Karlsruher Institut für Technologie

IEKP-KA/2016-07

Packaging and assembly technologies for the pixel detector upgrade and measurement of $\tau\tau$ final states with the CMS experiment at the LHC

Fabio Colombo

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
KIT – Karlsruher Institut für Technologie

genehmigte

DISSERTATION

von

M.Sc. Fabio Colombo
aus Merate

Tag der mündlichen Prüfung: 1. Juli 2016

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Ulrich Husemann

Abstract

The work presented in this thesis has been carried out in the context of $H \to \tau\tau$ measurements with the CMS experiment at the CERN Large Hadron Collider, for which the CMS pixel detector is a key component. In view of a complete replacement of the pixel detector, planned for the end of 2016, a substantial fraction of the pixel modules for the future detector has been assembled and tested. An in-house packaging process for the interconnection of the silicon sensor substrate with the front-end chips via solder bumps has been developed and optimized. Quality control procedures for the module assembly line have been defined. The produced modules have been found to possess excellent mechanical and electrical properties.
In a parallel development of the algorithms used for reconstruction and identification of hadronic tau decays τ_{h}, a new multivariate discriminator trained for rejecting electrons misidentified as τ_{h} candidates has been developed. The performance of the new discriminator and of the overall CMS tau identification sequence has been validated in the first $2.3\,\mathrm{fb}^{-1}$ of data collected by the experiment in proton-proton collisions at a center of mass energy of 13 TeV. A measurement of the τ_{h} energy scale in data has been performed as well. A deeper validation of the tau identification performance has been obtained in a global determination of the $\mathrm{Z} \to \tau\tau$ cross section, of the τ_{h} identification efficiency and of the τ_{h} energy scale. The extracted value for $\sigma(\mathrm{Z} \to \tau\tau) \times \mathrm{BR}(\mathrm{Z} \to \tau\tau)$ has been found to be well compatible with the NNLO theoretical prediction. As a by-product of the measurement, an additional estimate of the τ_{h} identification efficiency and of the τ_{h} energy scale is extracted with respective precisions of 3.4% and 1%, finding good agreement with the measurements performed using different techniques. The results prove the overall performance of the CMS τ_{h} identification and help in constraining the most important systematic uncertainties affecting future $H \to \tau\tau$ measurements.

Contents

Chapter 1

Introduction

This first chapter gives a brief introduction to the Standard Model of particle physics, together with a description of the electroweak symmetry breaking mechanism introduced by Brout, Englert and Higgs to explain the origin of the masses of the weak bosons and of the fundamental fermions. The second part gives an overview of the Large Hadron Collider, currently one of the most advanced playgrounds to investigate and probe the high energy domain of particle physics. Finally, some general remarks about the physics at hadron colliders are presented, together with a few experimental concepts (e.g., the luminosity of a particle collider, or the definition of "pile-up" interactions) which will be useful for a deeper understanding of the following chapters.

1.1 The Standard Model of particle physics

1.1.1 Forces and particles

In the current best description of the physical world which surrounds us, matter is ultimately made of elementary and indivisible constituents (*particles*) which interact among themselves via four fundamental forces, or interactions. The *gravitational force* is responsible for the mutual attraction of massive objects, the *electromagnetic force* is responsible for the interaction between particles possessing an electric charge, the *weak nuclear force* is responsible for radioactive decays and the *strong nuclear force* for the stability of the atomic constituents (protons and neutrons). While a complete inclusion of the gravitational force in this general scheme still has to be achieved, the electromagnetic, weak and strong interactions can be nicely framed in a theory called Standard Model of particle physics[1]. This is a relativistic quantum field theory (QFT) based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry group, it has perturbative behaviour at high energies and it is renormalizable.
The Standard Model includes 12 elementary particles of spin $\frac{1}{2}$ known as fermions, each one with the corresponding antiparticle having opposite quantum numbers and the same couplings. The fundamental fermions are divided in two groups: six *leptons* and six *quarks*, with each group further divided into three families, or generations, exhibiting a similar physics behaviour (see table 1.1).
The three lepton families are the electron (e), the muon (μ) and the tau (τ), each one with an

[1] Even if the lack of a satisfactory inclusion of the gravitational force into the Standard Model is an important missing piece for the theory, it should be noted that, given the extremely small mass of the fundamental particles, the strength of this force is negligible compared to the other three.

associated neutrino (ν_e, ν_μ, ν_τ).
The six quarks are labelled according to their "flavour" as: up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Quarks have a fractional electric charge ($Q/|e| = \frac{2}{3}, -\frac{1}{3}$). Moreover, they carry an additional charge, referred to as "colour", which can assume three possible values (red, blue, green). A phenomenon called "colour confinement" results in quarks being bound to one another, forming colour-neutral composite particles named *hadrons*, containing either a quark and an antiquark $q\bar{q}$ (*mesons*) or three quarks $qq'q''$ (*baryons*). Protons and neutrons are by far the most abundant baryons in the universe and are made of three quarks, in the combinations "uud" and "ddu", respectively.

1^{st} generation		2^{nd} generation		3^{rd} generation		$Q/\|e\|$
e	$511\,\mathrm{keV}/c^2$	μ	$105.7\,\mathrm{MeV}/c^2$	τ	$1.777\,\mathrm{GeV}/c^2$	-1
ν_e	~ 0	ν_μ	~ 0	ν_τ	~ 0	0
u	$2.3^{+0.7}_{-0.5}\,\mathrm{MeV}/c^2$	c	$1.275 \pm 0.025\,\mathrm{GeV}/c^2$	t	$173.21 \pm 0.51 \pm 0.71\,\mathrm{GeV}/c^2$	2/3
d	$4.8^{+0.5}_{-0.3}\,\mathrm{MeV}/c^2$	s	$95 \pm 5\,\mathrm{MeV}/c^2$	b	$4.18 \pm 0.03\,\mathrm{GeV}/c^2$	$-1/3$

Table 1.1: Fundamental spin-$\frac{1}{2}$ fermions, with their respective mass and electric charge, expressed in units of elementary charge: leptons (first row) and quarks (second row). The numerical values for the quark masses and their uncertainties are taken from [1].

The quantum field operators associated with fermions are four-components Dirac spinors, denoted as ψ in the following. Introducing the Weyl representation of the γ matrices:

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \qquad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

where the σ^i are the 2×2 Pauli matrices, the adjoint spinor can be built as $\bar{\psi} = \psi^\dagger\gamma^0$, with the dagger notation † representing the conjugate transpose. For a fermion of mass m, the free Lagrangian (i.e., without introducing any kind of interaction) is:

$$\mathcal{L}_{\mathrm{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi\,, \tag{1.1}$$

from which the following equation of motion (Dirac equation) can be derived:

$$(i\gamma^\mu\partial_\mu - m)\psi = 0\,. \tag{1.2}$$

For convenience, the Dirac spinor ψ is usually separated into the left-handed and the right-handed spinors $\psi = \psi_L + \psi_R$, obtained applying the projection operators P_L and P_R:

$$\psi_L = P_L\psi = \frac{1}{2}(1-\gamma^5)\psi\,, \qquad \psi_R = P_R\psi = \frac{1}{2}(1+\gamma^5)\psi\,. \tag{1.3}$$

The left and the right handedness of the spinors is called "chirality".
In the Standard Model, interactions between fermions are mediated by the exchange of spin-1 particles (bosons).
The $SU(2)_L \times U(1)_Y$ group is associated with electroweak interactions, which are the unified description of electromagnetic and weak interactions. The long-range electromagnetic interaction is mediated by the massless photon, while the short-range weak force carriers are the massive W^+, W^- and Z bosons. The $SU(2)_L$ gauge bosons couple only to the left-handed components ψ_L of the fermion fields, leading to the observed parity-violating character of weak interactions[2].

[2] A parity transformation is defined as the reversal of the spatial coordinates of a certain state. All fundamental interactions of elementary particles, with the exception of the weak interactions, are symmetric under parity.

The $U(1)_Y$ gauge boson couples to both the left-handed and the right-handed components. The left-handed projections of the fermion fields are grouped into $SU(2)_L$ doublets:

$$f_L = \left\{ \begin{pmatrix} \nu_\mathrm{e} \\ \mathrm{e} \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix}_L, \begin{pmatrix} \mathrm{c} \\ \mathrm{s} \end{pmatrix}_L, \begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix}_L \right\},$$

while the right-handed components are $SU(2)_L$ singlets:

$$f_R = \{\mathrm{e}_R,\ \mu_R,\ \tau_R,\ \mathrm{u}_R,\ \mathrm{d}_R,\ \mathrm{c}_R,\ \mathrm{s}_R,\ \mathrm{t}_R,\ \mathrm{b}_R\}\,.$$

A "weak isospin" charge $T = \frac{1}{2}$ is associated to each doublet: neutrinos and up-type quarks possess a third isospin component $T_3 = \frac{1}{2}$, while e_L, μ_L, τ_L and the down-type left-handed quarks have component $T_3 = -\frac{1}{2}$. The singlets f_R are conventionally assigned a null weak isospin charge ($T = 0$). The absence of right-handed neutrinos in f_R will be discussed later.
The $SU(3)_C$ colour group is associated with the strong interaction between quarks, which is described by quantum chromodynamics (QCD). The three different colour states in which each quark appears belong to a $SU(3)_C$ triplet, while leptons are colourless singlets. The quanta of the strong interaction field, called gluons, have spin 1, zero mass and carry colour charge themselves, forming eight possible colour states ("colour octet"). This has the consequence that gluons can also interact among each other, leading to the already mentioned phenomenon of quark confinement.
A free boson having mass m and spin 0 is represented in QFT by a complex scalar field ϕ, whose dynamics is described by the following Klein-Gordon Lagrangian, from which the Euler-Lagrange equation of motion can be extracted:

$$\mathcal{L}_\mathrm{KG} = (\partial_\mu\phi)^\dagger(\partial^\mu\phi) - m^2\phi^\dagger\phi \qquad \Rightarrow \qquad (\Box + m^2)\phi = 0\,. \tag{1.4}$$

For vector (i.e., spin-1) bosons, the associated representation is a vector field A_μ, whose dynamics is described by:

$$\mathcal{L}_\mathrm{Proca} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2A_\mu A^\mu \qquad \Rightarrow \qquad \partial_\mu F^{\mu\nu} + m^2A^\nu = 0\,, \tag{1.5}$$

where the antisymmetric field strength tensor is defined as $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$.
So far, only a non-interacting theory made of fermions and force mediators has been introduced. The Standard Model approach to account for interactions between particles is the requirement of local gauge invariance of the Lagrangian, as explained in the next section.

1.1.2 Gauge invariance principle and particle interactions

The requirement of a symmetry in the Lagrangian translates into a conservation of charges, via Noether's theorem, and allows for the introduction of new fields and interactions.
In QFT it proves very convenient to require that the Lagrangian is invariant under some symmetry transformation groups. For example, in the simplest case of the Abelian symmetry group $U(1)$, the transformation is just a phase multiplication having the form $\psi \to \psi' = e^{ie\alpha}\psi$, where e is some constant. If also the phase α is constant in time and space, this is referred to as a "global" phase transformation, whereas if it differs from point to point ($\alpha = \alpha(x)$) it is a "local" phase transformation.
The Dirac Lagrangian (1.1) that describes a free fermion field is invariant under global phase transformations because, trivially, $\partial_\mu\psi' = e^{ie\alpha}\partial_\mu\psi$ and the $e^{ie\alpha}$ term cancels out the $e^{-ie\alpha}$ which comes from the adjoint spinor $\bar{\psi}$. Physically speaking, this implies that in the system there is

a conserved current which, arranging the constants in the appropriate way, can be interpreted as the electromagnetic current $J_e^\mu = -e\bar{\psi}\gamma^\mu\psi$ and a conserved charge, obtained integrating the fourth component of the current on the three-dimensional space.
If the phase is made dependent from x^μ, the Lagrangian picks up an extra term under the transformation and the invariance is lost. One way to recover it, is to apply a "minimal substitution", replacing every derivative ∂_μ with the *covariant derivative*:

$$D_\mu \equiv \partial_\mu - ieA_\mu \,,$$

where the new field A_μ changes under the transformation as: $A_\mu \to A_\mu + \partial_\mu\alpha(x)$.
Therefore, the price paid in order to preserve the invariance of the Lagrangian is the introduction of a new vector field A_μ, called "gauge field", which needs its own free term $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ into the full Lagrangian. Moreover, expanding the covariant derivative, it is easy to see that this field couples with the fermion field ψ through a term of the form $-J_e^\mu A_\mu$.
It is important to notice that, in order not to spoil the local gauge invariance, the new field A_μ is required to be massless (that is, the term in equation (1.5) that depends from m^2 must be dropped). This is exactly the quantum field description of the electromagnetic interaction (QED), where the massless spin-1 boson A_μ is identified as the photon.
The generic transformation $\psi \to \psi' = U\psi$, where U is a $N \times N$ unitary matrix ($U^\dagger U = \mathbb{1}$) can also be considered. Any unitary matrix can be written in the form $U = e^{iH}$, where H is Hermitian ($H^\dagger = H$). Moreover, the most general Hermitian $N \times N$ matrix can be decomposed in the form $H = \theta \cdot \mathbb{1} + \sum_k \alpha_k \cdot t_k$, where $\mathbb{1}$ is the unit matrix and t_k are $N^2 - 1$ matrices that can be identified with the generators of the group. In group-theoretical language, this means that $U(N) = U(1) \times SU(N)$.
Every element of the $SU(N)$ group can be written as:

$$S = S(\vec{\alpha}) = e^{ig\alpha_k t_k} \qquad\qquad k = 1, \ldots, N^2 - 1 \,,$$

where g is a constant that will determine the interaction strength of the field.
As before, it is straightforward to prove the invariance under a global $SU(N)$ transformation of any Lagrangian which depends from some field ψ and from the derivatives $\partial_\mu\psi$. Requiring the invariance to hold also locally turns out to be more subtle compared to the $U(1)$ case because, in general, the symmetry group can be non-Abelian. This means that the generators follow the non trivial commutation relations $[t_i, t_j] = if_{ijk}t_k$, where the f_{ijk} are called "structure constants" of the group.
Adapting the previous idea, the invariance can be restored introducing a set of new vector fields A_μ^k and replacing the usual derivative with the covariant derivative:

$$D_\mu \equiv \partial_\mu - ig\mathcal{A}_\mu \qquad \text{with} \qquad \mathcal{A}_\mu = \sum_{k=1}^{N^2-1} t^k A_\mu^k \,.$$

Substituting the covariant derivative into equation (1.1), the Lagrangian reads:

$$\begin{aligned} \mathcal{L}_{\text{Dirac}} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^\mu t^k A_\mu^k\psi \,, \end{aligned}$$

where the last term expresses the coupling between the fermion field and the new vector fields. The field $\mathcal{A}_\mu$ (or, equivalently, the $N^2 - 1$ gauge fields A_μ^k) must be assigned a transformation rule such that $D_\mu\psi \to S(D_\mu\psi)$.
This translates in the requirement that:

$$\mathcal{A}_\mu \to \mathcal{A}'_\mu = S\mathcal{A}_\mu S^{-1} - \frac{i}{g}(\partial_\mu S)S^{-1} \tag{1.6}$$

and

$$A^k_\mu \to A'^k_\mu = A^k_\mu + \partial_\mu \alpha^k - 2g f^{ijk} \alpha^i A^j_\mu \,. \tag{1.7}$$

Finally, in order to give these gauge fields a free term, a tensor $F_{\mu\nu}$ antisymmetric in its spatial indexes has to be introduced. The definition $-igF^k_{\mu\nu} t^k = [D_\mu, D_\nu]$ or, more explicitly

$$F^i_{\mu\nu} \equiv \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g f^{ijk} A^j_\mu A^k_\nu$$

preserves local gauge invariance.
The ingredients mentioned so far are sufficient to describe the particle interactions at quantum level: what is needed are two non-Abelian symmetries and an Abelian one. First of all, a local $U(1)$ phase invariance is introduced in the Lagrangian, calling B_μ the gauge field associated with it. A second invariance, under a set of non-Abelian transformations that form a $SU(2)$ group, leads to the introduction of three W^i_μ fields ($i = 1, 2, 3$), one for each of the generators $\tau^i/2$, where τ^i is just another common notation for the set of 2×2 complex Hermitian and unitary Pauli matrices. The third invariance, also non-Abelian, under a set of transformations that form an $SU(3)$ group, requires the introduction of eight G^a_μ fields ($a = 1, \ldots, 8$).
The general transformation is then given by:

$$U = \exp\left[i \left(g' \beta(x) \frac{Y}{2} + g \alpha^i(x) \frac{\tau^i}{2} + g_s \gamma^a(x) \frac{\lambda^a}{2} \right) \right] \tag{1.8}$$

and the covariant derivative, which ensures the invariance of the theory under all the three transformations, takes the form:

$$D_\mu \equiv \partial_\mu - i g' \frac{Y}{2} B_\mu - i g \frac{\tau^i}{2} W^i_\mu - i g_s \frac{\lambda^a}{2} G^a_\mu \,, \tag{1.9}$$

where the scalar Y and the matrices[3] τ^i and λ^i are the generators for the $U(1)$ hypercharge, $SU(2)$ weak isospin and $SU(3)$ colour charge groups, respectively. The way fermions behave under gauge transformations depends on the charge they carry with respect to each interaction.

- $SU(3)_C$: only quarks have colour charge, and appear as colour triplets under $SU(3)$ transformations. Other leptons transform as colour singlets.
- $SU(2)_L$: recalling the chiral decomposition of ψ into the ψ_L and ψ_R spinors (equation (1.3)), the weak isospin charge is experimentally found to be different for left and right-handed particles. Left-handed fermions transform as isospin doublets, while right-handed ones are singlets of 0 weak isospin, and therefore do not interact with the gauge bosons of this symmetry group. This chiral nature of the weak isospin transformations has an immediate consequence. Fermion mass terms in the Lagrangian are written as

  $$-m\bar{\psi}\psi = -m\bar{\psi}\left(P_L + P_R\right)\psi = -m\left(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L\right) \,, \tag{1.10}$$

  which manifestly violate gauge invariance, since ψ_L is a member of an isospin doublet and ψ_R is a singlet. Therefore, fermion mass terms cannot be included into the theory in this naive way.
- $U(1)_Y$: the $U(1)$ hypercharge induces transformations as singlets and is non-zero for all fermions except for the right-handed neutrinos. As a convention, the corresponding quantum number for left-handed leptons is chosen to be $Y_L = -1$.

[3] The λ^a are the Gell-Mann traceless and Hermitian matrices.

Since right-handed neutrinos do not couple to any of the previously introduced interactions, they can be regarded as "sterile" and are not included into the theory.
Restricting to the electroweak sector alone, the Lagrangian must also include terms for the free gauge fields, which can be written as:

$$-\frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\,. \tag{1.11}$$

Following the $SU(2)$ algebra outlined above:

$$\begin{aligned} W^i_{\mu\nu} &\equiv \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon_{ijk}W^j_\mu W^k_\nu\,, \\ B_{\mu\nu} &\equiv \partial_\mu B_\nu - \partial_\nu B_\mu\,. \end{aligned} \tag{1.12}$$

From the above equation for $W_{\mu\nu}$, self-interaction terms among the gauge bosons, due to the non-Abelian character of the $SU(2)$ gauge symmetry, are visible.
Recalling equation (1.7), the following relations which express the transformation law for the vector gauge fields can be derived:

$$\begin{aligned} \vec{W}_\mu \to \vec{W}'_\mu &= \vec{W}_\mu + \partial_\mu\vec{\alpha}(x) - g\left(\vec{\alpha}(x)\times\vec{W}_\mu\right)\,, \\ B_\mu \to B'_\mu &= B_\mu + \partial_\mu\beta(x)\,. \end{aligned} \tag{1.13}$$

Unlike strong interactions, identified with the $SU(3)_C$ symmetry group, the $U(1)_Y$ and $SU(2)_L$ gauge interactions do not directly correspond to the electromagnetic and weak forces, respectively. The observed interactions are instead a manifestation of the combined $SU(2)_L \times U(1)_Y$ gauge group, where the physical fields A_μ, Z_μ and $W^\pm_\mu$ for, respectively, the photon, the Z boson and the $\mathrm{W}^\pm$ bosons, arise as combinations of the gauge fields according to:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right)\,, \\ \begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} &= \begin{pmatrix} \cos\theta_\mathrm{W} & \sin\theta_\mathrm{W} \\ -\sin\theta_\mathrm{W} & \cos\theta_\mathrm{W} \end{pmatrix}\begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}\,, \end{aligned} \tag{1.14}$$

where θ_W is the weak mixing angle (or Weinberg angle). The relation between electric charge, hypercharge and weak isospin is given by the Gell-Mann-Nishijima formula:

$$Q = T_3 + \tfrac{Y}{2}\,.$$

Up to this point, not only are the fermions forced to be massless, but also gauge boson mass terms are not allowed, since an explicit term $\propto \frac{1}{2}W^i_\mu W^\mu_i$ or $\frac{1}{2}B_\mu B^\mu$ cannot be included in the full Lagrangian. This works perfectly fine for the photon and the gluon, which are actually found to be massless in nature, but is experimentally contradicted by the fact that the mediators of the weak force have to be massive. An elegant solution for the conflict between massless particles, as required by the theory, and massive fermions and vector bosons is provided by the spontaneous breaking of the symmetry.

1.1.3 Spontaneous symmetry breaking and Higgs mechanism

If a theory is described by a Lagrangian which possesses a given symmetry, but its physical "ground state" (that is, the state with the lowest energy) does not, the symmetry is said to be spontaneously broken.
An important consequence of the spontaneous symmetry breaking in QFT, which happens when

the original symmetry is continuous, is the appearance of massless and scalar (spin-0) particles. This statement is the result of a theorem which goes under the name of Nambu-Goldstone theorem [2] and the newly appeared scalar fields are therefore referred to as "Goldstone bosons". The number of Goldstone bosons of the broken theory coincides with the number of continuous symmetries which are broken by the choice of a specific ground state.

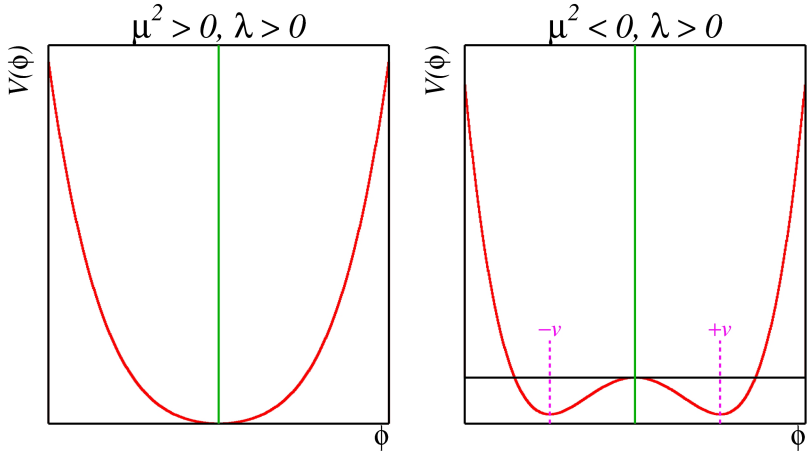


Figure 1.1: Form of the potential $V(\phi)$ depending on the sign of μ^2, positive on the left and negative on the right.

In the Standard Model, an external field is needed to break the electroweak gauge symmetry and this role is taken by the Higgs field [3–5]. In order to generate masses for the three gauge bosons $W^{\pm}$ and Z, without generating a photon mass, at least three degrees of freedom are necessary. The simplest solution is to add a complex $SU(2)$ doublet of scalar fields with hypercharge $Y = 1$:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{1.15}$$

This doublet has no colour charge and therefore it will not affect the $SU(3)_C$ sector. The Lagrangian for the Higgs field is given by:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi) \qquad \text{with} \qquad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi \right)^2 , \tag{1.16}$$

which is invariant under $SU(2)_L \times U(1)_Y$.
The expected form of the potential is sketched in figure 1.1. For $\mu^2 > 0$ the scalar potential has a global minimum at $\langle 0|\Phi|0\rangle = \Phi_0 = 0$, which would not break the electroweak gauge symmetry. For $\mu^2 < 0$ the potential has a circle of degenerate minima at

$$\langle 0|\Phi|0\rangle = \Phi_0 = -\frac{\mu^2}{2\lambda} = \frac{1}{2} v^2 \qquad \text{where} \qquad v \equiv \sqrt{-\frac{\mu^2}{\lambda}} . \tag{1.17}$$

The term v is the vacuum expectation value (VEV) of the field Φ.
The spontaneous breaking of the $SU(2)$ symmetry consists in choosing a particular ground state, around which the Higgs field $\Phi(x)$ is expanded. The particular vacuum chosen is:

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{1.18}$$

The field Φ_0 has charge under the weak isospin and the weak hypercharge symmetry:

$$T_3\Phi_0 = -\frac{1}{2}\Phi_0 \qquad \text{and} \qquad Y\Phi_0 = \Phi_0\,.$$

On the other hand, since Φ_0 is neutral, the $U(1)_Q$ symmetry remains unbroken, that is

$$Q\Phi_0 = \left(T_3 + \frac{Y}{2}\right)\Phi_0 = 0 \qquad \Rightarrow \qquad \Phi_0 \to \Phi_0' = e^{i\delta(x)\frac{Q}{2}}\Phi_0 = \Phi_0\,. \tag{1.19}$$

Thus, $SU(2)_L$ and $U(1)_Y$ are completely broken separately, but the product group $SU(2)_L \times U(1)_Y$ is not: after the symmetry breaking, a residual symmetry generated by Q will remain. This pattern of symmetry breakdown is described in formula as: $SU(2)_L \times U(1)_Y \to U(1)_Q$.
Labeling the fluctuation of the real scalar fields ϕ_1, ϕ_2, ϕ_3 and ϕ_4 around the minimum as θ_2, θ_1, H and $-\theta_3$, the doublet can be written as[4]:

$$\begin{aligned}
\Phi(x) &= \frac{1}{\sqrt{2}}\begin{pmatrix} \theta_2(x) + i\theta_1(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix} \\
&\simeq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 + i\theta_3/v & i(\theta_1 - i\theta_2)/v \\ i(\theta_1 + i\theta_2)/v & 1 - i\theta_3/v \end{pmatrix}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \\
&\simeq \frac{1}{\sqrt{2}} e^{i\frac{2\theta^i(x)}{v}\frac{\tau^i}{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}.
\end{aligned} \tag{1.20}$$

Thanks to the $SU(2)$ invariance of the Lagrangian, the three fields $\theta_i(x)$ in equation (1.20) can be gauged away with a transformation $U = \exp\left(-i\frac{2\theta^i(x)}{v}\frac{\tau^i}{2}\right)$: these are the massless Goldstone bosons, which do not explicitly appear in the final Lagrangian. This particular gauge fixing in which the Goldstone boson components are set to zero, making the number of scalar degrees of freedom minimal, is often called "unitary gauge".
Expanding the scalar Higgs field Lagrangian (1.16) around Φ_0 using

$$\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.21}$$

one finds:

$$\begin{aligned}
\mathcal{L}_{\text{Higgs}} &= \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}2v^2\lambda H^2 - \frac{1}{3!}6v\lambda H^3 - \frac{1}{4!}6\lambda H^4 \\
&+ \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{-\dagger} W^{-\mu} + \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{+\dagger} W^{+\mu} \\
&+ \frac{1}{2}\frac{v^2(g^2 + g'^2)}{4}\left(\frac{gW_\mu^3 - g'B^\mu}{\sqrt{g^2 + g'^2}}\right)^2 + 0 \cdot \left(\frac{g'W_\mu^3 + gB^\mu}{\sqrt{g^2 + g'^2}}\right)^2 \\
&+ \frac{1}{4}\left(2vH + H^2\right)\left[g^2 W_\mu^- W^{+\mu} + \frac{1}{2}(g^2 + g'^2)\left(\frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}}\right)^2\right].
\end{aligned} \tag{1.22}$$

In the first line, originated from the expansion of the potential $V(\Phi)$, the kinetic term for the Higgs boson, its mass term and the Higgs boson self-interaction terms are visible. The Higgs mass itself is equal to $m_{\text{H}} = v\sqrt{2\lambda}$. The term v can be put in relation to the Fermi constant G_F and therefore estimated from precise muon lifetime measurements as $v = \left(\sqrt{2}G_F\right)^{-1/2} \simeq 247\,\text{GeV}$.

[4] In the following expression only small fluctuations around the minimum are considered.

The parameter λ remains a free parameter of the model, meaning that the Higgs mass is not predicted by the theory.
In the second line, coming from the kinetic term $(D_\mu\Phi)^\dagger (D^\mu\Phi)$, the $\mathrm{W}^\pm$ vector bosons can be identified in the linear combination of the gauge bosons $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$. The process of spontaneous symmetry breaking allows them to acquire mass.
The third line provides the right mass terms for the observed Z and γ vector bosons[5]. The first linear combination of the gauge fields W^3_μ and B_μ comes with an appropriate mass term and it is therefore interpreted as the massive Z boson. The second combination of fields is orthogonal to the first one and it is added by hand with an associated null mass.
The results can be therefore interpreted as:

$$\begin{aligned} m_\mathrm{W} &= \frac{1}{2}vg \qquad \text{with} \qquad W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right), \\ m_\mathrm{Z} &= \frac{1}{2}v\sqrt{g^2+g'^2} \qquad \text{with} \qquad Z_\mu = \frac{gW^3_\mu - g'B_\mu}{\sqrt{g^2+g'^2}}, \\ m_\gamma &= 0 \qquad \text{with} \qquad A_\mu = \frac{g'W^3_\mu + gB_\mu}{\sqrt{g^2+g'^2}}. \end{aligned} \tag{1.23}$$

The gauge bosons have "eaten" the three massless Goldstone bosons, acquiring mass. The degrees of freedom of the Goldstone bosons are in fact needed: once the gauge bosons become massive, an additional degree of freedom is required in order to allow them to have a longitudinal polarization. The unbroken $U(1)_Q$ symmetry causes the photon to remain massless.
The mixing of W^3_μ and B_μ yielding the physical force carriers can be interpreted as a rotation of parameter θ_W, where

$$\frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_\mathrm{W} \qquad \text{and} \qquad \frac{g'}{\sqrt{g^2+g'^2}} = \sin\theta_\mathrm{W}. \tag{1.24}$$

Therefore, the following relation between the weak bosons masses can be inferred:

$$m_\mathrm{Z} = \frac{m_\mathrm{W}}{\cos\theta_\mathrm{W}}. \tag{1.25}$$

Finally, in the last line of the Lagrangian (1.22), the cubic and quartic couplings of the Higgs boson to the weak gauge bosons can be deduced. In particular, the coupling of one single Higgs boson to a pair of W or Z bosons is given by:

$$\begin{aligned} g_\mathrm{HWW} &= gm_\mathrm{W} = \frac{2m^2_\mathrm{W}}{v}, \\ g_\mathrm{HZZ} &= \frac{g}{2\cos\theta_\mathrm{W}}m_\mathrm{Z} = \frac{m^2_\mathrm{Z}}{v}, \end{aligned} \tag{1.26}$$

which is proportional to the square of the boson masses. The full Standard Model Lagrangian (neglecting the colour part) can be finally written as

$$\mathcal{L}_\mathrm{SM} = \mathcal{L}_\mathrm{GWS} + \mathcal{L}_\mathrm{Higgs}, \tag{1.27}$$

[5]The numerical factor $\sqrt{g^2+g'^2}$ has been introduced in order to normalize the combinations of gauge fields $gW^3_\mu - g'B_\mu$ and $g'W^3_\mu + gB_\mu$.

where its electroweak part (representing the Glashow-Weinberg-Salam model of electroweak unification) is given by:

$$\begin{aligned}\mathcal{L}_{\mathrm{GWS}} = &-\frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\ &+ i\bar{\nu}_L\gamma^\mu\partial_\mu\nu_L + i\bar{e}_L\gamma^\mu\partial_\mu e_L + i\bar{e}_R\gamma^\mu\partial_\mu e_R + \\ &+ i\bar{f}_L\gamma^\mu\left(-ig\frac{\tau^i}{2}W^i - ig'\frac{Y}{2}B_\mu\right)f_L + i\bar{e}_R\gamma^\mu\left(-ig'\frac{Y}{2}B_\mu\right)e_R\,.\end{aligned} \tag{1.28}$$

Re-expressing the interaction part of the above Lagrangian in terms of the physical fields and writing explicitly the covariant derivative, one obtains:

$$\begin{aligned}\mathcal{L}^{\mathrm{int}}_{\mathrm{GWS}} &= \mathcal{L}^{\mathrm{int}}_{\mathrm{CC}} + \mathcal{L}^{\mathrm{int}}_{\mathrm{NC}} \\ &= \left\{eJ^{\mathrm{em}}_\mu A^\mu + \frac{g}{\cos\theta_{\mathrm{W}}}J^{\mathrm{Z}}_\mu Z^\mu\right\} + \left\{\frac{g}{\sqrt{2}}\left(J^+_\mu W^{+\mu} + J^-_\mu W^{-\mu}\right)\right\}\end{aligned} \tag{1.29}$$

for the neutral and charged part respectively. The electromagnetic coupling constant e has been introduced, identifying $e = g\sin\theta_{\mathrm{W}}$. The following currents have also been defined:

$$\begin{aligned}J^{\mathrm{em}}_\mu &= Q\bar{f}\gamma_\mu f \\ J^{\mathrm{Z}}_\mu &= \frac{1}{2}\bar{f}\gamma_\mu\left(c^f_{\mathrm{V}} - c^f_{\mathrm{A}}\gamma_5\right)f \qquad \text{with} \qquad c^f_{\mathrm{V}} = T_3 - 2Q\sin^2\theta_{\mathrm{W}},\; c^f_{\mathrm{A}} = T_3 \\ J^+_\mu &= \frac{1}{2}\bar{\nu}\gamma_\mu\left(1-\gamma_5\right)e\,.\end{aligned} \tag{1.30}$$

Fermion masses

An attractive feature of the Standard Model is that the same Higgs doublet which generates W and Z masses is also sufficient to give mass to leptons and quarks. For the lepton sector, for instance, the following Lagrangian can be added (for each lepton generation ℓ):

$$\mathcal{L}^{\ell}_{\mathrm{Yukawa}} = -G_\ell\left[(\bar{\ell}_L\Phi)\ell_R + \bar{\ell}_R(\Phi^\dagger\ell_L)\right]\,, \tag{1.31}$$

where the Higgs doublet has exactly the required $SU(2)_L \times U(1)_Y$ quantum numbers to couple to $\bar{\ell}_L\ell_R$. After the spontaneous symmetry breaking, inserting equation (1.21) into the (1.31), one obtains:

$$\begin{aligned}\mathcal{L}^{\ell}_{\mathrm{Yukawa}} &= -\frac{G_\ell}{\sqrt{2}}\left\{\left(\bar{\nu}_\ell,\bar{\ell}\right)_L\begin{pmatrix}0\\ v+H\end{pmatrix}\ell_R + \bar{\ell}_R\,(0, v+H)\begin{pmatrix}\nu_\ell\\ \ell\end{pmatrix}_L\right\} \\ &= -\frac{G_\ell}{\sqrt{2}}\left\{v(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L) + (\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L)H\right\} \\ &= -\frac{G_\ell}{\sqrt{2}}\left\{v\bar{\ell}\ell + \bar{\ell}\ell H\right\}\,.\end{aligned} \tag{1.32}$$

It is now easy to see from equation (1.32) that an appropriate choice of the coupling factor G_ℓ (called "Yukawa coupling") can generate the required lepton mass:

$$m_\ell = \frac{G_\ell \cdot v}{\sqrt{2}}\,. \tag{1.33}$$

The coupling of a fermion to the Higgs boson is therefore proportional to its mass m_ℓ and this property has important consequences for the Higgs production and decay.
The Yukawa Lagrangian for the lepton sector can be rewritten as

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -m_\ell \left\{ \bar{\ell}\ell + \frac{1}{v}\bar{\ell}\ell H \right\} . \tag{1.34}$$

This technique allows to generate a mass term for leptons and down-like quarks. For up-like quarks and neutrinos, a different Higgs doublet has to be introduced, defined as:

$$\Phi_c = i\tau_2 \Phi^* = \begin{pmatrix} \phi^{0\,*} \\ -\phi^{+\,*} \end{pmatrix} . \tag{1.35}$$

1.1.4 Experimental confirmations of the theory

After having been finalized around the mid-1970s in its current formulation, the Standard Model has proven to be an extremely successful theory and its predictions have been confirmed in a number of experiments carried out during the last four decades.
The second and third quark generations have been predicted before their existence could be experimentally estabilished (the most massive quark, the top quark, was discovered only in 1995 [6]). A crucial confirmation of the theory came in 1983 with the discovery of the W and Z bosons at the UA1 and UA2 experiments at the Sp$\bar{\text{p}}$S proton-antiproton collider built by the European Organization for Nuclear Research (CERN) [7–10]. After that, the properties of the Z boson have been extensively measured in the 1990s by the experiments at the Large Electron-Positron Collider (LEP) at CERN and at the Stanford Linear Collider (SLC) at SLAC, in what has been one of the most accurate tests of the Standard Model predictions performed so far [11]. A particle closely resembling the properties of the last missing piece of the puzzle, the elusive Higgs boson, has been finally observed in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN [12, 13]. The combined searches of the two experiments showed evidence for the existence of a boson whose mass has been measured to be $m_{\text{H}} = 125.09 \pm 0.21 \pm 0.11\,\text{GeV}/c^2$ [14]. A detailed investigation of the properties of the newly discovered boson is ongoing, with the results being so far compatible within the uncertainties with what is predicted by the Standard Model [15].
Despite its extraordinary success, the Standard Model still leaves many questions unanswered and both experimental and theoretical considerations point towards the need of finding an extension of the theory. The experimental confirmation or refusal of the different models which try to achieve this ambitious goal will be the most intense research activity in the next decades of particle physics.

1.1.5 A note about unit conventions

From this point on and throughout the following chapters, the International System of Units (SI) will be used for all the common physical quantities, together with a few non-SI units such as the "electronvolt" (eV), as a unit of measure of atomic and particle energies ($1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J}$). The *cross section*, an important physical quantity which can be viewed as a measure of the probability of interaction between subatomic particles, has the dimensions of an area and is expressed in another non-SI unit named "barn" (b), with $1\,\text{b} = 10^{-28}\,\text{m}^2$. The standard prefixes to the unit names and symbols are used to indicate multiples and fractions of the units (e.g., μm, MeV, etc.).

The usual particle physics convention in which the speed of light c and the reduced Plank constant $\hbar$ are set to unity, namely:

$$c = \hbar = 1 \tag{1.36}$$

is adopted in the text. This has the consequence that energies, momenta and masses, if expressed in electronvolts or in multiple of this unit, have the same unit of measure (i.e., MeV, GeV, TeV).

1.2 The Large Hadron Collider

The Large Hadron Collider (LHC) [16] is a proton-proton and heavy-ion collider located in a 27 km long circular tunnel placed about 100 m underground, across the Franco-Swiss border near Geneva, Switzerland. It was built by CERN with the aim of investigating the high energy domain of particle physics through collisions of two opposite beams of hadrons, with an energy in the center of mass of 14 TeV (one order of magnitude higher than the previous generation of particle accelerators) and a nominal *instantaneous luminosity* of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In accelerator physics, the instantaneous luminosity of a collider is a quantity having the dimensions of a flux and expressing the number of particles crossing a unit of area in one second. As a comparison, the proton-antiproton collider called Tevatron built at Fermilab, near Chicago, which can be regarded as the flagship hadron collider before the start of the LHC operations and where the top quark has been discovered, had a center of mass energy of about 2 TeV and reached a maximum instantaneous luminosity $\mathcal{L} = 4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.
The unprecedented energy scale of the LHC allows accurate experimental tests of electroweak symmetry breaking in the Standard Model and provides further understanding of the properties of the already known particles enabling, for example, deeper studies of top quark physics or of the violation of nature's symmetries such as the CP-symmetry[6]. However, the most ambitious goal of the LHC research program is to find hints of some new physics going beyond the Standard Model, verifying or refuting models which predict the existence of supersymmetric particles at the TeV mass scale, shining new light on the nature of dark matter or even discovering some signatures belonging to more exotic new phenomena as, for example, extra dimensions or other experimental signs which would point to a modification of gravity at the TeV scale. Furthermore, the LHC will also provide high energy beams of lead (Pb) ions at energies over 30 times higher than the previous accelerators, allowing physicists to further extend the study of QCD matter under extreme conditions of temperature and density.
In order to achieve such high energies, the proton beams are prepared by the existing CERN accelerator complex (see figure 1.2). A linear accelerator (LINAC2) brings the energy of the protons up to 50 MeV and injects them into the Booster, where they are accelerated up to 1.4 GeV. The energy is then further increased up to 25 GeV in the Proton Synchrotron (PS). The last pre-acceleration step is carried out in the Super Proton Synchrotron (SPS), from which the protons are finally injected into the LHC at an energy of 450 GeV. Here, eight radio-frequency resonant cavities oscillating at 400 MHz accelerate the protons to their final energy with "kicks" of 0.5 MeV per turn. Since the protons are accelerated only in some specific phase intervals of the cavity's electric field, they naturally tend to stabilize themselves in bunch-like structures. Each LHC beam can contain up to 2808 bunches, having approximately 10^{11} protons each. In this design configuration, the time interval between two adjacent bunches is 25 ns, which corresponds to a frequency of 40 MHz. The beams are kept on their path through the use of 1232 dipole magnets and focused with additional 392 quadrupole magnets. The magnets work

[6] A CP-transformation is defined as the simultaneous application of a parity transformation and of a charge-conjugation transformation, where each charge q is replaced with a charge $-q$.

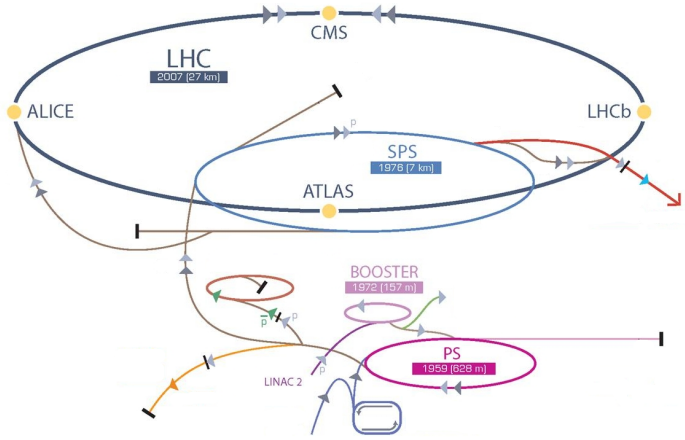


Figure 1.2: Schematic overview of the LHC ring with the injectors chain and the four main experiments. Adapted from [17].

in the superconducting regime and therefore have to be kept at a temperature of 1.9 K by means of super-fluid helium. Since collisions occur between particles of the same charge, two separate beam pipes are required, with two opposite magnetic field configurations.

At the four beam intersection points the main experiments have been installed. Two of them, CMS (Compact Muon Solenoid [18]) and ATLAS (A Toroidal LHC ApparatuS [19]), are general purpose detectors, one (LHC-b [20]) deals with the study of CP-violation and rare decays in flavour physics while the last one, ALICE (A Large Ion Colliding Experiment [21]), is devoted to the investigation of high energy ion physics. Three other experiments are located along the ring, sharing the intersection points with the previous four. The TOTEM (TOTal Elastic and diffractive cross section Measurement [22]) and the LHC-f experiments [23] are positioned on either side of the CMS and ATLAS detectors, respectively. The first is devoted to forward physics, to the measurement of the total proton-proton cross section and to the study of the proton structure, the latter uses particles coming from forward scatterings as a source for simulating cosmic rays in laboratory conditions. Finally, a smaller experiment named MoEDAL (Monopole and Exotics Detector At the LHC [24]), located near the LHC-b experiment, searches for magnetic monopoles and other highly ionizing stable (or pseudo-stable) massive particles.

1.2.1 LHC timeline

At the end of March 2010 the first collisions took place between beams of 3.5 TeV each, at a center of mass energy of 7 TeV (half of the nominal one) and at an instantaneous luminosity around $10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The same energy was kept for all the 2011 run, with the instantaneous luminosity rising up to $4 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at the end of the year. In 2012 the energy of the beams was increased up to 4 TeV each (8 TeV in the center of mass) to profit from the larger Higgs production cross section at this energy and the peak instantaneous luminosity reached $\sim 8 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [25]. These data taking periods (also known as Run I) led to major scientific results, among which the most outstanding was the already mentioned observation of a new boson with a mass around 125 GeV and whose properties are found to be compatible with those predicted by the Brout-Englert-Higgs mechanism. In February 2013 the first long shutdown of the LHC (often referred to as LS1) began, in order to prepare the collider for operations at higher energy and luminosity. During the following two-year period, a number of interventions has been carried out on the machine. The main goal was to consolidate the circuits of the superconducting

magnets, in order to ensure safe operations at higher energy. Additional repair, maintenance and consolidation work has been performed on the LHC as well as on the injector chain and experiments. The accelerator complex has been slowly powered up and commissioned in the spring of 2014, starting from the PS. The LHC started delivering physics data again in June 2015, with the unprecedented center of mass energy of 13 TeV. This point marks the start of the Run II data-taking period, which is planned to continue until the end of 2018, interrupted only by technical stops at the end of each year [26].

1.3 Physics at hadron colliders

A proton is a composite particle made of quarks and gluons, which in the high energy regime are usually referred to with the collective term of *partons*. While at low energy scales (i.e., large length scales) a baryon can be viewed as composed only of three "valence quarks" (up and down in the case of a proton), at high energies (i.e., small length scales) observations show the presence of additional "sea quarks" and gluons, in addition to the valence quarks. The probability density for finding a parton i carrying a fraction xp of the hadron momentum p, with x ranging between 0 and 1, at a certain energy scale Q^2 is given by parton distribution functions, or PDFs, $f_i(x, Q^2)$. High energy collisions between hadrons can therefore be viewed as interactions between the partons which constitute the colliding particles.
In figure 1.3 it can be seen that the total proton-proton cross section at the LHC, for an energy in the center of mass equal to 13 TeV, is around $\mathcal{O}(100\,\text{mb})$. This implies an interaction rate of about 10^8 Hz when running at an instantaneous luminosity of $10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$. In most of the cases, these events consist of long-distance collisions between the two incoming protons: the momentum transfer is small and particle scattering at large angle is suppressed. The particles produced in the final state of such interactions have large longitudinal momenta but small transverse momenta ($\approx 500\,\text{MeV}$) and most of the collision energy escapes down the beam pipe, undetected. These soft scattering events represent by far the majority of pp interactions, but are of no interest for the ATLAS and CMS physics program. Much more interesting are events in which a so-called *hard scattering* process takes place between two partons and the resulting products have high transverse momenta: it is in these kind of events that massive particles are produced and where one hopes to find signs of new physics.
A pictorial representation of a hard scattering process is shown in figure 1.4: H_1 and H_2 are the two incoming hadrons (protons, for example), each one with a momentum p_i. The two partons which interact carry a fraction $x_i p_i$ of the hadron momentum. The parton model formalism allows to write the cross section for a generic hard scattering process as:

$$\sigma_{H_1H_2}(p_1, p_2) = \sum_{i,j \text{ partons}} \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \; f_i^{(H_1)}(x_1, \mu_F^2) f_j^{(H_2)}(x_2, \mu_F^2) \; \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \alpha_S(\mu_R^2), \mu_R^2) \,. \tag{1.37}$$

The short distance partonic cross section $\hat{\sigma}$ can be calculated as a perturbative expansion in the running coupling constant α_S, which depends upon the renormalization scale μ_R. The partonic cross section is process-dependent and is a function of the effective center of mass energy $\hat{s} = x_1 x_2 s$, where s is the center of mass energy of the incoming hadrons. The terms f_i are the above defined PDFs and represent the non-perturbative part of equation (1.37). Due to their universality (i.e., they do not depend upon the particular process considered), the parton distribution functions are usually extracted from deep inelastic scattering experiments, where elementary particles (electrons, for example) are used to probe the inside of hadrons. Finally, the term μ_F is a factorization scale, which can be thought of as the scale that separates the long

and short-distance physics.

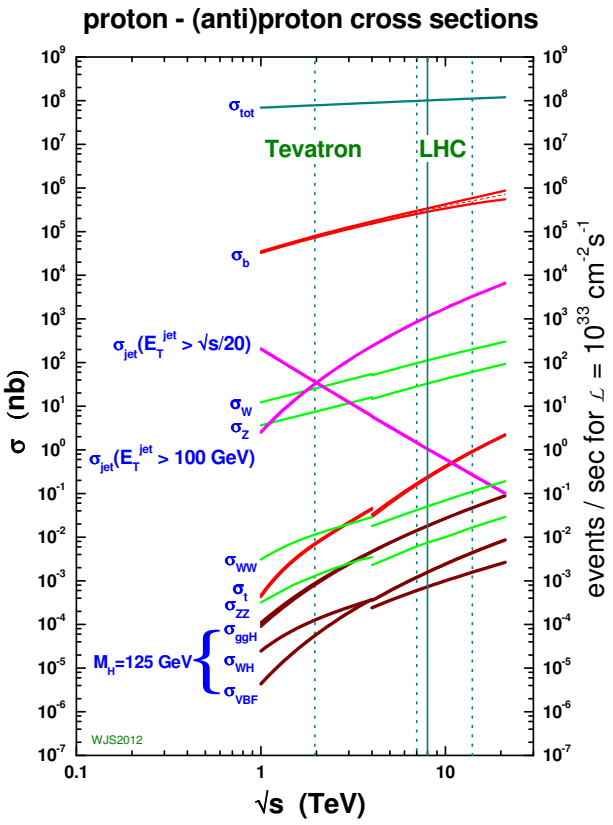


Figure 1.3: Standard Model cross sections at the LHC for some reference processes. Note the logarithmic scale on the vertical axis. Interaction rates are expressed in events per second [27].

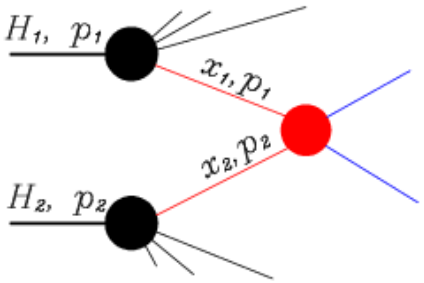


Figure 1.4: A graphical representation of a generic hard scattering process with undefined final states.

1.3.1 Underlying event and pile-up interactions

In a hadronic collision, the main hard interaction sketched in figure 1.4 is often accompanied by radiative effects, in the form of initial or final state radiation. In addition to this, further particles are always produced in the interactions between the other partons of the colliding

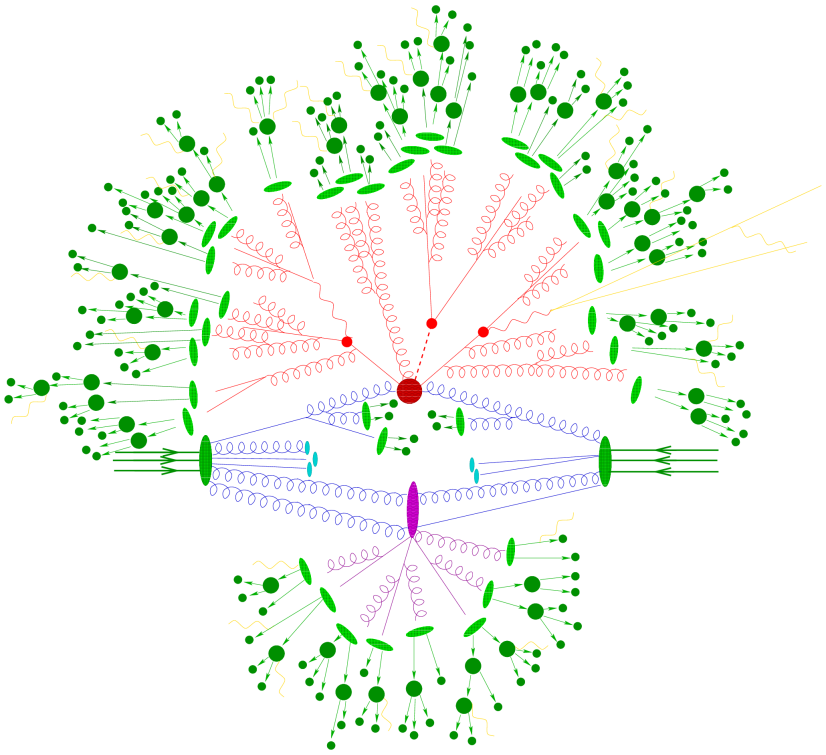


Figure 1.5: A graphical representation of a generic hadron-hadron collision. The red blob in the center represents the main hard interaction, surrounded by the radiation coming from the incoming partons (blue lines) and by the outcoming decay products (red lines). The transition between partons and hadrons is represented by the light green blobs, while the dark green blobs indicate hadron decays. The lower purple blob (with its own initial and final state radiation) indicates a secondary interaction between the proton remnants and is therefore part of the so called underlying event [28].

protons or by the hadronization of the beam remnants. As a result, the observed final state of an hadronic collision can be extremely complicated, as illustrated in the sketch of figure 1.5. Usually, the extra activity of a hadronic collision which can not be uniquely disentangled from the initial and final state radiation of the hard interaction is referred to as *underlying event* (UE).

Moreover, in general more than one proton pair of the colliding beams interact for each bunch crossing. The products of these additional collisions overlap with the ones coming from the main hard interaction and are referred to as *pile-up* (PU). If the additional interactions come from the same bunch crossing, it is common to refer at them as "in-time" pile-up, otherwise they are defined as "out-of-time" pile-up. The most obvious effect of pile-up is the presence, inside a given event, of many different vertices where the particles produced by these interactions originate. These vertices, including the one of the main interaction, are normally called *primary vertices* to distinguish them from the *secondary vertices* defined by the decays of unstable particles having a measurable decay length (τ leptons or hadrons containing a bottom quark, for example).

The expected number of additional pile-up interactions in a given bunch crossing i is expressed by the following equation:

$$\left(N_{\mathrm{PU}}^{\mathrm{true}}\right)_i = \frac{\mathcal{L}_i \cdot \sigma_{\text{min. bias}}}{n_b \cdot f}\,, \tag{1.38}$$

where $\mathcal{L}_i$ is the instantaneous luminosity of that specific bunch crossing and the denominator represents the bunch crossing rate, given as the product between the number of bunches n_b and the revolution frequency f ($f = 11.246\,\mathrm{kHz}$ at the LHC). The term $\sigma_{\text{min. bias}}$, or minimum-bias

cross section, can be seen as an estimate of the inelastic proton-proton cross section.

1.3.2 Luminosity and event rate

The concept of instantaneous luminosity in a collider, which has already been introduced when describing the nominal parameters of the LHC, can be more rigorously expressed by the following equation, where it is assumed that the colliding bunches have identical Gaussian profiles and that the collisions between bunches are "head-on":

$$\mathcal{L} = \frac{N_1 N_2 \cdot f \cdot n_b}{4\pi\sigma_x\sigma_y}\,. \tag{1.39}$$

Here, N_i is the number of particles in each colliding bunch, f is the revolution frequency, n_b is the number of bunches in one beam and σ_x, σ_y are the RMS of the transverse bunch sizes in the horizontal (bend) and vertical directions.
Given a certain physics process having a cross section σ, the expected event rate is then simply given by:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma\,. \tag{1.40}$$

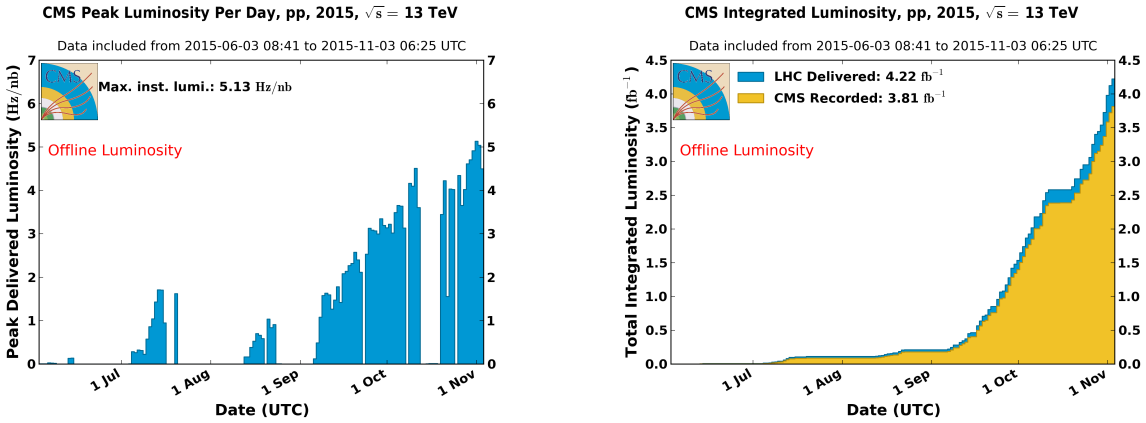
The time integral of the instantaneous luminosity gives what is normally called "integrated luminosity":

$$L = \int \mathcal{L}(t)dt\,. \tag{1.41}$$

Conventionally, the integrated luminosity is expressed in units of "barn", where $1\,\mathrm{b} = 10^{-24}\,\mathrm{cm}^2$, or in submultiples of this unit. Using the integrated luminosity, the number of expected events for the considered process is then:

$$N = L \cdot \sigma\,. \tag{1.42}$$

The small cross section of many interesting physics processes (e.g., the production of Higgs bosons, whose cross section is plotted, as a function of energy, in figure 1.3) makes it necessary to move towards higher luminosities, to be able to collect an higher amount of events in a given period of time. From the experimental point of view, the price to pay is represented by a more complicated final state since, as it is clear from equation (1.38), the number of interactions also grows proportionally to the luminosity.
During the 2015 LHC operations at a center of mass energy of 13 TeV, the peak instantaneous luminosity delivered to the CMS experiment each day has the behaviour represented in the plot of figure 1.6(a) and reached a maximum value of $5.13 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ before the technical stop at the end of the year. Note that in the plot the instantaneous luminosity is expressed in the equivalent units of Hz/nb. The cumulative distribution of the instantaneous luminosity is represented in figure 1.6(b). The total integrated luminosity recorded by the CMS experiment during the 2015 run amounts to $3.81\,\mathrm{fb}^{-1}$.



(a) (b)

Figure 1.6: Left: peak instantaneous luminosity delivered to the CMS experiment each day, during the LHC proton-proton collisions at a center of mass energy of 13 TeV in 2015. Right: cumulative luminosity as a function of the day, delivered to (blue) and recorded by CMS (orange) during stable beams [25].

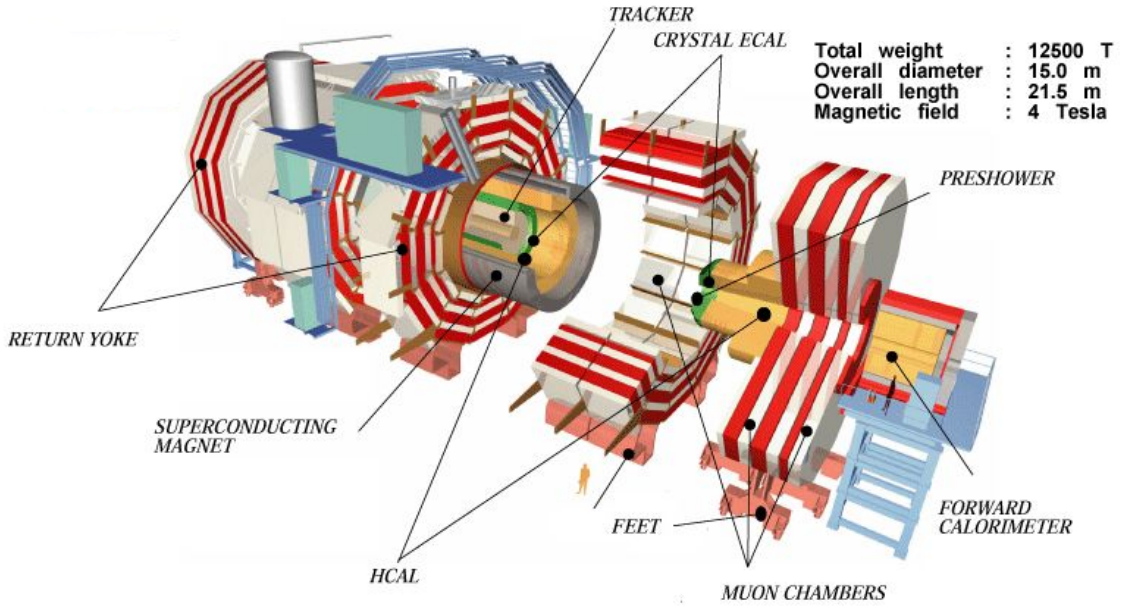
Chapter 2

The CMS detector at the LHC

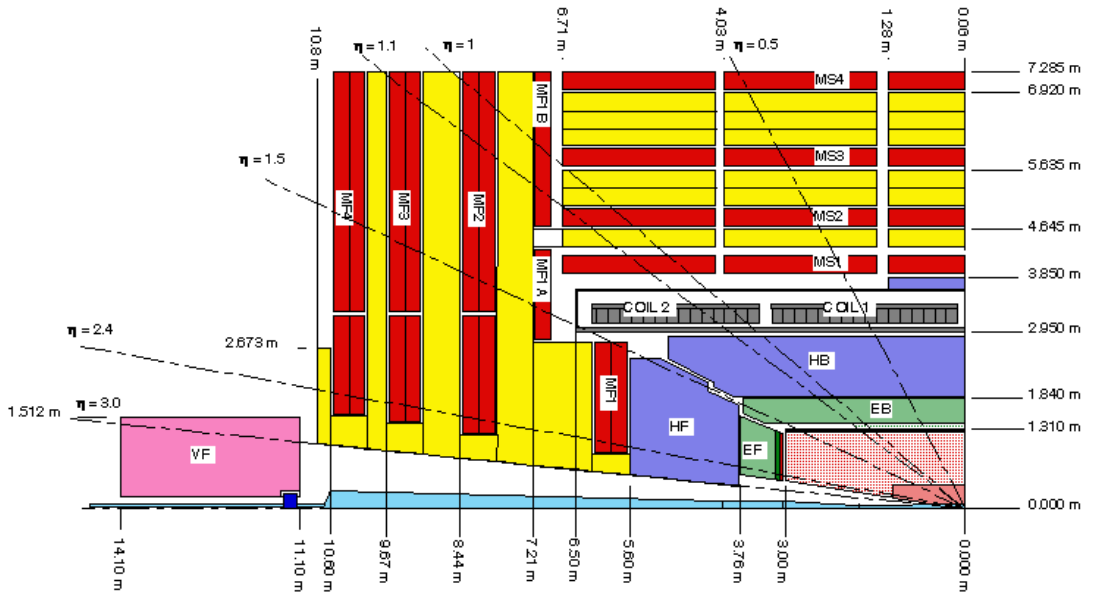
This chapter introduces the CMS experiment at the LHC, defining its physics goals and briefly describing the structure and working principles of its different subdetectors. Since part of the work done in the context of this thesis deals with the upgrade of the CMS pixel detector, a deeper discussion of the CMS pixel system is postponed to chapter 3. The second part of this chapter illustrates the fundamental concepts behind the CMS particle-flow approach and how the detector properties described in the first sections are exploited for reconstructing the main physics objects: electrons, photons, muons, jets and missing transverse energy. A detailed discussion of the τ reconstruction algorithm is left for chapter 5, where the personal contributions to the reconstruction and identification of τ hadronic decays during Run II are presented.

