sigma$ interval are obtained by plotting the quantity $q = -2 \ln \mathcal{L}$ as a function of μ, shifting it such that its minimum is zero, interpolating with the line $q = 1$, and reading the corresponding values of μ. It can be shown that this procedure applies not only for the likelihood of normal distributed variables, but also for any other likelihood function that can be mapped to a parabola with a change of variable.

B.3 Hypothesis testing

In high energy physics, the two most common situations to face when interpreting the outcome of a search for a signal are the following:

1. an excess of data is observed over the background-only expectation, whose significance needs to be quantified;
2. no excess (or a modest excess) of data is observed, indicating that the signal cross section is likely smaller than some upper limit, that can be calculated.

In both cases, there is the need to discriminate between two hypotheses, e.g. the presence or absence of the signal, by means of a test statistic.

B.3.1 Definition of the test statistics

The Neyman-Pearson lemma states that the test statistic providing the best discriminating power is the likelihood ratio. In the LHC experiments, the profile likelihood ratio $\lambda(\mu)$ defined as:

$$\lambda(\mu) = \frac{\mathcal{L}(\vec{n}|\mu, \hat{\vec{\theta}}_\mu)}{\mathcal{L}(\vec{n}|\hat{\mu}, \hat{\vec{\theta}})} \tag{B.7}$$

is used to build the test statistic q_μ [155]:

$$q_\mu = -2\ln\lambda(\mu) = -2\ln\frac{\mathcal{L}(\vec{n}|\mu, \hat{\vec{\theta}}_\mu)}{\mathcal{L}(\vec{n}|\hat{\mu}, \hat{\vec{\theta}})}, \text{ with } 0 \leq \hat{\mu} \leq \mu. \tag{B.8}$$

In the numerator of the test statistic the signal strength μ is fixed to the value under test, and the nuisance parameters $\hat{\vec{\theta}}_\mu$ are fitted to maximise the likelihood for the given value of μ, while in the denominator both the signal strength and the nuisances are varied to find their best values yielding the maximum likelihood. Bounds on the signal strength $\hat{\mu}$ are set to avoid negative signal rates and to obtain one-sided confidence intervals not detached from 0. From a physics point of view, the latter bound avoids that a fitted μ value larger than the one under test disfavours the signal hypothesis. Discriminating between the "signal-plus-background" and the "background-only" hypotheses based on the observed data is equivalent to discriminate between a given value of $\mu > 0$ and $\mu = 0$. To do this, the values of the nuisance parameters maximising the likelihood under each hypothesis $\hat{\vec{\theta}}_\mu^{obs}$ and $\hat{\vec{\theta}}_0^{obs}$ have to be found. Then, the probability for the test statistic $\tilde{q}_\mu$ to be larger than the observed value $\tilde{q}_\mu^{obs}$ under the signal-plus-background hypothesis can be calculated:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|\hat{\mu} = \mu) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\vec{\theta}}_\mu^{obs}) d\tilde{q}_\mu, \tag{B.9}$$

where $f(\tilde{q}_\mu|\mu, \hat{\vec{\theta}}_\mu^{obs})$ is the pdf of the test statistic for the value of μ under test and the nuisance parameters $\hat{\vec{\theta}}_\mu^{obs}$, and can be obtained by generating pseudo-experiments (toys) with a Monte Carlo method. The quantity p_μ defined in Eqn. B.9 can be recognised as the p-value for the signal-plus-background hypothesis. Similarly, the p-value for the background-only hypothesis can be calculated as

$$p_b = P(\tilde{q}_\mu < \tilde{q}_\mu^{obs}|\hat{\mu} = 0) = \int_0^{\tilde{q}_0^{obs}} f(\tilde{q}_\mu|0, \hat{\vec{\theta}}_0^{obs}) d\tilde{q}_\mu. \tag{B.10}$$

B.3.2 Significance calculation

The significance of an excess of data over the background-only expectation is quantified by the p-value for the hypothesis $\mu = 0$, that is the probability that the observed excess arises from a pure statistical fluctuation of the backgrounds. This can be calculated starting from the test statistics $\tilde{q}_\mu$ defined in Eqn. B.8, where the value under test is $\mu = 0$ [155]:

$$\tilde{q}_0 = -2 \ln \frac{\mathcal{L}(\vec{n}|0, \hat{\vec{\theta}}_0)}{\mathcal{L}(\vec{n}|\hat{\mu}, \hat{\vec{\theta}})}, \text{ with } \hat{\mu} \geq 0, \tag{B.11}$$

Following the same approach of Eqn. B.9, the p-value for the hypothesis $\mu = 0$ under test is

$$p_0 = P(\tilde{q}_0 \geq \tilde{q}_0^{obs} | \hat{\mu} = 0) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_0|0, \hat{\vec{\theta}}_0^{obs}) d\tilde{q}_0. \tag{B.12}$$

The p-value is converted into a significance Z considering a one sided gaussian tail. In this approach, the 5σ significance (Z = 5) corresponds to $p_0 = 2.8 \times 10^{-7}$. The p-value calculated as described above is called "local" p-value, because is valid only when the signal model is well defined, for example it is a resonance peak at a given known mass. If, instead, the signal is searched over a large mass range, the probability of observing an excess is larger, and the p-value needs to be corrected for the so called "look elsewhere effect" [158].

