

DISS. ETH NO. 25436

INSTITUTE FOR PARTICLE PHYSICS AND ASTROPHYSICS
ETH ZÜRICH



STUDY OF THE DIFFERENTIAL HIGGS BOSON PRODUCTION IN THE
DIPHOTON DECAY CHANNEL WITH CMS AND CHARACTERISATION AND
CALIBRATION OF DETECTOR MODULES FOR THE PHASE₁ UPGRADE OF
THE PIXEL DETECTOR

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
Vittorio Raoul Tavolaro
Laurea Magistrale in Fisica
Sapienza Università di Roma
born on 02.04.1989
citizen of Gambarogno (TI)

accepted on the recommendation of
Prof. Rainer Steffen Wallny, examiner
Prof. Günther Dissertori, co-examiner
Dr. Federico Ferri, co-examiner

ABSTRACT

The data collected by the CMS experiment at CERN in proton-proton collisions during 2016 are analyzed to characterize the Higgs boson in its decay channel to two photons. The signal strengths relative to the standard model cross section are measured inclusively and for individual production modes, employing an event categorization optimized for the best precision. Production cross sections are then measured in a fiducial phase space, differentially with respect to the Higgs boson production and decay kinematic properties. The cross section is also measured inclusively in the fiducial phase space, as well as in sub-regions enriched in events originating from specific production modes ($t\bar{t}H$, VH). For these measurements, an approach aiming at minimizing the model dependence of the measurements is adopted. The measurements are found to be in agreement with the standard model predictions, within experimental and theoretical uncertainties.

A new upgraded pixel detector has been installed in the Compact Muon Solenoid (CMS) experiment in early 2017. The upgraded detector, while introducing several improvements with respect to the previous system, allows highly-efficient data taking at the increasing instantaneous luminosities of the LHC. The procedures designed to thoroughly test and calibrate the 1200+ detector modules are presented in this thesis, together with the quality criteria defined to identify modules suitable for installation in the final system. The results of the module production are reviewed, with a focus on modules employed for the construction of the second layer of the central part of the detector.

RIASSUNTO

I dati raccolti dall'esperimento CMS presso l'acceleratore LHC in collisioni protone-protone nel 2016 sono analizzati al fine di produrre una caratterizzazione del bosone di Higgs nel canale di decadimento in due fotoni. L'intensità del segnale rispetto alla sezione d'urto prevista dal modello standard è misurata inclusivamente e per i diversi meccanismi di produzione, utilizzando una categorizzazione degli eventi ottimizzata per ottenere la migliore precisione. Inoltre, le sezioni d'urto di produzione sono misurate in uno spazio delle fasi fiduciale. Tali misure sono differenziali nelle variabili cinematiche della produzione e del decadimento del bosone di Higgs. La sezione d'urto è misurata sia inclusivamente nello spazio delle fasi fiduciale, sia in sue sotto-regioni popolate in prevalenza da eventi originati attraverso specifici meccanismi di produzione ($t\bar{t}H$, VH). Per queste misure, è stato adottato un approccio inteso a minimizzare la dipendenza dal modello teorico. Le misure sono in accordo con le previsioni date dal modello standard, entro le incertezze sperimentalistiche.

Un nuovo rivelatore a pixel di silicio è stato installato nell'esperimento CMS all'inizio del 2017. Il nuovo rivelatore permette di mantere un'alta efficienza nella presa dati alle crescenti luminosità istantanee dell'LHC e, al tempo stesso, introduce numerosi miglioramenti rispetto al precedente sistema. Le procedure per una completa caratterizzazione e una precisa calibrazione degli oltre 1200 moduli del rivelatore sono descritte in questa tesi, assieme ai criteri di qualità definiti per identificare i moduli per l'installazione nel rivelatore. I risultati della produzione dei moduli sono discussi, concentrandosi sui moduli utilizzati per la costruzione del secondo strato della parte centrale del rivelatore.

CONTENTS

1	INTRODUCTION	1
I THE CMS EXPERIMENT AT THE LHC		
2	LHC AND THE CMS DETECTOR	5
2.1	The Large Hadron Collider	5
2.2	The CMS Experiment	7
2.2.1	The tracking system	8
2.2.2	Electromagnetic Calorimeter	9
2.2.3	Hadronic Calorimeter	10
2.2.4	The Muon system	10
2.2.5	Trigger system and computing	11
3	PARTICLE RECONSTRUCTION AT CMS	13
3.1	Particle Flow	13
3.2	Muons	14
3.3	Photons	14
3.4	Electrons	15
3.4.1	Photon energy	15
3.5	Jets	16
3.6	Jets from b quarks	17
3.7	Missing transverse momentum	18
3.8	The Tag and Probe method	19
II THE DECAY OF THE HIGGS BOSON INTO TWO PHOTONS		
4	THE THEORETICAL FRAMEWORK	23
4.1	Selected aspects of the Standard Model of Particle Physics	23
4.2	The Higgs sector	25
4.3	Higgs boson production at the LHC	27
4.3.1	The diphoton decay channel	28
4.4	Event generation and simulation	29
5	PROPERTIES OF THE HIGGS BOSON IN THE DIPHOTON DECAY CHANNEL	33
5.1	Introduction	33
5.2	Analysis strategy	35
5.3	Trigger selection	35
5.4	Vertex identification BDT	36
5.5	Vertex probability BDT	36
5.6	Photon preselection	37
5.7	Photon identification BDT	38
5.8	Diphoton classifier BDT	39
5.8.1	Validation with $Z \rightarrow ee$	46
5.9	Event categories in diphoton BDT output	48
5.10	Event categories for individual production modes	51
5.10.1	tH categories	51
5.10.2	VH categories	52

5.10.3	VBF categories	53
5.10.4	Categorization order	53
5.11	Signal model	54
5.12	Background model	54
5.13	Systematic uncertainties	55
5.14	Results	57
III FIDUCIAL AND DIFFERENTIAL MEASUREMENTS OF THE HIGGS BOSON PRODUCTION AT THE LHC		
6	INTRODUCTION	69
6.1	Introduction	69
6.2	Mass resolution estimator	70
6.3	Event selection	73
6.4	Event categorization	74
6.5	Phase space definitions	75
6.6	Observables	76
6.7	Statistical analysis	77
6.8	Systematic uncertainties	81
6.9	Results	81
IV QUALIFICATION OF DETECTOR MODULES FOR THE PHASE 1 UPGRADE OF THE CMS PIXEL DETECTOR		
7	SILICON TRACKING DETECTORS	95
7.1	Tracking detectors	95
7.2	Interaction of charged particles with silicon	95
7.3	The p-n junction and signal collection	96
7.4	Radiation-induced effects	98
7.5	Pixel spatial resolution	98
7.6	Hybrid pixel concept	99
8	CMS PHASE 1 PIXEL UPGRADE	101
8.1	Motivation	101
8.2	Design of the new Pixel detector	101
8.3	The digital readout chip	103
8.3.1	The PUC and the ROC periphery	103
8.4	Readout format	105
8.5	Detector modules	107
9	MODULE PRODUCTION	109
9.1	Module fulltest	109
9.1.1	Pretest	109
9.1.2	PixelAlive	110
9.1.3	Noise via S-curve	110
9.1.4	BumpBonding	111
9.1.5	Trimming	111
9.1.6	Pulse-Height optimization	112
9.1.7	GainPedestal	115
9.1.8	Readback mechanism and calibration	116
9.1.9	IV curve	118

9.2	Module qualification	119
9.2.1	Qualification setup	119
9.2.2	Configuration	119
9.2.3	Grading scheme	122
9.2.4	X-ray tests and grading	124
9.3	Qualification results for module production	125
9.3.1	ROC functionality	125
9.3.2	Pixel defects	127
9.3.3	Sensor quality	133
9.3.4	Functionality under X-ray	133
9.3.5	Other failure modes	134
9.3.6	Summary of module failures	137
9.3.7	Potential impact on resolution of C-grade modules	137
10	SUMMARY	143
	BIBLIOGRAPHY	159

INTRODUCTION

The data collected by the Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS) experiments during the Run 1 of the Large Hadron Collider (LHC) allowed the discovery of a new particle with a mass of 125 GeV in 2012. The observations and measurements of its properties up to now confirm its compatibility with the Standard Model (SM) predictions for the Higgs boson. The establishment of this discovery represents a major advance in the comprehension of particle physics and in particular in the generation mechanism of particle mass in the SM, concluding a decades-long search for this particle after its theorization in 1964. During the Run 2 of the Large Hadron Collider (LHC), the characterization of the new particle continues and deepens, exploiting the data collected at the unprecedented center of mass energy of 13 TeV. In this thesis, after a description of the experimental setup, i.e. the LHC accelerator and the CMS detector, in Chapter 2 and of the event reconstruction and object identification in Chapter 3, the Higgs boson and its role in the SM are introduced in Chapter 4. Particular attention is given to the Higgs boson decay to two photons, whose final state can be reconstructed with high precision in CMS, that played a key-role in the discovery and offers a vast potential in the measurement of the Higgs boson properties. The analysis of the data collected in 2016 by the CMS experiment studying the $H \rightarrow \gamma\gamma$ channel is described in Chapter 5, presenting the most precise characterization to date of this decay mode in CMS.

Furthermore, the large data sets collected in Run 2 allow the study of the Higgs boson production differentially with respect to the production and decay kinematic properties with improved precision, probing extensively the properties of the Higgs boson in its decay to a diphoton pair. These measurements, presented in Chapter 6, employ a dedicated event categorization defined in a *fiducial* phase space. The fiducial approach, allowing reduced uncertainties related to full phase space extrapolation, assumes a great importance in the transition from the discovery of the Higgs boson to the era of precision measurements.

The innermost sub-system of CMS is a pixel silicon detector. A review of the main concepts in the use of semiconductors in tracking detectors is given in Chapter 7. The CMS pixel detector was replaced with an upgraded version of the system at the end of the 2016 data taking. The upgrade project, called *phase 1* upgrade, was foreseen in order to maintain and improve the performance in hit efficiency, tracking and b-tagging of the detector at instantaneous luminosities beyond the initial design parameters of the LHC, which have been surpassed since 2016. To achieve this, improved front-end electronics and readout have been developed. The new system also introduces several improvements, among which 4-hit coverage in the full tracking volume, compared to the 3-hit coverage of the previous system. The motivations and the features of the phase 1 pixel upgrade detector are explored in Chapter 8. The construction of the new detector required the production of more than 1200 detector modules. Thorough qualification and calibration of the modules is essential to assure the quality and performance of the assembled detector: for this reason, in-depth testing procedures and stringent quality criteria have been studied and implemented. In Chapter 9, a review of these procedures is presented, along with the

results of the qualification of the modules produced for the second layer of the central part of the detector.

The thesis is completed by the summary of the main results presented in Chapter 10.

Part I

THE CMS EXPERIMENT AT THE LHC

This Part describes the Large Hadron Collider at the Conseil Européen pour la Recherche Nucléaire (CERN) and introduces one of its four experiments, the Compact Muon Solenoid detector. The reconstruction of different kinds of particles arising from the hadronic collision is discussed and the performance in the determination of their kinematic properties is examined.

LHC AND THE CMS DETECTOR

In this chapter, the LHC accelerator and the CMS detector are introduced. The various subsystems of the particle detector are described and, finally, an illustration of the trigger system and of computing infrastructures is provided.

2.1 THE LARGE HADRON COLLIDER

The Large Hadron Collider is a circular accelerator designed to accelerate hadrons to energies in the TeV range. It is installed in the 26.7 km tunnel, located in the vicinity of Geneva, that previously hosted its predecessor, the CERN Large Electron Positron (LEP) collider.

In its main operation mode, the machine accelerates protons, stored in two beams circulating in opposite directions. A shorter period in the accelerator schedule is dedicated to the operations with lead ions, used to produce p-Pb or Pb-Pb collisions. The particles are prevented from tangentially departing from the quasi-circular trajectory thanks to the very strong magnetic field, generated by 1232 superconducting dipole magnets along the ring. The magnets employ niobium-titanium coils, brought to superconductive state by cooling them below 2 K with liquid helium. Higher order magnets with more complex geometries of the magnetic field are used to focus the particle beams and steer them into collision at the 4 designated interaction points, where an equal number of detectors are installed, named A Large Ion Collider Experiment (ALICE), ATLAS, CMS and Large Hadron Collider beauty (LHCb).

The LHC is the final stage of a chain of smaller accelerators, as shown in Fig.2.1, which act as injectors that gradually accumulate the hadrons and ramp up their energy. The protons are obtained by stripping electrons off hydrogen atoms. They are then accelerated by a linear accelerator, the LINAC2, up to 50 MeV, then sequentially injected into the Proton Synchrotron Booster (PSB) (1.4 GeV), the Proton Synchrotron (PS) (25 GeV) and the Super Proton Synchrotron (SPS) (450 GeV). The acceleration mechanism of the PS accelerator is also responsible for shaping the structure of the beam in bunches, containing each a nominal number of protons of 1.15×10^{11} , spaced in time by 25 ns, corresponding to a frequency of the collisions in the LHC of 40 MHz. The protons are finally injected into the LHC through two transfer lines. Each proton beam is brought to its final energy of 3.5 TeV (in the years 2010-2011), 4 TeV (2012) and 6.5 TeV (2015-present) by means of a system of 16 radio frequency cavities. Up to 2808 bunches per beam can be stored in the machine. A detailed description of the design and the operations of the LHC machine can be found in [42].

The high focusing of the beams and the elevated number of protons per bunch at LHC result in a number of proton-proton interactions per bunch crossing typically larger than one. This quantity is referred to as *pileup*. The positions of such collisions are Gaussianly distributed in the longitudinal direction around the nominal interaction point, defining the so-called *luminous region* of the interaction point.

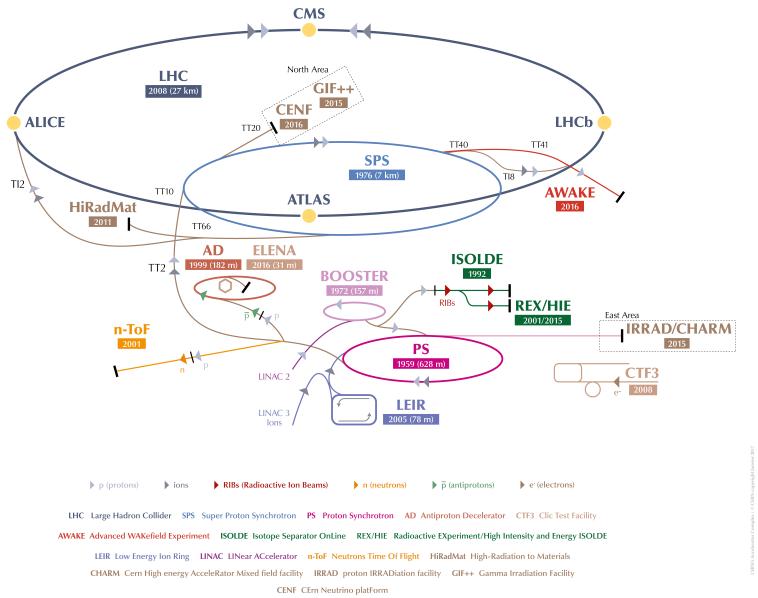


Figure 2.1: The CERN accelerator complex [77].

The number of particles delivered per unit of area and per unit of time by the machine is referred to as *instantaneous luminosity* and it is measured in units of $\text{cm}^{-2}\text{s}^{-1}$. In the hypothesis of two symmetric beams, as at the LHC, with equal number of bunches N_b and of particles per bunch N , colliding particles head-on with a revolution frequency f , and assuming Gaussian profiles for both dimensions of the transverse section of the beams with widths σ_x and σ_y , the instantaneous luminosity, \mathcal{L} , can be expressed as [59]:

$$\mathcal{L} = \frac{N^2 N_b f}{4\pi \sigma_x \sigma_y} \quad (2.1)$$

The integral over time of such quantity is referred to as *integrated luminosity* and it is measured in units of cm^{-2} or, by introducing the *barn* (b) unit ($1\text{b} = 10^{-24}\text{cm}^2$), in inverse femto barns (fb^{-1}).

The LHC operations so far are conventionally divided in two *Runs*. The so-called *Run 1* corresponds to the collisions produced at 7 and 8 TeV of energy in the center of mass between 2010 and 2012. The LHC delivered an integrated luminosity of 6.1 fb^{-1} at 7 TeV and 23.3 fb^{-1} at 8 TeV, with a peak instantaneous luminosity of $0.77 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [30].

After a first period of technical shutdown (Long Shutdown 1 (LS1)), the LHC resumed its operations in 2015 at 13 TeV of energy in the center of mass. The integrated luminosity delivered amounts to 4.2 fb^{-1} (2015), 40.8 fb^{-1} (2016) and 51.0 fb^{-1} (2017). This period of data taking, called *Run 2*, will continue until the end of 2018, when a second stop will take place (Long Shutdown 2 (LS2)). During Run 2, the LHC established a new world record in peak instantaneous luminosity for pp collisions, by reaching $2.06 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [99], corresponding to more than twice the design instantaneous luminosity. A summary of

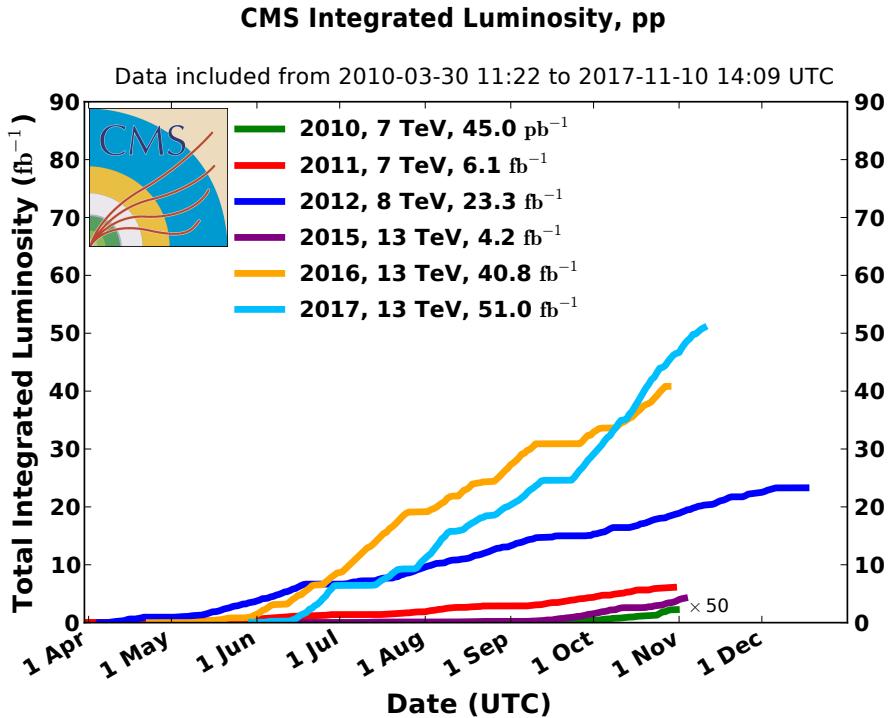


Figure 2.2: Integrated luminosity per year delivered to CMS in proton-proton collisions [31].

the performance of the LHC in terms of integrated luminosity per year in Run 1 and Run 2 is shown in Fig. 2.2.

2.2 THE CMS EXPERIMENT

The Compact Muon Solenoid detector, shown in Figure 2.3, is a general purpose detector placed at one of the interaction points of the LHC. It is designed to measure the properties of the particles produced in pp collisions with high precision and almost hermetic geometric coverage, in order to study a broad range of processes and pursue the search for new physics. It has a cylindrical shape, extending for 22 m in length and 15 m in diameter, and it is divided in a central region (*barrel*) and two forward regions (*end caps*). One of the key features of the CMS detector is its superconducting solenoid magnet, able to generate a magnetic field with a strength of 3.8 T. The inner bore of 6 m allows the positioning of the entire tracking and calorimetry systems inside of it. This configuration provides measurements of the energy and trajectories of particles unaltered by the energy loss and the multiple scattering in the material of the solenoid.

The CMS detector is composed by several sub-systems with cylindrical structure, arranged at different radii from the beam axis. The innermost component is the pixel and tracker system. Moving outwards, we encounter the Electromagnetic Calorimeter and the Hadronic Calorimeter. Outside the magnet bore and embedded in its steel return yoke, the muon system is placed, consisting of Drift Tubes (DT), Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC).

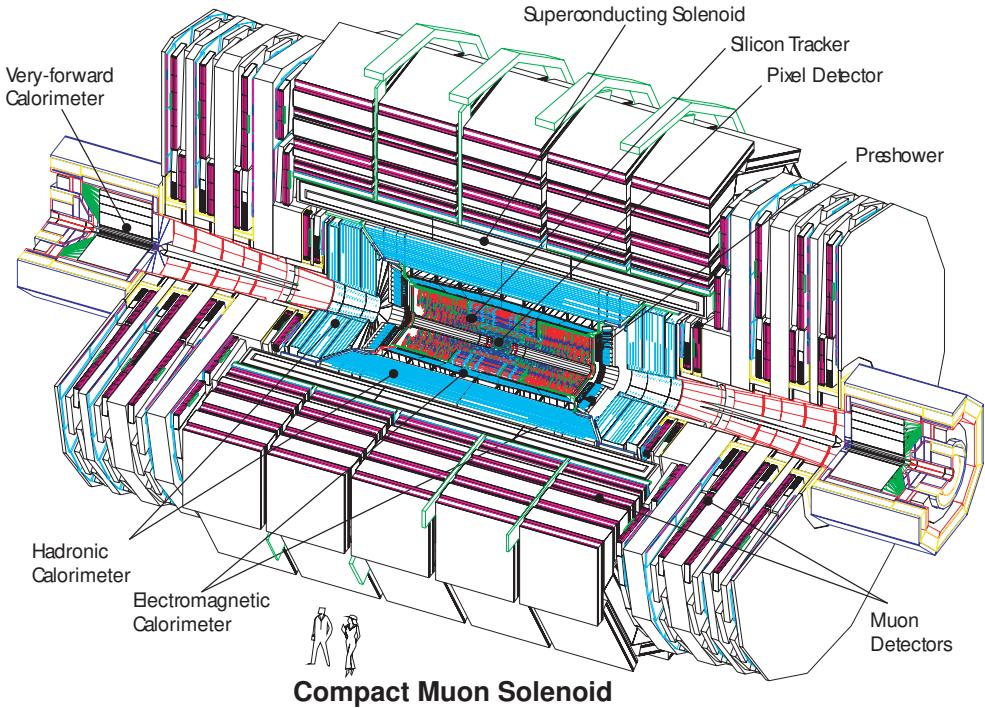


Figure 2.3: An exploded view of the CMS detector [29].

The coordinate system chosen to describe the geometry of the detector has its origin placed in the geometrical center of the detector. The x-axis points to the center of the LHC ring, the y-axis points upward, perpendicularly to the LHC plane, and the z-axis lies along the anticlockwise-beam direction, making the system right-handed. The azimuthal and polar angles (ϕ and θ) are measured in the x-y plane and from the positive z-axis, respectively. The *pseudorapidity* $|\eta|$, defined as $-\ln[\tan(\theta/2)]$, is often preferred to θ , given that differences in pseudorapidity are Lorentz-invariant for longitudinal boosts.

A complete description of the CMS detector and its components can be found in Ref. [18].

2.2.1 The tracking system

Two full-silicon sub-detectors form the tracking system of CMS, which aims at measuring the transverse momentum of charged particles from their reconstructed trajectory and at identifying primary collision vertices and secondary vertices from b quark and τ decays. Both sub-systems provide coverage up to $|\eta| = 2.5$ and ensure transverse momentum resolution of 1 – 2% for particles of 100 GeV up to $|\eta| = 1.6$ and better than 10% above $|\eta| = 1.6$.

2.2.1.1 The Pixel detector

The innermost part of the CMS tracking system is a silicon pixel detector, with about 66 million pixels of size $150 - 100 \mu\text{m}^2$ in the z and $r - \phi$ direction, respectively. The fine segmentation ensures adequate handling of the high flux (~ 1000) of particles per bunch crossing in the vicinity of the interaction point. The barrel region is composed

by three layers, with radii between 4.2 cm and 10.2 cm, while each endcap contains two disks, placed at 34.5 cm and 46.5 cm from the interaction point, providing coverage up to $|\eta| = 2.5$. The Pixel sub-system was replaced at beginning of 2017 by a new detector of similar conception and improved performance, with increased number of layers and disks and faster digital chip readout. Part iv of this thesis is dedicated to the description of the new system and its construction phase. The results presented in Parts ii and iii are obtained through the analysis of data collected by CMS in 2016 and so they rely on the configuration described here for the Pixel sub-system.

2.2.1.2 The Strip detector

At radii between 20 cm and 110 cm, several layers of silicon strips surround the Pixel detector. The strip pitch varies with radius, going from $80 \mu\text{m}$ in the first layer to $184 \mu\text{m}$ in the end caps. The strip geometry provides one-dimensional resolution information, and two-dimensional resolution is restored by arranging sensors with rotated orientation in adjacent layers.

2.2.2 Electromagnetic Calorimeter

The goal of the Electromagnetic Calorimeter (ECAL) is to precisely measure the energy of electrons and photons. To do so, it employs about 76000 lead tungstate crystals to compose a homogeneous calorimeter. The choice of lead tungstate was driven by its small Molière radius (2.2 cm) and its very short radiation length (0.89 cm), which enable a compact configuration that can be accommodated inside the magnet volume. Furthermore, lead tungstate provides very fast scintillation that allows the emission of more than 80% of the light before the subsequent bunch crossing. The ECAL Barrel (EB) region has an inner radius of 129 cm and an η coverage that extends to 1.479. In this region the crystals have an inner face with a size of $22 \times 22 \text{ mm}^2$ and a length of 23 cm. The ECAL Endcap (EE) extends the coverage up to $|\eta| = 3.0$ and contains crystals with a front face of $28.6 \times 28.6 \text{ mm}^2$ and a length of 22 cm. The low-yield scintillation light produced by the crystals is detected and internally amplified by Avalanche Photo Diode (APD) in the EB and Vacuum Photo-Triode (VPT) in the EE, to deal with different levels of radiation and configurations of the magnetic field.

A sampling calorimeter, composed by two planes of silicon micro-strip detectors and two lead absorbers, called *pre-shower*, is placed in front of the EE, to improve neutral pion identification. The layout of the ECAL system is show in Figure 2.4.

The ECAL relative energy resolution, σ_E/E , is estimated to be, from test beam performance [98]:

$$\sigma_E/E = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E} \quad (2.2)$$

where $a = 2.8\%$ (stochastic term) accounts for statistical fluctuations in the light efficiency and quantum efficiency, $c = 12\%$ (noise term) represents electronic noise and $b = 0.30\%$ (constant term) is the effect of several operational conditions, such as temperature and high-voltage stability, radiation levels, and the amount of material upstream the ECAL. This estimate corresponds to an energy resolution of 1% for photons and electrons with energy larger than 10 GeV.

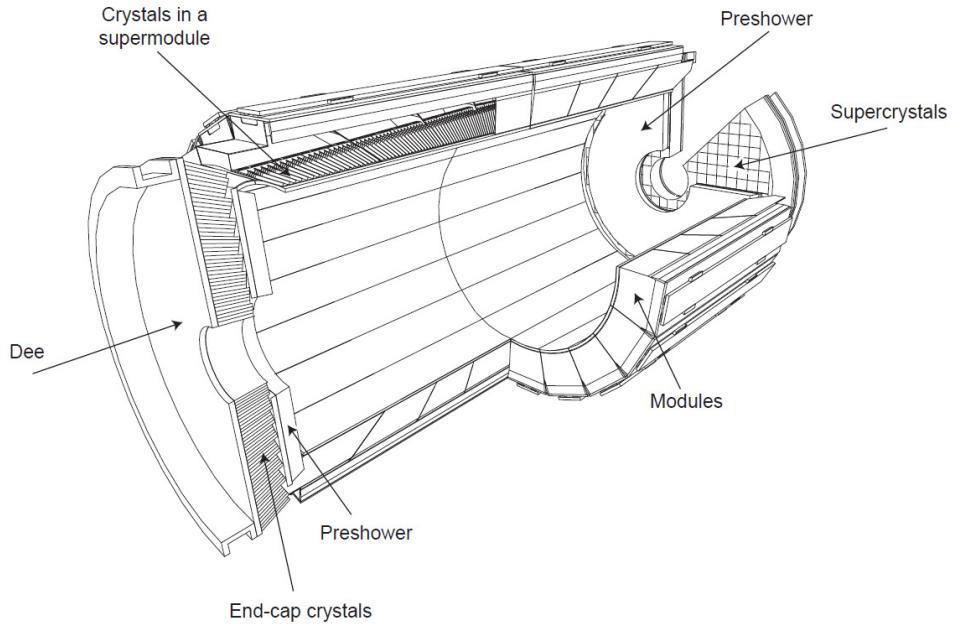


Figure 2.4: Layout of the CMS ECAL [19].

2.2.3 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is placed between the ECAL and the magnet solenoid. Its scope is the reconstruction of jets and the measurement of their energy. It adopts a sampling configuration, with 4 mm thick tiles of plastic scintillators interleaved with planes of brass absorbers. Optical fibers, Hybrid Photo Diode (HPD) and Silicon PhotoMultiplier (SiPM) are used to read out the scintillation light. The central part is composed by a HCAL Barrel (HB), covering the region $|\eta| < 1.4$, complemented by the HCAL Outer (HO), that uses the solenoid as an additional absorber in the region $1.4 < |\eta| < 1.26$. The endcap part HCAL Endcap (HE) covers the area between $1.3 < |\eta| < 3.0$, with the HCAL Forward (HF) extending the coverage to the region $3.0 < |\eta| < 5.0$. The energy resolution is governed by the same form described in Equation 2.2. The stochastic term a has been determined in test beam measurements [40] to be 115%, while the constant term b is 5.5%. The noise term c is found to be negligible compared to the other contributions.

2.2.4 The Muon system

The Muon system employs three different types of gaseous detector in order to identify muons and precisely determine their momentum. In the barrel region ($|\eta| < 1.2$), DTs are used, while in the end caps ($|\eta| < 2.4$) CSCs are employed. Both DTs and CSCs provide excellent spatial resolution, crucial for the measurement of the muon momentum. RPCs are installed up to $|\eta| = 2.1$ and are used for triggering events with high-momentum muons, thanks to their fast response and good time resolution, but provide less precise position information compared to DTs and CSCs.

2.2.5 Trigger system and computing

The size of the events recorded by the CMS experiment (~ 1.5 MB) and the high bunch crossing rate of the LHC of 40 MHz lead to a flux of data too large to be written on storage supports. This rate is reduced thanks to a trigger system, which is responsible of deciding which events are interesting for further analysis and which, instead, can be discarded. The system is structured in two layers:

- the Level-1 (L_1) Trigger, which takes a decision based on rough information from the calorimeters on energy deposits (trigger towers) and from the muon system. Custom front-end electronics are employed to guarantee fast response and optimal decision time. The data are kept in buffers while the decision is pending, allowing for a latency time of $3.2\ \mu s$. The L_1 Trigger reduces the event rate from 40 MHz to about 100 kHz.
- the High Level Trigger (HLT) takes in input the data rate reduced from the L_1 trigger and further reduces the event rate down to about 1 kHz. This reduction is achieved through a software layer which runs on a computing cluster which uses commercial Central Processing Units (CPUs). The decision at this level is taken based on a basic reconstruction of the final-state objects and on more elaborate calculations compared to the L_1 trigger. The precision of the reconstruction is limited by the time available to complete the decision, which amounts to about 300 ms on average. Several HLT selection paths are run in parallel, exploring the potential interest for further analysis in an event according to a range of possible criteria. The selection is performed on different types of objects and with different thresholds depending on the event topology sought by the Physics analyses. The various HLT paths are subject to a global OR logic, which selects the event if at least the decision of one of the paths is positive.