2.1 The CMS detector

The Compact Muon Solenoid (figure 2.1) is a general purpose detector located at the interaction point number 5 along the LHC ring [31]. It has a cylindrical symmetry around the beam pipe, with an overall length of 22 m and a diameter of 15 m and it is made of a sequence of substructures dedicated to the measurement of energy and momentum of all the collision products. Starting from the beam interaction region and moving outwards, the main subdetectors that are contained in CMS are: the tracking system, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the muon chambers. One major design choice was to keep the tracker and both the calorimeters inside the 13 m long and 6 m wide superconducting solenoid [32], hence the name "compact". The muon system is placed outside, embedded in the iron magnetic yoke that sustains the structure and drives back the 3.8 T magnetic field produced by the solenoid.
Every part of the overall detector has been designed in order to meet the goals of the LHC physics program. The innermost tracking system allows the reconstruction of charged-particle momentum and interaction vertices, as well as the offline identification of τ leptons and the decay of hadrons containing a bottom quark. The high granularity of the calorimeters provides good spatial resolution, good resolution on the energy measurement of electrons and photons and the possibility to exploit cluster shape-based methods to reject hadrons misidentified as electrons. Moreover, the magnetic field produced by the solenoid allows, through the curvature of their tracks, an excellent measurement of the momenta of charged particles: the goal during the design phase was to reach 10% momentum resolution for muon transverse momenta of 1 TeV.
In a typical proton-proton collision, the fractions x_a and x_b of the parent proton momenta which



(a) general cutaway view



(b) longitudinal section of one quadrant

Figure 2.1: Views of the Compact Muon Solenoid experiment at the LHC. Adapted from [29, 30].

are carried by the interacting partons are in general different, and the rest frame of the hard collision is therefore boosted along the beam line with respect to the laboratory frame. The reconstruction of the boost of the system would require the full reconstruction of the remnants of the colliding protons, which in practice is not possible, due to the presence of the beam pipe and the instrumentation placed at small angles with respect to the beam line. Because of the unknown energy balance along the beam line, proton collisions are usually studied in a convenient coordinate system which has been estabilished such that the origin is centered at the nominal collision point inside the experiment, the z-direction is parallel to the beam line, the y-direction is vertical and the x-direction is horizontal, pointing towards the center of the ring. The azimuthal angle, ϕ, is measured around the beam line in the x-y plane, starting from the x-axis, while the polar angle θ is measured from the z-axis. The polar angle is usually expressed in terms of the pseudorapidity η, defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.1}$$

The pseudorapidity is in fact the high-energy limit of the rapidity y, which is defined as:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right). \tag{2.2}$$

The advantage of this coordinate frame is the Lorentz invariance of transverse quantities and differences in y, under Lorentz boosts along the beam line. As a consequence, a solid angle in (y,ϕ) space is also invariant under longitudinal boosts. The momentum and energy transverse to the beam direction, conventionally denoted by $p_{\mathrm{T}} = \sqrt{p_x^2+p_y^2}$ and $E_{\mathrm{T}} = \sqrt{E_x^2+E_y^2}$, are computed from the x and y components.
To reach the nominal high luminosities, up to 2808 bunches per beam, containing about 1.1×10^{11} protons each and having a small transverse size of $15\,\mu$m, are collided every 25 ns[1]. This calls for a powerful online event selection process ("trigger") in order to reduce the huge event rate in a way which makes it suitable for storage and subsequent analysis. The CMS trigger system will be briefly described in section 2.1.5.

During the no-beam period called LS1 (between 2013 and 2014), all CMS infrastructures and subdetectors benefited from consolidation, maintenance and upgrade interventions, which were crucial in order to operate the detector at the higher energies and collision rates of 2015. The most important ones will be outlined when describing the corresponding subdetectors in the following sections.

2.1.1 The silicon tracker

The CMS tracker [33] is the innermost CMS subdetector and has been built with a cylindrical geometry surrounding the interaction point, with a total length of 5.8 m and a diameter of 2.5 m. In the outer part it consists of 10 layers of silicon microstrip detectors, completed by three layers of silicon pixel detectors closer to the interaction region. An overview of the system is shown in figure 2.2, where all the tracker subcomponents are clearly visible: the Tracker Outer Barrel (TOB), the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker End Caps (TEC) and the pixel detector. The entire tracker covers the region of pseudorapidity $|\eta| < 2.5$, with a barrel-endcap transition region at $0.9 < |\eta| < 1.4$.

[1] For all the LHC Run I period, up to 2012, pp collisions have been delivered every 50 ns, corresponding to twice the nominal bunch time separation.

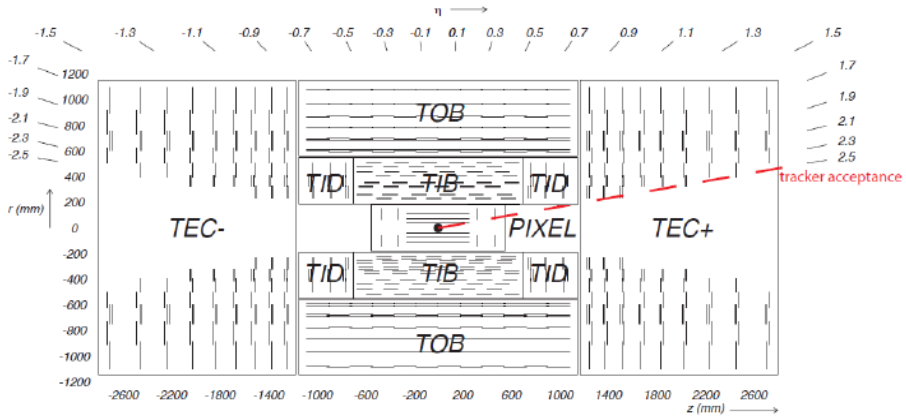


Figure 2.2: The different sub-systems of the CMS silicon tracker. Each line represents a detector module. Adapted from [31].

The choice of silicon as sensitive material was driven by the harsh environment with which the tracker has to cope. At the nominal LHC running conditions, approximately 20 inelastic pile-up interactions will be superimposed on the main hard scattering interaction between the two incoming partons [16]. This means that an average of about 1000 particles will emerge from the interaction region for each bunch crossing. Therefore, a detector technology featuring high granularity and fast response is required, so that trajectories can be identified reliably and attributed to the correct bunch crossing. These features imply however a high density of in-situ electronics, which in turn requires an efficient cooling system. This is in contrast to the general requirement of keeping the amount of material in a tracking system as low as possible, to prevent radiation of photons from deflected charged particles (bremsstrahlung), photon conversion into e^+e^- pairs and multiple scattering in the material itself. In this respect, a compromise had therefore to be found. The intense particle flux will also cause severe radiation damage to the tracking system, so radiation hardness is an additional important requirement.
The pixel detector is the innermost part of the CMS tracking system and is placed close to the interaction region to improve the measurement of the impact parameter[2] of charged-particle tracks and the positions of decay vertices. Given its importance when discussing the work done in the context of this thesis for the future CMS pixel Phase I upgrade, a more exhaustive description of this subdetector will be the topic of section 3.3.
The barrel part of the silicon strip tracker extends up to a radius of 116 cm and is completed at both ends by endcaps made of 3 + 9 disks for each side. It is built according to a modular structure, with a total of 15148 strip detector modules covering an area of approximately 200 m^2. The TIB (four layers) and the TID (three disks at each end) extend up to a radius of 55 cm and deliver up to four r-ϕ measurements using 320 μm thick microstrip sensors, with the strips parallel to the beam axis in the barrel and radial on the disks. The hit resolution, as measured with normally incident cosmic muons, is between 17 μm and 28 μm, depending on the layer [34]. The TIB/TID subsystem is surrounded by the Tracker Outer Barrel, which consists of six layers of 500 μm thick microstrip sensors, providing another six r-ϕ measurements with hit resolution between 23 μm and 40 μm [34]. The TOB extends between $z = \pm$ 118 cm. Beyond this, the Tracker End Cap covers the remaining region, with nine disks for each end, each one made of up to seven rings of silicon microstrip detectors. Finally, the modules in the first two layers and rings, respectively, of TIB, TID, and TOB as well as rings 1, 2 and 5 of the TECs carry a second microstrip detector module, mounted back-to-back with an angle of 100 mrad, in order

[2]The impact parameter of a track is geometrically defined as the distance between the vertex from where the track originates and the linearization of the track itself.

to provide a measurement of the second coordinate (z in the barrel and r on the disks). In this way, the layout ensures at least nine hits inside the acceptance region, with at least four of them being two-dimensional.

During LS1 the main task involving the tracker was to reduce the operation temperature significantly, in order to mitigate the detector damage induced by radiation. Since a lower temperature can lead to condensation and therefore to potential damage for detectors and services, a better dry-gas supply and new humidity barriers have been installed.

2.1.2 The electromagnetic calorimeter - ECAL

The EM calorimeter [35] is made of nearly 83000 lead tungstate ($PbWO_4$) crystals arranged into a barrel structure which covers the pseudorapidity region $|\eta| < 1.479$ and two endcaps on both sides ($1.479 < |\eta| < 3$), as showed in figure 2.3. The lead tungstate has been chosen as scintillating material because of its radiation hardness, high density ($8.28\,\mathrm{g/cm^3}$), short radiation length (0.89 cm) and fast response. The scintillation decay time is comparable with the LHC bunch crossing time: at a nominal rate of 40 MHz there will be an interaction every 25 ns, which is the time required for a crystal to emit about 80% of its scintillation light. However, the disadvantage of this technology is that $PbWO_4$ presents a low light-yield (from 50 to 80 γ/ MeV) which makes it necessary to use intrinsic high-gain photodetectors, capable of operating in a high magnetic field. Avalanche PhotoDiodes (APDs) and Vacuum PhotoTriodes (VPTs) are used: the former glued to the back side of each barrel crystal, the latter to endcap crystals.

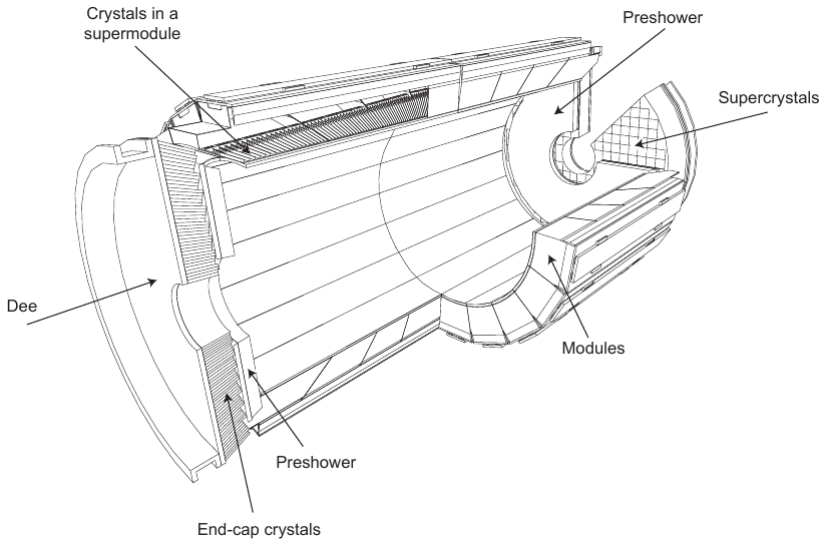


Figure 2.3: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules and supermodules, with the preshower in front of the endcap [31].

The barrel (EB) is made of 61200 crystals, arranged in a cylindrical shape with an inner radius of 1.29 m. Every crystal has a truncated pyramidal shape that, except for a small offset, points towards the interaction vertex, with a depth of 23 cm (corresponding to a depth of 25.8 radiation lengths X_0) and a front surface that covers the range of 1° in the η-ϕ plane ($22 \times 22\,\mathrm{mm^2}$). The EB crystals are grouped in arrays of 5×2 elements, called submodules, assembled in a thin glass-fiber alveolar structure. A module is built arranging from 40 to 50 submodules. Eventually, four modules are assembled with metallic cross plates in between to form the biggest unit of the barrel part, the so-called supermodule. In total there are 36 supermodules, 18 for each

side of the interaction point, each one covering 20° in ϕ.
The two endcaps (EE) consist of identically shaped crystals having a larger front face of 28.6 $\times$ 28.6 mm^2 and a shorter length of 22 cm, corresponding to $\sim 25\,X_0$. They are grouped into carbon-fiber structures of 5 $\times$ 5 elements, called supercrystals. Each endcap is divided into two halves, or Dees, holding 3662 crystals each. The crystals and the supercrystals are arranged in a rectangular x-y grid, with the crystals pointing beyond the interaction point, to avoid that the cracks between crystals are aligned with particle trajectories.
A preshower (ES) is placed in front of EE crystals with the aim of providing a position measurement of the electromagnetic shower with high accuracy and identifying the photons produced in $\pi^0 \to \gamma\gamma$ decays. The preshower is a sampling calorimeter with two layers: thin lead radiators are used to initiate the shower, while silicon strip sensors are placed beyond them to measure the hit position of the shower. The total thickness of the preshower is 20 cm, with a material depth corresponding to about $3\,X_0$, in a fiducial region of $1.653 < |\eta| < 2.6$.
When a photon or electron hits the bulk material of the scintillator, an electromagnetic shower develops inside the calorimeter and every crystal emits an amount of light directly proportional to the energy released by the shower inside it. To reconstruct the overall energy of the shower (and of the particle which has initiated it) clusters of square matrices of crystals are considered, centered around the most energetic one. The total energy is then estimated as the sum of the energies measured by all the elements of the cluster. The energy resolution can be parametrized as a function of the electron/photon energy, expressed in GeV, as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2\,, \tag{2.3}$$

where S is a stochastic term to which, for example, statistical fluctuations in the photoemission contribute; N is a noise term, due to electronics, digitization or pile-up; C is a constant term to which, for example, the non-uniformity in the light collection or the accuracy of the intercalibration constants between different crystals contribute. The target value for the constant term in equation (2.3) is 0.5% in both barrel and endcaps [35]. Measurements performed both with test beam electrons [36] and in-situ [37] show that the mean stochastic term is 2.8% and the mean constant term is 0.3%, therefore reaching and exceeding the design goals.

2.1.3 The hadronic calorimeter - HCAL

A high-energy, coloured quark or gluon emitted in a hard proton-proton collision does not appear in the detector. As it reaches distances from the rest of the proton of a few femtometers, the strong force potential favours the radiation of softer, often collinear, gluons and quarks, until a point is reached where a non-perturbative transition causes the partons to combine into colourless hadrons. The result is a spray of more or less collimated particles, referred to as "jet" which, due to energy conservation, reflects at some level the energy and the flight direction of the initial parton.
The purpose of the CMS Hadronic CALorimeter [38] is indeed to measure the energy and direction of hadronic jets, relying on the hadronic and electromagnetic showers initiated by nuclear interactions. It is a sampling calorimeter divided into three main subdetectors: a Hadron Barrel (HB), a Hadron Endcap (HE) and a Very Forward calorimeter (HF), as sketched in figure 2.4. The hadron calorimeter barrel is radially restricted between the outer extent of ECAL (at a radius of 1.77 m from the beam line) and the inner extent of the magnet coil (at a radius of 2.95 m). This constrains the total amount of material which can be installed to absorb the hadronic shower. The barrel part extends until $|\eta| = 1.3$, the endcap part reaches $|\eta| = 3$ while,

beyond it, forward hadron calorimeters are placed 11.2 m away from the interaction point on both z-ends, extending the pseudorapidity coverage down to $|\eta| = 5.2$.

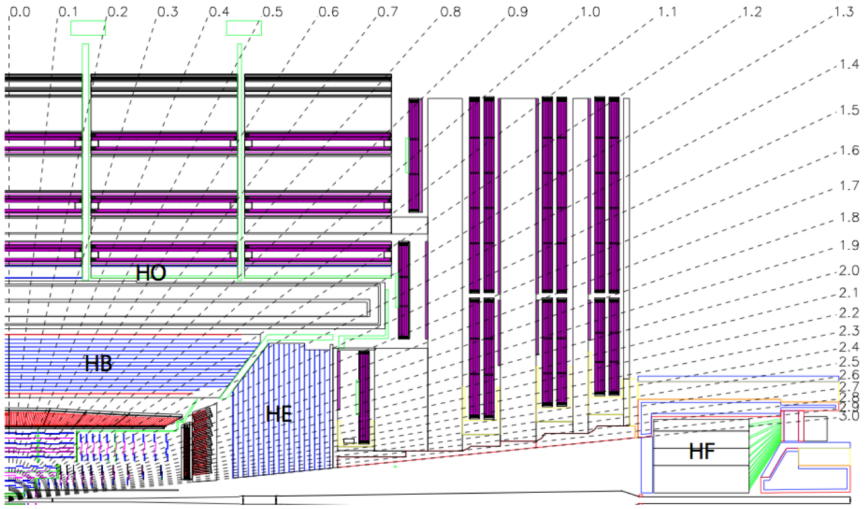


Figure 2.4: Longitudinal view in (r, z) of one quadrant of the CMS detector, showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [31].

The barrel and the endcap parts are made of absorbing layers of brass and plastic scintillators. The scintillation light is converted by wavelength-shifting fibers embedded in the scintillator tiles and sent via clear fibers to photodetectors (hybrid photodiodes, or HPDs) which are able to operate also in high magnetic fields. In the barrel, the total absorber thickness at 90° ($|\eta| = 0$) is 5.8 interaction lengths (λ_I) and it increases with the polar angle θ as $1/\sin\theta$. When the preceeding subdetectors are included, the number of interaction lengths is ~ 8, for centrally produced hadrons. Since approximately 5% of the energy of a 300 GeV pion would be deposited beyond the HB limit, an outer hadron calorimeter (HO) made of one or two layers of scintillators is placed beyond the solenoid and before the first muon layer, to improve shower containment. The total depth of the calorimeter system is thus extended to $\sim 11.8\,\lambda_I$. In the endcap, the total length of the calorimeter, including ECAL crystals, is 10 interaction lengths.
The forward calorimeter (HF) is made of quartz fibers embedded in steel. The choice of the active material was driven by the fact that the forward calorimeter will experience extremely high particle fluxes, especially at the largest rapidities. The signal is generated when charged particles above the Cherenkov threshold (190 keV for electrons) generate Cherenkov light, a small fraction of which is collected and read out with photomultipliers.
The good hermeticity of this system and the great pseudorapidity range covered allow the collection of most of the transverse energy of the event. This is mandatory in order to determine the *missing* transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) with the greatest possible precision. The missing transverse energy is defined as that part of momentum that is missing in the transverse plane of the beam and that has to be added in order to obtain an overall balance of zero p_{T}, consistent with momentum conservation. When possible sources of mismeasurement due to imperfections in the detector are corrected, a non-zero missing energy represents the most common signature of non-detectable particles, like SM neutrinos or neutral long living particles included in many topologies of physics beyond the Standard Model.
Two effects limit the hadron calorimeter resolution: non-containment of the shower and non-compensation. A non-compensating calorimeter has a response ratio $e/h > 1$, where e and h are the calorimeter responses to the electromagnetic and hadronic components of a shower. This causes an intrinsic sensitivity to fluctuations in the electromagnetic component and affects the

resolution. According to test beam results, the combined energy resolution of the CMS barrel calorimeter system (ECAL+HCAL) for pions is:

$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\%\,, \tag{2.4}$$

where the energy is measured in GeV [39]. Additional measurements have been performed in-situ using cosmic ray muons and LHC beams, improving the calibration of the system [40].

During LS1, several photodetectors, including all of those in the HCAL outer calorimeter (HO), have been replaced by better ones. The HO calorimeter switched from HPDs to silicon photomultipliers (SiPMs), which have a larger signal-to-noise ratio and allow the HO to identify muons more efficiently. For the forward calorimeter (HF) photomultiplier tubes (PMTs) have been chosen.

2.1.4 The muon system

The CMS muon system [41] has three main functions: muon identification, momentum measurement and triggering. It is placed outside the magnetic coil, embedded in the iron return yoke of the magnet itself. The return field (whose value goes from ~ 1.8 T in the barrel region to ~ 2.5 T in the endcaps) allows the measurement of momentum and charge via the bending of charged particle tracks. This is particularly important for muons with transverse momenta in the TeV range, for which the complementary tracker measurement degrades. Moreover, the iron yoke serves also as hadron absorber, helping to correctly identify muons. The muon system consists of three types of gaseous detectors, organized in four layers in the barrel part and four layers in the endcap region, with each layer providing track segments reconstruction from few distributed hits. These tracks will be later combined with the information coming from the tracker to form a complete muon track. A sketch of the muon system is represented in figure 2.5.

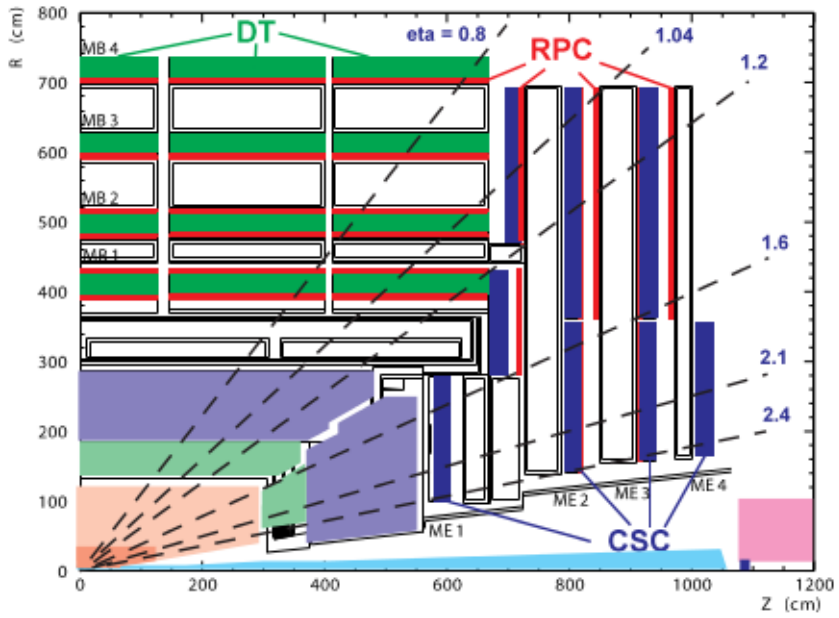


Figure 2.5: View in (r, z) of a quarter of the CMS muon system layout [42].

In the barrel part, where the muon rate is moderately low, the neutron-induced background is small and the return magnetic field is uniform, the chosen detection technology is represented

by Drift Tubes (DTs). These are gaseous detectors consisting of a long aluminium cell filled with gas, with an anode wire in the center that collects ionization charges. The detectors are organized in four layers of "chambers" which follow the cylindrical geometry of CMS, and are arranged in such a way that a muon traverses at least three of them. The chambers are segmented along the z-axis by the five wheels of the yoke and divided in twelve sectors on the ϕ plane, for a global coverage from $|\eta| = 0$ to $|\eta| \sim 1.3$. Each chamber is made of twelve layers of Drift Tubes grouped in three sub-units called superlayers. Two of the superlayers have anode wires parallel to the beam line, providing a measurement of the r and ϕ coordinates; the third one is placed perpendicularly between the others and provides the z measurement. Each station is designed to measure muon positions with 100 μm resolution in the (r-ϕ) plane and 150 μm in the (r-z) plane [41].
In the endcaps, where the muon rates and the background levels are higher and the magnetic field is larger and non-uniform, the muon system is equipped with Cathode Strip Chambers (CSCs), covering the pseudorapidity interval from $|\eta| = 0.9$ to $|\eta| = 2.4$. These are multi-wire proportional chambers, shaped in a trapezoidal form which follows the endcap geometry. Inside the chamber each cathode plane is segmented into strips running across wires. When a muon crosses the chamber, the avalanche developed on a wire induces a charge on several strips of the cathode plane. In every CSC, two coordinates per plane are therefore made available by the simultaneous and independent detection of the signal induced by the same particle on the wires and on the strips.
Resistive Plate Chambers (RPCs) are coupled to the previously described detectors both in the barrel and in the endcap. They are gaseous parallel-plate detectors made of bakelite and operating in avalanche mode. Their space resolution is very poor, due to the large width of the strips in which each plane is segmented (from 2.3 cm to 4.1 cm). They have however an exceptional time resolution (~ 3 ns), comparable to that of scintillators, and are therefore used mainly for triggering purposes and for an unambiguous identification of the bunch crossing.
The performance of the muon system has been measured in-situ and has been found to meet the design goals: the spatial resolution per chamber is 80-10 μm in the DTs, 40-150 μm in the CSCs and 0.8-1.2 cm in the RPCs; the achieved time resolution per chamber is 3 ns or better for all the three subsystems; the efficiency for reconstructing hits and track segments originating from muons traversing the muon chambers is in the range 95-98% [43].

During LS1 some major interventions have been carried out on the muon system. A fourth disk of RPCs and a set of 72 new Cathode Strip Chambers have been installed, completing the design of the endcap region as it was foreseen in the CMS technical design report. The readout granularity of the CSCs in the first endcap layer has also been improved, to increase the muon identification performance in the higher-rate environment of 2015.

2.1.5 The CMS trigger

The total proton-proton cross section at an energy in the center of mass of 14 TeV is expected to be approximately 100 mb. At the nominal LHC luminosity, this means that the expected event rate is about 10^9 inelastic events/s. Storing all these collision events is definitely impractical and, even if it were possible to do so, the dominating fraction of them is made of soft pp interactions, which are not interesting for the CMS physics program. Therefore, a trigger system has been developed with the purpose of providing a large rate reduction factor, whilst maintaining a high efficiency on potentially useful events [44, 45].
The total output rate is reduced by about seven order of magnitudes to $\mathcal{O}(100)$ Hz thanks to a two-level system: a Level-1 (L1) Trigger, which consists of custom-designed, largely pro-

grammable electronics, and a High Level Trigger (HLT), which is a software system running on a farm of commercial processors, including over 13000 CPU cores.
The Level-1 Trigger is hardware-based and has been designed to analyze each 25 ns bunch crossing on a coarse-grain scale, within a latency of no more than 3.2 μs. To deal with these short timescales, it employs calorimetric and muon informations only, since the tracker algorithms are too slow for this purpose. Therefore, the L1 trigger is organized into a calorimeter and a muon trigger, whose information is transferred to the global trigger which takes the accept-reject decision. The calorimeter trigger is based on trigger towers (arrays of 5 crystals in ECAL which match the granularity of the HCAL towers) grouped in calorimetric regions each one consisting of 4×4 towers. The calorimeter trigger identifies the best four candidates for each of the following classes: electrons and photons, central jets, forward jets and τ-jets. The information is then passed to the global trigger, together with the measured $E_{\mathrm{T}}^{\mathrm{miss}}$. The muon trigger is performed separately for each muon detector. The information is then merged and the best four muon candidates are transferred to the global trigger. At the end, the rate of selected events is reduced down to $\mathcal{O}(100)$ kHz.
The High Level Trigger completes the reduction of the output rate down to $\mathcal{O}(100)$ Hz. The idea of the HLT software is the regional reconstruction on demand, which means that only those objects in the useful regions are reconstructed and the uninteresting events are rejected as soon as possible. To the first level, which employs only the information coming from the muon system and the calorimeters, a second level is added, where the information coming from the tracker pixels is made available, followed by a third one, which exploits the full event information. Flexibility is maximized since there is complete freedom in the selection of the data to access, as well as in the sophistication of the algorithms, usually referred to as "HLT paths".

2.2 Object reconstruction at CMS

In this section, a description of the the most important high-level physics objects used in CMS analyses is presented. Due to their particular importance in the context of this thesis, a description of the τ leptons reconstruction is postponed to chapter 5.

2.2.1 The particle-flow approach

The particle-flow (PF) approach is a whole-event reconstruction technique whose purpose is the reconstruction and identification of all the high-level objects produced in each pp collision (namely, charged hadrons, photons, neutral hadrons, muons and electrons) with an optimized combination of all the informations coming from the different CMS subdetectors [46–48]. The resulting list can then be used to construct a variety of higher level objects and observables, such as jets, missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$), τ leptons, etc.
While no substantial changes are expected for the reconstruction of high-energy electrons and muons compared to the "standalone" reconstruction presented in section 2.2.2 and 2.2.3, the particle-flow approach allows to significantly improve the resolution of jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ with respect to a standard, purely calorimetry-based jet reconstruction. Since only about 35% of a jet's energy is carried by photons or neutral long-living hadrons (neutrons, Λ baryons, etc.), for the remaining 65% carried by charged particles the coarse HCAL information can be combined with the more precise tracker momentum measurements, thus allowing for a largely better jet reconstruction.
Charged hadrons are identified as charged particle tracks neither identified as electrons nor as muons, while neutral hadrons are identified as HCAL energy clusters not linked to any charged

hadron trajectory or, if they overlap with charged particles in the calorimeters, as ECAL and HCAL energy excesses with respect to the expected charged hadron energy deposit.
The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energy, corrected for zero-suppression effects and calibrated for the nonlinear response of the calorimeters. The energy of neutral hadrons is obtained from the corresponding calibrated ECAL and HCAL energy.

2.2.2 Electron and photon reconstruction

For each electron reaching the ECAL surface, an electromagnetic shower develops within the ECAL crystals. Most of the electron energy is, in general, collected within a small matrix of crystals centered around the one which has been hit, but the situation could be much more complicated: interactions with the tracker material cause electrons to lose part of their energy radiating photons (bremsstrahlung). As the electron energy degrades, the effect of the magnetic field is to enhance the bending of the electron trajectories, which ultimately results in a spread of emitted photons along the azimuthal ϕ coordinate. Therefore, to obtain an accurate measurement of the electron energy in correspondence of the hard scattering vertex, it is essential to collect as many of the bremsstrahlung photons as possible.
This is the purpose of the first stage in the electron reconstruction sequence, which goes under the name of superclustering [49]. The starting point is a "seed", defined as the crystal containing most of the energy deposited in any considered region and having at least 1 GeV (0.18 GeV) of deposited transverse energy in the barrel (endcap). In the barrel, the algorithm looks for 5×1 arrays of crystals in the η-ϕ plane, each with a total energy of at least 100 MeV. The arrays are aligned with the seed crystal along η and extend up to ± 17 crystals away from the seed one, in ϕ. Contiguous arrays are grouped together into clusters and later collected into groups of clusters, called superclusters. In the endcap, the energy is collected in clusters of 5×5 crystals around the seed one and then grouped together into superclusters.
Once a supercluster is found, the reconstruction proceeds with the track-building stage. One of the possibilities consists in back-propagating the energy and position of the supercluster in the magnetic field, under both $+|e|$ and $-|e|$ charge hypotheses (with $|e|$ the elementary charge), down to the nominal vertex and look for compatible hits in the tracker detector. Once a pair of compatible hits is found, an electron pre-track seed is built. Starting from seeds, compatible hits are searched for on the next available silicon layers, using the combinatorial Kalman filter algorithm [50]. In this pattern-recognition problem, the probability of energy losses due to material effects is taken into account. This procedure is iterated until the last tracker layer, unless no hit is found in two subsequent layers. A minimum of five hits is finally required to create a track.
In the final stage, the supercluster and track informations are merged. The energy measurement E_{sc} provided by the electromagnetic calorimeter can be combined with the tracker momentum measurement p_{tk} to improve the estimate of the electron momentum at the interaction vertex for low energy particles. The improvement is expected to come both from the opposite behaviour of the intrinsic calorimetry and tracking resolutions as a function of E or p (see figure 2.6) and from the fact that p_{tk} and E_{sc} are differently affected by the bremsstrahlung radiation.

Photons (e.g., the ones coming from π^0 decays or from electron bremsstrahlung) are identified as ECAL energy clusters not linked with the extrapolation to ECAL of any charged particle trajectory. Their energy is directly obtained from the ECAL measurement and corrected for the variation of shower containment in the clustered crystals and for the shower losses of photons that convert before reaching the calorimeter [51].

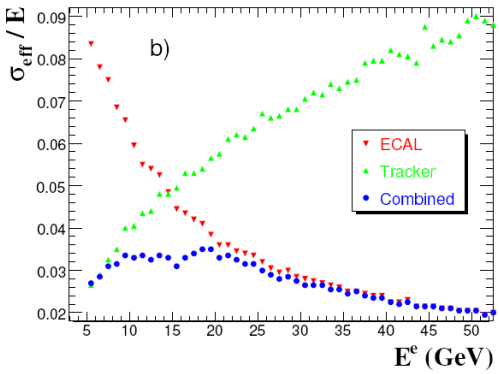


Figure 2.6: Fractional resolution (effective RMS) as a function of the generated energy E^e as measured with the ECAL supercluster (downward arrows), the electron track (upward arrows) and the combined track-supercluster (circles) [42].

2.2.3 Muon reconstruction

The standard muon reconstruction sequence is performed in three stages: a local reconstruction inside every muon subdetector, a standalone reconstruction in the muon system and a global reconstruction using informations coming from the whole detector [52].
In the first stage, local pattern recognition algorithms start from single hits and build track stubs separately in each subdetector (CSCs in the endcap and DTs in the barrel): the result is a three-dimensional segment associated with a single muon layer. In the second stage, track parameters are propagated from one to the next layer of the detectors, working from inside out and taking into account material effects (ionization and bremsstrahlung). A suitable χ^2 cut is applied on the result of the fit between different stubs in order to reject bad combinations and the procedure is iterated until the outermost surface of the muon system is reached. Inclusion of RPC measurements helps in the reconstruction of low p_T muons and of those which escaped through the spaces between the chambers, leaving only one fired DT station. The final track is then extrapolated to the point of closest approach to the beamline. Due to the large amount of material traversed to reach the muon spectrometers, the momentum resolution as measured in the muon chambers is degraded by multiple scattering. In the last stage (global reconstruction), muon trajectories are extended until the outermost layer of the tracker system (silicon strips and pixels) and an η-ϕ region of interest is defined. The compatibility between the muon track and the track parameters of the reconstructed silicon trajectories is checked on a χ^2 basis and, if the result is found in agreement, a global fit is performed with all the hits in the tracker and muon systems.
Figure 2.7 shows how the additional information provided by the muon tracking system is valuable for the momentum reconstruction of high-energy muons ($p \gtrsim 200\,\text{GeV}$), for which the tracker-only momentum measurement degrades. For low and medium-p_T muons, instead, the resolution of the tracking system is the dominant one.

2.2.4 Jet reconstruction

From the experimental point of view, hadronic jets are clusters of tracks and energy deposits, localized in a defined region of the detector, which are merged together following the prescriptions of a jet clustering algorithm.
Given the fact that the cross section for producing an extra gluon in the final or initial state is divergent in the soft ($p_T^g \to 0$) and collinear limit, the algorithms for jet clustering must satisfy a few requirements, which are needed in order to provide finite theoretical predictions. The two conditions to be respected are the following:

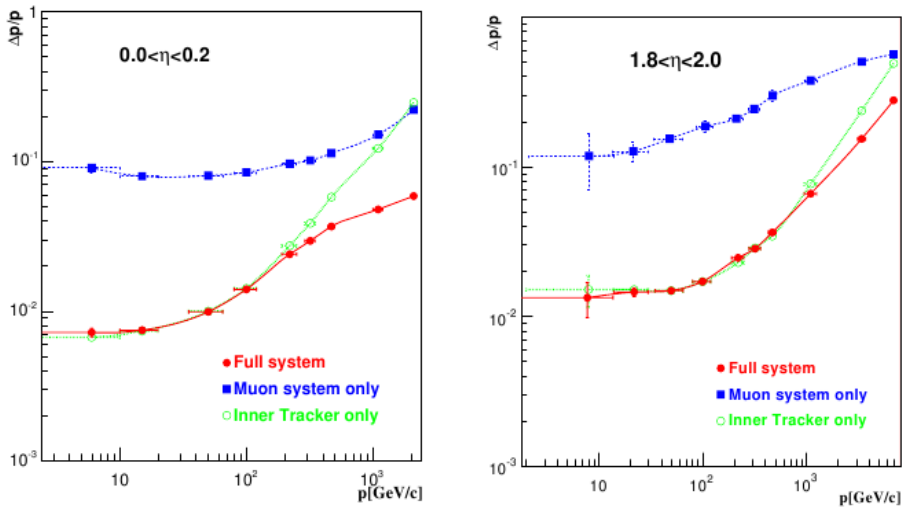


Figure 2.7: The muon momentum resolution versus p using the muon system only, the inner tracker only, or both ("full system"). Left: central barrel region ($|\eta| < 0.2$); right: endcap region ($1.8 < |\eta| < 2$) [42].

- *collinear safety*: the outcome of the jet algorithm must not change if a particle of momentum p is substituted by two collinear particles of momentum $p/2$;
- *infrared safety*: the outcome of the jet algorithm must not change if an infinitely soft particle is added (or subtracted) to the list of particles to be clustered.

In CMS, the adopted clustering algorithm, which respects the two criteria described above, is the so-called anti-k_{T} algorithm [53]. This algorithm proceeds via the definition of two distances for each particle i in the list of particles, namely:

$$\begin{aligned} d_{ij} &= \min\left(\frac{1}{p_{\mathrm{T}i}^2}, \frac{1}{p_{\mathrm{T}j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}\,, \\ d_{iB} &= \frac{1}{p_{\mathrm{T}i}^2}\,. \end{aligned} \tag{2.5}$$

The quantity d_{ij} can be interpreted as the "distance" between the particle i and a generic other particle j among those still to be clustered, while d_{iB} represents the "distance" between the particle i and the beam line. The quantity ΔR_{ij} is the distance between the two particles in the η-ϕ plane and is computed as:

$$\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}\,. \tag{2.6}$$

The constant R, or radius, is a parameter of the algorithm.
For each particle i, the algorithm looks if there is another particle j such that d_{ij} is smaller than d_{iB}. If this happens, then particles i and j are merged by adding their four-momenta, otherwise the particle i is promoted to a jet. The whole procedure is iterated until there are no more particles to merge.
A key feature of this clustering algorithm is that, due to the inverse dependence on the momentum, soft particles tend to cluster with hard ones before they cluster among themselves. If a hard particle has no hard neighbours, then it will simply accumulate all the soft particles within a circle of radius R, resulting in a perfectly conical jet.
The most common radius parameters adopted in CMS (and therefore the approximate jet sizes

in the η-ϕ plane) are either $R = 0.4$ or $R = 0.5$. The momentum associated to the jet is computed as the vectorial sum of the momenta of all the particles contained in it.
The PF approach to the event reconstruction allows for a natural definition of jet objects. Once well reconstructed leptons are excluded from the particle list, all the remaininig objects can be clustered into jets. The natural consequence of this approach is that jets and leptons are conceptually disentangled, since the same energy deposits or tracker hits cannot belong to different physics objects.
Figure 2.8 shows the typical composition of the jet energy in terms of different particle types. In the central part of the detector, where the tracker allows for charge measurements, the largest fraction of an event energy is carried by charged hadrons ($\sim 65\%$). Only about 2% is carried by electrons, while neutral hadrons ($\sim 10\%$) and photons ($\sim 25\%$) share the remaining part. Outside the tracker acceptance, instead, no distinction can be made between charged and neutral particles. Here, the vast majority of the jet energy is carried by hadronic candidates, with purely electromagnetic objects contributing 10% or less.

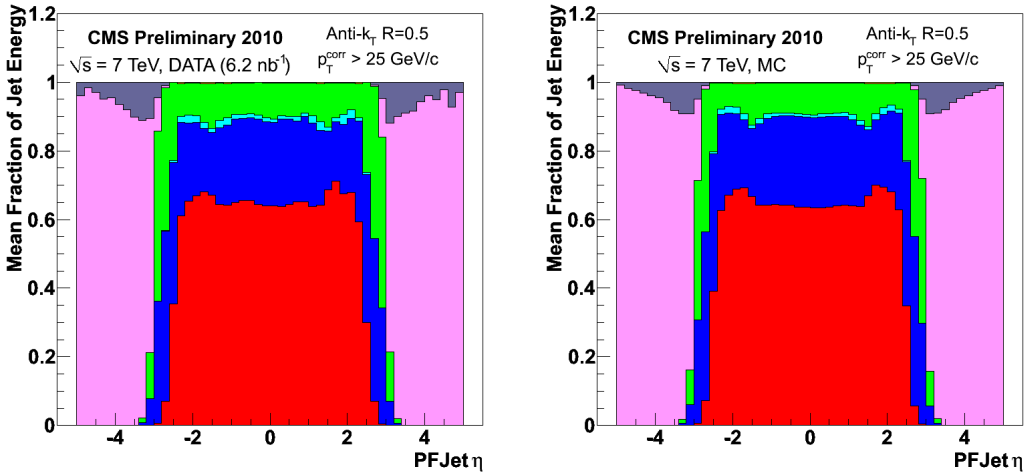


Figure 2.8: Reconstructed jet energy fractions as a function of pseudorapidity in data (left) and in simulation (right). From bottom to top, in the central region: charged hadrons, photons, electrons, and neutral hadrons. In the forward regions: hadronic deposits, electromagnetic deposits [47].

Jet energy corrections

A set of jet energy corrections has to be applied on reconstructed jets so that they can be used as high-level physics objects. These corrections relate the measured energy of the jet to the experimentally inaccessible true energy of the final state particle jet (jet energy scale).
The jet correction scheme adopted in CMS is factorized into subsequent steps, each of them addressing a different physics aspect. The output of each step is a multiplicative correction factor $\mathcal{C}$, to be applied on each component of the raw jet four-momentum [54].

- **Offset corrections**: the purpose of this first step is to remove from the jet the additional energy coming from spurious particles produced in secondary pp interactions within the same bunch crossing or from the underlying events that randomly overlap with the jet area. This correction is determined both in data and in simulation on an event-by-event basis. The charged component of a jet within the tracker acceptance can be removed from the jet,

calculating the impact parameter of all jet particles: those which are not compatible with the hard scattering vertex of the event are not considered in the jet clustering algorithm. To further remove the contribution of neutral particles, or to correct jets with $|\eta| > 2.4$, different techniques can be used used. In one of these, referred to as "jet area method", all particle-flow candidates are re-clustered using a different algorithm (k_{T} instead of anti-k_{T}) and after adding a large number of very soft "ghost" particles uniformly to the event. The median energy density[3] ($\rho_{\mathrm{PU}} = E/\Delta\eta/\Delta\phi$) of the many pseudo-jets so produced is taken as the estimate of the pile-up plus underlying event energy density for that event, and is subtracted from real jets, after being multiplied for the jet area (roughly πR^2) [55].

- **Monte-Carlo based corrections**: after pile-up mitigation, all the other effects on the jet energy scale are simultaneously corrected with simulation-based correction factors. Simulated multi-jet events are created using so-called Monte-Carlo generators, processed through a detailed simulation of the CMS detector and passed to the same jet reconstruction algorithms used on real data. A "jet response" variable $\mathcal{R}$ is then defined as the ratio between the transverse momentum of the reconstructed jet and the one of the generated jet:

$$\mathcal{R} = \frac{p_{\mathrm{T}}^{\mathrm{j,\,reco}}}{p_{\mathrm{T}}^{\mathrm{j,\,gen}}} \tag{2.7}$$

  The jet response might differ from unity because of reconstruction efficiencies, detector effects or out-of-cone effects (i.e., particles belonging to the hadronic shower but not contained in the clustering cone). A correction factor is evaluated as the inverse of the jet response, as a function of the jet's p_{T} and η. Additional data-driven corrections are applied on data, to account for differences with respect to the simulation.

- **Relative corrections**: these corrections are meant to correct for non-uniformities in the different CMS subdetectors by equalizing the jet response along η to the center of the barrel. Di-jet events with one jet in the barrel and one in the forward region are used for this purpose.

- **Absolute corrections**: this last correction factor correctly sets the jet's absolute energy scale. It is derived from γ + jet or $\mathrm{Z}(\rightarrow \mathrm{ee}/\mu\mu)$ + jet events, where the reference photon or the products of the Z boson decay can be measured with better accuracy in the electromagnetic calorimeter or in the muon chambers.

Jet tagging

The properties of a jet can be exploited in order to estimate (or "tag") the origin of the jet. From the practical point of view, "tagging algorithms" translate the jet properties into a set of real numbers, or *discriminators*, associated to each jet and expressing the probability of the hadronic shower to have been initiated by a gluon or by a quark of a certain flavour.
An important group of these tagging algorithms, which goes under the name of "b-tagging algorithms", consists of a set of methods which aim at identifying the jets originating from the decays of hadrons containing bottom quarks [56]. The resulting discriminators have higher values for jets having a higher probability of being initiated by bottom quarks (b-jets) and lower values for jets which are more likely to be initiated by lighter quarks or gluons. The b-tagging algorithms rely on specific features of hadrons containing bottom quarks, which reflect themselves on the properties of b-jets, namely:

[3] The median is used since it is only little affected by the presence of few real high-energy jets in the event.

- **long lifetime**: hadrons containing bottom quarks have a mean lifetime of $\tau = 1.5 \times 10^{-12}\,\mathrm{s}$ and a mean decay length of approximately $\gamma c\tau \sim \gamma \cdot 450\,\mu\mathrm{m}$. As a consequence, the tracks associated with b-jets present an higher value of the impact parameter (IP), which is defined as the distance between the vertex of the hard scattering interaction and the linearization of the track, performed from the point of closest approach to the jet axis. Often, in IP-based tagging algorithms, the significance of the impact parameter $\mathrm{S_{IP}} = \mathrm{IP}/\sigma_{\mathrm{IP}}$ is used, where σ_{IP} is the uncertainty associated with the IP measurement;
- a **displaced secondary vertex** with respect to the vertex of the hard interaction, where the secondary vertex is defined as the decay vertex of the b-hadron. This property follows directly from the longer mean decay length;
- the presence of **semileptonic decays** of b-hadrons, featuring the presence of leptons ($\mathrm{e}^{\pm}/\mu^{\pm}$) having high transverse momenta p_{T} with respect to the jet axis.

Among the b-tagging algorithms which are available in CMS, the Combined Secondary Vertex (CSV) is one of the most used. This sophisticated algorithm uses a large set of variables which are known to distinguish b-jets from jets initiated by gluons or lighter quarks. The impact parameter significance, jet kinematics and secondary vertex information are combined using multivariate techniques and a one-dimensional discriminator is extracted.

2.2.5 Missing transverse energy

In general, a vector $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ can be defined as the negative of the vector sum of the transverse momenta of all final state particles in the event. Since the initial pp collision occurs between two partons of negligible transverse momentum ($\lesssim 1\,\mathrm{GeV}$), under the hypothesis that all the stable or long-lived detectable final state particles are properly reconstructed $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ coincides with the sum of the four-momenta of all undetectable particles (i.e., neutrinos or particles such as neutralinos in more exotic scenarios beyond the Standard Model). In practice, this is not true because the detector is not perfectly hermetic and a fraction of the total event energy is unavoidably lost in the beam pipe or only coarsely reconstructed in the forward calorimeters. As a result, the measured $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is only an approximation of the transverse momenta of the undetected particles. The magnitude of the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ vector is denoted as $E_{\mathrm{T}}^{\mathrm{miss}}$ and referred to as missing transverse energy.
The missing transverse energy is an important quantity in both Standard Model measurements involving W bosons, top quarks or τ leptons, as well as in searches for new physics, especially in supersymmetry, extra dimensions and collider-based dark matter searches. Its reconstruction is particularly sensitive to detector effects and to the precise calibration of all the other physics objects (e, μ, τ, photons and jets) as well as to additional pile-up interactions.
Several algorithms have been developed in CMS in order to reconstruct $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ [57]. Since all detector information is included in the PF-based reconstruction, a simple and straightforward definition of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is the negative of the vector sum of the transverse momenta of all the particle-flow candidates of the event (PF $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$). Jet factorized energy corrections are propagated to the missing transverse energy computation in order to improve its resolution. Another less common definition, Calo $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, can be calculated from the calorimetric deposits and their directions relative to the center of the detector, after correcting the result for the muon deposits in the calorimeters.
More recent algorithms, which are specifically designed to mitigate the effects of large numbers of additional pile-up interactions on the missing transverse momentum resolution, are based on

a set of multivariate regressions (MVA $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$). The MVA is trained on $\mathrm{Z} \to \mu\mu$ events without genuine sources of missing transverse energy and takes as inputs the recoil of the Z boson, the $\sum E_{\mathrm{T}}$ of five different $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ definitions, as well as additional jet and vertex-related variables. A measurement of the performance of these novel algorithms shows indeed a sizeable reduction in the dependence of the missing transverse energy resolution on the number of pile-up interactions, consistent both in data and in simulated events [58, 59].
The MVA $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ definition played an important role in past analyses with neutrinos in the final state, such as $H \to \tau\tau$ searches and is also the reconstruction algorithm which has been used for the measurements presented in this thesis.

Chapter 3

The CMS pixel detector and the pixel Phase I Upgrade

In this chapter, a description of the fundamentals of semiconductor detectors and their application in particle physics is given. In order to facilitate the understanding of later chapters and sections, the reader is also introduced to the basic concepts of interconnection and packaging technologies between sensor substrates and readout electronics, with bump bonding being the example relevant for this thesis. From this, the discussion moves towards a very specific representative of semiconductor detectors: the pixel detector of the CMS experiment at the LHC. The current system is described in all its main components, with particular emphasis on the barrel part (BPIX). Finally, the changes to the current detector which are planned to be implemented during the forthcoming 2016/2017 extended year-end technical stop are presented.

3.1 Working principles of semiconductor detectors

Semiconductor detectors are an attractive option for applications where high granularity and good spatial and/or energy resolution are mandatory, such as particle physics, medical diagnostics or photon science. Among the semiconductor materials which can be used as detection medium, silicon has a predominant role and its usage can profit from decades of technological developments, driven by the microelectronics industry. In the following, the basics of semiconductor physics are explained, showing how these materials can be used for the detection of particle interactions with matter.

3.1.1 Interaction of radiation and particles with matter

In an extreme simplification, practically all particle detectors consist of some kind of active medium, where particles or radiation interact, and a collection mechanism which converts the products of the interaction in a measurable quantity (usually, an electrical signal). Different detectors and detection techniques can be employed, depending on the experimental conditions and the purpose of the measurement. Semiconductor detectors, in particular, are sensitive to the electromagnetic interactions of charged particles and photons.

Charged particles

Heavy charged particles (i.e., particles much heavier than electrons) interact with matter mainly via ionization and excitation of the atoms of the material, losing their energy in a continuous way. The mean energy loss per distance travelled in the medium depends on the velocity $v = \beta c$ of the particle, on its electric charge z and on the properties of the material and is described by the Bethe formula[1]:

$$-\frac{dE}{dx} = 4\pi r_e^2 m_e c^2 \cdot \frac{z^2 N_A Z \rho}{\beta^2 A} \cdot \left[\ln\left(\frac{2 m_e c^2 \beta^2}{I \cdot (1-\beta^2)} \right) - \beta^2 + \frac{\delta(\beta\gamma)}{2} \right], \tag{3.1}$$

where r_e is the classical electron radius, m_e is the electron mass, N_A is the Avogadro number and Z, A and ρ are, respectively, the atomic number, the atomic weight and the density of the material. The parameter I represents the mean excitation energy of the atomic electrons and is usually regarded as an empirical constant. The term $\delta(\beta\gamma)$ is a density-effect correction factor which becomes important at high energies. As can be seen from figure 3.1, the above equation is valid over a broad range of energies but fails at very low energies (around the energy of the atomic electrons) where additional corrections become necessary, as well as at high energies ($\beta\gamma \approx 500$), where radiative effects start to dominate. At $\beta\gamma \approx 2-3$ the energy loss has a minimum which is with good approximation independent of the interacting particle and the traversed material. Particles having an energy in this range are usually referred to as minimum ionizing particles.

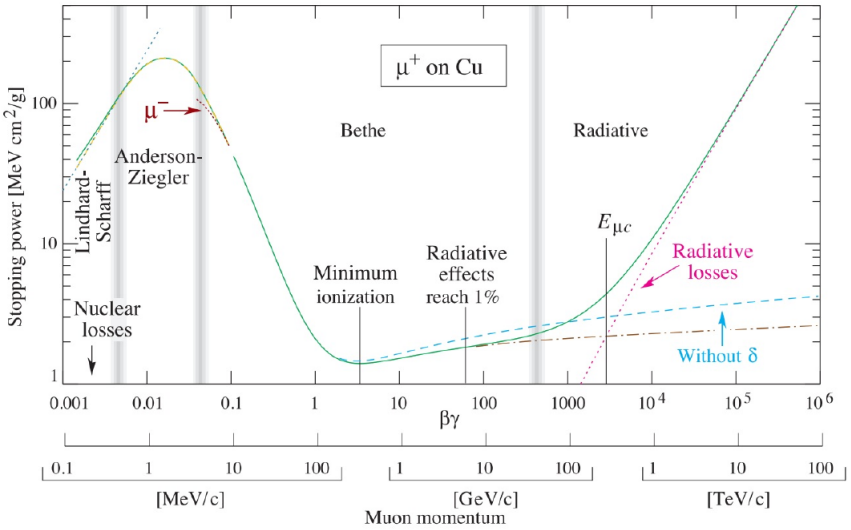


Figure 3.1: Mean energy loss for antimuons interacting with copper, as a function of the energy. The $\beta\gamma$ range between approximately 0.1 and 500 is described by the Bethe formula [1].

For lighter charged particles, such as electrons, the Bethe equation is not valid. For electrons having an energy above $\approx 10\,\text{MeV}$ the dominating energy loss mechanism is bremsstrahlung: the emission of radiation due to the acceleration of charged particles. In the case of electrons interacting with matter, the acceleration is given by the deflection force due to the Coulomb field of the surrounding nuclei.

[1] The energy loss per distance travelled is often defined as the "stopping power" of the material.

Photon interactions

Photons interact with matter via three different effects, whose relative contributions depend on the photon energy. At energies below 100 keV the dominant process is the complete absorption of the photon by an atom, followed by the emission of one of the inner shell electrons (*photoelectric effect*). All the energy is transferred to the electron in the form of ionization and kinetic energy. The vacant position in the atomic shell is filled by outer electrons and the process is therefore accompanied by the emission of low energy photons or Auger electrons.
At energies around 1 MeV the dominant interaction process is the inelastic scattering of the photon with one of the quasi-free atomic electrons of the outer shells (*Compton scattering*). The photon is not absorbed but loses part of its energy, which is transferred to the electron in the form of kinetic energy. The shift in the photon wavelength (and frequency) is given by:

$$\lambda' - \lambda = \frac{h}{m_e c}(1 - \cos\theta) \tag{3.2}$$

and results in a continuous photon spectrum with an end-point at the highest possible transferred energy, corresponding to a backscattered photon at $\theta = \pi$.
For higher energies, the dominating process is *pair production*, which consists in the creation of an e^+e^- pair in the presence of a nucleus. The process has an energy threshold at twice the electron rest mass ($2\,m_e = 1022\,\text{keV}$). The electron of the created pair is usually absorbed in the material, while the positron annihilates with one of the surrounding electrons, emitting two photons having an energy of 511 keV each. In general, all the energy of the first energetic photon is reabsorbed in the material, leading experimentally to a distinct energy peak as in the photoelectric effect. If one or both of the photons generated by the positron annihilation escape, however, "escape peaks" at lower energies can also be observed.

3.1.2 The energy-band model

The energy levels of the electrons in an atom are arranged in a discrete pattern and their values are given by the solutions of the Schrödinger equation for a central potential. When two atoms join together to form a molecule, the energy levels of their electrons combine into other different discrete levels, some of which are splitted into doublets. When more and more atoms ($\mathcal{O}(10^{23})$) are brought together, as in a solid, the electronic energy levels are so dense that they can be approximated to continuous *energy bands*. The remaining levels which are not "covered" by any band and are thus forbidden are usually referred to as *band gaps*.
In general, at thermodynamic equilibrium, the average number of electrons in a state of energy E is proportional to the Fermi-Dirac distribution:

$$f(E) = \frac{1}{1 + e^{(E-E_F)/k_B T}}\,, \tag{3.3}$$

where $k_B T$ is the product of the Boltzmann constant and the temperature and E_F is the *Fermi level*. It is easy to see that in the limit $T \to 0\,\text{K}$, the value of $f(E)$ tends to zero if $E > E_F$ and tends to unity otherwise. This can be interpreted as the fact that at the absolute zero temperature all the energy levels are filled, consistently with the Pauli exclusion principle, up to E_F, and there are no electrons with energies greater than the Fermi level.
The electrical properties of a material depend on the location of E_F within the band structure. If the Fermi level lies inside a large band gap ($\Delta E > 3\,\text{eV}$) and far away from allowed levels, the material behaves as an insulator. In a conductive material E_F is located inside at least one band. Semiconductors have a peculiar configuration: they are similar to insulators, in the sense

that the Fermi level lies inside a band gap, but the gap is smaller ($\Delta E \approx 0.20\,\text{eV}$ - $2\,\text{eV}$) and the energy bands are closer to the Fermi level. Usually, the highest energy band completely filled with electrons at $T = 0\,\text{K}$ is called *valence band*, while the band immediately above is called *conduction band*. Therefore, the previous statement can be rephrased stating that insulators are those materials in which the separation gap between valence and conduction band is large enough to prevent electrons from being thermally excited into the energy levels of the conduction band. Conductors are materials for which the two bands overlap, while in semiconductors the separation gap between conduction and valence band is small and excited electrons can easily occupy states in the conduction band, creating electron vacancies in the valence band (called "holes"). The configuration of valence and conduction bands for different materials is sketched in figure 3.2.

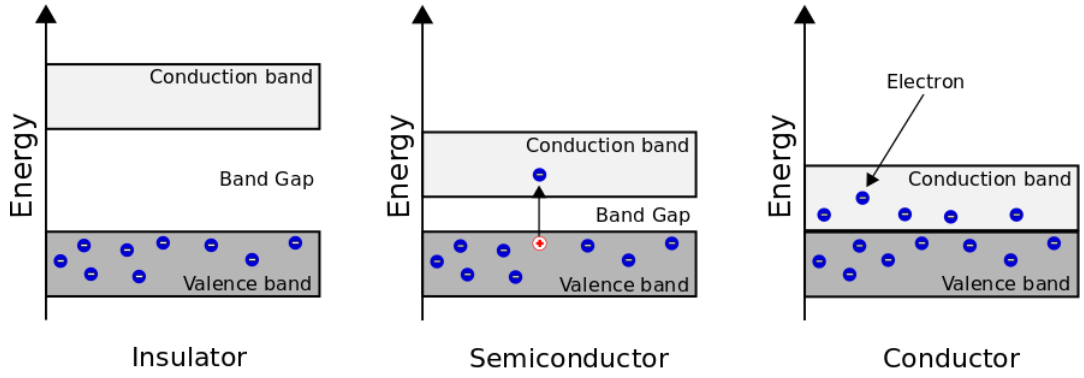


Figure 3.2: Structure of the energy bands in different materials. Insulators (left) have a large energy gap between valence and conduction band. Semiconductors (center) have a smaller gap, of the order of few eV, and electrons can be thermally excited into the conduction band, leaving vacancies (positive holes) in the valence band. Conductors (right) have no band gap and the electrons can freely enter into the conduction band.

Among the semiconductor materials, the ones which are mostly used for particle detection are silicon (Si) and germanium (Ge). Germanium has a gap between valence and conduction band equal to $\Delta E = 0.74\,\text{eV}$ at $T = 0\,\text{K}$, while Silicon has an energy gap slightly larger: $\Delta E = 1.17\,\text{eV}$ at $T = 0\,\text{K}$ and $\Delta E = 1.124\,\text{eV}$ at $T = 300\,\text{K}$. The energy required for creating an electron/hole pair is actually greater than the band gap, since part of the energy deposited in the material is lost, for example in the collective excitations of the atomic lattice (phonons). For silicon, the electron/hole pair creation energy is equal to $3.6\,\text{eV}$. Comparing this number with the energy necessary for the ionization of a gas molecule ($\sim 30\,\text{eV}$) it can be immediately seen that, on average, the number of produced charge carriers in a semiconductor is approximately ten times higher than in a gaseous detector and this property has a beneficial effect on the energy resolution of the device.
However, these properties are not sufficient for allowing a piece of pure (or intrinsic) semiconductor to work as a detector medium. In practice, the situation is complicated by the presence of free charge carriers inside the material. In an intrinsic semiconductor and in the absence of external ionizing radiation, the electron/hole pairs are generated by thermal excitations of the electrons into the conduction band. In this situation, the concentration of negative and positive charge carriers is equal ($n_i = p_i$) and both contribute to the electrical properties of the material. When an external electric field is applied to a semiconductor, the electrons and the holes start to migrate, in a combination of a random thermal velocity and a drift velocity parallel to the applied field. At low or moderate applied fields, the drift velocity of the electrons and holes is

proportional to the field itself, where the proportionality factor is referred to as *mobility*. It is interesting to note that in a semiconductor the mobility of holes and electrons has roughly the same order of magnitude, while in a gaseous detector the mobility of electrons is much larger than the one of ionized atoms.
The electrical resistivity of the material depends on the concentration of charge carriers and their mobilities and, for an ideally pure silicon material at room temperature, is equal to approximately 230 kΩ·cm [60].
Considering a hypothetical detection device made of a piece of intrinsic silicon having the geometrical dimensions of $1 \times 1\,\mathrm{cm}^2$ and a thickness of $100\,\mu$m, this translates into an electrical resistance of $R = 2.3\,\mathrm{k}\Omega$. If we now imagine to place an ohmic contact on both the surfaces and to apply an external potential of 100 V, the current flowing in the material would be $\sim 4.3 \times 10^{-2}$ A. In comparision, a minimum ionizing particle with an energy of 1 MeV interacting with the silicon bulk would produce roughly 10^5 electron/hole pairs which, on a time scale of 10 ns (the typical readout time for a detector of that thickness, see [60]), would give rise to a current of $\sim 10^{-6}$ A: four orders of magnitude lower than the one originating from the free charge carriers created by thermal excitations.
The only way to employ pure semiconductors as particle detectors is to keep the temperature of the device sufficiently low: this is indeed the working principle of high-purity germanium (HPGe) detectors, which can be operated as high-resolution energy spectrometers when cooled down to 77 K with liquid nitrogen. The key point for the successful application of semiconductors as detector medium also at higher temperatures, however, comes from the artificial alteration of the electrical properties of the material by the controlled introduction of impurities into the atomic lattice (a procedure known as *doping*). When two differently doped semiconductors are combined together, it is possible to create a spatial region which is reasonably free from free charge carriers.

3.1.3 Semiconductor doping

Doping is a mechanism in which additional elements are inserted into the lattice structure of a pure semiconductor, changing its electrical properties. Silicon, for example, is a group IV element and its atoms have four valence electrons which are involved in covalent bonds with the surrounding atoms of the lattice. If an element of group V (such as phosphorus (P) or arsenic (As)) is added in the atomic structure of silicon, one of the five valence electrons of the added element would be weakly bounded to the lattice and would therefore behave as a quasi-free electron. In this case, the added impurities (or dopants) are referred to as *donors* and the doped semiconductor is said to be of *n-type*. Adding a group III element such as boron (B) would create a "vacancy" due to the presence of only three valence electrons. If an electron is captured to fill this vacancy, it will participate in a covalent bond which is weaker than the others, because one of the two participating atoms is a trivalent impurity. From the electrical point of view, the vacancy would therefore behave as an additional quasi-free positive charge: a hole. In this case the dopants are called *acceptors* and the doped semiconductor is said to be of *p-type*.
In the energy-band model described in the previous section, this can be explained with the fact that dopants create additional energy levels inside the gap between the valence and conduction bands of the semiconductor (see figure 3.3). Donors introduce energy levels slightly below the conduction band: the additional electrons can easily be excited into the conduction band and contribute to the overall conductivity. Oppositely, acceptors introduce energy levels close to the valence band: electrons from the valence band can be easily promoted to this new energetic states and leave behind additional quasi-free holes. With these new levels, the position of the Fermi level shifts accordingly, towards the valence or the conduction band.