B.3.3 Limits setting

If the significance of the signal is low, it indicates that the its cross section is likely lower than some value, that is called upper limit. The observed limit on μ can then be derived from the CL_s value [155]:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}. \tag{B.13}$$

If, for $\mu = 1$, $CL_s \leq \alpha$, the signal-plus-background hypothesis is excluded with $(1 - \alpha)$ confidence level (C.L.). Vice versa, the upper limit on μ at a $(1 - \alpha)$ C.L. is the value of μ such that $CL_s = \alpha$. To set exclusion limits, the convention is to choose $\alpha = 0.05$ so that the exclusion is at 95% C.L.

The median expected upper limit on μ (at 95% C.L.) for the background-only hypothesis can be derived by generating toy experiments for $\mu = 0$, calculate CL_s and the value of $\mu \equiv \mu^{95\%}$ such that $CL_s(\mu^{95\%}) \leq 0.05$. By repeating the procedure several times, a distribution of $\mu^{95\%}$ is obtained. From its cumulative, the median expected upper limit is identified as the 50% quantile, and the $\pm 1\sigma$ error bands are given by the 16% and 84% quantiles.

Bibliography

[1] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, doi:10.1103/PhysRevD.98.030001.

[2] M. Peskin and D.V. Schröder, *An Introduction to quantum field theory*, Addison-Wesley (1995).

[3] I.J.R. Aitchison and A.J.G. Hey, *Gauge Theories in Particle Physics*, Taylor & Francis (2004).

[4] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235, doi:10.1080/00411457108231446, english translation by M. A. Tavel.

[5] R.K. Ellis, W.J. Stirling, and B.R. Webber, *QCD and collider physics*, Cambridge University Press (2003).

[6] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, Westview Press (1997).

[7] F. Halzen and A.D. Martin, *Quarks and Leptons: an Introductory Course in Modern Particle Physics*, Wiley (1984).

[8] W. Pauli, *Dear radioactive ladies and gentlemen*, Phys. Today **31N9** (1978) 27, `https://cds.cern.ch/record/83282/files/meitner_0393.pdf`.

[9] E. Fermi, *Tentativo di una Teoria Dei Raggi Beta*, Il Nuovo Cimento **11** (1934) 1, doi:10.1007/BF02959820.

[10] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2.

[11] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264.

[12] A. Salam, *Weak and Electromagnetic Interactions*, *Proc. of 8th Nobel Symposium*, Almqvist & Wiksell, Stockholm (1968).

[13] T. Nakano and K. Nishijima, *Charge Independence for V-particles*, Progress of Theoretical Physics **10** (1953) 5 581, doi:10.1143/PTP.10.581.

[14] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, Il Nuovo Cimento (1955-1965) **4** (1956) 2 848, doi:10.1007/BF02748000.

[15] C.S. Wu et al., *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413, doi:10.1103/PhysRev.105.1413.

[16] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652, doi:10.1143/PTP.49.652.

[17] S.L. Glashow, J. Iliopoulos and L. Maiani, *Weak interactions with Lepton-Hadron symmetry*, Phys. Rev. D **2** (1970) 1285, doi:10.1103/PhysRevD.2.1285.

[18] Y. Nambu, *Axial Vector Current Conservation in Weak Interactions*, Phys. Rev. Lett. **4** (1960) 380, doi:10.1103/PhysRevLett.4.380.

[19] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, Phys. Rev. **117** (1960) 648, doi:10.1103/PhysRev.117.648.

[20] J. Goldstone, *Field Theories with Superconductor Solutions*, Il Nuovo Cimento **19** (1961) 154, doi:10.1007/BF02812722.

[21] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, Phys. Rev. **130** (1963) 439, doi:10.1103/PhysRev.130.439.

[22] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.

[23] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.

[24] P. W. Higgs, *Broken Symmetries, Massless Particles and Gauge Fields*, Phys. Lett. **12** (1964) 2 132, doi:10.1016/0031-9163(64)91136-9.

[25] G.S. Guralnik, C.R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[26] J. Goldstone, A. Salam and S. Weinberg, *Broken symmetries*, Phys. Rev. **127** (1962) 965, doi:10.1103/PhysRev.127.965.

[27] J. Ellis, *Higgs Physics*, *Proceedings, 2013 European School of High-Energy Physics*, 117 (2015), doi:10.5170/CERN-2015-004.117.

[28] D. Griffiths, *Introduction to elementary particles*, Wiley-VCH (2004).

[29] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys. Rept. **457** (2008) 1, doi:10.1016/j.physrep.2007.10.004.