The data selected by the HLT system are stored and distributed for further analysis on a multi-tier computing architecture. The tier-0 computing center is placed at CERN. It stores the raw data and performs a first event reconstruction. Data are then distributed to several tier-1 centers, where backup copies are stored and full event reconstruction is run. Tier-1s provide data with high availability for the later stages of the architecture, which involves also tier-2 and tier-3 computing centers. These stages perform further analysis and reduction of the data collected and are employed for the production of samples of simulated events, whose main elements are introduced in Section 4.4.

The reconstruction techniques and algorithms deployed for the reconstruction of particles at CMS are introduced in this Chapter. The reconstruction of photons and the determination of their energy is explored in more detail, given their pivotal role in the results presented in the later Chapters of this thesis.

3.1 PARTICLE FLOW

The Particle Flow (PF) algorithm is designed to identify each stable particle produced in the collision, namely electrons, photons, muons, charged and neutral hadrons. Rather than referring to an individual sub-detector for the identification of a specific kind of particle, it relies on the combination of information from all sub-detectors. This approach is particularly helpful in supplementing the charged hadron identification and energy measurement in the HCAL, limited by the rather coarse energy resolution (see Section 2.2.3), with information from the tracking system and the ECAL. The particle-level information obtained through the PF algorithm can then be combined to build more complex objects, such as jets and p_T^{miss} , whose definition is given in Section 3.7.

The PF algorithm can be briefly illustrated as follows, while a detailed description of the algorithm can be found in [87]. The first step resides in the combination of the tracker hits into tracks and of energy deposits in the calorimeters into energy clusters. The tracking procedure follows an iterative strategy, where hits are gradually excluded for the next steps as they are assigned to a reconstructed track. The tracks with high quality requirements are reconstructed first, yielding a lower probability to reconstruct unphysical tracks from the remaining hits, then the requirements gradually are relaxed in further steps of the procedure, allowing a high efficiency in track reconstruction. The calorimetric clustering is seeded by local maxima in the distribution of the deposited energy. Clusters are grown starting from such seeds by adding neighboring cells according to their distance from the seed, until their energy falls below a given threshold.

The PF algorithm then performs a linking between tracks, clusters and track segments in the muon system. The procedure exploits then the linked objects to identify iteratively the different particles. First, muons are identified if a track and clusters are compatible with a track in the muon system and if the momentum measured with only the tracker system is compatible with the momentum determined from the full track. The tracks and clusters are then associated to a muon and removed from the list of linked objects. Then, electrons are reconstructed by combining a track with the matched ECAL clusters, including those from bremsstrahlung photons, which are in turn removed from the list. After that, the remaining tracks of charged particles are associated to charged hadrons, and the p_T of their track is compared to the energy of their HCAL cluster: if the calorimetric energy is found in excess compared to the tracker p_T measurement, such surplus of energy is further identified as a neutral hadron. Finally, calorimeter clusters not associated to a track are identified as photons. Similarly to the previous step, the energy of the HCAL

cluster is compared to the one of the ECAL cluster and the possible excess is identified as a neutral hadron.

3.2 MUONS

Muon candidates can be identified using only information extracted from the tracker (*Tracker Muons*), or the muon spectrometer (*Standalone Muons*), or matching a Standalone Muon to a compatible track (*Global Muons*). Muons used in this analysis are required to be reconstructed as Global Muons and to be identified as muons by the PF algorithm. Furthermore, they have to satisfy a set of criteria, defined as *Tight Muon Selection* [25], which rejects muons from the in-flight decays and cosmic muons, while retaining a high efficiency in selecting prompt muons. The selection includes requirements on the χ^2 of the track fit, the number of muon segments in the muon chambers, the number of hits in the fitted tracker track, and on the transverse and longitudinal impact parameter of the muon track with respect to its vertex.

3.3 PHOTONS

The energy deposits in the ECAL are grouped into clusters and Super Clusters (SCs) to reconstruct photons, with a clustering algorithm that allows almost complete recovery of the energy also for photons that convert to an electron-positron pair in the material upstream of the ECAL. A detailed description of the algorithm can be found in Ref. [78].

The clustering step accounts for the distribution of the energy of a single electromagnetic shower across several crystals. A 3×3 (5×5) matrix of crystals, centered around the unconverted incident photon, retains around 94% (97%) of its energy. For such reason, energy deposits in the ECAL are grown into clusters from seed crystals by topological aggregation of neighboring cells with a deposited energy in excess of a given threshold. This threshold is set at about two standard deviations of the electronic noise in the ECAL, corresponding to 80 MeV in the barrel and, depending on pseudorapidity, up to 300 MeV in the end caps. Adjacent clusters can share the energy of the same crystal, assuming a Gaussian transverse profile of the electromagnetic shower.

The reconstruction, though, needs to account for the conversion of roughly half of the photons into electron-positron pairs, due to the interaction with the material upstream the ECAL. The action of the magnetic field separates the positron and electron trajectories in the ϕ direction, leading to energy deposits spread in the ϕ direction and well collimated in the η direction, creating a widening of the shower profile. The *Mustache* algorithm takes advantage of the η - ϕ correlation in the cluster topology to merge groups of clusters into SCs, with improved containment of the electromagnetic shower, spatial resolution and energy measurement. In the case of converted photons, the conversion tracks are reconstructed based on information from the ECAL and the tracker, connected to a common vertex and then matched to compatible ECAL clusters. An important variable in the discrimination of converted and unconverted photons is R_9 , defined as the ratio of the energy contained in the 3×3 matrix surrounding the SC seed and the total energy of the SC. Unconverted photons have high values of R_9 , while converted photons and nonprompt photons are assigned lower values.

Other important variables characterizing the cluster shape and used in the analyses presented in this thesis are $\sigma_{ij|\eta}$, which describes the second momentum of the distribution of the energy deposits in the η direction measured in terms of crystal cells, and S_4 , defined as the ratio between the maximum energy of the 2×2 matrices containing the highest-energy crystal and the energy of the 5×5 matrix centered around the seed of the SC.

3.4 ELECTRONS

The reconstruction of electrons relies on information both from the silicon tracker and the ECAL.

Electrons radiate a significant (33% for $\eta \sim 0$ and 86% for $\eta \sim 1.4$) fraction of their energy through the interaction with the material between the interaction point and the ECAL. Such bremsstrahlung radiation is spread in the ϕ direction, due to the bending of the electron track under the action of the magnetic field, while their dispersion in the η direction is negligible for electrons with p_T larger than 5 GeV. The same clustering algorithms described in Section 3.3 are employed to handle such profile in the $\eta\text{-}\phi$ plane of the electromagnetic shower. These algorithms indeed provide clustering and reconstruction without making an assumption on the nature of the candidate particle associated to energy deposits. This allows in particular the reconstruction of electrons from the abundant $Z \rightarrow e^+e^-$ production as photons and their use to measure reconstruction and identification efficiencies, as described in the section 3.8.

The bremsstrahlung radiation, together with the multiple scattering interaction of the electrons, has also the effect of inducing non-Gaussian tails in the probability of their propagation from one layer to another of the tracker. To deal with this situation, the gaussian-sum-filter (gsf) is used to perform the track reconstruction, describing the probability density function as a sum of Gaussian distributions.

Reconstructed electrons are then separated from photon conversions, jets with a large fraction of electromagnetic activity, and semi-leptonic decays of b and c quarks by means of a Multi-Variate Analysis (MVA) [39] which combines information on the shape of the ECAL clusters and their matching to the track. The *loose* selection working point of this MVA, corresponding to an efficiency in selecting electrons of 90%, is used to accept electrons that are used in the analyses presented in this thesis.

3.4.1 Photon energy

The energy of photons is computed from the sum of the raw energy of the clustered crystals, with the addition of the pre-shower energy in its region of coverage. This energy is corrected for changes in the response of the crystals over time [21]. In particular, the exposure to ionizing radiation induces a loss of transparency in the crystals due to the creation of defects (color centers) in the lattice and thus lower light yields. Thermal annealing, though, constantly reduces the number of such defects in the crystals, leading to partial recovery of the transparency. A dedicated laser system is set up to monitor the transparency of the crystals during run time. It injects light of a known wavelength in each crystal and provides a transparency correction, allowing a transparency precision of 0.1% in the EB [74]. The response of the ECAL is then equalized across crystals in the

η and ϕ coordinates thanks to *intercalibration* corrections derived with techniques based on the ϕ -symmetry of the energy deposits in the detector, or on the reconstruction of the invariant mass in $\pi^0/\eta^0 \rightarrow \gamma\gamma$ decays, or on the measurement of the momentum of isolated electrons from $Z \rightarrow e^+e^-$ and $W \rightarrow e\nu$ decays.

The photon energy is then further corrected to account for the containment of the electromagnetic shower in the calorimeter, as well as for the energy dispersion in the detector due to conversions and their interaction with the material. Such corrections are derived using a semi-parametric regression technique, trained on simulated events. The target of the regression is the probability distribution function of the ratio between the true simulated energy and the reconstructed one. The distribution is parametrized as a function with a Gaussian core and two power law tails, an extended form of the Crystal Ball function [47]. The regression is thus able to provide an estimator for the mean and the width of the Gaussian distribution for each photon, which are taken as the energy correction and the energy resolution estimator, σ_E/E , respectively. The input variables include cluster shape variables, the coordinates of the SC and local information on the clusters and seeds. In the EE, the ratio of the pre-shower energy to the raw SC energy is additionally included. To account for residual effects on the energy scale due to Pile Up (PU) interactions and Underlying Event (UE) (the remnants of the scattering interaction, involving processes with lower transfer of transverse momentum), the number of primary vertices and the median energy density in the event, ρ , are included as inputs.

The invariant mass peak in $Z \rightarrow e^+e^-$ events, where electrons are reconstructed as photons, is then used to correct residual discrepancies in the energy response between data and simulation. In data, the energy scale is corrected so as to align the position of the mass peak to the true mass value of the simulation. In simulation, a Gaussian smearing correction is added to the energy resolution estimator to reproduce the width of the peak observed in data. The corrections are derived as functions of time (scale only), pseudorapidity, and R_9 .

The energy resolution of the photons varies according to the region where the candidate is reconstructed, as shown in Figure 3.1: in the barrel section of the ECAL, unconverted photons or photons converting near the inner face of the ECAL have an energy resolution of about 1% for energies in the range of tens of GeV. Other photons reconstructed in the EB have a resolution of about 1.3% up to a pseudorapidity of $|\eta| = 1$. The energy resolution rises to about 2.5% at $|\eta| = 1.4$. In the EE, unconverted or late-converting photons have an energy resolution of about 2.5%, while the remaining photons reconstructed in EE have resolutions between 3 and 4% [67].

3.5 JETS

The result of the hadronization of partons in the detector is a collimated jet of particles. Individual particles of the jet are identified by the PF algorithm and can be clustered under the hypothesis that they belong to the same parton shower [10].

In order to reject spurious signals arising from noise patterns in the HCAL, a clustered jet is selected only if it meets a set of criteria, based on the total number of constituents, the number of neutral constituents, and on the fraction of energy carried by different type of PF candidates (neutral hadrons, charged hadrons, muons, electrons, and photons).

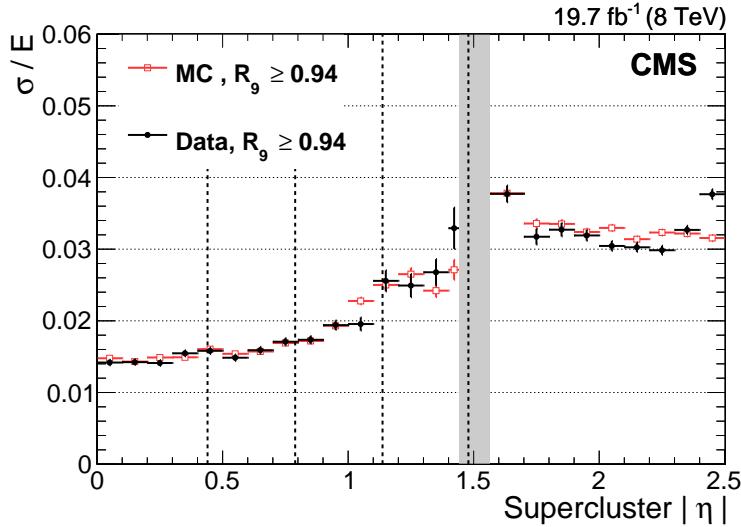


Figure 3.1: The relative photon energy resolution for photons with $R_9 > 0.94$ as a function of the pseudorapidity of the photon. The resolution is measured in $Z \rightarrow e^+e^-$ events in data (black points) and simulation (red squares). The module boundaries in the EB are indicated by the vertical dashed lines, while the grey area corresponds to the transition area between EB and EE [67].

The selection requirements vary according to the pseudorapidity of the jet, to account for different performance of the detector.

Furthermore, the contamination from jets arising from PU interactions is reduced by using a Boosted Decision Tree (BDT) [68] that combines information on the object multiplicity in the jet, the topology of the jet, and, in the region covered by the tracker, the compatibility of the jet tracks with the primary vertex.

The jet momentum is determined as the vectorial sum of all particle momenta in the jet. Residual contributions to the energy from PU interactions are corrected with an offset correction. Additional corrections to the jet energy are derived from simulation to bring the average measured response of jets to that of particle level jets. In situ measurements of the momentum balance in di-jet, photon+jet, Z +jet, and multijet events are used to account for any residual difference in jet energy scale in data and simulation. Additional selection criteria are applied to each jet to remove jets potentially dominated by anomalous contributions from various sub-detector components or reconstruction failures. After corrections, the measurement of jet momentum is found to agree with its true value in simulation within 5–10% over the whole p_T spectrum and detector acceptance. The jet energy resolution amounts typically to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV, to be compared to about 40%, 12%, and 5% obtained when the calorimeters alone are used for jet clustering.

3.6 JETS FROM B QUARKS

The longer lifetime of the b quarks enables the identification of jets arising from their hadronization among others. Such lifetime, in fact, allows the secondary vertex, corresponding to the decay of the b hadron, to be reconstructed inside the pixel detector and its distance with respect to the primary one to be measured. An MVA technique, the

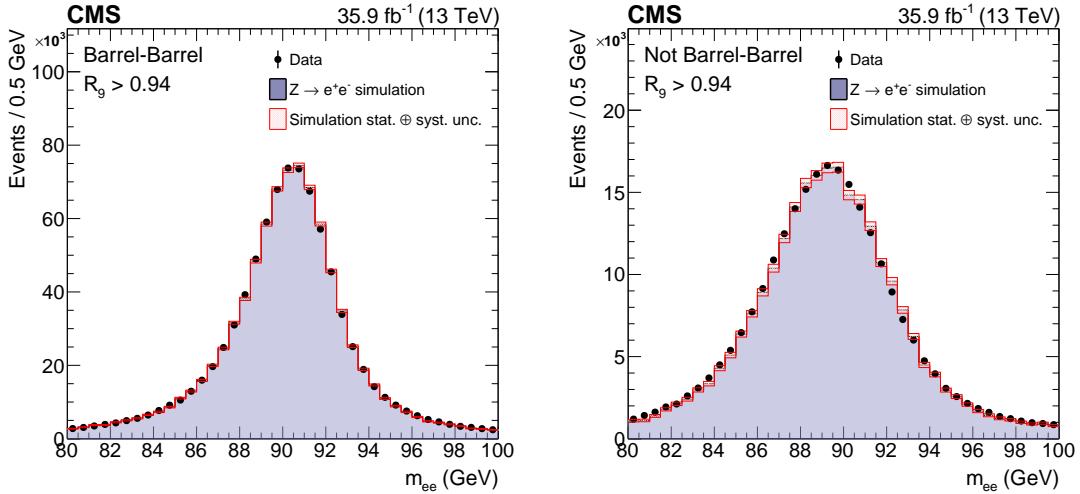


Figure 3.2: The invariant mass distributions for $Z \rightarrow e^+e^-$ is compared between data and simulation events, where electrons are reconstructed as photons. The distributions are shown in two particular bins where the corrections are derived, corresponding to events with both electrons in EB (left), and the remainder of the events (right), and with $R_9 > 0.94$. The distribution in simulation is normalized to the integral in data for $87 < m_{e^+e^-} < 93 \text{ GeV}$ [27].

Combined Secondary Vertex (CSV) b-tagging algorithm [22, 88, 89], combines information on the secondary vertex and assigns higher values to jets more likely to originate from a b quark. The performance of the tagging algorithm is evaluated in terms of efficiency in the identification of b-jets and misidentification probability of jets from light quarks as b-jets. Different values of the former (latter) correspond to three working points defined on the output of the discriminator: *loose*, 85% (10%); *medium*, 70% (1%); *tight*, 55% (0.1%). All three working points are used in distinct parts of the analyses presented in this thesis.

3.7 MISSING TRANSVERSE MOMENTUM

The energy balance in the transverse plane in the initial state of the collision is exploited to provide a measurement of the transverse momentum of particles that escape detection, as neutrinos or neutral stable Beyond the Standard Model (BSM) particles. This is achieved through the missing transverse momentum, \vec{p}_T^{miss} , defined as the negative sum of the transverse momenta of all PF candidates.

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{PF cand.}} p_T \quad (3.1)$$

The resolution in the measurement of \vec{p}_T^{miss} is limited by the impact of the so-called *instrumental p_T^{miss}* , which arises from imperfect measurement of the p_T of the PF candidates and from spurious detector signals. The corrections of the energy scale and resolution of jets are propagated to the measurement of \vec{p}_T^{miss} . Further details on the reconstruction and performance of missing transverse momentum can be found at [79]. In the analyses presented in this thesis, the magnitude of the vector \vec{p}_T^{miss} will be indicated with p_T^{miss} .

3.8 THE TAG AND PROBE METHOD

The Tag and Probe (T&P) technique allows the selection of a pure sample of objects by exploiting the decay of a resonance, with the goal of measuring selection efficiencies. The method is very general and separates the decay products of the resonance in a *tag* and a *probe* systems. The tag system is selected with stringent selection criteria which, via the constraint of the resonance mass, result in an enhancement of the purity of the samples of probe candidates, without any selection being directly applied to them. The invariant mass spectrum of the resonance enables, moreover, the fit and subtraction of the residual contamination from possible backgrounds. To extract the efficiency of a given selection, such fit is performed separately for events where the probe fails or passes the selection whose efficiency is being measured. The resonant yields can be extracted both in data and simulated events, such that the ratio between efficiencies in data and simulation, called also *efficiency scale factor*, can be computed. Due to the absence of abundant production of resonances decaying to photons with high mass, in the case of analyses using photons as the ones presented in Parts ii and iii, the T&P method for photon selection efficiencies is applied to $Z \rightarrow e^+e^-$ events, where the electrons are reconstructed as photons. One electron is required to pass the *tag-selection*, making the other the probe. This allows to measure the efficiencies of all the selections applied in the analyses described in this thesis, except for selection criteria that aim at discriminating electrons from photons (see Section 5.6). In such cases, the rarer production of $Z \rightarrow \mu^+\mu^-\gamma$ is used, where the $\mu\mu$ pair represents the tag system, allowing the selection of a sample of probe photons.

Part II

THE DECAY OF THE HIGGS BOSON INTO TWO PHOTONS

After introducing the theoretical set-up of modern Particle Physics, in particular the spontaneous symmetry breaking mechanism, this Part presents the study of the Higgs boson in its decay to two photons in pp collisions at $\sqrt{s} = 13$ TeV.

4

THE THEORETICAL FRAMEWORK

This chapter introduces the theory of particle physics, known as the Standard Model. It describes its main aspects, with a focus on the spontaneous symmetry breaking mechanism and the Higgs boson sector. The last Section delineates the techniques that exploit the theoretical comprehension to allow the numerical simulation of the processes studied at the LHC.

4.1 SELECTED ASPECTS OF THE STANDARD MODEL OF PARTICLE PHYSICS

The modern understanding of the physical phenomena relies on the description of the fundamental particles and of their interactions by means of Quantum Field Theory (QFT), with invariance under local gauge transformations and transformations of the Poincaré group [56]. The action of the operators associated to the quantum fields on the vacuum creates the corresponding particles, with properties defined by their quantum numbers. The choice of the fields defines which particles are considered elementary in the theory. The SM associates a field to each particle known as a fundamental block of matter and to the mediators of the interactions between them. The former are fermions and are grouped into leptons (electron, muon, tau, and the respective neutrinos) and quarks (up, down, charm, strange, top, bottom), while the latter are gauge bosons (photon, for the electromagnetic force, W and Z bosons for the weak force, and gluon for the strong force). Additional scalar fields are introduced in the theory to generate the masses of the gauge bosons and fermions through the spontaneous symmetry breaking mechanism. In its minimal version, the latter predicts an additional spin-0 particle, the Higgs boson (H).

The fundamental particles of the SM and their roles are summarized in Figure 4.1.

The electromagnetic and weak interactions are described by a renormalizable QFT invariant under transformations of the symmetry group $SU(2)_L \times U(1)_Y$, where the subscript L indicates the weak isospin, whose current couples only to left-handed fields, and Y the weak hypercharge. The unified description of the two forces formulated by Glashow [49], Salam [86], and Weinberg [102], based on the model of Yang and Mills [104], reflects the non-conservation of parity in weak interactions, observed experimentally for the first time in the decay of Cobalt-60 atoms [103], by using different representations for the left-handed and the right-handed leptons in the SM Lagrangian. In fact, while the left-handed leptons are represented as doublets of $SU(2)_L$, $(\nu_L \ l_L)$, the right-handed ones are singlets, (l_R) . This structure is replicated in three flavor families (electronic, muonic, and tauonic), as summarized in Figure 4.1.

Quarks are also subject to the weak force and are as well organized in three families of flavor, with the first family, $(u \ d)$, composing the ordinary matter, while the two additional families, $(c \ s)$ and $(t \ b)$, not occurring in stable matter, are produced at particle colliders and in cosmic radiation. Quarks have fractional electric charge, equal to $2/3$ for up-type quarks and to $-1/3$ for down-type ones. These electro-weak eigenstates of quarks do not coincide with the mass eigenstates: the two are connected by a unitary

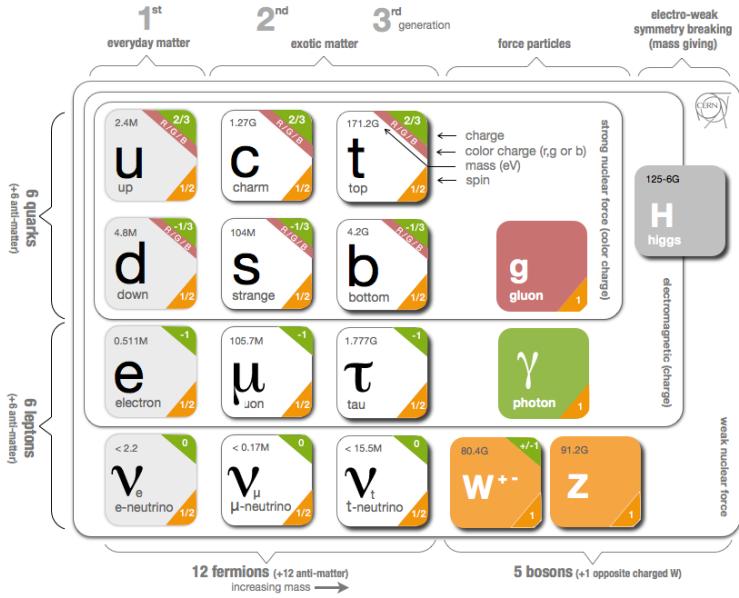


Figure 4.1: The elementary particles of the SM of Particle Physics. Quarks and leptons are organized in three *flavor* families, each containing an up-type and a down-type quark, and an electrically charged lepton and a neutrino. The mediators of the forces are gauge bosons: the diagrams indicate the particles subject to the interaction mediated by a given boson. Adapted from [80].

transformation, the Cabibbo-Kobayashi-Maskawa matrix, which expresses the mixing between the two sets of eigenstates. A more detailed description of the mixing mechanism is given in [15, 69].

The strong interaction between quarks is associated with an additional quantum number, the *color* charge, and its respective symmetry group, $SU(3)_C$. This sector of the SM, the Quantum Chromo Dynamics (QCD), describes the structure of the strong interaction, mediated between color charged particles by eight combinations of color and anti-color states, the gluons. The coupling constant for the strong force, α_S , can be introduced in analogy with the fine structure constant of electrodynamics. The study of the renormalization of the QCD theory reveals that the dependence of α_S on the momentum exchanged in the interaction, Q^2 , is such that α_S decreases for large Q^2 , contrarily to the case of Quantum Electro Dynamics (QED). This property corresponds to the *asymptotically free* behavior of quarks under the strong force: for high energy interactions, the coupling constant tends to zero, and perturbative approaches to the calculation of interaction amplitudes are possible; at low energies, instead, non-perturbative effects have to be taken into account. A second consequence of the running behavior of α_S is the *confinement* of the color charge, which can be qualitatively illustrated as follows: if we imagine of trying to isolate a coloured quark in, for example, a quark-antiquark bound state by means of excitation energy, as the distance between the particles increases, so does the intensity of the force, up to a point where it is more energetically favorable to create a new quark-antiquark pair in between the original quark and anti-quark, rather than further increasing the distance, leading to the formation of new bindings. The same mechanism is responsible for the *hadronization* process of high energy colored particles produced in

particle collisions, which evolve by gradually losing energy from a high-energy, perturbative regime to low-energy processes that produce color screening and the formation of composite states of quarks, named hadrons. The possible combinations of quarks for the formation of such composite particles are regulated at first order by the rules of symmetry of the $SU(3)$ group. In particular, baryons are composed by three quarks according to the Kronecker product of the fundamental representation of the group, nominally $3 \otimes 3 \otimes 3$, while mesons are obtained from the product of the fundamental and anti-fundamental representations, $3 \otimes \bar{3}$, corresponding to a quark-anti-quark structure. Further details on QCD, the scaling of α_S , and the structure of hadrons can be found in [56].

4.2 THE HIGGS SECTOR

The theory of Glashow, Salam, and Weinberg introduced earlier in this Section, while providing a successful description of the unified electroweak interaction, it shows a critical weakness by predicting the intermediate bosons to be massless. This aspect of the theory is in clear contrast with the physical experience of massive matter and, in particular, it breaks the interpretation of the electroweak sector in the low energy limit, where the mass of the bosons must be large enough such that these particles do not bring a visible contribution to the weak decays of neutrons and pions. The addition of *ad hoc* mass terms for the mediators to the Lagrangian of the theory is forbidden by the Gauge invariance of the theory and, even under the assumption of explicit breaking of the Gauge symmetry, similarly to the violation of the nuclear isospin by the mass difference between proton and neutron, such terms lead to a non-renormalizable theory. Furthermore, the inclusion of explicit mass terms is not viable also in the case of fermions: such terms would violate the Gauge symmetry by coupling the left-handed and the right-handed components of the fields, which transform differently under the symmetry of the theory.

The solution to the problem of the mass of elementary particles came from the *spontaneous symmetry breaking* mechanism, introduced by Goldstone [51, 52] and established in relativistic gauge theories by Higgs [60], Englert and Brout [41]. This mechanism occurs when the Lagrangian of the theory is invariant under a continuous global symmetry but one of the fields is not, having an expectation value on the vacuum different from zero. This corresponds to a configuration where the minimum of the potential breaks the symmetry, which is no longer realized in the particles' spectrum. To illustrate this mechanism, let us consider a theory with a complex field, ϕ , with quartic interaction:

$$\mathcal{L}_\phi = \delta_\mu \phi^\dagger \delta^\mu \phi - \mu^2 |\phi|^2 - \lambda |\phi|^4 \quad (4.1)$$

which, for $\mu^2 < 0$ and $\lambda > 0$, has the minimum of the interaction potential for $\phi_0 = v = \sqrt{\frac{-\mu^2}{2\lambda}} e^{i\theta}$. Under these conditions, the minimum has an arbitrary phase, which describes rotations along the extreme of the potential: none of the single minimum values respects the symmetry, but the symmetry is reflected in the locus of points of the minima. By reparametrising the field as $\phi = \frac{v + \sigma + i\eta}{\sqrt{2}}$, the Lagrangian in Eq. 4.1, can be rewritten, in the unitary gauge choice, as:

$$\mathcal{L}_\phi = \frac{1}{2} (\delta_\mu \sigma)^2 - \frac{1}{2} (2\lambda v^2) \sigma^2 + \frac{1}{2} (\delta_\mu \eta)^2 - \lambda v \sigma (\eta^2 + \sigma^2) - \frac{\lambda}{4} (\eta^2 + \sigma^2)^2 \quad (4.2)$$

where it can be seen how the spontaneous breaking of the symmetry introduces a mass of $2\lambda\nu^2$ associated to the scalar field σ and a new massless particle η , the Goldstone boson.

When promoting the global symmetry to a local one, the spontaneous symmetry breaking implies that the corresponding gauge field gains mass. This formulation of the model has still 4 degrees of freedom: the degree of freedom associated to the η field of the Goldstone boson can be absorbed, with an appropriate gauge choice, by the vectorial field, allowing its longitudinal polarization and thus its interpretation as a massive spin-1 particle, accounting for 3 degrees of freedom. The fourth degree of freedom is provided by the scalar particle with mass $2\lambda\nu^2$, called the Higgs boson (H).

In the $SU(2)_L \times U(1)_Y$ theory, this mechanism is actuated via the introduction of a doublet of complex scalar fields. Since such mechanism is repeated for each broken generator of the symmetry, in its application to the electroweak theory care is taken to leave the vacuum state electrically invariant, such that the photon would remain massless. After the symmetry breaking, the electroweak symmetry group survives in the form of $U(1)_Q$, where Q is the electric charge that remains exactly conserved. The mechanism is not only successful in providing masses for the intermediate bosons, but can also be applied to explain the masses of the fermion fields f , by introducing Yukawa interaction terms to ϕ , of the form $-g_f \bar{f}_L \phi f_R + h.c.$, which are now invariant under the symmetry of the Lagrangian.

The Higgs boson mass is an undetermined parameter in the electroweak theory, but it is directly sensitive to the energy scale Λ up to which the theory should be valid. The calculation of the Higgs boson mass involves terms quadratic in Λ and in the couplings to fermions and to gauge bosons, which can become arbitrarily large. For the mass of the Higgs boson to be much smaller than the Planck scale, counter-terms of the same order have to be involved in the renormalization process, producing a fine-tuned cancellation that results in the mass and the couplings measured experimentally. This process is in contrast with the *naturalness* principle [61], which would require the ratios between parameters of a model to be of order of unity. In general, the degree to which naturalness is violated can be inspected by precisely measuring the couplings and the properties of the Higgs boson. Conversely, if no deviation from the SM is observed by a precise measurement of the Higgs boson couplings, a lower bound on the scale Λ can be extracted [43].