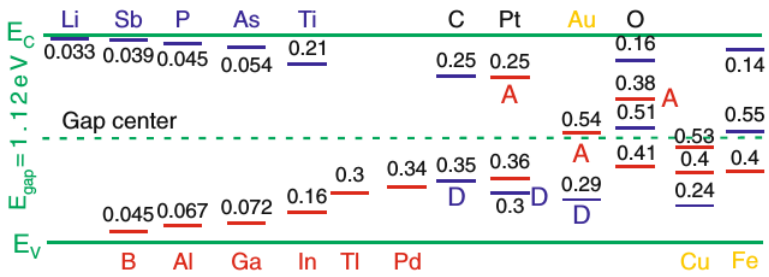


Figure 3.3: Additional energy levels introduced into the silicon band gap by dopants: the letter "A" stands for acceptors, "D" for donors. The additional energy level created by boron doping, for example, lowers the excitation energy from 1.12 eV to 0.045 eV [60].

After doping, the concentration of charge carriers becomes completely dominated by the added impurities ($n \approx N_D$ for donors and $p \approx N_A$ for acceptors) but an equilibrium relation between electrons and holes still holds and is the same as for intrinsic materials: $np = n_i p_i$.

3.1.4 The pn-junction

A *pn-junction* is a combination of two oppositely-doped semiconductors. Because of the different Fermi levels E_F, when an n-type semiconductor is joined to a p-type semiconductor[2], the electrons from the n-doped side of the junction diffuse towards the p-side, which has a lower E_F, creating a net current which stops as soon as a common Fermi level across the pn-junction has been estabilished (see figure 3.4).

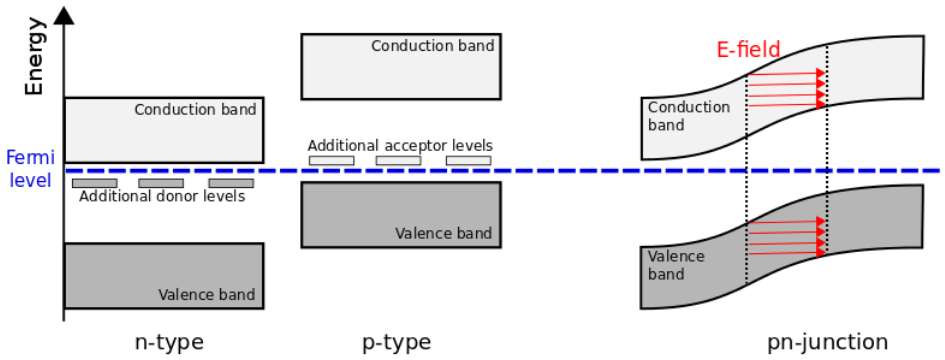


Figure 3.4: Energy band structure of n-type and p-type doped semiconductors, with the additional energy level introduced into the band gap by the dopants. When the two types are combined into a pn-junction, a common Fermi level is estabilished.

More intuitively, the diffusion of electrons from the n-side into the p-side leaves behind a positive space charge in the form of ionized donor atoms, which are not free to move inside the lattice. Conversely, the diffusion of holes into the n-side leaves a negative space charge due to the acceptor sites which picked up an extra electron. These fixed space charges create an electric field which acts in the direction of diminishing the tendency for further diffusion. At the interface between

[2]In practice, a simple mechanical connection between n-type and p-type materials is not sufficient, since the leftover gaps would be huge compared to the interatomic lattice spacing. A pn-junction is normally formed inside a single crystal, changing the impurity content from one side of the junction to the other.

the two oppositely-doped semiconductors, the recombination of electrons and holes creates a *depletion region* where no free charge carriers are present. The extension of the depletion region into the donor and acceptor sides depends on the relative concentration of the impurities in both regions (see figure 3.5).

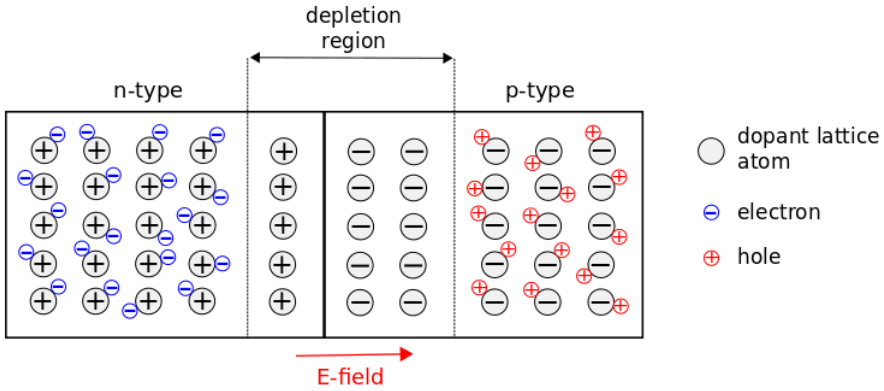


Figure 3.5: A pn-junction in thermal equilibrium. The fixed space charge of the lattice atoms as well as the electon and holes are sketched. The recombination of electrons and holes in the oppositely doped regions creates a depletion zone which is free from mobile charge carriers.

The presence of this depletion region has important consequences for the operation of a pn-junction as a detection medium. The electron/hole pairs formed after the interaction of ionizing radiation with the material will tend to drift in opposite directions, following the electric field of the depletion region and their motion will give rise to a measurable electric signal, proportional to the number of created pairs and, therefore, to the energy of the incident radiation.
When an external bias voltage is applied to the pn-junction, two configurations are possible.

1. **Forward bias**. The p-doped side of the junction is connected to the anode (positive voltage) and the n-side to the cathode (negative voltage). In this situation, the electrons from the n-side and the holes from the p-side would drift towards the junction and, since they are the majority carriers in both regions, the conductivity across the junction will be greatly enhanced. As soon as the external voltage exceeds the small E-field inside the depletion region, the pn-junction breaks down and the region without free charge carriers dissolves.

2. **Reverse bias**. The n-doped side is connected to the anode and the p-doped side to the cathode. In this case, the minority carriers (holes in the n-side and electrons in the p-side) tend to move across the junction but, since their concentration is low, the reverse current across the junction will be small. The applied reverse bias accentuates the potential difference across the junction and causes the depletion region to increase.

Based on what has been illustrated above, it is clear that the optimal configuration for operating a pn-junction as a particle detector is to apply a reverse bias which is high enough to completely deplete the device, since the depletion region constitutes the active medium of the detector. In particular, the width W of the depletion region increases with the applied reverse bias as:

$$W = \sqrt{2\epsilon(\mathrm{V_{bias}}/eN_D)}\,, \tag{3.4}$$

where $\epsilon = \epsilon_0\epsilon_r$ is the product of the vacuum permittivity and the relative permittivity of the semiconductor material, e is the electron charge and N_D is the dopant concentration.

The small current which flows in the device when operated in reverse bias mode is called *leakage current* and originates from the flow of minority carriers across the junction: it has a strong dependence on the temperature, and increases linearly with the applied bias voltage until the depletion is maximal. Modifications in the lattice structure due to external factors such as radiation damage also have strong influence on the magnitude of the leakage current.
A more detailed derivation of the properties of the reverse bias junction can be found in textbooks such as [60, 61].

3.1.5 Working principle of pixel detectors

In its simplest form, a pixel detector consists of a pn-junction with segmented electrode implants in the silicon bulk. In planar pixel detectors, the segmented implants are located on the surface of the silicon and doped oppositely with respect to the bulk (for example, p-doped implants on an n-doped bulk). By applying a reverse bias, the depleted region can be extended over the full bulk material such that the active area of the detector corresponds to the whole geometrical thickness of the silicon. The electron/hole pairs, generated by the interaction of particles or radiation with the depleted silicon bulk, drift towards the two faces of the bulk material, following the lines of the applied field and generating a measurable electrical signal. If the implants are p-doped, the charge carriers which are collected are holes, while for n-doped implants electrons are collected. In both cases, in planar geometries the charge carriers need to traverse roughly the whole bulk material before reaching the implants and this has an influence on the charge collection speed. To improve on this, other geometries are under investigation, for example creating 3D dopant implantations which reach through almost the whole bulk material.
Usually pixel detectors are designed following a hybrid approach where the active silicon material (normally referred to as the *sensor*) is connected to an external pixelated *front-end chip* which fits precisely the sensor area. In general, the main functionalities of the front-end readout electronics consist in collecting the charge from the sensor and passing the resulting signal to a preamplifier and a pulse-shaper. The signal is then compared to an adjustable threshold and, if it exceeds the threshold level, is converted into a digital value by an Analog-Digital Converter (ADC) and stored in a buffer until the readout takes place. The hybrid design has not only the obvious advantage of allowing the separate optimization of the processes for sensors and readout chip wafers, but is also beneficial when it comes to irradiation studies, since the radiation effects on the sensor and on the readout electronics can be disentangled and studied separately. On the other hand, for such an approach to work reliably, it is necessary to implement high-density vertical interconnection techniques between the sensor and readout electronics. The bump bonding technique, which will be described in detail in section 3.2, is an exemplary solution for this kind of assembly problems. Figure 3.6 shows a schematic cross section of a hybrid pixel detector cell, where the sensor is mechanically and electrically connected to the front-end chip via a solder bump.

A completely different approach consists in building both electronics and sensor in the same technological process. The possibility of employing such monolithic pixel detectors also for particle physics experiments is being actively investigated and might prove itself to be a successful concept in the near future, replacing the more traditional hybrid design [62].
The interaction of a particle with the pixelated area of the active material of the detector and the conversion of the generated charge into an electrical signal results in a precise point (or hit) localized in a two-dimensional plane. If the pixel detector is arranged into a sequence of planes (or layers), the track of the incident particle can be reconstructed interpolating the hits on subsequent layers by means of sophisticated fitting algorithms. This explains the importance of

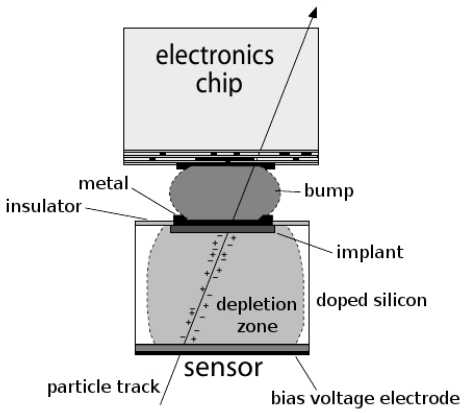


Figure 3.6: Schematic view of a pixel unit cell in a hybrid pixel detector. The ionizing particle crosses the fully depleted region of the sensor, generating electron/hole pairs which, moving in the electric field, produce a measurable electrical signal. The thickness of the depletion region depends on the bias voltage applied to the backside of the sensor. Adapted from [62].

pixel detectors as tracking devices as well as their fundamental role in providing a precise spatial measurement, which can then be used for the identification of track vertices. This is especially useful in situations where the track multiplicity is high (for example, in heavy ion collisions) or when studying the decays of long-living particles (for example, for the identification of the decay vertex of a hadron containing a bottom-quark).
The spatial resolution of a pixel detector depends primarily on the geometrical length and width of the pixel cells (pitch). An improvement can be obtained orientating the sensor such that an impinging particle crosses more than one pixel cell for each layer. In this way, a more precise estimate of the impact position can be obtained interpolating between the signals induced on different cells (charge sharing). Of course, if the charge carriers are shared among too many adjacent pixels the overall performance will soon start to degrade since low signals would be hardly distinguishable from the noise. Charge sharing between more than two pixels is therefore avoided. Another way of enhancing charge sharing, which is equivalent to tilting the surface of the pixels, consists in bending the direction of the interacting particles, for example placing the detector inside a magnetic field. Charged particles get deflected by the Lorentz force and will impact on the pixel surface with a non perpendicular angle. It is worth noting that the same Lorentz force is experienced also by the charge carriers drifting inside the active sensor material and this effect contributes to the charge sharing as well.

3.2 Bump bonding interconnection technology for pixel detectors

The production of hybrid pixel detectors for particle physics experiments relies on the possibility to precisely connect the active detection medium (the pixelated depleted silicon sensor) with the readout electronics, while keeping the pixel cell pitch around 100 μm or possibly even lower. So far, one of the most successful technologies which allowed to achieve this goal is the so called "flip-chip" bonding. Flip-chip bonding, which was commercially developed by IBM in the 1960s under the name of C4 as a packaging solution for integrated circuits, is a method for connecting semiconductor devices to external circuitry, based on small solder bumps which are deposited on the chip itself during the final wafer processing steps. This packaging technology allows to

achieve the fine-grained and high interconnection densities which are mandatory for medical or high energy physics (HEP) applications and which would be simply unreachable with other connection technologies such as, for example, the ones which make use of thin metallic wires (*wire bonding*). The solder bumps provide both an electrical and mechanical connection between the sensor and the readout electronics, allowing to produce more compact assemblies. One of the drawbacks of this approach is that it often requires complicated and possibly expensive wafer processing steps, which are necessary in order to deposit the small bumps of soldering material on the wafer surface.
The most widely used processes for bump deposition require the topmost substrate of the chip (usually an aluminized bonding pad for each pixel cell) to be prepared with a set of different metallic layers, which are often referred to as Under Bump Metallization (UBM). The UBM ensures the adherence of the actual solder material and prevents intermetallic diffusion. A typical UBM consists of a metallic layer of wettable material, for example nickel, which acts as the actual "glue" between the substrate and the soldering metal. The thickness of this layer is usually around 2-3 μm. Other intermediate layers made of refractory metals[3] such as tungsten or titanium can be added as a barrier in order to prevent the nickel from diffusing inside the underlying silicon and to create dangerous shorts which might destroy the chip. The topmost layer of the UBM consists usually of a thin surface gold layer which protects the nickel from oxidation. The metallic layers can be deposited electrochemically or by sputtering, with the second option being generally more expensive due to the requirement of photolithographic production steps.
On top of the UBM, the soldering material can be deposited via different methods [63]. The procedure pionereed by IBM with the C4 technology is based on **evaporation**: the solder is evaporated on a wafer covered by a metallic shadow mask, which is later removed. The metal vapors condense on the wafer and small bump structures are selectively formed in the regions corresponding to the holes left in the mask. The minimum bump diameter depends on the mechanical properties of the mask and is limited to values above 100 μm, making this technology not suitable for HEP applications. To overcome this limitation, thick photoresist layers patterned with UV light (a process known as photolithography) can be used instead of metallic masks. The evaporation is done similarly to the C4 process and the photoresist is subsequently etched, leaving only the grown structures. Indium bumps, evaporated on both the sensor and the readout chip substrate following the above procedure, have been used as soldering metal for the current CMS pixel detector and proved to work reliably [64].
A good alternative to evaporated bumps is **electroplating**: in this case, previously deposited UBMs are used as a plating base, with the areas to be bumped carved out of a photoresist layer via a lithographic step. The bumps are then deposited thanks to an electrochemical reaction inside a solvent. The photoresist is then removed and the plating base etched. During this process, a voltage needs to be applied to the chip in order to induce the deposition and the chip itself needs therefore to be conductive. A more detailed description of a bump deposition process using electroplating will be given in section 4.1.1, since this is the method employed for the bumping of the readout chips used at Karlsruhe Institute of Technology for the production of the modules for the pixel Phase I Upgrade.
Another possibility is given by **electroless plating** techniques, which are based on spontaneous chemical reactions between a metal and reactants in a bath. These methods are appealing because photolithography is not required and no external electrical current is needed. However, electroless plating is mainly suitable for producing very small bumps or UBM layers depositions, since the growth of larger structures is more difficult to control.

[3] Refractory metals are a class of metals which possess a very high melting point (above 1850 °C) and high hardness at room temperature.

Concerning the actual soldering material, both single metals, like indium mentioned above, as well as mixtures of different metals can be used. In the latter case, the mixture has often *eutectic* composition, meaning that the single components are mixed in a well-defined ratio, such that the global system melts as a whole at a specific temperature. Should this not be the case, the different components of the solder would melt (and solidify) at different temperatures, influencing the uniformity of the material and, possibly, its physical behaviour. One of the mostly used eutectic alloys for soldering is made of a mixture of 63% tin (Sn) and 37% lead (Pb). Recently, the tendency in the industry to move towards "lead-free" components has triggered deeper investigations in alternative soldering materials, such as tin (96.5%)/silver (3.5%) or tin (95.8%)/silver (3.5%)/copper (0.7%) mixtures.
After deposition on the substrate, the solder studs have often the shape of a more or less deformed cilinder. At this point, the wafer can be heated up above the liquid phase of the solder under a controlled atmosphere (a process called **reflow**): due to surface tension the cylinders will collapse into truncated spheres and maintain this shape also after being cooled down to room temperature again. Depending on the metal used, the melting point varies over a broad range: for example, the melting point for indium bumps is $\sim 156\,^{\circ}\mathrm{C}$, while it is $183\,^{\circ}\mathrm{C}$ for eutectic SnPb bumps. In general, lower melting points are favourable, since the risk of metal diffusion between solder, UBM metals and substrates during the reflow phase gets mitigated and less mechanical stress due to thermal gradients is put on the reflowed structures.
It is worth noting that all the deposition methods described so far include masking, photolithography, stripping or removal processes which are optimized for wafer-scale productions and less suitable for R&D projects or small in-house prototyping (for example, on the scale of single chip bumping). There are however other methods available on the market, which include **physical ball transfer technologies** such as ball jetting, where single solder spheres are dropped from a ball reservoir through a capillary, melted inside a placing nozzle by means of laser pulses and then precisely jetted onto the substrate, where they solidify. It is important to note that even if the bump deposition itself does not require a lithographic step, the process profits from a previous UBM deposition to ensure the solder adherence. A solder jetting procedure of this type is used at DESY[4] to bump the sensors for the CMS pixel Phase I Upgrade [65].
Another interesting option, inspired by wire bonding techniques, is to use **gold studs** as bump metals: instead of making a wire loop between two pads, only the first contact is made and the rest of the wire is then cut off slightly above. This gives a stud-like structure with a characteristic "pig-tail", whose diameter depends on the size of the wire. This method has the advantage of not requiring complicated UBM deposition processes, since the gold studs are ultrasonically bonded directly onto the aluminium surface of the chip pads. The costs per bump are however higher than other bumping technologies, due to material costs. Despite this, this technique remains appealing, especially for low-volume prototyping (single-chip bumping) or R&D activities, as proven also by research activities carried out at Karlsruhe Institute of Technology [66, 67].
Depending on the material used and on the desired final separation between sensor and readout chip, different strategies are possible: deposit the solder balls on both the sensor and the front-end chip sides, deposit the solder only on the chip side while preparing the sensor substrate with a UBM deposition (if required) or deposit the solder only on the sensor side. Once the substrates have been prepared with solder and/or UBM layers deposition, the real bonding process can start, as sketched in figure 3.7.

The chip is flipped such that the bumped area is on the bottom side (hence the name of the technique), positioned over the substrate (for HEP assemblies this is typically the sensor side) and precisely aligned with respect to it, usually matching some unique reference patterns. The

[4] Deutsches Elektronen-Synchrotron.

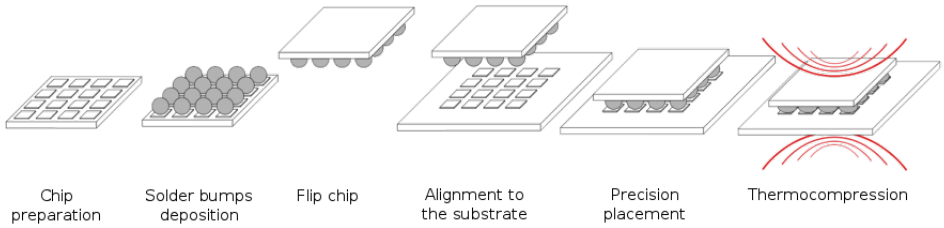


Figure 3.7: Illustration of a general flip-chip bonding process. The chip on which the solder bumps have been previously deposited is flipped and carefully aligned with respect to the substrate. The two components are then joined together by thermocompression. Adapted from [68].

two components are then merged in one single assembly through a mechanical compression force, usually accompanied by an increase in temperature, which can go up to 70-90% of the solder melting temperature (thermocompression). At this stage, not only is it mandatory for the alignment to be very precise, but the planarity between sensor and readout chip should also be kept under control, in order to avoid tilt angles that might lead to electrical shorts or to open connections. Usually, automated pick-&-place and bonding machines having a positioning accuracy of the order of few μm are employed for these kinds of procedures. The temperature needs also to be tuned carefully and to be distributed uniformly along the surfaces of the components, avoiding strong thermal gradients between the different parts: especially in the case of different sensor/ROC materials, the difference in the coefficients of thermal expansions translates directly into mechanical stress, leading to possible damage.
After bonding, the joined assembly (which at this stage already possesses a reasonable mechanical connection strength) usually undergoes a reflow process where it is heated up above the liquid phase of the solder under a controlled atmosphere, following a thermal profile optimized according to the bump metal and to the heat tolerance of the single components. The aim of the reflow process is to improve the mechanical connection between the bonded parts, as well as to profit from the self-alignment between sensor and readout chip, which is driven by the surface tension of the molten soldering material during the liquid phase. For achieving a high connection quality, it is essential to remove any metal oxidation from the surfaces to be soldered and to prevent re-oxidation. Since the use of flux, as done in regular soldering, needs to be avoided here because it leaves residuals which are difficult or impossible to remove, fluxless soldering techniques are applied, usually assisted by appropriate gaseous reducing agents. The atmosphere of the reflow chamber is therefore an important parameter for the process. Among the gaseous reducing agents, one of the most used is formic acid (HCOOH), which has the advantage of being active at typical reflow temperatures and decomposing itself without leaving contaminants. The reaction with the oxide is initiated at temperatures above 150 °C, where formic acid reacts with the metal oxide (MeO) giving metal formates and water [69]:

$$\mathrm{MeO} + 2\mathrm{HCOOH} \rightarrow \mathrm{Me(COOH)_2} + \mathrm{H_2O}\,. \tag{3.5}$$

If the temperature is above 200 °C, the reaction results in the reduced metal (Me), plus gaseous carbon dioxide and water:

$$\begin{aligned} \mathrm{Me(COOH)_2} &\rightarrow \mathrm{Me} + 2\mathrm{CO_2} + \mathrm{H_2} \\ \mathrm{H_2} + \mathrm{MeO} &\rightarrow \mathrm{Me} + \mathrm{H_2O}\,. \end{aligned} \tag{3.6}$$

Since the regrowth of metal oxides has to be prevented, the presence of oxygen or air in the reflow chamber needs to be avoided, for example by flushing inert gases such as nitrogen (N_2) and by creating an overpressure inside the chamber itself.

The peak temperature reached during the reflow process is usually around 30-40 °C above the metal melting point and is kept only for a short period of time (below 60 seconds), in order to avoid component damage or an excessive growth of intermetallic layers, which can lead to joint brittleness (see also section 4.1.4.3). The peak temperature phase in the reflow is then followed by a cooling-down phase in which the solder joints gradually solidify and the assembly is brought back to much lower temperatures, where it can be safely handled.

3.3 The CMS pixel detector

A specific example of a semiconductor detector for high energy physics is given by the pixel detector of the CMS experiment currently collecting data at the Large Hadron Collider.
The pixel detector is the innermost part of the CMS tracking system, closest to the interaction region. It has been designed and built in order to provide precise tracking of charged particles in the r-ϕ plane as well as in the z direction, small impact parameter resolution and good vertex reconstruction. Like many other subdetectors in CMS, it has a cylindrical structure, with a barrel region (referred to as BPIX) around the beam pipe, closed at the ends by two endcaps (FPIX). Each pixel cell has a size of $100 \times 150\,\mu\text{m}^2$, with the lower dimension in the r-ϕ direction for the barrel and in the r direction for the endcaps: this pixel size has been chosen in order to achieve a similar resolution in all the spatial directions [31]. The whole pixel system covers a pseudorapidity range of $|\eta| < 2.5$, and its layout is presented in figure 3.8.

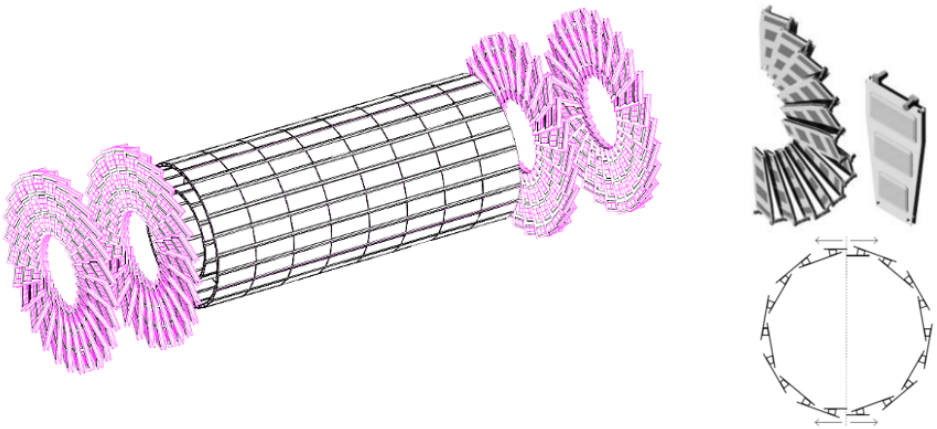


Figure 3.8: Overview of the CMS pixel detector (left). The two smaller inserts in the picture show: the turbine-like geometry of one disk of the forward pixel system and an FPIX plaquette (top, right); a radial cut of the first barrel layer, with the mechanical division in half-shelves and the staggering of the ladders which hold the BPIX modules (bottom, right) [33].

The barrel region consists of three layers placed at an average distance from the beam line equal to 44, 73 and 102 mm respectively and having a total length in the z direction equal to 530 mm. The FPIX disks extend from a radius of 60 mm to 145 mm from the beam line and are placed on each side of the center of the CMS detector, at distances $z = \pm\,345$ mm and $z = \pm\,465$ mm. The barrel layers and the endcap disks have been arranged in order to provide three tracking points over almost all the covered η range. These points are used to reconstruct a three-dimensional vertex in space (which is fundamental for defining the primary vertex of the interaction, as well as the secondary vertices of the decay of long-living particles such as τ leptons or hadrons containing bottom quarks), to form seed tracks for the outer track reconstruction and for the High Level Trigger.

The pixel system is placed inside the CMS solenoid magnet, with the 3.8 T magnetic field lines perpendicular to the endcaps and parallel to the barrel layers. The resulting Lorentz drift of the charged particles leads to charge sharing between adjacent pixels. This sharing effect helps in achieving a measured position resolution around 18 μm (in the r-ϕ direction) for high energetic cosmic muons interacting with the detector; a value which is significantly lower than $d/\sqrt{12}$, with d being the pixel pitch, namely the resolution expected for a binary readout [70]. Charge sharing is enhanced by a clever choice of the detector layout: in the FPIX, the blades which support the active detecting material are tilted by 20 ° and arranged in a turbine-like geometry, in order to ease the crossing of the particles with the detector at a non-normal angle (see the insert in figure 3.8).
Both the BPIX and the FPIX are built following a modular structure. In the BPIX there are 768 modules: 672 are full modules, while the remaining ones are half modules, installed in the boundary region where the two half-shelves in which the barrel structure is mechanically subdivided are joined together. The total number of pixels in the barrel section is approximately 48 million, covering an active area of 0.78 m^2. The forward region has a more complicated structure: each disk is divided into so called "plaquettes", having different geometrical dimensions, in order to cover the radial geometry without leaving gaps. There are five different types of plaquettes, for a total of 672 in the whole forward detector. The number of channels is approximately 18 million, distributed over an area of 0.28 m^2.
In the BPIX, each bare module[5] consists of a silicon sensor, connected via indium bumps to 16 readout chips (ROCs) arranged in a 2 × 8 grid. The half modules manufactured for the boundary region between the half-shelves of the barrel have only a single row made of 8 ROCs. Each chip consists of a matrix of 52 × 80 pixels, where the first number specifies the columns and the second is the number of rows. The modules in the FPIX have five different sizes, one for each of the five types of plaquettes, and contain therefore a different number of ROCs: 1 × 2, 2 × 3, 2 × 4, 1 × 5 or 2 × 5 (the notation is always in the form row × columns). Despite the difference in geometry and layout, the single readout chip design of the FPIX is completely equivalent to the ones installed in the barrel layers.

3.3.1 Pixel modules

For the sake of simplicity, the structure of a CMS pixel module will be outlined using a BPIX module as reference. The FPIX concept is in principle very similar, with only some modifications ascribable mainly to the different geometries involved. An exploded view of a BPIX module is shown in figure 3.9.

This design represents a good example of a hybrid pixel detector. The core consists of a silicon sensor, connected to the readout chips via bump bonding. A flexible, low-mass, three layer printed circuit board, called High Density Interconnect (HDI), is glued on the back side of the sensor and connected to each single ROC via wire bonds. The Token Bit Manager (TBM) as well as the power and signal cables are connected to the HDI. The module design is completed by silicon nitride (Si_3N_4) base strips, which are needed to fix the module to the mechanical support structure. A fully equipped BPIX module, including the base strips, has a size of 66.6 × 26.0 mm^2, a weight of 2.2 g (without cables) and consumes 2 W of power.

[5] From now on and especially when discussing in detail the production of pixel modules for the Phase I Upgrade, the term "bare module" will refer to a yet incomplete module made only of a sensor and readout chips, without the additional components which complete it into a full module.

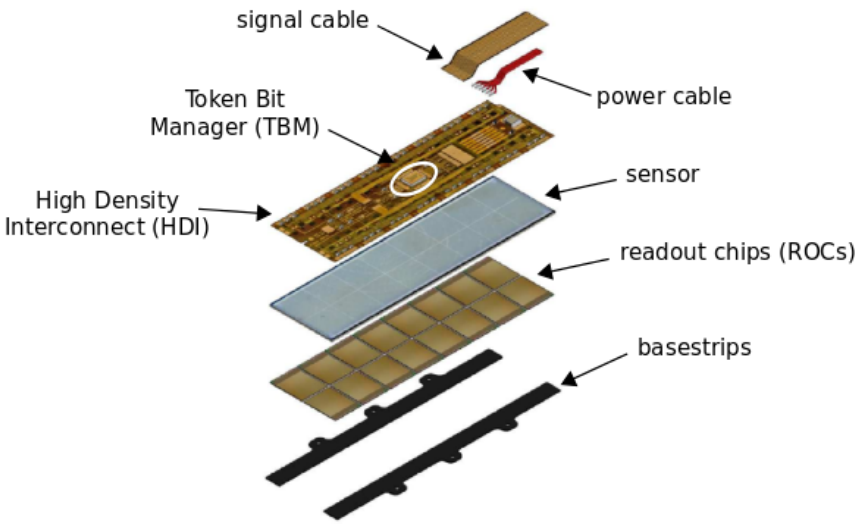
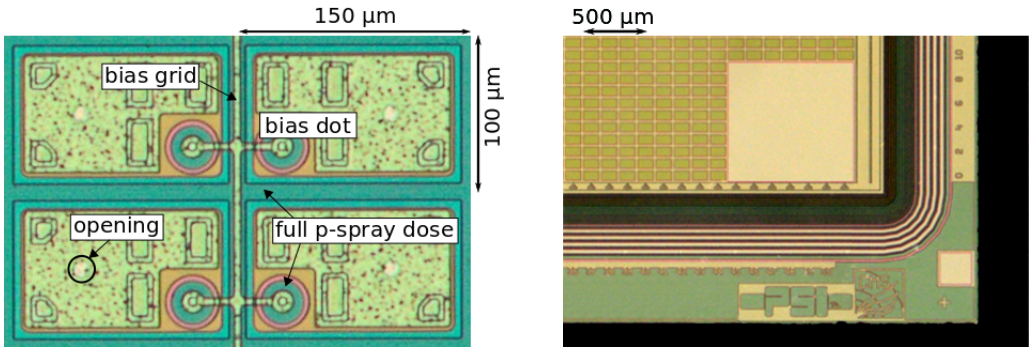


Figure 3.9: Exploded view of a BPIX module with all its subcomponents. The sensor and the readout chips are connected via bump bonding and together form a bare module. The TBM is glued and wire bonded to the HDI, which in turn is glued on top of the sensor and wire bonded to the 16 ROCs. Two base strips for the mechanical connection to the support structure complete the assembly. Adapted from [70].

Sensor technology

Given the close distance to the interaction region, the pixel detector has to sustain very high particle fluences, especially in the innermost layer. A radiation-hard design was therefore mandatory. For this reason, the technology which has been chosen for the sensor consists of highly doped n implants (n^+) on a lightly n-doped bulk. This implies that the collected charges are electrons, which is advantageous since they have a higher mobility compared to holes and are less prone to trapping, therefore ensuring a high signal charge at moderate bias voltages ($\leq 600\,\mathrm{V}$), also after high hadron fluences. In a silicon sensor, a Lorentz angle can be defined as the angle between the drift direction of the charge carriers under the influence of an external magnetic field and the direction of the electric field applied to the sensor. It can be seen that electrons have a larger Lorentz angle and this, together with the choice of the detector geometry described in the introductory paragraph, leads to easier charge sharing between adjacent pixels, thus improving the spatial resolution. The disadvantage of the n^+-in-n technology is that the sensor requires double sided processing, since the back side needs to be p-doped in order to create a pn-junction: this leads to higher costs compared to single-side processing. Another drawback of the n^+-in-n design is that an electron accumulation cloud tends to form near the negative dopants on the front side of the sensor, leading to possible shorts between adjacent pixels. To prevent this, the pixel cells are isolated thanks to a moderated p-spray technique (for the BPIX) or to open p-stops (for the disks of the FPIX), which guarantees good separation among the pixel cells also after high radiation doses. On the back side of the sensor, a guard ring scheme has been designed to ensure that all the sensor edges are at ground potential [71].
Four cells of a CMS sensor for the barrel modules are shown in figure 3.10(a). Most of the visible pixel area on the front side of the sensor is covered with the collecting electrode formed by the n-implants. The darker frame around each pixel corresponds to the region where the p-spray reaches the full dose. In the corner of each cell, a *bias dot* consisting of a circular n-implant isolated by the surrounding electrode is clearly visible. These dots are connected to a conductive line forming a *bias grid* which provides a high resistance punch-through connection to all pixels. Thanks to the grid, all the pixels can be connected to a common potential, allowing the

possibility to perform on-sensor I-V measurements when one contact supplying the bias voltage is placed on the sensor back side and another on any point of the bias grid. During this test, the leakage current (I) is measured as a function of the applied bias voltage (V). Since an increased leakage current is an indication for the presence of defects, this straightforward sensor-level characterization is fundamental to identify faulty sensors before bump bonding them.
In a full module, the importance of the bias grid connection is reduced, since each pixel cell gets a ground reference via the readout chip. The bias grid can nevertheless play a role in case of missing bumps or faulty connections, since it prevents large field gradients to be generated around the unconnected pixel. In a full module, the bias grid can be either left floating or connected to a fixed potential. In the KIT production process for the CMS Phase I Upgrade, the bias grid is connected to the ground potential supplied by the readout chip via an additional bump connection (*bias bump*). The whole surface of the pixel cells is covered by a protective passivation layer, which is opened above the small circular spots in the center of each cell to reveal the aluminized pad on top of which the Under Bump Metallization is later deposited.
The back side of the sensor (figure 3.10(b)) consists of a large p-implant where the reverse bias is applied, surrounded by a concentric guard ring scheme which gradually drops the voltage to zero until the dicing line, preventing dangerous sparks between the sensor and the readout chip. Each BPIX sensor is made of 285 μm thick diffused oxygen float zone (DOFZ) silicon and has a rectangular shape with a size of 66.6 $\times$ 18.6 mm^2 (except for the half sensors needed for the boundary modules, which have only half of the width). The pixelated matrix contains 66560 pixels, divided into 2 $\times$ 8 partitions (one for each readout chip) of 4160 pixels each, arranged in a 52 $\times$ 80 grid. The size of the edge pixels, which cover the space between two adjacent ROCs is twice the normal pixel size while the corner pixels have four times the normal size. The depletion voltage before irradiation is 50-60 V and the bulk resistivity is around 3.7 kΩ·cm, after processing.



(a) (b)

Figure 3.10: Left: detailed view of four adjacent pixel cells, as seen from the front side of the sensor. Right: sensor corner as viewed from the back side. A large testing pad and the concentric guard rings are clearly visible.

The pixel readout chip

The electrical signal developing in the depleted region of the sensor is read out by a readout chip (ROC) connected to the sensor via bump bonding. The chip currently in use in the CMS pixel, called PSI46V2, is a customized application-specific integrated circuit (ASIC) fabricated in a commercial 0.25 μm 5-layer complementary metal oxide semiconductor (CMOS) process:

a technology which is radiation tolerant when following special design rules [72]. Each ROC is divided into a matrix of 52 × 80 pixel unit cells (PUC), has a size of 7.9 × 9.8 mm^2 and a thickness of approximately 175 μm. The primary goal of the ROC is to amplify, shape and buffer the charge signal coming from the sensor, as well as to provide an individually adjustable threshold to each pixel cell. Only the signals above the adjustable threshold are read out. The ROC verifies also the Level-1 trigger information, discarding the hits which are not associated with a Level-1 trigger. In case of a positive trigger decision, the information containing the signal height, the pixel address and a time stamp information is sent to the chip periphery. The readout of the signals from the pixel unit cells is organized in 26 double columns.
For calibration purposes, an electrical signal can also be injected in each pixel cell, either directly, through a 4.8 fF capacitor connected to the the amplifier, or via the sensor side. The direct signal injection can be used to adjust the pixel threshold, equalizing the response of each pixel to a common reference level. The indirect signal coming from the sensor side can be used to test the quality of the bump bonding connection.
More about the geometry and design of the readout chip will be given in section 3.4, when describing its upgraded version.

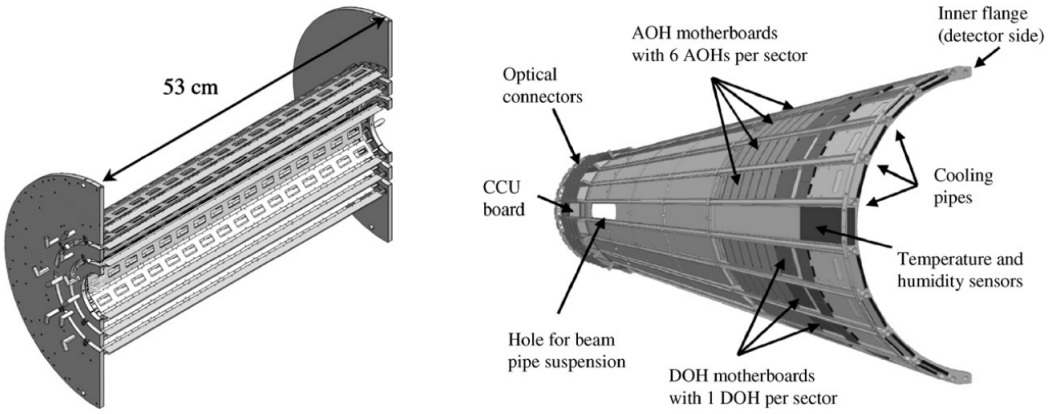
The Token Bit Manager

As the name suggests, the Token Bit Manager (TBM) chip is the component which is responsible for initiating the readout of the ROCs by distributing the token pass for each incoming Level-1 trigger, collecting the information coming from each group of ROCs and sending them to the front-end electronics via an optical link. It also takes care of distributing clock and reset signals to the ROCs.
In the BPIX, each detector module is equipped with a TBM, which is glued and wire bonded to the HDI, and which controls the readout of 8 or 16 ROCs (depending on the layer). In the FPIX, TBMs are mounted on the disk blades and control 21 or 24 ROCs (depending on the blade side).

3.3.2 Detector mechanics and cooling

The barrel pixel system is made of detector modules mounted on a support structure which follows a cylindrical geometry around the beam pipe, closed at both z-ends by a support frame (see figure 3.11(a)). The total length is 530 mm, which is much shorter than the surrounding silicon strip tracker. This means that supply tubes are needed in order to carry services along the beam pipe, from the panels located outside the tracking volume. The supply tubes also host readout electronics and control systems (figure 3.11(b)). The total length of the BPIX when the supply tubes are included becomes 5.6 m. The support structure is made of 0.3 mm thick aluminium cooling tubes, on which the 0.24 mm carbon fibre blades which hold the modules are glued. The support frames at both ends are built with fibreglass filled with foam and covered by carbon fibres, and are equipped with printed circuit boards connected to the module cables. The supply tubes carry the electrical power lines, the electronic control signals, the optical signals as well as the cooling fluid and are made of 0.1 mm thick stainless steel tubes, with additional aluminium flanges. The length of each supply tube is 2204 mm.

The forward pixel detector is located inside the BPIX supply tubes, mounted on separate rails. The mechanical structure is divided in four half-cylinders, where each half-cylinder contains two half-disks at a distance of 345 mm and 465 mm from the interaction point. The half-disks have twelve “U”-shaped cooling channels each, with 0.5 mm thick beryllium trapezoidal panels



(a) (b)

Figure 3.11: Left: drawing of half of the pixel barrel support structure. Right: a supply tube half-cylinder [73].

attached to them, on both sides. The panels, rotated in a turbine-like geometry, form the blades which hold the actual detector material (i.e., the plaquettes). The half-cylinders also contain all the electrical and mechanical infrastructures.
To maintain the temperature at a constant level and remove the heat produced by the power consumption of the detector (including the leakage current in the sensors), a liquid phase cooling with C_6F_{14} is used. There are 18 cooling loops in total, 10 for the barrel and 4 for each of the two endcaps.
Power for the pixel detector is supplied by commercial modular systems, part of which are installed in crates placed in the proximity of the detector to reduce power loss on the cables. In the crates there are two types of electronic cards: one (2.5 V/7 A) feeds the services on the supply tubes, while the other has two low voltage lines (1.75 V/7 A and 2.5 V/15 A) for ROC operation and two high voltage lines (600 V/20 mA) for sensor biasing.

3.3.3 Material budget

Any material placed inside the path of the particles coming from the interaction point and traversing the detector can lead to multiple scattering, nuclear interactions, photon conversion or photon emission (bremsstrahlung) and, in case of charged particles, reduce the accuracy of the track reconstruction. In addition, any energy loss in the tracker material spoils the energy measurement in the calorimeters behind. For these reasons, it is necessary to keep the amount of material (the so called "material budget") in the tracker as low as possible.
Usually the material budget is expressed in units of radiation length (x/X_0), where a radiation length X_0 is defined as the mean distance in which an electron loses all but $1/e$ of its energy because of bremsstrahlung (e is Euler's constant).
As can be seen from the plot in figure 3.12, the material in the pixel barrel is equal, on average, to less than 10% of a radiation length in the central region. Sensors and ROCs contribute to the material budget for roughly one third, while the support structure and the cooling fluid contribute about 50%. The main contribution of the barrel to the overall material budget is at $|\eta| > 1.2$ and is due to the support frame at both z-ends and to the inner part of the supply tube.
A large amount of material is associated with the current FPIX detector: most of the material between $1.2 < |\eta| < 2.4$ is located in the half-disks, while material between $2.4 < |\eta| < 3.6$ is

due to cables and to the cooling system.

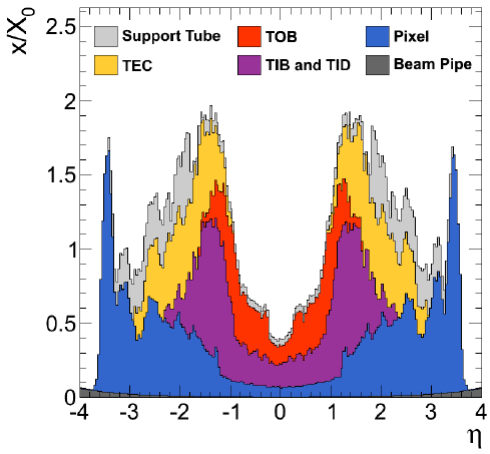


Figure 3.12: Material budget of the CMS tracker, in units of radiation length X_0, as a function of the pseudorapidity η. The contribution of the pixel detector (BPIX and FPIX) is shown in blue [50].

3.3.4 Performance of the pixel detector

The CMS pixel detector has performed very well so far, during all data taking periods. On average, more than 95% of the channels have been active, and the detector has been used to form seeds for the offline track reconstruction, as well as for fast tracking online in the High Level Trigger, performing vertex reconstruction, electron/photon-discrimination, τ identification, muon reconstruction and b-tagging.
The single hit efficiency measured in data is always above 99%, with a decreasing trend when the instantaneous luminosity increases. This is caused by a dynamic inefficency due to limits in the internal buffer of the readout chip, as well as temporary losses of modules. The track reconstruction efficiency has been measured using Z boson decays into muon pairs, as a function of the number of primary vertices and pseudorapidity: it is always high but slowly degrades as the number of pile-up events increases. When the number of vertices is above 40, the degradation becomes more rapid because of buffering issue in the readout chip. A dip in the efficiency has also been observed in a critical region around $|\eta| = 1.5$, where there is an overlap between BPIX and FPIX services. This effect is intended to be mitigated in the pixel Phase I Upgrade, when part of the BPIX services will be moved further out in the longitudinal direction, outside of the active tracking volume (see section 3.4).
The resolution in the measurement of the impact parameter ranges between 30 μm and 80 μm in the central region, depending on the transverse momentum, while it gets worse at higher $|\eta|$ because of the material that the particle needs to cross. A modulation of the resolution in ϕ, due to the periodicity of the mechanical structure, has also been observed. More details about these measurements and the overall performance of the pixel detector can be found in [50, 70].

3.4 The upgraded pixel detector

The need for a complete replacement of the CMS pixel detector arises from considerations based on the current performance of the LHC and on the future schedule of the whole accelerator complex. After the shutdown period LS1 (during the years 2013-2014), the center of mass energy has been increased to 13 TeV. If the performance of the machine follows the expectations, it is likely that the original goal of delivering an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with a 25 ns bunch spacing, will be achieved relatively soon. This will result in an average of about 25

inelastic interactions for each bunch crossing (pile-up) and corresponds to the scenario for which the CMS experiment, and all its subdetectors, has been designed for. If the behaviour of the LHC meets these predictions, the peak luminosity is therefore expected to exceed the nominal value by a factor of two before the next long shutdown (LS2, starting approximately at the beginning of 2019). Being the innermost part of the experiment, these running conditions will put the performance of the current pixel system under pressure. An inefficency of about 4% is expected in the first layer already at the design luminosity and is mainly caused by losses associated with the limited buffer capacity of the readout mechanism. Under the same assumptions in the Level-1 trigger accept rate, the loss in the innermost layer increases to 16% simply scaling the instantaneous luminosity by a factor two (for a 25 ns bunch crossing scenario) and becomes even worse with a 50 ns bunch spacing. Of course, these losses in the readout have direct impact on the physics performance of the tracking system, reducing substantially the tracking efficiency and increasing the probability of wrongly reconstructing tracks (fake rate), as can be seen from the plots in figure 3.13.

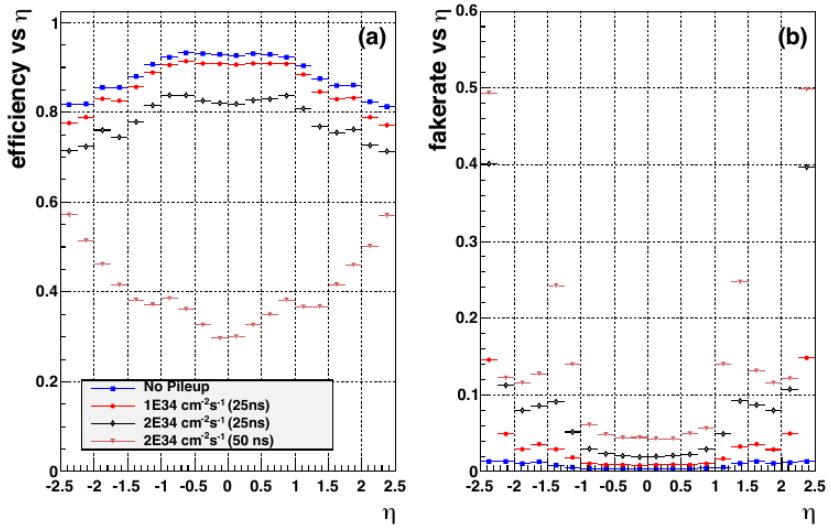


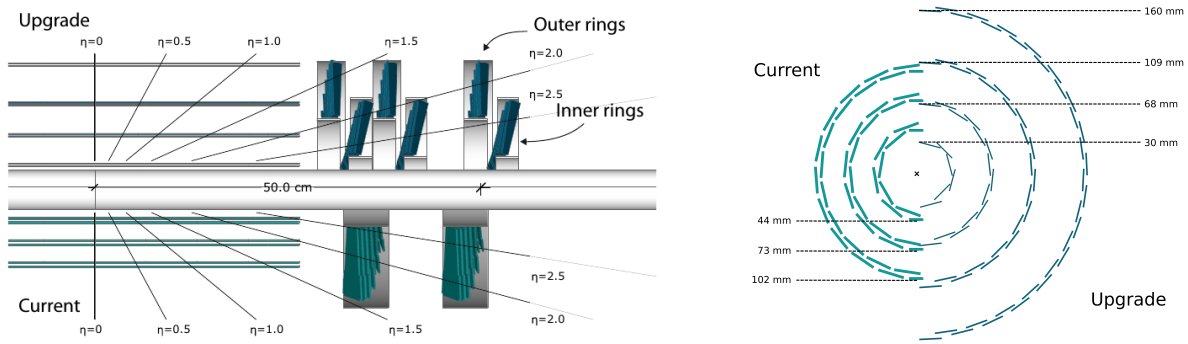
Figure 3.13: Performance of the current pixel detector in simulated $t\bar{t}$ events: efficiency (a) and fake rate (b). The four scenarios correspond to the ideal case of no pile-up, to the design condition of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ instantaneous luminosity with 25 ns bunch spacing and to the running conditions at double the luminosity and 25 ns or 50 ns bunch spacing. The progressive degradation of the performance, especially for the last scenario, is clearly visible [74].

A replacement of the whole pixel detector is mandatory and planned, as part of the "CMS Phase I Upgrade" program, during the 2016/2017 extended year-end stop of the LHC. The strategy of decoupling the pixel upgrade from the long shutdown periods will allow CMS to fully exploit the increase in performance of the LHC and to make full use of the integrated luminosity delivered to the experiments before LS2 [74].

3.4.1 Changes compared to the current system

The upgraded pixel system is designed to sustain an instantaneous luminosity up to $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and an average number of pile-up interactions exceeding 50, while maintaining the high tracking performance of the current detector.
The upgrade program will have an impact on all the subsystems of the detector, spanning from mechanics and service equipment to the readout electronics. The barrel structure will be ex-

tended with the addition of a fourth layer at a radius of 160 mm from the beam line, while the endcap regions will be redesigned to accommodate three disks for each z-end, instead of the current two. They will be placed at the coordinates $z = \pm 291$ mm, $z = \pm 396$ mm and $z = \pm 516$ mm and each disk will have a radial coverage ranging from 45 mm to 161 mm. This modification in the geometry will allow to have full four-hit coverage up to an $|\eta|$ range of 2.5, improving the track reconstruction and adding some redundancy to the system (see figure 3.14(a)).
Thanks to a reduction in the CMS central beam pipe diameter[6], the innermost barrel layer will be moved closer to the interaction region, at an average radius of 30 mm instead of the current 44 mm (figure 3.14(b)). This modification will have a beneficial effect on the impact parameter resolution, which is a key ingredient for vertexing and for the identification of long-living particles in high pile-up conditions. From the geometrical point of view, each one of the disks of the forward pixel will be divided into an inner and an outer ring, which will be equipped with blades supporting the radially arranged modules. Like the current FPIX, the blades are rotated by 20 ° in a turbine-like geometry designed to enhance charge sharing among adjacent pixels. In addition, the inner rings will have a further 12 ° tilt towards the interaction point. The overall layout of the barrel system is basically unchanged, although the division in half-barrels has been redesigned in order to avoid the need of special half modules for the boundary regions. The new fourth layer at a larger radius will comprise a total of 512 modules while, given the smaller distance from the beam pipe, the number of modules needed for the inner layer will be reduced down to 96. The second and the third layer will require roughly the same number of modules of the current BPIX (224 and 352, respectively).



(a) (b)

Figure 3.14: Layout difference between the current pixel detector and the upgraded version. Left: longitudinal view. Right: radial cut of the BPIX system. Note the closer distance of the innermost layer with respect to the interaction point and the absence of half modules in the upgraded system [74].

Overall, the total number of pixels will increase up to around 79 million. Despite the higher number of components and the addition of a completely new layer, the material budget of the future pixel detector will be lower than the current one. This reduction will be achieved thanks to a light-weight mechanical support made of Airex foam and carbon fibre sheets and a redesign of the cooling system, which will move from the C_6F_{14} liquid cooling of the present detector to a two-phase CO_2 cooling. A relocation of much of the passive material (e.g., electronic boards and connectors) outside of the η range of the tracking volume will also be decisive in reducing the material budget inside the active region (see figure 3.15).

The data loss in the readout chip due to limited buffering in high luminosity running will be

[6]The new beam pipe has already been installed in the Summer of 2014, as a part of the interventions on CMS scheduled during the first long shutdown of the LHC (LS1).

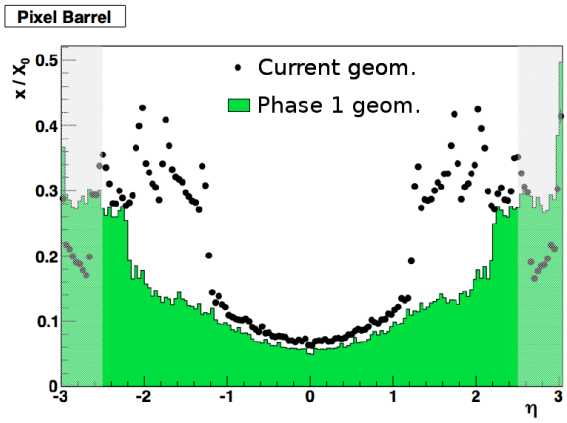


Figure 3.15: Amount of material in the pixel active volume in units of radiation length X_0 as a function of the pseudorapidity η, for the current detector (black points) and for the upgraded version (solid histogram). Adapted from [74].

cured by a new version of the chip, named PSI46digV2.1respin [75]. A picture of the new readout chip is shown in figure 3.16.
The readout mechanism is still organized in double columns of 2 × 80 pixel unit cells each, but the time stamp and data buffers are extended: the upgraded ROC has 24 time stamp buffers and 80 data buffers for each double column (they are respectively 12 and 32 in the current PSI46V2 chip version). Moreover, the pulse height encoding is moved from analog to digital, thanks to an 8-bit on-chip ADC. The advantages of the increased buffers are completed by a new Token Bit Manager design, which ensures a readout speed of 160 Mbit/s, faster than the current 40 MHz analog coding.

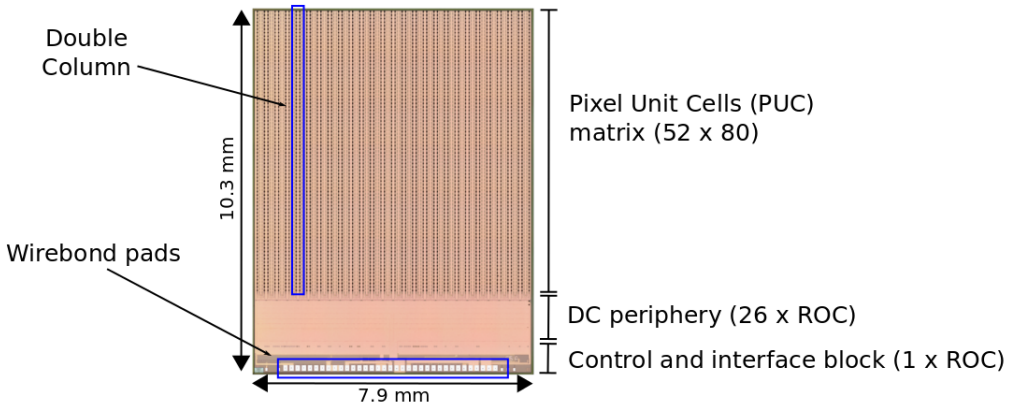


Figure 3.16: PSI46digV2.1respin readout chip for the CMS pixel upgrade. The electrical connection of each chip to the rest of the module is achieved via wire bonds connected to the 35 aluminium pads in the lower part of the ROC.

These modifications will allow the maximum rate that the ROC can sustain without becoming inefficient to increase from the current 100 MHz/cm^2 to 250 MHz/cm^2. The buffer extension has also the geometrical consequence that the size of the readout chip periphery increases, causing the new digital ROC to be roughly 500 μm longer than the current analog version.
The design of the new BPIX modules resembles very closely the current one, as can be seen in figure 3.17. The power and signal cables are replaced with one single twisted-pair cable for signal transmission, power and detector bias voltage. Due to the geometrical increase of the ROC size, the new modules are expected to be slightly wider compared to the ones currently installed in CMS and will measure 66.6 × 27.0 mm^2, including the base strips.

Finally, the upgrade will also bring a reduction in the number of different module types and interfaces, which is particularly relevant for the forward detector. In the FPIX, only one module

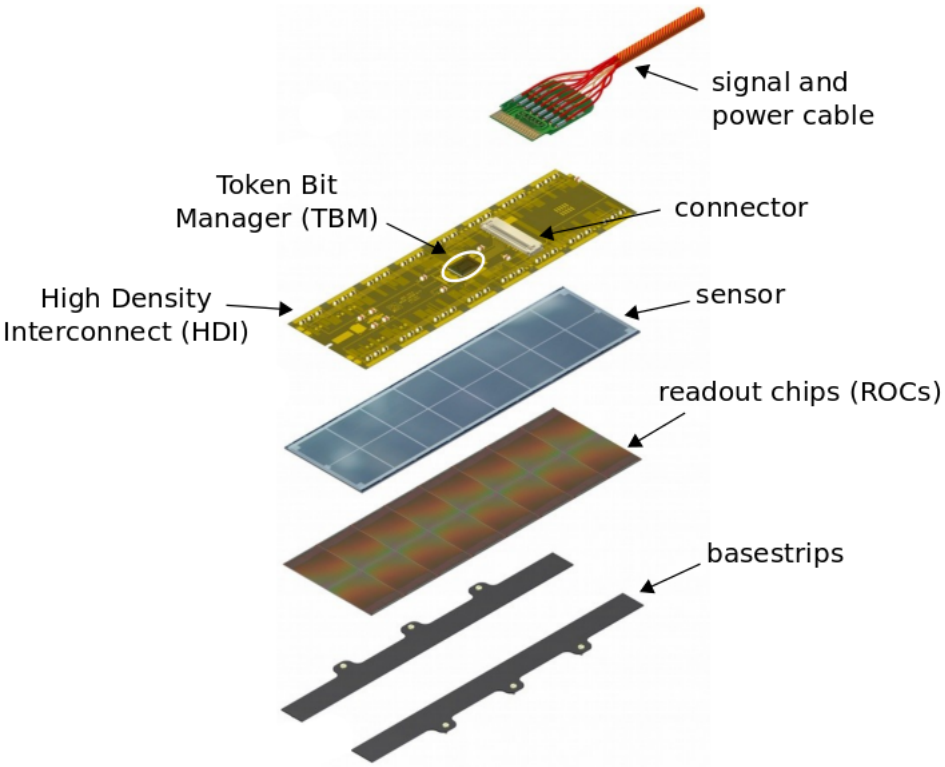


Figure 3.17: Exploded view of an upgraded BPIX module. The design in the picture is implemented for the modules to be installed in the second to fourth layer. The modules for the innermost layer need to satisfy more stringent requirements, both in terms of space and performance, and have therefore a slightly different design, which will not be detailed here [74].

with 2×8 ROCs as in the barrel is going to be used, and it will have a radial orientation in order to improve the r-ϕ resolution (the current detector has $100 \times 150\,\mu\text{m}^2$ pixels oriented at $90\,^\circ$ compared to the upgraded one). This change in the layout is intended to allow a simplification of the production and of the maintenance of the system.
Despite the many modifications briefly described in the paragraphs above, it is also obvious that the new pixel system will have to profit as much as possible from the already existing surrounding infrastructures, since a complete replacement of all the cable plant which serves the current detector is out of question, because of cost and time constraints. On the other hand, due to the increase in the number of readout channels, the front-end power consumption will increase by a factor close to two, while the power loss on the cables increases by a factor of four. A new powering scheme has therefore been developed, which is based on providing a higher voltage on the existing cables (therefore reducing the power losses on the cable itself) and then stepping it down to the 1.6 V and the 2.2 V needed by the digital readout chips, thanks to custom DC-DC converters placed on the service infrastructure on the detector [76].
For what concerns the project organisation, the forward part of the future detector will be built in USA, like the current one, while the production of the new barrel will be shared among four Europe-based consortia, one of which involves also Taiwan. The first and second layer will be produced by a Swiss consortium composed of PSI[7], ETH[8] Zurich and University of Zurich. The third layer will be assembled by CERN and other institutes and universities located in Italy, Taiwan and Finland. The construction of the fourth barrel layer falls under the respon-

[7]Paul Scherrer Institute.
[8]Eidgenössische Technische Hochschule.

sibility of a German consortium composed of DESY, the University of Hamburg (UHH), the Rheinisch-Westfälische Technische Hochschule (RWTH) in Aachen and the Karlsruhe Institute of Technology (KIT). Half of the modules for the layer 4 quota are physically assembled at the Karlsruhe Institute of Technology and then tested and calibrated at RWTH Aachen, while DESY and University of Hamburg share production and testing of the second half.

Chapter 4

Production of pixel modules for the CMS pixel Phase I Upgrade at KIT

This chapter will give an overview of the Karlsruhe Institute of Technology assembly line for the production and testing of pixel modules for the CMS Phase I Upgrade. The first part of the chapter describes the construction of bare modules made of a silicon sensor bump bonded to the readout chips in detail. The following section gives an overview of the testing and qualification procedures needed to assess the quality of the produced bare modules. The final part briefly describes the steps which are needed to complete the bare modules into fully operational modules. The chapter is concluded by an overview of the KIT module production campaign, including production yields and statistics.
Producing and qualifying detector modules for a particle physics experiment is a complex task which involves interactions with a number of external vendors and companies. The complete KIT production workflow chain, from single components to the final, fully equipped, barrel modules, is outlined in the diagram of figure 4.1. Two KIT institutes are actively involved in the project: the "Institut für Experimentelle Kernphysik" (IEKP) and the "Institut für Prozessdatenverarbeitung und Elektronik" (IPE).