[30] C. B. W. Lee, C. Quigg and H.B. Thacker, *Weak interactions at very high energies: The role of the Higgs-boson mass*, Phys. Rev. D **16** (1977) 1519, doi:10.1103/PhysRevD.16.1519.

[31] W. J. Marciano and S. Willenbrock, *Radiative corrections to heavy-Higgs-scalar production and decay*, Phys. Rev. D **37** (1988) 2509, doi:10.1103/PhysRevD.37.2509.

[32] S. Dawson and S. Willenbrock, *Unitarity constraints on heavy Higgs bosons*, Phys. Rev. Lett. **62** (1989) 1232, doi:10.1103/PhysRevLett.62.1232.

[33] M. Lüscher and P. Weisz, *Scaling laws and triviality bounds in the lattice ϕ^4 theory: (I). One-component model in the symmetric phase*, Nuclear Physics B **290** (1987) 25, doi:10.1016/0550-3213(87)90177-5.

[34] T. Hambye and K. Riesselmann, *Matching conditions and Higgs mass upper bounds revisited*, Phys. Rev. **D55** (1997) 7255, doi:10.1103/PhysRevD.55.7255.

[35] G. Degrassi et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, JHEP **08** (2012) 098, doi:10.1007/JHEP08(2012)098.

[36] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803.

[37] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables* (2011), doi:10.5170/CERN-2011-002.

[38] CMS Collaboration, *Search for the associated production of the Higgs boson and a vector boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV via Higgs boson decays to τ leptons*, Submitted to JHEP (2018), `https://cds.cern.ch/record/2637846`.

[39] CMS Collaboration, *Search for the associated production of a Higgs boson with a top quark pair in final states with a τ lepton at $\sqrt{s}$* = 13 TeV, Technical Report CMS-PAS-HIG-17-003 (2017), `https://cds.cern.ch/record/2257067`.

[40] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector* (2016), doi:10.23731/CYRM-2017-002.

[41] T. Asaka and K. Hikasa, *Four-fermion decay of Higgs bosons*, Phys. Lett. B **345** (1995) 36, doi:10.1016/0370-2693(94)01599-8.

[42] LHC Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties* (2013), `https://cds.cern.ch/record/1559921`.

[43] LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL, *Search for the standard model Higgs boson at LEP*, Phys. Lett. B **565** (2003) 61, doi:10.1016/S0370-2693(03)00614-2.

[44] Tevatron New Physics Higgs Working Group (CDF, D0), *Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data* (2012), `arXiv:1207.0449`.

[45] M. Baak et al., *Updated Status of the Global Electroweak Fit and Constraints on New Physics*, Eur. Phys. J. C **72** (2012) 2003, doi:10.1140/epjc/s10052-012-2003-4.

[46] 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, doi:10.1016/j.physletb.2012.08.020.

[47] 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, doi:10.1016/j.physletb.2012.08.021.

[48] CMS Collaboration, *Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV*, JHEP **06** (2013) 081, doi:10.1007/JHEP06(2013)081.

[49] ATLAS and CMS Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV*, JHEP **08** (2016) 045, doi:10.1007/JHEP08(2016)045.

[50] J. Ellis and T. You, *Updated Global Analysis of Higgs Couplings*, JHEP **06** (2013) 103, doi:10.1007/JHEP06(2013)103.

[51] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) 08 S08001, doi:10.1088/1748-0221/3/08/S08001.

[52] J. L. Caron, *LHC Layout* (1997), AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[53] J. L. Caron, *Cross Section of LHC Dipole.* (1998), AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[54] J. L. Caron, *Magnetic Field induced by the LHC Dipole's superconducting Coils.* (1998), AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[55] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) 08 S08003, doi:10.1088/1748-0221/3/08/S08003.

[56] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) 08 S08004, doi:10.1088/1748-0221/3/08/S08004.

[57] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) 08 S08005, doi:10.1088/1748-0221/3/08/S08005.

[58] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) 08 S08002, doi:10.1088/1748-0221/3/08/S08002.

[59] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) 08 S08007, doi:10.1088/1748-0221/3/08/S08007.

[60] LHCf Collaboration, *The LHCf detector at the CERN Large Hadron Collider*, JINST **3** (2008) 08 S08006, doi:10.1088/1748-0221/3/08/S08006.

[61] MoEDAL Collaboration, *The Physics Programme Of The MoEDAL Experiment At The LHC*, Int. J. Mod. Phys. A **29** (2014) 1430050, doi:10.1142/S0217751X14300506.

[62] A. Bejar and L. Rossi, *HiLumi LHC Technical Design Report: Deliverable: D1.10*, Technical report (2015), http://cds.cern.ch/record/2069130.

[63] M. Lamont, *Status of the LHC* **455** (2013) 1 012001, doi:10.1088/1742-6596/455/1/012001.

[64] CMS Collaboration, *Public CMS Luminosity Information*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[65] CMS Collaboration, *CMS detector design*, http://cms.web.cern.ch.