The Higgs boson arising as a signature of the spontaneous symmetry breaking mechanism, formulated in the 1960s, was discovered in 2012 by the CMS and ATLAS experiments [2, 24], based on proton-proton collisions at center of mass energy of 7 and 8 TeV. Since then, a vast set of measurements and characterizations of the newly discovered particle has been expanding, to accurately determine its properties and test them against the predictions of the SM [76]. The measurements of production and decay rates so far, including those described in this thesis, show consistency with the theoretical expectations. The mass of the boson, which is not predicted by the model, has been measured by the CMS and ATLAS experiments based on the data collected during the Run1 of the LHC to be $125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV [1]. Similar measurements carried out on Run2 data at 13 TeV confirm the agreement between measurements and predictions for the Higgs boson.

4.3 HIGGS BOSON PRODUCTION AT THE LHC

The structure of the interaction of the Higgs boson and its peculiar role in the theory are responsible for its couplings to the other SM particles to be proportional to their mass. Such couplings are completely determined from theory, once the Higgs boson mass is fixed. This feature governs the relative size of the cross sections of the possible production mechanisms. In the following, the main production channels in proton-proton collisions are illustrated. Numerical values are provided assuming a mass of the Higgs boson of 125 GeV and proton-proton collisions at the center of mass energy of 13 TeV, as illustrated in Figure 4.2.

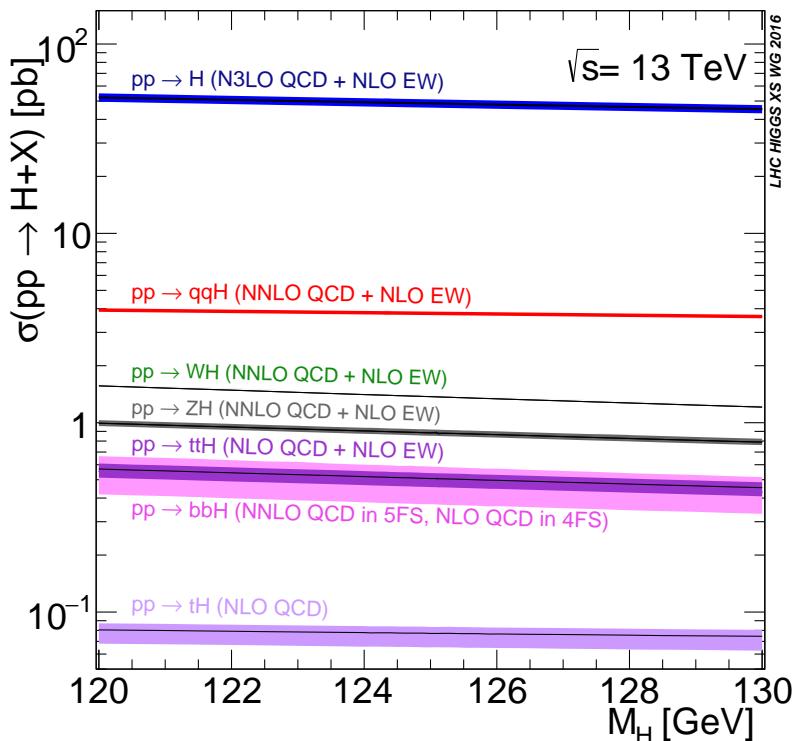


Figure 4.2: The production cross section for the SM Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV is shown as a function of the mass of the Higgs boson, separately for different production mechanisms [35].

The dominant production mechanism at the LHC is the gluon-gluon fusion (ggH). The coupling of the gluons to the Higgs boson occurs via a fermionic loop, dominated by heavy quarks, in particular the top quark. The process thus starts already at loop-level in the Born approximation, but its cross section is enhanced by the large probability of probing gluons in the incoming protons, as shown by the Parton Distribution Functions (PDFs) distribution in Figure 4.4. The most recent calculations of the ggH cross section include contributions up to Next-to-Next-to-Next to Leading Order (N₃LO) [8] and yield a result of 48.52 pb.

The vector boson fusion (VBF) process arises from the scattering of two quarks, which exchange a W or a Z boson, which in turn radiates the Higgs boson. The process can also be visualized as the incoming quarks radiating two vector bosons, which *fuse* in

the vertex where the Higgs boson is radiated, hence the name of the mechanism. The outgoing quarks will have a transverse momentum of the order of a fraction of the vector boson mass, leading to the production of two highly energetic jets with a large rapidity separation. This feature can be experimentally exploited to tag the VBF events and enable their study. The peculiar QCD radiation structure of this process, related to emissions from the incoming quarks and thus channeled towards high rapidities, provides additional lever arm to reject background processes. The radiation pattern is also responsible for rather different distributions between VBF and ggH+2 jets production. In particular, the azimuthal separation between the two jets and the rapidity of the Higgs boson with respect to the average rapidity of the two jets (Zeppenfeld variable [82]) have a strong discriminating power in separating the two distributions. The differential measurement of these quantities, presented in Part 6, is thus sensitive to the relative contribution of the two processes. The process is computed at Next-to-Next to Leading Order (NNLO) and Next to Leading Order (NLO) accuracies in QCD and electroweak approximation, respectively, yielding a predicted cross section of 3.779 fb [35].

The third most relevant production mechanism at the LHC is the associated production of a vector boson and the Higgs boson or *Higgs-Strahlung* (VH). At lowest order, it involves a quark-anti-quark pair to produce an off-shell vector boson, which radiates the Higgs boson. Compared to VBF, whose initial state is composed by valence quarks of the proton, the size of VH contribution is penalized by the presence of an anti-quark. According to the decay channel of the vector boson, different techniques can be applied to use the decay products as handles for the tagging of VH events. The cross section for this process is calculated to be 2.2514 pb [35], with NNLO and NLO accuracies in QCD and electroweak approximation, respectively.

Finally, the associated production of a quark-anti-quark top pair with the Higgs boson (ttH) has a very small cross section, but it is of particular interest since it allows to explore the direct coupling between the Higgs boson and the top quark. Several diagrams contribute to this process, whose cross section is calculated at NLO accuracy and found to be 0.5065 fb [35].

A summary of the Feynman diagrams contributing to the ggH, VBF, VH, and ttH production mechanisms is given in Figure 4.3.

As discussed for the production modes, the couplings of fermions and gauge bosons to the Higgs boson are proportional to their masses and determined once the Higgs boson mass is fixed. This is reflected in the fact that the Higgs boson decays with the largest probability into the heaviest particles made available by kinematic constraints. The possible channels include decays to fermion-anti-fermion pairs, vector boson pairs, and decays to massless particles (photons, gluons) via a loop containing massive particles.

4.3.1 The diphoton decay channel

Of particular interest is the decay channel to two photons, which, despite the small branching fraction (2.7×10^{-3}), provides a clear and fully reconstructable signature with two high energetic photons, thanks to the photon energy resolution of the ECAL allowing the building of a narrow mass peak over a continuous background. These features made this channel one of the *golden* channels for the discovery and study of the Higgs boson, and the analyses presented in Parts 5 and 6 explore further this decay mode. The prin-

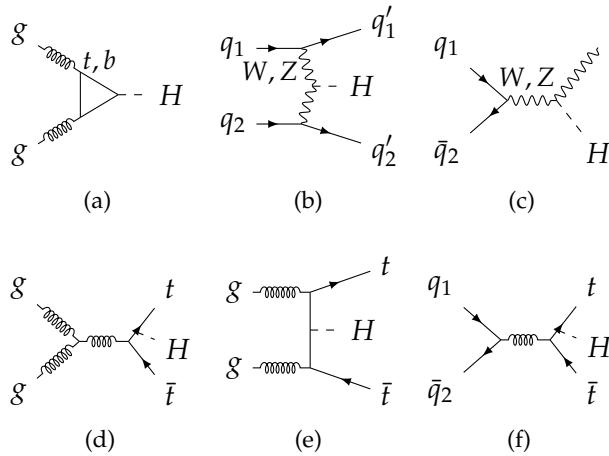


Figure 4.3: The Feynman diagrams for the four main production mechanisms in proton-proton collisions at the LHC: ggH (a), VBF (b), VH (c), and ttH (d, e, f).

cipal Feynman diagrams contributing to this decay mode correspond to the fermionic loop, dominated by the top quark, and to the loop containing W bosons, which interfere destructively with each other. The loop-induced nature of this decay mode makes its study particularly interesting to indirectly test the presence of BSM particles through their contribution to the decay loop.

The final state, characterized by two photons of large momentum, can be mimicked by other, largely more abundant processes of the SM, which constitute the background contaminations of the measurements performed in this channel. These can be grouped into:

- irreducible backgrounds: production of two photons in the hard-scattering interaction, also called *prompt-prompt* production. The contribution of this process, having two genuine photons in the final state, cannot be reduced by means of a photon selection with increased purity. It is nevertheless possible to partially reduce this contribution by exploiting the different kinematic properties of the diphoton pair compared to that produced by the Higgs decay, as illustrated in Section 5.8;
- reducible backgrounds: production of photon + jet or multi-jet events. Jet fragments, in particular collimated photons resulting from the decay of neutral hadrons (π^0 , η^0), can give rise to a large electromagnetic activity in the jet and can be identified as photons by reconstruction algorithms. It is possible to distinguish between such nonprompt photons, also indicated as *fake photons* or *fakes*, and prompt photons by exploiting the difference in the lateral development of their electromagnetic showers in the ECAL and in their isolation levels. A photon identification algorithm, described in Section 5.7, is set up for this purpose and used in the analyses presented in Chapters 5 and 6.

4.4 EVENT GENERATION AND SIMULATION

The description of the hadron collisions and of the interactions of elementary particles provided by the SM can be used to build software that simulates the events produced

in the LHC collisions. These programs employ a statistical approach based on Monte Carlo (MC) methods to scan the phase space of the simulated processes and to approximate integrals not solvable analytically. For this reason, they are also indicated as MC programs or generators. The simulation method relies on the separate description of a few ingredients: the probabilistic sampling of the structure of the colliding protons; the calculation of the cross section of the partonic interaction; the hadronization process of coloured final state particles. This model is summarized by the *factorization theorem*, which is expressed as follows

$$\begin{aligned} \sigma(pp \rightarrow X; p_1, p_2) = \\ \sum_{ij} \int dx_1 dx_2 f_i^p(x_1, \mu_F) f_j^p(x_2, \mu_F) \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \hat{x}, \mu_F^2, \mu_R^2) D(\hat{x}, X, \mu_f) \end{aligned} \quad (4.3)$$

The hadronic cross section σ of the production of a final state X from two initial protons of momenta p_1 and p_2 is expressed as the product of three distinct elements: the cross section $\hat{\sigma}_{ij}$ of the partonic process producing a state \hat{x} from two partons i and j ; the probabilities of finding such two partons with a given momentum fraction of the protons, expressed by the PDFs f_i^p ; the fragmentation function D of the partonic state \hat{x} into the hadronic state X , dependent on the fragmentation energy scale μ_f . The fragmentation functions are estimated in the calculations by parton shower models, which approximate the description of the radiation of soft partons and their branching into pairs, and by hadronization models, which describe the formation of final state hadrons. The cross section is then obtained by integrating over all the possible momentum fractions and by summing over the parton combinations contributing to the process described. The renormalization and factorization scales are indicated by μ_R and μ_F , respectively. The renormalization scale appears as an energy dependence in the process of renormalization of ultraviolet divergences in the cross section calculation, while the factorization scale is introduced to cure divergences arising from infrared emission of soft and collinear particles.

The PDFs are extracted from data with different techniques and available in different sets. An illustration of the behavior of the PDF at different scales is given in Figure 4.4. The simulated samples used in the analyses presented in this thesis are obtained with the NNPDF3.0 PDFs [13], whose methodology is validated by closure test on pseudo-experiments.

The simulated signal events are generated with **MADGRAPH5_AMC@NLO v2.2.2** [7], which calculates matrix element contributions to the production of the Higgs boson and two additional jets up to NLO accuracy in perturbative Quantum Chromo Dynamics (pQCD). The parton level is interfaced to the **PYTHIA8.205** [91] generator to describe the hadronization process, applying the FxFx merging scheme [45]. PYTHIA also takes care of simulating activity arising from remnants of additional softer partonic interactions, using the CUETP8M1 UE tune.

An alternative simulated sample of ggH events is obtained with the **POWHEG** program [6] [12], which includes the simulation of only one additional jet at matrix element level. This sample provides a second benchmark for comparison of the measurements, presented in Part iii, of cross sections inclusive in the number of jets or differential in the kinematic of the first additional jet in the event.

The ggH events, generated both with **MADGRAPH5_AMC@NLO** and **POWHEG**, are weighted as functions of the p_T of the Higgs boson and the number of jets in the event to match the

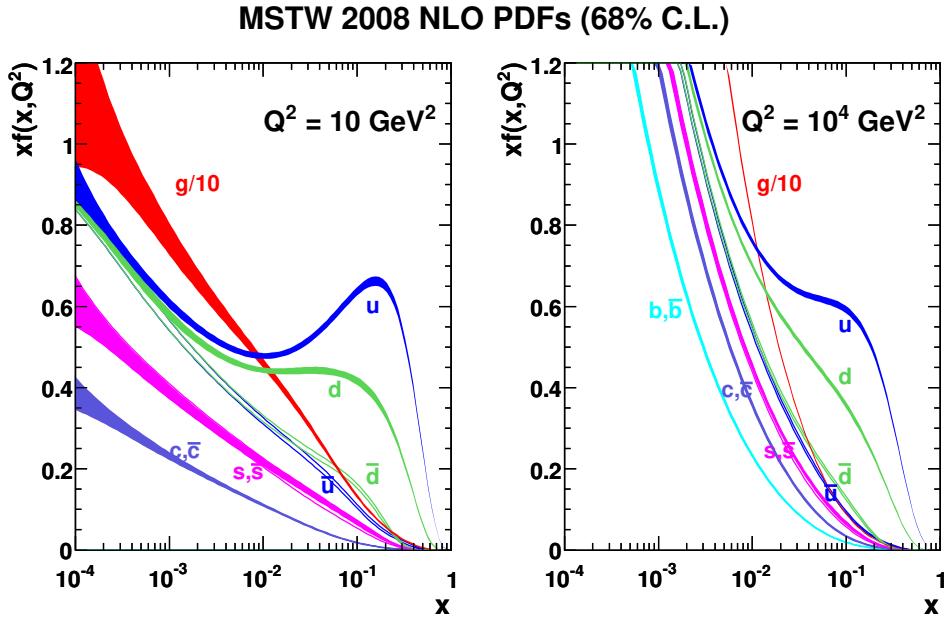


Figure 4.4: The Martin-Stirling-Thorne-Watt parametrization of the PDFs as a function of the momentum fraction of the proton, for two different energy scales [75].

prediction of the `POWHEG NNLOPS` program [57][65]. This program has the advantage of predicting at NNLO accuracy, both the differential cross section with respect to the QCD radiative effects and the normalization of the inclusive cross section [58].

The dominant prompt-prompt background is simulated with the `SHERPA v.2.2.1` [50] generator, which includes the Born processes with up to three additional jets and the box processes at Leading Order (LO) accuracy. Reducible prompt-fake and fake-fake backgrounds are accounted for with simulation of $\gamma + \text{jet}$ and multijet events, entirely modeled with `PYTHIA`. The efficiency in the generation of jets with a large electromagnetic component is increased by applying a filter which requires a signal associated to photons, electrons, or neutral hadrons with a p_T of at least 15 GeV.

Other processes are simulated and used for validation of the analysis tools and measurement of selection efficiencies: `MADGRAPH5_AMC@NLO` is used to simulate $W\gamma$ and $Z\gamma$ events at LO and Drell-Yan events at NLO.

The simulation of the events is completed by their interaction with a detailed model of the response of the CMS detector, implemented with the `GEANT4` [5] package.

PROPERTIES OF THE HIGGS BOSON IN THE DIPHOTON DECAY CHANNEL

In this Chapter, the measurement of the properties of the Higgs boson in the $\gamma\gamma$ decay channel are presented. After a brief description of the motivation for these studies, the experimental techniques deployed to precisely characterize the Higgs boson in this channel are illustrated, followed by a discussion of the results.

5.1 INTRODUCTION

As illustrated in Part 4, the SM predicts the existence of a scalar boson particle connected with the mass generation mechanism. After a decades-long quest, a new particle of this kind, with a mass of about 125 GeV, was discovered by the CMS [20, 23] and ATLAS [2] collaborations in 2012, using proton-proton collision data produced by the LHC at center-of-mass energies of 7 and 8 TeV. The $H \rightarrow \gamma\gamma$ channel, along with the $H \rightarrow ZZ^* \rightarrow 4l$ and $H \rightarrow WW^*$ decay modes, played a central role in the achievement of this historical result. The discovery of the particle was followed by a vast campaign of studies to ascertain its properties. The characterization covered all the production and decay modes accessible with the full data set delivered to both experiments during the Run1 of the LHC, consisting of about 6 fb^{-1} (7 TeV) plus about 23 fb^{-1} (8 TeV), as described in Section 2.1. The final results of these measurements combine the ATLAS and CMS analyses [3, 4] to provide the best precision achievable and indicate that the new particle is compatible with the SM predictions, within experimental and theoretical uncertainties.

The Run2 of the LHC entails proton-proton collisions at a 13 TeV center-of-mass energy. The new collision energy, compared to the 8 TeV of the Run1, gives rise to an increase in the production cross section of the Higgs boson by a factor 2 to 4, depending on the production mode, as shown in Figure 5.1.

This effect enabled to soon re-discover the Higgs boson at 13 TeV in 2016 with just 12.9 fb^{-1} of collected data [96, 100]. The most recent measurements are based on the full data set collected in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1} . The analysis of such data set allowed to surpass the precision of the Run1 measurements performed by the individual Collaborations and constitutes the most detailed and precise study of the diphoton decay channel in CMS to date. The analysis presented here provides measurements of the Higgs boson production rates with respect to the SM predictions, also referred to as *signal strengths*, as well as measurements of the coupling modifiers to fermions and bosons, and effective coupling modifiers to photons and gluons, in the so-called κ framework [71]. The precise determination of rates and couplings, indeed, allows the constraint of possible deviations from the Higgs boson predicted by the SM. After its principal results, this analysis is often referred to as the *coupling* analysis of the diphoton decay channel of the Higgs boson, and this convention will be observed in this thesis.

These results are presented in a recently-submitted publication [27], whose discussions, text and illustrations are closely followed in the present Chapter of this thesis.

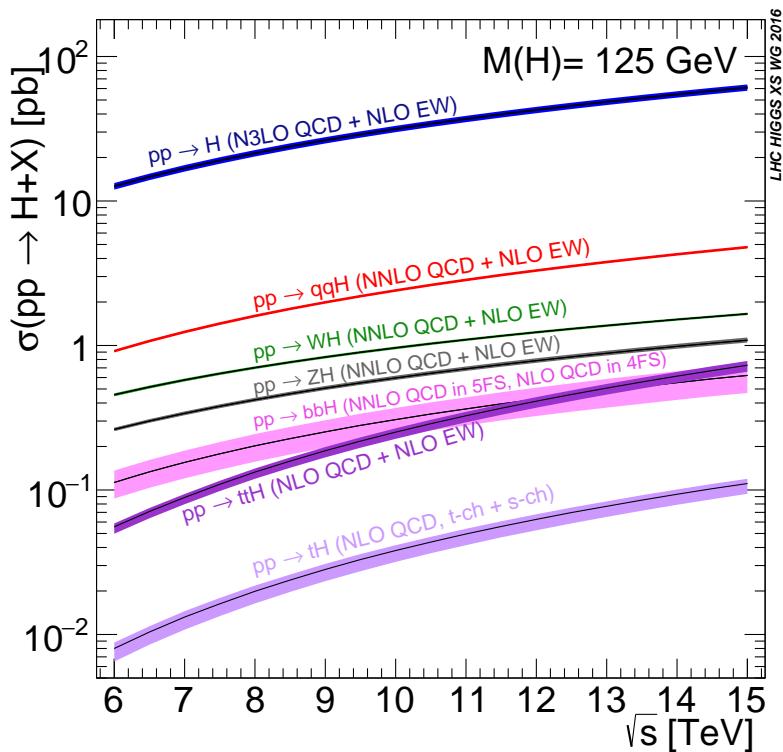


Figure 5.1: The production cross section, for different production mechanisms in proton-proton collisions, for a Higgs boson with a mass of 125 GeV as a function of the center-of-mass energy [35].

5.2 ANALYSIS STRATEGY

The signal extraction relies on the diphoton invariant mass distribution of the selected events. The Higgs boson production gives rise to a narrow peak in this distribution that can be statistically separated from the continuous, smoothly-falling distribution due to the background processes. The analysis sensitivity is improved by splitting the events into categories, of two kinds:

- categories targeting a specific production mode. These categories are defined requiring final state features peculiar to a particular production mode: forward jets for VBF, top decay products (such as leptons, missing transverse momentum and b jets) for ttH, and Z/W decay products for VH. These distinctive characteristics allow the reduction of the contaminating backgrounds and thus a better signal-to-background (S/B) ratio. At the same time, these categories allow sensitivity to couplings of the Higgs boson to different kinds of particles.
- generic event categories, which aim at selecting events with higher probability of originating from a signal process, reducing background contributions, and at providing improved S/B based on the mass resolution information and on the diphoton kinematic features. These categories tend to be populated by the most abundant production process, ggH.

Frequent use of multivariate techniques, in addition to those already introduced in Part i for object reconstruction and identification, is done in the definition of the categories. They are employed for the identification of events with signal-like kinematics, for improved tagging of forward jets in VBF events, for the reduction of ggH contamination in VBF categories, and for identification of ttH events among high-jet multiplicity events. Furthermore, multivariate classifiers are used to identify the primary vertex of $H \rightarrow \gamma\gamma$ events and to estimate its probability of being the correct diphoton vertex, as well as to separate prompt photons from misidentified jets. Since simulated events are used for training of these MVAs, imperfect modeling of background and signal processes might result in sub-optimal performances of the discriminators. On the other hand, since the parametrization of the background shape is obtained directly from data, the same mismodelling has no impact or bias on the correctness of the results extracted.

5.3 TRIGGER SELECTION

The $H \rightarrow \gamma\gamma$ decay is characterized by the presence of two isolated highly energetic photons. Based on this feature, a selection is tuned at the level of the trigger system, described in Section 2.2.5, and an HLT path is defined to extract and collect these events from the collisions. The trigger selection requires the presence of two photon candidates and has asymmetric thresholds on their transverse energy, set at 30 and 18 GeV for the higher- p_T and the lower- p_T photon of the pair, also referred to as leading and sub-leading photon, respectively. The photons are then required to satisfy a loose calorimetric identification based on the shape of the electromagnetic showers, a loose isolation requirement, and the ratio between the energy deposits in the HCAL and the ECAL for the photon candidate must fall below a threshold. The efficiency of the trigger selection is measured in

$Z \rightarrow e^+e^-$ events with the Tag and Probe technique, introduced in Section 3.8 and it is found to be $> 99\%$ on the events selected for the analysis.

5.4 VERTEX IDENTIFICATION BDT

The identification of the primary vertex has a direct role in the estimation of the diphoton invariant mass resolution and therefore it has a crucial role in the performance of the analysis. However, since the photons do not deposit energy by ionization in the tracker, the vertex assignment in a diphoton event has to rely on the exploitation of the tracks of charged particles recoiling against the diphoton system. A multivariate classifier based on a BDT is used to combine the most discriminant features of the ensemble of recoiling tracks, in particular:

- the sum of the squared transverse momenta, $\sum_i |\vec{p}_T^i|^2$;
- the sum of the projections of the charged PF candidates p_T on the direction of the $\gamma\gamma$ system in the transverse plane, $-\sum_i (\vec{p}_T^i \cdot \frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|})$,
- the scalar p_T imbalance between the diphoton and the charged particle tracks, $(|\sum_i \vec{p}_T^i| - p_T^{\gamma\gamma}) / (|\sum_i \vec{p}_T^i| + p_T^{\gamma\gamma})$,

where the index i runs over all the charged particle-flow candidates associated with a given vertex and the diphoton transverse momentum $\vec{p}_T^{\gamma\gamma}$ is calculated with respect to the same vertex.

If either of the photons converted into an electron-positron pair in the material upstream the ECAL, additional information from the conversion track is used by the algorithm.

The accuracy required in the vertex identification is 1 cm in the longitudinal direction. Vertex displacements smaller than 1 cm, indeed, are found to bring a negligible contribution to the mass resolution, compared to the contribution arising from the calorimetric energy resolution of the individual photons. A vertex assignment is thus considered correct if the identified vertex lies within 1 cm along z with respect to the true one. The algorithm described has an average vertex finding efficiency of 80%.

5.5 VERTEX PROBABILITY BDT

A second BDT is trained to estimate the event-by-event probability of a correct vertex assignment. To do so, the algorithm uses information on the number of vertices, the output value of the vertex identification BDT for the three highest-scoring vertices, the distances between the identified vertex and the second and third choices, the number of reconstructed conversion tracks, and the diphoton transverse momentum.

The performance of both vertex BDTs are validated in $Z \rightarrow \mu^+\mu^-$ events, as shown in Figure 5.2 (left), where the vertex finding is operated after the removal of the muon tracks, mimicking a diphoton event without conversions. The performance in presence of conversions is similarly studied using $\gamma + \text{jet}$ events. An uncertainty is associated with the discrepancies between data and simulation and propagated to the later stages of the

analyses. The efficiency of selecting a vertex within 1 cm of the true one is also measured for simulated signal events as a function of the p_T of the diphoton system, and it is compared to the average probability of a correct vertex selection as estimated by the vertex probability BDT, as shown in Figure 5.2 (right).

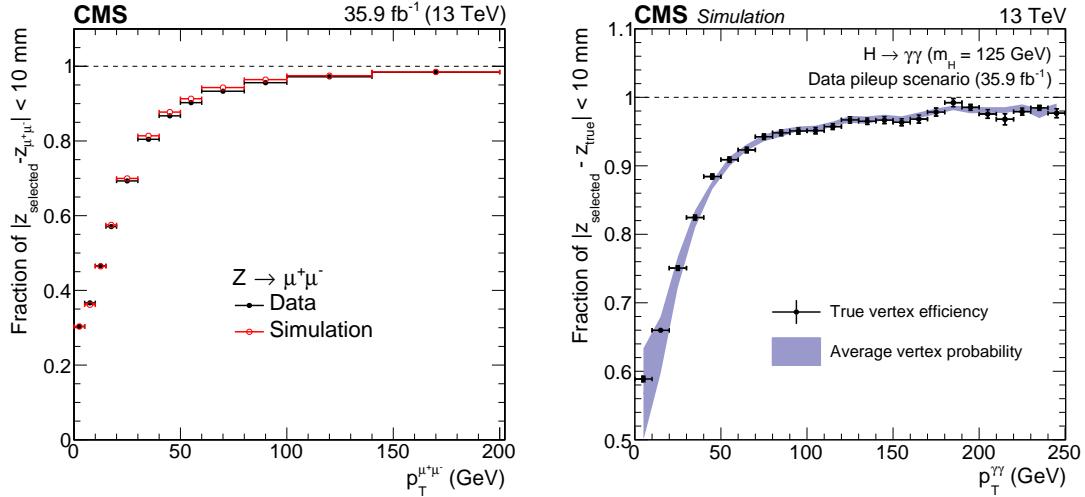


Figure 5.2: (Left) Validation of the vertex identification BDT in data and simulation for $Z \rightarrow \mu^+\mu^-$ events, where the muon tracks are removed to mimic a diphoton event. (Right) Vertex finding efficiency on signal events as a function of the diphoton p_T , compared to the average vertex probability BDT score.

5.6 PHOTON PRESELECTION

A loose set of requirements is applied to all photon candidates. Such requirements are similar, though slightly more stringent, than those enforced by the trigger selection described in Section 5.3. The preselection thus ensures that both data and simulation are restricted to a common phase space. The leading (sub-leading) photon is required to have a $p_T > 30$ (20) GeV and a pseudorapidity smaller than 2.5, with the exclusion of the region $1.44 < |\eta| < 1.57$, corresponding to the transition region between EB and EE where the lateral containment of electromagnetic shower is imperfect. Further requirements are applied with different thresholds depending on the $|\eta|$ and R_9 of the photon candidate, as summarized in Table 5.1. These involve selection on the ratio of the energy in the HCAL behind the supercluster to the energy of the supercluster (H/E) to reject hadrons, on the lateral extension of the shower ($\sigma_{\eta\eta}$) defined as the energy spread in the 5×5 crystal matrix centered on the crystal with the highest energy in the supercluster, on the photon isolation (\mathcal{I}_{ph}) computed as the sum of transverse energy of photons inside a cone of $\Delta R = 0.3$ around the photon candidate, and on the track isolation (\mathcal{I}_{tk}) corresponding to the sum of the transverse momentum of all tracks in a hollow cone around the photon with (inner) outer radius of $\Delta R = 0.3$ (0.04). Additionally, all photons have to satisfy either $R_9 > 0.8$ and $\mathcal{I}_{ch} < 20$ GeV, or $\mathcal{I}_{ch}/p_T < 0.3$, where \mathcal{I}_{ch} is the charged-hadron isolation, calculated as the sum of the transverse momenta of charged particles in a cone around the photon of $\Delta R = 0.3$. Finally, electrons are vetoed, by rejecting the photon candidate

if its supercluster is matched to an electron track with no missing hits in the innermost tracker layers.

	R_9	H/E	$\sigma_{\eta\eta}$	\mathcal{I}_{ph} (GeV)	\mathcal{I}_{tk} (GeV)
Barrel	[0.5, 0.85]	< 0.08	< 0.015	< 4.0	< 6.0
	> 0.85	< 0.08	–	–	–
End caps	[0.8, 0.90]	< 0.08	< 0.035	< 4.0	< 6.0
	> 0.90	< 0.08	–	–	–

Table 5.1: Summary of preselection requirements on shower shape variables and isolation quantities.

The efficiency of the preselection criteria is measured with the Tag and Probe technique, described in Section 3.8, applied to $Z \rightarrow e^+e^-$ events for all requirements, except the electron-veto, for which $Z \rightarrow \mu\mu\gamma$ events are used. The ratio between the efficiencies measured in data and simulation is found to be close to one for all photons and it is used to correct the signal yields estimated from simulation in the analyses.

5.7 PHOTON IDENTIFICATION BDT

As illustrated in Section 4.3.1, photons produced in the jet fragmentation or collimated photons from π^0 , η^0 hadron decays represent nonprompt and fake photon production that pollutes the selection of prompt photons from the Higgs boson decay. These contributions are rejected by exploiting their different electromagnetic shape profiles and isolation levels. A BDT is trained for such purpose with the following inputs:

- shower shape variables. The most discriminating ones (the width in η of the SC, R_9 , S_4 , $\sigma_{i\eta i\eta}$, defined in Section 3.3) are corrected in simulation to attenuate the discrepancy with data;
- isolation variables, \mathcal{I}_{ph} and \mathcal{I}_{ch} ; in order to reject misidentified photons originating from a vertex different than the selected one, \mathcal{I}_{ch} is also computed with respect to the vertex associated with the largest isolation sum and given as input to the BDT;
- the pseudorapidity and the energy of the ECAL supercluster associated with the photon candidate;

The effect of pileup on the above quantities is accounted for in the discriminator by including the median energy density per unit area in the event, ρ , among the input variables.