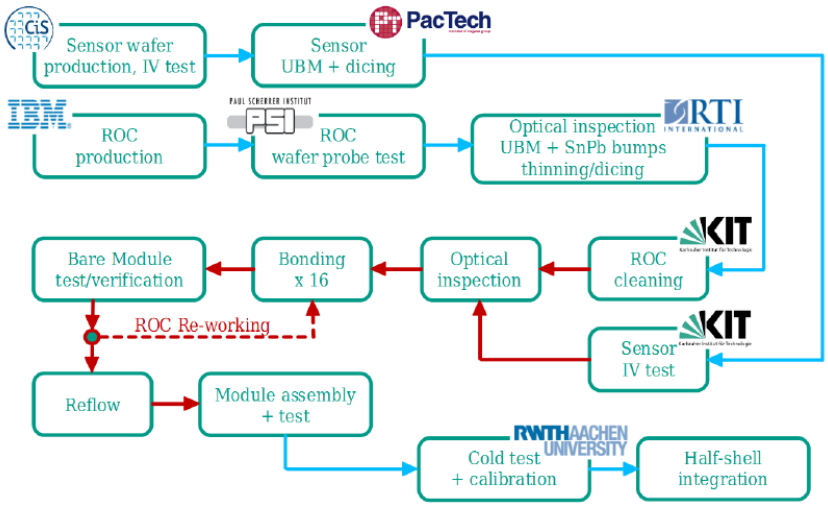


Figure 4.1: The diagram shows the workflow for the production of pixel barrel modules, for the quota assigned to KIT. Next to each block, the name of the responsible institute or external company is given. The operations contained in the blocks connected by the red arrows take place at KIT.

4.1 Production of pixel bare modules

This section describes the assembly line for the bare module production, where the readout chips are connected to the sensor using the bump bonding technology.
While handling such sensitive devices the amount of external contaminations needs to be kept as low as possible. All the processes that are described in the following, including material inspection, cleaning of the components, the actual flip-chip assembly and the reflow of the bonded bare modules are therefore performed inside a $4 \times 4\,\mathrm{m}^2$ class 1000 clean room[1] located inside the IPE facilities of KIT, Campus North.

4.1.1 Sensors and readout chips

The silicon sensors for the whole CMS production are centrally manufactured in Erfurt, Germany, by CiS Forschungsinstitut für Mikrosensorik GmbH, on 4" silicon wafers. The wafers for the KIT quota are then sent to a company called PacTech - Packaging Technologies GmbH based in Nauen, Germany, for further processing. Here, an electroless process deposits a UBM on the $30\,\mu$m wide circular aluminium pads of each pixel cell and on the contact pads located in the sensor periphery for sensor biasing and testing. The UBM is made of a thin 200 nm layer of palladium, a $5\,\mu$m nickel layer and a very thin 50 nm surface layer of gold[2]. On each wafer, different structures are manufactured, including three sensors having the size of a bare module and ten single chip sensors. These components are cut out from the full wafer and shipped to KIT. The module-sized sensors are used for the actual pixel module production, while the single chip sensors are mainly used for test-assemblies, necessary for tuning the processes (especially in the first pre-production phase) and for monitoring the quality of the assembly chain while the module production is ongoing. The single chip sensors are also useful as R&D material and for the production of small samples used for irradiation studies. Figure 4.2 shows a magnified picture of a detail of the sensor pixel array, after the UBM deposition.
Since the mechanical stress due to the UBM deposition processes or to the dicing can damage the sensor, each die is re-tested at KIT after delivery: the sensors which are classified as good are bonded, while sensors having a lower quality are discarded or used for mechanical test assemblies. The criteria for judging the quality of a sensor are based on the behaviour of the I-V curve, which is obtained measuring the leakage current as a function of the reverse bias voltage applied to the sensor. A sensor is considered "good" if the leakage current at 150 V is lower than $2\,\mu$A and if the ratio between the leakage current at 150 V and the one at 100 V is less than 2. The rationale behind these working points is that the sensors used for CMS modules must be able to maintain reasonably low leakage currents also when an increase of the applied bias voltage (initially around 50-60 V) will be required for counteracting the effects of radiation damage [60].

The digital PSI46digV2.1respin readout chips are centrally produced at IBM on 8" wafers using a commercial $0.25\,\mu$m CMOS process. The full wafers, each one containing 244 readout chips, are electrically tested at the Paul Scherrer Institute (PSI) in Villigen, Switzerland, and subsequently distributed to all the production consortia. The KIT lot is directly shipped to an external

[1] Clean room classification is based on the number and size of particles allowed in a certain volume of air. According to the US FED-STD-209E standard, a clean room of class N has at maximum N particles of size $0.5\,\mu$m or larger per cubic foot of air. A class 1000 clean room in the above standard corresponds roughly to a class ISO 6 clean room in the more recent ISO 14644-1 standard.

[2] The technical name of this particular UBM deposition process is ENEPIG, which stands for "Electroless Nickel, Electroless Palladium and Immersion Gold".

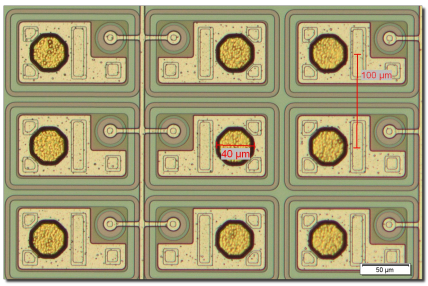


Figure 4.2: Front side of the pixelated silicon sensor before bonding. The topmost layer of the under bump metallization can be clearly recognized as a 40 μm wide golden pad at the center of each pixel cell.

company called RTI International, based in the USA, where spherical eutectic SnPb bumps having a diameter of approximately 30 μm are placed on the wafer. The bump deposition is based on electroplating and follows different steps, which are briefly summarized in the following points and illustrated in figure 4.3 [77].

1. The original passivation layer is removed and the wafer is re-passivated with a polymer (benzocyclobutene, BCB) that protects the surface during subsequent processing, providing a uniform and planar base and acting as a stress buffer between the wafer substrate and the bumps. Holes corresponding to the I/O metal pads which need to be bumped are created in the passivation layer. The wire bond pads on the chip periphery are kept closed until the end of the wafer processing.

2. UBM layers are deposited on the wafer via electroless plating. The UBM consists, moving from the metal pad of the chip towards the surface, of a thin TiW layer which acts as a diffusion barrier, a few μm thick nickel layer as a wettable metal and a thin surface gold layer which prevents nickel oxidation. The thickness of the last layer is kept as low as possible, in order to avoid possible problems given by gold embedding into the actual soldering material.

3. After UBM deposition, the wafer is covered with a thick photoresist layer which is then patterned in a photolithographic process to create small openings above the UBM pads.

4. The bump material is deposited on the wafer via electroplating, controlling the amount of deposited material (and hence the final bump size) varying the electroplating current. The process is stable enough to ensure a high bump uniformity over the whole wafer.

5. After the electroplating process, the photoresist layer is chemically removed in a solvent cleaning step and the parts of the UBM which are not covered by the bump are etched.

6. The SnPb cylinders undergo a reflow procedure and are reshaped into truncated spheres having a diameter of 30 μm. This reflow step helps in creating a solid intermetallic connection between the solder material and the UBM.

7. Finally, the aluminium wire bond pads, which are later needed to connect the readout chips to the HDI during the assembly of a full module, are made accessible by opening the BCB passivation layer above them in a dedicated dry etching process.

After bumping, a thick photoresist layer is applied on top of the ROC wafer in order to protect the bumps during the subsequent processing steps and during transport, as well as to ease chip handling and manipulation at KIT. In order to reduce the overall material budget of the final assembly, the ROC wafer is thinned down to a thickness of approximately 175 μm, removing unneccessary material from the back side. The single chips, having a dimension of 7.9 x 10.3 mm^2 each, are finally diced out from the wafer and placed in carriers specially designed for handling

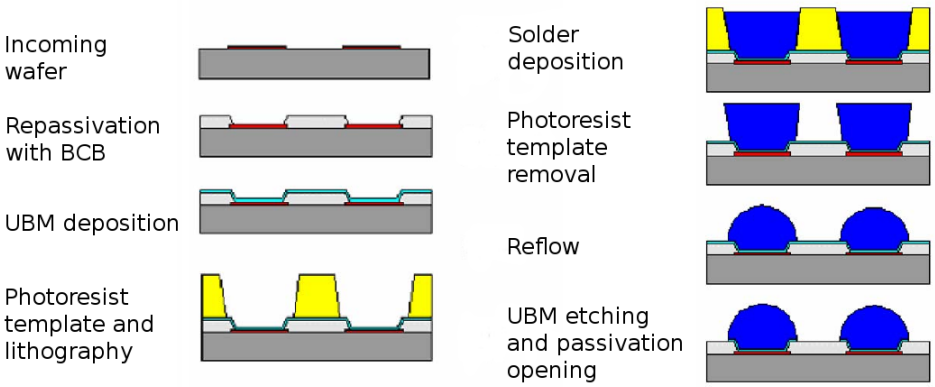


Figure 4.3: Cross-sectional view of the RTI bumping process. Adapted from [77].

sensitive devices, colloquially referred to as gel-packs from the name of the company (Gel-Pak®) which produces them. Before shipping the gel-packs to KIT, all the bumped chips are preliminarily inspected, looking for major defects such as wrong dicing, residual of seed metals, enlarged, miss-shaped or missing bumps. A more detailed inspection is carried out at KIT before the flip-chip assembly, as described in the next paragraph.

4.1.2 Cleaning of components and optical inspection

Despite the clean operating environment, it is sometimes the case that the components used for the module production suffer from pollution coming from sub-optimal transport conditions, from the tools used for manipulating them or from the operators themselves. In this case, a cleaning procedure become necessary. Different types of contamination are best removed by different solvents. For the removal of inorganic contaminants, especially salt or ionic compounds, pure water (flushed or in a bath) is usually the best choice. This can be combined - if required and if allowed - with some additional mechanical action, for example in an ultrasonic bath, where a generator working at frequencies between 20 and 400 kHz creates compression waves in a solvent, whose very localized pressure helps in removing surface contaminants. For organic pollution, acetone is usually used. Isopropyl alcohol (IUPAC name, 2-propanol) is also a standard cleaning fluid, largely employed in the microelectronic industry thanks to the ability of dissolving a wide range of non-polar compounds.
In the KIT production line, cleaning is a mandatory procedure for removing the protective photoresist layer from the readout chips received from RTI after the bump deposition. The cleaning is performed following a procedure tuned in order to achieve the best result in terms of amount of residuals on the chip surface, while reducing the direct manipulation of the delicate bumped chips as much as possible.
First of all, the single ROCs are removed from the delivered gel-packs using a vacuum tool. Thanks to the protective layer, manipulation on the bumped side is, at this point, safe. Handling the chips with other instruments such as pliers or tweezers needs to be avoided, since these operations are more likely to cause fractures in the 175 μm thick silicon or to destroy the bumps, especially near the chip periphery. The ROCs are then placed on a custom-designed tray made of a type of thermoplastic material named Polyoxymethylene (POM)[3], which has small conductive particles embedded into the polymer structure in order to make it safe against electrostatical discharges (ESD-safe). This material has been chosen looking for a compromise

[3]The commercial name of this thermoplastic material is POM-ELS.

between softness, resistance to the solvents used in the cleaning sequence, ESD-protection and good mechanical properties. Especially the last criterion is mandatory in order to be able to mill the material with standard industrial CNC machines. Small cavities with a depth of approximately 300 μm and the geometrical dimensions of a readout chip (plus 150 μm in each direction) are carved in the tray and can accommodate one ROC each, for a total of 35. The idea is to use the tray to keep the readout chips in position during the cleaning procedure and to move them around without touching the delicate solder bumps. A teflon grid fixed on an aluminium frame is mounted on top of the tray to avoid chip flotation during the cleaning steps.
The readout chips positioned on the tray are cleaned with a sequence of three different solvent baths. The first bath consists of ultrapure ($> 99\%$) acetone, which quickly dissolves the photoresist layer and other possible organic contaminants. Since acetone is a non polar compound, it is not possible to mix it with water and an intermediate bath with ultrapure ($> 99\%$) isopropyl alcohol is therefore needed for washing away the acetone residuals. The transition between two wet environments is mandatory in order to avoid that fast evaporation leaves stains and residuals on the surface of the chips. The cleaning process is completed by a sequence of two baths in deionized and filtered water. To ease the evaporation of the water from the readout chips, the tray is finally placed inside a vacuum chamber heated to 70 °C, until it is completely dry. The whole cleaning sequence lasts approximately 30 minutes, with roughly half of the overall time taken by the final drying inside the vacuum chamber. More details about the cleaning procedure, including mechanical drawings of the tools used, can be found in [66].
While finalizing the above mentioned procedure, the quality of the results was carefully checked inspecting a cleaned chip under a scanning electron microscope (SEM). An exemplary result, shown in the picture of figure 4.4, proves that after cleaning only tiny darker residual spots are visible on the bump surface. Due to their far too small size, their influence on the bond strength or on the electrical connection between the readout chip and the sensor is negligible.

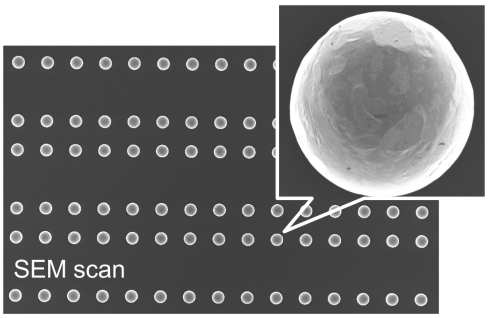


Figure 4.4: SEM image of double columns of bumps on a readout chip and a detailed view of a single bump, after the cleaning process (images: A. Jung, ITEP-KIT).

Before proceeding with the bonding, the tray containing the cleaned chips is placed under an optical microscope and each ROC is inspected looking for possible defects which might reduce the quality of the later assemblies. Apart from the overall quality of the cleaning and the presence of residuals on the chip surface, it is important to assess also the goodness of the chip dicing and to check for the presence of miss-shaped bumps (due to imperfect solder deposition) or for missing bumps. The optical microscope located in the production clean room is equipped with a movable motorized table and runs a software which allows to collect a sequence of magnified pictures for each ROC and to stitch them together to form a single, high-resolution, image of each chip. These images are stored on disk for future reference and fed to a custom-made pattern recognition software which, looking at each picture, automatically searches for the presence of any of the defects mentioned above, presenting them to the operator who decides if the quality of the readout chip is sufficient for using it for a module assembly.
The efficiency of this cleaning and inspection phase has been found to be around 90%, with the main source of defects being the presence of environmental contamination, silicon splinters or cleaning residuals on the chip surface, or defects in the BCB layer. During the whole production

campaign, the number of missing bumps for each chip has been observed to be very low, also after the cleaning phase, and only in 1% of the cases more than 5 missing bumps (corresponding to 0.12% of the total bumps on a chip) have been found on individual ROCs.
The POM tray has been designed such that, after the cleaning and the optical inspection, a vacuum plate can be mounted on the back side. Once the vacuum is applied, the cleaned chips remain well in position inside the carved cavities and the whole tray can be flipped on top of a gel-pack having the same size as the tray itself. When the vacuum is switched off, the chips are released on the soft surface of the gel-pack with the bumped area facing down, arranged in a regular 7×5 matrix and ready for being picked up by the bonding machine.
A simpler optical inspection is performed also on the sensors, looking for possible damages such as scratches on the guard ring structure, defects in the UBM deposition or in the passivation layer.

4.1.3 Flip-chip bonding and reflow

While the basic general principles of thermocompression bonding have been discussed in section 3.2, this section will describe the details of the bonding procedure developed for the pixel module assembly line at KIT and the optimization of the bonding and reflow parameters.

4.1.3.1 Bonding

Flip-chipping and bonding of the CMS readout chips onto the sensor substrate is performed inside the class 1000 clean room using a "FINEPLACER® femto" bonding machine manifactured by Finetech GmbH in Berlin, Germany [78]. An overview of the machine is given in figure 4.5.

The central part of the machine consists of a motorized *bonding table*, movable in the (x, y, ϕ) directions and equipped with a high resolution drive for precise movements along the z-axis. On the left and right side, two presentation plates are mechanically connected to the table and host the gel-pack where the cleaned components to be picked up and bonded are placed. In the central part of the bonding table, where the actual bonding takes place, a strong vacuum circuit, created by a Venturi system, keeps the substrate in position during the process. The bonding area of the table, which has a surface of approximately $100 \times 100\,\mathrm{mm}^2$, can be heated up to a temperature of 400 °C thanks to an external heating module and cooled down by a custom-designed system working with compressed air. A *placement arm* is located above the table, perpendicular to its surface, and is joined to a pivot point having a single rotatory degree of freedom around the x-axis. A direct current motor moves the arm in a 90 ° swing movement, while an inclination sensor monitors the arm position. At the far end of the placement arm, a mechanical interface holds the *bond head*, which consists of a custom-made heatable rectangular tool designed to have exactly the geometrical dimensions of a CMS PSI46digV2.1respin readout chip and equipped with a vacuum circuit for holding the components. The inner structure of the tool is essentially a movable mechanical joint which is able to adapt itself to the surface underneath and then maintain the acquired orientation with a high degree of precision during the whole bonding process. A *movable camera* runs along the x coordinate of the table and is equipped with a 60 mm long light splitter which allows the simultaneous visualization of the substrate lying on the table and of the component held by the bond head. The usage of a light splitter for visualization and alignment of the components is typical of these type of bonding machines and crucial for the quality of the process. On old bonding machines, the alignment was usually performed manually by the operator, while on more recent machines, such as the one in use, the alignment is done by sophisticated pattern-recognition software. The bonder

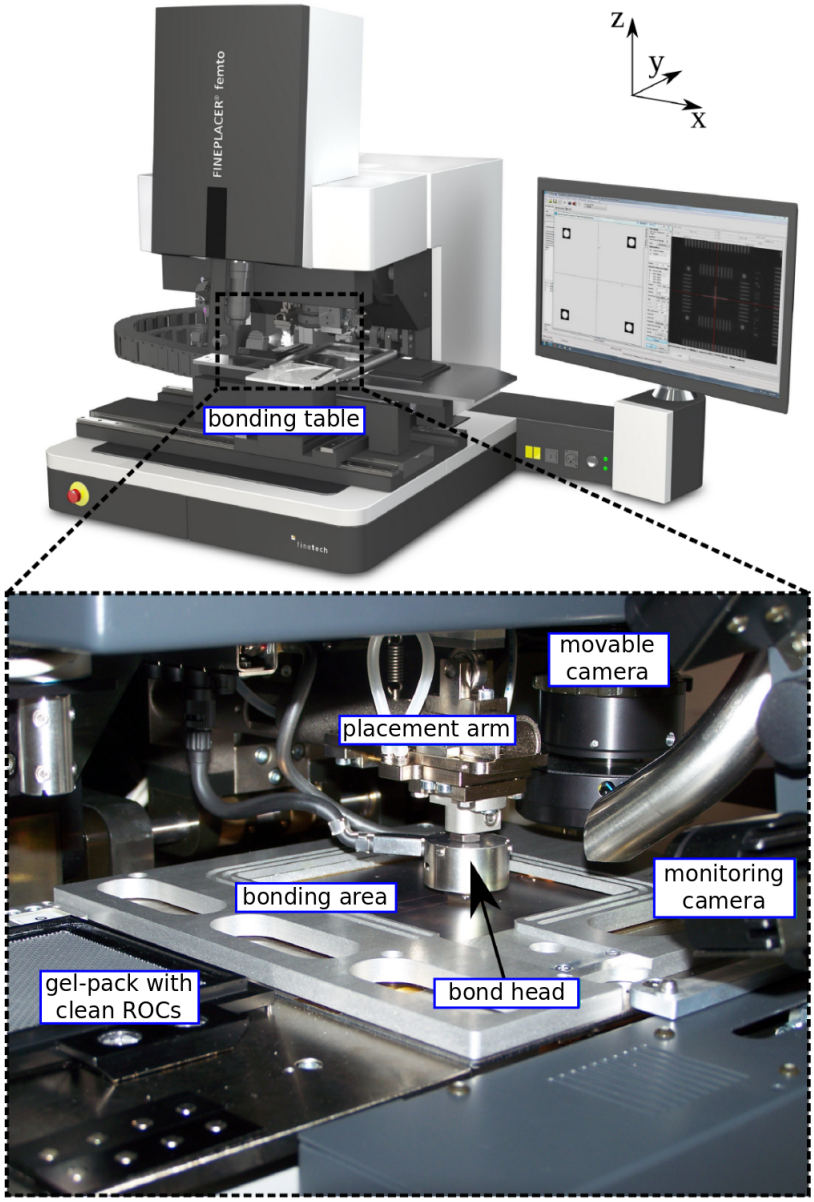


Figure 4.5: General overview of the Finetech "FINEPLACER® femto" bonding machine and a closer view of the bonding table, with the placement arm in the touch-down position at 0 ° with respect to the table surface [78].

is completed by other extensions such as an antivibrational table, a computer to program and control the process and one additional camera placed in front of the table, which allows the user to monitor the bonding sequence. Additionally, the bonding area can be completely covered with a metallic plate having a circular hole with the size of the bond head in the center. This means that when the placement arm is in the horizontal position, at 0 ° with respect to the table surface, a closed chamber is created around the assembly. This allows to apply an in-situ reflow procedure to the bonded components, thanks to connections between the machine and additional gas modules (nitrogen and formic acid). This possibility was largely exploited during the first development and testing phase to tune a suitable reflow profile, while for the actual production of pixel modules the usage of an external reflow oven was preferred due to the higher capacity in terms of modules which can be accommodated in the chamber for each reflow and

to a better atmosphere control during the reflow process.
The bonding force which can be applied by the placement arm is user-defined and ranges from 0.5 N to 500 N. According to the technical specifications, the placement accuracy of the machine is around $\pm 0.5\,\mu$m, while the planarity resolution is below $4\,\mu$rad.
The procedure developed at KIT for flip-chip bonding the CMS pixel assemblies (single chips or full bare modules) consists of a sequence of steps, which are outlined below.

1. A sensor substrate, having the size of a full CMS module or of a single chip, is placed at the center of the bonding table, with the UBM side facing up, and is held in position by the built-in vacuum system.

2. The machine recognizes the two lower corners of the sensor substrate and memorizes their relative coordinates with respect to the internal reference system.

3. The gel-pack containing the cleaned and selected readout chips is placed on one of the side plates of the table. The table then moves such that the readout chip to be picked up lies exactly under the placement arm. Since the back side of the chips consists of a flat silicon surface, without any visible structure, this alignment procedure can rely only on the ROC corners: this is why the dicing of the chips itself has a non-negligible impact on the goodness of the bonding process and the cutting quality needs to be carefully checked while inspecting the chips in the pre-bonding activities described in section 4.1.2. After the alignment, the arm goes in the touch-down position at $0\,^\circ$ with respect to the table surface and picks up one of the ROCs, holding it in position thanks to the vacuum system.

4. Since the soft surface of the gel-pack does not guarantee the flatness of the bond head during the pick-up procedure mentioned above, the chip is lowered again on a very flat surface made of a ceramic material and then picked up a second time by the placement arm, applying an extra force of 4 N. This ensures that the movable tool of the bond head can adapt itself to the chip surface. An optimization procedure has been performed to obtain good parallelism between the ceramic surface and the bonding table.

5. The machine performs a pattern recognition procedure, using some unique structures located on the lowermost row on the sensor and on the ROC sides as a reference. The bonding table position is recursively adjusted until the user-defined accuracy of the alignment between the above defined reference points is successfully reached.

6. The arm lowers onto the substrate for the actual bonding process. When reaching the contact point, the table moves upwards, applying an user-defined force profile which is monitored by a force sensor located on the placement arm. At the same time, the bonding tool is heated up to a target temperature of 140 °C. The temperature of the bonding table is kept fixed to the same value during the whole bonding sequence, in order to avoid thermal stress on the assembly. In figure 4.6, the temperature and force profiles of the bonding process are shown as a function of time. The force profile has a slow build-up and a characteristic "step-like" behaviour: the first step at 30 N helps in estabilishing good planarity between the chip and the surface before applying a full bonding force of 100 N (corresponding to 24 mN per bump). The bonding of one single chip lasts 85 seconds.

7. The placement arm goes back to the the initial position, at $90\,^\circ$ with respect to the table surface, and the bond head is cooled down to 65 °C.

8. In case of bonding a full module, the whole picking, aligning and placing procedure is repeated until all the 16 ROCs are connected to the substrate[4].

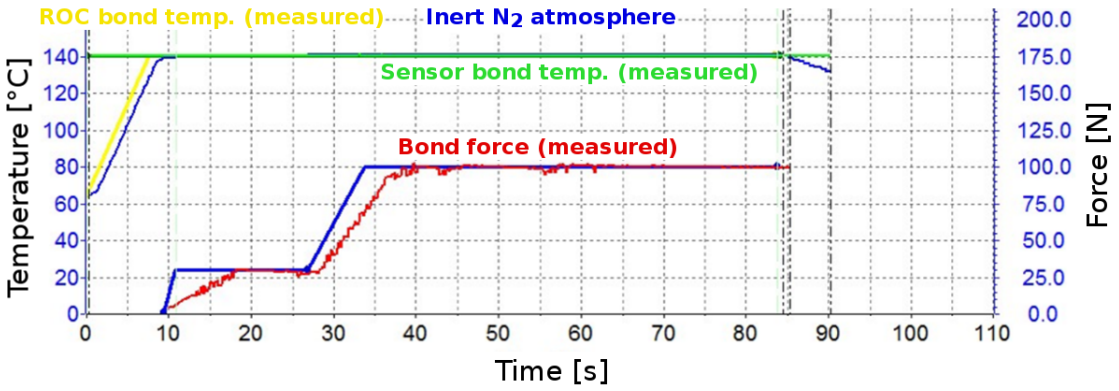


Figure 4.6: Force and temperature profiles used for the bump bonding process at KIT. The thick blue and the red lines show, respectively, the set and measured force. The blue and yellow lines represent the set and measured temperature on the bond head, while the overlapping dark green and light green lines are the set and measured temperature on the bonding table.

During all the steps described above, the chamber where the bonding takes place is continuously flushed with N_2 ($5\,\ell$/min) to avoid metal oxidation at high temperatures.
The complete sequence, including the pick-up of the chips, the flattening procedure on the ceramic plate, the actual bonding and the final cooling of the bond head lasts approximately 4 minutes for each readout chip. This means that a complete bare module consisting of 16 ROCs is normally bonded (excluding the final reflow) in approximately one hour.

4.1.3.2 Reflow

For the sake of clarity, the reflow procedure on the CMS pixel modules is described at this point of the discussion, following the same approach of section 3.2. In the chronological order of the module production workflow, however, the reflow of the bare modules is executed only *after* electrically testing them in a custom-designed probe station (see section 4.2). The decision of arranging the working sequence in this way was taken based on the fact that, in case of issues emerging from the bare module test, it is much easier to "rework" a problematic module (i.e., to un-mount an electrically bad ROC and to bond another chip in its place) if the mechanical connection with the substrate has not been improved yet with a reflow. In fact, at this point, a successful rework procedure can be performed by simply placing the module on the bonding table, heated at a temperature around 130 °C, and removing the problematic ROC by hand or using the vacuum of the bond head. Moreover, it was observed that the mechanical strength at room temperature of a module without reflow is still sufficient for standing the mechanical stress of the probing without suffering from damage.
At first, the reflow of the bonded assemblies was performed on the FINEPLACER® machine itself, immediately after the bonding sequence and on one module at a time. This approach was particularly attractive due to the high flexibiliy of the machine in switching between different

[4]For technical reasons related to the pick-up procedure, only the eight ROCs of the lowermost row of the module are bonded in sequence. The sensor is then manually turned by 180 ° and the second row is completed in a second sequence, positioning the remaining eight chips.

reflow atmospheres (nitrogen or nitrogen/formic acid mixtures) and because no further manipulation of the module was required until the end of the reflow process. The possibility to reflow one single module at a time, however, would have turned into an undesirable bottleneck in the production flow, since no further bonding activities would have been possible on the machine during the reflow of a module. Before the actual production phase, therefore, a new reflow oven was installed in the class 1000 clean room. This model of reflow oven (SRO-700, manufactured by ATV Technologie GmbH in Vaterstetten, Germany [79]) is equipped with a reflow chamber heatable up to 450 °C, one N_2/HCOOH mixture line, two N_2 gas lines (which are used also for cooling) and a pump mechanism which allows to perform reflows under vacuum (< 5 mbar). The inner area of the reflow chamber measures approximately $23 \times 22\,\mathrm{cm}^2$ and can therefore easily accommodate six or more bare modules at the same time, thus reducing the overall time needed for the reflow sequence.

4.1.4 Optimization of bonding and reflow parameters

Before fixing the final bonding profiles used for the production of CMS pixel modules which have been shown in figure 4.6, some optimizations have been performed on test materials, checking the results by means of mechanical and electrical tests, as well as by destructive and non-destructive inspections. This section presents the main results of these investigations.
Test material consisted usually of ROCs or sensors diced out from "dummy" wafers, i.e., plain silicon substrates without any real electronics on them. These have been bumped according to the offered processes by RTI as well as by other vendors, which were evaluated in the first R&D phase. When it was not possible to do otherwise, also real components have been used for tests and optimizations, choosing the ones with the worst electrical quality or with mechanical defects (surface scratches, large number of missing or deformed bumps, etc.) which made them not suitable for module production.

4.1.4.1 Assessment of the quality of the assemblies

A variety of well-defined procedures exists to test the quality of microelectronic devices and are usually encoded into widely used standards, to ease the comparison between different manufacturers[5]. One of the most immediate methods to check the connection strength between a semiconductor device and the substrate to which it is attached consists in applying an increasing controlled force perpendicular to the surface of the assembly, until the breaking point is reached[6] (pull-test). It is important for the force to be applied in the normal direction, since the bumps are less resistant to transverse stress and force components going in this direction would cause a "domino-like" breaking of the assembly, where only a fraction of the bumps is breaking at the same time. Together with the numerical value of the force required for the separation, also the breaking mechanism is an indication of the bonding quality. Apart from the obvious situation where no connection has been established at all and the two parts look like before the bonding after the separation, the breaking point might happen anywhere between the top metal layer (usually aluminium or compounds with aluminium like AlSiCu), the UBM metals and the solder. In general, from the point of view of bonding, the optimal breaking mechanism is the one which completely lifts off the top metal layer from the silicon underneath, since this

[5] The most used standard, called MIL-STD-883, developed by the Department of Defense of the United States of America, consists of a collection of test methods for microelectronic devices, grouped into mechanical, electrical and environmental tests [80].

[6] This corresponds to the prescriptions of method 2031.1 from MIL-STD-883.

indicates that both the connections between metallization and UBM and between UBM and solder are stronger. In a bonding configuration where bumps are deposited only on one side (for example on the ROC side), a breaking mechanism which leaves a bit of solder on both surfaces is also good, since this means that a strong intermetallic connection between the bump and the UBM on the opposite substrate has been achieved and it is stronger than the solder itself. From the technical point of view, all the pull-tests have been performed on a DAGE 4000 bondtester manufactured by Nordson [81] and equipped with a cartridge able to apply forces up to 10 kg[7]. Before the test, the assembly needs to be prepared by firmly connecting the two faces, respectively, to an aluminium plate and to an aluminium die by means of Araldite glue. The plate is then clamped to the base plate of the DAGE machine, while the die is screwed to the cartridge. The force is measured by the machine during the whole force build-up, until the breaking point is reached. The two parts can then be inspected with an optical microscope to investigate the breaking mechanism.
A destructive inspection of the inner structure of the single bumps can be obtained cutting the assembly across the bumped area (cross section). Since a regular dicing procedure would most likely destroy structures at the micrometer level, the sample (typically a single chip) needs to be embedded into an epoxy resin. After curing, the block of resin is grinded, removing material until the region of interest (for example a region close to the bumped area) is reached. At this point, the resin is polished with different abrasive papers, lowering gradually the grain size of the sheets up to a few μm and accompanying the mechanical action with water or with specific abrasive solutions. The polished resin block containing the sample can then be inspected with an optical microscope or subjected to an SEM scan.
X-ray scans are among the best methods for inspecting the assemblies in a non-destructive way: they allow to probe the alignment between the chip and the substrate without separating them and, if the resolution is high enough, even to check the density distribution of the material inside each single bump, looking for possible mechanical or structural anomalies.

4.1.4.2 Bonding temperature and force

The first tests to set the values of temperature and force in the bonding process have been performed on dummy samples equipped with daisy-chain structures. A daisy-chain consists of a series of short conductive lines running between two neighbouring bumps, on both the dies to be mounted. These lines are designed in such a way that only after bonding they will be connected to each other into a closed loop, usually ending with small testing pads placed in the periphery of the chip (see figure 4.7).

Test material with daisy-chains and eutectic SnPb bumps has been provided by VTT (Valtion Teknillinen Tutkimuskeskus), a Finnish research and technology center which was evaluated by KIT as a possible vendor for the bumping of sensors and readout chips. The VTT daisy-chains are designed to run into 26 different closed loops (one for every double column of the readout chip), with each loop ending into four conductive pads. On these contact points, a "four-point probing" technique can be applied: a small current is supplied via needles contacted with two of the pads and the voltage drop is read out from the other two pads. Ohm's law allows to extract the impedance of the daisy-chain, as $R = \mathrm{V/I}$. This rather simple measurement already allows to get important information on the quality of the assembly since, thanks to the conductive chains, not only the mechanical strength of the assembly can be tested, but also the goodness of the process in terms of electrical connection between the two dies: missing or defective bumps would result into a change of the electrical properties of the whole chain, which would manifest itself

[7]For these type of tests is customary to express forces in units of kilogram-force (1 kg $\equiv$ 9.81 N).

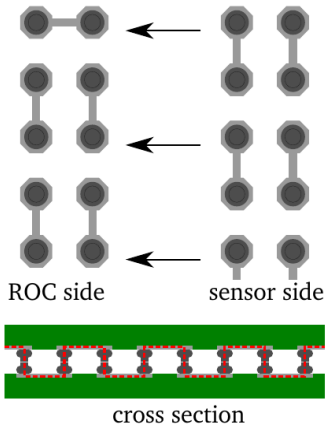


Figure 4.7: Intuitive illustration of the working principle of a daisy-chain: the partial connections between adjacent bumps on the sensor and ROC side form a closed loop after the mechanical connection of the two parts [66].

as higher measured values of the double column impedance ($R \sim \mathrm{k}\Omega$) or as open connections ($R = \infty$). As an example, two set of impedance measurements on two different ROCs are presented in figure 4.8. The left plot shows a ROC with no apparent electrical problems and a value of R which is basically constant across all the double columns; the right plot shows a readout chip which presents a broad region with high values of the impedance ($> 20\,\Omega$), one double column with $R > 1\,\mathrm{k}\Omega$ and one open chain. An inspection after the pull-test showed that the open chain was due to a missing bump on the ROC side, while the broad region with high resistance was traced back to a group of defective bumps. The chain with very high impedance was ascribed to an imperfect connection between two bumps of the chain, where a splinter on the ROC side made a punctual connection with the bump on the opposite surface.

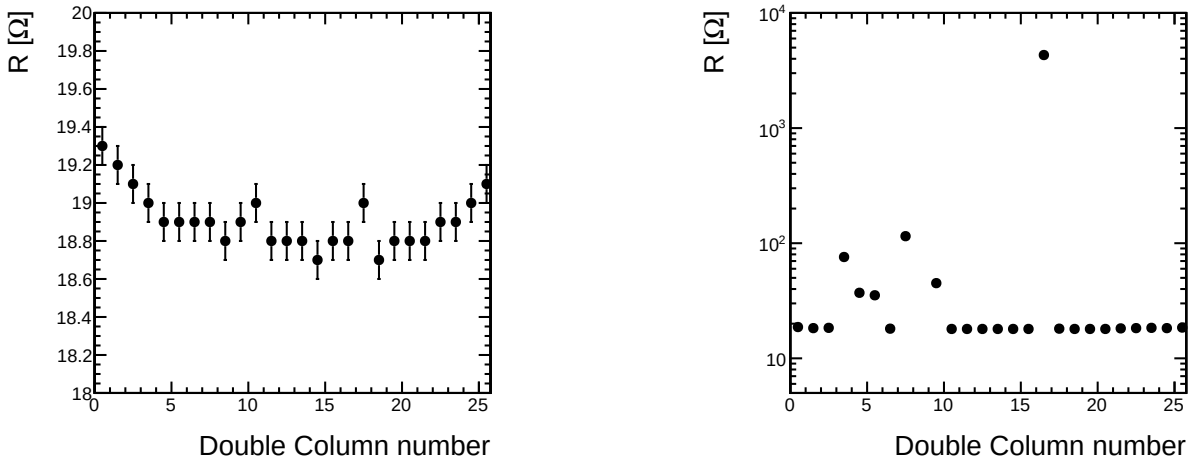


Figure 4.8: Impedance measurement on the daisy-chains of two different test assemblies. Left: no mechanical or electrical problems detected in the connection. Right: the assembly presents a broad region with high impedance values, one double column with very high impedance due to an imperfect connection between two bumps of the chain and one open chain (double column number 9) due to a missing bump. Note the logarithmic scale on the vertical axis of the right plot.

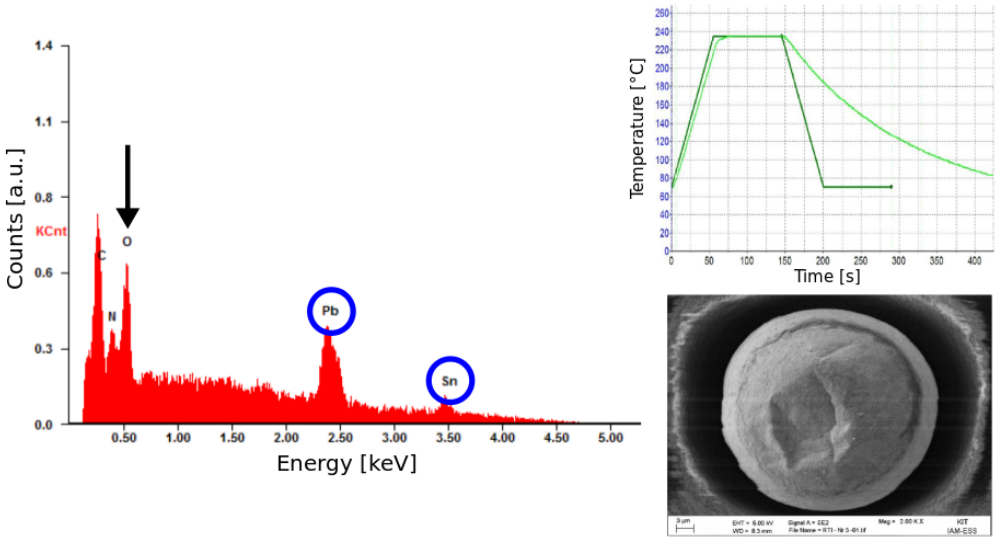
To test the dependence of the interconnection quality on the bonding force, a full testing bare module made of 16 ROCs with daisy-chain structures has been produced at the beginning of the test phase, applying to each group of ROCs different bonding forces, ranging from 60 N to 100 N. Looking at the results on the different chips in terms of connection strength during the pull-test and at the electrical properties of the chains, no difference between the ROCs bonded with different forces was noticed. This indicates that, if the bonding temperature is suitable,

even relatively low forces ($\sim 60\,\mathrm{N}$) are sufficient to break the native oxidation layer on top of the bumps and to create reliable mechanical connections. The bonding force, however, has direct impact on the separation gap between the ROC and the sensor and this, in turn, influences the electrical test of the bare module, changing the air gap capacitance seen by the calibration signal coming from the sensor side (see section 4.2). Taking this in account, the maximum value of the bond force during the process was set at 100 N, with the step-like behaviour already mentioned. For eutectic SnPb bumps, such as the ones deposited by RTI on the readout chips, the chosen bonding temperature is usually around 100 °C-150 °C: these temperatures are lower than the eutectic SnPb liquid point (at 183 °C), but high enough to soften the soldering metal and ease the diffusion onto the substrate UBM during the thermocompression phase. A small number of test bondings at much lower temperatures (50 °C) has been tried on dummy samples during the very first preliminary studies, but the results in terms of mechanical connection of the assemblies have not been satisfactory. The temperature of the process has been set to 140 °C: a value which was observed to guarantee an homogeneous intermetallic diffusion between bump and UBM. The same temperature is chosen for both the bonding table and the bond head in order to avoid additional stress due to a mismatch of the thermal conditions.

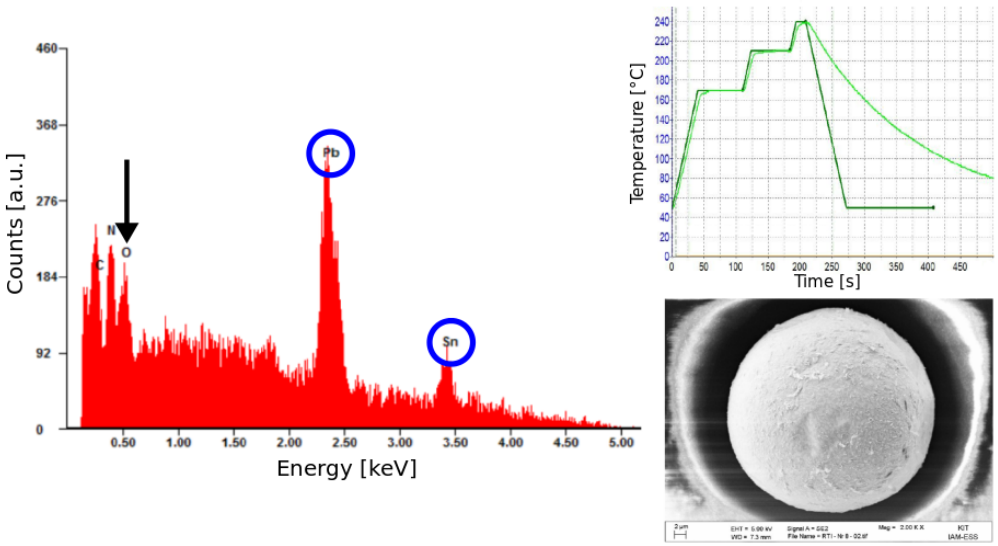
4.1.4.3 Optimization of the reflow parameters

In general, finding an optimal reflow configuration is more difficult than setting up a working bonding process, since the first one depends on many parameters, most of which are strongly correlated among themselves: the atmosphere composition in the reflow chamber, the temperature profile, the maximum value of temperature reached (peak temperature), the time above liquidus of the bumps (or dwell time) and the speed of the final cool-down.
The most suitable temperature profile was investigated fixing both the atmosphere (N_2/HCOOH mixture) and the peak temperature (235 °C) and performing test reflows using some readout chips as material (not bonded to any substrate) on top of which spherical eutectic SnPb bumps had been deposited and already reflowed once by the vendor. It was observed that, after reflowing the chip with a single-step profile (as in figure 4.9(a)), a lot of oxygen was present on the surface of the bumps, as shown by an Energy-dispersive X-ray spectroscopy (EDX) analysis, and that the bump shape was highly irregular. This was interpreted as a consequence of poor oxide removal due to an imperfect metal-oxide reduction reaction during the reflow phase: the presence of oxygen in the liquid drop reduces the liquid surface tension and affects the shape of the bumps. With a multi-step temperature profile (figure 4.9(b)) the presence of oxygen on the bump surface after the reflow was highly reduced and the shape improved clearly. It should be noted that the absolute value of the peaks in the EDX spectra is not a meaningful quantity here, because of the different acquisition time of the two measurements. Nevertheless, the difference between the two plots can be appreciated looking at the relative height of the oxygen peak with respect to the Sn and Pb peaks. It is also interesting to see that the metal composition of the surface does not respect the eutectic mixure of 63% Sn and 37% Pb of the bumps, since the lead peak is higher than the tin one. Although a more quantitative estimate would have been difficult to perform, this observation goes in the same direction of what reported in the existing literature, regarding the lead-enrichment in the surface composition of tin-lead alloys, when they are placed in an oxidizing environment [82]. The temperature profile of the reflow was therefore set to a double-step rise, with the first plateau at 150 °C, followed by a second rise towards the peak temperature. This also allows better control of the temperature rise and avoids possible thermal shocks due to a sudden temperature increase of the reflowed assembly.

The peak temperature and the dwell time are two strongly correlated parameters, since the



(a)

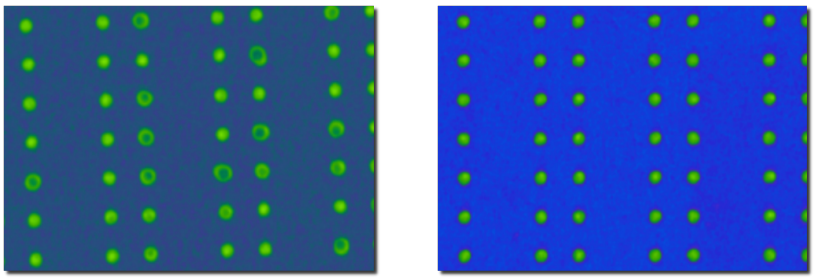


(b)

Figure 4.9: The left graph shows the X-ray photon spectrum obtained from an EDX analysis. Each peak is associated with a chemical element present in the sample. The oxygen peak is pointed out by the black arrow. The top right plot shows the reflow profile under test, with the dark green curve representing the set temperature and the light green curve the measured one. The bottom right plot is an SEM image of a bump surface. Top: single-step reflow profile. Bottom: multi-step profile (SEM images and EDX spectra: A. Jung, ITEP-KIT).

maximum cool-down speed of both the bonder and the reflow oven is limited to a few °C/s by the technical specifications of the machines and cannot be exceeded. High peak temperatures would therefore translate into longer time intervals in which the bumps remain in the liquid phase. It was empirically observed that the length of the dwell time was strongly correlated with the formation of *micrometric voids* located inside the reflowed bumps. These voids are among the typical defects which might appear when performing a sub-optimal reflow of solder joints: the current theoretical interpretation is that the voids are formed as a result of coalescence of much smaller gas bubbles, which are trapped inside the bumps during the liquid phase [83].

The bubbles are generated by the reducing action of the flux used during the solder or, in case of fluxless reflows (as in our operating condition) by the gases present inside the reflow chamber. It should be mentioned that the presence of voids inside the bumps, even when affecting a high percentage of the solder joints ($\sim 70\%$) was never observed to reduce the mechanical strength of the assembly as a whole in a dangerous way. The concern was however that these localized defects could have turned into a possible source of physical damage in the long run, because of the inhomogeneous distribution of current density in the solder bumps leading to bump failures [84]. By performing a set of different tests with the same reflow profile but different peak temperatures (and therefore different dwell times) it was experimentally observed that, for dwell times below 95-100 seconds, virtually void-free assemblies could be successfully produced. The X-ray images of two different bonded and reflowed assemblies are shown in figure 4.10, the false colour scale being a function of the material density. It is clearly visible that the material density of the bumps in figure 4.10(a) is not uniform and small areas with lower density, interpreted as voids, are well visible, while an optimized process leads to a higher bump density uniformity (figure 4.10(b)).



(a) (b)

Figure 4.10: Left: X-ray scan of an assembly reflowed with a profile having a long dwell time (~ 225 s). Right: result of a reflow with a dwell time < 100 s. The colour scale is a function of the material density of the bumps (images: RJL Micro & Analytic GmbH, Karlsdorf-Neuthard, Germany).

A not too-long dwell time is also beneficial for reducing an excessive formation of *intermetallic layers*. These intermetallic layers, which facilitate the actual bond between the solder joint and the substrate, are created during the reflow and the cooling process. The higher the peak temperature and the longer the time above liquidus are, the more these intermetallics are formed. In this process, however, there is a tradeoff: usually a too thick intermetallic layer is responsible for a degradation in the connection strength, since this is also the most brittle part of the solder [85]. This aspect could be experimentally verified with an assembly reflowed with a peak temperature of 240 °C and a dwell time of more than 4 minutes, which showed a mechanical strength during the pull-test less than half of the value found on assemblies reflowed with the profile of figure 4.10(b).
The final reflow profile is shown in figure 4.11. When moving to the SRO-700 reflow oven, it was decided to exploit the possibility of performing reflows under vacuum, switching on the N_2/HCOOH mixture only for a short period of time around the temperature peak, in order to improve the metal-oxide reduction and the bump shaping in a formic acid atmosphere. The fastest possible cool-down speed is achieved by flushing the chamber with a N_2 flux at the maximum available capacity (the section indicated with "Max. cooling" in figure 4.11). A reflow process takes approximately 30 minutes, but from the point of view of the production workflow this is an acceptable time, given the fact that typically at least six bare modules can be reflowed at the same time.

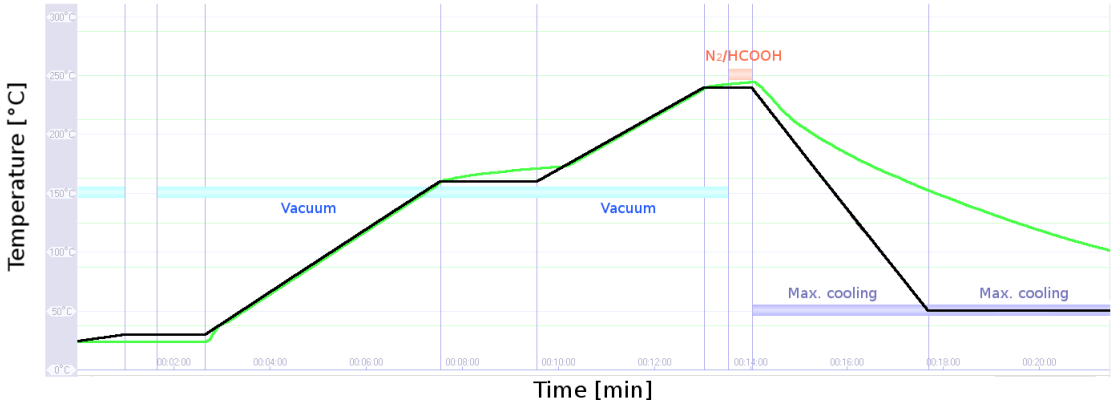


Figure 4.11: Temperature profile used for the reflow process at KIT. The black and the green lines show, respectively, the set and measured temperature. During the temperature rise, the reflow chamber is evacuated until the inner pressure reaches ~ 3 mbar. On the temperature peak (240 °C), a N_2/HCOOH mixture is injected in the chamber. The final cooling is achieved flushing the chamber with N_2.

4.1.5 Results

After the optimizations described in the previous paragraphs, the pixel bare modules assembled at the Karlsruhe Institute of Technology have been observed to possess excellent mechanical properties. The connection strength probed via pull-test is usually greater than 10 kg/ROC, with minimum values around 8 kg/ROC, corresponding to at least 1.9 cN for each bump [67]. As a rough term of comparison, which is nevertheless influenced by the fact of using a different soldering material, a value of 0.2 cN was found for each indium bump during the production of the modules for the current CMS pixel detector [64]. For the same eutectic SnPb process, RTI quotes a value of 4.5 cN/bump for the resistance to shear stresses, although a corresponding number for the pull-test strength is not explicitly stated [77]. The most observed breaking mechanism consists of bump disconnection on the ROC side, indicating a reliable intermetallic connection between the solder alloy and the UBM layers.
After reflow, the gap between the sensor and ROC is approximately 24 μm. The assemblies possess good planarity between sensor and readout chip, with a misplanarity between the region near the ROC pads and the opposite side below 0.3 mrad. The cross section inspection shows that the bumps do not present an excessive intermetallic growth and that they are free from major voids (figure 4.12). This result is confirmed also by more detailed X-ray scans.

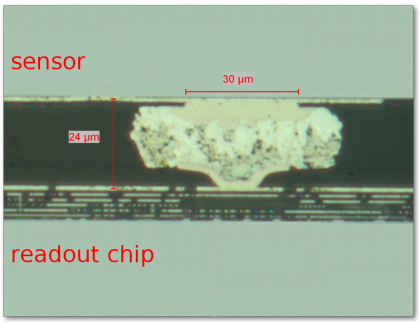


Figure 4.12: Optical image of a bonded assembly cross section. Different structures are clearly visible: the 40 μm wide UBM on the sensor side, deposited on top the the 30 μm wide aluminium pad, the UBM on the ROC side and the CMOS layers in the readout chip.

4.2 Bare module electrical test

Before proceeding with the assembly, all bare modules are tested on a probe station designed and built at KIT and located in the same class 1000 clean room where the flip-chip assembly takes place. The purpose of the test is to check the leakage current of the sensor after the bonding process, the electrical functionality of each readout chip and to estimate the number of missing or faulty bump bond interconnections. The results are then summarized in a grade, which is assigned to each bare module.
The reader is referenced to [86] for an exhaustive description of the test setup, its optimization and the details of the testing procedure, especially for what concerns the custom-made KIT bump bond test.

4.2.1 Bare module test sequence

The probe station, simplified down to its basic parts, consists of a motorized positioning stage which mounts a 1 cm thick dielectric teflon chuck. The bare module is placed on the chuck, with the bonded ROCs on the teflon surface and the back side of the sensor facing up. The reverse bias for sensor depletion is provided by an external power supply, contacted with one of the pads on the sensor back side via a tungsten needle. A steel plate supports a needle probe card which, during the test, estabilishes temporary electrical connections with the wire bond pads of a single readout chip. The card is connected to a Digital Test Board, a custom FPGA[8] board which communicates with the chip and reads the digitized outcoming signals. The test board is interfaced to a commercial PC where both the testing software and the software controlling the probe station run. To ensure a standardized procedure and the ability to share results and expertise, the test software, called pXar[9], is common to all the production centers, although it is possible to implement user-defined test routines, as done for the bump bond test described in section 4.2.1.3. The softwares controlling the probe station and the external powering have been written at KIT, and depend on the National Instruments LabWindows^{TM} [87] and National Instruments Vision [88] software packages. An optical microscope is used for the visual control of the bare module position and for checking the contact with the needle card. The image of the microscope is recorded by a camera mounted on the microscope. A second camera is used for pattern recognition during the alignment procedure (see later).
Great care has been put in reducing the electrical noise conditions of the probe station, by using appropriate shielding and by filtering the noise coming from the power line and from the bias supply line.
The test procedure is organized in different steps, which are repeated for each readout chip in turn.

1. The bare module is placed on the dielectric chuck and inserted in the probe station. A vacuum mechanism holds the module in place during the whole testing sequence.

2. The bare module is automatically aligned such that each one of the 35 wire bond pads of one ROC is precisely located under the corresponding needle of the probe card. The alignment is performed measuring the position coordinates of the first and the last chip of the row: the coordinates of the remaining chips are then obtained by linear interpolation. This alignment procedure relies on a pattern recognition software which compares the

[8] Field Programmable Gate Array, a configurable integrated circuit where complex logic can be implemented at hardware level.
[9] Pixel eXpert Analysis Readout.

image recorded by the camera with a reference model. The alignment of the opposite row is obtained based on the first row, after having rotated the module by 180 ° in ϕ.

3. The probe card needles are contacted with the first chip of the module and an I-V characterization of the sensor is performed (section 4.2.1.1).

4. The electrical functionality and the programmability of the ROC is checked running a standard test procedure via pXar (section 4.2.1.2).

5. A custom test for checking the quality of each of the 4160 bump bond interconnections is run (section 4.2.1.3).

6. The complete set of results coming from the above test procedures is uploaded to a database and a grade is automatically assigned to the bare module.

7. The bare module is removed from the probe station and is ready for being reflowed or reworked, depending on the outcome of the above tests.

The full test procedure lasts approximately 3 minutes for each ROC, for a total of 45 minutes for each bare module.

4.2.1.1 I-V characterization

The I-V sensor characterization on the probe station is executed contacting one readout chip (typically, the first one) with the needle card. The bias voltage is applied from the sensor back side, while the ground reference is provided by the ROC itself. The reverse bias voltage is ramped up in steps of 10 V, from 0 V to 250 V, with a 3 s time interval between each step. The leakage current is measured at the end of each time interval. The I-V scan needs to be performed only once for each bare module. This check is essential since, in general, the stress induced by the bonding process and/or possible damages due to the direct manipulation of the module (e.g., scratches on the guard ring structure) can modify the behaviour of the I-V curve measured during the pre-bonding tests on the sensors (section 4.1.1).
During the module production at KIT, it has been observed that some of the sensors which had been characterized as "good" during the in-house test presented an increased leakage current ($> \mu$A) after bump bonding, which influenced the bare module grading. To avoid losing all the readout chips bonded to one of these problematic sensors, an alternative workflow between bump bonding and testing was implemented already from the early phase of the module production. A single ROC is bonded on the sensor in the first position and the I-V behaviour of the incomplete bare module is tested on the probe station. If the results are satisfactory the bare module returns to the bonding machine for being completed with the remaining 15 ROCs, otherwise it is discarded. This simple procedure allowed to spot problems already at an earlier stage and to save a significant amount of material.
After completing the bare modules with the remaining 15 ROCs, a tendency of the good sensors to maintain their characteristics was observed. The current interpretation for the appearance of the problematic sensors points therefore in the direction of an incomplete testing mechanism *before* bonding, more than towards damage arising from the bonding itself. Before flip-chipping, the sensors are tested with a simple procedure where the voltage is applied between the high voltage pad and the outer guard ring (both on the back side), while a closer reproduction of the operating conditions would require to apply the bias between the pad on the back side and the punch-through bias grid on the front side. Since this two-side contact mechanism is difficult to implement for the pre-bonding checks, the above mentioned partial bonding and

testing procedure offers at least a workaround for mitigating the potentially undesired effects of an incomplete test.

4.2.1.2 Electrical test of the ROC functionality

During the electrical test, pXar routines are executed on the readout chip to check its programmability and the functionality of each pixel cell. Many of these tests rely on the ROC feature which allows to inject a calibration signal directly into the preamplifier input, as illustrated in figure 4.13.

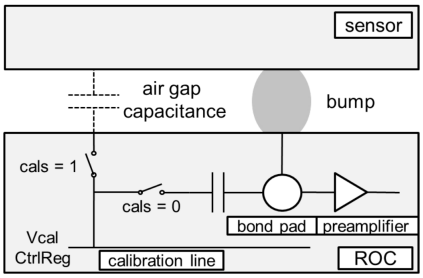


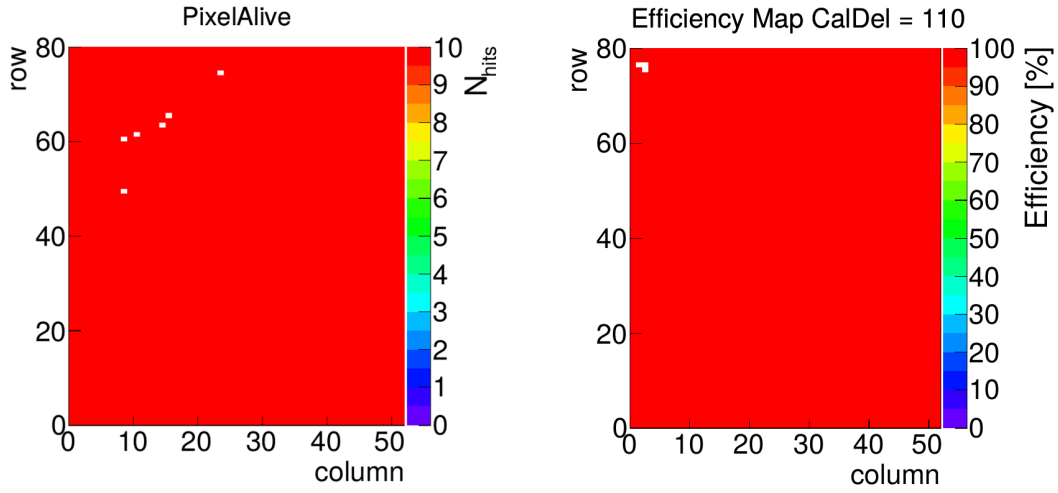
Figure 4.13: Internal calibration mechanism of the PSI46digV2.1respin readout chip. Setting the value of the parameter named `cals`, a calibration signal can be injected directly into the preamplifier or through the sensor, via the air gap capacitance. The parameters named `Vcal` and `CtrlReg` control the height of the calibration signal [86].

At first, ROC parameters such as, for example, the value of the threshold are optimized and saved for being used in the following.
One test routine sends 10 calibration signals in each of the pixel unit cells of the ROC. Exactly 10 hits are expected in return for a perfectly working pixel cell (see figure 4.14(a)). If the pixel is not working, no hits are observed and the pixel is defined as "dead". Pixel cells having a high noise level will observe a number of hits which is increased compared to the injected signals (possibly up to a saturation level where no further hits are registered) or decreased (in case of negative noise which reduces the signal below the threshold). In a second routine, the possibility of "masking" each pixel is tested: this possibility can be exploited, if needed, in case of noisy pixels. Finally, it is checked that the calibration signal is detected in the correct pixel: if no signal is observed or the signal is detected in another pixel, the pixel fails the test.

4.2.1.3 Bump bond test

The quality of the bump bond interconnection is probed sending a calibration signal through the sensor, via the air gap capacitance between the sensor and the readout chip (see figure 4.13). This kind of test is strongly dependent on the bump bonding technology and on the size of the ROC/sensor separation gap. For this reason, a custom-made bump bond test has been developed at KIT and is specifically tuned for the RTI bumping and the KIT bonding process. Ultimately, the requirement for an optimal bump bond test is to deliver results which are compatible with the ones obtained with tests closer to the real operating conditions of a module, for example tests where charge is generated in the sensor by irradiation with X-rays or charged particles coming from a radioactive source. These testing procedures, however, are difficult to implement at this level, due to practical, safety-related and legal constraints.
The KIT bump bond test is organized in two parts. In the first part, the optimal threshold level is set. This is defined as the minimal threshold at which the noise is still suppressed (less than five pixels out of 4160 are noisy). In the second part, while keeping this threshold level, 100 calibration signals for each pixel are sent through the sensor and read out. To take into account the systematic difference in the time of injection for pixels located in different regions of the ROC, this procedure is repeated varying the time at which the calibration signals are injected.

An efficiency map is then obtained measuring the efficiency of each pixel, where the efficiency is defined as the ratio between the number of calibration signals received and injected (see figure 4.14(b)). Pixels with a missing bump or with defective electronics have an efficiency much lower than good ones. Based on the collected experience and on the results of cross-checks with X-ray irradiation tests, a discrimination cut is set at 50% efficiency.



(a) (b)

Figure 4.14: Left: exemplary result of the pXar routine checking the functionality of each pixel cell of a ROC: working pixels register all the 10 calibration signals which are sent to the preamplifier. The 6 pixels in the top left region of the chip are defective. Right: exemplary efficiency map obtained from the bump bond test algorithm, for a certain value of the parameter (`CalDel`) controlling the time injection of the calibration pulse. Three missing or defective interconnections can be seen in the top left corner of the chip. The two maps are obtained from two different readout chips and are therefore not related to each other. Plots from [86].

During the commissioning of the probe station and the first part of the KIT pixel module production, the results given by the custom bump bond test have been cross-checked with another testing algorithm developed at DESY and with the results obtained from the X-ray irradiation on full modules. In both cases, a good correlation in the number of found defects has been measured, with the KIT bump bond test being slightly more conservative than the DESY one (i.e., classifying also some pixels which were responsive during the X-ray scan as defective) [86].

4.2.1.4 Bare module grading

After the upload of the test results to a central production database, a grade is automatically generated for each bare module. The grading scheme has three levels: A (best grade), B and C (worst) and takes into account the following informations.

- Sensor leakage current, extrapolated to room temperature, from the I-V characterization:
  - Grade A: $I(150\,V) \leq 2\,\mu A$ and $I(150\,V)/I(100\,V) \leq 2$
  - Grade B: $2\,\mu A < I(150\,V) \leq 10\,\mu A$ and $I(150\,V)/I(100\,V) \leq 2$
  - Grade C: $I(150\,V) > 10\,\mu A$ or $I(150\,V)/I(100\,V) > 2$

- Number of defective pixel channels, from the electrical test and the bump bond test:
  - Grade A: ≤ 41 defective pixel channels per ROC (1%)
  - Grade B: ≤ 164 defective pixel channels per ROC (4%)
  - Grade C: > 164 defective pixel channels per ROC
- Digital current of each ROC:
  - Grade A: $I_{dig} \leq 65\,mA$
  - Grade C: $I_{dig} > 65\,mA$

The grade assigned to the module is the worst of the single grades for the above points.
In case major defects found on the bare module are located on a single problematic readout chip, it is possible to apply a rework procedure on the bonding machine and replace the defective ROC with a new one, as described in section 4.1.3.2. If at least one of the readout chips of the module has been reworked, a "minus" sign is added to the grade for internal reference (e.g., A−).