[66] CMS Collaboration, *Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays*, JINST **5** (2009) T03021, doi:10.1088/1748-0221/5/03/T03021.

[67] CMS Collaboration, *Description and Performance of Track and Primary-vertex Reconstruction with the CMS Tracker*, JINST **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009.

[68] CMS Collaboration, *CMS Physics: Technical Design Report, Vol. 1: Detector Performance and Software*, CERN, Geneva (2006), http://cds.cern.ch/record/922757.

[69] Y. Allkofer et al., *Design and Performance of the Silicon Sensors for the CMS Barrel Pixel Detector*, NIM A **584** (2008) 25, doi:10.1016/j.nima.2007.08.151.

[70] CMS Collaboration, *Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons*, JINST **5** (2009) T03007, doi:10.1088/1748-0221/5/03/T03007.

[71] CMS Collaboration, *Tracker Material Budget Plots*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots (2014).

[72] A. Zabi, *Test Beam Results on the Performance of the CMS Electromagnetic Calorimeter*, Technical report, CERN, Geneva (2006), http://cds.cern.ch/record/988640.

[73] CMS Collaboration, *Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data*, JINST **5** (2009) T03012, doi:10.1088/1748-0221/5/03/T03012.

[74] 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.

[75] CMS Collaboration, *The CMS trigger system*, JINST **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020.

[76] J. Wittmann et al., *Design and performance of the phase I upgrade of the CMS Global Trigger*, JINST **12** (2017) 01 C01046, doi:10.1088/1748-0221/12/01/C01046.

[77] F. Krauss, *Sketch of a* $t\bar{t}h$ *event.*, http://www.opensciencegrid.org/wp-content/uploads/2014/05/event.jpg.

[78] D. S. Drell and T. M. Yan, *Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies*, Phys. Rev. Lett. **25** (1970) 316, doi:10.1103/PhysRevLett.25.316.

[79] G. Curci, W. Furmanski and R. Petronzio, *Evolution of parton densities beyond leading order: The non-singlet case*, Nuclear Physics B **175** (1980) 1 27, doi:10.1016/0550-3213(80)90003-6.

[80] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **2014** (2014) 7 79, doi:10.1007/JHEP07(2014)079.

[81] P. 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.

[82] S. Alioli P. Nason C. Oleari and E. Re, *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.

[83] S. Alioli K. Hamilton P. Nason C. Oleari and E. Re, *Jet pair production in POWHEG*, JHEP **04** (2011) 081, doi:10.1007/JHEP04(2011)081.

[84] S. Alioli P. Nason C. Oleari and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, JHEP **04** (2009) 002, doi:10.1088/1126-6708/2009/04/002.

[85] T. Sjostrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) Supplement C 159, doi:10.1016/j.cpc.2015.01.024.

[86] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. C **53** (2008) 3 473, doi:10.1140/epjc/s10052-007-0490-5.

[87] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, JHEP **2012** (2012) 12 61, doi:10.1007/JHEP12(2012)061.

[88] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070.

[89] CMS Collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, Eur. Phys. J. C **76** (2016) 3 155, doi:10.1140/epjc/s10052-016-3988-x.

[90] Geant4 Collaboration, *Geant4 - a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 3 250 , doi:10.1016/S0168-9002(03)01368-8.

[91] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, JINST **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003.

[92] CMS Collaboration, *Track Reconstruction in the CMS Tracker* (2006), CMS-NOTE-2006-041, `https://cds.cern.ch/record/934067`.

[93] K. Rose, *Deterministic Annealing for Clustering, Compression, Classification, Regression and related Optimisation Problems*, Proceedings of the IEEE **86** (1998) 2210, doi:10.1109/5.726788.

[94] R. Frühwirth, W. Waltenberger and P. Vanlaer, *Adaptive Vertex Fitting*, Technical Report CMS-NOTE-2007-008, CERN, Geneva (2007), `https://cds.cern.ch/record/1027031`.

[95] 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.

[96] CMS Collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*, JINST **13** (2018) 06 P06015, doi:10.1088/1748-0221/13/06/P06015.

[97] W. Adam and others, *Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC*, J. Phys. C **31** (1995) 2005, doi:10.1088/0954-3899/31/9/N01.

[98] CMS Collaboration, *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at* $\sqrt{s}$ *= 8 TeV*, JINST **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005.

[99] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*k_t *Jet Clustering Algorithm*, JHEP **4** (2008) P63, doi:10.1088/1126-6708/2008/04/063.

[100] D0 Collaboration, *Observation of Single Top Quark Production*, `http://www-d0.fnal.gov/Run2Physics/top/singletop_observation`.

[101] CMS Collaboration, *Identification of b quark jets at the CMS Experiment in the LHC Run 2* (2016), CMS-PAS-BTV-15-001, `http://cds.cern.ch/record/2195743`.