A flat identification efficiency as a function of η and energy of the photon is a desirable feature in order to reduce the model-dependence of the photon selection. After transforming, prior to the BDT training, the distributions of these two observables such that they have the same shape for prompt and misidentified photons, the efficiencies are observed to be flat.

The output of the photon identification BDT for the lowest-scoring photon of the pair in data and simulation is shown in Figure 5.3 (left). The events are required to have a diphoton invariant mass between 100 GeV and 180 GeV and to fulfill the preselection

requirements. The BDT assigns high values (close to 1) to photons from signal events and from events simulating the irreducible background coming from prompt diphoton production, while low values (close to -1) are assigned to photons from simulated background events where at least one photon originates from jet fragmentation.

The agreement between data and simulation for the photon identification BDT is evaluated on $Z \rightarrow e^+e^-$ events, with electrons reconstructed as photons, as shown in Figure 5.3 (right). A systematic uncertainty is assigned to the shape of the discriminator, estimated in order to cover the largest observed discrepancy between data and simulation.

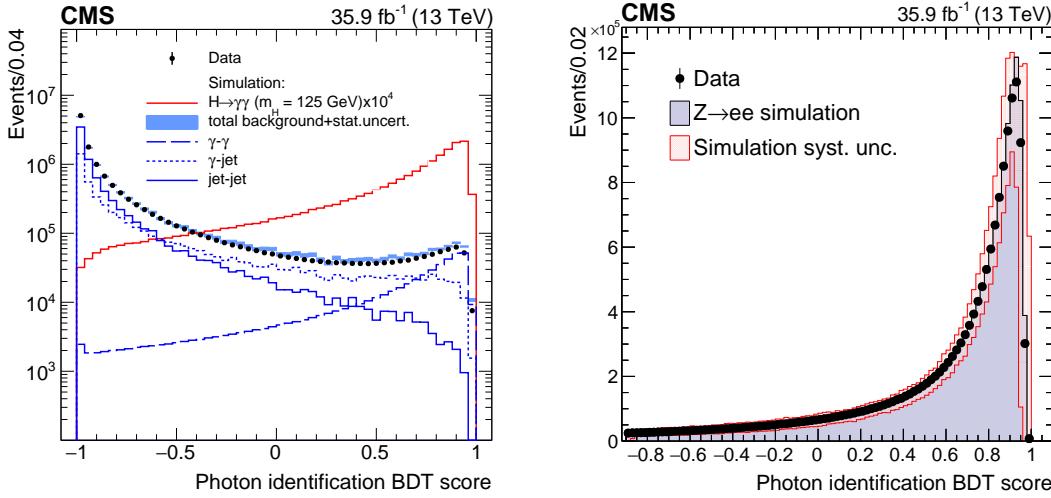


Figure 5.3: (Left) The photon identification BDT score is compared between data (points) and simulation (histograms), separately for signal (red) and background (blue) events. The BDT output for the lowest-scoring photon of the pair is shown. (Right) The same distribution is shown for $Z \rightarrow e^+e^-$ events in data and simulation, where electrons are reconstructed as photons. The red hashed region shows the systematic uncertainty assigned to the shape of the discriminator.

5.8 DIPHOTON CLASSIFIER BDT

The precision of the measurements extracted is improved by dividing the events into categories. The most performant of such categories are characterized by a higher signal-to-background ratio and better mass resolution. The primary event categorization is performed via a multivariate classifier that aims at separating events containing a diphoton pair with signal-like kinematic characteristics, good invariant mass resolution and photon identification BDT scores typical of prompt photons.

For this purpose a BDT is trained, using the Toolkit for Multivariate Data Analysis with ROOT (TMVA) package [92], with the following input variables:

- the transverse momenta for both photons, rescaled for the diphoton mass, $p_T/m_{\gamma\gamma}$;
- the pseudorapidities of both photons, η^1 and η^2 ;
- the cosine of the angle between the two photons in the transverse plane, $\cos(\Delta\phi)$;
- the identification BDT score for both photons;

- the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using the correct primary vertex ($\sigma_{rv}/m_{\gamma\gamma}$);
- the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using an incorrect primary vertex ($\sigma_{wv}/m_{\gamma\gamma}$);
- the output of the vertex probability BDT (p_{vtx}).

The events must not be selected by the classifier according to their invariant mass: given that the BDT is trained under a specific mass hypothesis for the signal ($m_H = 125$ GeV) and that the background processes are instead continuously distributed over the invariant mass range used for the analysis, this feature would be highly discriminant and exploited by the classifier, if available. Such a biased selection would result in distorted invariant mass distributions that would compromise the technique of the signal extraction fit, described in Section 5.11. The set of input variables is thus chosen such that the BDT is blind to the mass information. For the same purpose, all dimensional quantities are rescaled by the invariant mass of the diphoton system.

The relative invariant mass resolution estimators are computed by propagating the photon energy resolution estimator, under the assumption that the functional form of the latter is Gaussian. In the scenario where the correct vertex has been selected for the reconstruction of the invariant mass, no additional contributions are considered for the computation of the invariant mass resolution, since in this case the resolution in the measurement of the energy of the photons is the dominant contribution. The relative mass resolution in the right-vertex case is thus calculated as:

$$\sigma_{rv}/m_{\gamma\gamma} := \frac{1}{2} \sqrt{\left(\frac{\sigma_{E1}}{E_1}\right)^2 + \left(\frac{\sigma_{E2}}{E_2}\right)^2} \quad (5.1)$$

For those events where an incorrect primary vertex is selected, on the other hand, the estimator of the mass resolution includes an additional term, σ_{vtx} , given by the displacement between the correct and the selected primary vertex. Such displacement is distributed as a Gaussian with width $\sqrt{2}\sigma_Z^{beamspot}$, where $\sigma_Z^{beamspot} = 3.7$ cm is the average longitudinal width of the luminous region of the CMS interaction point in 2016. The term σ_{vtx} can be computed analytically given the impact positions of the two photons in the calorimeter and used to define:

$$\sigma_{wv}/m_{\gamma\gamma} := \sqrt{(\sigma_{rv}/m_{\gamma\gamma})^2 + (\sigma_m^{vtx}/m_{\gamma\gamma})^2} \quad (5.2)$$

The distributions of the BDT input variables are shown Figure 5.4 for data and simulation, where a satisfactory level of agreement is found. Residual mis-modeling does not affect the soundness of the analysis, but would only make the classification and categorization of the events sub-optimal. Events from data have to satisfy the same selection simulated events undergo prior to the training of the classifier, described below.

In order to improve the performance of the classifier, the information that the signal-to-background ratio is inversely proportional to mass resolution is injected in its training. To such purpose, signal events are weighted by a combination of the two mass resolution estimators, combined via the probability of a correct vertex choice, as follows:

$$w_{sig} = \frac{p_{vtx}}{\sigma_{rv}/m_{\gamma\gamma}} + \frac{1 - p_{vtx}}{\sigma_{wv}/m_{\gamma\gamma}} \quad (5.3)$$

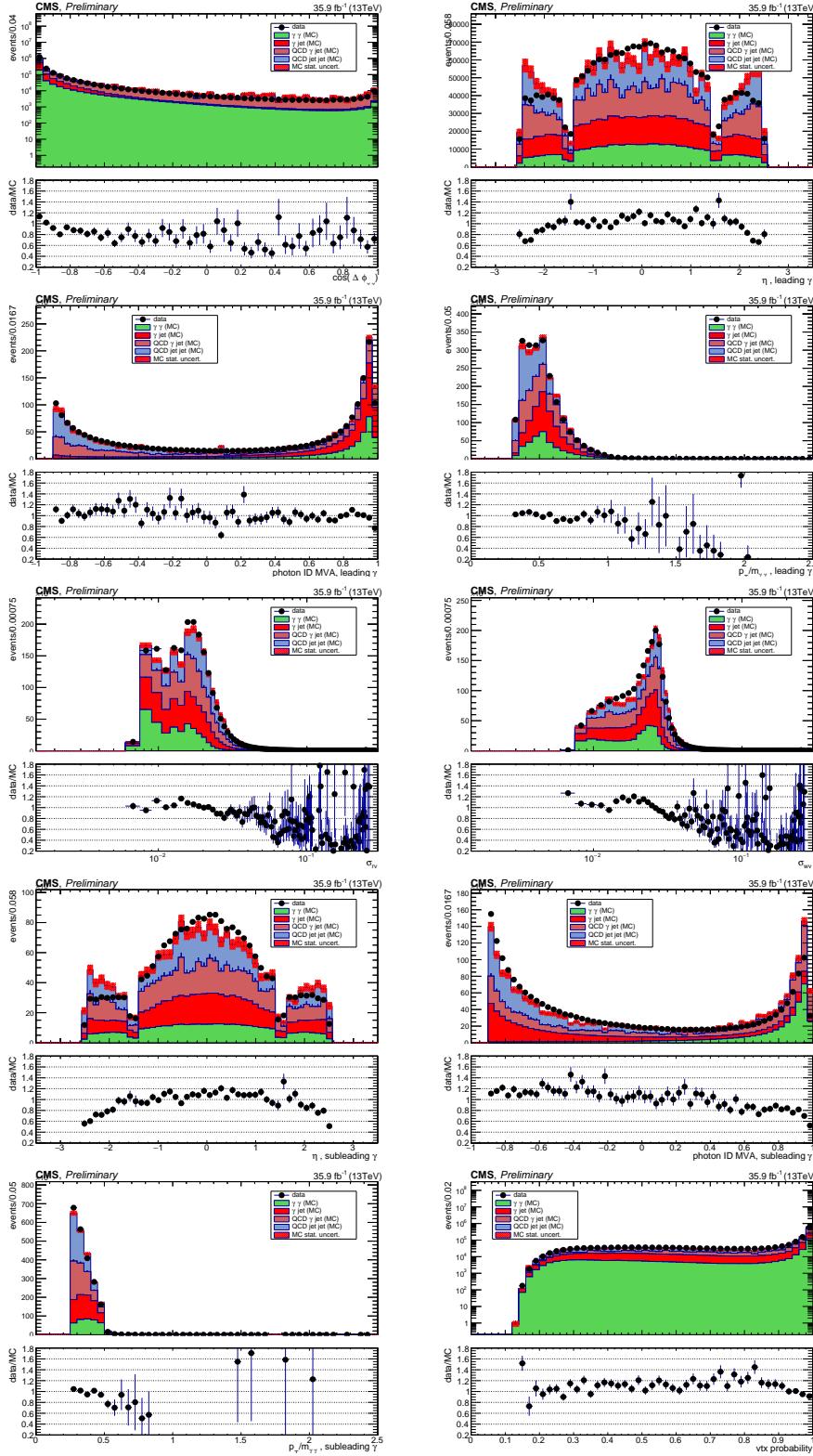


Figure 5.4: Data-simulation comparisons for BDT input variables (photon kinematic variables, mass resolution estimates under correct and incorrect vertex hypotheses, cosine of the angle in ϕ between the two photons, vertex probability). For every variable, the comparison between the cross-section-weighted mixture of simulated (stacked histograms) samples with data (black points), normalized to the area, is shown, together with the ratio plot.

Events with a better mass resolution will be given more importance by the BDT.

The BDT is trained on events with a diphoton invariant mass between 100 and 180 GeV passing, on top of the preselection, a loose requirement on the photon identification BDT output (> -0.9 , corresponding to 99% efficiency on signal events) and a selection on their transverse momentum, scaling with the invariant mass of the diphoton system, set to $p_T > 1/3(1/4)m_{\gamma\gamma}$ for the leading (sub-leading) photon. These p_T selections scaling with $m_{\gamma\gamma}$ allow less stringent cuts for lower $m_{\gamma\gamma}$ diphoton pairs, preventing shaping of the mass distribution at the lower end of the selected mass spectrum.

The four main production mechanisms of the SM Higgs boson (ggH, VBF, VH and ttH) are simulated for a mass $m_H = 125$ GeV and used, each weighted by its cross section, to compose the signal training sample. The `POWHEG` program has been chosen for the simulation of the ggH, VBF and ttH events, since it brings the advantage of having only positive per-event generator weights, which, as opposed to negative weights, are more properly handled in the training procedure. The VH events are simulated with the `MADGRAPH5_AMC@NLO` program, leading to a fraction of events with negative weight. The handling of such events in the training of the classifier is technically non trivial and can cause instabilities in the BDT results [92]. Given the small cross section of the VH process, compared to the other production modes, and considering that only a fraction of its events have a negative weight, such events were ignored in the training procedure. This choice privileges stability of the result over optimal performance in the classification of the VH events, the latter not being the primary goal of the classifier and being addressed by dedicated selections downstream in the analysis strategy (see Section 5.10.2). The events with negative weight, though, are included in the samples used to evaluate and validate the performance of the BDT.

The background training sample is the cross-section weighted mixture of SM backgrounds including the irreducible prompt-prompt production and the reducible contributions from events with at least one photon arising from the misidentification of a jet.

The background composition and the signal-to-background ratio vary as a function of the photon kinematics, with the relative contributions from prompt-prompt, prompt-fake and fake-fake components assuming different magnitudes as a function of the kinematics.

Three quarters of the signal and background samples are used to train the classifier, while the remaining quarter forms the *test sample*, used in later stages of the analysis. Such split is necessary in order to assess the performance of the classifier from a statistically independent sample and to avoid bias due to over-training in later category extraction and optimization.

The fake-fake background simulated sample has much lower statistical power compared to the other background samples. The event weights for this sample have correspondingly a larger magnitude. Such a hierarchy in the size of the event weights of the samples is the cause of numerical instability of the training procedure. In order to avoid this effect, the fake-fake simulated events are down-weighted by a factor 25.

This is a tolerable adjustment to the training configuration, since the output variable of the classifier in TMVA, proportional to the signal-to-background-ratio, is related to the ratio of the normalized probability densities for signal and background events, $p_{sig}(x)/p_{bkg}(x)$. The addition or subtraction of even a large number of background events in a region poorly populated by signal events will not change the signal-to-background discriminating power against the remaining background, but only the numerical value of $p_{sig}(x)/p_{bkg}(x)$, since the fraction of total background in the region is different, and

thus the numerical value of the BDT output variable, which has no physical meaning. The fake-fake events, indeed, are expected and found to mostly populate the low-score region of the BDT output, i.e. the most background-like region, as shown in Figure 5.9, where the BDT score distribution is shown separately for each process of the background. The overall effect of the scaling of the weights of the fake-fake sample is thus to reduce the fraction of total background in the training which survives in the signal-like region and, therefore, changing the numerical values assigned by the TMVA classifier and the output shape. The above difference is mainly technical and is a tolerable effect due to the down-weighting of fake-fake events.

The classification performed by the BDT is studied by inspecting the input variables in different bins of signal efficiency defined on the output of the BDT. By doing so, it is possible to map bins of the score of the BDT to regions of its input variables. The definition of the bins in signal efficiency on the BDT output is shown in Figure 5.5. The same bins are shown for the BDT inputs in Figure 5.6, maintaining the same color key.

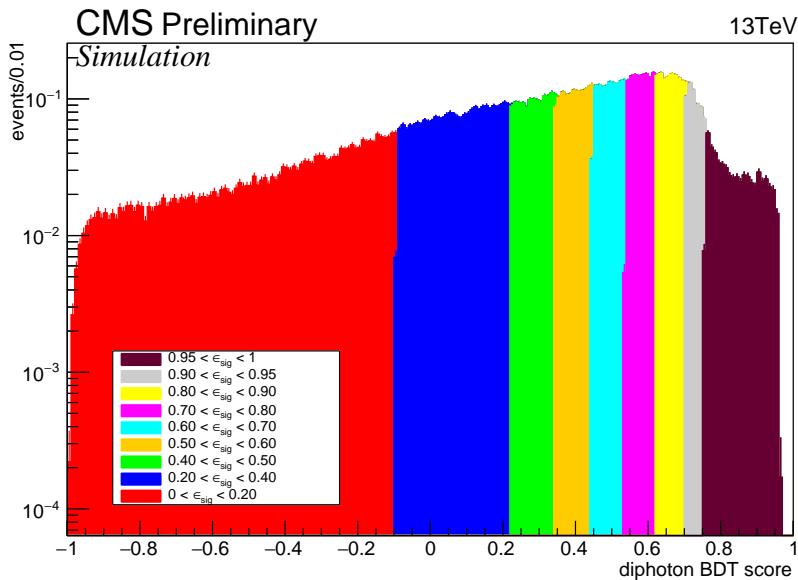


Figure 5.5: Distribution for simulated signal events of the BDT output. Different colors correspond to different bins of signal efficiency with respect to a selection performed on the BDT output.

It can be seen that the classifier assigns high scores to events where the diphoton pair reflects the characteristics of the signal. These events indeed have a boosted kinematic configuration, have higher p_T , principally low pseudorapidity values, good mass resolution and high scores of the photon identification BDT.

The shape of the invariant mass distribution is also examined in the signal efficiency bins defined above. In Figure 5.7 (left), such distribution is shown for simulated background events, normalized to unity in each efficiency bin. The distribution is smooth in all bins, confirming that the BDT is not classifying events according to their invariant mass. The same distributions are shown for simulated signal events in Figure 5.7 (right), where it is made evident that the classifier is assigning a higher score to diphoton pairs with

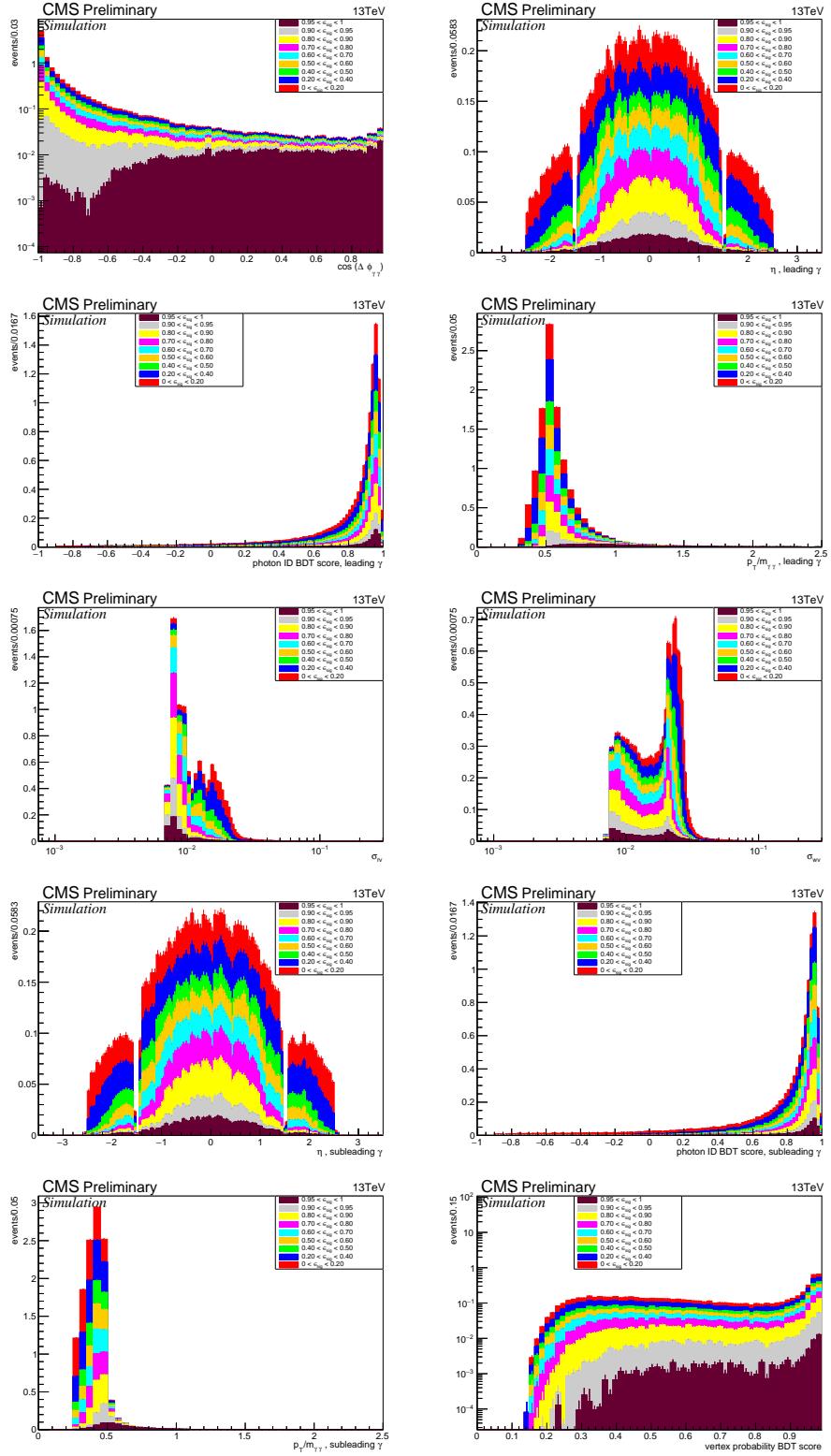


Figure 5.6: Distributions for simulated signal events of input variables (leading photon kinematic variables, mass resolution estimates under correct and incorrect vertex hypotheses, cosine of the angle in ϕ between the two photons). The different colors identify different bins of signal efficiency with respect to a selection performed on the BDT output. It can be seen that events classified in efficiency quantiles corresponding to higher BDT scores have higher p_T , more central pseudorapidities and higher photon identification BDT scores for both photons, narrower mass resolution, and photons closer in the ϕ angle.

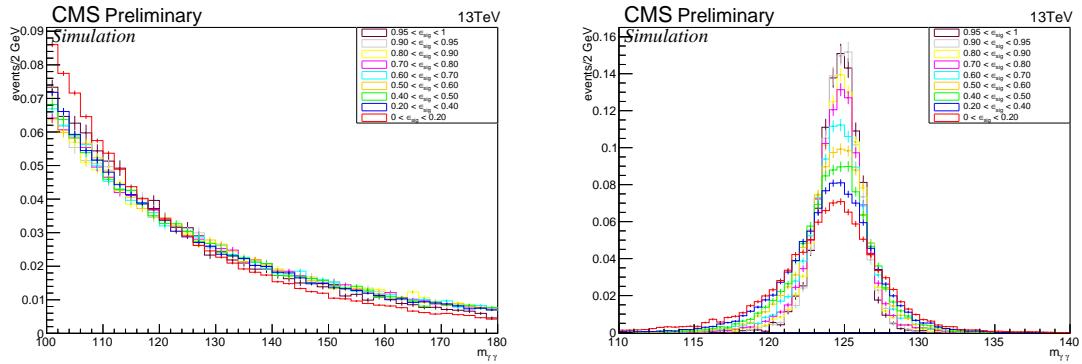


Figure 5.7: The diphoton mass distributions, normalized to unity in every signal efficiency bin, for background (left) and signal (right) events.

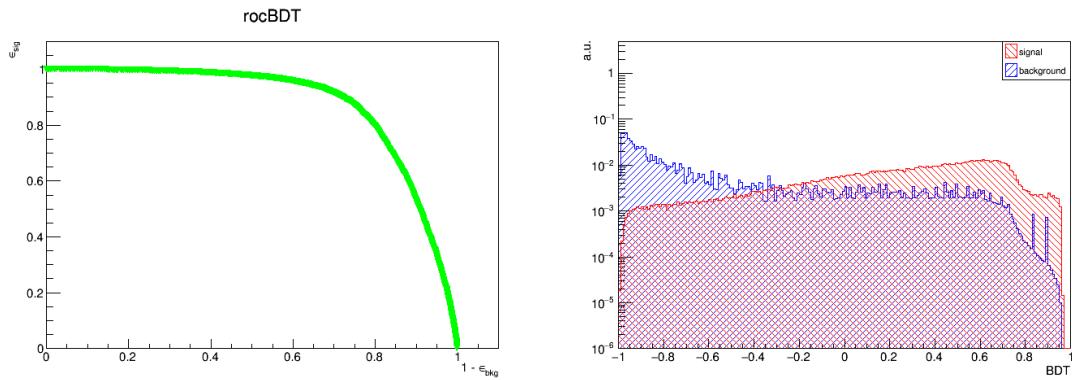


Figure 5.8: The classifier ROC curve (left) and the shapes normalized to unity of the BDT output variable for simulated signal and background events (right).

better invariant mass resolution, since the distributions become narrower as the signal efficiency bin becomes higher.

The discriminating power of the BDT is summarized in Figure 5.8, where the Receiver Operating Characteristic (ROC) curve of the classifier (left) and the shapes, normalized to unity, for simulated signal and background events (right) are shown.

The distributions of the BDT output for individual processes composing the signal and background samples are shown in Figure 5.9. The VBF, VH, and ttH production mechanisms of the Higgs boson have a more boosted topology, compared to ggH [35], which results in higher BDT scores for the events simulating these processes, consistently with the analysis of classifier's selection on its inputs, illustrated earlier in this section. As far as the background samples are concerned, the prompt-prompt component is associated to higher scores of the BDT, given that they are characterized by a better invariant mass resolution compared to the prompt-fake and fake-fake components and that both photons of the pair obtain high scores from the photon identification BDT, while events simulating the prompt-fake and fake-fake components tend to cluster towards lower values of the classifier.

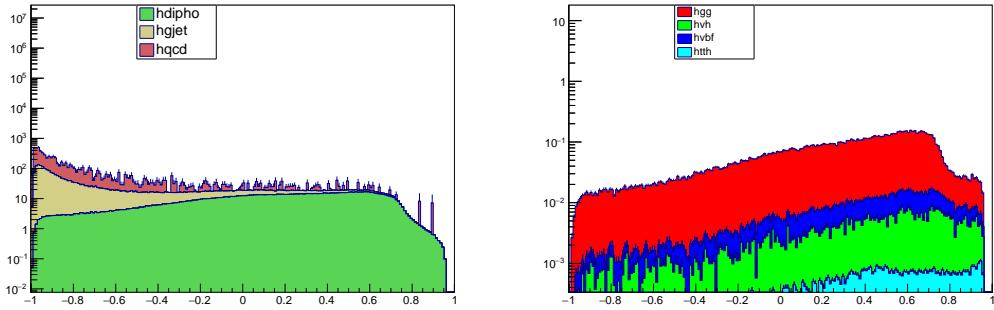


Figure 5.9: Shape of the BDT output variable split for different background (left) and signal (right) samples. Background events with two prompt photons tend to have higher BDT scores. VBF, VH, and ttH processes tend to accumulate to higher BDT scores compared to gluons fusion.

The comparison between data and simulation is also shown for the BDT output distribution in Figure 5.10. The BDT score is here transformed such that it has a flat distribution between 0 and 1 for the sum of the simulated signal processes. Such transformation facilitates the visualization of signal efficiencies directly from the distribution.

5.8.1 Validation with $Z \rightarrow ee$

The signal modeling described in Section 5.11 employs simulation to extract the signal shape and normalization. Monte Carlo programs are expected to model correctly, within the theoretical uncertainties, generator level kinematic variables of the diphoton system. However, other quantities, in particular invariant mass resolution estimators and photon identification BDT scores, are also sensitive to an accurate simulation of the detector geometry and response.

It is thus important to corroborate the fact that the simulation modeling of the signal in data is accurate enough for the signal extraction to be viable and reliable. In order to do so, events from the Z decay to electrons can be used, where the electrons are reconstructed as photons. The extent to which the $Z \rightarrow e^+e^-$ events are representative of the signal processes is limited by the intrinsic differences between the Z boson and the Higgs boson: the kinematics of the decay products is different, due to the different spin of the two bosons; the natural width of the Z boson (2.5 GeV) is comparable to the experimental invariant mass resolution, whereas the Higgs boson natural width (4 MeV) is entirely negligible in this context, resulting in the Z boson being less sensitive to imperfect modeling of the mass resolution estimators.

The comparison between data and simulation for $Z \rightarrow e^+e^-$ events with a $p_T > 40$ GeV is shown in Figure 5.11 for the BDT input variables and in Figure 5.12 for the BDT score, respectively. The electrons are reconstructed and selected as photons, except for the electron-veto condition that is inverted. The agreement between data and simulation is found to be generally good.

Systematic uncertainties are assigned, to correctly account for discrepancies in the inputs, to the simulation of the photon identification BDT and the per-photon energy esti-

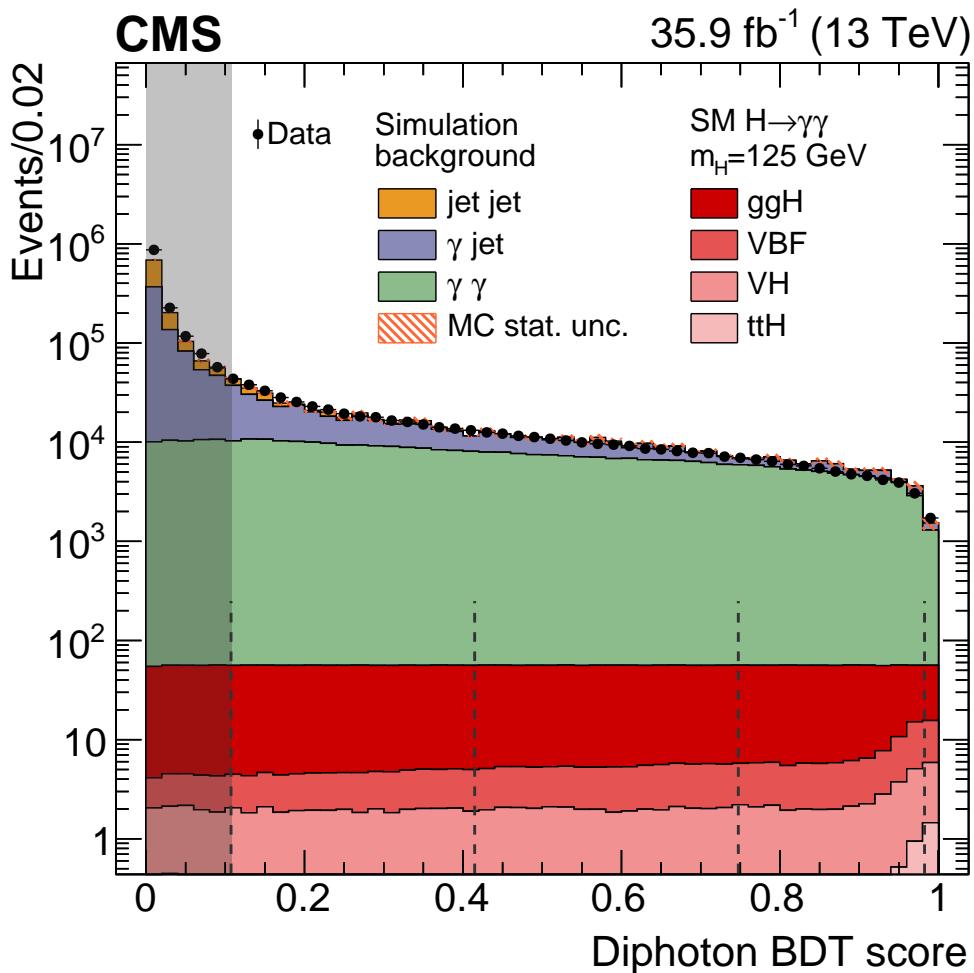


Figure 5.10: The score of the diphoton BDT for events with two photons in data (black points) and simulated signal (red shades) and background (histograms) events. The distribution has been transformed to have a flat distribution between 0 and 1 for the sum of the simulated signal processes. The vertical dashed lines represent the boundaries of the categories defined for the analysis, described in Section 5.9, and the grey shaded area indicates the events discarded by the categorization.

mators, with the latter being propagated to the invariant mass resolution estimators. The impact of these systematic uncertainties on the BDT output, shown by the red-shaded band in Figure 5.12, is found to adequately cover the data-simulation mis-modeling.