4.2.2 Results from the KIT bare module production

During a production campaign which lasted approximately one year, from May 2015 to May 2016, a total of 437 pixel bare modules have been bonded and tested at KIT. Of these, 406 (corresponding to 92.9% of the total) have been classified as grade A during the bare module test on the probe station, 14 (3.2%) as grade B and 17 (3.9%) as grade C. The production rate, expressed in units of tested bare modules per week, is shown in the left plot of figure 4.15, where the colour code distinguishes the modules according to their grade. The time periods where no bare modules could be assembled are related to temporary interruptions in the production flow, caused by delays in the delivery of the bumped wafers. A cumulative distribution of the testing rate, showing the total number of produced bare modules as a function of time, is presented in the right plot of figure 4.15. Overall, the yield of high-quality assembled bare modules, obtained adding up the A and B-graded modules, is larger than 96%: a result which proves the outstanding performance of the KIT bare module assembly chain in terms of material selection, quality control and reliability of the bonding and reflow processes which have been developed.
The reasons which caused the few low grades are almost exclusively related to sensor problems. The majority of B-graded bare modules present a value of the leakage current measured during the bare module test on the probe station between $2\,\mu A$ and $10\,\mu A$. An even larger value of the measured leakage current ($I(150\,V) > 10\,\mu A$) or an earlier "breakdown" of the sensor during the ramp-up of the reverse bias voltage are the main failure modes for the C-graded bare modules. Only in a limited number of cases the low grading is caused by mechanical damages (typically, on the guard ring structure) or by electrical problems on one or more readout chips. For the A-graded bare modules, the average sensor leakage current is measured to be $0.9\,\mu A$ and a tendency is observed in maintaining similar values of leakage current both before and after the bump bonding process. This indicates that the mechanical stresses induced on the sensors during the direct manipulation and the bonding sequence are negligible.
The reason for the low number of problems related to the readout chip electronics is that localized defects such as dead double columns or large regions with irresponsive pixels could be often successfully solved substituting the defective chip via a rework procedure. Among the 437 produced bare modules, 86 (19.7% of the total) have been reworked. In more than 80% of the cases, the rework procedure improves the classification of the problematic module, promoting B or C-graded modules to the highest grade.

The quality of the bump bond interconnections is observed to be excellent. The average number of defective pixels per module (that is, the combined effect due to missing bump bond connections, dead pixels or pixels with high noise level), as measured during the bump bond test, varies between 4 (for the A-graded modules) and 9 (for the lower grades). In about 30% of the bare modules exactly zero defective pixels are found. In the worst case, the maximum number of defective pixels found on a bare module has been 62, which corresponds to less than 0.1% of the total interconnections. A very good agreement is observed between the results provided by the bump bond test developed at KIT and the testing algorithm provided by DESY.
A total of 37 ROC wafers have been used during the production campaign and, in general, from 11 to 12 bare modules could be bonded with the readout chips coming from a single wafer. Considering that, on average, 90% of the 244 readout chips of a wafer are declared as electrically good during the wafer-level test performed at PSI before bumping (see section 4.1.1), this indicates that approximately 87% of the functioning ROCs have been used for the assemblies. The remaining 13% convolutes the fraction of readout chips which have been discarded during the quality checks performed after the cleaning procedure, the ones which have been damaged during bumping, thinning or dicing and the ones which have been bonded and subsequently reworked.
Of the 475 good quality sensors used during the production, around 8% had to be put aside due to defects found during the optical inspection or because they failed the single-ROC test described in section 4.2.1.1 and had to be discarded before being completely bonded.
The non-negligible number of nominally good readout chips and sensors which had to be rejected clearly illustrates the advantage of having robust quality control procedures and of designing test routines as close as possible to the real usage conditions of the final assemblies.

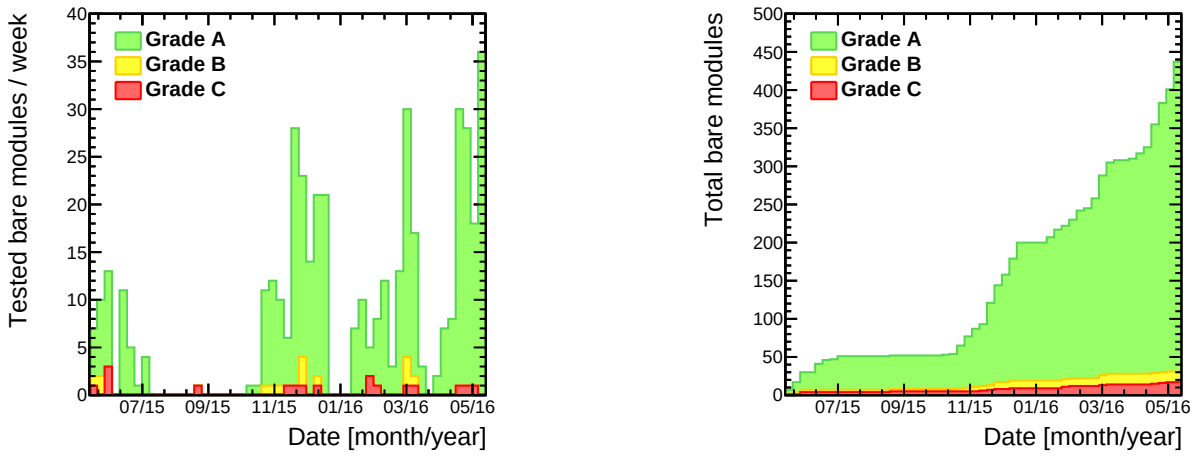


Figure 4.15: Left: number of bare modules assembled and tested during each week of the production campaign. Right: total number of produced bare modules. The colour of the histograms reflects the grade computed for each bare module based on the results of the probe station test.

4.3 Full modules assembly line

After production, testing and grading, the bare modules are packed into sealed boxes and safely moved from the IPE clean room to a class 100000 clean room located inside the IEKP facilities at KIT Campus North. Here, if the grading from the electrical test is equal or better than B,

they are assembled into full modules. This section presents a brief overview of all the steps of the full module assembly line, starting from the single components.

4.3.1 TBM and HDI mounting and testing

The High Density Interconnect (HDI) consists of a low X_0 flex print circuit made of a Kapton film with surface mounted components. All the HDIs are centrally produced by a company called Hightec MC AG based in Lenzburg, Switzerland, and sent to each production center after a simple optical inspection where the absence of major mechanical defects is checked. In the design of the modules for the second to fourth layer (as in figure 3.17) each HDI is equipped with one single Token Bit Manager (TBM) chip, which is glued and wire bonded onto the flex surface. The TBM chips, produced by IBM, are electrically tested at wafer level before being diced and delivered to the production centers.
At KIT, each TBM is glued on top of an HDI using a two-component mixture of Araldite 2011 glue. Precise positioning and alignment during glueing is ensured by custom designed mechanical jigs which can accommodate up to 8 TBMs and HDIs for each glueing session. After allowing the glue to cure, the TBM is electrically connected to the flex print circuit using a standard wedge-to-wedge wire bonding technology at room temperature. The wirebonding machine, a Bondjet 710 manufactured by Hesse & Knipps, connects the pads on the HDI with the TBM via ultrasonic welding, using AlSi wires with a diameter of 25 μm. The quality of the process is assessed by an optical inspection on each single bonded component, while the mechanical strength of the connection has been probed on test material by means of a pull-test procedure[10], obtaining results larger than 10 g/bond.
Once a reliable mechanical and electrical connection with the TBM has been estabilished, each HDI is electrically tested. The purpose of this test is to check the TBM-to-HDI communication, before using the flex print circuit to produce a complete pixel module: glueing the HDI on the sensor side of a bare module is an irreversible process and a single faulty component might, at this step of the production chain, compromise the whole assembly, turning the efforts into a waste of good quality sensors and readout chips. During the test procedure the HDI, fixed on a mechanical jig, is connected via a needle card equipped with 30 spring-loaded probing needles to a relay board. Each needle is mechanically put in contact with a test pad on the HDI circuitry. The relays of the board are controlled by a test software running on a commercial PC. At the same time, a custom FPGA running on a Digital Test Board connected to the HDI executes a sequence of 15 different tests which, among the others, check the clock and serial data signals coming from the TBM chip. The output signals are driven to an external oscilloscope, where they are visually inspected by the operator. The test routine is concluded with an high voltage test where the HV is routed to the HDI and the current is measured. This ensures that the line which ultimately brings the bias voltage to the sensor works properly.
After completion, each HDI is graded with a binary pass/fail grade, depending if all the signals have been correctly readed on the oscilloscope and if their amplitudes are measured to be within 90% of the nominal value[11].
Only the flex print circuits which successfully pass the electrical test are used to complete the bare modules into full modules.

[10] This test follows the prescriptions of method 2011.9 from MIL-STD-883.

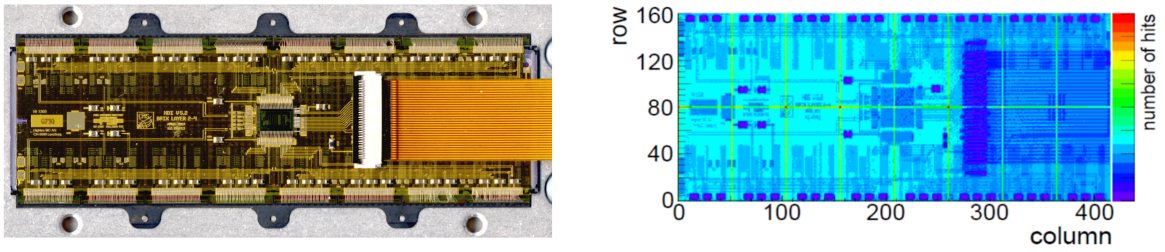
[11] For the signals which are probed in this test, this corresponds to the requirement that the amplitude has to be larger than 20 mV.

4.3.2 Module assembly

The first step consists in mounting the silicon nitride base strips on the bottom part of the bare module, glueing them to the bonded readout chips. Alignment during this procedure is maintained thanks to the usage of a special glueing jig.
In a second sequence, one of the HDIs passing the above described test is glued on the back side of the silicon sensor, after having precisely aligned the two components based on reference marks engraved on the sensor and on the HDI surface. The process is monitored by a movable microscope connected to a camera while a coordinate measuring machine precisely checks the alignment of the components after the glue has been allowed to cure out. The module is then completed connecting the 16 readout chips to the glued HDI via wire bonds. The same machine and technique used for the TBM-to-HDI bonding is employed. An additional AlSi wire connecting the HDI to the sensor pad provides the bias voltage needed for operating the sensor as a particle detector. The readout and signal cable is finally connected to the HDI via a connector previously soldered on the HDI surface. A picture of a fully assembled barrel module produced at KIT is shown in figure 4.16(a).
Having different glueing stations working in parallel allows to mitigate the impact of the glue curing time (typically, several hours), therefore maintaining an high assembly rate during the whole production campaign. The typical assembly rate has been around 12-18 full modules/week, with a peak of 24 modules/week reached around the middle of the production.

4.4 Testing and final qualification

A final optical inspection is performed at KIT on the complete module, followed by basic electrical tests. These include a repetition of the I-V characterization as well as some of the functionality checks already performed during the bare module test. In a further measurement, the module is placed into a box where it is irradiated with an X-ray beam, while the signals are read out. This allows to obtain a definive assessment about the number of pixels which are properly working and the ones which are defective due to faulty bump bond interconnections or problems in the pixel unit cells of the readout chip. An exemplary efficiency map is shown in figure 4.16(b).



(a) (b)

Figure 4.16: Left: picture of a fully assembled pixel module. Right: example of an efficiency map obtained irradiating a full module with an X-ray beam. The colour scale is a function of the number of photons collected by each cell: defective pixels or missing bump bond connections manifest themselves as white dots in the map. It is interesting to note how the photon collection efficiency is modulated by the inactive components placed in front of the sensor, such as the surface mounted components on top of the HDI, the readout cable and the TBM. The edge pixels covering the space between two adjacent ROCs and the corner pixels present a higher efficiency due to their larger size.

If no major defects are found in this final test, the modules are shipped to RWTH Aachen, as shown in the workflow diagram of figure 4.1. Here the complete and final qualification of the pixel modules produced at KIT is performed. This qualification includes a measurement of the leakage current at room temperature and a test inside a cold-box, where the modules are cycled ten times between −25 °C and +17 °C. This test ensures that mechanical stresses due to temperature gradients similar to the ones that must be sustained in the actual experimental conditions do not have an impact on the quality of the interconnections. An electrical test similar to what has been described in section 4.2 and consisting of an I-V characterization of the sensor, of a test of the ROC functionality and adjustability and of a check of the interconnection quality is performed both at −20 °C and +17 °C.
The full module qualification is followed by an high-rate X-ray test, whose results are also converted into a grade similar to the one used for the bare modules and uploaded to the central production database. The final step consists of a complete module calibration with monochromatic secondary X-ray photons coming from a set of metallic targets irradiated with a primary X-ray beam.

4.4.1 Results from the complete pixel module production at KIT

The full module assembly line operated in parallel to the bare module production, successfully concluding its operations the first week of June 2016. Out of the 420 A and B-graded bare modules, 409 have been assembled into full modules. Part of the remaining ones has been sent to other production centers due to a lack of available and tested HDIs around the end of the production, while the rest has not been completed into full modules due to small mechanical defects or multiple reworks on individual readout chips. A graphical summary of the production campaign is shown in figure 4.17.

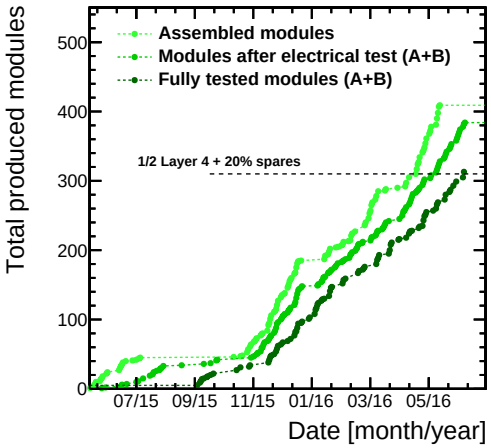


Figure 4.17: The light green graph indicates the progress in the number of A and B-graded bare modules which have been assembled (409 in total). The middle graph and the dark green graph show, respectively, the number of A and B-graded full modules after the electrical test at KIT and the total number of A and B-graded full modules after the high-rate X-ray test in Aachen. The horizontal dashed line indicates the number of fully assembled and tested good modules which had to be reached, corresponding to half the layer 4 quota and 20% spares. Module testing in Aachen is still ongoing.

The horizontal dashed line indicates the target of 310 good-quality full modules, corresponding to half of the layer 4 quota (256 modules) and 20% spare modules in addition. The light green graph indicates the progress in the number of A and B-graded bare modules which have been completed into full modules, for a total of 409 at the end of the campaign. The middle graph shows the total number of full modules which obtained a grade equivalent to A or B after the electrical test performed at KIT. A lower grade has been assigned to the remaining modules due to occasional issues in the wirebonding of one or more readout chips to the HDI, in the connection between the signal cable and the HDI or because of problems in the readout chips

found during the electrical test routines. The dark green graph indicates the total number of A and B-graded full modules after the high-rate X-ray test in Aachen. A certain delay between the assembly at KIT and the final qualification in Aachen is to be expected and module testing in Aachen is therefore still ongoing. The targeted goal of producing 310 good-quality complete pixel modules has nevertheless already been achieved and exceeded around the second week of June 2016.

Chapter 5

Validation of the tau identification at CMS using Drell-Yan events

The general improvements in the physics performance which are introduced by upgrades in the detector must necessarily be complemented by a parallel evolution of the techniques used for the reconstruction and identification of physics objects. This is especially true for those objects, such as jets or τ leptons, whose reconstruction is challenged by the fact of involving many different subdetectors.
The motivation behind these efforts is always deeply connected with the physics goals of the experiment. As an example, the outstanding CMS tau reconstruction performance has been the main pillar on which many results produced by the collaboration in the data collected during the Run I operations of the LHC were built. Among others, this led to the evidence of a 125 GeV Higgs boson decaying to a pair of tau leptons [89]. During the very first part of this Ph.D. project, contributions have been provided to a multivariate analysis designed for the $H \to \tau\tau \to \mu\mu$ final state, which has been included, together with the other $\tau\tau$ final states, in the CMS public result cited above. More details about this analysis can be found in [90]. On the same Run I dataset, more and more stringent exclusion limits have been established by the CMS collaboration on the search for additional Higgs bosons, which are predicted by different supersymmetric extensions of the Standard Model [91–94].
When increasing the center of mass energy of the proton-proton collisions to 13 TeV, it is essential to validate the performance of both the old and the newly developed tau identification algorithms under such different experimental conditions, before applying them in the context of searches for new physics or in other Standard Model measurements.
After a general introduction to the properties of tau leptons (in section 5.1), section 5.2 familiarizes the reader with the main ideas behind the tau reconstruction in CMS, introducing the basic aspects of the Hadron Plus Strips algorithm. Section 5.3 describes in more detail the development of an improved discriminator for the rejection of electrons mimicking the signature of hadronically decaying taus, evaluating the discriminator's performance on events simulated according to the Run II conditions. Section 5.4 explains how the well understood Drell-Yan process $q\bar{q} \to Z/\gamma^* \to \tau\tau$ can be used as a Standard Model reference for validating the performance of the algorithms in the data collected in 2015, during the first part of the 13 TeV LHC run. The rest of the chapter presents a set of measurements performed on Drell-Yan events: the measurement of the $e \to \tau_h$ misidentification rate, of the efficiency of a multivariate tau isolation discriminator and of the tau energy scale. Finally, an alternative approach for extracting the tau identification efficiency and the tau energy scale is presented as a by-product of an in-situ

measurement of the $Z \to \tau\tau$ cross section, obtained from the statistical combination of three $\tau\tau$ final states.

5.1 Tau decay modes

With a mass of $m_\tau = 1.777\,\text{GeV}$ [1], tau leptons are the only leptons heavy enough to decay into hadrons. Table 5.1 summarizes the main tau decay modes and their relative branching ratios. In about one third of the cases, taus decay leptonically into an electron or muon plus two neutrinos. The remaining are hadronic decay modes (referred collectively as τ_h in the following), in which taus decay typically into either one or three charged pions or kaons, plus one neutrino and up to two neutral pions. Neutral pions decay into photon pairs ($\pi^0 \to \gamma\gamma$) on a timescale of $10^{-16}\,\text{s}$. From the experimental point of view, the photons coming from neutral pion decays have a non negligible probability of converting into e^+e^- pairs in the tracker detector before reaching the electromagnetic calorimeter.
The reconstruction and identification of electrons and muons coming from leptonic tau decays follows the techniques outlined in section 2.2, while the undetectable neutrinos contribute to the event signature as a genuine source of missing transverse energy. The reconstruction of the visible products of the hadronic tau decays is instead based on a specific algorithm, usually referred to as the Hadron Plus Strips (HPS) algorithm. A brief overview of the HPS algorithm is given in the following. For a more detailed description the reader is pointed to [95].

Decay Mode	Resonance	BR [%]	Reconstructed Decay Mode
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8	
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4	
All leptonic modes		35.2	
$\tau^- \to h^- \nu_\tau$		11.5	h^-
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0	$h^-\pi^0$
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	10.8	$h^-\pi^0\pi^0$
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8	$h^-h^+h^-$
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8	
Other hadronic modes		1.8	
All hadronic modes		64.8	

Table 5.1: Approximate branching ratios of the main tau decay modes [1]. The symbol $h^\pm$ is used in the table and throughout the text to generically indicate a charged pion or kaon. The last column lists the decay modes which are reconstructed by the HPS algorithm. Charge-conjugated decays have the same branching ratios and, for simplicity, are not included in the table.

5.2 The Hadron Plus Strips algorithm

The HPS algorithm exploits the CMS particle-flow event interpretation in terms of unambiguous collections of high-level objects (charged and neutral hadrons, photons, muons and electrons). It is seeded by jets reconstructed using the anti-k_T algorithm described in section 2.2.4, with a parameter $R = 0.5$ (during Run I) or $R = 0.4$ (during Run II). The reconstruction of the hadronically decaying taus proceeds looking for any combination of the charged and neutral particles contained in the jet compatible with the main tau decay modes listed in table 5.1.

Neutral pions are reconstructed clustering the photons and electrons contained in the jet into η-ϕ strips. Only electrons and photons having a $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ are considered for the clustering. As the name suggests, the strips have a larger size in the ϕ dimension compared to the η dimension in order to take into account the bending of $\mathrm{e}^+\mathrm{e}^-$ conversion pairs coming from $\pi^0 \to \gamma\gamma$ decays in the 3.8 T solenoidal magnetic field of the detector. In the reconstruction algorithm implemented during Run I, the strips had a fixed η-ϕ size of 0.05×0.20. The η dimension of the strip has been made p_{T}-dependent in the version of the HPS algorithm developed for Run II to collect low-p_{T} electrons coming from photon conversion and falling outside the fixed dimension of the strip. After the clustering, all strips having a sum of transverse momenta of electrons and photons larger than 2.5 GeV are retained and considered as π^0 candidates. The π^0 candidates and the charged constituents of the jets are then combined under different decay mode hypotheses and each hypothesis is tested based on the invariant mass of the constituents and requiring the charged particle tracks of the $h^{\pm}h^{\mp}h^{\pm}$ decay mode to be compatible with the same vertex hypothesis. The algorithm aims at reconstructing all the main hadronic decay modes listed in the last column of table 5.1, namely: $h^{\pm}$, $h^{\pm}\pi^0$, $h^{\pm}\pi^0\pi^0$ and $h^{\pm}h^{\mp}h^{\pm}$. In case more than one combination of charged particles and π^0 candidates satisfies a given decay mode hypothesis, only the combination with the largest total p_{T} is retained, such that at the end of the reconstruction sequence each tau-jet is unambiguously associated with one decay mode.
Figure 5.1 shows the correlation between the reconstructed and generated tau decay modes for hadronic tau decays in simulated $\mathrm{Z}/\gamma^* \to \tau\tau$ events generated according to the expected Run II conditions. A set of minimal selections is applied on the kinematic of the τ_{h} candidates, requiring their transverse momentum to be larger that 20 GeV and the absolute value of the pseudorapidity to be smaller than 2.3. An additional selection based on an isolation discriminator, whose motivation will be clarified in the next paragraph, reduces the probability that the τ_{h} candidates originate from misidentified jets. From the correlation table it can be seen that, after these selection requirements, the probability of correctly reconstructing the decay mode is about 90%. In particular, the signature of $h^{\pm}h^{\mp}h^{\pm}$ decays is almost always correctly reconstructed, while taus decaying into a single hadron are wrongly reconstructed as $h^{\pm}\pi^0$'s in around 10% of the cases, due to electromagnetic energy deposits near the tau direction being incorrectly clustered into strips. This effect has a dependence on the pile-up conditions and is therefore expected to be slightly larger (in the order of 1-2%) during Run II, compared to the Run I expectation.

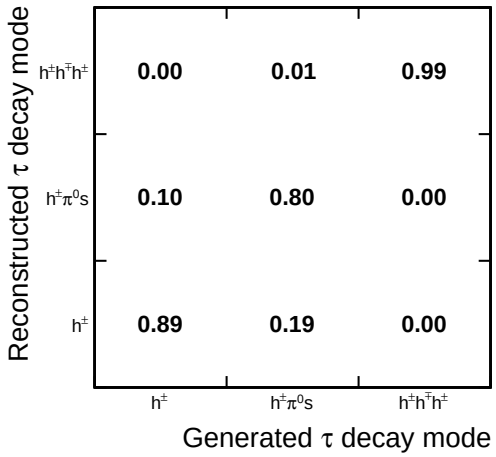


Figure 5.1: Correlation between reconstructed and generated tau decay modes, for hadronic tau decays in simulated $\mathrm{Z}/\gamma^* \to \tau\tau$ events. The numbers indicate the fraction of τ_{h}'s generated with decay mode x which are reconstructed in the decay mode y. Reconstructed τ_{h} candidates are required to match within $\Delta R < 0.3$ to generated hadronic taus, to be reconstructed in one of the decay modes of table 5.1, to pass the tight WP of the MVA isolation discriminator (see later in the text) and to satisfy $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$. Due to the large number of τ_{h} candidates, the statistical error on the numbers is negligible.

At this step, only approximately 60% to 85% (depending on the transverse momentum) of the reconstructed τ_{h} candidates come from genuine tau decays, while a sizeable contribution is given by quark or gluon jets being misidentified as tau decay products. A set of tau *discriminators*,

based on various tau properties, has been developed to distinguish between hadronically decaying taus and jets, electrons or muons mimicking a tau signature. A typical handle to reduce jet misidentification consists for example in exploiting the fact that hadronic taus and their decay products are usually isolated with respect to the rest of the event, in contrast to jets initiated by quarks or gluons, which are characterized by a much larger hadronic activity in their surrounding. A simple but robust "cut-based" tau isolation discriminator can be computed summing the transverse momenta of the charged particles contained in a signal cone around the reconstructed tau direction and applying a correction to account for the neutral particles coming from pile-up interactions. A more refined discriminator can be built combining a set of tau-related variables into a multivariate (MVA) discriminator. Among others, lifetime-related variables are also included to exploit the fact that the tau lifetime ($\gamma c\tau = \gamma \cdot 87\,\mu\text{m}$) is large enough to distinguish the secondary vertex of tau decays, thanks to the good performance of the CMS tracker. Other discriminators make use of the properties of tau-initiated calorimeter showers or track-related variables to separate genuine tau decays from electrons and muons mimicking a tau signature.
Usually, a set of working points (WP) is defined on the output of each discriminator, based on its expected *efficiency* on genuine τ_h decays and on the *misidentification rate*, defined as the probability to incorrectly identify other physics objects (quark/gluon jets, electrons or muons) as τ_h candidates.
The following section describes in detail a particular example of discriminator designed to reduce the $\text{e} \to \tau_\text{h}$ contamination, which has been implemented for the Run II τ_h identification. Discriminating against $\mu \to \tau_\text{h}$ misidentification is in general less challenging, due to the fact that muons, being more massive compared to electrons, tend to radiate much less on their way through the detector material, therefore mainly influencing the τ_h reconstruction only in the $h^\pm$ decay mode. Moreover, the signature in the muon chambers is subject to lower uncertainty compared to calorimetric deposits. The techniques used for suppressing $\mu \to \tau_\text{h}$ misidentification will not be discussed further. More details about muon discrimination, together with the complete set of discriminators developed for Run I and their expected and measured performance, can be found in [95].

5.3 MVA-based discriminator against electrons

Electrons have a non-negligible probability of being misidentified as the charged decay products $h^\pm$ of a hadronic tau decay or, if they emit bremsstrahlung photons, of mimicking the signature of the $h^\pm\pi^0$ decay mode. This is especially true for isolated electrons coming from the decays of Z or W bosons, which have a high probability of passing the isolation-based tau identification criteria mentioned in the previous section. In the tau identification implemented during Run I, the rejection of electrons misidentified as τ_h candidates was achieved with the help of a dedicated multivariate discriminator [95]. For Run II, an improved version of the discriminator has been developed, with the aim of further reducing the $\text{e} \to \tau_\text{h}$ contamination while maintaining high efficiency for genuine tau decays over a wide range of tau transverse momenta.
The discriminator is based on a boosted decision tree (BDT) [96] which has been trained to separate electrons from hadronic tau decays. As input, the algorithm uses an ensemble of calorimetry-based variables to discriminate electromagnetic from hadronic showers, in combination with photon-related variables which are sensitive to the emission of bremsstrahlung along the tau track having the largest p_T (also referred to as "leading" track), together with other track-related variables.
In detail, the following variables are used as input for the BDT:

- The electromagnetic energy fraction $E_{\mathrm{ECAL}}/(E_{\mathrm{ECAL}}+E_{\mathrm{HCAL}})$, defined as the ratio between the amount of energy deposited in the ECAL and the sum of the ECAL and HCAL energy deposits which are associated with the decay products of the τ_{h} candidate.

- E_{ECAL}/p and E_{HCAL}/p, defined as the ratios between the ECAL and HCAL energies associated with the leading track of the τ_{h} candidate and the leading track momentum.

- $m_{\tau}^{\text{in-sigcone}}$, the visible mass of the τ_{h} candidate, computed summing the four-momenta of photons and charged particles inside a τ_{h} *signal cone* having a p_{T}-dependent R size around the tau direction, where R is defined as:

$$R \equiv \begin{cases} 0.1 & \text{if} \quad p_{\mathrm{T}} < 30\,\mathrm{GeV} \\ \dfrac{3.0}{p_{\mathrm{T}}/\,\mathrm{GeV}} & \text{if} \quad 30\,\mathrm{GeV} \leq p_{\mathrm{T}} \leq 60\,\mathrm{GeV} \\ 0.05 & \text{if} \quad p_{\mathrm{T}} > 60\,\mathrm{GeV}\,. \end{cases} \tag{5.1}$$

  This variable cone size takes into account the fact that the decay products of high-p_{T} taus are more collimated than the decay products of low-p_{T} taus.

- $F_{\mathrm{brem}} \equiv |p_{\mathrm{in}} - p_{\mathrm{out}}|/p_{\mathrm{in}}$, where p_{in} and p_{out} are the momenta, measured from the track curvature at the innermost and outermost position in the tracker, of the tracks reconstructed using the Gaussian-sum filter (GSF) algorithm: a particular track-reconstruction algorithm which has been developed with the aim of providing a more accurate treatment of the bremsstrahlung energy loss distribution of particles, especially electrons, interacting with matter [97].

- The ratio $E_{\mathrm{ECAL}}/p_{\mathrm{in}}$ between the total ECAL energy and the inner track momentum. The numerator includes the contribution coming from the bremsstrahlung photons reconstructed as ECAL clusters tangent to the track.

- $N_{\mathrm{hits}}^{\mathrm{GSF}}$, the number of valid hits of the track reconstructed by the GSF algorithm.

- $(N_{\mathrm{hits}}^{\mathrm{GSF}} - N_{\mathrm{hits}}^{\mathrm{KF}})/(N_{\mathrm{hits}}^{\mathrm{GSF}} + N_{\mathrm{hits}}^{\mathrm{KF}})$, where $N_{\mathrm{hits}}^{\mathrm{GSF}}$ and $N_{\mathrm{hits}}^{\mathrm{KF}}$ are, respectively, the number of valid hits in the tracker detector which are associated with the track reconstructed by the GSF or Kalman filter (KF) algorithms. The Kalman filter is the standard algorithm used for track reconstruction in CMS, as described in [50]. This combined variable is particularly sensitive to the emission of hard bremsstrahlung photons.

- The χ^2 per degree-of-freedom (DoF) of the track fit performed with the GSF and the KF algorithms.

- The number of photons in any of the strips associated with the τ_{h} candidate.

- $\sqrt{\sum(\Delta\eta)^2 \cdot p_{\mathrm{T}}^{\gamma\ \text{in/out-sigcone}}/\,\mathrm{GeV}}$ and $\sqrt{\sum(\Delta\phi)^2 \cdot p_{\mathrm{T}}^{\gamma\ \text{in/out-sigcone}}/\,\mathrm{GeV}}$ computed from the p_{T}-weighted square of the distance in η and ϕ between each photon included in a strip and the leading track of the τ_{h} candidate.

- $\sum E_{\gamma\ \text{in/out-sigcone}}/E_{\tau}$, the fraction of τ_{h} energy carried by photons.

The photon-related variables of the last three bullet points in the above list are computed separately for photons inside and outside the τ_{h} candidate signal cone in order to increase their separation power. The set of input variables is then completed by the p_{T} and η of the hadronic tau candidate and by the ratio between the transverse momentum of the τ_{h} leading track and

the τ_h transverse momentum. Additionally, the distances in η and in ϕ between the GSF track and the nearest boundary between ECAL modules are also added to the input variables.
Not all the input variables are defined for each τ_h candidate: the strip and photon-related variables, for example, are defined only for the τ_h candidates reconstructed in one of the decay modes with neutral pions. The candidates are therefore split into different categories and a separate BDT is trained for each category. The categorization depends on the presence of a reconstructed electron in the vicinity of the τ_h candidate and on the number of strips associated with the reconstructed τ_h. Since some of the quantities listed above are expected to differ depending on the pseudorapidity, due to differences in the detector or in the reconstruction algorithms, a further categorization is performed between the τ_h candidates reconstructed in the barrel ($|\eta| < 1.479$) and the ones reconstructed in the detector endcap ($1.479 < |\eta| < 2.3$). In total, eight categories are used for the training of the anti-e discriminator. The number of input variables included in each BDT varies from 10 to 32, depending on the category.
Some of the most discriminating input variables are presented in figure 5.2, where the distributions, normalized to unit area, are shown for genuine τ_h candidates in simulated $Z/\gamma^* \to \tau\tau$ events and electrons in simulated $Z/\gamma^* \to ee$ events. The τ_h candidates entering in these distributions are required to pass the loose working point of the cut-based isolation and do not have to be matched with loosely identified electrons. Additionally, they must have $p_T > 20$ GeV, $|\eta| < 2.3$ and have to be reconstructed in one of the decay modes listed in the last column of table 5.1. Some characteristic features can be recognized in these distributions. In the visible mass distribution, for example, it can be seen how the first peak around the electron mass (at 0.5 MeV) is indeed populated almost exclusively by electrons, while the second peak at the pion rest mass (140 MeV) has a dominant τ_h contribution, although the electron contamination is non negligible. Electrons clearly tend to cluster towards higher values of the E_{ECAL}/p variable, while τ_h candidates have the opposite tendency. In the photon fraction distribution, electrons populate the region close to zero, while the shape for τ_h candidates is more uniform across the whole range.

The BDTs have been trained using 13 TeV simulated data samples, generated according to the expected LHC scenario for Run II. The training samples include simulated $Z/\gamma^* \to \tau\tau$, $Z/\gamma^* \to ee$, $W \to \tau\bar{\nu}_\tau$, $W \to e\bar{\nu}_e$, $t\bar{t}$, $H \to \tau\tau$ (both Standard Model Higgs and as predicted by supersymmetric models, for different mass hypotheses), $Z' \to \tau\tau$, $Z' \to ee$, $W' \to \tau\bar{\nu}_\tau$ and $W' \to e\bar{\nu}_e$ events. The Z' and W' samples, in particular, are important to cover the kinematic region where the τ_h candidates have high p_T (above 300 GeV). A basic introduction to the software libraries used for event simulation, including references to the generators used for producing the above mentioned samples is given later in the text, in section 5.4.1.
Reconstructed hadronic tau candidates are considered as "signal" ("background") in case they are matched within $\Delta R < 0.3$ to a generated hadronic tau decay (electron). Before training the BDT, the input signal and background candidates are required to pass the loose working point of the cut-based isolation. To avoid training against well reconstructed electrons, artificially improving the performance of the BDT, the τ_h candidates are required not to be matched with loosely identified electrons, within a distance $\Delta R < 0.3$. Candidates reconstructed in the uninstrumented region between ECAL barrel and endcap ($1.446 < |\eta| < 1.558$) are also rejected.
An exemplary distribution showing the output of one of the trained BDTs is presented in figure 5.3, for τ_h candidates in simulated $Z/\gamma^* \to \tau\tau$ events and electrons in simulated $Z/\gamma^* \to ee$ events. The BDT output consists of a real number which is closer to +1 if the object's properties resemble those of a genuine hadronic tau decay and closer to −1 otherwise. From the figure it is clear that the discriminating power of the BDT output is higher than the one provided by each individual variable of figure 5.2.
When training a binary classifier such as a BDT, particular care needs to be put in avoiding the

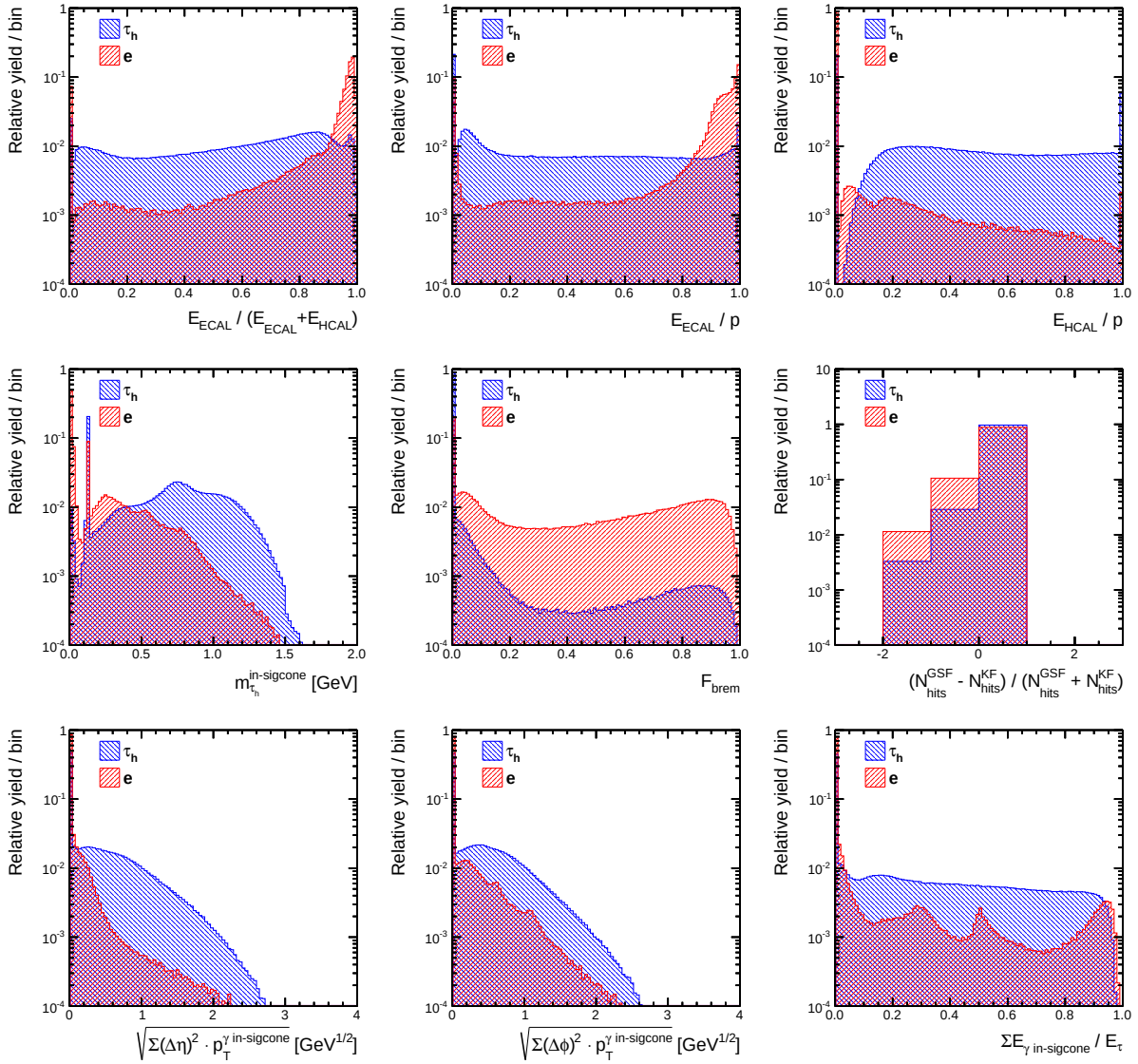


Figure 5.2: Distributions, normalized to unit area, of some of the observables used as input variables for the MVA-based anti-e discriminator, for τ_{h} candidates in simulated $\mathrm{Z}/\gamma^* \to \tau\tau$ events (blue) and electrons in simulated $\mathrm{Z}/\gamma^* \to \mathrm{ee}$ events (red). The τ_{h} candidates are required to pass the loose working point of the cut-based isolation and to be not matched with loosely identified electrons. They must have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 2.3$ (avoiding the transition region between ECAL barrel and endcap) and have to be reconstructed in one of the decay modes $h^{\pm}$, $h^{\pm}\pi^0$, $h^{\pm}\pi^0\pi^0$ or $h^{\pm}h^{\mp}h^{\pm}$.

situation where the classifier learns to model differences between signal and background which are specific to the particular training sample used. This effect, called "overtraining", results in a perfect classification performance on the training sample, which dramatically changes when applying the classifier on a separate set of events. The absence of overtraining for the BDT used for the anti-e discriminator is verified evaluating the discriminator on a test sample orthogonal to the one used for training. The separation power of a binary classifier is normally translated into a Receiver Operating Characteristic (ROC[1]) curve: a graphical representation illustrating the performance of the classifier in terms of signal efficiency and background rejection for various

[1]Not to be confused with the identical acronym used in chapter 3 and 4 referring to the front-end readout chips of the pixel detector.

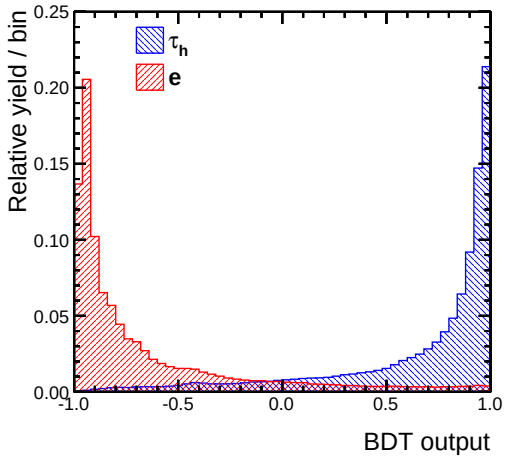


Figure 5.3: Output distribution, normalized to unit area, of the BDT trained on τ_h candidates reconstructed in one of the decay modes with π^0's and with pseudorapidity in the barrel region of the detector. The BDT is clearly able to separate τ_h candidates in simulated $Z/\gamma^* \to \tau\tau$ events (blue) from electrons in $Z/\gamma^* \to ee$ events (red).

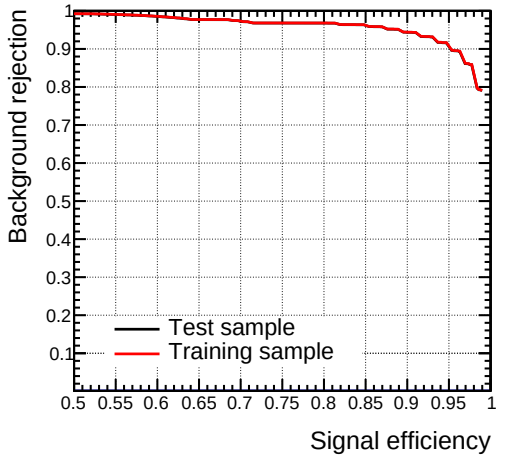


Figure 5.4: ROC curve for the MVA-based anti-e discriminator, evaluated from the training (red line) and test (black line, almost perfectly covered by the red one) samples. The ROC curve is drawn computing, for each value of the selection on the BDT output, the efficiency on signal (τ_h candidates matched to generated hadronic tau decays) and the rejection of background (τ_h candidates matched to generated electrons). The performance in terms of background rejection extracted from the ROC curve can not be directly compared with the values in table 5.2, due to the different sample admixture on which the discriminator is trained.

selections on the classifier's output. The ROC curves obtained from the BDT output of the MVA-based anti-e discriminator, evaluated on both the training and test samples, are shown in figure 5.4. From the perfect overlap between the two curves it can be concluded that the performance of the discriminator on the training sample can be reproduced also on an independent test sample, thus ensuring that the BDT is far from an overtraining condition.

Different working points are defined on the BDT output, according to the efficiency for hadronic taus to pass the discriminator. The selection on the BDT output for each working point is computed as a function of the τ_h candidate transverse momentum, in four coarse p_T-regions: below 60 GeV, between 60 GeV and 100 GeV, between 100 GeV and 200 GeV, above 200 GeV. This ensures that, for a given working point, the efficiency of the discriminator on tau decays remains reasonably constant also moving towards higher τ_h transverse momenta.
The expected performance of the multivariate discriminator is presented in figure 5.5, where the efficiency of the discriminator on τ_h decays is shown for each working point at different p_T-regimes, in simulated $Z/\gamma^* \to \tau\tau$ and $Z' \to \tau\tau$ events. The τ_h candidates have $p_T > 20$ GeV and $|\eta| < 2.3$ and are required to pass, in addition to the given working point of the discriminator, also the loose working point of the cut-based isolation. The $e \to \tau_h$ misidentification rate as a function of the electron transverse momentum is shown in figure 5.6, for simulated $Z/\gamma^* \to ee$ events where the electrons have $p_T > 20$ GeV and $|\eta| < 2.3$.
The numerical values for the efficiency and $e \to \tau_h$ misidentification rate of the individual working points are listed in table 5.2.

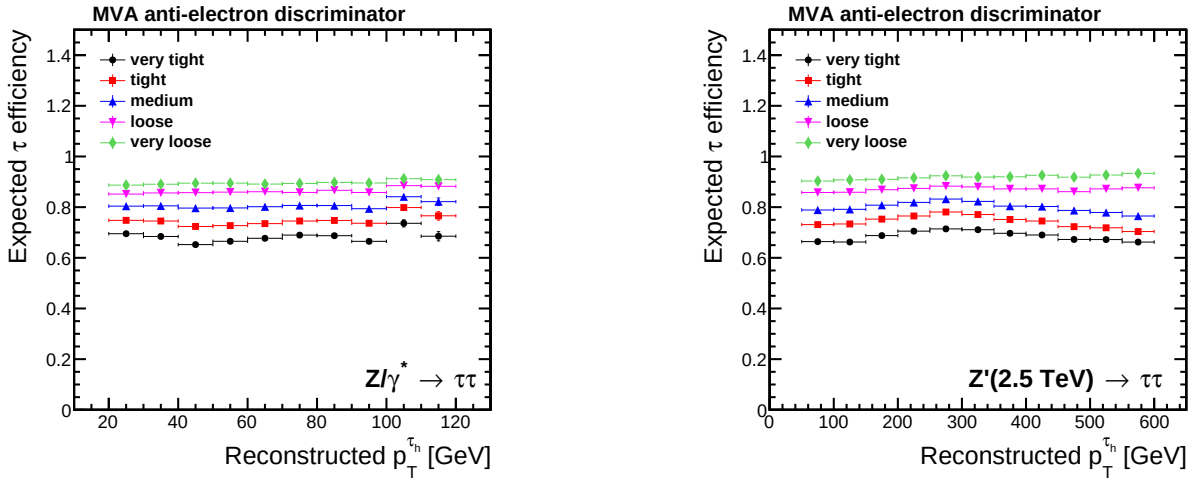


Figure 5.5: Expected efficiency for τ_h decays in simulated $Z/\gamma^* \to \tau\tau$ (left) and $Z' \to \tau\tau$ (right) events to pass the very loose, loose, medium, tight and very tight working point of the MVA-based anti-e discriminator. The τ_h candidates are required to be reconstructed in one of the decay modes $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$ or $h^\pm h^\mp h^\pm$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$ and to pass the loose working point of the cut-based isolation. The efficiency is shown as a function of the reconstructed τ_h transverse momentum.

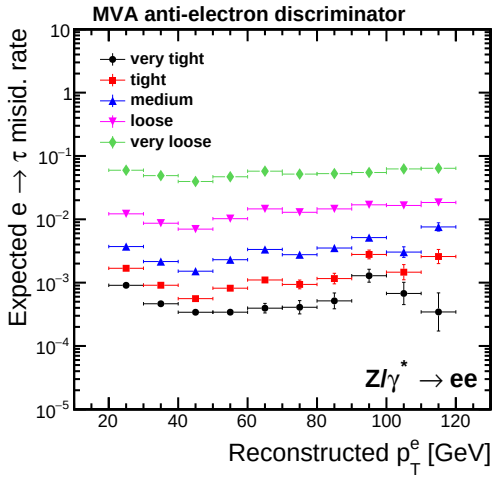


Figure 5.6: Expected probability for electrons of $p_T > 20$ GeV and $|\eta| < 2.3$ to pass the very loose, loose, medium, tight and very tight working point of the MVA-based anti-e discriminator in simulated $Z/\gamma^* \to ee$ events. The misidentification rate is shown as a function of the reconstructed electron transverse momentum.

A comparison between the previous version of the multivariate discriminator, as used during Run I, and the new version developed for Run II is shown in figures 5.7 and 5.8, for the tight working point of the discriminator. The same selections used in the plots of figure 5.5 and 5.6 are applied. The new discriminator has an efficiency on τ_h decays very close to the old one for transverse momenta up to 70 GeV, which tends to remain constant for higher p_T values, due to the implementation of p_T-dependent working points. To the large improvement observed at high p_T, as in the right plot of figure 5.7, contributes also the fact that the Run I discriminator has not been re-trained on 13 TeV samples. When applying this discriminator on 13 TeV simulated events an under-performance is therefore expected, especially in the kinematic regions which are more difficult to model. This also clearly illustrates the need of re-training the discriminators on samples which are closer to the expected experimental situations, as soon as the running conditions change or the reconstruction algorithms evolve. The misidentification rate, evaluated on a sample of simulated $Z/\gamma^* \to ee$ events, is expected to be approximately 40% lower, depending on the working point, compared to what was obtained using the previous version of the discriminator. The reduction is particularly clear in the region below 70 GeV, which corresponds

WP	Efficiency		$e \to \tau_h$ misidentification rate
	$Z/\gamma^* \to \tau\tau$	$Z'(2.5\,\text{TeV}) \to \tau\tau$	$Z/\gamma^* \to ee$
Very Loose	89.0%	92.8%	4.93×10^{-2}
Loose	85.5%	86.7%	0.97×10^{-2}
Medium	80.3%	78.0%	2.49×10^{-3}
Tight	74.3%	72.2%	1.02×10^{-3}
Very Tight	68.5%	66.3%	0.53×10^{-3}

Table 5.2: Expected efficiency and misidentification rate for the individual working points of the MVA-based anti-e discriminator. The statistical uncertainty on the numbers is negligible.

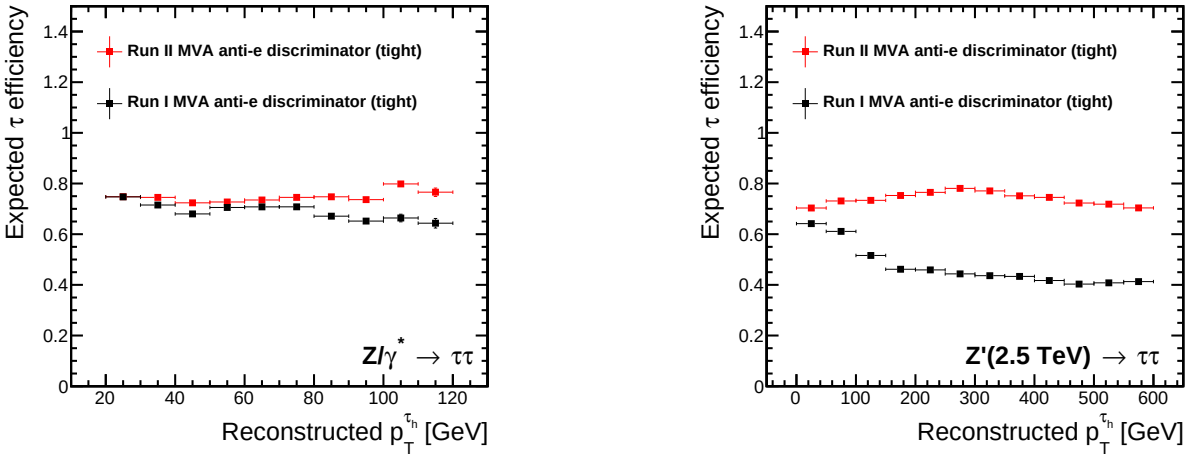


Figure 5.7: Expected efficiency for τ_h decays in simulated $Z/\gamma^* \to \tau\tau$ (left) and $Z' \to \tau\tau$ (right) events to pass the tight working point of the MVA-based anti-e discriminator, for the version used during Run I (black points) and the version developed for Run II (red points). The τ_h candidates are required to be reconstructed in one of the decay modes $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$ or $h^\pm h^\mp h^\pm$, to satisfy $p_T > 20$ GeV and $|\eta| < 2.3$ and to pass the loose working point of the cut-based isolation. The efficiency is shown as a function of the reconstructed τ_h transverse momentum.

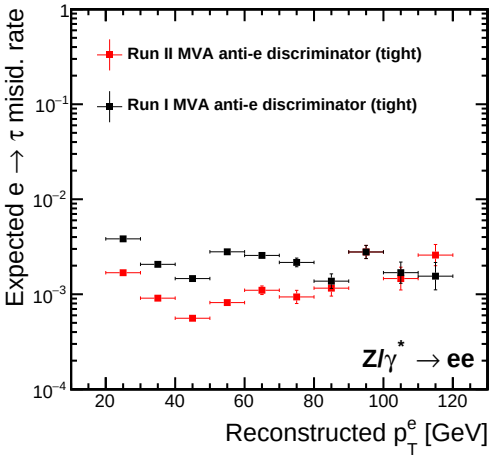


Figure 5.8: Expected probability for electrons of $p_T > 20$ GeV and $|\eta| < 2.3$ to pass the tight working point of the MVA-based anti-e discriminator in simulated $Z/\gamma^* \to ee$ events, for the version of the discriminator used during Run I (black points) and the version developed for Run II (red points). The misidentification rate is shown as a function of the reconstructed electron transverse momentum.

to the typical kinematic range for τ_h candidates coming from $Z/\gamma^* \to \tau\tau$ decays.
A measurement of the $e \to \tau_h$ misidentification rate in the data collected by CMS during the 2015 proton-proton run is presented later in the text, in section 5.5.

5.4 Drell-Yan event selection

The performance of the τ_h reconstruction and identification algorithms can be evaluated in data on events where Z bosons are produced in the scattering of two partons (Drell-Yan process). Leptonic decays coming from a Z boson, whose mass and branching ratios have been measured with extreme precision in e^+e^- colliders such as LEP (at CERN) and SLC (at the Stanford Linear Accelerator Center), offer a well-known reference which can be exploited to test the performance of the τ_h identification in CMS. This section lists the background processes which mimic a Drell-Yan signature and describes the selections applied for the identification of the physics objects used in the analysis, as well as the corrections applied to the simulation to correct for known mismodelling effects.

5.4.1 Event topology and data samples

Z bosons are produced at the LHC via the Drell-Yan process [98]. The leading order production mechanism is $q\bar{q} \to Z/\gamma^*$. As shown in table 5.3, Z bosons decay into opposite-charge pairs of leptons (ee, $\mu\mu$, $\tau\tau$) with a branching ratio of approximately 3.4% for each lepton flavour, into a pair of light quarks hadronizing into jets (branching ratio $\sim 70\%$) or into three generations of neutrino pairs (branching ratio $\sim 20\%$) [11].
In the $\tau\tau$ decay mode, each of the τ leptons decays, in turn, into purely leptonic final states ($\tau \to e\,\overline{\nu}_e\,\nu_\tau$ or $\tau \to \mu\,\overline{\nu}_\mu\,\nu_\tau$) or into hadronic final states τ_h. The combination of all decay modes gives the six possible final states which are reported, together with their relative branching ratios, in table 5.4.

Decay Mode	**BR [%]**
$Z \to q\bar{q}$	69.9
$Z \to \nu\bar{\nu}$	20.0
$Z \to e^+e^-$	3.4
$Z \to \mu^+\mu^-$	3.4
$Z \to \tau^+\tau^-$	3.4

Table 5.3: Branching ratios for all Z decay modes, including hadronic and invisible decays [11].

Decay Mode	**BR [%]**
$Z \to \tau^+\tau^- \to e^+e^-\,\nu_e\,\bar{\nu}_\tau\,\overline{\nu}_e\nu_\tau$	3.1
$Z \to \tau^+\tau^- \to \mu^+\mu^-\,\nu_\mu\,\bar{\nu}_\tau\,\overline{\nu}_\mu\nu_\tau$	3.0
$Z \to \tau^+\tau^- \to e^+\mu^-\,\nu_e\,\bar{\nu}_\tau\,\overline{\nu}_\mu\nu_\tau$	6.2
$Z \to \tau^+\tau^- \to e^+\tau_h^-\,\nu_e\,\bar{\nu}_\tau\,\nu_\tau$	23.1
$Z \to \tau^+\tau^- \to \mu^+\tau_h^-\,\nu_\mu\,\bar{\nu}_\tau\,\nu_\tau$	22.6
$Z \to \tau^+\tau^- \to \tau_h^+\tau_h^-\,\bar{\nu}_\tau\,\nu_\tau$	42.0

Table 5.4: Branching ratios for all the $Z \to \tau\tau$ decay modes [1]. Charge-conjugation in the final states is implied.

From now on, the undetectable neutrinos in the final state will be omitted in favour of a simplified notation: $\ell\ell$, $\ell\tau_h$ (with $\ell = e, \mu$) and $\tau_h\tau_h$.
Standard Model processes such as W+jets, $t\bar{t}$, single-top, diboson (WW, WZ, ZZ) and QCD multijet production have event topologies which can resemble that of a Drell-Yan process, therefore acting as background sources.

- **W($\to \ell\nu$)+jets**: the production of a single W vector boson decaying leptonically in association with additional jets due to quark or gluon radiation is an important source of background in the eτ_h and $\mu\tau_\mathrm{h}$ final states, if one jet is misidentified as a τ_h candidate.
- **$\mathrm{t\bar{t}}$**: top quark pairs are produced at LHC via the gluon fusion process $\mathrm{gg} \to \mathrm{t\bar{t}}$ or via quark annihilation $\mathrm{q\bar{q}} \to \mathrm{t\bar{t}}$ and decay with a branching ratio very close to unity into $\mathrm{t} \to \mathrm{Wb}$. The fully-leptonic final states where the two W bosons decay, respectively, into an electron and a muon plus a genuine source of missing transverse energy are the main source of background in the eμ final state.
- **Diboson**: the production of vector boson pairs (WW, ZZ, WZ) constitutes a small source of background, for example when the W or Z bosons decay into leptons or when the jet coming from hadronic decays is misidentified as a τ_h candidate.
- **Single-top**: single-top production is considered as a source of background together with the above ones. It should be however noted that this contribution is smaller compared to $\mathrm{t\bar{t}}$ or diboson production.
- **QCD multijet**: under this common definition, all the processes leading to final states with a large jet multiplicity are grouped. Although in this kind of events the only genuine leptons are the ones coming from hadron decays, it is still possible that one or more jets get misidentified as τ_h candidates or as leptons. Therefore these processes constitute another important source of background in the eτ_h and $\mu\tau_\mathrm{h}$ final states. Despite the fact that the misreconstruction probability is quite low, the background yield is still sizeable due to the huge multijet production cross section at hadron colliders (see also figure 1.3).

Data samples

The measurements described in the following sections have been performed in the data collected by the CMS experiment during 2015, when the LHC collider has been operated with a 25 ns separation between proton bunches. This data-taking period is identified as `Run2015D` and the total analyzed data sample corresponds to an integrated luminosity of $2.3\,\mathrm{fb}^{-1}$. For reference, the analyzed data samples and the corresponding run periods are summarized in table A.1 in the appendix.

Simulated samples

Simulated events have been generated with event generators based on Monte Carlo techniques. These are software libraries which simulate high-energy collisions, modelling the physics at very short distance scales (whose treatment is based on perturbation theory) as well as hadron formation and decay.
Drell-Yan events, $\mathrm{Z}/\gamma^* \to \ell\ell$ (with $\ell = \mathrm{e}, \mu, \tau$), have been generated with MadGraph5 [99]. Parton shower and hadronization have been performed with pythia8 [100] using the cuetp8m1 tune [101]. The same generator and tune have been used for simulating the process in which a W boson is produced in association with jets. Single-top and $\mathrm{t\bar{t}}$ events have been generated using powheg [102–104] interfaced to pythia8 for parton shower and hadron formation. Background samples for electroweak diboson production have been generated with MadGraph5_aMC@NLO using the FxFx merging scheme [105, 106]. All the generated events have been processed through a detailed simulation of the CMS detector which is based on the geant4 [107] software and reconstructed using the same algorithms and codes which are used

for data.
A complete list of all the Monte Carlo samples which have been used to simulate the signal and background contributions in the measurements presented in the following sections can be found in table A.2 and table A.3 in the appendix.

QCD multijet production is not included in the above list of simulated samples despite its importance, especially in the $\ell\tau_{\mathrm{h}}$ and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ final states. The reason is that, due to the huge production cross section and to finite computing resources, it is in general impractical, if not impossible, to simulate a sufficient number of QCD multijet events surviving all the analysis selections. Even overcoming this technical limitation would not solve the problem completely, since the analysis selections are aiming at a small portion of the total phase space, where the overall correctness of the simulation is not granted. The QCD background estimation, therefore, does not rely on the simulation but is based on a purely data-driven approach, as described in more detail in section 5.4.4.

5.4.2 Object identification

Moving towards a detailed description of the measurements performed to validate the τ identification performance in data, the following paragraphs describe the selections applied for the identification of the physics objects used in the analysis.

Vertices

The offline reconstruction of primary vertices begins with the reconstructed tracks, which are selected based on their compatibility with the beam spot, the number of hits in the tracker and the quality of the fits used to form track seeds out of the individual hits in the tracker detector. The tracks are then clustered into several primary vertex candidates, according to the z coordinate of the point of closest approach of the tracks to the z axis. A three-dimensional vertex fit is performed using the tracks of each primary vertex candidate, retaining only the candidates whose position is compatible with the beam line. Finally, the selected primary vertex of the event (PV) is defined as the vertex with the highest $\sum_i p_{\mathrm{T},i}^2$, where the index i runs on all the tracks associated with the vertex.

Electrons

Electrons are reconstructed following the particle-flow approach, as described in section 2.2.2. The electron candidates have to pass several requirements in order to be identified as genuine electrons.

- An identification variable is computed as the output of a multivariate classifier, which combines different properties of the electron candidate's track and electromagnetic energy deposit, among which: the ECAL cluster shape and width, the ratio between the energy deposit in ECAL and HCAL, the number of track hits and the quality of the track fit, plus a set of kinematic variables such as the ratio between energy and momentum. The classifier is trained on all electron candidates, irrespective of the trigger requirement, with the aim of discriminating between genuine electrons and objects (jets, for example) mimicking an electron signature. The value of this multivariate identification variable (referred to as

non-triggering MVA ID) is required to be above a tight working point having an efficiency of roughly 80% on genuine electrons.

- In order to make sure that the electron candidates come from the primary hard interaction and not from one of the many pile-up vertices, the z coordinate of the primary vertex and the z coordinate of the electron track vertex, evaluated with respect to the beam line, are required to lie within a distance of less than 0.2 cm. A selection on the absolute value of the transverse impact parameter calculated with respect to the primary vertex is also applied, requiring it to be smaller than 0.045 cm.
- In order to reject events in which the electron originates from a conversion of a photon into an e^+e^- pair inside the tracker material, the number of missed inner tracker layers of the electron track is required to be smaller or equal to one. In addition, any event in which the electron candidate is close in space to a partner track compatible with a photon conversion is rejected.

Muons

Muons are reconstructed following the particle-flow approach, as described in section 2.2.3. The muon candidates have to pass several requirements in order to be identified as genuine muons.