[102] CMS Collaboration, *Performance of heavy flavour identification algorithms in proton-proton collisions at 13 TeV at the CMS experiment* (2017), CMS-DP-2017-012, `http://cds.cern.ch/record/2263801`.

[103] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, JINST **6** (2011) 11 P11002, doi:10.1088/1748-0221/6/11/P11002.

[104] CMS Collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, JINST **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014.

[105] CMS Collaboration, *Jet energy scale and resolution performances with 13 TeV data* (2016), CMS-DP-2016-020, `http://cds.cern.ch/record/2160347`.

[106] Particle Data Group, *Review of Particle Physics*, Chin. Phys. **C40** (2016) 10 100001, doi:10.1088/1674-1137/40/10/100001.

[107] CMS Collaboration, *Performance of* τ*-lepton reconstruction and identification in CMS*, JINST **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001.

[108] CMS Collaboration, *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2* (2016), CMS-PAS-TAU-16-002, http://cds.cern.ch/record/2196972.

[109] CMS Collaboration, *Tau-Id performance with full 2016 dataset using $Z \to \tau_\mu \tau_h$ events* (2017), CMS-DP-2017-002, http://cds.cern.ch/record/2243476.

[110] CMS Collaboration, *Alignment of the CMS silicon tracker during commissioning with cosmic rays*, JINST **5** (2010) 03 T03009, doi:10.1088/1748-0221/5/03/T03009.

[111] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013.

[112] CMS Collaboration, *Measurement of the weak mixing angle with the Drell-Yan process in proton-proton collisions at the LHC*, Phys. Rev. D **84** (2011) 112002, doi:10.1103/PhysRevD.84.112002.

[113] N. Bartosik, *Associated Top-Quark-Pair and b-Jet Production in the Dilepton Channel at $\sqrt{8}$ TeV as Test of QCD and Background to tt+Higgs Production* (2015), PhD Thesis, DESY-THESIS-2015-035.

[114] V. Blobel and C. Kleinwort, *A new method for the high precision alignment of track detectors*, *Advanced Statistical Techniques in Particle Physics. Proceedings, Conference, Durham, UK, March 18-22, 2002* (2002), http://www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings//blobel1.pdf.

[115] V. Blobel, *Software alignment for tracking detectors*, Nucl. Instrum. Meth. A **566** (2006) 5, doi:10.1016/j.nima.2006.05.157.

[116] V. Blobel, *Millepede II: Linear Least Squares Fits with a Large Number of Parameters* (2007), http://www.desy.de/~blobel/Mptwo.pdf.

[117] CMS Collaboration, *Alignment of the CMS tracker with LHC and cosmic ray data*, JINST **9** (2014) 06 P06009, doi:10.1088/1748-0221/9/06/P06009.

[118] C. Kleinwort, *General Broken Lines as advanced track fitting method*, Nucl. Instrum. Meth. A **673** (2012) 107, doi:10.1016/j.nima.2012.01.024.

[119] J. Draeger, *Track based alignment of the CMS silicon tracker and its implication on physics performance* (2011), PhD Thesis, DESY-THESIS-2011-026.

[120] CMS Collaboration (CMS), *Alignment of the CMS Silicon Strip Tracker during stand-alone Commissioning*, JINST **4** (2009) T07001, doi:10.1088/1748-0221/4/07/T07001.

[121] CMS Collaboration, *CMS Tracker Alignment Performance Results 2016* (2017), CMS-DP-2017-021, http://cds.cern.ch/record/2273267.

[122] CMS Collaboration, *CMS Tracker Alignment Performance Results Summer 2016* (2016), CMS-DP-2016-063, https://cds.cern.ch/record/2221746.

[123] CMS Collaboration, *Technical Proposal for the Upgrade of the CMS Detector through 2020* (2011), https://cds.cern.ch/record/1355706.

[124] CMS Collaboration, *CMS Technical Design Report for the Pixel Detector Upgrade* (2012), http://cds.cern.ch/record/1481838.

[125] CMS Collaboration, *CMS Tracker Alignment Performance Results Start-Up 2017* (2017), CMS-DP-2017-055, http://cds.cern.ch/record/2297528.

[126] ALEPH Collaboration, *Observation of an excess in the search for the standard model Higgs boson at ALEPH*, Phys. Lett. B **495** (2000) 1, doi:10.1016/S0370-2693(00)01269-7.

[127] DELPHI Collaboration, *Final results from DELPHI on the searches for SM and MSSM neutral Higgs bosons*, Eur. Phys. J. C **32** (2004) 145, doi:10.1140/epjc/s2003-01394-x.

[128] L3 Collaboration, *Standard model Higgs boson with the L3 experiment at LEP*, Phys. Lett. B **517** (2001) 319, doi:10.1016/S0370-2693(01)01010-3.

[129] OPAL Collaboration, *Search for the standard model Higgs boson in e^+e^- collisions at $\sqrt{s}$ =192-209 GeV*, Phys. Lett. B **499** (2001) 38, doi:10.1016/S0370-2693(01)00070-3.