5.9 EVENT CATEGORIES IN DIPHOTON BDT OUTPUT

The diphoton BDT score is used primarily to define categories, without targeting an individual production mode. Such categories, called for this reason *Untagged* categories, are thus mostly populated by events produced via the most abundant Higgs boson production mechanism at the LHC, i.e. ggH.

The number of Untagged categories and the position of their boundaries is optimized using an iterative procedure, described below. For an n -category analysis, n boundaries are initially placed at equal distances in signal quantiles from each other, given the two fixed boundaries at the edges of the spectrum -1 and 1 . By using n boundaries, $n+1$ categories are defined, out of which only n are selected for the analysis and for the calculation of the expected performance of the categorization, while the events falling in the lowest-BDT score category are discarded. The positions of the boundaries are free to float and for each configuration of the categories explored the p-value of a signal-plus-background fit is computed. Signal and background models, simplified with respect to the full models described in Sections 5.11 and 5.12 to improve the speed of the computation, are extracted from simulation: the background model is extracted through an exponential shape fit to the diphoton invariant mass distribution in simulated background samples; the signal model is extracted through the fit to the same variable in simulated signal samples of a sum of Gaussian functions. The median expected significance of the signal hypothesis is then extracted from a fit to an Asimov toy dataset thrown from the signal-plus-background model simultaneously in all categories. The positions of the boundaries are adjusted minimizing the combined p-value. The optimization is repeated for an increasing number of categories, from $n = 2$ up to $n = 6$.

Since the procedure involves a large number of computations of the p-value, no systematic uncertainty is taken into account at this stage, as to reduce the CPU intensity of the optimization.

The impact on the expected significance of the analysis of splitting the events into categories is shown in Figure 5.13. Since the additional gain brought by the use of 5 categories, compared to 4 categories is only $\sim 1.2\%$, a choice is made to employ 4 categories in the analysis.

The optimised positions of boundaries defining these categories are found to be -0.405 , 0.204 , 0.564 , 0.864 , expressed in terms of the BDT output variable. The categories are labeled Untagged 0, 1, 2 and 3, in decreasing order of diphoton BDT values.

The category optimization is performed on the same simulated processes used for the training of the diphoton BDT, but selecting the test sample, kept statistically independent for this purpose. The category definition is in principle dependent on the mass of the Higgs boson. However, the simulated signal samples are generated for mass hypothesis of 125 GeV, which is very close to the measured world average [1]. Furthermore, the BDT classifier has been proved to be insensitive to the mass value (see previous Section).

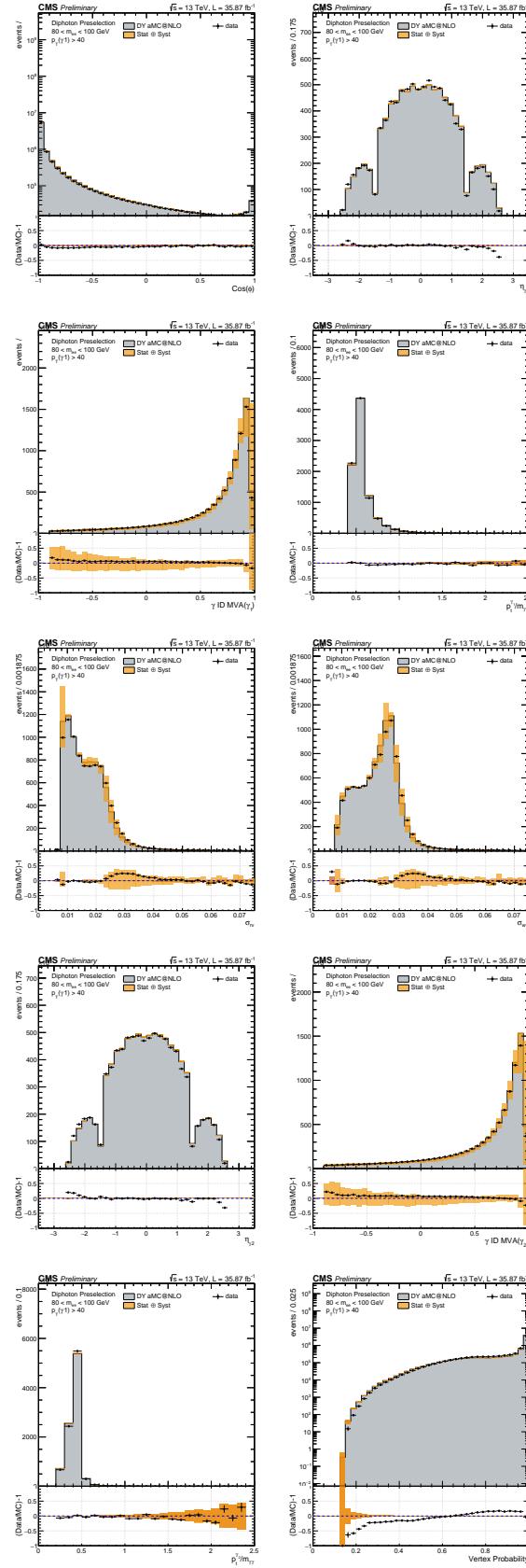


Figure 5.11: The BDT input variables (photon kinematic variables, mass resolution estimates under correct and incorrect vertex hypotheses, cosine of the angle in ϕ between the two photons, vertex probability) are compared between data (black points) and simulation (filled histogram) on electrons reconstructed as photons from Z boson decay.

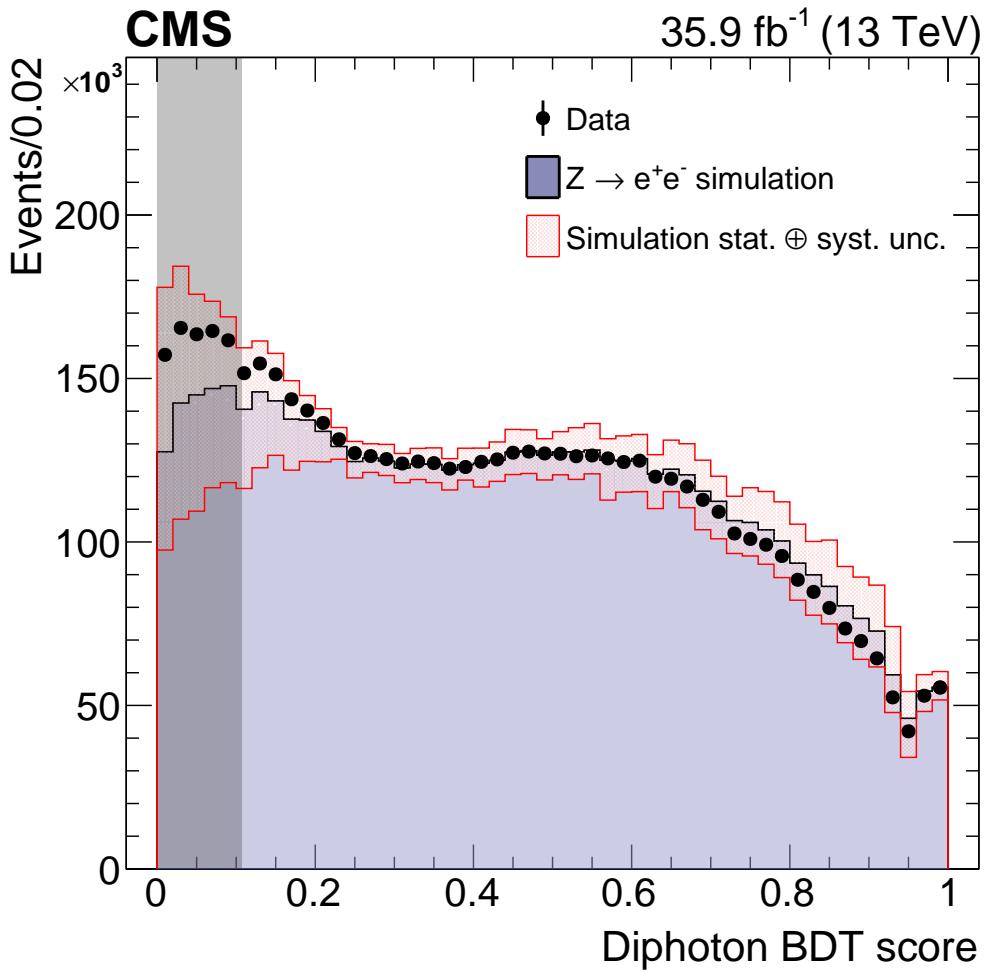


Figure 5.12: Data-simulation comparison for diphoton BDT output for electrons reconstructed as photons from the decay of the Z boson. The red band indicates the impact on the BDT output variable of the systematic uncertainties on the photon identification BDT variable and on σ_E/E .

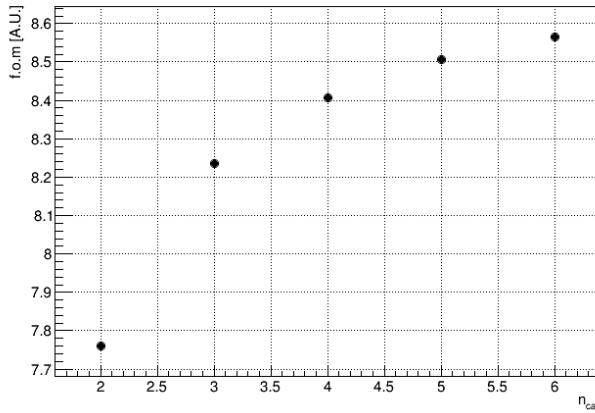


Figure 5.13: Expected significance as a function of the number of analysis categories. The estimate of the expected significance is approximated since it is extracted from a simplified signal+background fit. It can be seen how the relative gain in expected significance becomes of the order of 1% after $n = 4$.

5.10 EVENT CATEGORIES FOR INDIVIDUAL PRODUCTION MODES

The generic event categories described in Section 5.9 select events based uniquely on the characteristics of the diphoton pair, regardless of the production mechanisms from which they originate. Exclusive categories, which combine the diphoton information with topological information on the specific production mode, are defined for VBF, VH and ttH events.

5.10.1 $t\bar{t}H$ categories

Events produced via the ttH production mechanism feature a top quark-antiquark pair in association with the Higgs boson. The the top-flavored pair decays into two b quarks and two W bosons of opposite charge. Depending on the type of decay of each W boson, different event categories can be defined:

- semi-leptonic decays, where at least one W boson decays in the leptonic channel (ttH Leptonic):
 - diphoton BDT score greater than 0.11;
 - $p_T/m_{\gamma\gamma}$ greater than 1/2 (1/4) for the leading (sub-leading) photon;
 - at least one lepton, either electron or muon, as defined in Chapter 3, with a $p_T > 20$ GeV and $|\eta| < 2.4$. Electrons must satisfy a set of loose identification requirements, as described in Section 3.4, and they are rejected in the EB-EE transition region ($1.4442 < |\eta| < 1.566$). In order to reject misidentified electrons arising from $Z + \gamma \rightarrow e^+e^-\gamma$ events, the invariant mass of the electron and either of the two photons, $m_{e,\gamma}$, must be at least 5 GeV from the nominal Z boson mass. Muons must also satisfy a requirement on their relative isolation, based on the sum of the transverse energy of neutral hadrons and photons in a cone around the muon with a radius depending on the p_T of the muon;

- at least two jets with $|\eta| < 2.4$ and $p_T > 25$ GeV, and separated from each photon or lepton by a $\Delta R > 0.4$;
- at least one of the jets is identified as a b jet, according to the medium requirement of the CSV algorithm, as defined in Chapter 3;
- hadronic decays, where both W bosons decay to quarks (ttH Hadronic):
 - diphoton BDT score greater than 0.58;
 - $p_T/m_{\gamma\gamma}$ greater than 1/3 (1/4) for the leading (sub-leading) photon;
 - no leptons, defined as for the Leptonic category;
 - at least three jets in $|\eta| < 2.4$ with $p_T > 25$ GeV;
 - at least one of the jets is identified as a b jet, according to the loose requirement of the CSV algorithm, as defined in Chapter 3;
 - a selection on a BDT discriminant designed to enhance the tagging of ttH multi-jets events, which takes as input the number of jets with $p_T > 25$ GeV, the leading jet p_T , the two highest scores of the b-tagging CSV algorithm.

5.10.2 VH categories

The associated production of the Higgs boson with a Z or W boson leads to the presence of leptons, jets, and missing transverse momentum in the event. Different categories are defined according to the decay mode of the vector boson:

- leptonic Z decays (ZH Leptonic):
 - diphoton BDT score greater than 0.11;
 - $p_T/m_{\gamma\gamma}$ greater than 3/8 (1/4) for the leading (sub-leading) photon;
 - two same-flavor leptons, defined as for the ttH Leptonic category, with a combined invariant mass between 70 and 110 GeV;
 - electrons associated with a reconstructed conversion of a photon are vetoed, by requiring that, if the electron and the photon candidates share an ECAL supercluster, the angular separation between the center of the supercluster and the electron track is larger than 0.4;
- leptonic W decays (WH Leptonic):
 - diphoton BDT score greater than 0.28;
 - $p_T/m_{\gamma\gamma}$ greater than 3/8 (1/4) for the leading (sub-leading) photon;
 - at least one lepton, defined as for the ttH Leptonic category;
 - missing transverse momentum larger than 45 GeV;
 - at most two jets in $|\eta| < 2.4$, with $p_T > 20$ GeV, and with an angular separation from each photon or lepton larger than 0.4;
- loose selection of both W and Z decays (VH LeptonLoose), defined as the WH Leptonic category, except for the inverted requirement on the missing transverse momentum (< 45 GeV);

- leptonic decays of Z or W, with at least one missing lepton (VH MET)
 - diphoton BDT score greater than 0.79;
 - $p_T/m_{\gamma\gamma}$ greater than 3/8 (1/4) for the leading (sub-leading) photon;
 - missing transverse momentum larger than 85 GeV and with a separation in the ϕ angle from the diphoton system larger than 2.4;
- hadronic decays of Z or W (VH Hadronic)
 - diphoton BDT score greater than 0.79;
 - $p_T/m_{\gamma\gamma}$ greater than 3/8 (1/4) for the leading (sub-leading) photon;
 - at least two jets in $|\eta| < 2.4$ with $p_T > 40$ GeV, with a combined invariant mass compatible with the Z mass, i.e. between 60 GeV and 120 GeV;
 - $|\cos(\theta^*)| < 0.5$, where θ^* is the photon helicity angle in the Collins-Soper frame [32], to reduce ggH contamination.

5.10.3 VBF categories

The VBF topology is characterized by the presence of two jets with a large pseudorapidity gap, originating from the quarks radiating the vector bosons. The tagging of events containing jets with kinematical features compatible with those of the VBF production mode is achieved via a multivariate discriminant (di-jet BDT). The di-jet BDT is trained for VBF signal events against a background of ggH production with additional jets.

The di-jet BDT is further combined with the diphoton BDT and the diphoton transverse momentum, rescaled by the invariant mass, $p_T^{\gamma\gamma}/m_{\gamma\gamma}$, into an additional BDT classifier (VBF combined BDT). The VBF combined BDT is used to extract and optimize three categories, maximising the median expected significance, with a procedure similar to the one described in Section 5.9.

The selection of candidate VBF events is summarized as follows:

- $p_T/m_{\gamma\gamma}$ greater than 1/3 (1/4) for the leading (sub-leading) photon;
- photon identification BDT output larger than -0.2 , to further reject contamination from prompt-fake background events;
- two jets in $|\eta| < 4.7$, with $p_T > 30$ GeV (20 GeV) for the leading (sub-leading) one, and with a combined invariant mass larger than 250 GeV.

5.10.4 Categorization order

All the events fulfilling the preselection requirements and with $100 \text{ GeV} \leq m_{\gamma\gamma} \leq 180 \text{ GeV}$ are considered for the analysis and assigned to one event category.

A priority order is defined among the categories, with higher priority assigned to categories targeting individual production modes and to categories with low expected signal yields. Events fulfilling the selection requirements of multiple categories are therefore assigned only to the highest-ranking category, making the categories exclusive. The order chosen is the following: ttH Leptonic, ttH Hadronic, ZH Leptonic, WH Leptonic, VH

LeptonicLoose, VBF categories, VH MET, VH Hadronic, and Untagged. The VBF and Untagged categories are internally ranked from higher to lower values of their respective BDT discriminators.

5.11 SIGNAL MODEL

The shape of the signal peak in the diphoton invariant mass distribution is constructed from events simulating the four main production modes. The shape is modeled using the sum of up to 5 Gaussians as a functional form.

The signal model is constructed separately for each category and for each process. Such segmentation of the signal modeling allows the separate extraction of the corresponding components of the signal yields, laying the grounds for the measurements of per-process and per-category signal strengths.

Furthermore, the additional angular contribution to the invariant mass resolution arising from an incorrect vertex identification, as described in Section 5.4, leads to a significantly broader mass peak shape. For this reason, the distributions for events with a correct vertex assignment within 1 cm and the remainder of the events are modeled separately.

Each component of the signal model is built under different mass hypotheses for the Higgs boson, comprised between 120 and 130 GeV, using the same functional form. The parameters of the fit are then interpolated between the mass values, producing a continuous signal model as a function of m_H . In this way, the mass of the boson can be floated in the signal extraction fit.

The simulated signal events take into account differences between data and simulation in trigger, identification and selection efficiencies, as measured with data-driven techniques and described in Section 3.8. Photon shower shapes are tuned in order to take into account data-simulation discrepancy, as described in Section 5.7. Furthermore, events are weighted to take into account the different size of the luminous region in simulated and real conditions.

Each signal shape component is normalized to its expected signal yield and summed to obtain the final signal model in each category.

An example of the signal shape as a function of m_H for the Untagged 1 category is shown in Figure 5.14, left. The same model, evaluated at $m_H = 125$ GeV is shown in Figure 5.14, right. Two estimates of the width of the signal peak are also displayed on the figure: the σ_{eff} value, corresponding to half of the interval covering 68.3% of the integral of the distribution around its maximum (grey area); the full width at half maximum (FWHM) (double headed arrow).

5.12 BACKGROUND MODEL

The background model for the $m_{\gamma\gamma}$ distribution is extracted entirely from data and employs the discrete profiling method [36]. Such method was developed to include the choice of the background functional form as a discrete nuisance parameter in the likelihood used to fit the data. This provides a way to estimate the systematic uncertainty associated with the choice of a particular functional form directly from the fit, thus including it in the statistical component of the uncertainty.

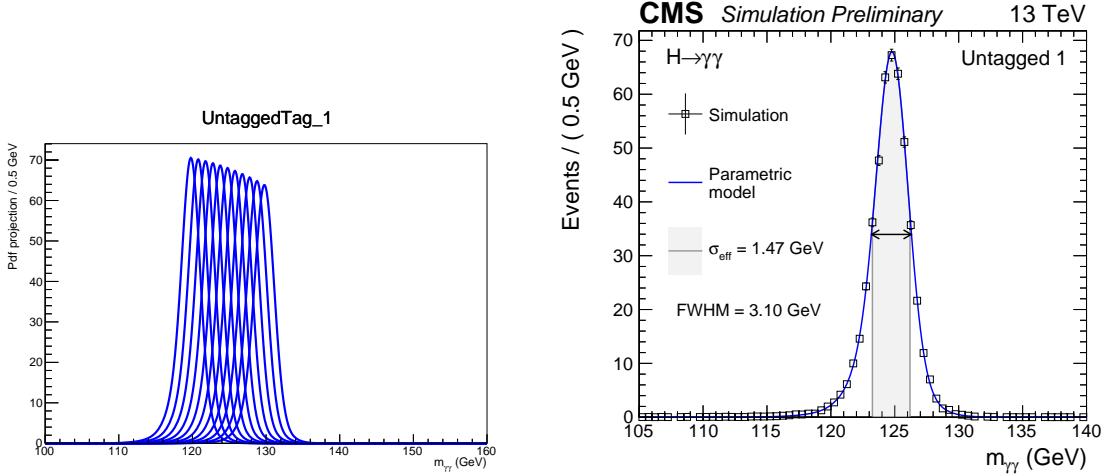


Figure 5.14: Signal shape for the Untagged 1 category as a function of m_H (left) and evaluated at $m_H = 125$ GeV (right). The points indicated the simulated signal events and the blue lines the fitted models.

Four families of candidate functions, able to qualitatively describe the falling diphoton invariant mass distribution, are considered: exponentials, Bernstein polynomials, Laurent polynomials, and power law functions.

When a signal-plus-background hypothesis is fitted to the data, the value of twice the negative logarithm of the likelihood is minimized. In this process, through the dedicated discrete nuisance parameter, all functions of the four families are tested as possible analytic descriptions of the background shape. In order to reduce the initial set of functions to a finite one, only functions satisfying a threshold in goodness of fit are considered. Furthermore, the highest order considered is determined by increasing the number of free parameters until the marginal improvement brought by using a function of the $(N+1)$ -th order, instead of the N -th order, falls below a predetermined threshold. Such improvement is quantified by the difference in χ^2 of the two fits. In order to account for the number of free parameters in the fit, N_B , this quantity is added as a *penalty term* to the minimized likelihood function. The minimized quantity, for each given candidate function B , is thus the *penalized* likelihood $\tilde{\mathcal{L}}_B$, defined as:

$$-2 \ln \tilde{\mathcal{L}}_B = -2 \ln \mathcal{L}_B + N_B, \quad (5.4)$$

where \mathcal{L}_B is the *unpenalized* likelihood function.

5.13 SYSTEMATIC UNCERTAINTIES

Several sources of systematic uncertainty are taken into account in the extraction of the results of this analysis. They can be grouped in two classes, depending on whether their origin has a theoretical or an experimental nature. The sources of systematic uncertainties taken into account with a theoretical origin are the following:

- QCD parameters, including: uncertainty on the renormalization and factorization scales; PDF uncertainties, included following the PDF4LHC prescription [14, 72] for

normalization and the MC2HESSIAN procedure [16] for migrations; uncertainty in the strong coupling constant, α_s ;

- uncertainty on the $H \rightarrow \gamma\gamma$ branching fraction, assessed to be about 2% [72];
- change in the UE and parton shower description, estimated using simulated samples with different tuning of the generators;
- ggH contaminations in categories with large Higgs boson p_T and high jet-multiplicity, given that the ggH samples include the simulation of the emission of two additional jets, limiting the reliability of the ggH prediction in this regime. In particular, for the ttH tagged categories, are included: a 10% uncertainty associated with the size of the simulated samples; additional uncertainty in the parton shower modeling at high jet multiplicity, estimated at 35% for jet-multiplicity larger than 5; uncertainty associated with the gluon splitting modeling, regulating the contamination from ggH events with real b jets, estimated using the discrepancy between data and simulation in the ratio $\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ at 13 TeV [90]. For the VBF categories, the following uncertainties are taken into account: uncertainty in jet multiplicity calculation, based on the STWZ [94] and BLPTW [73, 81, 94] prescriptions; uncertainties in the modeling of the Higgs boson p_T , resulting in 6–8% change in the ggH contamination; missing higher-order QCD terms in the theoretical calculations, affecting the migration of ggH events into VBF categories, estimated using an extension of the Stewart–Tackmann method [48, 93].

The experimental sources of systematic uncertainty included in the analysis are:

- photon energy scale and resolution, determined with $Z \rightarrow e^+e^-$, in particular:
 - uncertainties in the determination of the photon energy scale and resolution corrections. These uncertainties are assessed by varying the R_9 distribution, the energy regression settings and the selection of electrons used to derive the corrections. The propagation of these uncertainties, in bins of pseudorapidity and R_9 , leads to variations of the photon energy from 0.15% to 0.5%, depending on the photon category;
 - non-linear effects in the energy scale extrapolation between the Z boson mass and the H boson mass, evaluated using electrons from the decay of boosted Z bosons and found to have an impact of 0.1% of the photon energy;
 - uncertainty in the modeling of the fraction of scintillation light reaching the photo-detectors, accounting for the variation of transparency of the ECAL crystals, affecting the photon energy by 0.07%;
 - uncertainties in the electromagnetic shower simulation in GEANT4, evaluated to be 0.05% of the photon energy.
 - imperfect determination of the amount of material upstream the ECAL, assessed using dedicated simulated samples with variations of the material budget and found to affect the energy scale at most by 0.24%;
 - uncertainty in the modeling of the shower shape in simulation, estimated using samples with and without the corrections described in Section 5.7, with an impact between 0.01% and 0.15% on the energy scale, depending on the photon category.

- uncertainties in efficiencies of trigger selection and preselection, assessed in $Z \rightarrow e^+e^-$ events, and varying between 0.1% and 0.5% variation of the event yield, depending on the category;
- uncertainties in the photon identification BDT output and per-photon energy resolution estimator, assigned to cover the data-simulation discrepancies in $Z \rightarrow e^+e^-$ and propagated through the diphoton BDT;
- the determination of energy scale and resolution corrections for jets, whose uncertainty is responsible for event migrations between categories of the order of 10%;
- uncertainty in the missing transverse momentum, evaluated by varying within its uncertainty the p_T of the particle candidates, each according to its type, entering the calculation of p_T^{miss} ;
- the discrepancy between the identification of pileup jets in data and simulation, in events with a Z boson and one jet;
- the discrepancy between data and simulation in the identification efficiencies for electrons and muons, estimated with the tag-and-probe technique in Z events;
- uncertainty on the CSV b tagging discriminant, assessed comparing data and simulation;
- an uncertainty of 2% is assigned to the vertex selection efficiency, arising from the modeling of the UE and from the ratio of efficiencies between data and simulation in $Z \rightarrow \mu^+\mu^-$ events;
- the uncertainty in the integrated luminosity measurement for the data set analyzed, assessed to be 2.5% [28].

As described in Section 5.12, the discrete profiling method includes the systematic uncertainty associated to the background modeling in the statistical uncertainty of the likelihood fit. It thus does not appear among the sources of systematic uncertainties listed in this Section.

The systematic uncertainties with the largest impact on the results are the modeling of the photon ID BDT and per-photon energy estimator, the determination of the energy scale and resolution corrections, the jet energy scale correction, and the integrated luminosity.

The total impact of all the sources of systematic uncertainty is of a few percent on the inclusive signal strength measurement. For this measurement, even though characterized by the smallest statistical uncertainty among those in this Chapter, the systematic component of the uncertainty is found to be largely subdominant. A summary of the impact of the different sources of systematic uncertainties is given in Fig. 5.15.

5.14 RESULTS

The analysis presented in this Chapter provides the measurement of the properties of the Higgs boson in its diphoton decay channel. The signal strength modifier, μ , defined as the ratio between the observed signal yield and the corresponding SM expectation, is measured inclusively and individually for each production mode.

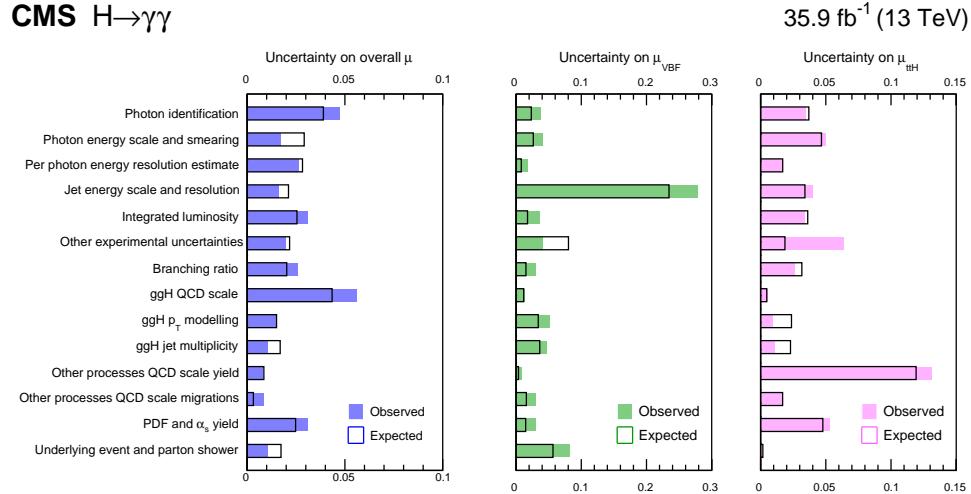


Figure 5.15: The different sources of systematic uncertainties and their impact on the signal strengths, inclusively and for the VBF and $t\bar{t}H$ production mechanisms. The solid (empty) bars show the observed (expected) results.

The expected number of signal events in each category is shown in Figure 5.16, where the contributions of individual production mechanisms are indicated in different colors. The figure also lists, for each category, two estimates of the width of the signal peak, σ_{eff} and the FWHM, as defined in Section 5.11, and the ratio between the number of signal events and the sum of signal and background events within one σ_{eff} around the peak.

The results are extracted through a maximum likelihood fit, simultaneously in all categories, to the diphoton invariant mass distribution. In addition to the signal strength modifier, the mass position of the signal peak is free to float in the fit.

The result of the signal-plus-background fit and the data are reported for the 4 Un-tagged categories in Figure 5.17. The same information is shown for the tagged categories in Figures 5.18 and 5.19 and summed over all the categories in Figure 5.20.

The total uncertainties are extracted by performing a likelihood scan around its minimum value, applying an asymptotic approach [11, 63, 83] using a test statistic based on the profile likelihood ratio ($q(\mu) = -2 \ln(\lambda(\mu))$, where μ is the signal strength modifier and λ the profile likelihood ratio function) [46]. The contribution of statistical uncertainties is evaluated by carrying out a likelihood scan where the nuisances associated with the different sources of systematic uncertainty are not allowed to float. The systematic uncertainty is then computed as the difference in quadrature between the total and the statistical uncertainties. The results of these scans for the measurement of the inclusive μ are shown in Figure 5.21.

The inclusive signal strength modifier is measured to be

$$\hat{\mu} = 1.18^{+0.17}_{-0.14} = 1.18^{+0.12}_{-0.11} (\text{stat})^{+0.09}_{-0.07} (\text{syst})^{+0.07}_{-0.06} (\text{theo}) \quad (5.5)$$

with a best fit mass of

$$\hat{m}_H = 125.4 \text{ GeV} \pm 0.3 \text{ GeV} = 125.4 \text{ GeV} \pm 0.2 \text{ GeV} (\text{stat}) \pm 0.2 \text{ GeV} (\text{syst}) \quad (5.6)$$

compatible with the world average of 125.09 GeV [1].