- It is necessary for the muon candidate to be identified as a muon by the particle-flow event reconstruction and to be reconstructed either as a global muon (i.e., after having matched the track reconstructed in the muon detector with the tracks reconstructed in the inner tracker) or as a tracker muon (i.e., considering all tracker tracks to be muon candidates and after having tested this hypothesis looking for compatible signatures in the calorimeters and in the muon detector). Additional track-quality and muon-quality selections are applied. This set of requirements is referred to as medium muon ID.
- The z coordinate of the primary vertex and the z coordinate of the muon inner track vertex, evaluated with respect to the beam line, are required to lie within a distance of less than 0.2 cm. A selection on the absolute value of the transverse impact parameter calculated with respect to the primary vertex is also applied, requiring it to be smaller than 0.045 cm.

An *absolute isolation* variable expressing the amount of charged or neutral activity around a generic lepton ℓ, can be defined as:

$$\mathrm{I}^{\ell} \equiv \sum_{\text{charged}} p_{\mathrm{T}} + \max\left(0, \sum_{\text{neutral}} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} - \frac{1}{2} \sum_{\text{charged, PU}} p_{\mathrm{T}}\right), \tag{5.2}$$

where $\sum_{\text{charged}} p_{\mathrm{T}}$ is the scalar sum of the transverse momenta of all electrons, muons and charged hadrons coming from the primary vertex of the event and contained in an isolation cone of size $R = 0.3$ centered on the lepton direction. The sums $\sum_{\text{neutral}} p_{\mathrm{T}}$ and $\sum_{\gamma} p_{\mathrm{T}}$ correspond to the same quantity (without vertex association), computed from the neutral hadrons and from the photons contained in the cone. The last sum in the expression in brackets is an estimate of the contribution of the neutral components (neutral hadrons and photons) coming from pile-up interactions. These are obtained from the charged pile-up component, using a correction factor $\frac{1}{2}$ corresponding to the ratio of neutral to charged hadron production in the hadronization process

of inelastic pp collisions, as estimated from simulation.
A *relative isolation* variable can be defined dividing equation (5.2) by the lepton transverse momentum:

$$\mathrm{I}^{\ell}_{\mathrm{rel}} \equiv \mathrm{I}^{\ell}/p^{\ell}_{\mathrm{T}}\,. \tag{5.3}$$

Isolation variables are a powerful handle to discriminate genuine electrons or muons from jets mimicking a leptonic signature in the detector, since the latter are much less isolated.

τ_{h} candidates

Hadronically decaying taus are reconstructed by the HPS algorithm explained in section 5.2. The τ_{h} candidates are required to pass several requirements in order to be identified as genuine hadronic taus.

- Hadronic tau candidates are required to be reconstructed in one of the decay modes $h^{\pm}$, $h^{\pm}\pi^0$, $h^{\pm}\pi^0\pi^0$ or $h^{\pm}h^{\mp}h^{\pm}$.
- The z coordinate of the τ_{h} candidate's leading track, evaluated with respect to the primary vertex, is required to lie within a distance of less than 0.2 cm from the primary vertex.

The discriminators which are applied on the τ_{h} candidate depend on the type of measurement which is performed since the background composition, especially in terms of lepton misidentification, changes according to the final state. The exact dicriminators and working points which have been used for the measurements presented in this thesis are listed explicitly in the corresponding sections.

Jets

Jets are reconstructed clustering the charged and neutral hadron candidates obtained from the particle-flow event reconstruction. The anti-k_{T} clustering algorithm described in section 2.2.4 is used, with the R parameter equal to 0.4. The complete set of recommended corrections are applied both in data and in the simulation. Residual corrections are applied on data to account for additional data/simulation differences. Only jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 4.7$ are considered.
To account for the fact that some genuine leptons (electrons, muons or τ_{h} candidates) might also be misidentified as jets, the distance ΔR between the jet and the leptons identified by the above mentioned criteria is required to be equal or larger than 0.5.

b-tagged jets

A jet is considered to be initiated by a bottom quark if the output of the Combined Secondary Vertex algorithm described in section 2.2.4 is above the medium working point. This working point is defined as the value of the selection applied on the discriminator for which the rate for misidentifying a jet initiated by a light quark as a b-jet is 1%. Given the fact that this b-tagging algorithm relies on track-based variables, the absolute value of the b-tagged jet pseudorapidity $|\eta|$ is required to be inside the tracker acceptance, namely smaller than 2.4.

Missing transverse energy

The missing transverse energy is computed for each event using a multivariate regression algorithm intended to reduce the effect of pile-up interactions on the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, as discussed briefly in section 2.2.5 (MVA $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$). The MVA has been trained on simulated events, generated for a center of mass energy of 13 TeV and a 25 ns bunch crossing period.

5.4.3 Data/simulation corrections

Pile-up description

The expected number of additional pile-up (PU) interactions in a given bunch crossing i is expressed by equation (1.38), where a value of 69 mb (with an estimated uncertainty of 5%) is assumed for $\sigma_{\text{min. bias}}$, as measured in CMS by a dedicated analysis.
The presence of additional interactions is taken into account in the simulation, superimposing simulated inelastic pp collisions to the hard scattering process in each event. The actual number of PU interactions is randomly taken from a Poisson distribution having the variable $N_{\mathrm{PU}}^{\mathrm{true}}$ of equation (1.38) as mean parameter. Although the simulation tries to reproduce the actual running conditions of the experiment as close as possible, it is nevertheless necessary to reweight the simulated events, in order to account for residual differences and match the two distributions exactly.
The effectiveness of the reweighting procedure is verified by the agreement between data and simulation in the distribution of the number of reconstructed primary vertices for each event (N_{PV}): an observable which is strongly correlated with the number of additional pile-up interactions.
The N_{PV} distribution is shown in figure 5.9, before and after applying the reweighting procedure, for a sample of events selected requiring the presence of a $\mu\tau_{\mathrm{h}}$ pair in the final state. Muons are selected requiring, in addition to the identification criteria of section 5.4.2, $p_{\mathrm{T}} > 19\,\mathrm{GeV}$, $|\eta| < 2.1$ and relative isolation $\mathrm{I}_{\mathrm{rel}}^{\mu} < 0.1$. The τ_{h} candidates are required to pass the tight working point of the cut-based discriminator against $\mu \to \tau_{\mathrm{h}}$ misidentification and the tight working point of the MVA isolation discriminator and to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$. The average number of reconstructed primary vertices $\langle N_{\mathrm{PV}} \rangle$ is around 10, while events with up to 30 reconstructed primary vertices are observed. From the distributions in the figure, it can be clearly seen that the agreement between data and simulation improves after the reweighting procedure, especially for events with a number of reconstructed vertices larger than 15.

Identification and isolation efficiencies

The efficiency of the identification and isolation requirements applied on electrons and muons is measured in data via a *tag-and-probe* technique: $\mathrm{Z}/\gamma^* \to \ell\ell$ (ℓ = e, μ) events are selected requiring the presence of a well-identified and isolated lepton ("tag") and of an additional loosely identified lepton ("probe") in the final state. Further selections, based on the electric charge and on the invariant mass of the pair, are applied on the tag-probe system in order to increase the probability that both objects originate from the same resonance. The efficiency of a given selection is then defined as the fraction of probes passing the identification or isolation selection under study with respect to the total number of probes:

$$\epsilon(p_{\mathrm{T}}, \eta) = \frac{\#\text{ probes passing Id/Iso requirement}}{\#\text{ all probes}}\,. \tag{5.4}$$

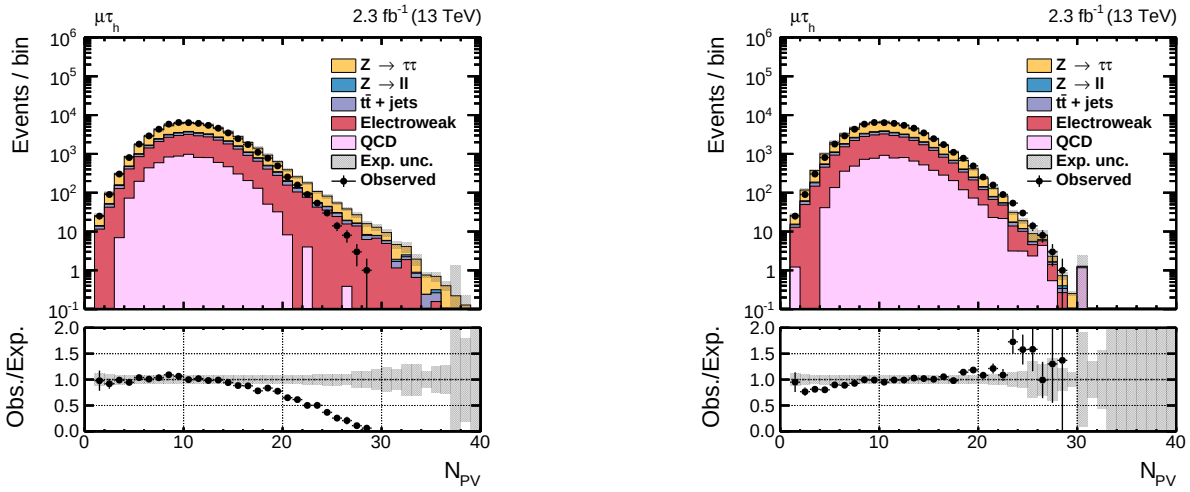


Figure 5.9: Distribution of the number of primary vertices, before (left) and after (right) reweighting the simulated events according to the value of $N_{\mathrm{PU}}^{\mathrm{true}}$. Events are selected requiring the presence of an oppositely-charged pair made of: a muon with $p_{\mathrm{T}} > 19\,\mathrm{GeV}$, $|\eta| < 2.1$, satisfying the identification criteria of section 5.4.2 and with $\mathrm{I}_{\mathrm{rel}}^{\mu} < 0.1$ and of a τ_{h} with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.3$, passing the tight WP of the cut-based discriminator against $\mu \to \tau_{\mathrm{h}}$ misidentification and the tight WP of the MVA isolation discriminator. Backgrounds are estimated according to the prescription of section 5.4.4. The electroweak background contribution includes events from W+jets, diboson and single-top production.

The number of probes in the numerator and denominator of equation (5.4) is extracted from a fit on the invariant mass of the dilepton tag-probe pair, since this variable gives a clear separation between the resonant Drell-Yan signal and the smoothly falling background. The signal is modelled in the fit by two asymmetric Gaussian functions, while the background is parametrized by an exponential function.
The very same measurement is performed on simulated events and the resulting efficiency is compared to the one measured in data. The ratio between the two is translated into data/simulation scale factors, parametrized according to the transverse momentum and pseudorapidity of the probe lepton: $w_{\mathrm{Id+Iso}}(p_{\mathrm{T}}, \eta) = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{simulation}}$. These scale factors are then used to correct the simulated events on an event-by-event basis.
Exemplary efficiency curves, as measured in data and on simulated events, are shown in figure 5.10, for probe electrons and probe muons in the barrel region of the detector passing the identification criteria of section 5.4.2 in addition to a selection on the relative isolation variable, namely: $\mathrm{I}_{\mathrm{rel}}^{\ell} < 0.1$ (with $\ell = \mathrm{e}, \mu$). The efficiency curves have a typical "turn-on" behaviour, where the efficiency increases as a function of the transverse momentum and reaches a constant value ("plateau") close to unity.
At the plateau, the efficiencies measured in data range from 86% to 90% (for electrons) and from 96% to 98% (for muons), depending on the pseudorapidity region in which the measurement is performed. The extracted identification and isolation scale factors range from 0.92 to 0.98, depending on the p_{T} and pseudorapidity of the probe lepton.

Trigger efficiencies

Data events contained in the `Run2015D` dataset have been stored after passing a set of Level-1 and HLT trigger decisions applied during data-taking. These triggers are usually designed to retain events based on the presence in the final state of online-identified leptons (electron, muon or τ_{h} candidates) which are isolated and have a transverse momentum above a certain trigger

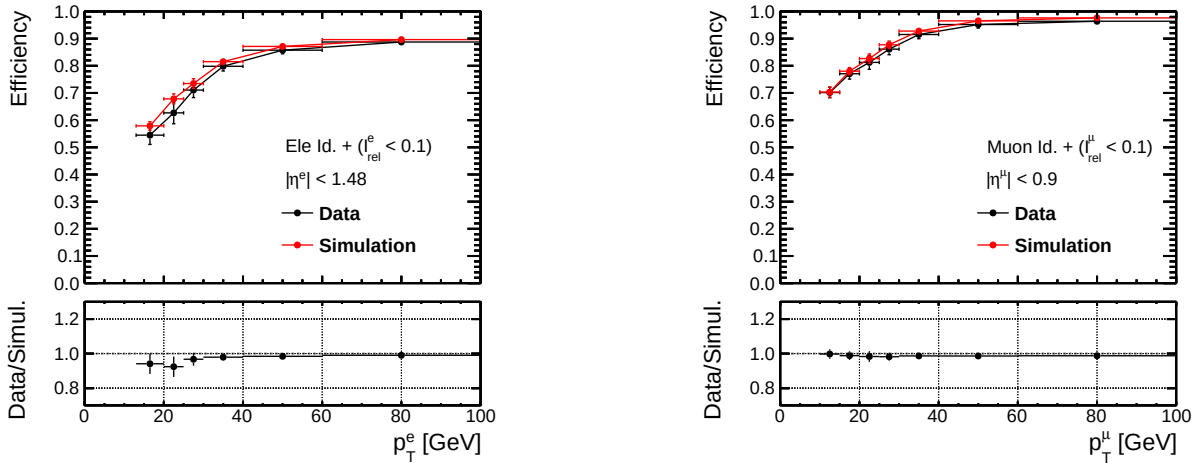


Figure 5.10: Efficiency, measured in data and simulation, to pass the electron/muon identification criteria of section 5.4.2 and $I^{\ell}_{\text{rel}} < 0.1$, for electrons (left) and muons (right) with $p^{\text{e}}_{\text{T}} > 13\,\text{GeV}$ and $p^{\mu}_{\text{T}} > 10\,\text{GeV}$, respectively, and pseudorapidity in the barrel region of the detector. The ratio in the bottom part of the plot represents the scale factor, as a function of the probe's transverse momentum. The horizontal error bars of the points correspond to the different p_{T}-regions in which the measurement is performed. The line connecting the points is a simple linear interpolation between the measurements.

threshold.
A set of triggers as close as possible to the ones deployed during data-taking is included in the simulation, such that trigger requirements can also be applied when running on simulated events. The actual triggers used for the measurements performed in the following sections depend on the targeted final state and will be discussed later. In general, when analyzing the data collected during 2015, unprescaled[2] single lepton triggers with the lowest possible p_{T}-threshold have been employed.
The efficiency of a given trigger is measured in data via a tag-and-probe technique very similar to the one used for measuring the identification and isolation efficiencies. In the simpler case of a single lepton trigger, tag leptons are required to satisfy some tight identification and isolation selections and to pass a "monitoring" trigger having a p_{T}-threshold higher than the trigger under study. This ensures that the leptons are in a kinematic region in which the trigger having a lower p_{T}-threshold is fully efficient, thus providing an unbiased measurement of the probe lepton efficiency. Probe leptons are required to pass the nominal offline identification and isolation selections (usually the same applied on the tag lepton). The fraction of probes passing also the tested trigger requirement defines the efficiency, as:

$$\epsilon(p_{\text{T}}, \eta) = \frac{\#\text{ probes passing trigger requirement}}{\#\text{ all probes}}\,. \tag{5.5}$$

Also in this case, the number of probes in the numerator and denominator of equation (5.5) is extracted from a fit on the invariant mass of the dilepton tag-probe pair.
After performing the same measurement on simulated events, the ratio between data and simulation efficiencies is translated into a set of scale-factors, parametrized as a function of transverse momentum and pseudorapidity of the probe lepton: $w_{\text{trig}}(p_{\text{T}}, \eta) = \epsilon_{\text{data}}/\epsilon_{\text{simulation}}$. These scale factors are then used to correct the simulated events on an event-by-event basis.

[2] A given trigger is said to be "prescaled", with prescale factor n, if the pass decision is taken only once every n events satisfying the criteria of the trigger. Prescaled triggers are useful for collecting samples of particularly high-rate triggers without clogging the data acquisition system.

Exemplary efficiency curves, as measured in data and in simulation, are shown in figure 5.11 for probe electrons and probe muons passing, respectively, the `HLT_Ele23_WPLoose_Gsf` and the `HLT_IsoMu18` single lepton triggers and with pseudorapidity in the barrel region of the detector. Both curves show a behaviour similar to what has been seen for the identification and isolation efficiencies, with the turn-on around the trigger p_{T}-threshold. Due to resolution effects, the turn-on has a behaviour which can be approximated by an error function.
At the plateau, the efficiencies of the single electron and muon triggers measured in data are above 90%. The extracted trigger scale factors range from 0.95 to 0.99, depending on the p_{T} and pseudorapidity of the probe.

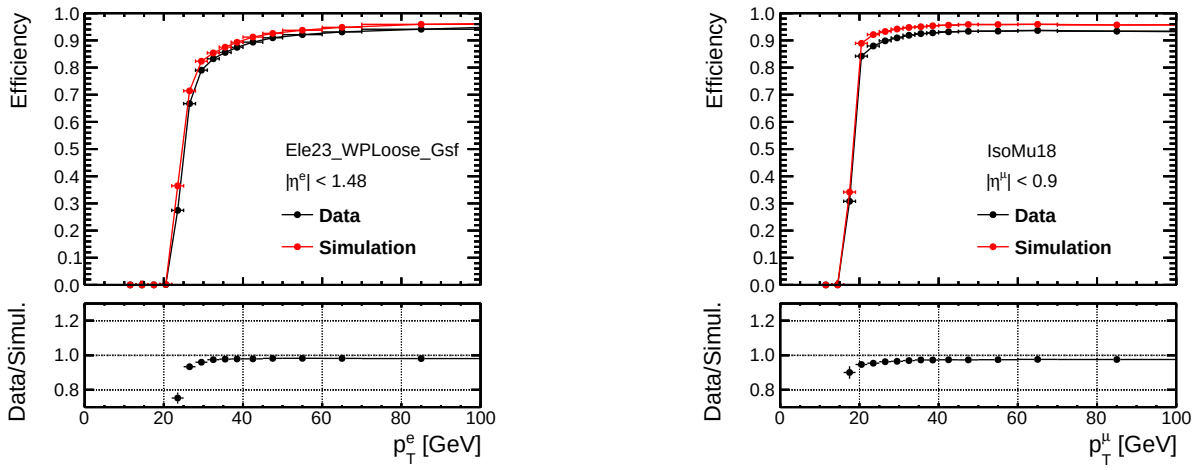


Figure 5.11: Efficiency, measured in data and simulation, to pass the `HLT_Ele23_WPLoose_Gsf` single electron trigger (left) and the `HLT_IsoMu18` single muon trigger (right) for probes having a pseudorapidity in the barrel region of the detector. The ratio in the bottom part of the plot represents the scale factor, as a function of the probe's transverse momentum. The horizontal error bars of the points correspond to the different p_{T}-regions in which the measurement is performed. The line connecting the points is a simple linear interpolation between the measurements.

5.4.4 Background estimation methods

When constructing the distribution of a given observable, the shape for the Drell-Yan and all other background processes, except QCD, are taken from the simulation. The normalization is determined from the cross section of the specific sample, as reported in table A.2 and table A.3, applying to each simulated event a weight computed as:

$$w = \frac{\sigma \cdot \epsilon_{\mathrm{sim.}}}{N^{\mathrm{gen}}_{\mathrm{sim.}}}\,, \tag{5.6}$$

where $N^{\mathrm{gen}}_{\mathrm{sim.}}$ is the total number of generated events and $\epsilon_{\mathrm{sim.}}$ is the efficiency of the selections which might be applied during event generation.
An exception is given by the W+jets background, whose contribution is derived from data using a W+jets enriched control region.
This enriched region is defined based on the fact that the genuine missing transverse energy $\vec{E}^{\mathrm{miss}}_{\mathrm{T}}$ coming from neutrinos produced in τ decays tends to be collinear with the visible products of the decay itself. In W+jets events where one jet is misidentified as a τ_{h}, however, due to the high mass of the W boson, the neutrino (and therefore the reconstructed $\vec{E}^{\mathrm{miss}}_{\mathrm{T}}$) is preferably

opposite to the lepton direction in the transverse plane. These considerations can be translated into a *transverse mass* variable, defined as:

$$m_{\mathrm{T}} \equiv \sqrt{2\, p_{\mathrm{T}}^{\ell}\, E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos(\Delta\phi))}\,, \tag{5.7}$$

where p_{T}^{ℓ} is the transverse momentum of the lepton and $\Delta\phi$ is the difference in the azimuthal angle ϕ between the lepton direction and the direction of the missing transverse energy vector $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. As can be seen from figure 5.12, where the shapes are normalized to unit area, this variable has a clear separation power between the $\mathrm{Z}/\gamma^{*} \to \tau\tau$ contribution and the W+jets background. Transverse mass distributions for the event selections described later in the text, showing the contribution of the W+jets background with respect to the others, are presented in figures D.2 and D.3 in the appendix.

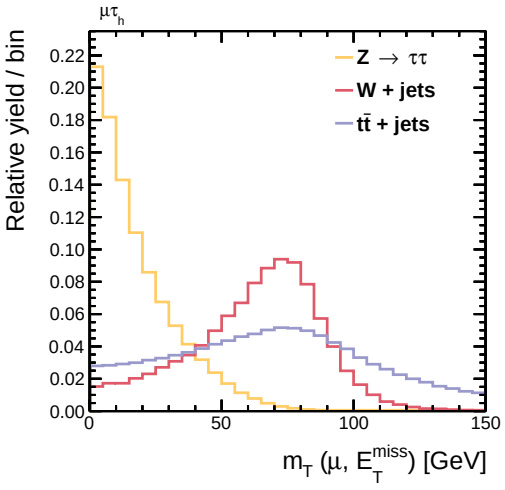


Figure 5.12: Distribution, normalized to unit area, of the transverse mass variable m_{T} in simulated $\mathrm{Z}/\gamma^{*} \to \tau\tau$, W+jets and $\mathrm{t\bar{t}}$ events. The events are selected requiring the presence of an oppositely-charged pair made of a muon with $p_{\mathrm{T}} > 19\,\mathrm{GeV}$, $|\eta| < 2.1$, satisfying the identification criteria of section 5.4.2 and with $\mathrm{I}_{\mathrm{rel}}^{\mu} < 0.1$ and of a τ_{h} with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.3$, passing the tight WP of the cut-based discriminator against $\mu \to \tau_{\mathrm{h}}$ misidentification and the tight WP of the MVA isolation discriminator.

A W+jets enriched control region is defined requiring the events to have a transverse mass above 70 GeV. The number of W+jets events in the control region is computed from data, after subtracting the remaining backgrounds. The yield is then translated to the low-m_{T} region applying a scale factor between low-m_{T} region and control region which is derived from the simulation.

The shape and normalization of the QCD multijet background are extracted from data in a control region in which the objects of the selected $\tau\tau$ pair have the same electric charge. The contribution of all other backgrounds, as estimated from the simulation (including a data-driven yield correction, for the W+jets background), is subtracted from the same-sign control region in this procedure. This control region is shown in the left plot of figure 5.13, for a sample of events selected requiring the presence in the final state of a $\mu\tau_{\mathrm{h}}$ pair with the same electric charge. Muons are required to have $p_{\mathrm{T}} > 19\,\mathrm{GeV}$, $|\eta| < 2.1$, to satisfy the identification criteria of section 5.4.2 and $\mathrm{I}_{\mathrm{rel}}^{\mu} < 0.1$, while the τ_{h} candidates must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.3$, pass the tight working point of the cut-based discriminator against $\mu \to \tau_{\mathrm{h}}$ misidentification and the tight working point of the MVA isolation discriminator. By construction, the agreement between data and the sum of all backgrounds in this control region is perfect.
The final QCD background contribution in the opposite-sign signal region is then obtained applying a scale factor computed in a QCD-enriched control region where the electrons and muons are required to be loosely isolated ($0.1 < \mathrm{I}_{\mathrm{rel}}^{\mathrm{e}/\mu} < 0.2$). The right plot in figure 5.13 shows the distribution of the invariant mass computed from the muon and the visible products of the hadronic tau decay in this QCD-enriched region, after having corrected the QCD yield

by the above mentioned scale factor. The satisfactory agreement between data and the sum of all processes proves that no substantial shape-altering effects are introduced in the QCD distribution, when going from the same-sign to the opposite-sign region.

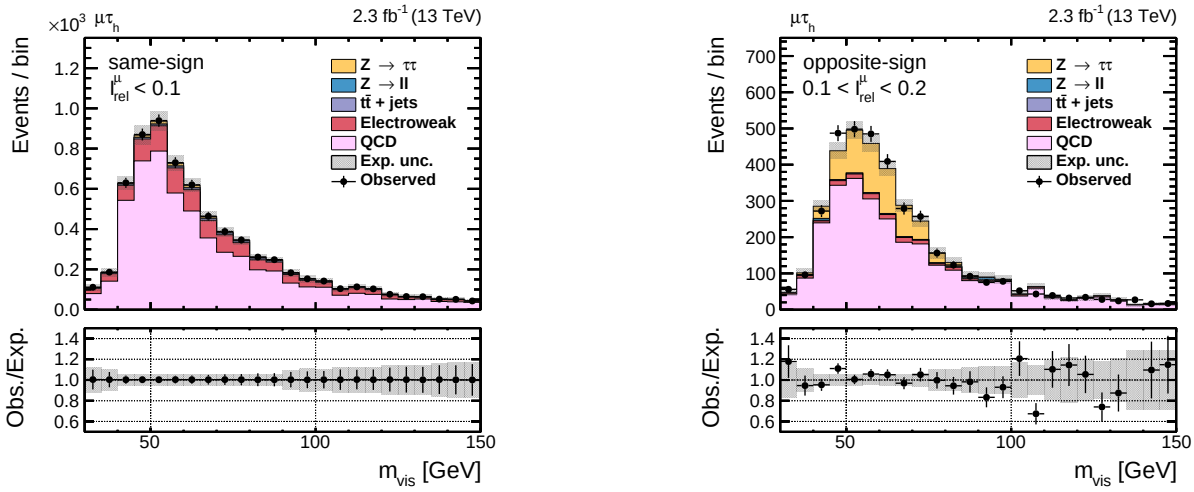


Figure 5.13: Distribution of the invariant mass computed from the muon and the visible products of the hadronic tau decay. Left: control region where the muon and τ_h candidate have the same electric charge. The QCD contribution is extracted from data in this region and therefore the agreement between data and the sum of all backgrounds is, by construction, perfect. Right: control region where the muon and τ_h candidate have opposite electric charge and the muon is loosely isolated ($0.1 < I^{\mu}_{rel} < 0.2$). Events are selected requiring a muon with $p_T > 19\,GeV$, $|\eta| < 2.1$, satisfying the identification criteria of section 5.4.2 and of a τ_h with $p_T > 20\,GeV$, $|\eta| < 2.3$, passing the tight WP of the cut-based anti-μ discriminator and the tight WP of the MVA isolation discriminator. The electroweak background includes events from W+jets, diboson and single-top production.

5.4.5 Signal extraction

The measurements presented in the following are based on fitting the distribution of some discriminating variable observed in data with shape templates for signal and background processes, in order to extract an estimate for the signal contribution. The chosen discriminating variable is typically a mass variable, although other choices are possible. The reconstruction of the invariant mass of $\tau\tau$ decays ($m_{\tau\tau}$) is complicated by the presence of at least two neutrinos in the final state. The algorithm developed at CMS for the $\tau\tau$ invariant mass reconstruction relies on a likelihood-based approach where the information about the missing transverse energy E_T^{miss} and its resolution is combined with the momenta of the visible τ decay products. More details about this algorithm, referred to as SVFIT, can be found in [89].
The basic idea consists in constraining the six unknown parameters describing a hadronic τ decay[3] with four observables, corresponding to the components of the four-momentum of the system formed by the visible decay products of the τ lepton, as measured in the laboratory frame. In leptonic τ decays, where two neutrinos are present in the final state, there is one additional degree of freedom, which can be parametrized by the invariant mass $m_{\nu\nu}$ of the two-neutrino system. This leaves two (three) degrees of freedom for each hadronic (leptonic) τ decay, which are parametrized by some unknown parameters $\vec{a}$. The underconstrained problem of finding the

[3]These six parameters are usually defined to be the invariant mass of the visible decay products, the azimuthal and polar angle of the decay product system in the τ rest frame and the three boost parameters from the rest frame to the laboratory frame.

best value for $m_{\tau\tau} = m_{\tau\tau}(p_1^{\text{vis}}, p_2^{\text{vis}}, \vec{a}_1, \vec{a}_2)$ can be solved with a maximum likelihood approach, using the $E_{\text{T},x}^{\text{miss}}$ and $E_{\text{T},y}^{\text{miss}}$ components of the $\vec{E}_{\text{T}}^{\text{miss}}$ vector in the transverse plane, with their given experimental resolution, as additional constraints. A combined likelihood function is built as the product of three likelihood functions:

$$\mathcal{L}(p_1^{\text{vis}}, p_2^{\text{vis}}, \vec{E}_{\text{T}}^{\text{miss}}, \vec{a}_1, \vec{a}_2) = \mathcal{L}_1(p_1^{\text{vis}}, \vec{a}_1) \cdot \mathcal{L}_2(p_2^{\text{vis}}, \vec{a}_2) \cdot \mathcal{L}_\nu(\vec{E}_{\text{T}}^{\text{miss}}, \vec{a}_1, \vec{a}_2)\,. \tag{5.8}$$

The first two, $\mathcal{L}_1$ and $\mathcal{L}_2$, model the decay parameters $\vec{a}_1$ and $\vec{a}_2$ of the τ decays and are based on leading order matrix elements (for unpolarized leptonic τ decays) or on a two-body approximation (for hadronic τ decays). The third likelihood function, $\mathcal{L}_\nu$, quantifies the compatibility of a τ decay hypothesis with the measured $\vec{E}_{\text{T}}^{\text{miss}}$ in the event, under the assumption that the neutrinos coming from the τ decays are the only source of missing transverse energy:

$$\mathcal{L}_\nu(\vec{E}_{\text{T}}^{\text{miss}}, \vec{a}_1, \vec{a}_2) = \frac{1}{2\pi\sqrt{|C|}} \exp\left[-\frac{1}{2}\begin{pmatrix} E_{\text{T},x}^{\text{miss}} - \sum p_x^\nu \\ E_{\text{T},y}^{\text{miss}} - \sum p_y^\nu \end{pmatrix}^T \cdot C^{-1} \cdot \begin{pmatrix} E_{\text{T},x}^{\text{miss}} - \sum p_x^\nu \\ E_{\text{T},y}^{\text{miss}} - \sum p_y^\nu \end{pmatrix}\right]. \tag{5.9}$$

Here the $\vec{E}_{\text{T}}^{\text{miss}}$ resolution is represented by the covariance matrix C, which is estimated on an event-by-event basis using a dedicated algorithm [57].
The combined likelihood function is used to compute the probability

$$P(m_{\tau\tau}^i) = \int \delta\left(m_{\tau\tau}^i - m_{\tau\tau}(p_1^{\text{vis}}, p_2^{\text{vis}}, \vec{a}_1, \vec{a}_2)\right)\, \mathcal{L}(p_1^{\text{vis}}, p_2^{\text{vis}}, \vec{E}_{\text{T}}^{\text{miss}}, \vec{a}_1, \vec{a}_2)\, \mathrm{d}\vec{a}_1 \mathrm{d}\vec{a}_2\,, \tag{5.10}$$

as a function of the mass hypothesis $m_{\tau\tau}^i$. The best estimate $\hat{m}_{\tau\tau}$ for $m_{\tau\tau}$ is defined for each event as the value of $m_{\tau\tau}^i$ which maximizes $P(m_{\tau\tau}^i)$.
The invariant mass obtained from this estimate of the full $\tau\tau$ system has the advantage of completely reconstructing the resonance from which the $\tau\tau$ pair originates, in contrast to the simpler definition of visible mass (m_{vis}), which is computed solely from the visible components of the τ decays reconstructed by the HPS algorithm. The relative resolution on $m_{\tau\tau}$, as computed by the SVFIT algorithm, varies between 10% and 20% depending on the number of neutrinos in the $\tau\tau$ final state [89].

The likelihood function which is maximized in the fit procedure for signal extraction is built from the products of Poisson probability distributions, describing the probability of observing n_i events in a certain bin i of the distribution, given a number ν_i of events expected from the sum of signal and background processes in that bin:

$$\mathcal{L}(\vec{\mu}, \vec{\theta}) = p(\text{data}|\vec{\mu}, \vec{\theta}) \cdot p(\tilde{\vec{\theta}}|\vec{\theta}) = \prod_i \frac{{\nu_i}^{n_i}}{n_i!}\, e^{-\nu_i} \cdot p(\tilde{\vec{\theta}}|\vec{\theta})\,. \tag{5.11}$$

The parameter $\vec{\mu}$ in the likelihood function, referred to as parameter of interest (POI), is the parameter that has to be measured. A vectorial notation has been used to indicate that, in principle, it is possible to define more than a single parameter of interest when building a likelihood function. The remaining set of parameters, referred to as "nuisance" parameters, are grouped into the parameter $\vec{\theta}$. The number of expected events depend on $\vec{\mu}$ and on the value of the nuisance parameters.
The probability density function $p(\tilde{\vec{\theta}}|\vec{\theta})$ represents the probability of observing a set of values $\tilde{\vec{\theta}}$ in separate auxiliary measurements, given that the true value of the nuisance parameters is $\vec{\theta}$. During the likelihood maximization, nuisance parameters are treated following the approach described in [108, 109].
The value of $\vec{\mu}$ that maximizes the likelihood of equation (5.11) is taken as the best-fit estimate

of the parameters of interest and is indicated as $\hat{\vec{\mu}}$. The uncertainty on $\hat{\vec{\mu}}$ is extracted from the confidence bounds built from the likelihood function. In the simplest case in which the likelihood has a single parameter of interest, the values $\mu^{+1\sigma}$ and $\mu^{-1\sigma}$ are defined as the values of μ for which the negative logarithm of the likelihood function exceeds the maximum by half a unit:

$$-\ln \mathcal{L}(\mu^{\pm 1\sigma}, \hat{\theta}_{\mu^{\pm 1\sigma}}) = -\ln \mathcal{L}(\hat{\mu}, \hat{\theta}_{\hat{\mu}}) + 0.5 \tag{5.12}$$

In this procedure, the nuisance parameters are profiled, which means that the values $\hat{\theta}_{\mu^{\pm 1\sigma}}$ are chosen such that the likelihood function reaches its local maximum, under the constraint that $\mu = \mu^{+1\sigma}$ and $\mu = \mu^{-1\sigma}$, respectively.

5.4.6 Systematic uncertainties

Systematic uncertainties are usually conceptually divided into *normalization* and *shape-altering* uncertainties. To the first group belong all those uncertainties which are assumed to mainly affect the normalization of a given template, while preserving its shape: a typical example are uncertainties related to cross sections or scale factors. To the second group belong all those uncertainties which are expected to have an influence on the shape of the template. A classical example is given by the uncertainty on the properties of a certain physics object (the electron energy, for example) used to build a derived quantity (the invariant mass m_{ee} in $\mathrm{Z} \to \mathrm{ee}$ events). In the template fit, systematic uncertainties are represented by the nuisance parameters $\vec{\theta}$ in equation (5.11). From the technical point of view, constraints on the nuisance parameters that affect the normalization but not the shape of the templates are represented by log-normal probability density functions, whose width is controlled by a parameter k. For small uncertainties, a log-normal function with parameter $k = 1 + \epsilon$ is asymptotically identical to a Gaussian function with relative uncertainty ϵ. The log-normal function is however more appropriate for large uncertainties and, in contrast to a Gaussian, allows to avoid problems with nuisances which are physically constrained to be positively defined [108]. Shape-altering uncertainties are included as additional templates, computed after having shifted the affected quantity according to $\pm 1\sigma$, with σ being the expected uncertainty. This gives two additional up/down-wards shifted templates, in addition to the nominal shape. These are incorporated in the likelihood fit via a template morphing technique as detailed in [110] and constrained by Gaussian probability density functions.

Although the actual set of considered systematic uncertainties varies according to the measurement performed, a general list of normalization uncertainties is presented in the following to avoid repetitions. Peculiarities and additional uncertainties are discussed later in the text, if needed.

- **Luminosity**: a systematic uncertainty on the value of the integrated luminosity has a direct impact on all the processes whose normalization is not extracted via a data-driven approach. The overall uncertainty on the integrated luminosity measurement is estimated to be 2.7% [111].
- **Diboson cross sections**: the uncertainty on the cross section used for the normalization of the WW, WZ and ZZ processes is taken to be 10%, corresponding to the theory uncertainty associated with the cross section calculations.
- $\mathrm{t\bar{t}}$ **cross section**: a 6% uncertainty is attributed to the value of the cross section used for the normalization of the $\mathrm{t\bar{t}}$ process, as given by the sum in quadrature of the scale

uncertainty and the uncertainty associated with the choice of the PDF functions and the α_S value.

- **Single-top cross section**: the uncertainty on the cross section used for the normalization of the single-top process is assumed to be 4%, given by the sum in quadrature of the scale uncertainty and the PDF and α_S uncertainties.
- **Drell-Yan cross section**: an uncertainty of 4% is assumed on the value of the cross section used for the normalization of the $Z \to \ell\ell$ ($\ell = e, \mu, \tau$) processes, corresponding to the theory uncertainty associated with the cross section calculation.
- **W+jets cross section**: a theory-driven uncertainty of 4% is associated with the cross section value of the W+jets process.
- **W+jets extrapolation**: an additional 20% systematic uncertainty is associated with the extrapolation method from the control region at high m_T which is used for computing the W+jets normalization, as described in section 5.4.4. This value, which is the dominant source of uncertainty affecting the W+jets process, is motivated by measurements of the jet $\to \tau_h$ misidentification rate.
- **QCD estimation**: a systematic uncertainty between 10% and 15%, depending from the $\tau\tau$ final state, is associated with the scale factor computed in a QCD-enriched control region used for the normalization of the QCD background, as described in section 5.4.4.
- **Electron and muon identification/isolation/trigger efficiency**: the total uncertainty on the trigger, identification and isolation efficiencies, as estimated during the procedure described in section 5.4.3, is set to 4% (for electrons) and 3% (for muons) and applied to all the processes with genuine electrons or muons in the final state.
- **Tau identification efficiency**: this uncertainty is assumed to be 5% following a set of measurements of the identification discriminators performed on the 2015 dataset, such as the one presented in section 5.6, and applied to all the processes with genuine taus in the final state.
- $e \to \tau_h$ **misidentification rate**: the uncertainty on the measurement of the electron misidentification rate, which is applied to the $Z \to ee$ process in $e\tau_h$ final states, varies from 5% to 30% depending from the discriminator working point, as estimated in the analysis presented in section 5.5.
- $\mu \to \tau_h$ **misidentification rate**: a measurement of the $\mu \to \tau_h$ misidentification rate in data has not yet been completed by the CMS collaboration and a very conservative uncertainty of 100% is associated with the $Z \to \mu\mu$ process in $\mu\tau_h$ final states.

An additional systematic uncertainty is included for each bin of each template to account for possible fluctuations due to the finite number of events in the bin, following the proposal in [112]. Effectively, this corresponds to the introduction of a shape-altering uncertainty which is allowed to modify the shape of each template, shifting each bin upwards or downwards according to a 1σ shift in the statistical uncertainty associated with the bin. These class of uncertainties are referred to as **bin-by-bin** uncertainties.

5.5 Measurement of the electron misidentification rate in Z → ee events

The e → τ_h misidentification rate has been measured in data collected by the CMS experiment during the 2015 data taking period at a center of mass energy of 13 TeV. The analysis strategy consists in measuring the number of $Z/\gamma^* \to ee$ events which are reconstructed as $Z/\gamma^* \to \tau\tau \to e\tau_h$, therefore estimating the fraction of electrons which are misidentified as τ_h candidates. This is achieved thanks to a tag-and-probe technique where $Z/\gamma^* \to ee$ events are selected requiring the presence of a well identified and isolated electron (the tag) and a loosely identified τ_h candidate (the probe). Depending on whether the probe passes or fails a selection based on the anti-e discriminator under study, the events are divided into two categories called *pass* and *fail*, respectively. The number of $Z/\gamma^* \to ee$ events in each category, $N_{\text{pass}}^{\text{probe}}$ and $N_{\text{fail}}^{\text{probe}}$, is extracted from a simultaneous template fit in the two anticorrelated categories.
Defining the misidentification rate as ϵ, the number of probes in each category can be expressed as:

$$\begin{aligned} N_{\text{pass}}^{\text{probe}} &= \epsilon \cdot N_0^{\text{probe}} \\ N_{\text{fail}}^{\text{probe}} &= (1-\epsilon) \cdot N_0^{\text{probe}} , \end{aligned} \tag{5.13}$$

with N_0^{probe} being the total number of probes. The parameter $\mu = \epsilon'/\epsilon$, defined as the ratio between the misidentification rate before and after the fit, is taken to be the parameter of interest with respect to which the likelihood function of equation (5.11) is maximized. The number of probes in each category, after the fit, can be expressed as a function of μ and ϵ as:

$$\begin{aligned} N_{\text{pass}}^{\text{probe}'} &= \epsilon' \cdot N_0^{\text{probe}} = \mu\epsilon \cdot N_0^{\text{probe}} = \mu \cdot N_{\text{pass}}^{\text{probe}} \\ N_{\text{fail}}^{\text{probe}'} &= (1-\epsilon') \cdot N_0^{\text{probe}} = (1-\mu\epsilon) \cdot N_0^{\text{probe}} = \frac{(1-\mu\epsilon)}{(1-\epsilon)} \cdot N_{\text{fail}}^{\text{probe}} . \end{aligned} \tag{5.14}$$

From equation (5.14) it is straightforward to see that, with this definition of the normalization in the two anticorrelated categories, the total number of probes is conserved before and after the fit: $N_{\text{pass}}^{\text{probe}} + N_{\text{fail}}^{\text{probe}} = N_{\text{pass}}^{\text{probe}'} + N_{\text{fail}}^{\text{probe}'} = N_0^{\text{probe}}$.
Signal extraction is based on templates built from the visible mass m_{vis} of the tag-probe pair, computed using the energy and the momentum reconstructed by the electron reconstruction algorithm for the tag and the visible energy and momentum reconstructed by the HPS algorithm for the probe. The visible mass is chosen as discriminating variable because of the clear separation between $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \tau\tau \to e\tau_h$ events.

5.5.1 Object identification and event selection

Data events are collected by the unprescaled `HLT_Ele23_WPLoose_Gsf` single electron trigger, which was deployed during all the `2015D` data-taking period. Tag electrons are required to pass the tight set of identification selections described in section 5.4.2 and to have a relative isolation $I_{\text{rel}}^{e} < 0.1$. The transverse momentum of the tag electron candidate is chosen to be 1 GeV larger than the HLT threshold, in order to avoid turn-on effects due to the fact that the trigger is not yet fully efficient in the low-p_T region. The electron p_T selection is therefore set to 24 GeV. The absolute value of the electron pseudorapidity $|\eta|$ is required to be smaller than 2.1.
In addition to the requirements listed in section 5.4.2, τ_h candidates are required to pass the loose working point of the MVA isolation discriminator. In order to reduce the already small fraction of muons misidentified as τ_h, the τ_h candidate is also required to pass the loose working point of the cut-based discriminator against $\mu \to \tau_h$ misidentification. The transverse momentum of the

τ_h candidate is required to be larger than 20 GeV and the absolute value of the pseudorapidity to be smaller than 2.3.
Requiring the ΔR distance in the η-ϕ plane between the tag electron and the τ_h candidate to be equal or larger than 0.5 ensures that the probes are clearly separated from the tag electrons. The tag-probe system is finally required to have opposite electric charge.

The same trigger requirement, identification and isolation criteria are applied on the simulated events for the processes listed in table A.2 and A.3. Scale factors taking into account mismodelling in the pile-up description as well as data/simulation differences in the identification, isolation and trigger selections, as explained in section 5.4.3, are used to correct the simulated events.

After the selections described above, the selected event sample includes a non-negligible contribution coming mainly from W+jets and $t\bar{t}$ backgrounds. This contamination is reduced requiring the events to pass a selection on the transverse mass computed from the tag electron and the missing transverse energy: $m_T(e, E_T^{miss}) < 30$ GeV.

Visible mass templates, in the region 60 GeV $< m_{vis} <$ 120 GeV, are built from the $Z/\gamma^* \to ee$ and the $Z/\gamma^* \to \tau\tau$, W+jets, $t\bar{t}$, single-top, diboson (WW, WZ, ZZ) and QCD multijet processes. The fail category, highly dominated by $Z/\gamma^* \to ee$ events, is used solely to constrain the overall normalization of the $Z/\gamma^* \to ee$ contribution: the fit in this category is performed on the inclusive number of events under the Z peak, between 60 GeV and 120 GeV.
The shape templates for the $Z/\gamma^* \to ee$ and all background processes, except QCD, are taken from the simulation. For the normalization of the W+jets contribution, a correction derived from data in a background-dominated control region at high $m_T(e, E_T^{miss})$ is applied according to the procedure outlined in section 5.4.4. The shape and normalization of the QCD multijet background are derived from data in a control region where the tag and the probe have the same electric charge, as explained in section 5.4.4.
The basic systematic uncertainties outlined in section 5.4.6 are included as nuisance parameters $\vec{\theta}$ in the template fit. The systematic uncertainty associated with the $e \to \tau_h$ misidentification rate is excluded from the uncertainty model, since it is estimated in this measurement. An additional normalization uncertainty of $\pm 3\%$ is associated with the $Z/\gamma^* \to ee$ process only, to disentangle possible differences between the $Z/\gamma^* \to ee$ and the $Z/\gamma^* \to \tau\tau$ normalization.
The following systematic uncertainties affecting the shape of the templates are taken into account as well:

- **Tag electron energy scale**: the energy scale of tag electrons is varied by $\pm 2\%$ and the resulting difference in the m_{vis} template is considered as a shape-altering uncertainty.
- **Probe electron energy scale**: the energy scale of probes matched to generated electrons, namely τ_h candidates originating from a misidentified electron, is varied by $\pm 5\%$ and the difference in the m_{vis} template is considered as a shape-altering uncertainty.
- **Probe tau energy scale**: the energy scale of probes matched to generated τ_h's is varied by $\pm 3\%$, a conservative value consistent with the τ_h energy scale measurement presented in section 5.7, and the difference in the m_{vis} template is considered as a shape-altering uncertainty.
- **Electron energy resolution**: the uncertainty on the electron resolution is taken into account as an additional shape-altering uncertainty, varying the difference between the generated and the reconstructed visible mass of the tag-probe pair upwards and downwards by 10%. This value is more conservative than the electron energy resolution measured by

CMS (2-5%, depending on the pseudorapidity [37]) and justified by the fact that, in this measurement, the tag-probe pair invariant mass is built from physics objects which satisfy much looser identification criteria, compared to the ones from which the electron energy resolution has been derived.

Given the fact that, as explained in section 5.3, different BDTs are trained for the anti-e discriminator depending on the pseudorapidity of the τ_h candidates, the measurement is performed separately for probes in the barrel ($|\eta| < 1.460$) and in the endcap ($|\eta| > 1.558$).

5.5.2 Results

The visible mass distributions for the category where the probes have a pseudorapidity in the barrel region of the detector and pass the anti-e discriminator are shown in figure 5.14 for the loose, medium, tight and very tight working points.
A complete summary of the $e \to \tau_h$ misidentification rates, as measured for all the working points of the discriminator, is given in table 5.5. Depending on the working point, the measured $e \to \tau_h$ misidentification rates are around the percent or permill level. For probes in the barrel region, the misidentification rates measured in data exceed the expectation, with the difference between data and simulation increasing for the tight to very tight working points. A similar tendency is observed for probes in the endcap region passing the very loose and loose working points of the discriminator. For endcap probes passing tighter working points the data/simulation ratios are compatible with unity, although the uncertainties associated with the measurement are large, due to the limited statistics. Overall, the trend observed in the measured misidentification rates is similar to what has been measured in 8 TeV data, using the previous version of the MVA-based anti-e discriminator [95]. The uncertainty associated with the measured data/simulation ratios varies from 5% to 30%, depending from the discriminator working point. A graphical comparison between the misidentification rates measured in data and the expectation is shown in figure 5.15.
The measured data/simulation ratios can be taken into account in future physics analyses targeting final states with τ_h candidates by applying suitable correction factors to the simulated events. The uncertainties associated with the misidentification rate measurement are then included as systematic uncertainties in the model used for the statistical interpretation of the data.

5.6 Measurement of the τ_h identification efficiency in $Z \to \tau\tau \to \mu\tau_h$ events

A tag-and-probe technique very similar to the one introduced in section 5.5 for the measurement of the $e \to \tau_h$ misidentification rate can also be employed to measure the efficiency of a given τ_h discriminator in data. The targeted final state is $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ which, due to the presence of the muon, presents less experimental difficulties compared to the $e\tau_h$ final state.
A sample of $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events is selected requiring a well identified and isolated muon (the tag) and a loosely identified τ_h candidate (the probe). Events are then categorized into a *pass* and *fail* category, depending on whether the probe passes or fails the τ_h-identification discriminator under study. In the model on which the statistical interpretation of the data is based, the parameter ϵ of equation (5.14) represents now the efficiency of the given discriminator to identify genuine τ_h decays. The number of $Z/\gamma^* \to \tau\tau$ events in each category is extracted

ECAL barrel ($\lvert\eta\rvert < 1.460$)			
WP	$e \to \tau_h$ misid. rate (Simulation)	$e \to \tau_h$ misid. rate (Data)	Data/Simul.
Very Loose	$(5.92 \pm 0.59) \times 10^{-2}$	$(5.99 \pm 0.59) \times 10^{-2}$	$1.01^{+0.05}_{-0.01}$
Loose	$(1.12 \pm 0.12) \times 10^{-2}$	$(1.22 \pm 0.13) \times 10^{-2}$	1.10 ± 0.03
Medium	$(2.75 \pm 0.29) \times 10^{-3}$	$(3.94 \pm 0.42) \times 10^{-3}$	1.43 ± 0.10
Tight	$(1.16 \pm 0.12) \times 10^{-3}$	$(1.89 \pm 0.20) \times 10^{-3}$	$1.63^{+0.18}_{-0.15}$
Very Tight	$(0.67 \pm 0.07) \times 10^{-3}$	$(1.10 \pm 0.12) \times 10^{-3}$	$1.65^{+0.28}_{-0.26}$
ECAL endcap ($\lvert\eta\rvert > 1.558$)			
WP	$e \to \tau_h$ misid. rate (Simulation)	$e \to \tau_h$ misid. rate (Data)	Data/Simul.
Very Loose	$(7.05 \pm 0.70) \times 10^{-2}$	$(7.79 \pm 0.77) \times 10^{-2}$	$1.11^{+0.05}_{-0.01}$
Loose	$(1.10 \pm 1.16) \times 10^{-2}$	$(1.17 \pm 0.12) \times 10^{-2}$	1.06 ± 0.05
Medium	$(2.26 \pm 0.24) \times 10^{-3}$	$(2.12 \pm 0.23) \times 10^{-3}$	$0.94^{+0.15}_{-0.14}$
Tight	$(0.95 \pm 0.10) \times 10^{-3}$	$(0.95 \pm 0.10) \times 10^{-3}$	$0.99^{+0.34}_{-0.25}$
Very Tight	$(0.54 \pm 0.06) \times 10^{-3}$	$(0.52 \pm 0.06) \times 10^{-3}$	$0.96^{+0.57}_{-0.56}$

Table 5.5: Probability for electrons in $Z/\gamma^* \to ee$ events to pass the different working points of the MVA-based anti-e discriminator, divided into "barrel" and "endcap" according to the probe pseudorapidity $|\eta|$. For each working point, the $e \to \tau_h$ misidentification rate is defined as the fraction of probes passing the given discriminator with respect to the total number of probes.

from a simultaneous template fit, maximizing a likelihood function having the ratio $\mu = \epsilon'/\epsilon$ between the efficiency before and after the fit as parameter of interest. The templates for the fit are built from the distribution of the visible mass m_{vis}, computed from the energy and momentum reconstructed by the muon reconstruction algorithm for the tag and the visible energy and momentum reconstructed by the HPS algorithm for the probe.
In the following, the efficiency of the MVA isolation discriminator is measured in data. A crucial requirement for the measurement is the ability to control the shape and normalization of the W+jets and QCD multijet backgrounds, which represent by far the dominant contribution in the category where the probes fail the isolation discriminator.

5.6.1 Object identification and event selection

Events are collected by the unprescaled `HLT_IsoMu18` single muon trigger, which was deployed during all the `2015D` data-taking period. Muons are identified as described in section 5.4.2 and required to have a relative isolation $I^{\mu}_{\text{rel}} < 0.1$. The transverse momentum of the muon candidate is chosen to be larger than 19 GeV, 1 GeV above the HLT threshold. The absolute value of the muon pseudorapidity $|\eta|$ is required to be smaller than 2.1.
The τ_h candidates, reconstructed by the HPS algorithm and satisfying the identification criteria of section 5.4.2, are required to have $p_T > 20$ GeV and $|\eta| < 2.3$. The contamination coming from electrons and muons which are wrongly reconstructed as hadronically decaying taus is suppressed requiring the τ_h candidate to pass the MVA discriminator against electrons described in section 5.3 and the cut-based discriminator against muons: the loose WP is applied on the anti-e discriminator, while the tight WP is used for the discriminator against muon misidentification. The distance ΔR between the muon and the τ_h candidate is finally required to be larger than 0.5 and the pair must have opposite electric charge.

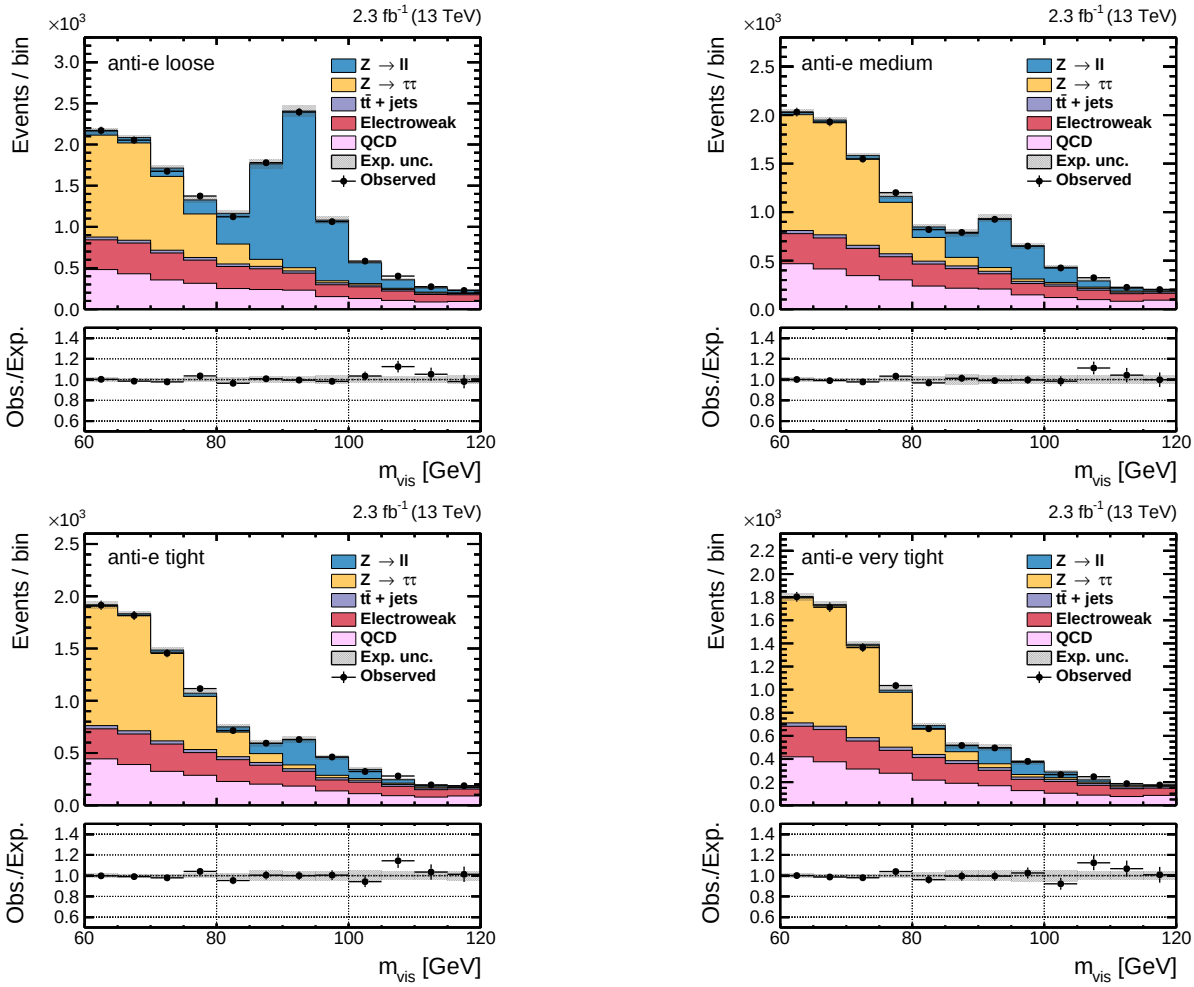


Figure 5.14: Distributions, after the fit, of the visible mass of the tag-probe pair in the pass category, for the loose (top left plot), medium (top right), tight (bottom left) and very tight (bottom right) WP of the anti-e discriminator. The probe is reconstructed in the barrel region of the detector ($|\eta| < 1.460$). The electroweak background contribution includes events from W+jets, diboson and single-top production.

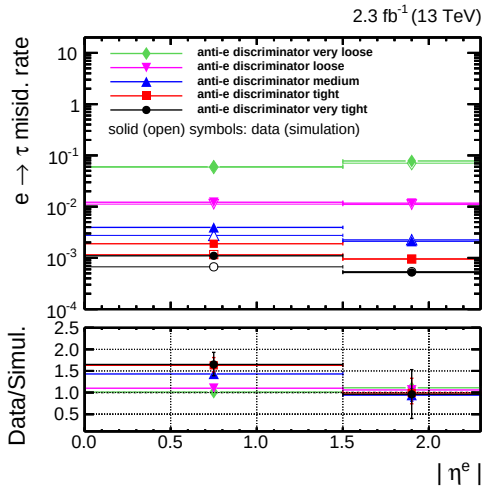


Figure 5.15: Probability for electrons in $Z/\gamma^* \to ee$ events to pass the different working points of the MVA-based anti-e discriminator, as a function of the electron pseudorapidity. The measured $e \to \tau_h$ misidentification rates (solid symbols) are compared to the simulation (open symbols). The ratio between data and simulation in the bottom part of the plot corresponds to the values in the last column of table 5.5.

The same trigger requirement, identification and isolation criteria are applied on the simulated events for the processes listed in table A.2 and A.3, correcting the mismodelling between data

and simulation with the scale factors described in section 5.4.3.

The contamination from W+jets and $t\bar{t}$ backgrounds is reduced requiring the events to pass a selection on the transverse mass, computed from the muon candidate and the missing transverse energy: $m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) < 40\,\mathrm{GeV}$.
Another help in discriminating between genuine $\tau\tau$ events and W+jets and $t\bar{t}$ backgrounds is offered by the two following quantities:

$$\begin{aligned} p_{\zeta}^{\mathrm{vis}} &= \vec{p}_{\mathrm{T},1} \cdot \hat{\zeta} + \vec{p}_{\mathrm{T},2} \cdot \hat{\zeta} \\ p_{\zeta}^{\mathrm{miss}} &= \vec{E}_{\mathrm{T}}^{\mathrm{miss}} \cdot \hat{\zeta} \end{aligned} \tag{5.15}$$

As sketched in figure 5.16, ζ is the bisector of the angle between the directions of the visible τ decay products in the transverse plane. The quantities $\vec{p}_{\mathrm{T,i}} \cdot \hat{\zeta}$ and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}} \cdot \hat{\zeta}$ are, respectively, the projections of the transverse momenta of the visible decay products and of the missing transverse energy vector onto the bisector. In genuine $\tau\tau$ events the neutrinos are emitted preferably in the same direction of the visible components of the decay, hence $p_{\zeta}^{\mathrm{miss}}$ tends to be collinear with p_{ζ}^{vis}. In W+jets events where one jet is misidentified as a τ_{h} the neutrino is preferably opposite to the lepton direction in the transverse plane and not correlated with the direction of the misidentified jet, resulting in a corresponding diluted correlation between the two projections. The same holds true for the more complicated $t\bar{t}$ decay topology. The projections can be linearly combined in the variable:

$$D_{\zeta} \equiv p_{\zeta}^{\mathrm{miss}} - 0.85 \cdot p_{\zeta}^{\mathrm{vis}}\,. \tag{5.16}$$

which maximizes the separation between $\tau\tau$ events and W+jets and $t\bar{t}$ backgrounds. Shapes, normalized to unit area, of the D_{ζ} variable in Drell-Yan, W+jets and $t\bar{t}$ simulated events are shown in figure 5.17. Additional distributions of the D_{ζ} variable, showing the relative contributions of the different backgrounds for events selected according to the requirements described in section 5.8, are presented in figures D.1, D.2 and D.3 in the appendix.
Based on the above considerations, an additional selection is applied on the D_{ζ} variable, requiring it to be larger than $-25\,\mathrm{GeV}$.
Another handle for the reduction of the $t\bar{t}$ background is given by a selection on the number of reconstructed b-tagged jets: genuine $t\bar{t}$ events have two b-tagged jets in the final state which, due to the finite precision of the tagging algorithms, can be correctly or incorrectly identified. Events are required to have strictly zero b-tagged jets, where the tagging is performed using the algorithm described in 5.4.2.
A sub-dominant background contribution coming from prompt $\mathrm{Z}/\gamma^{*} \to \mu\mu$ decays is reduced rejecting the events with an opposite-charge muon pair, where the muons of the pair are required to be separated by $\Delta R > 0.15$ and to pass looser isolation and kinematic selections[4].

Visible mass templates are built from the $\mathrm{Z}/\gamma^{*} \to \tau\tau$ and the $\mathrm{Z}/\gamma^{*} \to \mu\mu$, W+jets, $t\bar{t}$, single-top, diboson (WW, WZ, ZZ) and QCD multijet processes. The shape templates for the $\mathrm{Z}/\gamma^{*} \to \tau\tau$ and all background processes, except QCD, are taken from the simulation. The normalization of the W+jets contribution and the shape and normalization of the QCD multijet background are derived from control regions in data, as explained in section 5.4.4.
The basic list of systematic uncertainties outlined in section 5.4.6 are included as nuisance parameters $\vec{\theta}$ in the template fit, except for the uncertainty on the τ identification efficiency. The uncertainty on the b-tagging efficiency and on the misidentification probability, which enters via the requirement on the number of b-tagged jets, are assumed to be equal to 1% and 8% respectively, based on the latest measurements performed in CMS [113]. Similar to what is done

[4]The muons of the pair are required to have $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$, while the selection on the relative isolation variable is relaxed to $\mathrm{I}_{\mathrm{rel}}^{\mu} < 0.3$.