[130] CDF Collaboration, *Search for a low mass Standard Model Higgs boson in the $\tau - \tau$ decay channel in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV*, Phys. Rev. Lett. **108** (2012) 181804, doi:10.1103/PhysRevLett.108.181804.

[131] D0 Collaboration, *Search for the standard model Higgs boson in tau lepton pair final states*, Phys. Lett. B **714** (2012) 237, doi:10.1016/j.physletb.2012.07.012.

[132] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117, doi:10.1007/JHEP04(2015)117.

[133] 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.

[134] W.J. Stirling, *13/8 TeV LHC luminosity ratios*, Private communication.

[135] CMS Collaboration (CMS), *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector*, Phys. Lett. **B779** (2018) 283, doi:10.1016/j.physletb.2018.02.004.

[136] NNPDF Collaboration, *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, Nucl. Phys. B **855** (2012) 153, doi:10.1016/j.nuclphysb.2011.09.024.

[137] L. Bianchini, J. Conway, E. K. Friis and C. Veelken, *Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques*, Journal of Physics: Conference Series **513** (2014) 2 022035, doi:10.1088/1742-6596/513/2/022035.

[138] CMS Collaboration, *Observation of the SM scalar boson decaying to a pair of τ leptons with the CMS experiment at the LHC* (2017), CMS-PAS-HIG-16-043, https://cds.cern.ch/record/2264522.

[139] L. and R. Chakravarti, *Handbook of Methods of Applied Statistics*, volume I, John Wiley and Sons (1967).

[140] CMS Collaboration, *Measurement of the differential cross section for $\mathrm{t\bar{t}}$ production in the dilepton final state at $\sqrt{s}$ = 13 TeV* (2016), CMS-PAS-TOP-16-011, https://cds.cern.ch/record/2140061.

[141] CMS Collaboration, *Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV*, Phys. Rev. D **95** (2017) 9 092001, doi:10.1103/PhysRevD.95.092001.

[142] D0 Collaboration, *Measurement of the Cross Section for Inclusive Z Production in Di-Muon Final States at $\sqrt{s}$ = 1.96 TeV* (2004), https://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/EW/E07/E07.pdf.

[143] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s}$ = 13 TeV*, JINST **13** (2018) 10 P10005, doi:10.1088/1748-0221/13/10/P10005.

[144] D. Jang, *Search for MSSM Higgs decaying to τ pairs in $p\bar{p}$ collision at $\sqrt{s}$ = 1.96 TeV at CDF* (2006), PhD Thesis, FERMILAB-THESIS-2006-11.

[145] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, JINST **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002.

[146] CMS Collaboration, *Performance of missing transverse momentum reconstruction algorithms in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the CMS detector* (2012), CMS-PAS-JME-12-002, http://cds.cern.ch/record/1543527.

[147] CMS Collaboration, *Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s}$ = 13 TeV*, Phys. Lett. B **772** (2017) 752, doi:10.1016/j.physletb.2017.07.047.

[148] CMS Collaboration, *Measurement of the WZ production cross section in pp collisions at $\sqrt{s}$ = 13 TeV*, Phys. Lett. B **766** (2017) 268, doi:10.1016/j.physletb.2017.01.011.

[149] D. de Florian, G. Ferrera, M. Grazzini and D. Tommasini, *Higgs boson production at the LHC: transverse momentum resummation effects in the $H \to 2\gamma$, $H \to WW \to \ell\nu\ell\nu$ and $H \to ZZ \to 4\ell$ decay modes*, JHEP **06** (2012) 132, doi:10.1007/JHEP06(2012)132.

[150] M. Grazzini and H. Sargsyan, *Heavy-quark mass effects in Higgs boson production at the LHC*, JHEP **09** (2013) 129, doi:10.1007/JHEP09(2013)129.

[151] J. Bellm et al., *Herwig++ 2.7 Release Note* (2013), https://arxiv.org/abs/1310.6877.

[152] I. W. Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, Phys. Rev. D **85** (2012) 034011, doi:10.1103/PhysRevD.85.034011.

[153] J. S. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, 115 (2011), doi:10.5170/CERN-2011-006.115.

[154] M. Cepeda et al. (Physics of the HL-LHC Working Group Collaboration), *Higgs Physics at the HL-LHC and HE-LHC*, Technical report, CERN (2018), `http://cds.cern.ch/record/2650162`.

[155] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Technical report, CERN, Geneva (2011), CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, `https://cds.cern.ch/record/1379837`.

[156] A. L. Read, *Linear interpolation of histograms*, Nucl. Instrum. Meth. **A425** (1999) 357, doi:10.1016/S0168-9002(98)01347-3.

[157] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Computer Physics Communications **77** (1993) 2 219, doi:10.1016/0010-4655(93)90005-W.

[158] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, Eur. Phys. J. **C70** (2010) 525, doi:10.1140/epjc/s10052-010-1470-8.