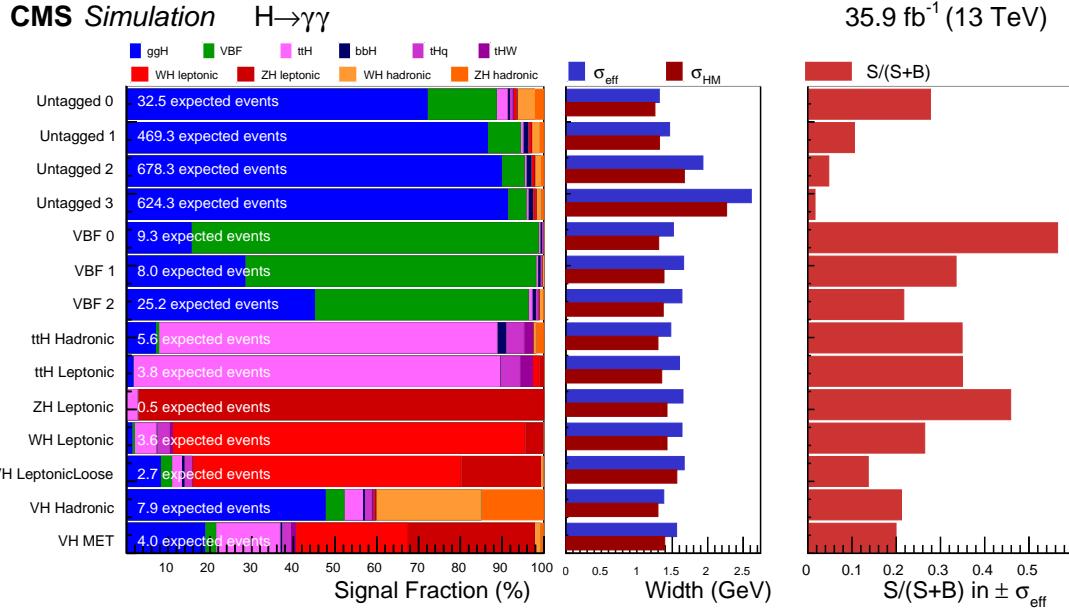


Figure 5.16: Number of expected signal events in each category, with the breakdown of the relative contributions of individual production mechanisms.

The measurement of per-production mode signal strength modifiers is summarized in Figure 5.22. The observed (expected) values for the VBF, VH, and ttH processes correspond to p -values of 4.2% (1.8%), 0.47% (12%), and 0.074% (7.3%), respectively.

The signal strength modifiers can also be grouped and defined for fermionic production modes (ggH and ttH), $\mu_{\text{ggH},\bar{\text{t}}\text{tH}}$, and bosonic ones (VBF, VH), $\mu_{\text{VBF},\text{VH}}$. The measurement of these two quantities is extracted through a two-dimensional scan of the likelihood, shown in Figure 5.23. The best fit values are $\mu_{\text{ggH},\bar{\text{t}}\text{tH}} = 1.19^{+0.22}_{-0.18}$ and $\mu_{\text{VBF},\text{VH}} = 1.21^{+0.58}_{-0.51}$.

The κ -framework [71] can be used to parametrize distortions of the SM predictions by using coupling modifiers, indicated with κ , in the calculation of the scattering amplitudes. The two-dimensional likelihood scan for the measurement of the Higgs boson coupling modifiers to fermions, κ_f , and to bosons, κ_V , is shown in Figure 5.24, left. Similarly, the measurement of the modifiers of effective coupling to photons, κ_γ , and to gluons, κ_g , is reported in Figure 5.24, right. The point $(\kappa_V, \kappa_f) = (1, -1)$ is inconsistent with the observed (expected) best fit value at 5.8 (7.0) standard deviations.

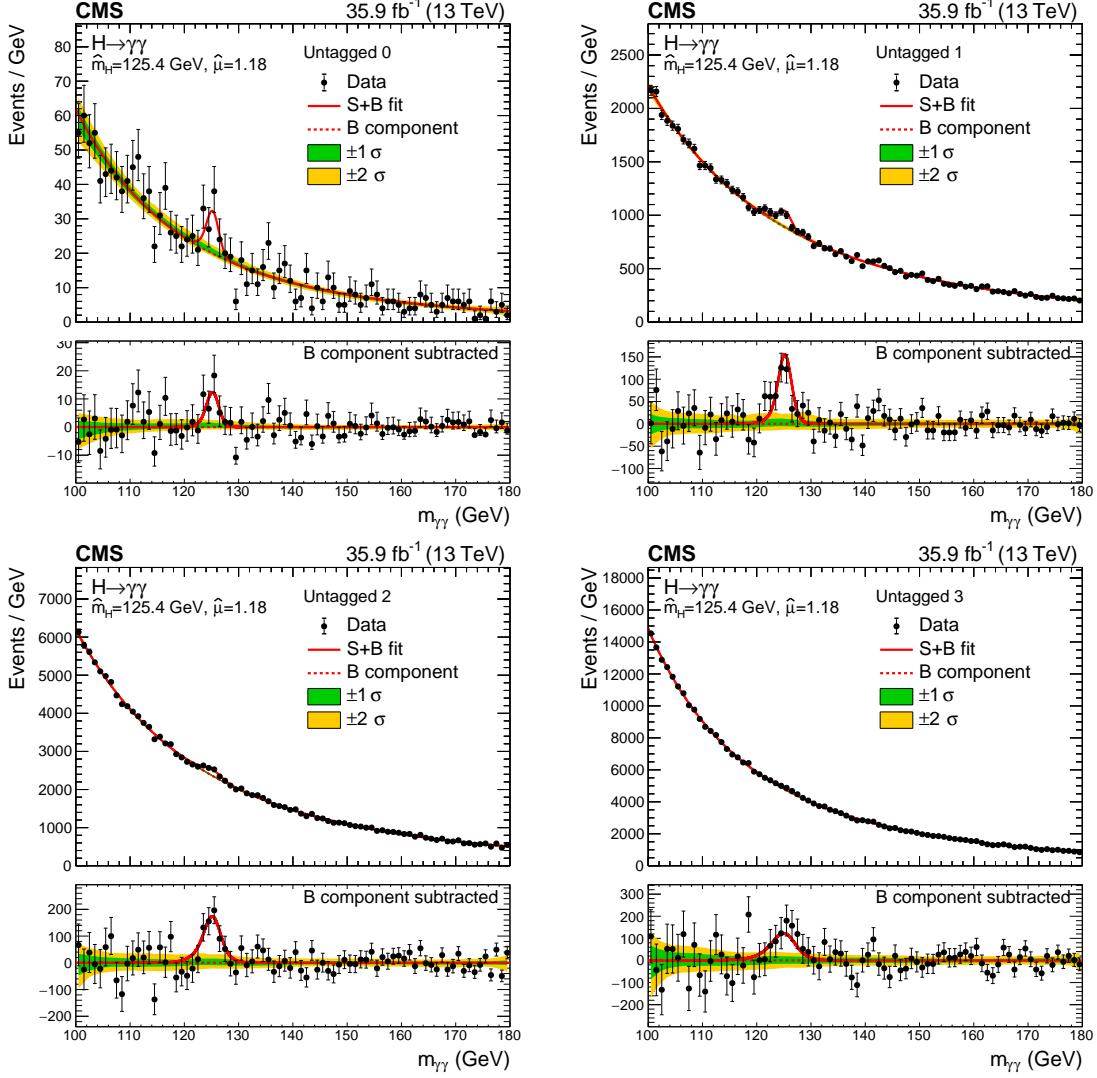


Figure 5.17: The signal-plus-background model fits and data for the four Untagged categories. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit.

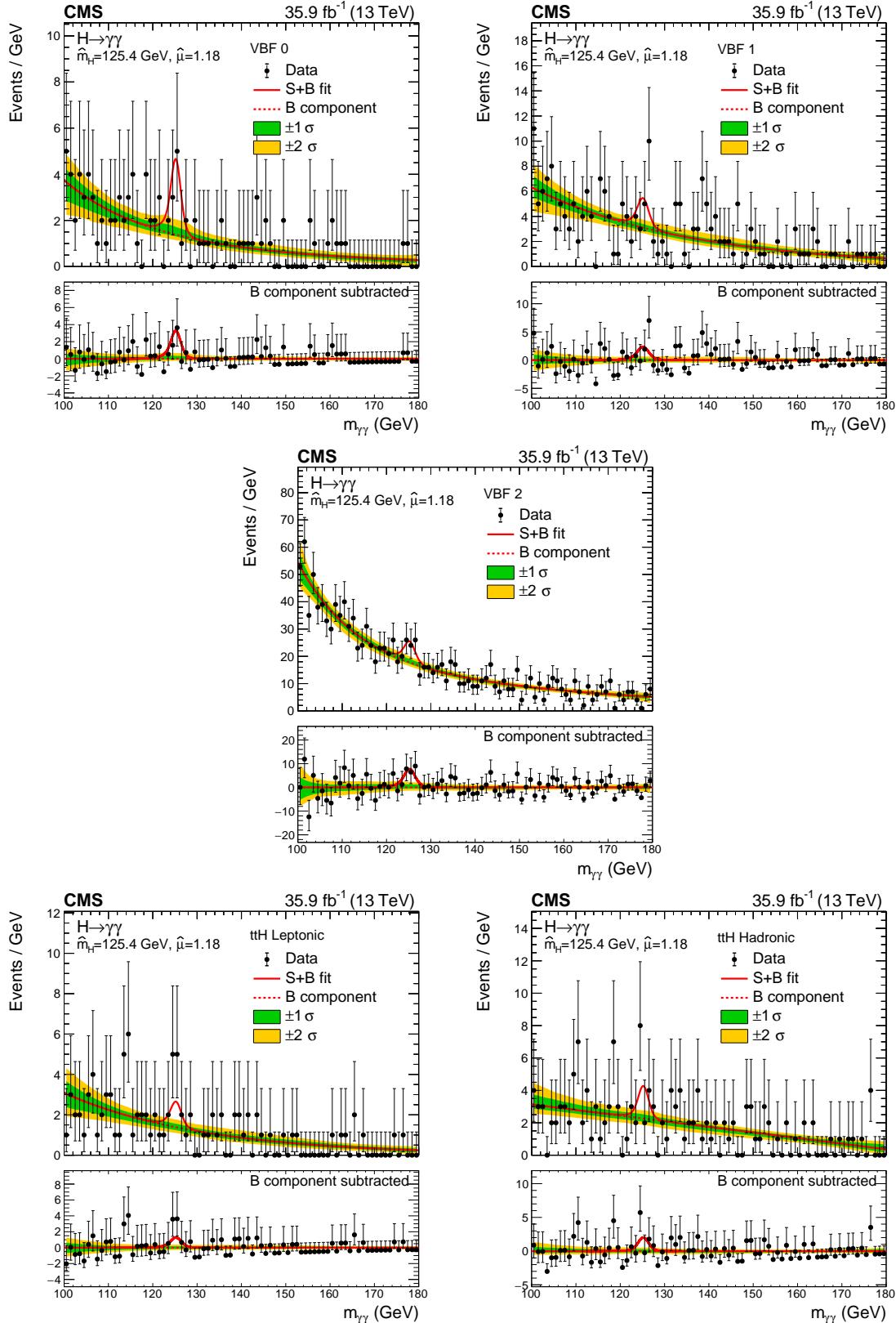


Figure 5.18: The signal-plus-background model fits and data for the VBF and ttH categories. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit.

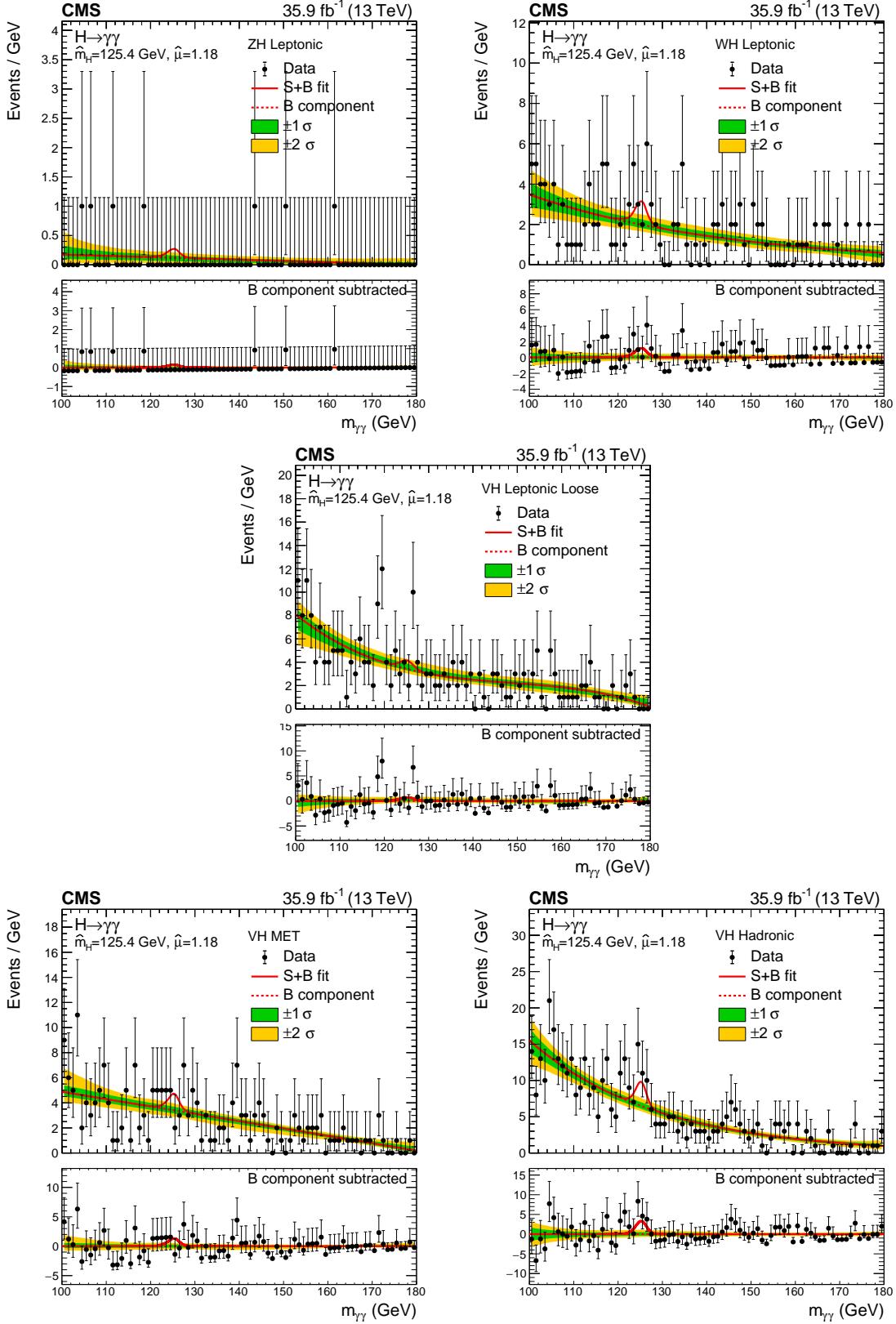


Figure 5.19: The signal-plus-background model fits and data for the VH categories. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit.

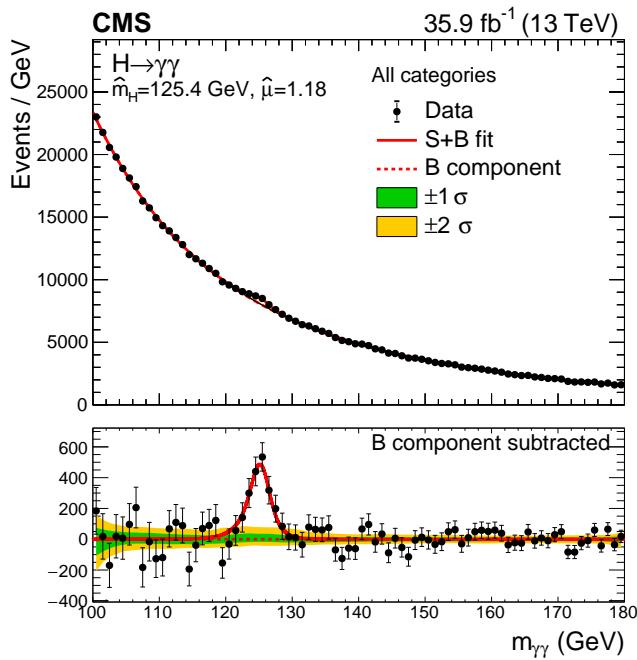


Figure 5.20: The signal-plus-background model fits and data, summer over all the categories of the analysis. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit.

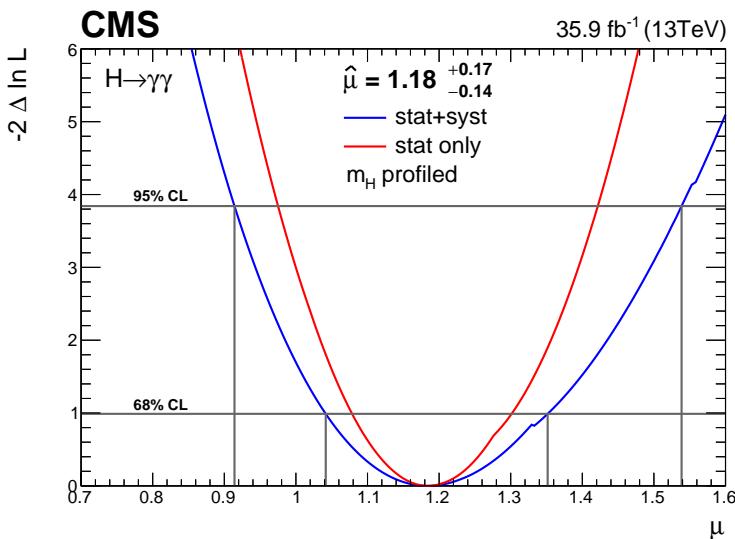


Figure 5.21: The likelihood scan for the signal strength where the value of the SM Higgs boson mass is profiled in the fit.

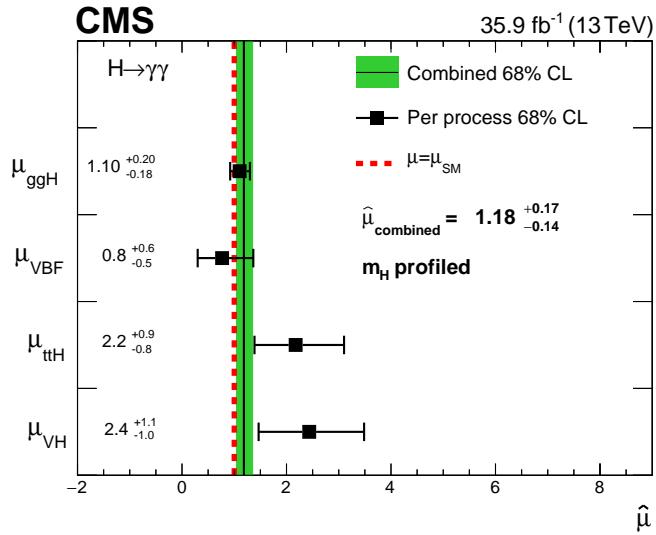


Figure 5.22: Signal strength modifiers measured for each process (black points), with the SM Higgs boson mass profiled, compared to the overall signal strength (green band) and to the SM expectation (dashed red line).

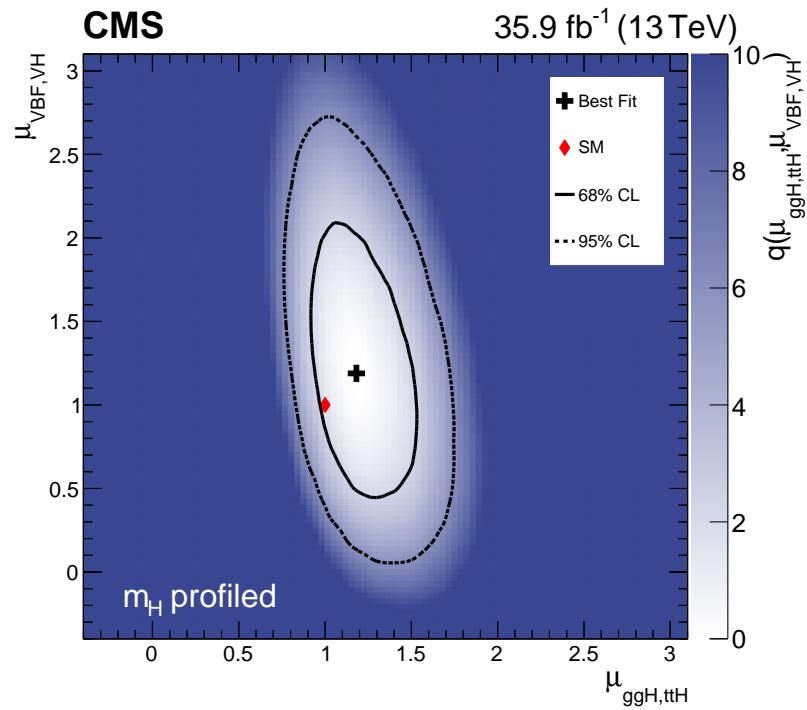


Figure 5.23: The best fit value (black cross) of the two-dimensional likelihood scan for $\mu_{ggH,ttH}$ versus $\mu_{VBF,VH}$ is compared to the SM expectation (red diamond). The solid (dashed) contour shows the 68% (95%) confidence region.

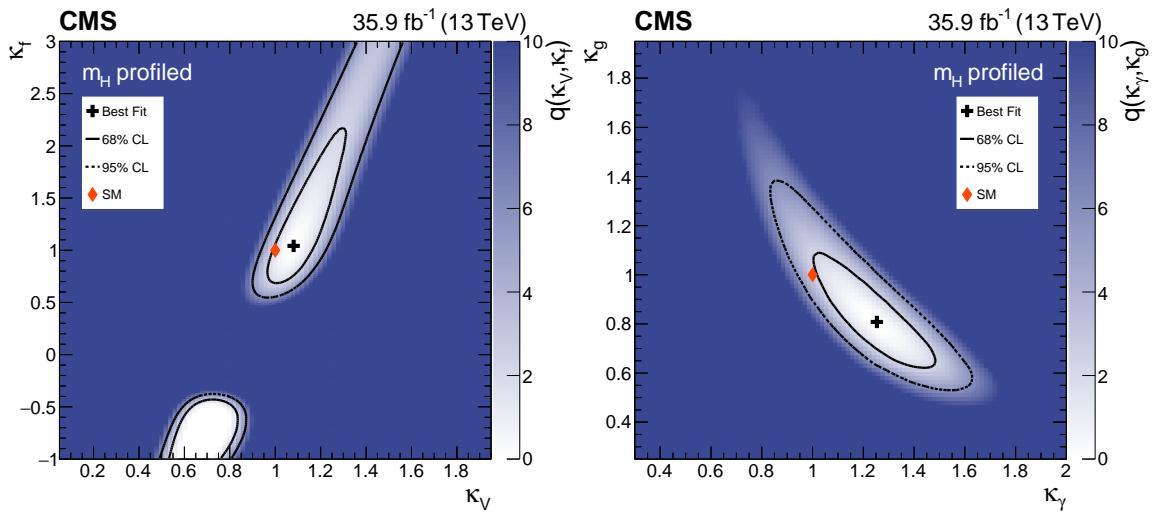


Figure 5.24: The best fit values (black crosses) of the two-dimensional likelihood scan for κ_f versus κ_V , on the left, and κ_g versus κ_γ , on the right, are compared to their respective SM predictions (red diamonds). The solid (dashed) contour shows the 68% (95%) confidence region.

Part III

FIDUCIAL AND DIFFERENTIAL MEASUREMENTS OF THE HIGGS BOSON PRODUCTION AT THE LHC

This Part reports the latest measurements of fiducial and differential cross sections for the production of the Higgs boson decaying to two photons, performed on pp collision data at $\sqrt{s} = 13$ TeV collected with the CMS experiment.

6

INTRODUCTION

This Chapter presents broadened and more extensive studies of the diphoton decay channel of the Higgs boson, by introducing the concept of differential and fiducial production cross sections. After a review of the specific tools developed for this kind of measurements, the results are presented and discussed.

6.1 INTRODUCTION

The measurements of the inclusive production rates and couplings presented in Chapter 5 employ a highly sophisticated categorization scheme and the full information provided by the simulation of events to optimize the performance of the analysis. While achieving the best statistical precision, this approach leads to a strong model-dependence of the results. A complementary approach consists in performing *fiducial* cross section measurements: in fact, the measurement of cross sections, in place of signal strength modifiers, allows to decouple the measurement from the underlying theory predictions. In particular, cross section measurements remain valid regardless of new theoretical calculations, with which they can be directly compared. The extraction of fiducial measurements entails the definition of a fiducial volume of the phase space to which the analysis is restricted. This volume typically corresponds to the most performant region of the detector and, at the same time, provides a fixed reference for theoretical predictions, which can be calculated after the measurement has been performed. Furthermore, the definition of a fiducial phase space allows to avoid uncertainties on the measurements due to extrapolation to the full phase space, which would rely on current theoretical prediction and would constitute an additional source of model-dependence.

The model-dependence of the measurements is further reduced, compared to the *coupling* analysis, by avoiding or minimizing the use of multivariate techniques, in favor of simpler approaches. The measurement of signal strength modifiers, in fact, relies on highly sophisticated tools with complex selection acceptances, preventing a straightforward interpretation as cross section measurements. Moreover, MVA approaches for event categorization, as that described in Section 5.8, are optimized relying on state-of-the-art theoretical predictions and simulations, causing the measurements to be strongly model-dependent and preventing their re-interpretation as cross sections while the calculation of predictions and simulation techniques advance. For this reason, a simpler categorization scheme is studied and introduced for the fiducial measurements described in this Chapter. This approach is found to provide largely model-independent results, at a limited cost in terms of statistical precision. This cost becomes less and less limiting with the larger and larger amount of data collected by the CMS and ATLAS experiments during the LHC Run2, making the measurement of fiducial cross sections a key element in the transition from the discovery of the Higgs boson to the era of precision measurements.

Furthermore, the measurement of *differential* distributions allows a more thorough test of the nature of the Higgs boson and its compatibility with the SM expectations. The

production cross section can be measured as a function of observables describing the kinematics of the diphoton system, as well as of the additional particles produced in association with it. The spectrum of different observables is sensitive to different theoretical aspects of the prediction computations, as it will be illustrated in Section 6.6. At the same time, the differential cross section measurements offer a powerful probe for deviations from the SM predictions.

The analysis presented in this Chapter shares a subset of its tools with the *coupling* $H \rightarrow \gamma\gamma$ analysis, described in Chapter 5. The Chapter will thus focus on the illustration of the tools and techniques specific to this analysis, before presenting its results. The results are as well described in a publication [26] which, at the time of writing this thesis, has been submitted to the Journal of High Energy Physics. The structure, discussions and illustrations of this publication are closely followed in the present Chapter.

6.2 MASS RESOLUTION ESTIMATOR

The categorization of the selected pair of photons relies on a relative invariant mass resolution estimator. The invariant mass of the pair, $m_{\gamma\gamma}$, is calculated from the energies of the individual photons, E_1 and E_2 , and their opening angle, α as follows:

$$m_{\gamma\gamma} = \sqrt{2E_1 E_2 (1 - \cos \alpha)} \quad (6.1)$$

When neglecting the angular contribution, the mass resolution can be computed by propagation of the resolution estimators of the photon energies, obtaining the expression already given in Equation 5.1. In the energy range of the photons selected in this analysis, the resolution of the energy measurement in the ECAL depends on the energy of the photon itself, via the stochastic and noise terms, introduced in Equation 2.2, as opposed to high-energy photons whose resolution is dominated by the constant term. Such a dependence is propagated to the mass resolution estimator described in Equation 5.1, which thus depends on the invariant diphoton mass itself.

The correlation between the plain mass resolution estimator and the mass makes it unsuitable for direct categorization, which would result in distortions of the diphoton invariant mass spectrum. In particular, since events with a higher invariant mass tend to have better mass resolution, the categories collecting photon pairs with good resolution would perform a biased selection that would tend to exclude events at the lower end of the mass window considered for the analysis. This would undermine the assumption of a smoothly falling background distribution, on which the background modeling described in Section 5.12 relies.

This effect can be avoided by removing such a correlation and building a decorrelated relative invariant mass resolution estimator, indicated with $\sigma_m/m|_{\text{decorr}}$. The decorrelation procedure exploits the property that, given a continuous probability distribution function $f(x)$, its Cumulative Distribution Function (CDF), defined as $y(x) = \int_{x_{\min}}^x f(x')dx'$, maps the random variable x into a uniformly distributed one y . This property can be applied to build a uniformly distributed mass resolution estimator in each bin of $m_{\gamma\gamma}$, by defining a mass dependent transformation

$$y(\sigma_{rv}/m_{\gamma\gamma}|m_{\gamma\gamma}) = \int_0^{\sigma_{rv}/m_{\gamma\gamma}} f((\sigma_{rv}/m_{\gamma\gamma})'|m_{\gamma\gamma}) d(\sigma_{rv}/m_{\gamma\gamma})' \quad (6.2)$$

where the conditional distribution of $\sigma_{rv}/m_{\gamma\gamma}$ on $m_{\gamma\gamma}$, $f((\sigma_{rv}/m_{\gamma\gamma})'|m_{\gamma\gamma})$, is obtained from simulation by histogramming $\sigma_{rv}/m_{\gamma\gamma}$ in bins of $m_{\gamma\gamma}$. The variable y , being uniformly

distributed in each of the bins of $m_{\gamma\gamma}$, does not depend on $m_{\gamma\gamma}$ anymore. A further transformation can be applied to y in order to obtain a quantity physically interpretable as a mass resolution. The inverse CDF at a fixed mass value serves to this purpose, where the value of $m_{\gamma\gamma} = 125$ GeV has been chosen. The final relative mass resolution estimator obtained through this procedure is denoted with the symbol $\sigma_m/m|_{\text{decorr}}$, where the subscript $\gamma\gamma$ is omitted to allow a simpler notation.

This technique is applied to background prompt-prompt simulated events and the transformation is derived separately for events where both photons are reconstructed in EB and for the other events. The effect of the transformation is verified by extracting the profile of the mass resolution estimator versus the mass, which is shown in Figures 6.1 and 6.2, for events with both photons in EB and the remainder of the sample, respectively. The negative correlation observed is entirely removed by the decorrelation transformation, which provides a flat profile in both event categories.

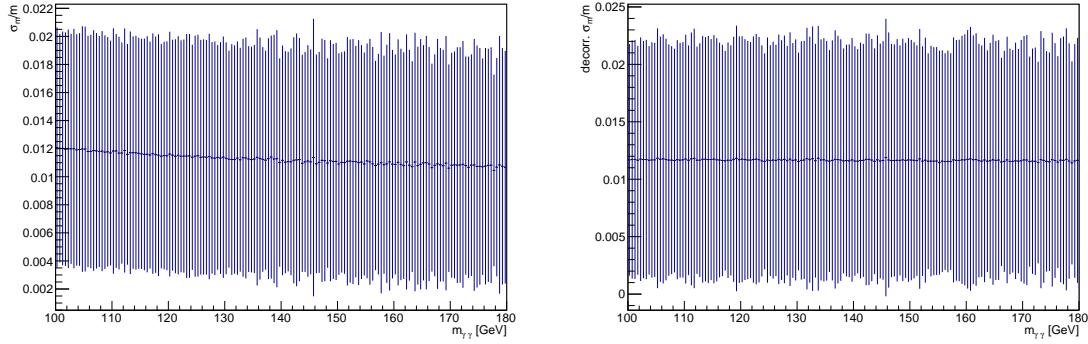


Figure 6.1: The profile correlation distribution of the (decorrelated) mass resolution estimator as a function of the di-photon mass is shown in the plot on the left (right), for events where both photons are reconstructed in EB. The points (bars) correspond to the mean (RMS) of the mass resolution estimator in a given mass bin.