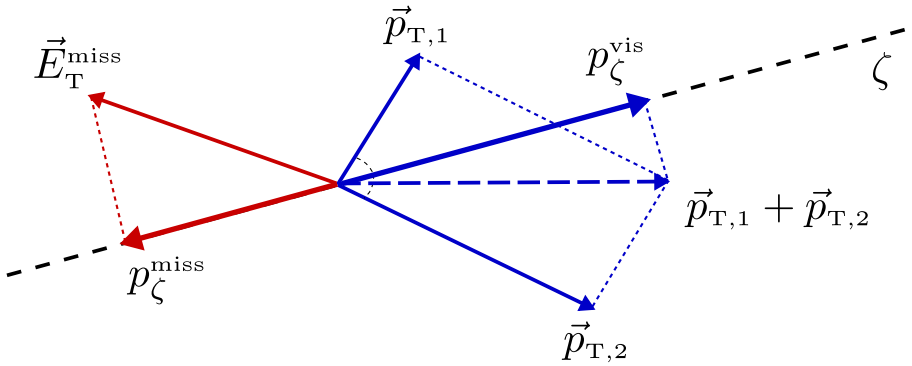


Figure 5.16: The sketch illustrates how the p_ζ^{vis} and p_ζ^{miss} variables are constructed starting from the direction of the visible τ decay products in the transverse plane (the blue vectors) and the missing transverse energy vector $\vec{E}_T^{miss}$ (in red).

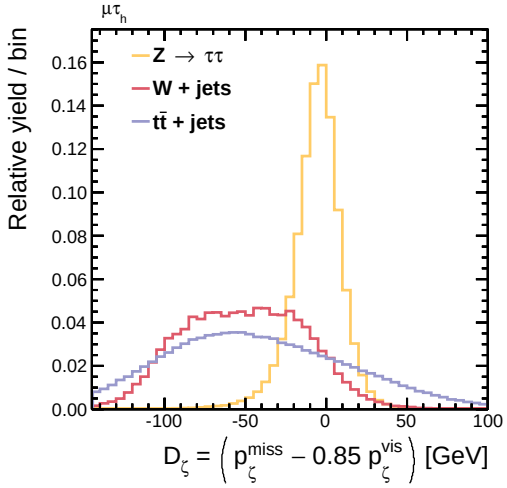


Figure 5.17: Distribution, normalized to unit area, of the D_ζ variable in simulated $Z/\gamma^* \to \tau\tau$, W+jets and $t\bar{t}$ events. The events are selected requiring the presence of an oppositely-charged pair made of a muon with $p_T > 19\,GeV$, $|\eta| < 2.1$, satisfying the identification criteria of section 5.4.2 and with $I_{rel}^{\mu} < 0.1$ and of a τ_h with $p_T > 20\,GeV$, $|\eta| < 2.3$, passing the tight WP of the cut-based discriminator against $\mu \to \tau_h$ misidentification and the tight WP of the MVA isolation discriminator.

for the measurement presented in section 5.5, the energy scale of probes matched to generated τ_h is varied by $\pm 3\%$ and the difference in the m_{vis} template is considered as an additional shape-altering uncertainty. In a separate tag-and-probe measurement of the $\mu \to \tau_h$ misidentification rate, an energy scale effect with a magnitude around 5% has been observed for muons misidentified as τ_h candidates: this effect is also incorporated as an additional shape-altering uncertainty associated with the $Z/\gamma^* \to \mu\mu$ template.

5.6.2 Results

Visible mass distributions for probes passing or failing the MVA-based isolation discriminator are shown in figure 5.18 for the loose, medium and tight working points. It is interesting to note how, in the categories where the probes do not pass the selection applied on the isolation discriminator, the $Z/\gamma^* \to \tau\tau$ contribution represents a small fraction of the total events, which are dominated by the W+jets and QCD multijet backgrounds.
A complete summary of the τ_h identification efficiencies, for each of the working points, is given in table 5.6. For this particular discriminator the measured efficiencies are well in agreement with the expectation for all the working points, resulting in data/simulation ratios compatible with unity. The overall uncertainty on the measurement is about 4%.

WP	MVA τ_h iso eff. (Simulation)	MVA τ_h iso eff. (Data)	Data/Simulation
Loose	0.868 ± 0.009	0.883 ± 0.008	1.02 ± 0.04
Medium	0.791 ± 0.014	0.795 ± 0.013	1.01 ± 0.04
Tight	0.707 ± 0.017	0.711 ± 0.017	1.00 ± 0.04
Very Tight	0.615 ± 0.019	0.612 ± 0.019	0.99 ± 0.03

Table 5.6: Probability for τ_h candidates in $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events to pass the different working points of the MVA-based isolation discriminator. For each working point, the efficiency is defined as the fraction of probes passing the given discriminator with respect to the total number of probes.

5.7 Measurement of the τ_h energy scale in $Z \to \tau\tau \to \mu\tau_h$ events

The energy scale (ES) of the τ_h candidates is defined as the ratio of the reconstructed τ_h energy to the true energy of the generated visible tau decay products, as defined in the simulation. This scale effect in the energy reconstruction constitutes an important source of systematic uncertainty for many analyses relying on the detection of τ_h decays in the final state, especially those where the $Z/\gamma^* \to \tau\tau$ process represents an irreducible source of background.
The τ_h-ES is measured selecting a sample of $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events and fitting the distribution of a variable sensitive to a scaling of the τ_h energy to the data. The visible tau mass m_{τ_h} reconstructed by the HPS algorithm depends linearly on the energy scale and can be used as discriminating variable for the τ_h candidates reconstructed in all the decay modes, except for the $h^\pm$ decay mode: in this case, in fact, the τ_h mass is fixed to the nominal pion mass during the reconstruction sequence and the discriminating power is therefore lost. The visible mass $m_{\text{vis}}(\mu\tau_h)$, built from the energy and momentum of the muon and from the visible components of the τ_h decay, is proportional to the square root of the energy scale and can be used as discriminating variable in all decay modes. These considerations justify the choice of performing the measurement in the $\mu\tau_h$ final state, where the muon used to build the visible mass $m_{\text{vis}}(\mu\tau_h)$ is less influenced by scale effects in the muon energy reconstruction compared to the electron in the $e\tau_h$ final state.
For the measurement, different templates are built for the $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ process, scaling the four-momenta of the τ_h candidates between -6% and $+6\%$, in discrete steps of 0.1%. It should be noted that this scaling affects also the event acceptance, due to the selection applied on the reconstructed transverse momentum of the τ_h candidates. The impact on the acceptance, which is around 5% for the largest applied scale variation, is taken into account in the measurement. To cover the intermediate values of the τ_h-ES where no templates are available, the discrete templates are linearly interpolated such that the transition between a template and the subsequent one is controlled by a continuous parameter. This technique is identical to what is applied in Higgs analyses such as [89,93] to interpolate between different signal mass hypotheses. The parameter controlling the variation of the templates is defined as the parameter of interest in the likelihood function maximized during the template fit. The best-fit value for the energy scale and the associated uncertainty are extracted after a scan of the likelihood as a function of the τ_h-ES variation. The best-fit is given by the minimum of the negative logarithm of the likelihood function, while the uncertainty is obtained from the points at which $-2\Delta\ln\mathcal{L} = 1$. No scale variation is applied on the background templates, since the fraction of genuine τ decays in the dominant W+jets and QCD backgrounds is found to be negligible.
The energy scale can be measured inclusively or separating the τ_h candidates according to the reconstructed decay mode into three categories: $h^\pm$, $h^\pm(\geq 1\pi^0)$ and $h^\pm h^\mp h^\pm$. Measuring the energy scale in different decay modes allows to disentangle possible differences between scaling effects introduced by the reconstruction of the charged components of the hadronic tau decay

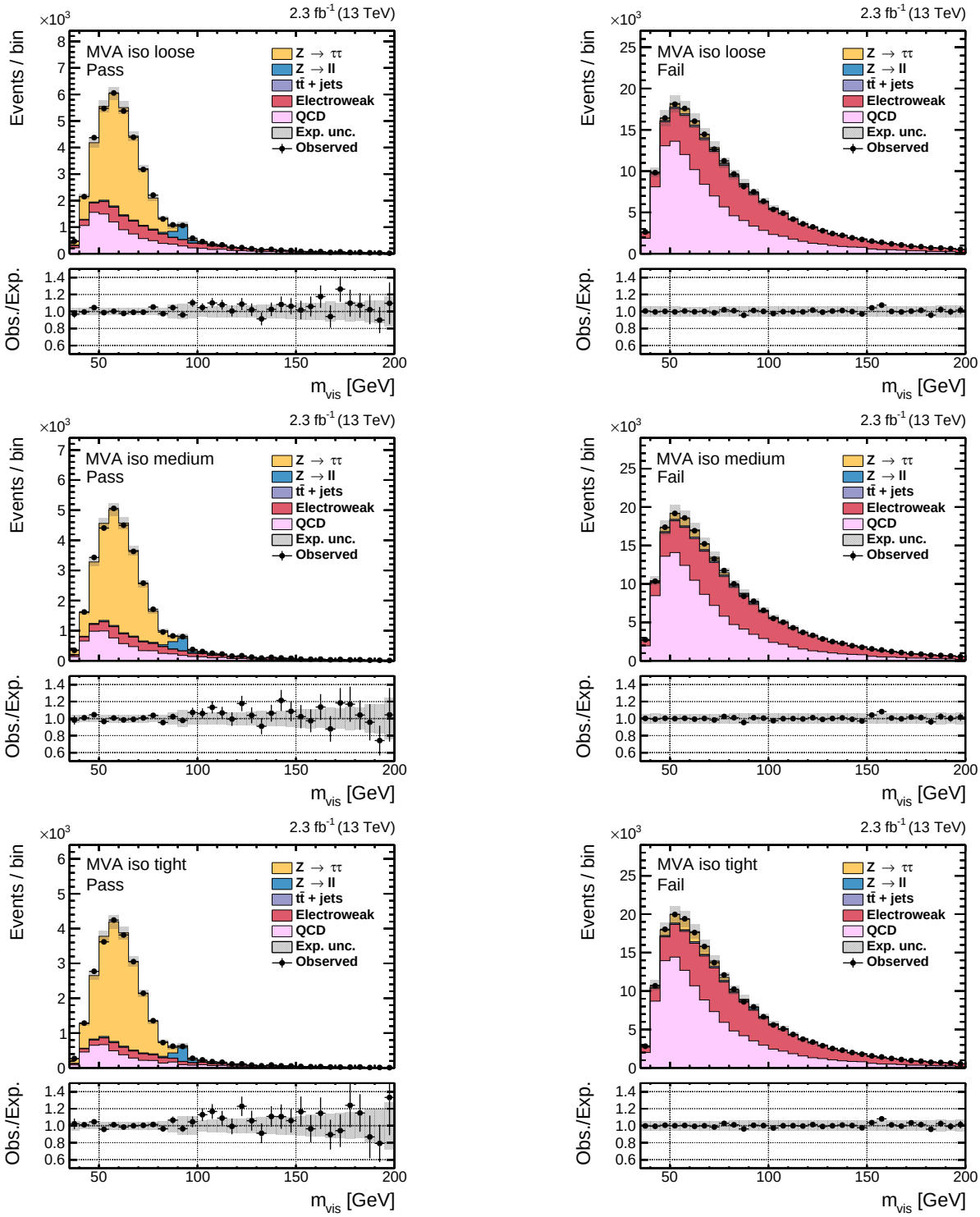


Figure 5.18: Distributions, after the fit, of the visible mass of the tag-probe pair, in the pass (left) and fail (right) categories for the loose (top plots), medium (central plots) and tight (bottom plots) WP of the MVA-based isolation discriminator. The electroweak background contribution includes events from W+jets, diboson and single-top production.

and the ones due to the reconstruction of the neutral component.
The evolution of the τ_h-ES with respect to the energy of the τ_h candidate can be probed fur-

ther, categorizing the candidates according to their reconstructed transverse momentum. The amount of data collected during 2015 is however approximately only 10% of the dataset collected by the experiment during Run I and a detailed measurement of the energy scale in different p_{T}-regions, similar to what presented in [95], is therefore challenging. A measurement is performed grouping all the decay modes in an inclusive category, dividing the events according to the transverse momentum of the τ_{h} candidate into two kinematic regions ($20\,\mathrm{GeV} < p_{\mathrm{T}} < 35\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 35\,\mathrm{GeV}$) and repeating the measurement in each one of them. For the reason mentioned above, candidates reconstructed in the $h^{\pm}$ decay mode are excluded from this measurement, when building the $m_{\tau_{\mathrm{h}}}$ templates.

5.7.1 Object identification and event selection

The event selection follows closely the criteria which have been applied for the tag-and-probe measurement presented in section 5.6, with the only difference that the τ_{h} candidates are now additionally required to pass the tight working point of the MVA isolation discriminator.
Events are divided in three different categories depending on the reconstructed τ_{h} decay mode and a measurement of the energy scale is performed both inclusively and for each category. Mass templates are built for the $\mathrm{Z}/\gamma^{*} \to \tau\tau$ and the $\mathrm{Z}/\gamma^{*} \to \mu\mu$, $\mathrm{t\bar{t}}$, single-top, diboson (WW, WZ, ZZ), W+jets and QCD multijet processes, extracting the shape and normalization of the last two backgrounds from control regions in data, as explained in section 5.4.4. All the correction factors discussed in the previous sections are applied to the simulated events.
The distribution of the reconstructed visible tau mass $m_{\tau_{\mathrm{h}}}$ is used for building the templates in the $h^{\pm}(\geq 1\pi^{0})$ and $h^{\pm}h^{\mp}h^{\pm}$ categories, while a measurement using $m_{\mathrm{vis}}(\mu\tau_{\mathrm{h}})$ is performed in all the three of them. The effect of an energy scale variation on the shape of the visible tau mass distribution is shown in figure 5.19.

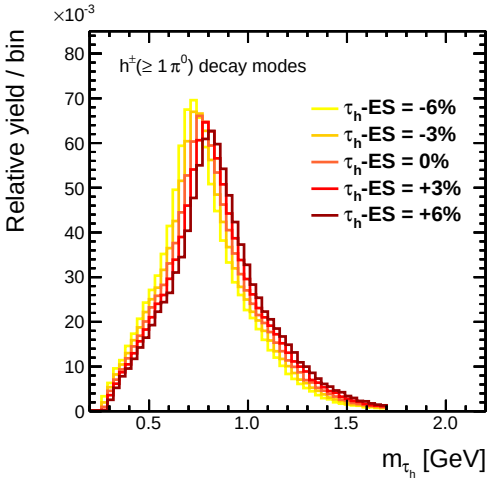


Figure 5.19: Distributions, normalized to unit area, of $m_{\tau_{\mathrm{h}}}$, the mass of the reconstructed τ_{h} candidate, for different τ_{h}-ES variations. The τ_{h} candidates are required to be reconstructed in the $h^{\pm}\pi^{0}$ or $h^{\pm}\pi^{0}\pi^{0}$ decay modes, to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$ and to pass the tight WP of the MVA isolation discriminator.

Visible mass distributions, illustrating the effect of a scale variation of the $\mathrm{Z}/\gamma^{*} \to \tau\tau$ templates on the agreement between data and simulation, are shown in figure 5.20, 5.21 and 5.22 for each decay mode category and three different τ_{h}-ES variations: −6%, 0% and +6%. From the ratio between data and simulation in the bottom part of the plots, it can be clearly seen that energy scale variations between −6% and +6% cover the best value of the energy scale in all the decay modes. Additional mass distributions are shown in appendix B.

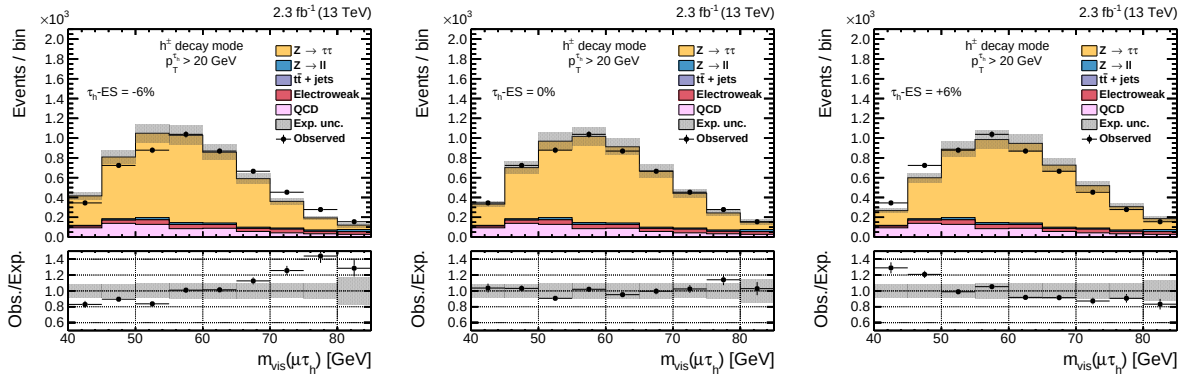


Figure 5.20: Distribution of $m_{\text{vis}}(\mu\tau_{\text{h}})$, the invariant mass reconstructed from the muon and the visible components of the τ_{h} decay, for different τ_{h}-ES variations: -6% (left), 0% (center) and $+6\%$ (right). The τ_{h} candidates are required to be reconstructed in the $h^{\pm}$ decay mode, to have $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.3$ and to pass the tight WP of the MVA isolation discriminator.

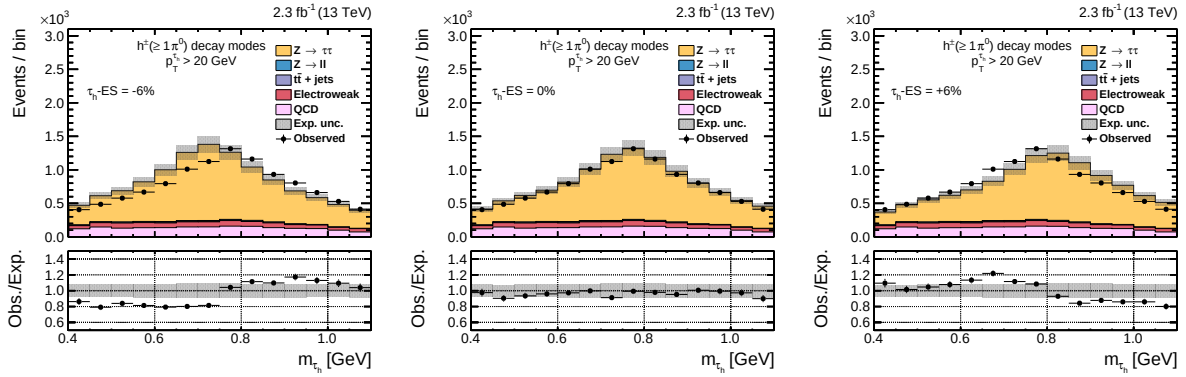


Figure 5.21: Distribution of $m_{\tau_{\text{h}}}$, the mass of the reconstructed τ_{h} candidate, for different τ_{h}-ES variations: -6% (left), 0% (center) and $+6\%$ (right). The τ_{h} candidates are required to be reconstructed in the $h^{\pm}\pi^0$ or $h^{\pm}\pi^0\pi^0$ decay modes, to have $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.3$ and to pass the tight WP of the MVA isolation discriminator.

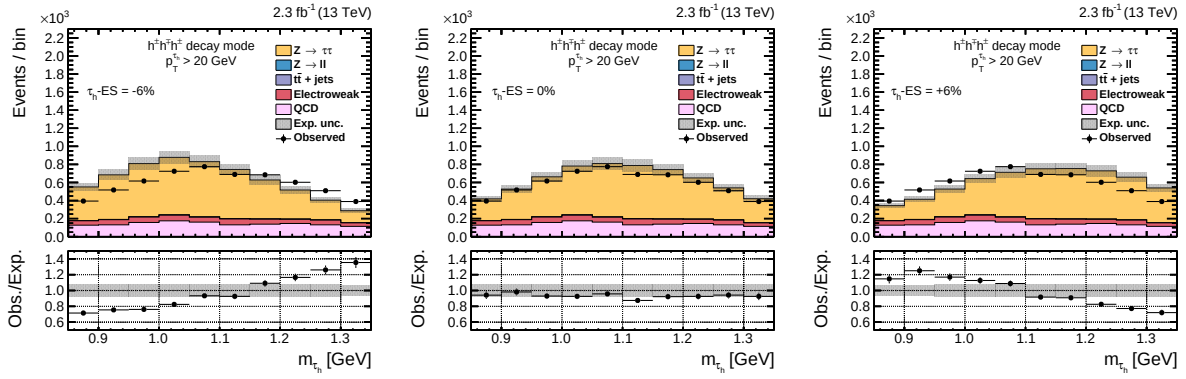


Figure 5.22: Distribution of $m_{\tau_{\text{h}}}$, the mass of the reconstructed τ_{h} candidate, for different τ_{h}-ES variations: -6% (left), 0% (center) and $+6\%$ (right). The τ_{h} candidates are required to be reconstructed in the $h^{\pm}h^{\mp}h^{\pm}$ decay mode, to have $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.3$ and to pass the tight WP of the MVA isolation discriminator.

5.7.2 Results

The results of the measurements are summarized in table 5.7, where the best-fit values of the τ_h-ES variation are presented for each decay mode category and p_T region. A graphical representation of the results is shown in figure 5.23. In the $h^\pm h^\mp h^\pm$ decay mode, the measurements performed on the two mass variables agree within the uncertainties while a 2σ tension is observed in the more complicated decay modes involving neutral pions. For the measurement in the $h^\pm$ category only one result, derived from the $m_{\text{vis}}(\mu\tau_h)$ templates, is extracted. The blue dotted line in figure 5.23 corresponds to a τ_h-ES variation of 0%. In general, the measurements performed on the m_{τ_h} variable, which depends linearly on the energy scale, are more sensitive compared to the ones computed from the $m_{\text{vis}}(\mu\tau_h)$ templates and this is reflected in the smaller size of the uncertainties associated with the measured scale variations. The results of the measurements performed in the inclusive decay mode category, for τ_h candidates reconstructed in different p_T-regions, are shown in figure 5.24. No indications are found for a dependence of the measured τ_h-ES variations on the transverse momentum, even though the precision of the measurement is limited by the large statistically-driven uncertainties.
With the current amount of data collected at 13 TeV, there is no strong indication for the need of correcting the simulation to take into account energy scale effects in the τ_h reconstruction in any particular decay mode. The dispersion around a 0% τ_h-ES variation of the best-fit values is translated into a conservative 3% systematic uncertainty on the τ_h energy scale, which reflects the complexity of the measurement. It has already been observed in the past that an increase in the amount of collected data has a beneficial effect on the quality of the likelihood scans, which leads to more precise results. Repeating this measurement adding the first 2016 data, therefore at least doubling the currently available dataset, will give more precise indications on the magnitude of the scaling effects and, most of all, will allow to probe the dependency of the τ-ES on the reconstructed τ_h momentum. The number of available $Z/\gamma^* \to \tau\tau$ simulated events and the overall correctness of the simulation influence the mass templates but on this side less substantial changes can be expected in the near future. More interesting would be the possibility to investigate the effect induced by a data-driven modelling of the $Z/\gamma^* \to \tau\tau$ contribution, achieved via an "embedding" technique already employed during Run I [95]. Future prospects for this analysis also include the study of a different discriminating variable. If the amount of collected data allows it, a combination of the visible τ_h mass and of $m_{\text{vis}}(\mu\tau_h)$ in a two-dimensional discriminator can for example be tested. An aspect that has been neglected so far is the fact that, as stated in section 5.2, the τ_h reconstruction in the different decay modes includes a selection on the visible mass of the candidate. Energy scale variations have an impact on this, since they move events inside and outside the mass selection window and allow for migrations between the different decay modes. A proper treatment of this aspect is technically very challenging because it requires to completely re-run the τ_h reconstruction sequence for each variation of the mass variable. A first step towards understanding the magnitude of this effect will consist in removing the candidates failing the mass selection requirement from the templates in the individual categories and studying the impact on the final result.

5.8 Measurement of the $Z \to \tau\tau$ cross section

The individual quantities determined so far can also be extracted in an analysis which aims at simultaneously measuring the inclusive Z production cross section in the $Z \to \tau\tau$ decay mode, the τ_h identification efficiency and the τ_h energy scale, exploiting the well-known Z boson signature as benchmark process. The advantage of this approach relies on the possibility of studying the

Decay Mode	Fit on $m_{\text{vis}}(\mu\tau_{\text{h}})$ [%]		Fit on $m_{\tau_{\text{h}}}(\mu\tau_{\text{h}})$ [%]	
All decay modes	$-0.8^{+0.4}_{-0.7}$		$-0.7^{+0.5}_{-0.6}$	
$h^{\pm}$	$+0.9^{+0.7}_{-1.0}$			
$h^{\pm}(\geq 1\pi^0)$	$-1.6^{+0.8}_{-0.9}$		$+0.5^{+0.5}_{-0.4}$	
$h^{\pm}h^{\mp}h^{\pm}$	$-1.1^{+1.2}_{-1.3}$		$-0.2^{+0.6}_{-0.6}$	
	$20 < p_{\text{T}} < 35\,\text{GeV}$	$p_{\text{T}} > 35\,\text{GeV}$	$20 < p_{\text{T}} < 35\,\text{GeV}$	$p_{\text{T}} > 35\,\text{GeV}$
All decay modes	$-2.2^{+1.2}_{-1.5}$	$-1.2^{+1.4}_{-0.5}$	$-0.6^{+0.8}_{-0.4}$	$-0.8^{+0.9}_{-1.1}$

Table 5.7: Results of the τ_{h} energy scale measurement, performed on the $m_{\text{vis}}(\mu\tau_{\text{h}})$ and $m_{\tau_{\text{h}}}$ distributions, for τ_{h} reconstructed in different decay modes and p_{T}-regions. When measuring the τ_{h}-ES for the combination of all decay modes using the $m_{\tau_{\text{h}}}$ variable, the candidates reconstructed in the $h^{\pm}$ decay mode are excluded from the templates. The results and uncertainties are expressed in percentages.

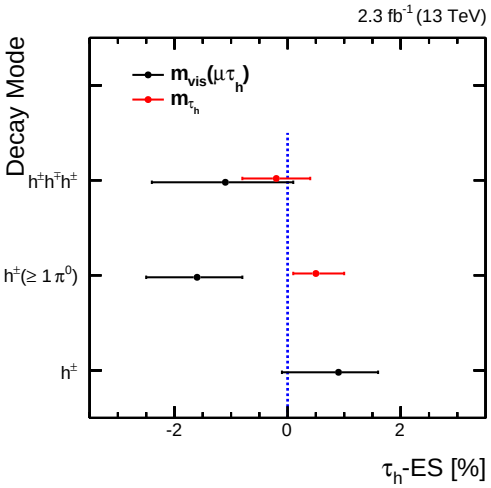


Figure 5.23: Results of the τ_{h} energy scale measurement, performed on $m_{\text{vis}}(\mu\tau_{\text{h}})$ and $m_{\tau_{\text{h}}}$, for τ_{h} candidates reconstructed in different decay modes. The blue dashed line corresponds to an energy scale variation of 0%.

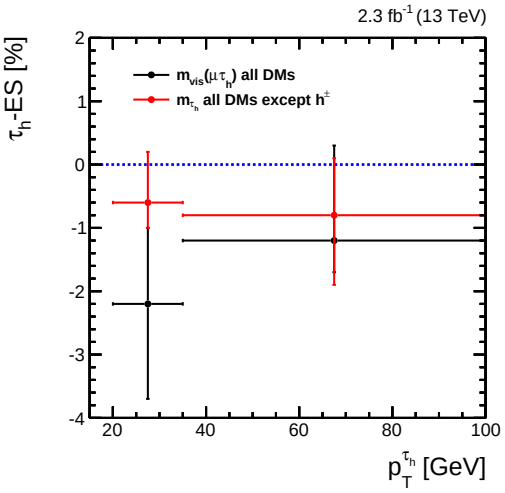


Figure 5.24: Results of the τ_{h} energy scale measurement, performed on $m_{\text{vis}}(\mu\tau_{\text{h}})$ and $m_{\tau_{\text{h}}}$, for τ_{h} candidates reconstructed in all the decay modes and in different ranges of the τ_{h} transverse momentum. When measuring the τ_{h}-ES using the $m_{\tau_{\text{h}}}$ variable, the candidates reconstructed in the $h^{\pm}$ decay mode are excluded. The blue dashed line corresponds to an energy scale variation of 0%.

combination of multiple $\tau\tau$ final states, which are affected by different systematic uncertainties and experimental issues, into a single measurement. This is an important and necessary milestone towards more complicated analyses, aiming at measuring the properties of the discovered Higgs boson in the decay mode with two τ leptons or at searching for additional resonances in this final state pointing to new physics beyond the Standard Model. A similar measurement, performed by the CMS collaboration in the data collected during the 7 TeV run in 2010, can be

found in [114].
From the experimental point of view, the different $Z \to \tau\tau$ final states listed in table 5.4 have advantages and disadvantages. The fully-hadronic decay mode $\tau_h\tau_h$ has by far the largest branching ratio but suffers from the contamination by QCD multijet events in which genuine taus come from hadron decays inside jets or in which one or more jets mimic the signature of a τ_h candidate. The ee and $\mu\mu$ final states have cleaner experimental signatures but this advantage is overcome by the low branching ratio and the overwhelming contamination coming from prompt $Z \to ee/\mu\mu$ decays, which are experimentally difficult to disentangle from the non-prompt decays of tau leptons. Moreover, the presence of four neutrinos in the final state worsens the resolution of the invariant mass $m_{\tau\tau}$. The fully-leptonic $e\mu$ final state has twice the branching ratio of the previous two and, due to the different lepton flavour, is ideally free from any $Z \to ee/\mu\mu$ contamination, even though the still limited branching ratio and the background from $t\bar{t}$ events remain an issue. The semileptonic decay modes $\ell\tau_h$ (with $\ell = e, \mu$) present the best compromise between branching ratio and a clear experimental signature. In this case, however, good control of the contamination coming from multijet events, as well as from $e \to \tau_h$ and $\mu \to \tau_h$ misidentification, is mandatory.
In what follows, only the $e\mu$, $e\tau_h$ and $\mu\tau_h$ final states are considered and, from now on, they will also be referred to as analysis "channels". Considering both semileptonic and fully leptonic final states is fundamental to allow the different channels to intercalibrate each other when measuring the τ_h identification efficiency (which affects the $\ell\tau_h$ channels but not $e\mu$), even if the τ_h identification efficiency and the $Z \to \tau\tau$ cross section are measured at the same time.

5.8.1 Object identification

In the $e\tau_h$ channel, events are triggered by the unprescaled `HLT_Ele23_WPLoose_Gsf` single lepton trigger, while for the $\mu\tau_h$ channel the unprescaled single lepton trigger `HLT_IsoMu18` is used. In the double-leptonic $e\mu$ channel, the events are selected if they pass either one of the following two combined triggers having asymmetric p_T-thresholds on the electron and muon legs: `HLT_Mu[17][8]_TrkIsoVVL_Ele[12][17]_CaloIdL_TrackIdL_IsoVL`.
Electron, muon and τ_h identification follows the same requirements outlined in section 5.4.2. In the $e\tau_h$ and $\mu\tau_h$ channels, a τ_h candidate with transverse momentum larger than 20 GeV and absolute value of the pseudorapidity smaller than 2.3 is required, together with an electron with $p_T > 24\,\text{GeV}$ and $|\eta| < 2.1$ or a muon with $p_T > 19\,\text{GeV}$ and $|\eta| < 2.1$. In the $e\mu$ channel an electron with $p_T > 13\,\text{GeV}$ and $|\eta| < 2.5$ and a muon with $p_T > 10\,\text{GeV}$ and $|\eta| < 2.4$ are required. If the event passes only one of the two combined triggers, the selection on the transverse momentum is raised to 18 GeV, for whichever object has fired the higher p_T leg of the trigger.
The impact of these p_T and $|\eta|$ selections on the $Z \to \tau\tau$ acceptance is summarized in table C.1 in the appendix, for all the three final states.
Depending on the channel, different additional selections are applied on the final state leptons.

$e\mu$ channel

Both the electron and muon are required to be isolated, with: $I^e_{rel} < 0.15$ and $I^\mu_{rel} < 0.15$, where I^e_{rel} and I^μ_{rel} are the relative electron and muon isolations, computed as in equation (5.3).

$e\tau_h$ channel

The electron is required to be isolated ($I^e_{rel} < 0.1$) and the τ_h candidate is required to pass the tight working point of the MVA isolation discriminator.
The contamination coming from electrons and muons which are misidentified as τ_h candidates is suppressed requiring the candidate to pass the MVA discriminator against electrons described in section 5.3 and the cut-based discriminator against muons. The tight working point is applied on the discriminator against electrons, while the loose working point is used for the discriminator against $\mu \to \tau_h$ misidentification.
In order to reduce $Z \to ee$ background, the event is rejected if there is an opposite-charge electron pair in which the electrons are separated by $\Delta R > 0.15$ and both pass loose identification requirements.

$\mu\tau_h$ channel

The muon is required to be isolated ($I^\mu_{rel} < 0.1$) and the τ_h candidate is required to pass the tight working point of the MVA isolation discriminator.
As in the $e\tau_h$ channel, the contamination coming from electrons and muons which are misidentified as τ_h candidates is suppressed via appropriate discriminators. The loose working point is applied on the MVA discriminator against electrons, while the tight working point is used for the cut-based discriminator against muon misidentification.
The $Z \to \mu\mu$ background contamination is reduced rejecting the events having an opposite-charge muon pair in which the muons are separated by $\Delta R > 0.15$ and both pass loose identification requirements.

5.8.2 Event selection

In order to avoid overlap between the different analyzed channels, the event is rejected if there are any additional electrons or muons passing looser identification requirements with respect to the ones described above. In this loose selection, the p_T requirement is lowered to 10 GeV for both electrons and muons, the selection on the relative isolation is relaxed to 0.3 and the electron multivariate identification variable is required to be above a working point having an efficiency of roughly 90% on genuine electrons. This ensures that, in the combination of all final states, no events are associated with more than one decay channel.
Finally, since the Z boson is electrically neutral, the leptons defining the channel are required to have opposite electric charge.

The transverse mass constructed from the electron or muon momentum and the missing transverse energy, as defined in equation (5.7), is a powerful variable for discriminating the $Z \to \tau\tau$ signal from the W+jet and $t\bar{t}$ backgrounds in the $\ell\tau_h$ channels (see also figure 5.12). A signal-enriched region in the $e\tau_h$ and $\mu\tau_h$ channels is defined requiring the events to have $m_T < 40$ GeV. In the $e\mu$ channel, where there are four neutrinos in the final state, the D_ζ variable defined in equation (5.16) is used to reduce the $t\bar{t}$ contamination, requiring it to be larger than -20 GeV.

The samples listed in table A.2 and table A.3 are used to simulate the signal and background processes, reweighting the simulated events to take into account possible mismodelling of the pile-up scenario or data/simulation differences in the efficiency of the identification, isolation and trigger requirements. For the more complicated trigger used in the $e\mu$ channel, the efficiency for each electron and muon leg is computed via a tag-and-probe measurement on $Z/\gamma^* \to$

$\ell\ell$ events similar to the one described in section 5.4.3, using double lepton triggers such as `HLT_Ele17_Ele12`. The efficiencies of the two legs are then combined according to the equation:

$$\begin{aligned}\epsilon(p_{\mathrm{T}},\eta) = \epsilon^{\texttt{Mu17}}(p_{\mathrm{T}},\eta)\cdot\epsilon^{\texttt{Ele12}}(p_{\mathrm{T}},\eta) + \epsilon^{\texttt{Mu8}}(p_{\mathrm{T}},\eta)\cdot\epsilon^{\texttt{Ele17}}(p_{\mathrm{T}},\eta) \\ - \epsilon^{\texttt{Mu17}}(p_{\mathrm{T}},\eta)\cdot\epsilon^{\texttt{Ele17}}(p_{\mathrm{T}},\eta)\,,\end{aligned} \tag{5.17}$$

which holds under the reasonable hypotheses that the triggers for both legs are uncorrelated and that the trigger with the lowest p_{T}-threshold is always more efficient than the one with higher threshold.
In the eμ channel, an additional kinematic weight is applied to the dominant $\mathrm{t\bar{t}}$ background to better match the simulated top quark p_{T} distribution observed in data with the theory prediction obtained at next-to-next-to-leading order (NNLO) in perturbation theory [115, 116]. This correction is neglected in the $\ell\tau_{\mathrm{h}}$ channels, where the $\mathrm{t\bar{t}}$ process contributes to less than 5% of the observed events.
The efficiencies of all the above mentioned selection criteria on $\mathrm{Z}\to\tau\tau$ events are summarized in table C.2, for all the analyzed final states.

5.8.3 Cross section measurement

Defining $\mathrm{Z}\to\tau\tau$ as the signal process, a simple model for the statistical interpretation of the data can be constructed using a dimensionless "signal strength modifier"

$$\mu = \frac{\sigma(\mathrm{pp}\to\mathrm{Z})\times\mathrm{BR}(\mathrm{Z}\to\tau\tau)}{\sigma_{\mathrm{exp}}(\mathrm{pp}\to\mathrm{Z})\times\mathrm{BR}_{\mathrm{exp}}(\mathrm{Z}\to\tau\tau)} \tag{5.18}$$

as parameter of interest in the likelihood function of equation (5.11). Here σ and BR are, respectively, the Z production cross section in pp collisions and the branching ratio for Z boson decays in the $\tau\tau$ final state. In the following, the branching ratios in equation (5.18) are absorbed in the cross section and a simplified notation $\mu = \sigma/\sigma_{\mathrm{exp}}$ is used. This strength modifier effectively controls the normalization of the $\mathrm{Z}\to\tau\tau$ signal process, with respect to the expectation given by σ_{exp}.
A template fit procedure, identical to the one which has been used for the other measurements presented in this work, is performed on mass templates constructed for the signal and background processes. In preparation for future measurements, especially those in the Higgs realm, the templates are produced from the distribution of the invariant mass $m_{\tau\tau}$ reconstructed by the SVFIT algorithm.

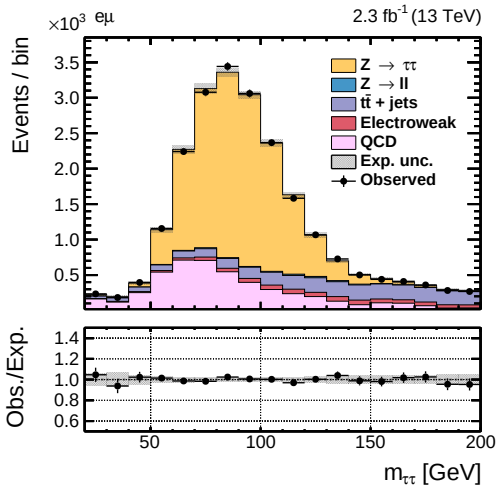
Systematic uncertainties are included as nuisance parameters $\vec{\theta}$ in the likelihood function. The systematic uncertainties considered for this measurement are largely based on the list given in section 5.4.6, although some meaningful modifications are needed, especially when considering the correlation between different uncertainties across the analyzed channels. The luminosity and the cross section uncertainties for the diboson, $\mathrm{t\bar{t}}$, single-top, W+jets and Drell-Yan processes are correlated among all channels. The theory-driven 4% uncertainty on the Drell-Yan cross section is however not associated with the $\mathrm{Z}\to\tau\tau$ process in this particular measurement but only with the $\mathrm{Z}\to\mathrm{ee}/\mu\mu$ contributions. The systematic uncertainties on the electron and muon identification, isolation and trigger efficiencies affect all the processes and final states with genuine electrons and muons and are correlated across the channels, neglecting the small differences due to the slightly different selections applied in each channel. The data-driven W+jets and QCD background estimation procedures are performed separately in each channel and the systematic uncertainties associated with these methods are kept uncorrelated in the fit.

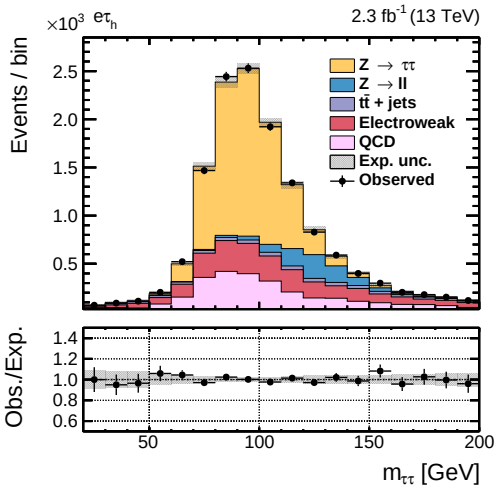
Systematic uncertainties on the $e \to \tau_h$ and $\mu \to \tau_h$ misidentification rates are included in the $e\tau_h$ and $\mu\tau_h$ channels, with a magnitude which varies depending on the working point considered. An uncertainty on the resolution and scale of the reconstructed $\vec{E}_T^{miss}$ is derived from studies of the hadronic recoil in $Z \to \mu\mu$ events in data and simulation: this results in additional 1-3% normalization uncertainties, depending on the process, which are correlated among all channels. The effect on the signal acceptance of theory uncertainties such as the choice for the parton distribution functions (PDFs), the α_S value and the renormalization and factorization scales μ_R and μ_F have been studied on simulated Drell-Yan events. The resulting effect, which is measured to be between 0.5% and 1.5%, is included as an additional systematic uncertainty correlated among all channels. The treatment of the uncertainty related to the τ_h identification efficiency requires some care, since there is an interplay between the τ_h identification efficiency and the signal strength modifier μ of equation (5.18), for the $e\tau_h$ and $\mu\tau_h$ channels. This uncertainty is split into a 3% contribution, which is kept uncorrelated between the $e\tau_h$ and $\mu\tau_h$ channels, and an additional component which is left unconstrained in the fit and is treated as a freely floating scale factor with respect to the nominal τ_h identification efficiency ($\tau^{ID}/\tau^{ID}_{exp}$).
Bin-by-bin uncertainties are included for all processes, to account for fluctuations due to the finite number of events in each bin of the templates used for signal extraction. A shape-altering systematic uncertainty related to the τ_h energy scale is set to 3%, affecting the invariant mass distribution of the $Z \to \tau\tau$ process in the $e\tau_h$ and $\mu\tau_h$ channels. Similarly, shape-altering uncertainties for the electron energy scale are included in the $e\mu$ channel, corresponding to a 1% uncertainty in the barrel and 2.5% in the endcap. For completeness, a shape-altering uncertainty related to the muon energy scale is included in the $e\mu$ channel as well, even though the expected contribution is only 1% across all the pseudorapidity range. Finally, the uncertainty on the reweighting of the $t\bar{t}$ simulated events based on the top quark p_T is included as an additional shape uncertainty in the $e\mu$ channel, varying the applied weights between no weight and two times the recommended weight.

Distributions, after the fit, of the invariant mass $m_{\tau\tau}$ are presented in figure 5.25, for all three channels. Additional plots, showing the distribution of the main observables, can be found in figure D.1, figure D.2 and figure D.3 in the appendix.
In figure 5.26 a scan of the negative logarithm of the likelihood ($-2\Delta\ln\mathcal{L}$) as a function of the parameter of interest $\mu = \sigma/\sigma_{exp}$ is shown. Unless explicitly stated, the other parameters of the likelihood are profiled during this procedure. This means that, for each point μ of the scan, the values $\vec{\theta}$ are chosen such that the likelihood $\mathcal{L}$ reaches its (local) maximum. In other words, a multi-dimensional maximization of the likelihood is performed in the vectorial space $\vec{\theta}$, for each point of the scan. For the plot of figure 5.26, three different scans have been performed, each time profiling a different set of nuisance parameters. In the first scan, all the nuisance parameters have been profiled for each value of μ, resulting in the black solid line. From this scan, the best-fit value and the total uncertainty on the parameter of interest can be determined as $\hat{\mu} = 0.936^{+0.053}_{-0.051}$, where the total uncertainty is extracted from the points at which $-2\Delta\ln\mathcal{L} = 1$. In a second scan of the negative log-likelihood, the uncertainty associated with the integrated luminosity has been fixed to its best-fit value, while all the other parameters are profiled. This results in the red dashed likelihood profile in figure 5.26. In the last scan, both the luminosity and all the other nuisance parameters have been fixed to their best-fit values, resulting in the blue dashed line. These different scans allow to decompose the total uncertainty in the luminosity, systematic (excluding the luminosity component) and statistical uncertainties as:

$$\hat{\mu} = 0.936\,^{+0.016}_{-0.015}(\text{lumi})\,^{+0.050}_{-0.047}(\text{syst})\,^{+0.010}_{-0.010}(\text{stat})\,.$$

From this, it can be seen that the precision on the result is dominated by systematic uncertainties.





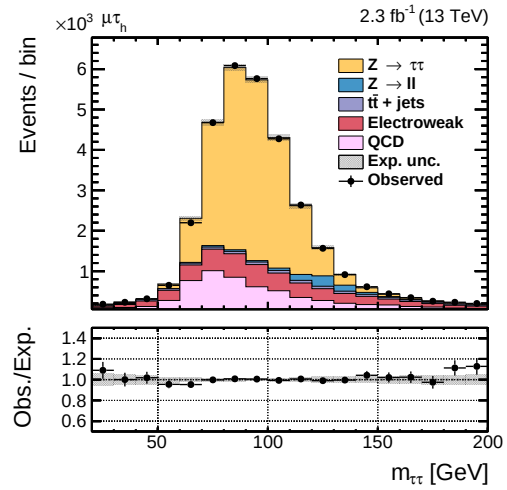


Figure 5.25: Distribution, after the fit, of the $\tau\tau$ pair invariant mass, in the eμ (top left), eτ_{h} (top right) and $\mu\tau_{\mathrm{h}}$ (bottom left) final states. Events are selected according to the criteria outlined in the text. The electroweak background contribution includes events from W+jets, diboson and single-top production.

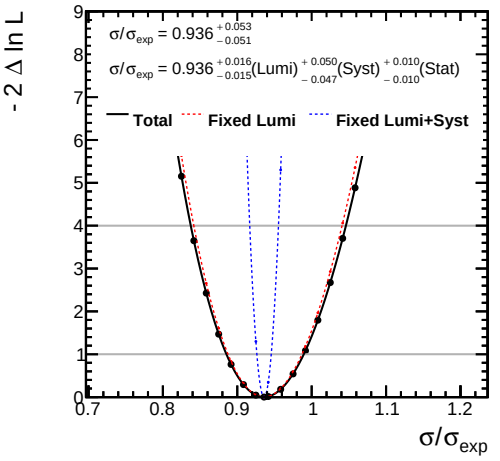


Figure 5.26: Scan of the negative log-likelihood $-2\Delta\ln\mathcal{L}$ as a function of the parameter of interest $\mu = \sigma/\sigma_{\mathrm{exp}}$. In the scan labelled as "total", all the nuisance parameters are profiled for each value of μ. In the scan represented by the dashed red line, all the nuisance parameters are profiled, except the luminosity parameter, which is fixed to its best-fit value. In the dashed blue scan, all the nuisance parameters are fixed to their best-fit values. The crossing of each curve with $-2\Delta\ln\mathcal{L} = 1$ defines the quoted uncertainties on the best-fit value.

This result can be converted into a cross section value via the following equation:

$$\sigma(\mathrm{pp} \to \mathrm{Z} + X) \times \mathrm{BR}(\mathrm{Z} \to \tau\tau) = \frac{\mathrm{N}^{\mathrm{sig}}}{\epsilon \cdot \mathcal{A} \cdot \mathrm{BR}' \cdot L}\,, \tag{5.19}$$

where $\mathrm{N}^{\mathrm{sig}}$ is the number of extracted signal events, $\mathcal{A}$ is the acceptance of signal events, ϵ is the combined efficiency of the identification, reconstruction and selection criteria on signal events,

BR' is the branching ratio of the considered $\tau\tau$ decay mode and L is the integrated luminosity. The measured cross section can be compared with the theoretical prediction for the inclusive Z production cross section at 13 TeV, computed at NNLO QCD with the program FEWZ 3.1 [117]. For the theoretical computation, the renormalization and factorization scales are set to the Z boson mass and a $60\,\text{GeV} < m_{\ell\ell} < 120\,\text{GeV}$ dilepton mass cut is applied. The theoretical value for the cross section times branching ratio is $1928 \pm 72\,\text{pb}$, where the main contribution to the theoretical uncertainty is given by the choice for the PDFs set. For a direct comparison, the number of extracted signal events N^{sig} is corrected for the fraction of events outside a $60\,\text{GeV} < m_{\tau\tau}^{gen} < 120\,\text{GeV}$ window on the mass of the generated leptons: $N = N^{sig} \cdot (1 - f_{out})$. The values computed for f_{out} are summarized in table C.3 in the appendix, for all the three analyzed final states.
The value of the cross section extracted from the fit is

$$\sigma(pp \to Z + X) \times BR(Z \to \tau\tau) = 1811 \pm 29\,(\text{lumi}) \pm 91\,(\text{syst}) \pm 18\,(\text{stat})\,\text{pb}\,.$$

This value is compatible with the NNLO theoretical prediction and consistent with a measurement of the inclusive Z production cross section in the $Z \to \mu\mu$ final state, performed by the CMS collaboration on the data collected at 13 TeV for a total integrated luminosity of $2.3\,\text{fb}^{-1}$ [118]. Figure 5.27 shows a comparison between the measured cross sections and the theoretical prediction. The results for the single $Z \to \tau\tau$ final states are obtained from a set of fits where individual signal strength modifiers $\mu = \sigma/\sigma_{exp}$ are defined separately in each analyzed channel. The likelihood function scans for these fits are reported in appendix D. The error bars associated with the points in figure 5.27 indicate the total uncertainty on the measurement. Statistical uncertainties are included in these error bars although, as already seen, they are the component which impacts less on the final precision of the $Z \to \tau\tau$ results and they are almost negligible in the $Z \to \mu\mu$ measurement. The inner error bars represent the component of the total uncertainty which is associated with the luminosity systematic. As it can be seen, this is the component currently dominating the precision of the result extracted from the $Z \to \mu\mu$ measurement. As expected from the much cleaner signature of this final state, the measurement performed in $Z \to \mu\mu$ events is more precise and even challenges the current precision of the theoretical prediction.

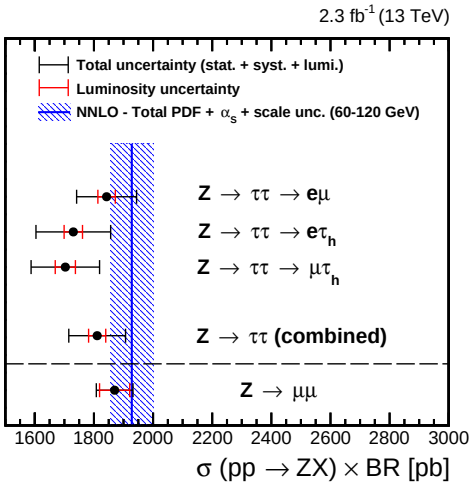


Figure 5.27: Summary of the measured $Z \to \tau\tau$ cross sections, in the individual analyzed final states and in the combination of all channels. The error bars associated with the points indicate the total uncertainty on the measurement. The inner error bar (in red) indicates the component of the total uncertainty which is due to the luminosity uncertainty. The measurements are compared with the theoretical prediction, represented by the blue vertical band. The value of the cross section measured in the $Z \to \mu\mu$ decay mode is taken from [118].

The impact of each nuisance parameter θ on the parameter of interest μ is defined as the shift $\Delta\mu$ that is induced if θ is fixed and brought to its $+1\sigma$ or -1σ post-fit values, while all other nuisance parameters are profiled as usual. This is useful not only to determine which nuisance parameters have the largest effect on μ, but also to obtain information about the correlation

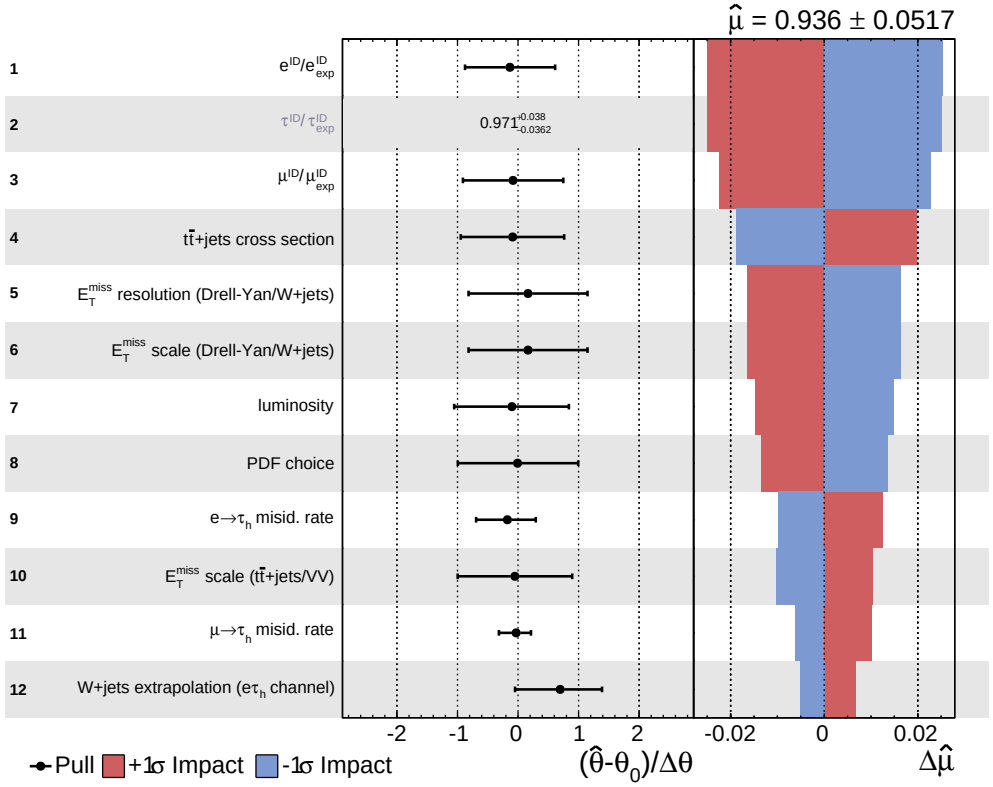


Figure 5.28: Impact of the nuisance parameters $\vec{\theta}$ on the parameter of interest $\mu = \sigma/\sigma_{\text{exp}}$. The left part of the plot shows the post-fit values $\hat{\theta}$ and the uncertainties $\Delta\theta$ of each nuisance parameter, with respect to the value θ_0 assumed prior to the fit. The right part of the plot shows the impact of a given uncertainty on the parameter of interest, defined as the shift $\Delta\hat{\mu}$ induced when varying the nuisance parameter by $\pm 1\sigma$. The nuisance parameters are ranked according to the impact that they have on the parameter of interest μ. For the sake of clarity, only the first 12 are shown.

between the nuisance parameters and the parameter of interest, since it can be verified if a positive shift in θ is translated into a positive variation of μ.
The impact of the most important nuisance parameters on $\mu = \sigma/\sigma_{\text{exp}}$ is shown in the right part of figure 5.28. The nuisance parameters associated with the uncertainties on the electron, muon and τ_{h} identification efficiency have the largest impact on the signal strength modifier (slightly above 2%), since they affect the normalization of all the processes with genuine leptons in the final state. As one could intuitively expect, these nuisance parameters are anticorrelated with the signal strength modifier. The $t\bar{t}$ cross section uncertainty influences the parameter of interest μ mainly via its relative importance in the eμ channel, where it is one of the most important backgrounds. The correlation with μ is interpreted as a result of the interplay with the lepton efficiencies: if the normalization of the $t\bar{t}$ contribution is increased, the effect on the lepton efficiency nuisances in the $\ell\tau_{\text{h}}$ channels tends to move in the opposite direction, which translates in a positive correlation with the signal strength modifier. Also moderate variations of the nuisance parameters associated with the $\vec{E}_{\text{T}}^{\text{miss}}$ scale and resolution have an impact on $\sigma/\sigma_{\text{exp}}$ which is among the top-ranked ones, albeit already below a 2% level.
The left part of figure 5.28 compares the post-fit value $\hat{\theta}$ of each nuisance parameter and the associated uncertainty $\Delta\theta$ with respect to the value θ_0 assumed prior to the fit. The quantity $(\hat{\theta} - \theta_0)/\Delta\theta$, represented by the black points and the associated error bars, is expected to be reasonably close to zero in case of a non-biased estimation of the nuisance parameter. This can

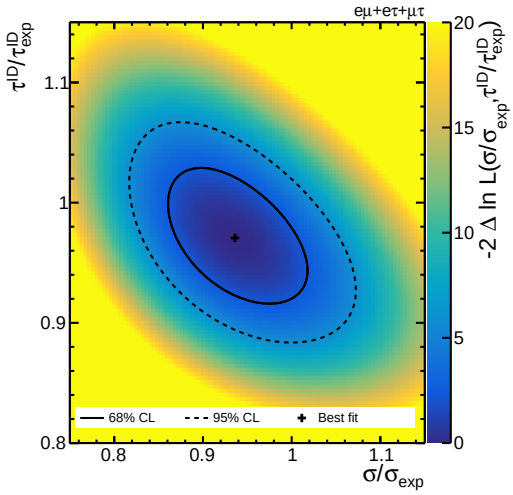


Figure 5.29: Likelihood contour for the joint parameter estimation of the cross section and the τ_h identification efficiency. The cross section and the τ_h identification efficiency are expressed with respect to the expectation σ_{exp} and τ^{ID}_{exp}, respectively. The 68% and 95% CL are represented by the solid and dashed black lines.

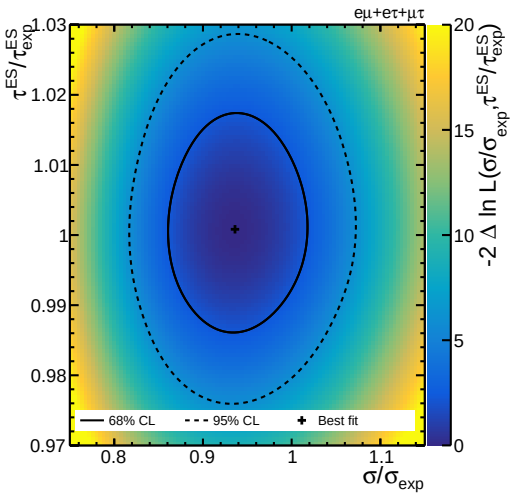


Figure 5.30: Likelihood contour for the joint parameter estimation of the cross section and the τ_h energy scale. The cross section and the τ_h energy scale are expressed with respect to the expectation σ_{exp} and τ^{ES}_{exp}, respectively. The 68% and 95% CL are represented by the solid and dashed black lines.

be confirmed for the results of this measurement. The size of the error bars, expressed in units of $\Delta\theta$ in the plot of figure 5.28, gives a useful indication of how strongly a single nuisance parameter is constrained by the fit, with respect to the prior knowledge of the uncertainty. It can be seen that the uncertainties associated with the e → τ_h and $\mu \to \tau_h$ misidentification rates are largely constrained during the fit procedure, with respect to the uncertainty assumed initially. This is expected, given the large value of the uncertainty associated with these systematics. Especially for the $\mu \to \tau_h$ misidentification rate, a very conservative uncertainty of 100% has been assumed (see section 5.4.6). As already stated, the nuisance associated with the τ_h identification efficiency is treated in the fit as a floating parameter, defined with respect to the expectation. The value assumed by this parameter after the fit, which is also reported in figure 5.28, is found to be equal to

$$\tau^{ID}/\tau^{ID}_{exp} = 0.971 \pm 0.038\,.$$

The extracted τ_h identification efficiency is well compatible with the expectation, with an associated uncertainty around 4%, equal or slightly smaller than the uncertainty usually obtained from tag-and-probe measurements in individual ττ final states, such as the one presented in section 5.6.

A two-dimensional scan of the negative log-likelihood can be performed on different combinations of parameters. Figure 5.29 shows the result of a scan performed on the signal strength modifier σ/σ_{exp} and on the parameter scaling the τ_h identification efficiency, $\tau^{ID}/\tau^{ID}_{exp}$. The solid and dashed black lines show the contours corresponding, respectively, to a coverage of 68% and

95%. As can be expected from their similar influence on the $Z \to \tau\tau$ signal, the two parameters are anticorrelated (correlation coefficient $\rho = -0.5$). Their effect on the overall $Z \to \tau\tau$ normalization can be disentangled only because of the inclusion into the measurement of the $e\mu$ channel, which is not influenced by the τ_h identification efficiency. Without this fully-leptonic channel, the two parameters would be highly anticorrelated ($\rho \approx -0.99$) and the ellipse of figure 5.29 would become almost degenerate.
Figure 5.30 shows a similar scan performed simultaneously on the signal strength modifier σ/σ_{exp} and on the parameter controlling the τ_h energy scale. This second parameter has been included in the fit model as a shape-altering uncertainty. It can be seen that the two parameters are uncorrelated (correlation coefficient $\rho = 0.02$). This is expected, given the fact that the signal strength modifier influences the overall $Z \to \tau\tau$ normalization, while the energy scale acts solely on the shape of the template.

5.8.4 Measurement of the τ_h identification efficiency and energy scale

A more precise estimate of the τ_h identification efficiency can be extracted assuming the value of the Z production cross section to be already known, as obtained from an additional measurement or from the theoretical prediction. In the following, the result of the CMS measurement in the $Z \to \mu\mu$ final state is used as a known constraint.
An additional 3.3% nuisance parameter, corresponding to the current precision of the $\sigma(pp \to Z + X) \times BR(Z \to \mu\mu)$ result [118], is associated to the $Z \to \tau\tau$ process while the cross section is fixed to the value measured in the $Z \to \mu\mu$ final state. A scan of the negative log-likelihood is performed with respect to $\tau^{ID}/\tau^{ID}_{exp}$, extracting the best-fit value and uncertainty on this scale factor. The result of the scan, which is shown in figure 5.31, gives

$$\tau^{ID}/\tau^{ID}_{exp} = 0.967 \pm 0.034\,.$$

The best-fit value is approximately 3% smaller than the expectation, although compatible within the uncertainty. This small tension with respect to the efficiency measurement performed in section 5.6, which showed a very good agreement between data and the expectation, can be explained with the fact that only a single discriminator has been probed in that particular measurement, while the $\tau^{ID}/\tau^{ID}_{exp}$ parameter here defined is sensitive to the combined effect of the full τ_h reconstruction and identification sequence including, for example, the decay mode reconstruction.

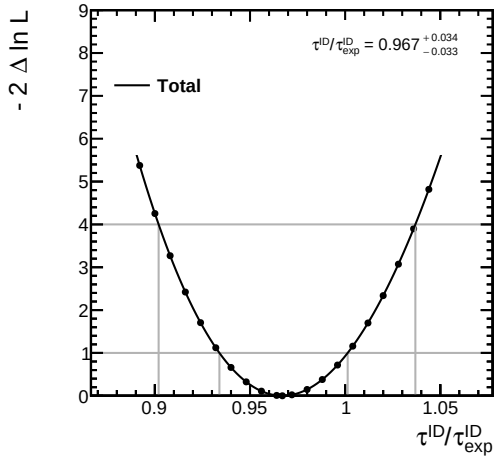


Figure 5.31: Scan of the negative log-likelihood $-2\Delta\ln\mathcal{L}$ as a function of the parameter of interest $\mu = \tau^{ID}/\tau^{ID}_{exp}$. All the nuisance parameters are profiled for each value of μ. The crossing of the solid black curve with $-2\Delta\ln\mathcal{L} = 1$ defines the quoted uncertainties on the best-fit value.