List of Figures

1.1 Standard Model elementary particles 6
1.2 Shape of the Higgs potential 15
1.3 Theory constraints on the Higgs boson mass 21
1.4 Stability of the Higgs potential 21
1.5 Feynman diagrams of Higgs boson production processes 22
1.6 Higgs boson production cross sections at the LHC 23
1.7 Feynman diagrams of loop-induced Higgs boson decays to $\gamma\gamma$, $Z\gamma$ and gg 24
1.8 Branching ratios of Higgs boson decays 25
1.9 Total decay width of the Standard Model Higgs boson 26
1.10 Higgs production and decay signal strengths for the combination of ATLAS and CMS Run 1 data 27
1.11 Higgs couplings to Standard Model particles from the combination of ATLAS and CMS Run 1 data 28

2.1 Schematic layout of the LHC. 33
2.2 Cross section of a LHC dipole magnet. 33
2.3 LHC beam pipes and dipole magnetic field lines. 34
2.4 Injection chain of the LHC. 34
2.5 Timeline of the LHC project. 35
2.6 Integrated luminosity delivered to CMS in 2015 and 2016. 36
2.7 Pileup profile of 2016 CMS data. 36
2.8 View of the CMS detector with all its subsystems. 38
2.9 Magnetic field of the CMS solenoid 39
2.10 Layout of the CMS tracker 40
2.11 Layout of the CMS pixel tracker 41
2.12 Material budget of the CMS tracker in units of radiation length X_0 42
2.13 Geometric view of one quarter of the CMS ECAL. 43
2.14 ECAL energy resolution 44
2.15 The CMS hadronic calorimeter 45
2.16 Layout of one quarter of the CMS muon system 46
2.17 Muon p_T resolution 47
2.18 Material thickness of the CMS detector. 48
2.19 Sketch of a $t\bar{t}H$ event produced from a proton-proton collision. 49

3.1 Particle interactions through the CMS detector 54
3.2 Sketch of a b-jet 64
3.3 Sketch of a hadronic tau decay 67
3.4 Size of the strips in η and ϕ used in the Hadron Plus Strips algorithm . . 68

3.5 Reconstructed τ_h mass and visible mass per decay mode 68
3.6 Tau identification ROC curves and efficiency as a function of p_T 69
3.7 Efficiency and fake rate of the τ_h anti-electron MVA discriminator 70

4.1 Concept of track-based alignment 76
4.2 Illustration of the telescope weak mode 78
4.3 Illustration of the twist weak mode 78
4.4 Hierarchical structure of the CMS tracker 79
4.5 Example of tracker geometries comparison 82
4.6 Example of distribution of median of residuals in the pixel tracker 83
4.7 Example of distribution of median of residuals in the strip tracker 84
4.8 Example of primary vertex validation 86
4.9 Example of $Z(\mu\mu)$ validation 87
4.10 Example of track-split validation 87
4.11 Geometry of the upgraded pixel detector 88
4.12 Comparison of the design and misaligned tracker geometries 90
4.13 Comparison of the realigned geometry after the pixel high level structures alignment wrt. ideal 94
4.14 Comparison of the realigned geometry after the pixel module level alignment wrt. ideal 95
4.15 Primary vertex validation for realignment scenarios with minimum bias events 96
4.16 Largest corrections after the pixel high-level structures alignment with 0 T cosmic ray tracks 102
4.17 Unbiased track-hit residuals for the barrel and forward pixel modules after alignment with 0 T cosmic ray tracks 103
4.18 Track-split validation after alignment with 0 T cosmic ray tracks 104
4.19 Unbiased track-hit residuals for the barrel and forward pixel modules after alignment with 3.8 T cosmic ray tracks 105
4.20 Track-split validation after alignment with 3.8 T cosmic ray tracks 106

5.1 Ratio of LHC parton luminosities for 13 TeV/8 TeV 109
5.2 Feynman diagram of the Drell-Yan process 113
5.3 Feynman diagram of W+jets production 113
5.4 Feynman diagrams of $t\bar{t}$ production processes 114
5.5 Distribution of the M_T variable in the $\mu\tau_h$ and $e\tau_h$ channels 120
5.6 Control distributions in the $\mu\tau_h$ channel 123
5.7 Control distributions in the $e\tau_h$ channel 127
5.8 Visible mass distribution in the $\mu\tau_h$ and $e\tau_h$ channels 128
5.9 Distributions of m_{vis} and $m_{\tau\tau}$ for the Higgs signal, in the $\mu\tau_h$ channel . . 130
5.10 Reconstructed Higgs boson transverse momentum in the $\mu\tau_h$ and $e\tau_h$ channels 130
5.11 Reweighing of LO Drell-Yan simulation 132
5.12 Effect of pileup reweighing correction in the $\mu\tau_h$ channel 136
5.13 Example of fit of the dimuon invariant mass 137
5.14 Example of fit of the dimuon invariant mass with FSR component 138
5.15 Muon identification and isolation scale factors 139
5.16 Electron identification and isolation scale factors 140