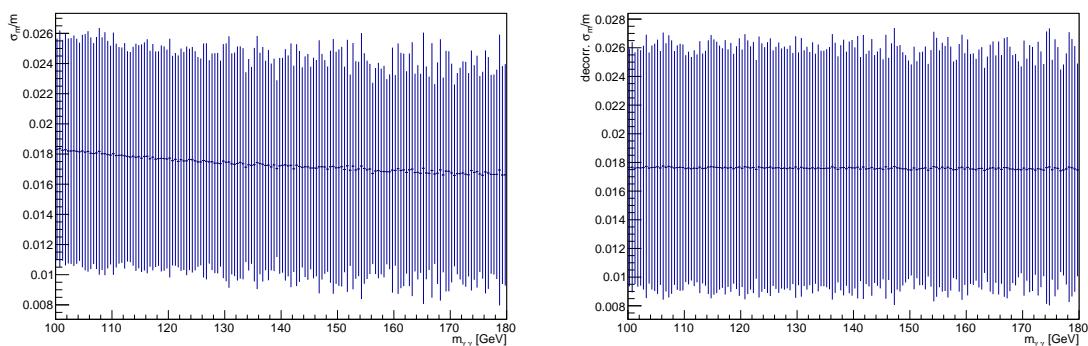


Figure 6.2: The profile correlation distribution of the (decorrelated) mass resolution estimator as a function of the di-photon mass is shown in the plot on the left (right), for events where at least one photon is reconstructed outside EB. The points (bars) correspond to the mean (RMS) of the mass resolution estimator in a given mass bin.

It is important to notice that the correlations with the pseudorapidities and the cluster shape features of the photons are respected by the transformation, preserving the power of the $\sigma_m/m|_{\text{decorr}}$ estimator to discriminate between events with high and low mass resolution. This effect is illustrated in Figure 6.3, where the distribution of $\sigma_m/m|_{\text{decorr}}$ in simulated background diphoton events is decomposed in various contributions, each corresponding to a particular selection in η and R_9 of the two photons. The multi-modal structure of the distribution shown in the Figure is the result of two effects: the intrinsic difference in energy resolution performance between the EB and the EE is responsible for the dip occurring in correspondence of $\sigma_m/m|_{\text{decorr}} \sim 0.011$, separating the distribution for events with both photons reconstructed in the EB from the distribution for events where at least one photon is reconstructed in the EE; the further splitting of these distributions is due to different smearing corrections being applied to the energy resolution of photons in different bins of (η, R_9) , as described in Section 3.4.1. These features will be exploited in the definition of event categories, as described in Section 6.4.

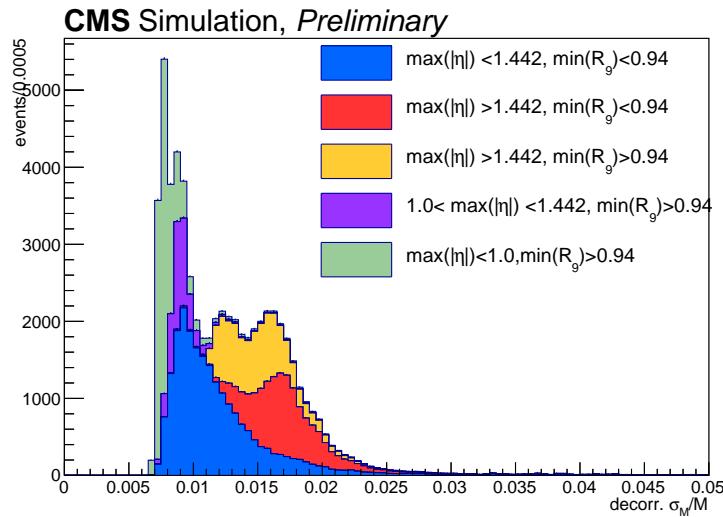


Figure 6.3: Distribution of $\sigma_m/m|_{\text{decorr}}$ in simulated diphoton background events. Each of the stacked histograms corresponds to different exclusive selections in pseudorapidities and R_9 of the two photons.

Events from $Z \rightarrow e^+e^-$ decays, where electrons are reconstructed as photons, are used to compare the modeling of $\sigma_m/m|_{\text{decorr}}$ in data and simulation on signal-like samples, following the same logic described in Section 5.8.1. The imperfect modeling of the electromagnetic shower shape variables in simulation [38] affects the level of agreement between data and simulation for the per-photon energy resolution estimator and, thus, for $\sigma_m/m|_{\text{decorr}}$. For this reason, the output of the semi-parametric regression, which uses them as inputs and which is detailed in Section 3.4.1, is recomputed after the cluster shape corrections, described in Section 5.7, are applied to simulation. The comparison of the $\sigma_m/m|_{\text{decorr}}$ estimator in data (dots) and simulation (histogram) is shown in Figure 6.4, for $Z \rightarrow e^+e^-$ events where both electrons are in EB (left) and the remaining events (right). The area depicted in gray corresponds to the region of events discarded by the categorization described in Section 6.4. The systematic uncertainty assigned to

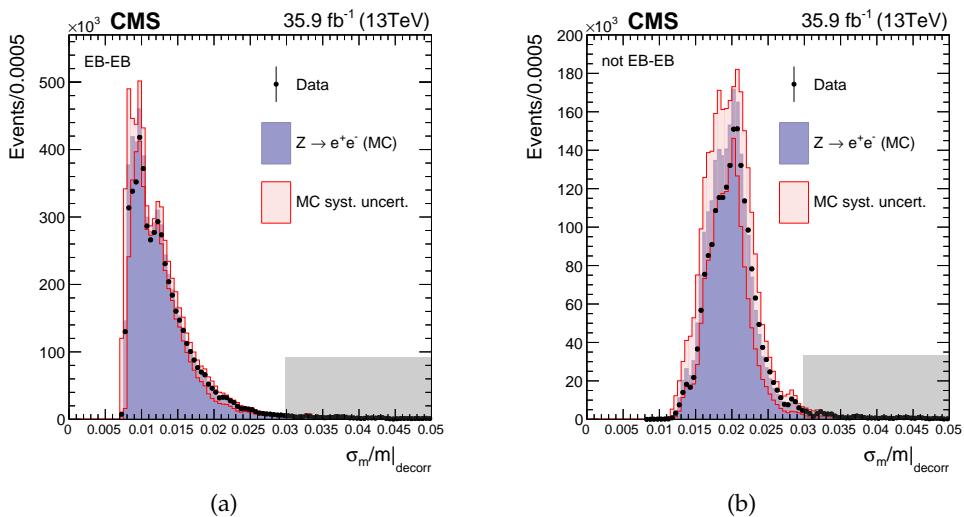


Figure 6.4: The decorrelated mass resolution estimator $\sigma_m/m|_{\text{decorr}}$ for data and simulated $Z \rightarrow e^+e^-$ events. The distributions are shown separately for events with both electrons in EB (left) and the remainder of the events (right). The shaded area represents the region of values of $\sigma_m/m|_{\text{decorr}}$ where events are not selected for the analysis.

the value of σ_E/E for each photon candidate, described in Section 5.8, is propagated to $\sigma_m/m|_{\text{decorr}}$ and it is shown by the red-shaded uncertainty band for simulation.

6.3 EVENT SELECTION

Photons are reconstructed using the algorithms described in Section 3.3 and are selected if they have a SC within $|\eta| < 2.5$, with the exclusion of the transition region between the EB and the EE, corresponding to $1.4442 < |\eta| < 1.566$. They then have to satisfy the criteria defined by the preselection, described in Section 5.6. Photons are further required to have transverse momentum divided by the diphoton mass greater than $1/3$ and $1/4$ for the leading and sub-leading photon, respectively. A minimum requirement is set for both photons on the score of the photon identification BDT, described in Section 5.7. This last selection criterion is optimized simultaneously with the definition of the event categories, described in Section 6.4.

Jets are clustered and selected as described in Section 3.5 and they are chosen if they have $p_T > 30$ GeV. Jets are required to be angularly separated from each of the photons by $\Delta R(\gamma, \text{jet}) > 0.4$, in order to avoid double-counting of photon candidates as jets. Different pseudorapidity selections for jets are used in different parts of the analysis: a more stringent one, which requires both jets to have $|\eta| < 2.5$, and one that extends the selection up to $|\eta| < 4.7$. The two collections are used to study differential observables requiring at least one or at least two selected jets in the event, respectively, as described in Section 6.5. The $|\eta| < 2.5$ collection contains jets that benefit from tracker information, resulting in better reconstruction quality and energy resolution; when applying the tighter requirement of at least two jets in the same event, the $|\eta|$ range is extended to 4.7 to in-

crease the selection acceptance. The same kinematic selection is applied to generator-level jets.

Jets selected within $|\eta| < 2.4$ and with $p_T > 30$ GeV are identified as b jets if they satisfy the tight working point requirement of the CSV tagging algorithm, introduced in Section 3.6. At the generator level, a jet is considered a b jet if it contains at least one clustered B hadron.

Leptons (electrons and muons) are selected if they have $p_T > 20$ GeV and $|\eta| < 2.4$. Furthermore, the minimum angular distance between a lepton and either of the photons is required to be $\Delta R(\gamma, l) > 0.35$. Electrons have to fulfill the selection described in Section 5.10.1. Muons are additionally requested to be isolated, by setting a maximum on the sum of the transverse energy, corrected for PU effects, of charged hadrons, neutral hadrons and photons in a cone of radius 0.4 around the muon. At the generator-level, leptons are required to satisfy the same kinematic selection. Furthermore, their momentum is “dressed”, i.e., the four-momenta of all photons in a cone of radius $\Delta R = 0.1$ around the lepton are added to the four-momentum of the lepton, allowing the recovery of photons emitted as final state radiation.

The events used in this analysis were collected using the same diphoton trigger described in Section 5.3. The efficiencies of the photon selection criteria are measured using the T&P technique, introduced in Section 3.8, analogously to what is done for the coupling analysis. While trigger and preselection efficiencies are unchanged compared to what already described there, the identification efficiencies are recomputed, given the different working point of the selection on the photon identification BDT for this analysis. These efficiencies are found to be 3% (5%) lower in data than in simulation for photons reconstructed in EB (EE) with $R_9 < 0.85$ (0.9), while they are compatible with unity within statistical uncertainties for photons with large values of R_9 . The efficiency scale factors are taken into account in the final signal modeling, to correct the signal yields estimated from simulation.

6.4 EVENT CATEGORIZATION

The events entering the analysis are requested to meet the preselection criteria, described in Section 5.6, and to have a diphoton invariant mass between 100 GeV and 180 GeV. The $\sigma_m/m|_{\text{decorr}}$ estimator, described in Section 6.2, is used to divide events into categories so as to maximize the analysis precision. The number of categories, as well as the position of their boundaries, is optimised with the same procedure described in Section 5.9. In addition a minimum requirement on the photon identification BDT score of each photon is optimised simultaneously with the category definition. The median expected significance of the signal hypothesis, computed with simplified signal and background models, is used as the figure of merit of the optimization, performed on simulated background and signal events, normalized to the integrated luminosity of the analyzed data set of 35.9 fb^{-1} . The median expected significance achievable with the current data set saturates when splitting the events in more than three categories, which is taken as the optimal number of categories. The optimized boundaries in $\sigma_m/m|_{\text{decorr}}$ for this configuration are found to be 0, 0.0084, 0.012, 0.030, and the associated minimum threshold on the photon identification BDT is 0.23, which corresponds to an efficiency of roughly 80% of selecting

a diphoton pair in signal events in the *baseline* fiducial phase space, defined in the next Section 6.6.

Thanks to the correlation of the $\sigma_m/m|_{\text{decorr}}$ with the pseudorapidities of the photons it is possible to ascertain the correspondence between the categories extracted via the optimization procedure and different geometrical regions of the detector. The distributions in Figure 6.3, described in Section 6.2, show that: the best resolution category (cat. 0) designates approximately events where both photons are reconstructed in the central region of EB ($|\eta| < 1.0$); the intermediate resolution category (cat. 1) collects events with both photons in EB, but at least one outside of its central region; the low resolution category (cat. 2) is populated by events with at least one photon reconstructed in the EE.

6.5 PHASE SPACE DEFINITIONS

The inherent characteristic of fiducial cross section measurements is that of being performed in a well-defined fiducial phase space, rather than extrapolated to the full phase space. While the selection requirements in data, listed in Section 6.3, are applied to reconstruction-level quantities, the definition of the fiducial phase space is set on generator-level quantities, to provide measurements comparable to the theoretical predictions, which do not account for the response of a specific detector. Given the finite resolution of this response, the two selections, though chosen to be as aligned as possible, do not exactly coincide, and event migrations can occur between the two: events belonging to the fiducial phase space can fail the analysis selection, defining its efficiency, and events fulfilling the selection criteria can originate from outside the acceptance of the fiducial phase space. This latter type of migrations, labeled as *outside acceptance* signal component, defines the proportions of the extrapolation occurring when interpreting the reconstructed signal yields as fiducial cross sections and, thus, the level of model dependence of the measurements on the theoretical predictions on which such an extrapolation relies.

The *baseline* fiducial phase space of this analysis, defined below, allows to confine these kind of contributions to below 1% of the total selected signal. This phase space is defined by requiring that generator-level photons have $|\eta| < 2.5$ and $p_T/m_{\gamma\gamma} > 1/3$ ($1/4$) for the first (second) ranking photon in p_T . The same kinematic selection is applied to reconstruction level quantities at the reconstruction level. Furthermore, photons are required to be isolated, by setting a maximum threshold of 10 GeV on the sum of the transverse energy of stable particles contained in a cone of radius $\Delta R = 0.3$ around the photon. This phase space definition is used for the measurement of the integrated fiducial cross section and for differential measurements as functions of observables pertaining to the diphoton system only. On top of this *baseline* phase space, further requirements can be applied to allow the measurements to be performed differentially in properties of additional objects produced in association with the Higgs bosons and to study specific production modes. A 1-jet sub-region of the phase space is specified by requiring at least one clustered jet within $|\eta| < 2.5$, selected as described in Section 6.3. This definition corresponds to $\sim 35\%$ of the *baseline* phase space and it is used to study observables involving one additional jet in the event. Observables requiring two additional jets, instead, are measured in a 2-jets phase space defined by requiring at least two jets, as defined in Section 6.3, within the extended pseudorapidity range $|\eta| < 4.7$, corresponding to $\sim 16\%$ of the *baseline* phase space. The VBF production mechanism is targeted by a dedicated phase space, labeled

VBF-enriched and defined as a subset of the *2-jets*, with selection criteria based on the main topological features of the VBF production. In particular, the *VBF-enriched* requires the pseudorapidity separation between the jets, $\Delta\eta^{jj}$, to be greater than 3.5 and their invariant mass, m^{jj} , to exceed 200 GeV. The size of such a phase space for signal events is $\sim 3.8\%$ of the *baseline* phase space.

Three additional phase space definitions are introduced, aiming at increasing the purity of signal events produced via a specific mechanism. These phase spaces, where the integrated fiducial cross section is measured, are defined by applying the following additional criteria, on top of the *baseline* requirements, on objects defined as in Section 6.3:

- at least one lepton, at least one b jet, labeled as *≥ 1 -lepton, ≥ 1 -b-jet*, selecting $\sim 1.7 \times 10^{-3}$ of the signal events in the *baseline* phase space;
- exactly one lepton, $p_T^{\text{miss}} \geq 100$ GeV, labeled as *1 -lepton, high p_T^{miss}* ($\sim 1.5 \times 10^{-3}$ of *baseline* phase space);
- exactly one lepton, $p_T^{\text{miss}} < 100$ GeV, labeled as *1 -lepton, low p_T^{miss}* ($\sim 7.4 \times 10^{-3}$ of *baseline* phase space).

The definitions of the *≥ 1 -lepton, ≥ 1 -b-jet* and *1 -lepton, high p_T^{miss}* phase spaces reproduce loosely the selection defined for categories, described in Section 5.10, targeting ttH and W($l\nu$)H production mechanisms, respectively. The *1 -lepton, low p_T^{miss}* selects the region of the phase space complementary in p_T^{miss} with respect to *1 -lepton, high p_T^{miss}* .

The definitions of the phase spaces introduced above are replicated at the reconstruction level by applying the same selections on the corresponding reconstruction-level quantities and objects.

6.6 OBSERVABLES

A comprehensive set of differential observables is chosen and studied in the relevant phase spaces. A summary of these observables, with the respective bin choice, is given in table 6.1.

The letter y indicates the rapidity of a particle or system of particles and the symbols j_1 and j_2 are used to denote the p_T -leading and p_T -sub-leading jet in the event, respectively.

Observables describing the kinematics of the Higgs boson are measured in the *baseline* phase space and are dominated by the ggH production mechanism. The $p_T^{\gamma\gamma}$ in ggH events is balanced by emission of QCD radiation and, thus, provides a sensitive probe to the modeling of the latter, as well as to the modelling of the Yukawa coupling of the Higgs boson. Similarly, the number of jets produced in association with the Higgs boson, N_{jet} , represents a complementary probe to the same theoretical aspects, as well as the modeling of different production processes. Furthermore, the different production mechanisms present distinctively diverse $p_T^{\gamma\gamma}$ spectra, with VH and ttH processes characterized by more boosted production on average, making the differential measurement of the $p_T^{\gamma\gamma}$ distribution sensitive to variations of the SM predictions and to contributions from BSM processes. The rapidity of the diphoton system, $|y^{\gamma\gamma}|$, is a sensitive probe to the Higgs boson production mechanism and the PDFs of the proton [35]. An important observable is the cosine of the photon helicity angle θ^* in the Collins-Soper frame [32], $|\cos(\theta^*)|$,

whose distribution is particularly sensitive to the spin and CP hypotheses of the decaying particle.

Observables describing the kinematic of the jets and their configuration relatively to the diphoton system are mostly sensitive to the modeling of the QCD radiation and to the relative contributions of different production mechanisms. In particular, while the *1-jet* phase space is still dominated by ggH production, in the *2-jet* phase space the contribution from VH and VBF processes becomes more significant. For this reason, a subset of the observables involving two jets ($p_T^{j_2}$, $|\Delta\phi^{j_1,j_2}|$, $|\Delta\phi^{\gamma\gamma,j_1,j_2}|$) is also measured in the *VBF-enriched* phase space.

The number of b jets, leptons and the missing transverse momentum are indicated with N_{jet}^b , N_{lepton} , and p_T^{miss} , respectively, and provide useful probes to contributions from production mechanisms, other than ggH, and to BSM effects. Contributions to the differential cross sections originating from BSM processes are, more in general, expected to be relevant in specific regions of the phase space depending on the alternative model considered.

The bin choice for the measurements has been determined as a trade-off between the number of parameters measured and their relative uncertainty. Care has also been taken to align the bin choice with that of similar measurements performed in other channels or on different data sets, in view of the combination of those results with the ones presented here.

For each subset of the *baseline* phase space, the events not contained in the subset, but part of the *baseline* phase space, are collected in an additional bin, labeled *underflow*. As for all other bins, this additional bin corresponds to an additional process at the generator level whose normalization is extracted in the final likelihood fit, and at the reconstruction level to an additional event category. The events in this bin are used in the final signal extraction, both for integrated and differential measurements, to account properly for event migrations between the *baseline* and the subset considered and to provide additional information used to constrain the nuisance parameters of the fit. In particular, this configuration allows the floating of the mass position of the signal peak, m_H , which can thus be inferred and constrained from data.

6.7 STATISTICAL ANALYSIS

For a differential observable with n_b kinematic bins, the analysis extracts the cross section in all bins via a simultaneous binned maximum likelihood fit to the $m_{\gamma\gamma}$ spectra in all event categories. The same bin choice used at the generator level, introduced in Section 6.6, is transferred at reconstruction level and applied to the events selected by the three event categories in $\sigma_m/m|_{\text{decorr}}$, with the consequence of increasing the final number of categories to $3n_b$. Furthermore, the misalignment between generated and reconstructed quantities causes event migrations across bins and categories. These migrations are measured by sorting simulated signal events in the matrix defined by the generator and reconstruction bins, also labeled *response matrix*. Most of the events populate the diagonal elements of this matrix, with the number and size of populated off-diagonal elements becoming larger with the worsening of the resolution of the detector in the observable studied. Two examples of such matrices are shown in Figure 6.5: the response matrices for the differential measurement as a function of $p_T^{\gamma\gamma}$ (top row) are mostly diagonal, thanks to the high resolution in the diphoton momentum, while the matrices for the N_{jet} observ-

Phase space	Observable	Bins											
Baseline	$p_T^{\gamma\gamma}$ (GeV)	0	15	30	45	80	120	200					
	N_{jet}	0	1	2	3	4	∞						
	$ y^{\gamma\gamma} $	0	0.15	0.3	0.6	0.9	2.5						
	$ \cos(\theta^*) $	0	0.1	0.25	0.35	0.55	1						
1-jet Baseline + ≥ 1 jet $p_T^j > 30$ GeV, $ \eta^j < 2.5$	$p_T^{\gamma\gamma}$ (GeV), $N_{\text{jet}} = 0$	0	20	60	∞								
	$p_T^{\gamma\gamma}$ (GeV), $N_{\text{jet}} = 1$	0	60	120	∞								
	$p_T^{\gamma\gamma}$ (GeV), $N_{\text{jet}} \geq 1$	0	150	300	∞								
	N_{jet}^b	0	1	2	∞								
2-jets Baseline + ≥ 2 jets $p_T^j > 30$ GeV, $ \eta^j < 4.7$	N_{lepton}	0	1	2	∞								
	p_T^{miss} (GeV)	0	100	200	∞								
	$p_T^{j_1}$ (GeV)	0	45	70	110	200	∞						
	$ y^{j_1} $	0	0.5	1.2	2	2.5							
VBF-enriched 2-jets + $\Delta\eta^{jj} > 3.5$, $m^{jj} > 200$ GeV	$ \Delta\phi^{\gamma\gamma,j_1} $	0	2.6	2.9	3.03	π							
	$ \Delta y^{\gamma\gamma,j_1} $	0	0.6	1.2	1.9	∞							
	$p_T^{j_2}$ (GeV)	0	45	90	∞								
	$ y^{j_2} $	0	1.2	2.5	4.7								
2-jets Baseline + ≥ 2 jets $p_T^j > 30$ GeV, $ \eta^j < 4.7$	$ \Delta\phi^{j_1,j_2} $	0	0.9	1.8	π								
	$ \Delta\phi^{\gamma\gamma,j_1,j_2} $	0	2.9	3.05	π								
	$ \bar{\eta}_{j_1,j_2} - \eta_{\gamma\gamma} $	0	0.5	1.2	∞								
	m_{j_1,j_2} (GeV)	0	100	150	450	1000	∞						
VBF-enriched 2-jets + $\Delta\eta^{jj} > 3.5$, $m^{jj} > 200$ GeV	$ \Delta\eta_{j_1,j_2} $	0	1.6	4.3	∞								
	$p_T^{j_2}$ (GeV)	0	45	90	∞								
	$ \Delta\phi^{j_1,j_2} $	0	0.9	1.8	π								
	$ \Delta\phi^{\gamma\gamma,j_1,j_2} $	0	2.9	3.05	π								

Table 6.1: The differential observables studied with the corresponding bins chosen, grouped by the region of fiducial phase space where the measurements are performed.

able (bottom row) show larger off-diagonal elements. These matrices are defined for each of the three categories in $\sigma_m/m|_{\text{decorr}}$ and labeled as K_k^{ij} , where i indicates the category under study, the column index k represents the generator-level bin and the row index j the reconstruction-level bin.

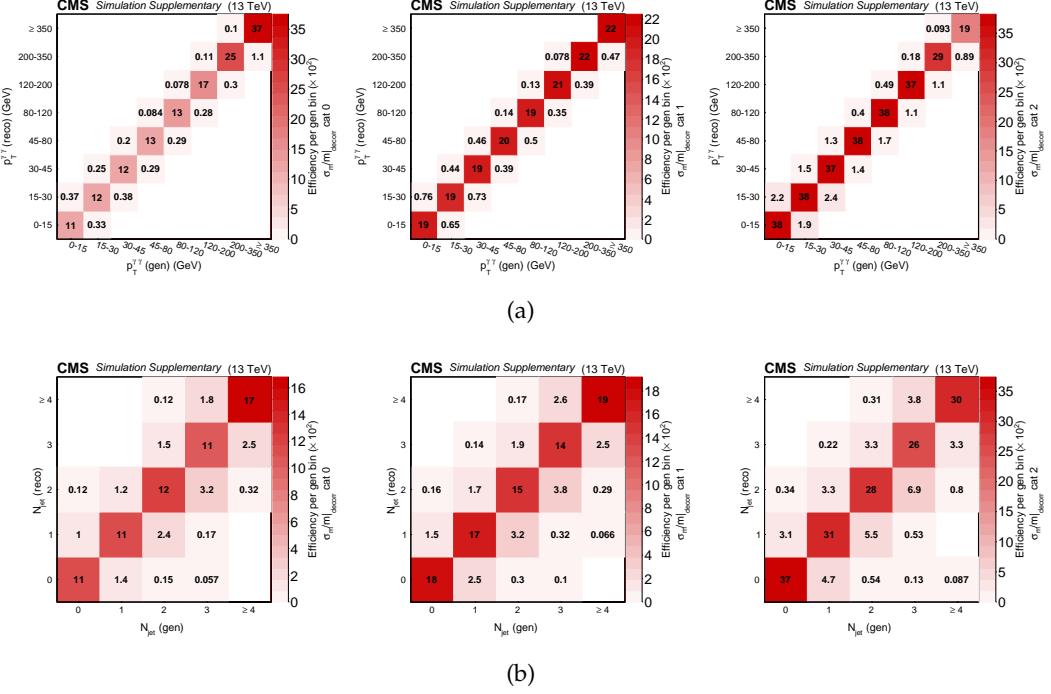


Figure 6.5: The response matrices for $p_T^{\gamma\gamma}$ (a) and N_{jet} (b). The matrices are derived separately for categories 0, 1, and 2 of $\sigma_m/m|_{\text{decorr}}$, shown from left to right. The numbers in the cells of the matrices indicate the efficiency of a cell in selecting events in the corresponding generator-level bin, across all $\sigma_m/m|_{\text{decorr}}$ categories.

The signal model, built with the same technique described in Section 5.11 for the coupling analysis, is constructed separately for each cell of these matrices, as well as for the signal components originating inside and outside the acceptance of the fiducial volume, and for events where a correct and incorrect vertex assignment is performed, according to the correct vertex identification definition given in Section 5.4. This fine segmentation of the signal shape modeling allows the direct extraction from the reconstructed diphoton mass spectra of the generator-level signal yields, by building the K matrices in the likelihood minimized by the fit. This corresponds to a matrix inversion unfolding [33].

The complete likelihood expression can be expressed as follows:

$$\begin{aligned} \mathcal{L}(\text{data} | \Delta \vec{\sigma}_{\text{fid}}, \vec{n}_{\text{bkg}}, \vec{\theta}_S, \vec{\theta}_B) = \\ \prod_{i=1}^{n_{\text{cat}}} \prod_{j=1}^{n_b} \prod_{l=1}^{n_{m_{\gamma\gamma}}} \left(\frac{\sum_{k=1}^{n_b} \Delta \sigma_k^{\text{fid}} K_k^{ij}(\vec{\theta}_S) S_k^{ij}(m_{\gamma\gamma}^l | \vec{\theta}_S) L + n_{\text{OA}}^{ij} S_{\text{OA}}^{ij}(m_{\gamma\gamma}^l | \vec{\theta}_S) + n_{\text{bkg}}^{ij} B^{ij}(m_{\gamma\gamma}^l | \vec{\theta}_B)}{n_{\text{sig}}^{ij} + n_{\text{bkg}}^{ij}} \right)^{n_{\text{ev}}^{ij}} \cdot \\ \text{Pois}(n_{\text{ev}}^{ij} | n_{\text{sig}}^{ij} + n_{\text{bkg}}^{ij}) \cdot \text{Pdf}(\vec{\theta}_S) \text{Pdf}(\vec{\theta}_B) \end{aligned} \quad (6.3)$$

where:

- $n_{\text{cat}} = 3$ is the number of categories in $\sigma_m/m|_{\text{decorr}}$ and $n_{m_{\gamma\gamma}}$ is the number of bins of the $m_{\gamma\gamma}$ distribution;
- $\Delta\vec{\sigma}^{\text{fid}} = (\Delta\sigma_1^{\text{fid}}, \dots, \Delta\sigma_{n_b}^{\text{fid}})$ is the vector of generator-level fiducial cross sections, multiplied by the branching fraction of the diphoton decay channel, which are the parameters of interest extracted in the fit;
- K_k^{ij} are the already introduced detector response matrices, normalized such that each element represents the probability that an event generated in the k -th kinematic bin is reconstructed in the j -th reconstruction-level bin and the i -th $\sigma_m/m|_{\text{decorr}}$ category;
- the functions S_k^{ij} and B^{ij} are the signal and background distributions in $m_{\gamma\gamma}$ for the bin ijk , normalized to unity, constructed with the methods illustrated in the Sections 5.11 and 5.12, respectively;
- L is the total integrated luminosity analyzed;
- $n_{\text{ev}}^{ij}, n_{\text{sig}}^{ij}, n_{\text{bkg}}^{ij}$ are the number of observed, signal and background events in the ij -th reconstruction-level category respectively;
- the term $n_{\text{OA}}^{ij} S_{\text{OA}}^{ij}$ indicates the *outside acceptance* signal contamination, described in Section 6.5. This contribution is estimated from simulation of the SM signal production and kept fixed in the signal extraction fit. In this way, the measured cross sections are corrected in the fit for such an effect;
- the parameters $\vec{\theta}_S$ and $\vec{\theta}_B$ are the nuisance parameters associated with the signal and background model, respectively.

The direct extraction of the generator-level cross sections $\Delta\vec{\sigma}^{\text{fid}}$ from the likelihood fit provides measurements unfolded to particle level. The events collected in the *underflow* bin and categories are taken into account in the signal extraction and the corresponding determination of uncertainties, as introduced in Section 6.6. The test statistics $q(\Delta\vec{\sigma}^{\text{fid}})$, based on the profile likelihood ratio [46], is asymptotically distributed as a χ^2 with n_b degrees of freedom[46] and can be expressed as:

$$q(\Delta\vec{\sigma}^{\text{fid}}) = -2 \cdot \log \left(\frac{\mathcal{L}(\Delta\vec{\sigma}^{\text{fid}} | \hat{\vec{\theta}}_{\Delta\vec{\sigma}^{\text{fid}}})}{\mathcal{L}(\Delta\hat{\vec{\sigma}}^{\text{fid}} | \hat{\vec{\theta}})} \right) \quad (6.4)$$

where $\vec{\theta} = (n_{\text{bkg}}, \vec{\theta}_S, \vec{\theta}_B)$. The best fit estimate of $\vec{\theta}$ ($\Delta\vec{\sigma}^{\text{fid}}$) is indicated with $\hat{\vec{\theta}}$ ($\Delta\hat{\vec{\sigma}}^{\text{fid}}$), and the best fit estimate of $\vec{\theta}$, conditional on the value of $\Delta\vec{\sigma}^{\text{fid}}$ is indicated with $\hat{\vec{\theta}}_{\Delta\vec{\sigma}^{\text{fid}}}$. The uncertainties and the correlation matrices are obtained from this test statistics. A nuisance parameter, m_H , common to the signal shapes in all categories, is used to describe the mass position of the signal peak. All the nuisance parameters, including m_H , are profiled in the maximization of the likelihood.