In figure 5.32 the parameter $\mu = \tau^{\mathrm{ES}}/\tau^{\mathrm{ES}}_{\mathrm{exp}}$, which modifies the τ_{h} energy scale with respect to the expectation, is defined as the parameter of interest of the likelihood and a scan similar to the above one is performed. Since this parameter has a pure shape-altering effect, fixing the cross section to the expectation has no influence on the result. The extracted scale factor is

$$\tau^{\mathrm{ES}}/\tau^{\mathrm{ES}}_{\mathrm{exp}} = 1.00 \pm 0.01\,.$$

The measured energy scale is perfectly consistent with the expectation. Translating the scale factor into an energy scale variation, the result is consistent with a 0% shift in the τ_{h} energy reconstruction. The associated 1% uncertainty has the same order of magnitude of what has been obtained from the measurement in the $\mu\tau_{\mathrm{h}}$ final state presented in section 5.7.

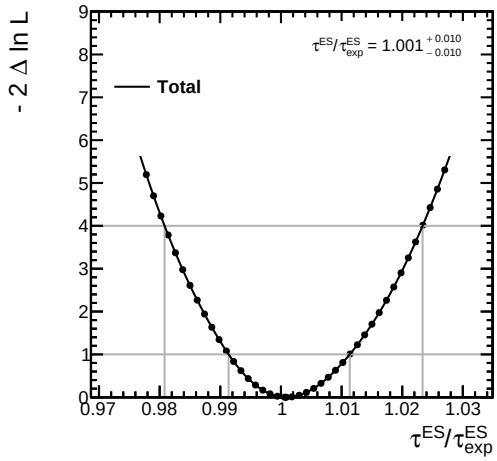


Figure 5.32: Scan of the negative log-likelihood $-2\Delta\ln\mathcal{L}$ as a function of the parameter of interest $\mu = \tau^{\mathrm{ES}}/\tau^{\mathrm{ES}}_{\mathrm{exp}}$. All the nuisance parameters are profiled for each value of μ. The crossing of the solid black curve with $-2\Delta\ln\mathcal{L} = 1$ defines the quoted uncertainties on the best-fit value.

Chapter 6

Conclusions

The excellent performance and stability of the CMS experiment at the Large Hadron Collider led to major scientific achievements already after very few years of activity. However, new experimental challenges driven by the brilliant progress in the LHC operations need to be faced in the near future. In particular, the expected increase in luminosity of the collider, which is intended to exceed the nominal value by a factor of two in the next years, calls for a complete replacement of the very inner core of CMS: the pixel detector. The new pixel detector introduces a number of improvements compared to the current one as, for example, a new design of the readout chip, an additional layer for improving vertexing and track reconstruction and an overall reduction of the material budget. The Karlsruhe Institute of Technology, as a production center, is responsible for the assembly of half of the pixel modules that will be arranged in the outermost barrel layer of the new detector. The modules are based on a hybrid approach in which the pixelated silicon sensor is mechanically and electrically joined to the readout electronics. In the context of this thesis, an in-house packaging process for the interconnection of the readout chips with the sensor substrate, using the bump bonding technology, has been developed and applied to the module production. Quality control procedures have been defined and implemented, with the purpose of identifying problems at an earlier stage in the bonding sequence thus improving the overall quality of the assemblies. An active contribution has been provided during the production of a total of more than 400 pixel bare modules composed of a sensor connected to readout chips, which have been found to possess excellent mechanical and electrical properties. The majority of these bare modules have been completed into fully equipped and tested pixel modules which, from the end of the year, will be integrated in the future detector.
Detector upgrades are however only a single aspect of a constant improvement process which has the ultimate goal of extending the physics reach of the CMS experiment in the quest for understanding the mysteries of nature. With the restart of operations in the summer of 2015, the LHC provided proton-proton collisions at the unprecedented center of mass energy of 13 TeV. Another fundamental activity consists therefore in developing new methods for the reconstruction and identification of the physics objects under these new running conditions. Tau leptons, in particular, are complex objects whose reconstruction involves different detector elements and which are prone to be wrongly identified. Real electrons radiating bremsstrahlung photons, for example, can be easily misidentified as the charged and neutral components of a hadronic tau decay τ_{h}, therefore constituting a sizeable source of background for all the analyses aiming at identifying tau leptons in their final states. A multivariate discriminator, trained for rejecting electrons which are wrongly reconstructed as τ_{h} candidates, has been developed and introduced in the CMS tau reconstruction sequence. The new discriminator proved to maintain a high efficiency on hadronic tau decays, while keeping the electron misdentification probability at the

percent or sub-percent level, depending on the chosen working point. The well-known Drell-Yan process has been used as a reference for validating the performance of the new anti-electron discriminator in the data collected by the CMS experiment in 2015, computing scale factors to account for the observed differences between data and simulation and evaluating the uncertainties associated with the measurement. Using a similar tag-and-probe technique, the efficiency of a multivariate τ_{h} isolation discriminator has been measured in data as well, finding good agreement with the expectations. One of the main sources of systematic uncertainties related to the reconstruction of hadronic tau decays, namely the τ_{h} energy scale, has been measured in data in all the reconstructed τ_{h} decay modes. Part of the results obtained in this set of measurements has been included in a document which summarizes the performance of the new CMS τ_{h} identification algorithms in the data collected in 13 TeV proton-proton collisions and which is about to be made public [119]. A deeper validation of the tau identification performance has been obtained in a global determination of the $\mathrm{Z} \to \tau\tau$ cross section, of the τ_{h} identification efficiency and of the τ_{h} energy scale, based on the statistical combination of three different $\tau\tau$ final states. The extracted value for $\sigma(\mathrm{Z} \to \tau\tau) \times \mathrm{BR}(\mathrm{Z} \to \tau\tau)$ has been found to be compatible with the NNLO theoretical prediction and consistent with a published CMS measurement of the inclusive Z production cross section in the $\mathrm{Z} \to \mu\mu$ final state. Constraining the value of the $\mathrm{Z} \to \tau\tau$ cross section with this independent measurement, a more precise estimate of the identification efficiency for hadronic tau decays and of the τ_{h} energy scale has been determined as a by-product, with respective precisions of 3.4% and 1%. The good agreement with the results which have been obtained using different techniques, such as the tag-and-probe measurement in $\mathrm{Z} \to \tau\tau \to \mu\tau_{\mathrm{h}}$ events, allows to draw a coherent picture of the CMS tau reconstruction and identification performance. These results prove the robustness of the identification algorithms and pave the way for future $H \to \tau\tau$ measurements, where the τ_{h} energy scale and the τ_{h} identification efficiency are the most important systematic uncertainties.

Appendix A

Datasets and simulated samples

A.1 Datasets

Analyzed data samples with the corresponding runs and the total integrated luminosity. Each run is unambiguously defined as a period of time inside a given LHC fill where the data acquisition system of the experiment has been constantly operational, without interruptions or resets. The datasets in which the events are stored are labelled as `SingleElectron`, `SingleMuon` or `MuonEG`, depending on the flavour of the lepton identified by the trigger. The label `MINIAOD` refers to a CMS-specific dataformat, whose details are not discussed in this context. The uncertainty on the total integrated luminosity has been estimated in a dedicated analysis [111].

complete dataset name: /[SingleElectron][SingleMuon][MuonEG]/XXX/MINIAOD		
/XXX/	run range	L [fb^{-1}]
/Run2015D-16Dec2015-v1/	256630 – 260627	2.301 ± 0.062

Table A.1: Summary of the analyzed datasets, together with the corresponding run ranges and the total integrated luminosity.

A.2 Simulated samples

In the following tables, the full set of simulated samples used for the measurements presented in chapter 5 are listed. Refer to the text and to the references therein for a documentation of the generator libraries. The label `MINIAODSIM` refers to a CMS-specific dataformat, whose details are not discussed in this context.
Table A.2 lists the full set of $\mathrm{Z}/\gamma^* \to \ell\ell$ (with $\ell = \mathrm{e}, \mu, \tau$) samples used for simulating the Drell-Yan process in all the three leptonic final states. Table A.3 lists the samples used for all the remaining processes.
The equivalent luminosity L^{eq} is computed from the total number of generated events and the cross section of the given process as:

$$L^{\mathrm{eq}} = \frac{N^{\mathrm{gen}}_{\mathrm{sim.}}}{\sigma} \,. \tag{A.1}$$

Drell-Yan samples

complete sample name: /XXX/RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12*/MINIAODSIM		
/XXX/	σ [pb]	L^{eq} [fb^{-1}]
/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/	18610	~ 3.3
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	6025.2	~ 41.0
/DY1JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	1231.4	~ 53.0
/DY2JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	404.8	~ 49.5
/DY3JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	123.8	~ 46.1
/DY4JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	66.6	~ 62.9
/DYJetsToLL_M-150_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	6.657	~ 913.2

Table A.2: List of the simulated samples used for the measurements described in the text, together with their cross sections and the equivalent integrated luminosity. In this table: Drell-Yan process, inclusive samples in different mass regions and exclusive samples, binned according to the jet multiplicity.

Other samples

complete sample name: /XXX/RunIIFall15MiniAODv2-PU25nsData2015v1_76X_mcRun2_asymptotic_v12*/MINIAODSIM		
/XXX/	σ [pb]	L^{eq} [fb^{-1}]
/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	61526.7	~ 0.8
/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	9644.5	~ 4.7
/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	3144.5	~ 9.6
/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	954.8	~ 20.0
/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	485.6	~ 18.5
/TT_TuneCUETP8M1_13TeV-powheg-pythia8/	831.76	~ 225.6
/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/	35.6	~ 28.1
/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/	35.6	~ 28.1
/ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1/	44.07	~ 74.9
/ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1/	26.2	~ 62.2
/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8/	11.95	~ 238.9
/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/	3.22	~ 4752.1
/ZZTo4L_13TeV-amcatnloFXFX-pythia8/	1.212	~ 8866.7
/WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/	49.997	~ 104.9
/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/	5.595	~ 4620.3
/WZJToLLLNu_TuneCUETP8M1_13TeV-amcnlo-pythia8/	4.666	~ 415.2
/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/	3.05	~ 558.6
/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/	10.71	~ 1843.4

Table A.3: List of the simulated samples used for the measurements described in the text, together with their cross sections and the equivalent integrated luminosity. In this table: W+jets, $t\bar{t}$, single-top and diboson (WW, WZ, ZZ) samples. The W+jets process is modelled by an inclusive sample and a set of exclusive samples, binned according to the jet multiplicity.

Appendix B

Supporting material for the τ_{h} energy scale measurement

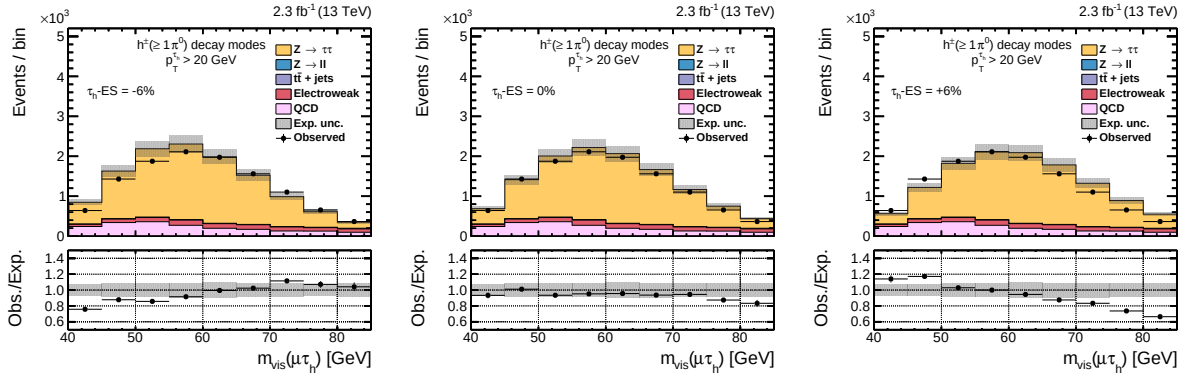


Figure B.1: Distribution of $m_{\mathrm{vis}}(\mu\tau_{\mathrm{h}})$, the invariant mass reconstructed from the muon and the visible components of the τ_{h} decay, for different τ_{h}-ES variations: -6% (left), 0% (center) and $+6\%$ (right). The τ_{h} candidates are required to be reconstructed in the $h^{\pm}\pi^{0}$ or $h^{\pm}\pi^{0}\pi^{0}$ decay modes, to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$ and to pass the tight WP of the MVA isolation discriminator.

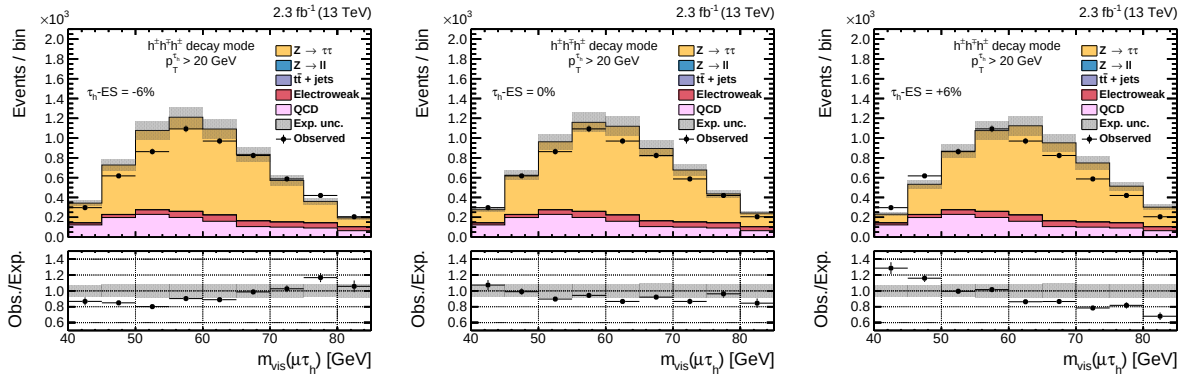


Figure B.2: Distribution of $m_{\mathrm{vis}}(\mu\tau_{\mathrm{h}})$, the invariant mass reconstructed from the muon and the visible components of the τ_{h} decay, for different τ_{h}-ES variations: -6% (left), 0% (center) and $+6\%$ (right). The τ_{h} candidates are required to be reconstructed in the $h^{\pm}h^{\mp}h^{\pm}$ decay mode, to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$ and to pass the tight WP of the MVA isolation discriminator.

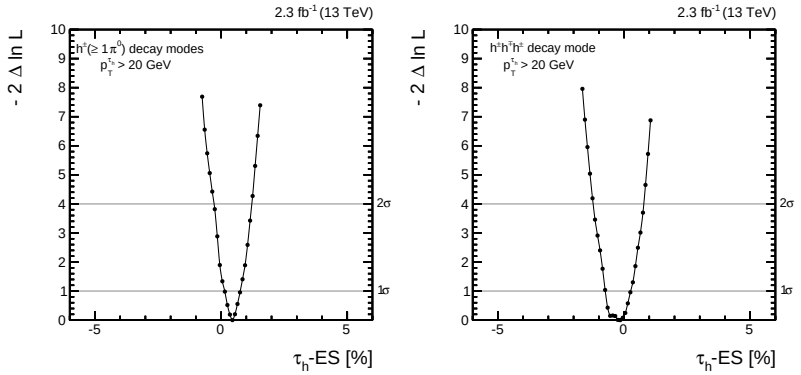


Figure B.3: Likelihood scans for the τ_h-ES measurement using the m_{τ_h} variable. The τ_h candidates are required to have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$, to pass the tight WP of the MVA isolation discriminator and to be reconstructed respectively in the $h^{\pm}\pi^0$ or $h^{\pm}\pi^0\pi^0$ decay modes (left) or in the $h^{\pm}h^{\mp}h^{\pm}$ decay mode (right).

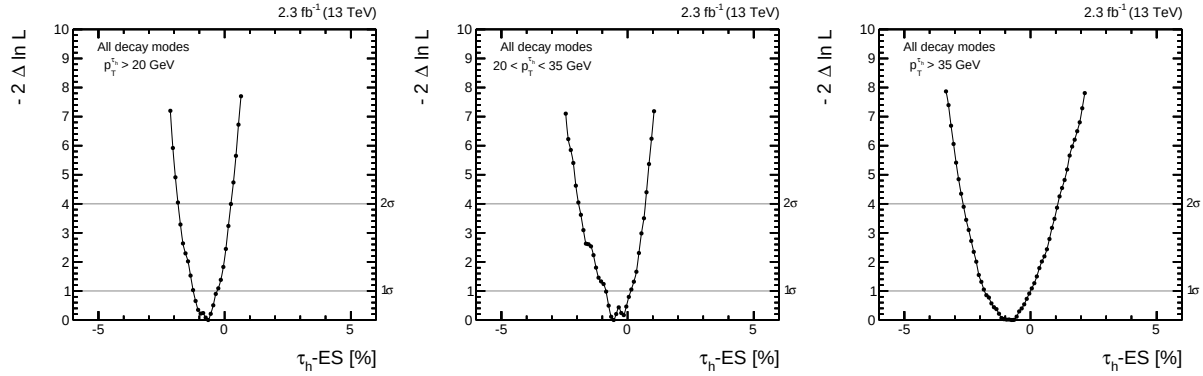


Figure B.4: Likelihood scans for the τ_h-ES measurement using the m_{τ_h} variable. The τ_h candidates are required to have $|\eta| < 2.3$, to pass the tight WP of the MVA isolation discriminator, to be reconstructed in any decay mode except the $h^{\pm}$ decay mode and to have $p_T > 20\,\text{GeV}$ (left), $20\,\text{GeV} < p_T < 35\,\text{GeV}$ (center) or $p_T > 35\,\text{GeV}$ (right).

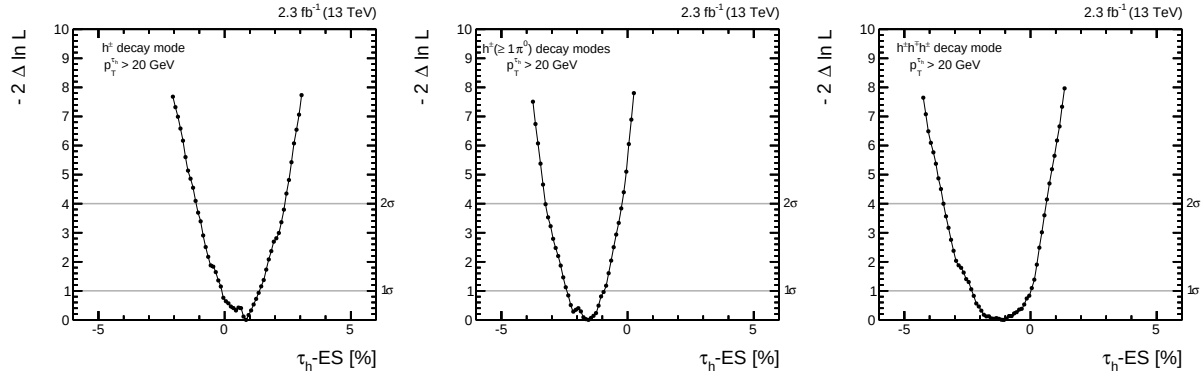


Figure B.5: Likelihood scans for the τ_h-ES measurement using the $m_{\text{vis}}(\mu\tau_h)$ variable. The τ_h candidates are required to have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$, to pass the tight WP of the MVA isolation discriminator and to be reconstructed respectively in the $h^{\pm}$ decay mode (left), in the $h^{\pm}\pi^0$ or $h^{\pm}\pi^0\pi^0$ decay modes (center) or in the $h^{\pm}h^{\mp}h^{\pm}$ decay mode (right).

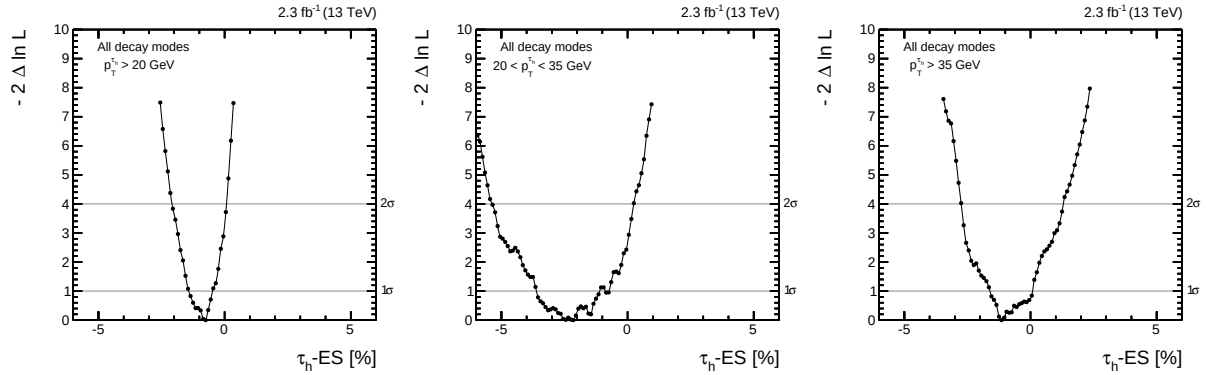


Figure B.6: Likelihood scans for the τ_{h}-ES measurement using the $m_{\mathrm{vis}}(\mu\tau_{\mathrm{h}})$ variable. The τ_{h} candidates are required to have $|\eta| < 2.3$, to pass the tight WP of the MVA isolation discriminator, to be reconstructed in any decay mode and to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ (left), $20\,\mathrm{GeV} < p_{\mathrm{T}} < 35\,\mathrm{GeV}$ (center) or $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ (right).

Appendix C

Signal acceptances and selection efficiencies on $\mathrm{Z} \to \tau\tau$ events

C.1 Signal acceptances

The signal acceptance $\mathcal{A}$, for a generic $\mathrm{Z} \to \tau\tau$ final state XX, can be computed on the simulated $\mathrm{Z}/\gamma^* \to \tau\tau \to XX$ Drell-Yan events whose generated $\tau\tau$ mass is in the region $60\,\mathrm{GeV} < m_{\tau\tau}^{\mathrm{gen}} < 120\,\mathrm{GeV}$, requiring the generator-level final state leptons to pass the p_{T} and $|\eta|$ selections applied in the analysis. The term "final state lepton" indicates the status of the simulated particle after the emission of initial or final state radiation.
The acceptance definition is summarized in the following equation:

$$\mathcal{A} \equiv \frac{\text{denominator} \quad \&\& \quad \text{passes } p_{\mathrm{T}}/|\eta| \text{ selections}}{\mathrm{Z}/\gamma^* \to \tau\tau \to XX \text{ on generator level} \quad \&\& \quad (60 < m_{\tau\tau}^{\mathrm{gen}} < 120)}\,, \tag{C.1}$$

where the "denominator" keyword is used as a shortcut to avoid repeating the whole expression in the denominator of equation (C.1).
Acceptances are computed for each one of the final states considered in the analysis ($\mathrm{e}\mu$, $\mathrm{e}\tau_{\mathrm{h}}$ or $\mu\tau_{\mathrm{h}}$), using the set of MadGraph5 Drell-Yan samples listed in table A.2. After requiring that $60\,\mathrm{GeV} < m_{\tau\tau}^{\mathrm{gen}} < 120\,\mathrm{GeV}$, events are selected if they contain: a generated final state electron with $p_{\mathrm{T}} > 13\,\mathrm{GeV}$ and $|\eta| < 2.5$ and a generated final state muon with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$ (for the acceptance definition in the $\mathrm{e}\mu$ channel); a generated final state electron with $p_{\mathrm{T}} > 24\,\mathrm{GeV}$ and $|\eta| < 2.1$ or a generated final state muon with $p_{\mathrm{T}} > 19\,\mathrm{GeV}$ and $|\eta| < 2.1$ and a generated τ_{h} with visible $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.3$ (for the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels, respectively). The acceptance values in all the analyzed final states are shown in table C.1. Due to the large number of simulated events, the statistical uncertainties on the numbers in the table are negligible ($< 1\%$).

	$\mathrm{e}\mu$	$\mathrm{e}\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{h}}$
Acceptance $\mathcal{A}$	0.129	0.046	0.075

Table C.1: Acceptance of the p_{T} and $|\eta|$ requirements on $\mathrm{Z} \to \tau\tau$ events, in the $\mathrm{e}\mu$, $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states.

C.2 Signal efficiencies

The combined efficiency ϵ, for a generic $Z \to \tau\tau$ final state XX, is defined on the simulated $Z/\gamma^* \to \tau\tau \to XX$ Drell-Yan events whose generated $\tau\tau$ mass is in the region $60\,\text{GeV} < m_{\tau\tau}^{\text{gen}} < 120\,\text{GeV}$ and where the generator-level final state leptons satisfy the acceptance selections defined above, requiring them to pass all the identification, isolation and selection criteria applied in the analysis.
The definition is summarized in the following equation:

$$\epsilon \equiv \frac{\text{denominator} \quad \&\& \quad \text{passes id/iso/selection requirements}}{\text{numerator of acceptance definition } \mathcal{A}}\,. \tag{C.2}$$

The "denominator" keyword is used as a shortcut to avoid repeating the whole expression in the denominator of equation (C.2).
The efficiency of the selections applied in each analyzed final state, divided into the single components and combined together, are shown in table C.2. Due to the large number of simulated events, the statistical uncertainties on the numbers in the table are negligible ($< 1\%$).

	$e\mu$	$e\tau_h$	$\mu\tau_h$
Trigger	0.673 (1.000)	0.673 (1.000)	0.886 (1.000)
e identification	0.559 (0.831)	0.611 (0.908)	-
μ identification	0.554 (0.991)	-	0.863 (0.974)
τ_h identification	-	0.585 (0.957)	0.831 (0.963)
Best pair selection	0.487 (0.879)	0.364 (0.622)	0.524 (0.631)
Opposite charge requirement	0.485 (0.996)	0.350 (0.962)	0.508 (0.969)
e isolation	0.465 (0.959)	0.338 (0.966)	-
μ isolation	0.435 (0.935)	-	0.488 (0.961)
τ_h MVA isolation discriminator	-	0.229 (0.678)	0.336 (0.689)
τ_h anti-e/μ discriminators	-	0.179 (0.782)	0.295 (0.878)
$Z/\gamma^* \to \ell\ell$ rejection	-	0.178 (0.994)	0.295 (1.000)
Veto on extra leptons	0.435 (1.000)	0.178 (1.000)	0.295 (1.000)
$D_\zeta > -20\,\text{GeV}$	0.389 (0.894)	-	-
$m_T < 40\,\text{GeV}$	-	0.151 (0.848)	0.265 (0.898)
Combined efficiency ϵ	0.389	0.151	0.265

Table C.2: Efficiency of the selections on $Z \to \tau\tau$ events, in the $e\mu$, $e\tau_h$ and $\mu\tau_h$ final states. The efficiency of each selection requirement is also shown. The number in brackets indicates the relative efficiency with respect to the selection applied immediately before.

C.3 f_{out} fraction

In the $Z \to \tau\tau$ cross section measurement, the number of extracted signal events N^{sig} has to be corrected for the small fraction of events outside the generator-level mass window. This event fraction is defined as f_{out} and computed as:

$$f_{\text{out}} \equiv 1 - \frac{\text{denominator} \quad \&\& \quad \text{passes } p_T/|\eta| \text{ selections} \quad \&\& \quad (60 < m_{\tau\tau}^{\text{gen}} < 120)}{Z/\gamma^* \to \tau\tau \to XX \text{ on gen. level} \quad \&\& \quad \text{passes id/iso/selection requirements}}\,. \tag{C.3}$$

The "denominator" keyword is used as a shortcut to avoid repeating the whole expression in the denominator of equation (C.3).
The values for f_{out} in all the analyzed final states are shown in table C.3.

	$e\mu$	$e\tau_h$	$\mu\tau_h$
f_{out}	0.029	0.069	0.062

Table C.3: Fraction of signal events outside the generator-level mass window $60\,\text{GeV} < m_{\tau\tau}^{\text{gen}} < 120\,\text{GeV}$, in the $e\mu$, $e\tau_h$ and $\mu\tau_h$ final states.

Appendix D

Supporting material for the $Z \to \tau\tau$ cross section measurement

Figures D.1, D.2 and D.3 show the distribution of the main observables in the eμ, eτ_h and $\mu\tau_h$ final states, respectively. Events are selected according to the requirements described in section 5.8, correcting the simulation to take into account possible mismodelling of the pile-up scenario or known data/simulation differences in the efficiency of the identification, isolation and trigger requirements. The backgrounds are estimated according to the prescription of section 5.4.4. The electroweak background contribution includes events from W+jets, diboson and single-top production.

Figure D.4 shows the scan of the negative log-likelihood $-2\Delta\ln\mathcal{L}$ as a function of a signal strength modifier $\mu = \sigma/\sigma_{exp}$ defined separately in each analyzed channel. In the model used for these scans, the nuisance parameter associated with the τ_h identification efficiency gets an additional 5% constraint. This is needed in order to be able to distinguish between the cross section scaling and a scaling of the τ_h identification efficiency, in the eτ_h and $\mu\tau_h$ channels. A separation of the total uncertainty on the best-fit value into a luminosity-related uncertainty and a systematic uncertainty, identical to what described in section 5.8, is performed.

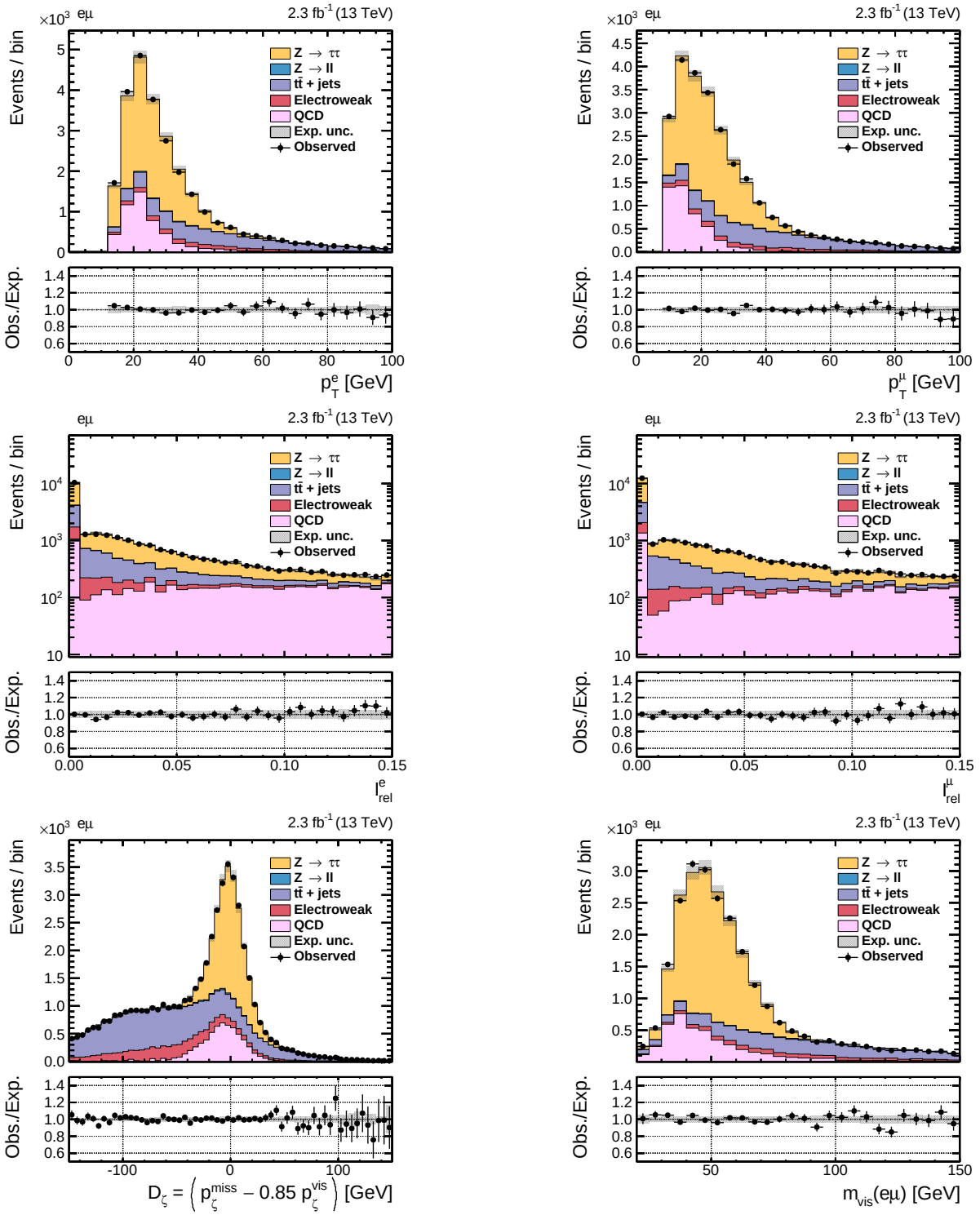


Figure D.1: Distribution of the main observables in the eμ channel. From top to bottom and from left to right: electron p_{T} and muon p_{T}, electron relative isolation $\mathrm{I}^{\mathrm{e}}_{\mathrm{rel}}$, muon relative isolation $\mathrm{I}^{\mu}_{\mathrm{rel}}$, D_ζ variable and invariant mass of the eμ pair. The events are required to pass all the selection criteria described in section 5.8. In the D_ζ plot, all the selections except the cut on this variable are applied.

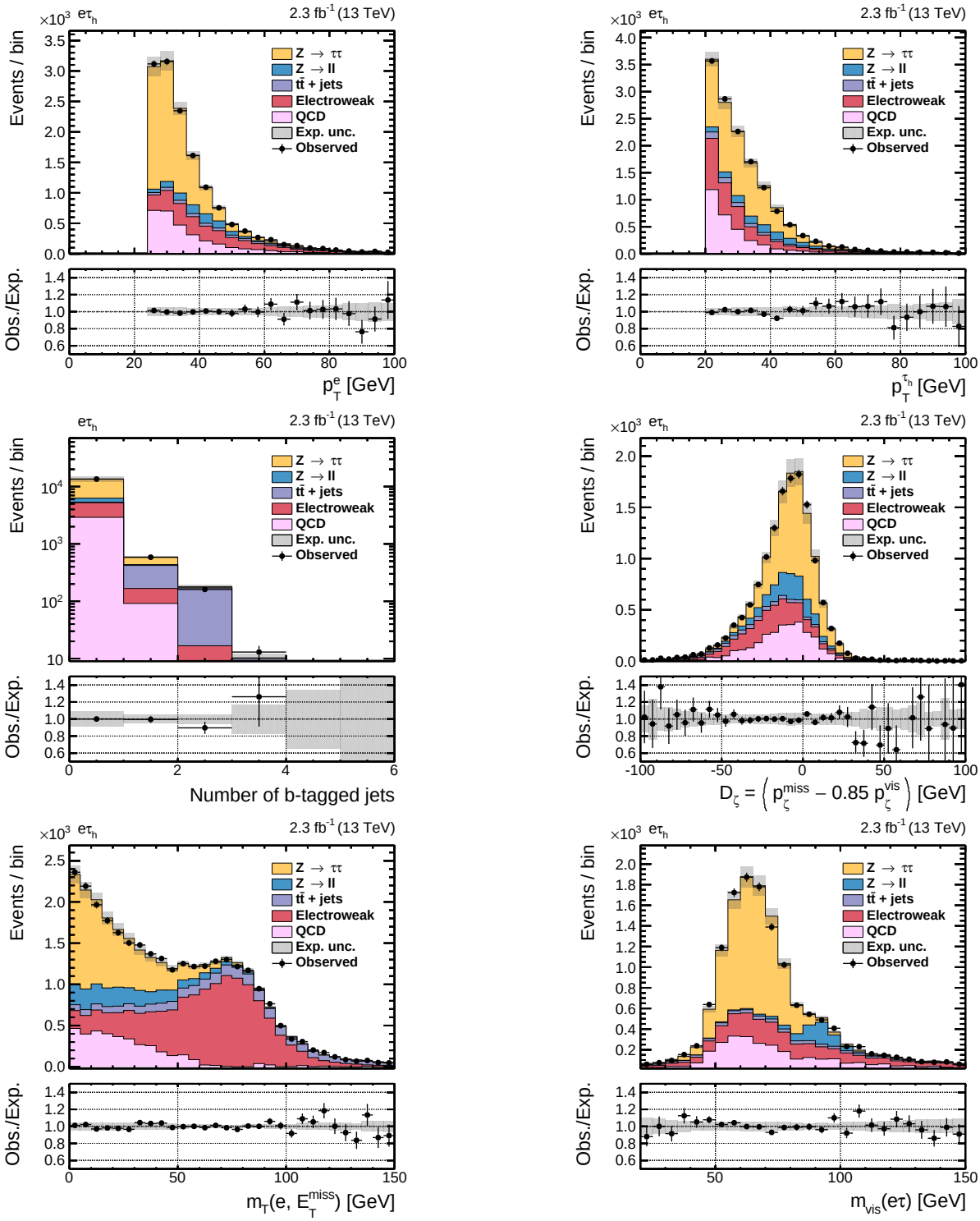


Figure D.2: Distribution of the main observables in the eτ_h channel. From top to bottom and from left to right: electron p_T and τ_h candidate p_T, number of b-tagged jets, D_ζ variable, $m_T(e, E_T^{miss})$ and invariant mass built from the electron and the visible components of the τ_h decay. The events are required to pass all the selection criteria described in section 5.8. In the $m_T(e, E_T^{miss})$ plot, all the selections except the cut on this variable are applied.

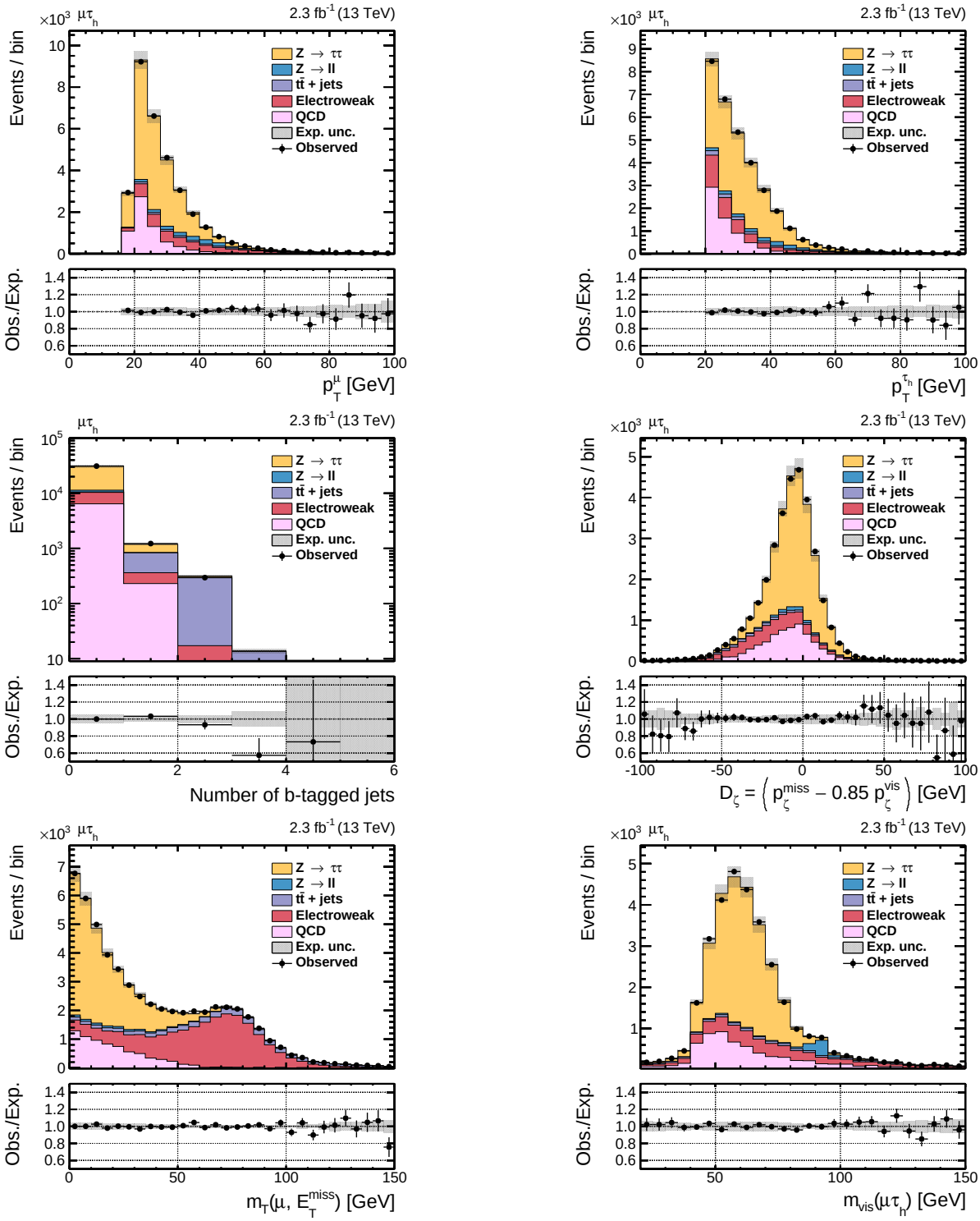
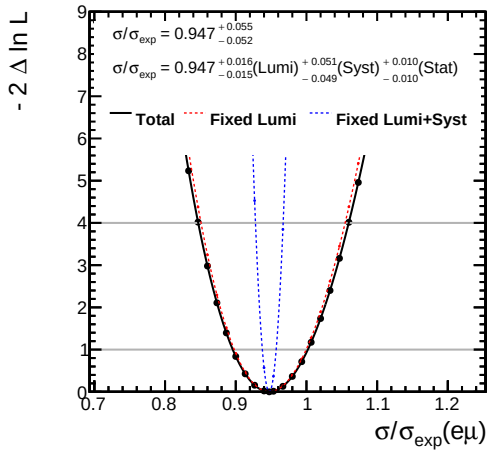
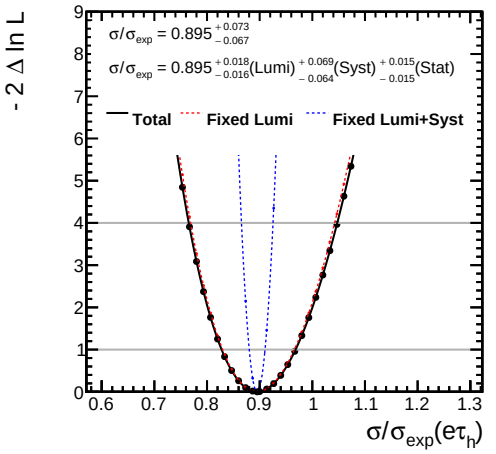


Figure D.3: Distribution of the main observables in the $\mu\tau_{\mathrm{h}}$ channel. From top to bottom and from left to right: muon p_{T} and τ_{h} candidate p_{T}, number of b-tagged jets, D_{ζ} variable, $m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}})$ and invariant mass built from the muon and the visible components of the τ_{h} decay. The events are required to pass all the selection criteria described in section 5.8. In the $m_{\mathrm{T}}(\mu, E_{\mathrm{T}}^{\mathrm{miss}})$ plot, all the selections except the cut on this variable are applied.





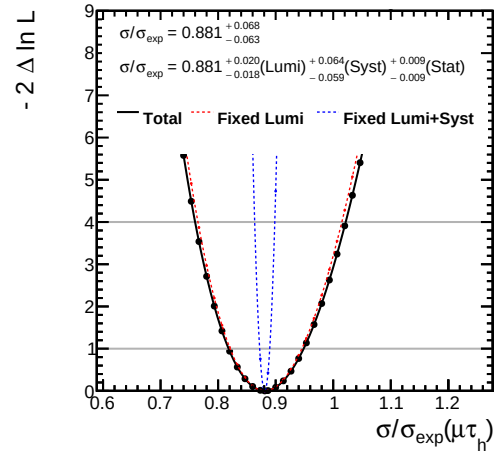


Figure D.4: Scan of the negative log-likelihood $-2\Delta\ln\mathcal{L}$ as a function of the parameter of interest $\mu = \sigma/\sigma_{\text{exp}}$, in each channel. In the scan labelled as "total", all the nuisance parameters are profiled for each value of μ. In the scan represented by the dashed red line, all the nuisance parameters are profiled, except the luminosity parameter, which is fixed to its best-fit value. In the dashed blue scan, all the nuisance parameters are fixed to their best-fit values. The crossing of each curve with $-2\Delta\ln\mathcal{L} = 1$ defines the quoted uncertainties on the best-fit value.

List of Figures

List of Tables

References

[1] Particle Data Group, "Review of Particle Physics", *Chin. Phys.* **C38** (2014) 090001. doi:10.1088/1674-1137/38/9/090001.

[2] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys. Rev.* **127** (1962) 965–970. doi:10.1103/PhysRev.127.965.

[3] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323. doi:10.1103/PhysRevLett.13.321.

[4] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509. doi:10.1103/PhysRevLett.13.508.

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587. doi:10.1103/PhysRevLett.13.585.

[6] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (1995) 2626–2631, arXiv:hep-ex/9503002. doi:10.1103/PhysRevLett.74.2626.

[7] UA1 Collaboration, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV", *Phys. Lett.* **B122** (1983) 103–116. doi:10.1016/0370-2693(83)91177-2.

[8] UA2 Collaboration, "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p} - p$ Collider", *Phys. Lett.* **B122** (1983) 476–485. doi:10.1016/0370-2693(83)91605-2.

[9] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider", *Phys. Lett.* **B126** (1983) 398–410. doi:10.1016/0370-2693(83)90188-0.

[10] UA2 Collaboration, "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p} - p$ Collider", *Phys. Lett.* **B129** (1983) 130–140. doi:10.1016/0370-2693(83)90744-X.

[11] SLD Electroweak Group, SLD Heavy Flavour Group, SLD Collaboration, LEP Electroweak Working Group, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, "Precision electroweak measurements on the Z resonance", *Phys. Rept.* **427** (2006) 257–454, arXiv:hep-ex/0509008. doi:10.1016/j.physrep.2005.12.006.

[12] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

[13] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021.

[14] ATLAS, CMS Collaboration, "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, arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.

[15] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. J.* **C75** (2015), no. 5, 212, arXiv:1412.8662. doi:10.1140/epjc/s10052-015-3351-7.

[16] O. S. Brüning et al., "LHC Design Report". CERN, Geneva, 2004. CERN-2004-003-V-1.

[17] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN". CERN, Geneva, 2008. CERN-DI-0812015.

[18] CMS Collaboration, "Technical proposal". CERN, Geneva, 1994. CERN-LHCC-94-38.

[19] ATLAS Collaboration, "ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN". CERN, Geneva, 1994. CERN-LHCC-94-43.

[20] LHCb Collaboration, "LHCb : Technical Proposal". CERN, Geneva, 1998. CERN-LHCC-98-004.

[21] ALICE Collaboration, "ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC". CERN, Geneva, 1995. CERN-LHCC-95-71.

[22] TOTEM Collaboration, "TOTEM, Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC: Technical Proposal". CERN, Geneva, 1999. CERN-LHCC-99-007.

[23] LHCf Collaboration, "Technical Proposal for the CERN LHCf Experiment: Measurement of Photons and Neutral Pions in the Very Forward Region of LHC". CERN, Geneva, 2005. CERN-LHCC-2005-032.

[24] MoEDAL Collaboration, "Technical Design Report of the MoEDAL Experiment". CERN, Geneva, 2009. CERN-LHCC-2009-006.

[25] CMS Collaboration, "Public CMS Luminosity Information". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2015. [Online; accessed 15th June 2016].

[26] The HL-LHC project, "LHC/HL-LHC Plan". http://hilumilhc.web.cern.ch/about/hl-lhc-project, 2015. [Online; accessed 15th June 2016].

[27] W. J. Stirling, "Parton luminosity and cross section plots". http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html, 2015. [Online; accessed 15th June 2016].

[28] S. Höche, "Introduction to parton-shower event generators", *TASI 2014: Journeys Through the Precision Frontier: Amplitudes for Colliders - Boulder, Colorado, USA* (2014) arXiv:1411.4085.

[29] The Compact Muon Solenoid Experiment, "CMS Detector Drawings". https://cmsinfo.web.cern.ch/cmsinfo/Media/Images/Detector/Detector%20Drawings/index.html, 2015. [Online; accessed 15th June 2016].

[30] R. S. Schmidt et al., "Performance of the Fast Beam Conditions Monitor BCM1F of CMS in the first running periods of LHC", *JINST* **6** (2011) C01004, arXiv:1012.3580. doi:10.1088/1748-0221/6/01/C01004.

[31] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[32] CMS Collaboration, "The CMS magnet project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. CERN-LHCC-97-010.

[33] CMS Collaboration, "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. CERN-LHCC-98-006.

[34] CMS Collaboration, "Commissioning and Performance of the CMS Silicon Strip Tracker with Cosmic Ray Muons", *JINST* **5** (2010) T03008, arXiv:0911.4996. doi:10.1088/1748-0221/5/03/T03008.

[35] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. CERN-LHCC-97-033.

[36] P. Adzic et al., "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST* **2** (2007) P04004. doi:10.1088/1748-0221/2/04/P04004.

[37] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, arXiv:1306.2016. doi:10.1088/1748-0221/8/09/P09009.

[38] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. CERN-LHCC-97-031.

[39] S. Abdullin et al., "The CMS barrel calorimeter response to particle beams from 2 GeV/c to 350 GeV/c", *Eur. Phys. J.* **C60** (2009) 359–373. doi:10.1140/epjc/s10052-009-0959-5.

[40] CMS Collaboration, "Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data", *JINST* **5** (2010) T03012, arXiv:0911.4991. doi:10.1088/1748-0221/5/03/T03012.

[41] CMS Collaboration, "The CMS muon project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997. CERN-LHCC-97-032.

[42] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006. CERN-LHCC-2006-001.

[43] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *JINST* **8** (2013) P11002, arXiv:1306.6905. doi:10.1088/1748-0221/8/11/P11002.

[44] CMS Collaboration, "CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems". Technical Design Report CMS. CERN, Geneva, 2000. CERN-LHCC-2000-038.

[45] CMS Collaboration, "CMS TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger". Technical Design Report CMS. CERN, Geneva, 2002. CERN-LHCC-2002-026.

[46] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$", *CMS Physics Analysis Summary* (2009). CMS-PAS-PFT-09-001.

[47] CMS Collaboration, "Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 TeV", *CMS Physics Analysis Summary* (2010). CMS-PAS-PFT-10-002.

[48] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with leptons from J/Ψ and W decays at 7 TeV", *CMS Physics Analysis Summary* (2010). CMS-PAS-PFT-10-003.

[49] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, arXiv:1502.02701. doi:10.1088/1748-0221/10/06/P06005.

[50] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, arXiv:1405.6569. doi:10.1088/1748-0221/9/10/P10009.

[51] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 08, P08010, arXiv:1502.02702. doi:10.1088/1748-0221/10/08/P08010.

[52] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, arXiv:1206.4071. doi:10.1088/1748-0221/7/10/P10002.

[53] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k_{T} jet clustering algorithm", *JHEP* **0804** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063.

[54] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011), no. 11, P11002, arXiv:1107.4277. doi:10.1088/1748-0221/6/11/P11002.

[55] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett.* **B659** (2007) 119–126, arXiv:0707.1378. doi:10.1016/j.physletb.2007.09.077.

[56] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2012), no. 04, P04013, arXiv:1211.4462. doi:10.1088/1748-0221/8/04/P04013.

[57] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, arXiv:1106.5048. doi:10.1088/1748-0221/6/09/P09001.

[58] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, arXiv:1411.0511. doi:10.1088/1748-0221/10/02/P02006.

[59] CMS Collaboration, "Multivariate Determination of the Missing Energy in the Transverse Plane (E_T^{miss}) at $\sqrt{s} = 13\,\text{TeV}$", *CMS Detector Performance Summary* (2015). CMS-DP-2015-042.

[60] F. Hartmann, "Evolution of Silicon Sensor Technology in Particle Physics", volume 231 of *Springer Tracts in Modern Physics*. Springer-Verlag Berlin Heidelberg, 2009.

[61] G. F. Knoll, "Radiation detection and measurement; 4th ed.". Wiley New York, 2010.

[62] L. Rossi, P. Fischer, T. Rohe, and N. Wermes, "Pixel Detectors. From Fundamentals to Applications". Springer-Verlag Berlin Heidelberg, 2006.

[63] M. Bigas, E. Cabruja, and M. Lozano, "Bonding techniques for hybrid active pixel sensors (HAPS)", *Nucl. Instr. Meth. Phys. Res. A* **574** (2007), no. 2, 392–400. doi:10.1016/j.nima.2007.01.176.

[64] C. Broennimann, F. Glaus, J. Gobrecht et al., "Development of an Indium bump bond process for silicon pixel detectors at PSI", *Nucl. Instr. Meth. Phys. Res. A* **565** (2006), no. 1, 303–308. doi:10.1016/j.nima.2006.05.011.

[65] E. Zakel, L. Titerle, T. Oppert and R. G. Blankenhorn, "High Speed Laser Solder Jetting Technology for Optoelectronics and MEMS Packaging", *Proceedings of the International Conference on Electronics Packaging - Tokyo, Japan* (2002).

[66] S. Kudella, "Gold-stud bump bonding interconnection technology for the CMS experiment". Master Thesis. Karlsruhe Institute of Technology, 2014. IEKP-KA/2014-11.

[67] M. Caselle, T. Blank, F. Colombo, A. Dierlamm, U. Husemann, S. Kudella, and M. Weber, "Low-cost bump-bonding processes for high energy physics pixel detectors", *JINST* **11** (2016), no. 01, C01050. doi:10.1088/1748-0221/11/01/C01050.

[68] University of California - Davis, Facility for Interconnect Technology, "General bonding procedure". http://fit.physics.ucdavis.edu/doku.php?id=shared:finetech, 2015. [Online; accessed 15th June 2016].

[69] W. Lin and Y. C. Lee, "Study of fluxless soldering using formic acid vapor", *IEEE Trans. Adv. Packaging* **22** (1999), no. 4, 592–601. doi:10.1109/6040.803451.

[70] CMS Collaboration, "Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons", *JINST* **5** (2010) T03007, arXiv:0911.5434. doi:10.1088/1748-0221/5/03/T03007.

[71] Y. Allkofer, C. Amsler, D. Bortoletto et al., "Design and performance of the silicon sensors for the CMS barrel pixel detector", *Nucl. Instr. Meth. Phys. Res. A* **584** (2008), no. 1, 25–41. doi:10.1016/j.nima.2007.08.151.

[72] H. C. Kästli, M. Barbero, W. Erdmann et al., "Design and performance of the CMS pixel detector readout chip", *Nucl. Instr. Meth. Phys. Res. A* **565** (2006) 188–194. doi:10.1016/j.nima.2006.05.038.

[73] H. C. Kästli, W. Bertl, W. Erdmann et al., "CMS barrel pixel detector overview", *Nucl. Instr. Meth. Phys. Res. A* **582** (2007), no. 3, 724–727. doi:10.1016/j.nima.2007.07.058.

[74] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade". Technical Design Report CMS. CERN, Geneva, 2012. CERN-LHCC-2012-016.

[75] H. C. Kästli, "Frontend electronics development for the CMS pixel detector upgrade", *Nucl. Instr. Meth. Phys. Res. A* **731** (2013) 88–91. doi:10.1016/j.nima.2013.05.056.

[76] L. Feld, M. Fleck, M. Friedrichs et al., "DC–DC powering for the CMS pixel upgrade", *Nucl. Instr. Meth. Phys. Res. A* **732** (2013) 493–496. doi:10.1016/j.nima.2013.06.053.

[77] A. Huffman, R. LaBennett, S. Bonafede and C. Statler, "Fine-pitch wafer bumping and assembly for high density detector systems", *IEEE Nucl. Sci. Conf. R.* **5** (2003) 3522–3526. doi:10.1109/NSSMIC.2003.1352670.

[78] Finetech GmbH & Co. KG, "FINEPLACER® femto Automated sub-micron Die Bonder", 2015. Datasheet [Online; accessed 15th June 2016].

[79] ATV Technologie GmbH, "SRO-700 Vacuum Reflow Oven", 2015. Datasheet [Online; accessed 15th June 2016].

[80] Department of Defense - United States of America, "Test Method Standard, Microcircuits (w/Change 5)", 2015. MIL-STD-883.

[81] Nordson Corporation, "DAGE 4000 Multipurpose Bondtester", 2015. Datasheet [Online; accessed 15th June 2016].

[82] G. C. Nelson and J. A. Borders, "Surface composition of a tin-lead alloy", *J. Vac. Sci. Technol.* **20** (1982), no. 4, 939–942. doi:10.1116/1.571647.

[83] L. Goenka and A. Achari, "Void formation in flip-chip solder bumps", *IEEE/CPMT Int. Electr. Manuf. Technol. Symp.* (1996) 430–437. doi:10.1109/IEMT.1996.559783.

[84] Z. Tang and F. Shi, "Effects of preexisting voids on electromigration failure of flip-chip solder bumps", *Microelectr. J.* **32** (2001), no. 7, 605–613. doi:10.1016/S0026-2692(01)00033-7.

[85] J. Pan, B. J. Toleno, T. Chou and J. D. Wesley, "The effect of reflow profile on SnPb and SnAgCu solder joint shear strength", *Solder. Surf. Mt. Tech.* **18** (2006), no. 4, 48–56. doi:10.1108/09540910610717901.

[86] B. Hiti, "Probe station for functionality tests of bare modules for the Phase I upgrade of the CMS Pixel Detector". Master Thesis. Karlsruhe Institute of Technology, 2015. IEKP-KA/2015-13.

[87] National Instruments, "LabWindows^{TM}/CVI Software". http://www.ni.com/lwcvi/, 2015. [Online; accessed 15th June 2016].

[88] National Instruments, "NI Vision Software". http://www.ni.com/vision/, 2015. [Online; accessed 15th June 2016].

[89] CMS Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *JHEP* **05** (2014) 104, arXiv:1401.5041. doi:10.1007/JHEP05(2014)104.

[90] T. Müller, "Analysis of Standard Model Higgs Boson Decays to Tau Pairs with the CMS Detector at the LHC". PhD Thesis. Karlsruhe Institute of Technology, 2015. IEKP-KA/2015-24.

[91] CMS Collaboration, "Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 231801, arXiv:1104.1619. doi:10.1103/PhysRevLett.106.231801.

[92] CMS Collaboration, "Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions", *JHEP* **10** (2014) 160, arXiv:1408.3316. doi:10.1007/JHEP10(2014)160.

[93] CMS Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with h $\to \tau\tau$", arXiv:1510.01181. Accepted for publication in Phys. Lett. B.

[94] CMS Collaboration, "Search for a very light NMSSM Higgs boson produced in decays of the 125 GeV scalar boson and decaying into τ leptons in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **01** (2016) 079, arXiv:1510.06534. doi:10.1007/JHEP01(2016)079.

[95] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS", *JINST* **11** (2016), no. 01, P01019, arXiv:1510.07488. doi:10.1088/1748-0221/11/01/P01019.

[96] A. Hoecker, P. Speckmayer, J. Stelzer et al., "TMVA - Toolkit for Multivariate Data Analysis", arXiv:physics/0703039.

[97] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (2005) N9–N20, arXiv:physics/0306087. doi:10.1088/0954-3899/31/9/N01.

[98] S. D. Drell and T.-M. Yan, "Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies", *Phys. Rev. Lett.* **25** (1970) 316–320. doi:10.1103/PhysRevLett.25.316.

[99] J. Alwall, M. Herquet, F. Maltoni et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, arXiv:1106.0522. doi:10.1007/JHEP06(2011)128.

[100] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, arXiv:0710.3820. doi:10.1016/j.cpc.2008.01.036.

[101] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", arXiv:1512.00815.

[102] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, arXiv:hep-ph/0409146. doi:10.1088/1126-6708/2004/11/040.

[103] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, arXiv:0709.2092. doi:10.1088/1126-6708/2007/11/070.

[104] 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, arXiv:1002.2581. doi:10.1007/JHEP06(2010)043.

[105] J. Alwall, R. Frederix, S. Frixione et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, arXiv:1405.0301. doi:10.1007/JHEP07(2014)079.

[106] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, arXiv:1209.6215. doi:10.1007/JHEP12(2012)061.

[107] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.

[108] LHC Higgs Combination Group, ATLAS Collaboration, CMS Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", *Technical Report* (2011). CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11.

[109] CMS Collaboration, "Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B710** (2012) 26–48, arXiv:1202.1488. doi:10.1016/j.physletb.2012.02.064.

[110] 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* (2011) arXiv:1103.0354.

[111] CMS Collaboration, "CMS Luminosity Measurement for the 2015 Data Taking Period", *CMS Physics Analysis Summary* (2016). CMS-PAS-LUM-15-001.

[112] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219–228. doi:10.1016/0010-4655(93)90005-W.

[113] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", *CMS Physics Analysis Summary* (2016). CMS-PAS-BTV-15-001.

[114] CMS Collaboration, "Measurement of the Inclusive Z Cross Section via Decays to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **08** (2011) 117, arXiv:1104.1617. doi:10.1007/JHEP08(2011)117.

[115] CMS Collaboration, "Measurement of differential top-quark pair production cross sections in the lepton+jets channel in pp collisions at 8 TeV", *CMS Physics Analysis Summary* (2013). CMS-PAS-TOP-12-027.

[116] CMS Collaboration, "Measurement of the differential ttbar cross section in the dilepton channel at 8 TeV", *CMS Physics Analysis Summary* (2013). CMS-PAS-TOP-12-028.

[117] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev.* **D86** (2012) 094034, arXiv:1208.5967. doi:10.1103/PhysRevD.86.094034.

[118] CMS Collaboration, "Measurements of inclusive and differential Z boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV", *CMS Physics Analysis Summary* (2016). CMS-PAS-SMP-15-011.

[119] CMS Collaboration, "Performance of reconstruction and identification of τ leptons in their decays to hadrons and ν_τ in LHC Run-2", *CMS Physics Analysis Summary* (2016). CMS-PAS-TAU-16-002 (to be published).

Acknowledgements
Danksagung
Ringraziamenti

There are many people who deserve to be thanked for the role they had in these years. First of all, many thanks to both my supervisors, Prof. Dr. Günter Quast and Prof. Dr. Ulrich Husemann, for giving me the opportunity to work in their research groups, for the careful revision of this manuscript and for the helpful comments which have been fundamental in giving the final shape to this thesis. Thanks to Dr. Michele Caselle, for sharing his expertise in microelectronics packaging and pixel detectors. Thanks to Dr. Roger Wolf, for the very detailed comments and suggestions on the last analysis chapter and for all the physics discussions that we had during the last months. Thanks to Dr. Andrew Gilbert, for unveiling some of the mysteries of combine, and to Dr. Joram Berger, for introducing me to Kappa and for guiding me in my first code developments there.

Many thanks to Simon and Benedikt, for the great times in the clean room and for their music taste. Thanks to all the people who came and went from the office in these years and in particular to Thomas, for all the interesting discussions on analysis and coding, to Dominik, Georg, Max, Raphael and Renè.

Thanks to all the great people and friends that I met in Karlsruhe and in particular to Silvia, Lorenzo, Valerio and Matteo. Double amount of thanks to Silvia and Lorenzo, who also took the time to review some of the chapters of this thesis.

Grazie ai miei genitori, a mia sorella e a tutta la famiglia per avermi supportato in tutte le decisioni, compresa quella di fare un dottorato all'estero.

Most of all, and above all, many thanks to Katharina for the chocolate, for turning strange sounds into a familiar language, and for being always there.

Karlsruhe,
June 2016

Erklärung der selbständigen Anfertigung meiner Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertationsschrift mit dem Titel

Packaging and assembly technologies for the pixel detector upgrade and measurement of $\tau\tau$ final states with the CMS experiment at the LHC

selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Fabio Colombo

Karlsruhe, den 15. Juni 2016