5.17 Muon trigger scale factors 142
5.18 Electron trigger scale factors 143
5.19 Effect of applying muon efficiency scale factors 144
5.20 Effect of applying electron efficiency scale factors 145
5.21 Components of the vector boson recoil 146
5.22 Effect of applying E_T^{miss} recoil corrections in $Z \to \mu\mu$ events 147

6.1 Signal and background 2D distributions from MC in the $\mu\tau_h$ channel . . 155
6.2 Control regions for the W+jets background 157
6.3 Control regions for the QCD background 158
6.4 Post-fit unrolled 2D distributions for the $\mu\tau_h$ channel 165
6.5 Post-fit unrolled 2D distributions for the $e\tau_h$ channel 166
6.6 Post-fit unrolled 2D distributions for the $\tau_h\tau_h$ channel 167
6.7 Post-fit unrolled 2D distributions for the $e\mu$ channel 168
6.8 Distribution of the decimal logarithm of the S/(S+B) ratio for all bins . . 169
6.9 Combined observed and predicted $m_{\tau\tau}$ distributions. 171
6.10 Observed and expected local p-value as a function of m_H 172
6.11 Best fit signal strength per category and channel 172
6.12 Scan of the negative log-likelihood as a function of κ_V and κ_f 173

List of Tables

2.1 Operational parameters of the LHC in 2016. 36
2.2 Main characteristics of the CMS tracker subsystems 42

3.1 Selection criteria for the medium working point of the particle-flow muon identification 60
3.2 Branching ratios of τ lepton decays 66

4.1 Selection of tracks used for tracker alignment 80
4.2 Assumed misalignment scenario after the pixel detector installation . . . 91
4.3 Mean and RMS of the differences of pixel modules positions before and after the realignment 93
4.4 Recovery from misalignment as a function of pileup 97
4.5 Recovery from misalignment as a function of the APE choice 98
4.6 Recovery from misalignment using cosmic ray events 100

5.1 Cross sections for Higgs boson production and Standard Model background processes at 8 and 13 TeV 110
5.2 Branching fractions of $\tau\tau$ final states 111
5.3 List of datasets used for the analysis 112
5.4 List of simulation samples for background processes 115
5.5 List of simulation samples for Higgs boson production processes 116
5.6 Offline event selection for the $\mu\tau_h$ channel 122
5.7 Offline event selection for the $e\tau_h$ channel 126
5.8 Extrapolation factors from SS to OS events for the QCD estimation . . . 134
5.9 Tau energy scale corrections 149

6.1 Signal categories definitions and variables used for signal extraction . . 154
6.2 Event yields the $\mu\tau_h$ and $e\tau_h$ channels categories 156
6.3 Overview of experimental uncertainties 163
6.4 Expected and observed event yields in the most sensitive bins of each channel 170

Acknowledgements

This doctoral work would not have been possible without the advice, help and encouragement of many people. This last page is for them.
First, I want to thank my Professor **Elisabetta Gallo** for giving me the opportunity to make my PhD in the DESY CMS group. I must admit I was sometimes surprised how much she trusted me, allowing a lot of freedom to organize my work the way it better suited to me. At the same time, she has always found the time to give me precious advices and feedback. It has been a great pleasure being part of her team. Her deep respect and understanding of all colleagues ensures a very relaxed and friendly atmosphere among the group, which I enjoyed very much.
I am deeply thankful to my supervisor **Alexei Raspereza**, who has guided me step-by-step into the Higgs to tau tau analysis. I am impressed by his immense knowledge and passion for research. The dedication and energy he puts in his work are really extraordinary, and it has been an honour to learn from him. I thank him for the time and patience he spent explaining me this or the other topic or line of code, as well as for his precious advices.
Then, I would like to thank the people with which I collaborated the most during my PhD time, in the context of the Higgs to tau tau analysis. In the first months, **Francesco** helped me to get familiar with the analysis techniques and technicalities. Then, **Teresa** joined me in the most exciting (and tiring) time of the analysis, sharing with me even the dirty work and my worries. During all my PhD time, I could exchange my findings and problems with **Yiwen**. I also want to thank the colleagues of the U.S. groups with which I worked in close contact, especially **Cècile**, **Laura** and **Tyler**.
Unlike my first expectations, the work on tracker alignment turned out to be very interesting and challenging. In this respect, I am thankful to **Gregor** and **Matthias** for introducing me to this task and for guiding me through all the details with immense patience. The expertise of **Claus** and **Rainer** has helped me to gain a deeper understanding of the CMS tracker alignment.
Finally, I want to thank my **family** for constantly supporting and encouraging me. I thank my parents, my sister and my brother, for trusting my intuition when I decided to move to Hamburg. Despite the distance, they always find a way to be there for me, and going back home even only for a few days is a great joy. The last and most important sentence is dedicated to **my men**, as I like to call them. They are *my* family, making me happy with their smiles and unconditional love.