6.8 SYSTEMATIC UNCERTAINTIES

The sources of systematic uncertainties are included in the likelihood as nuisance parameters and they are profiled in the fit. The sources considered are those discussed in Section 5.13 for the coupling analysis, in particular: integrated luminosity uncertainty, trigger and photon identification efficiency scale factors uncertainties, vertex identification efficiency uncertainty, uncertainty in photon energy scale and resolution determination, uncertainty in jet energy scale and resolution corrections, efficiency scale factors uncertainties for identification of jets, b-jets and leptons, uncertainty in the determination and correction of the missing transverse momentum.

The main difference in the assessment of the systematic uncertainties with respect to the coupling analysis concerns the theoretical systematic uncertainties affecting the total expected yields, which are here associated with the theoretical prediction and not with the measurements extracted. The impact on the measurements of the uncertainties coming from the choice of the QCD renormalization and factorization scales and of the PDF of the proton is studied and found to be negligible.

The systematic uncertainties are found to be largely subdominant in all the measurements performed compared to the statistical component. The measurement of the integrated cross section in the *baseline* phase space, characterized by the smallest statistical uncertainty among the measurements extracted ($\sim 13\%$), is affected by a systematic uncertainty of about 8%.

The uncertainties related to the jet energy scale and resolution corrections, while negligible for jet-inclusive measurements, produce sizable event migrations across bins for jet-related differential observables, ranging between 10% and 20%, depending on the bin.

6.9 RESULTS

As described in Section 6.7, the cross section measurements are extracted through a signal-plus-background fit to the diphoton invariant mass distribution, performed simultaneously in all the event categories. The mass distributions in data and the relative best signal-plus-background fit results are shown Figure 6.6 in the three $\sigma_m/m|_{\text{decorr}}$ categories for the measurement of the integrated fiducial cross section in the *baseline* phase space. The scan of the test statistics as a function of this cross section is shown in Fig. 6.7. The best fit value, corresponding to the minimum of the scan, is compared to the theoretical prediction (red band). The two-sided one-standard deviation uncertainties in the observed value are extracted from the likelihood scan as the values corresponding to an increase of q of 1 unity with respect to its minimum. This prescription yields the following result:

$$\hat{\sigma}_{\text{fiducial}} = 84 \pm 11 \text{ (stat)} \pm 7 \text{ (syst)} \text{ fb} = 84 \pm 13 \text{ (stat+syst)} \text{ fb}$$

The value of the profiled Higgs boson mass is compatible with the world average of 125.09 GeV [1].

The theoretical prediction for a Higgs boson with $m_H = 125.09$ GeV is computed taking the normalization of the signal yields for each Higgs boson production mechanism from the predictions in [35]. The sum of the main four production mechanisms yields the full phase space prediction, σ_{full} . The same reference [35] is used for the value of the

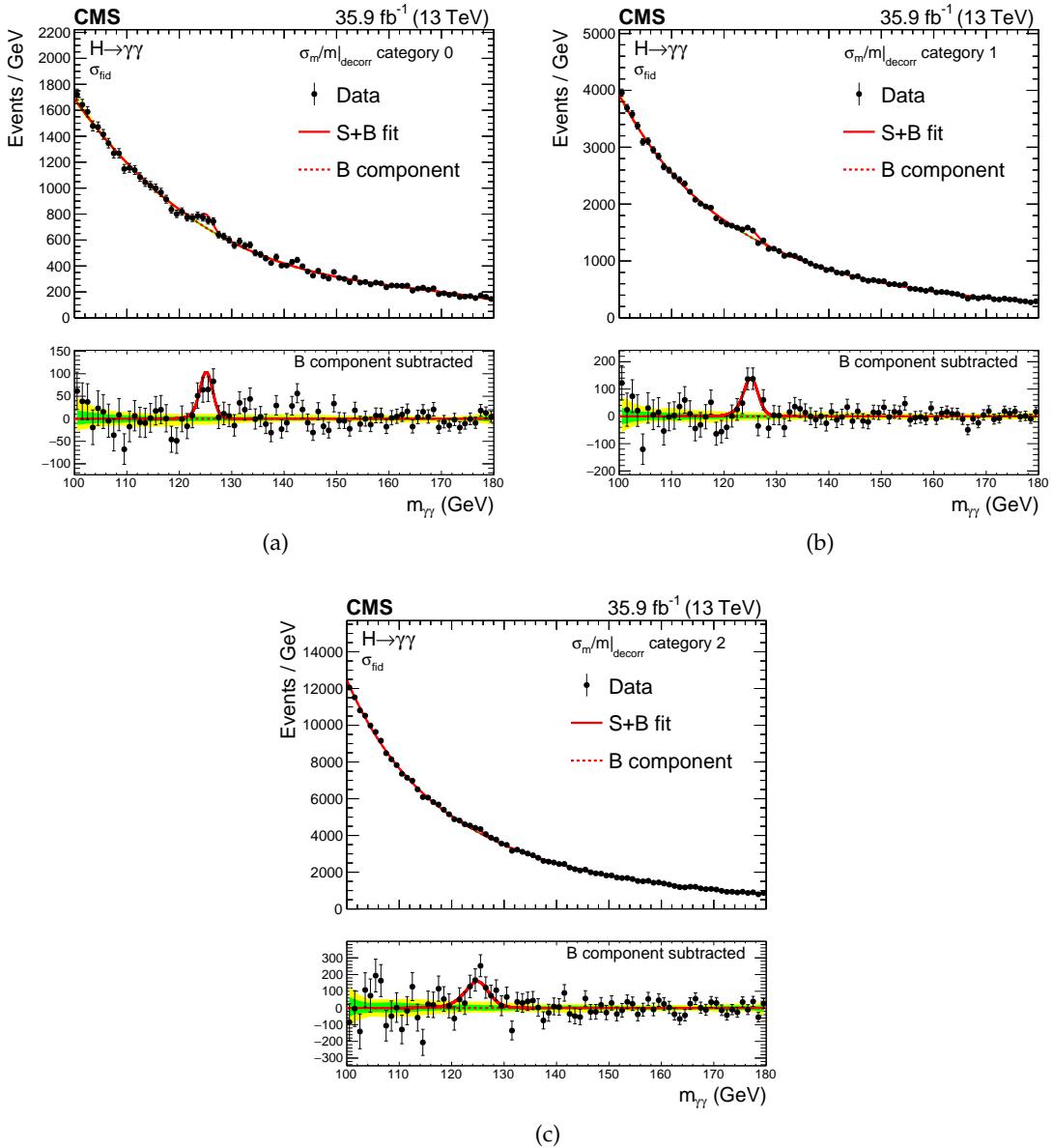


Figure 6.6: The diphoton mass spectrum in data (black points), together with the best signal-plus-background fit (red lines), for each $\sigma_m/m|_{\text{decorr}}$ category employed for the measurement of the inclusive fiducial cross section. The green (yellow) band indicates the one (two) standard deviation uncertainty in the background component.

branching fraction of the Higgs boson into two photons, $\text{BR}_{\gamma\gamma}$. To obtain the prediction corresponding to the phase space and selection of this analysis, signal events simulated with the `MADGRAPH5_AMC@NLO` program are used to compute the fiducial region acceptance for a Higgs boson with a mass of 125.09 GeV . The acceptance of the *baseline* phase space, \mathcal{A} , is estimated to be 0.60 with an uncertainty of 1% due to the choice of renormalization and factorization scales. This variable has a small dependence on the production mechanism and amounts to 0.60 , 0.60 , 0.52 , and 0.52 for ggH , VBF , VH , and

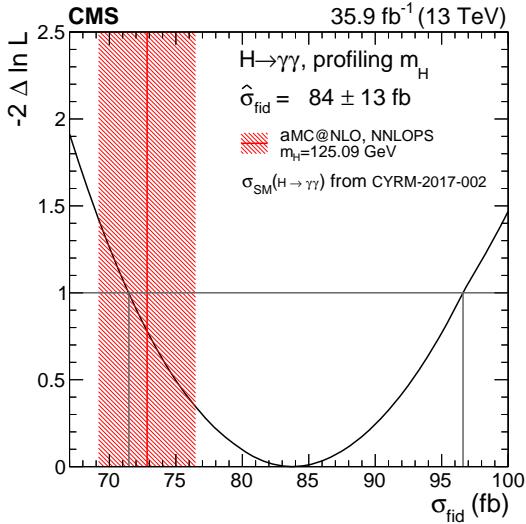


Figure 6.7: Likelihood scan (black curve) for the fiducial cross section measurement, where the value of the SM Higgs boson mass is profiled in the fit. The measurement is compared to the theoretical prediction (vertical red line), with which it is in agreement within uncertainties.

$t\bar{t}H$ production, respectively. The integrated fiducial cross section prediction is then obtained as the product of the three terms introduced:

$$\sigma_{\text{fiducial}}^{\text{theory}} = \sigma_{\text{full}} \text{BR}_{\gamma\gamma} \mathcal{A} = 73 \pm 4 \text{fb} \quad (6.5)$$

The uncertainty in this prediction is taken as the result of the propagation of the uncertainties in the per-production mechanism cross sections and in $\text{BR}_{\gamma\gamma}$ provided in [35]. Furthermore, the impact on the calculation of the acceptance of independently varying the renormalization and factorization scales by a factor of 2 and $1/2$, excluding the combinations $(1/2, 2)$ and $(2, 1/2)$, is added in quadrature as an additional source of uncertainty in $\sigma_{\text{fiducial}}^{\text{theory}}$. The acceptance \mathcal{A} has also been estimated with an alternative generator, `POWHEG`, both with and without re-weighting the ggH events to match the prediction of the `POWHEG` NNLOPS program. The alternative estimates are found to agree with the one obtained with `MADGRAPH5_AMC@NLO` within less than 1% and thus are not shown in the results. The measured value of $\hat{\sigma}_{\text{fiducial}} = 84 \pm 13 \text{ fb}$ is found in agreement within 1 standard deviation with the theoretical prediction of $\sigma_{\text{fiducial}}^{\text{theory}} = 73 \pm 4 \text{fb}$.

Using a generalization to an arbitrary number of bins of the signal extraction described for the integrated fiducial cross section (corresponding to a one-bin measurement), as described in Section 6.7, the measurement of the differential distributions of the fiducial cross section as functions of the observables described in Section 6.6 is performed. The Figures 6.8-6.13 compare, in each bin of the distributions, the best fit value measured in data (black dot) with the theoretical prediction. The total uncertainties on the observed value are indicated by the black bars, with the systematic component of the uncertainty shown separately by the red bar. The calculation of the theoretical predictions, obtained with a procedure analogous to that described for the integrated fiducial cross section, relies on different MC generators for the calculation of the shape of the

observables' spectra for the different production mechanisms. The contributions from VBF, VH and ttH production mechanisms (referred to as HX as a whole) are simulated with the `MADGRAPH5_AMC@NLO` generator (orange line). The predictions for the ggH component are added on top of the HX contribution and are obtained with the `MADGRAPH5_AMC@NLO` program (green histogram). An alternative prediction for the ggH production mechanism is obtained by weighting the events in the sample generated with `MADGRAPH5_AMC@NLO` to match the `POWHEG NNLOPS` prediction (red histogram), as explained in Sec. 4.4. The ggH contribution is also estimated using the `POWHEG` generator (blue histogram), which includes the simulation of only one additional jet produced in association with the Higgs boson, for those observables which are inclusive in the number of jets or involving only the first jet in the event.

The uncertainties taken into account in the estimation of these predictions, shown as bands in each bin of the histograms, are the uncertainties in the total integrated cross section for individual processes and on $\text{BR}_{\gamma\gamma}$. The maximum variation induced in the bin acceptance by the variation of renormalization and factorization energy scales is taken as the uncertainty in the acceptance of a given bin and added in quadrature to the uncertainty in the prediction in each bin. For the observable $|\Delta\phi^{\gamma\gamma,j_1j_2}|$, the spectrum is known to be not infrared-safe for values close to π [9], which correspond to a configuration where the Higgs boson and the di-jet system are close to back-to-back and the emission of soft collinear additional jets in ggH events is involved. Large theoretical uncertainties are associated with this regime, which are not taken into account here. This leads to the theoretical uncertainties obtained with scale variations tending to be underestimated in the large $|\Delta\phi^{\gamma\gamma,j_1j_2}|$ regime. The effect described is particularly relevant in the last bin of the spectrum corresponding to $|\Delta\phi^{\gamma\gamma,j_1j_2}| \in [3.05, \pi]$.

The differential cross sections are measured with a precision that varies greatly depending on the phase space and on the particular bin of the observable. The spectrum of the diphoton system transverse momentum, $p_T^{\gamma\gamma}$, is measured with an average precision of 40% in 8 bins, as shown in Figure 6.8 (top left). The other observables characterizing the diphoton system, $|y^{\gamma\gamma}|$ and $|\cos(\theta^*)|$, shown in Figure 6.8 (bottom right and bottom left), are measured with an average uncertainty of $\sim 35\%$, but with only 5 bins. The uncertainties in the measurement of the cross section as a function of the number of jets, N_{jet} , depend strongly on the bin considered, going from $\sim 25\%$ for the 0-jet bin up to $\gtrsim 100\%$ for the high jet multiplicity bins, as indicated in Figure 6.8 (top right). When moving to the 1-jet phase space, the reduced data set selected allows to extract only 4 bins for each observable with uncertainties around 40%, except $p_T^{j_1}$, for which 5 bins are defined with about 50% average uncertainty. The measurements in the 1-jet phase space are shown in Figure 6.9. The further reduction of the phase space when measuring observables involving two jets leads to a corresponding decrease in number of bins, 3, and increase in the average uncertainties, which range between $\sim 70\%$ and $\sim 90\%$. The combined mass of the two jets, $m_{j_1j_2}$, is the only observable of this group for which 5 bins are defined. The result of these measurements is shown in Figures 6.10 and 6.11. The differential measurements in the *VBF-enhanced*, shown in Figure 6.12, phase space are extracted keeping the same bin definition used in the 2-jets phase space, with uncertainties between $\sim 110\%$ and $\sim 150\%$. The double differential measurement as simultaneous function of $p_T^{\gamma\gamma}$ and N_{jet} is performed in 3×3 bins and yields uncertainties between $\sim 35\%$ and $\sim 60\%$, depending on the bin, and it is reported in Figure 6.13. The differential cross

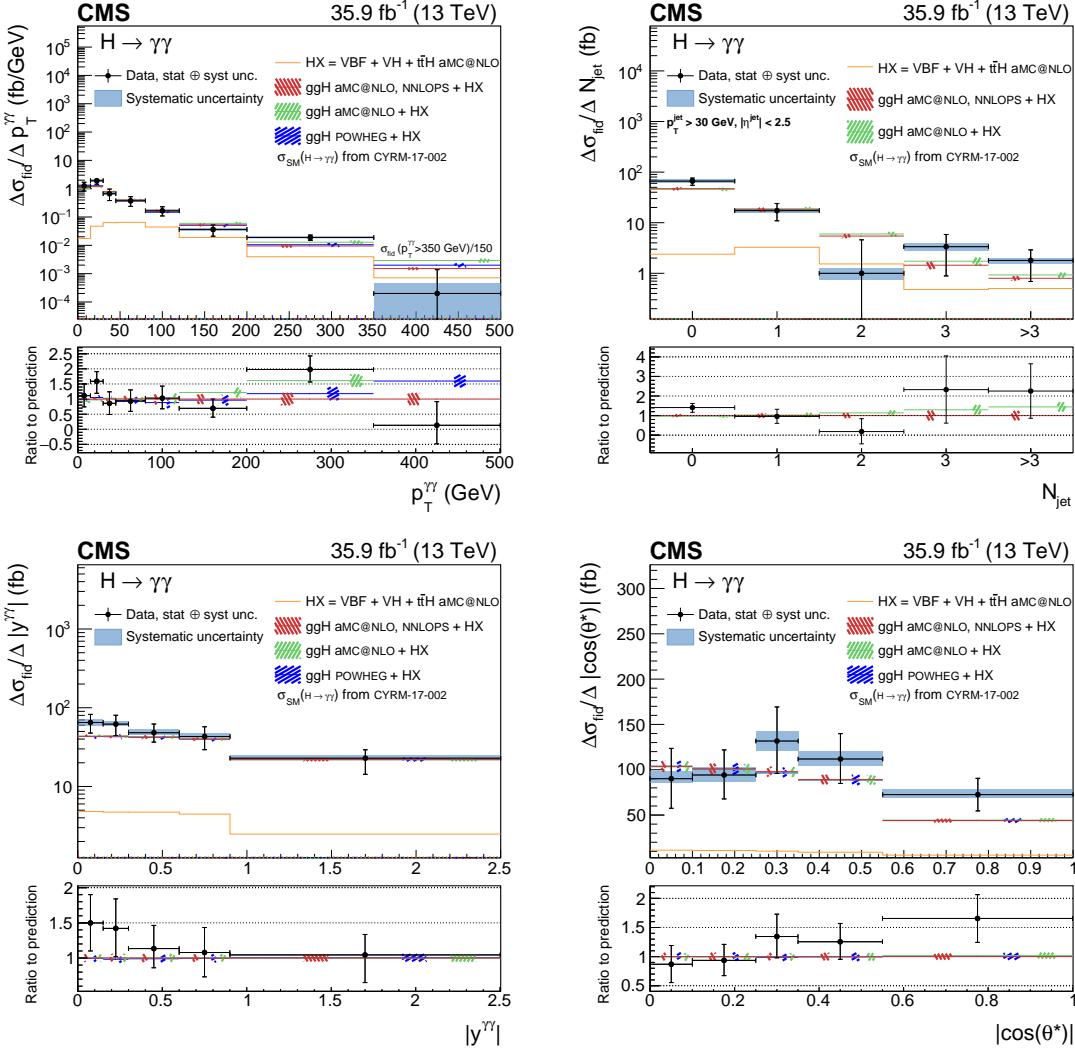


Figure 6.8: The measurement of the differential cross section for $p_T^{\gamma\gamma}$, N_{jet} , $|y^{\gamma\gamma}|$, and $|\cos(\theta^*)|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure.

sections extracted as functions of N_{jet}^b , N_{lepton} and p_T^{miss} are shown in Figure 6.14: the vast majority of events do not contain b-jets, nor leptons and have small values of p_T^{miss} . This is reflected in the uncertainties of the first bin of each of these observables being comparable with that of the measurement of the integrated fiducial cross section. The other bins, then, correspond to highly restrictive selections, where the cross section is measured with uncertainties around 200-250%. The results are found to be in agreement with the SM predictions within uncertainties.

The integrated fiducial cross section can also be extracted in very restricted sub-regions of the phase space, defined in Section 6.6 and corresponding to about only $\sim 10^{-3}$ of

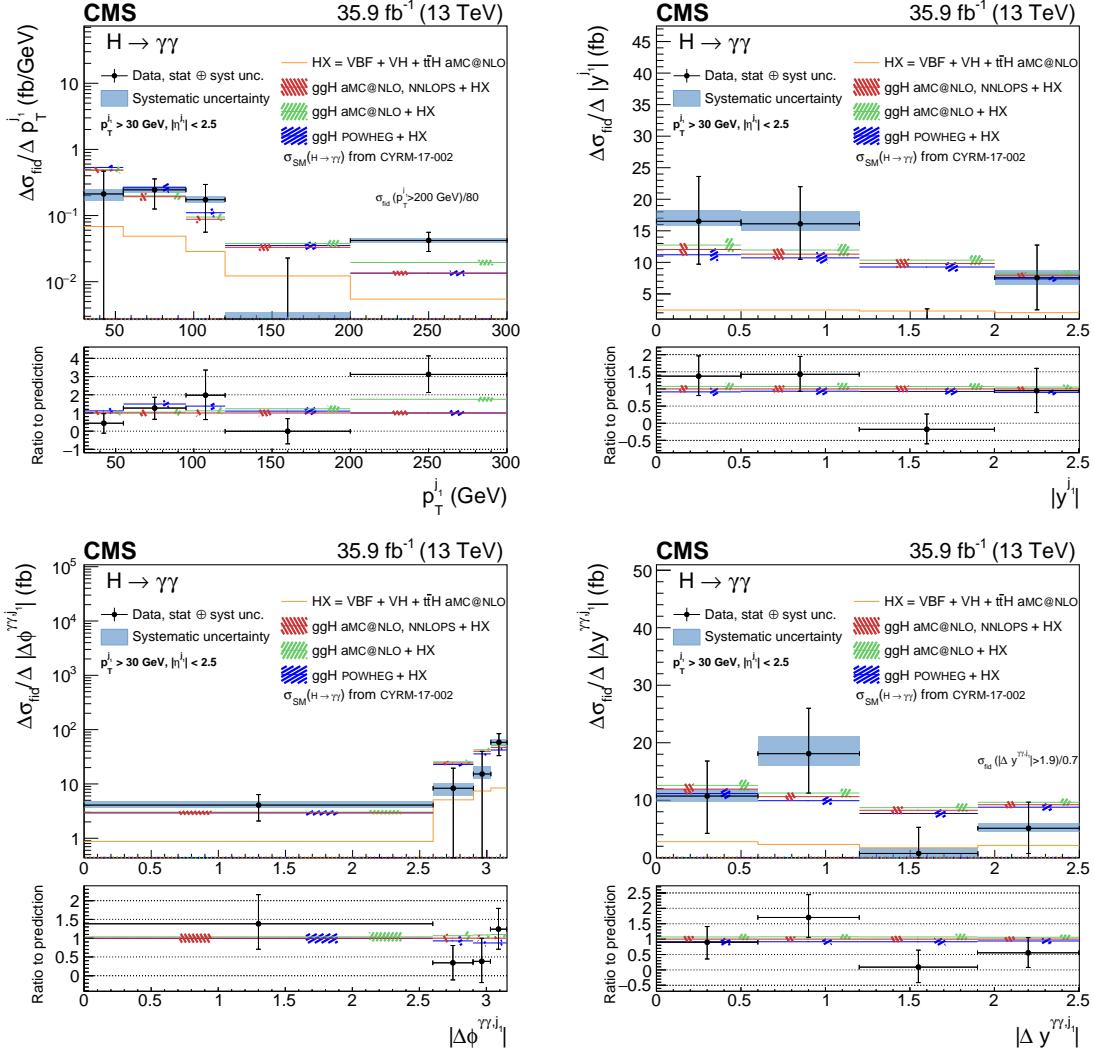


Figure 6.9: The measurement of the differential cross section for $p_T^{j_1}$, $|y^{j_1}|$, $|\Delta\phi^{\gamma\gamma,j_1}|$, and $|\Delta y^{\gamma\gamma,j_1}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure.

the *baseline* phase-space. Such regions, labeled ≥ 1 -lepton, ≥ 1 -*b*-jet, 1-lepton, high p_T^{miss} and 1-lepton, low p_T^{miss} , target individual production mechanisms of the Higgs boson. The results of these measurements are shown in Figure 6.15. The summary given in the Figure is made more comprehensive by reporting also the results of selected bins of the differential distributions. The measurements are compared to the theory predictions, whose values and uncertainties are obtained with the same methods described for the other observables earlier in this Section and relying on the `MADGRAPH5_AMC@NLO` generator for the computation of the acceptance. The uncertainties in the measurements range between

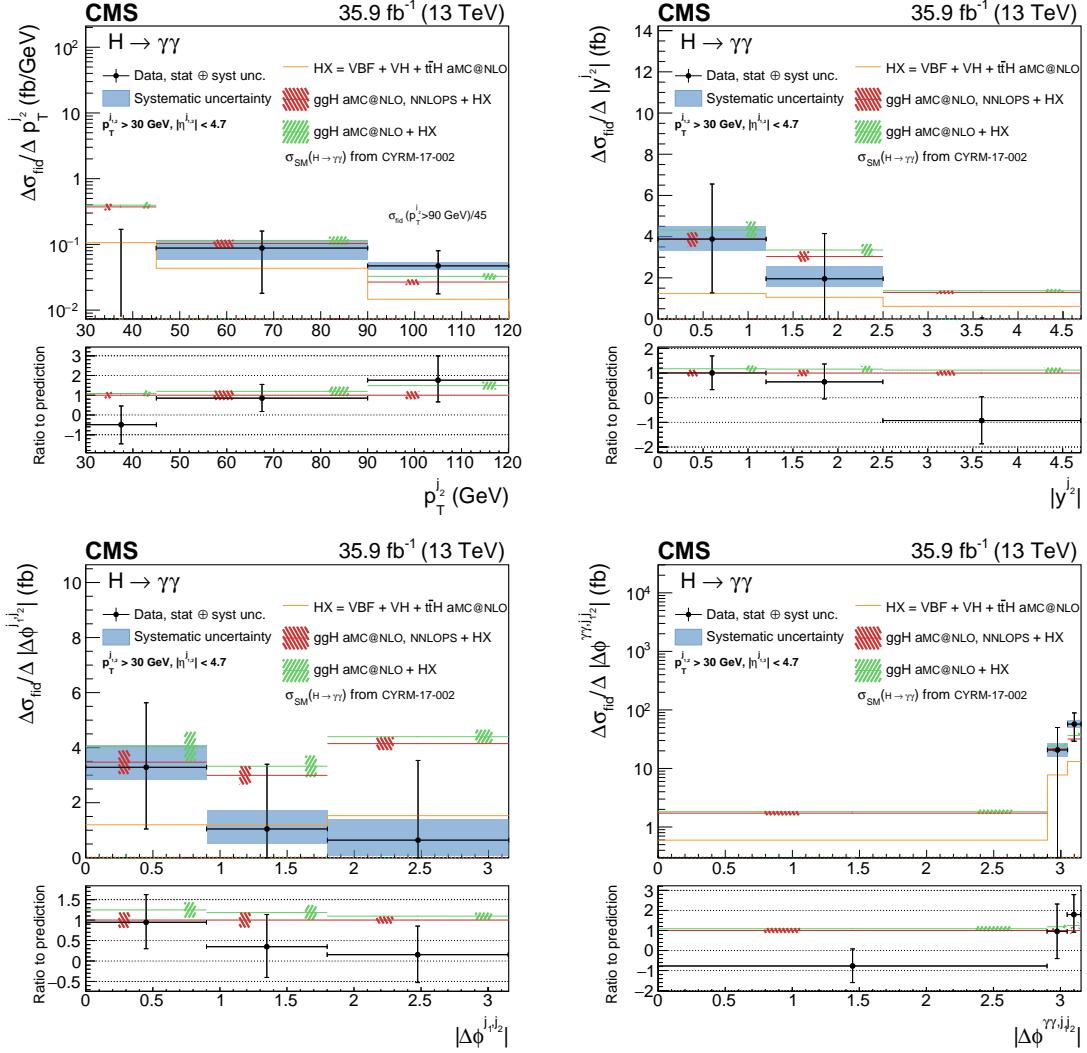


Figure 6.10: The measurement of the differential cross section for $p_T^{j_2}$, $|y^{j_2}|$, $|\Delta\phi^{j_1,j_2}|$, and $|\Delta\phi^{\gamma\gamma,j_1,j_2}|$.

The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure.

around 250% (≥ 1 -lepton, high p_T^{miss} and ≥ 1 -lepton, low p_T^{miss}) and $\sim 350\%$ (≥ 1 -lepton, ≥ 1 - b -jet). The measurements are found to be in agreement with the SM prediction.

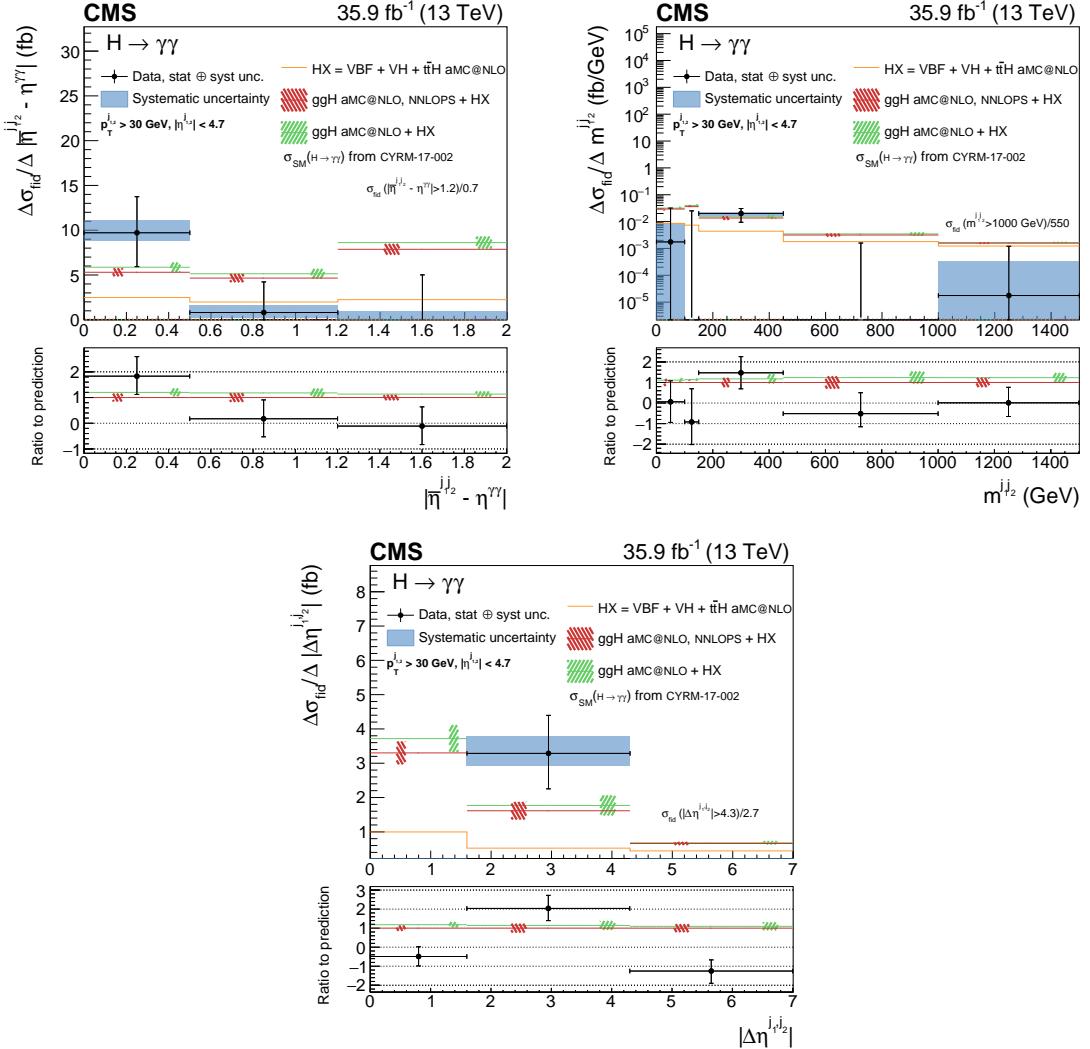


Figure 6.11: The measurement of the differential cross section for $|\eta_{j_1 j_2} - \eta_{\gamma\gamma}|$, $m_{j_1 j_2}$, and $|\Delta\eta_{j_1, j_2}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure.

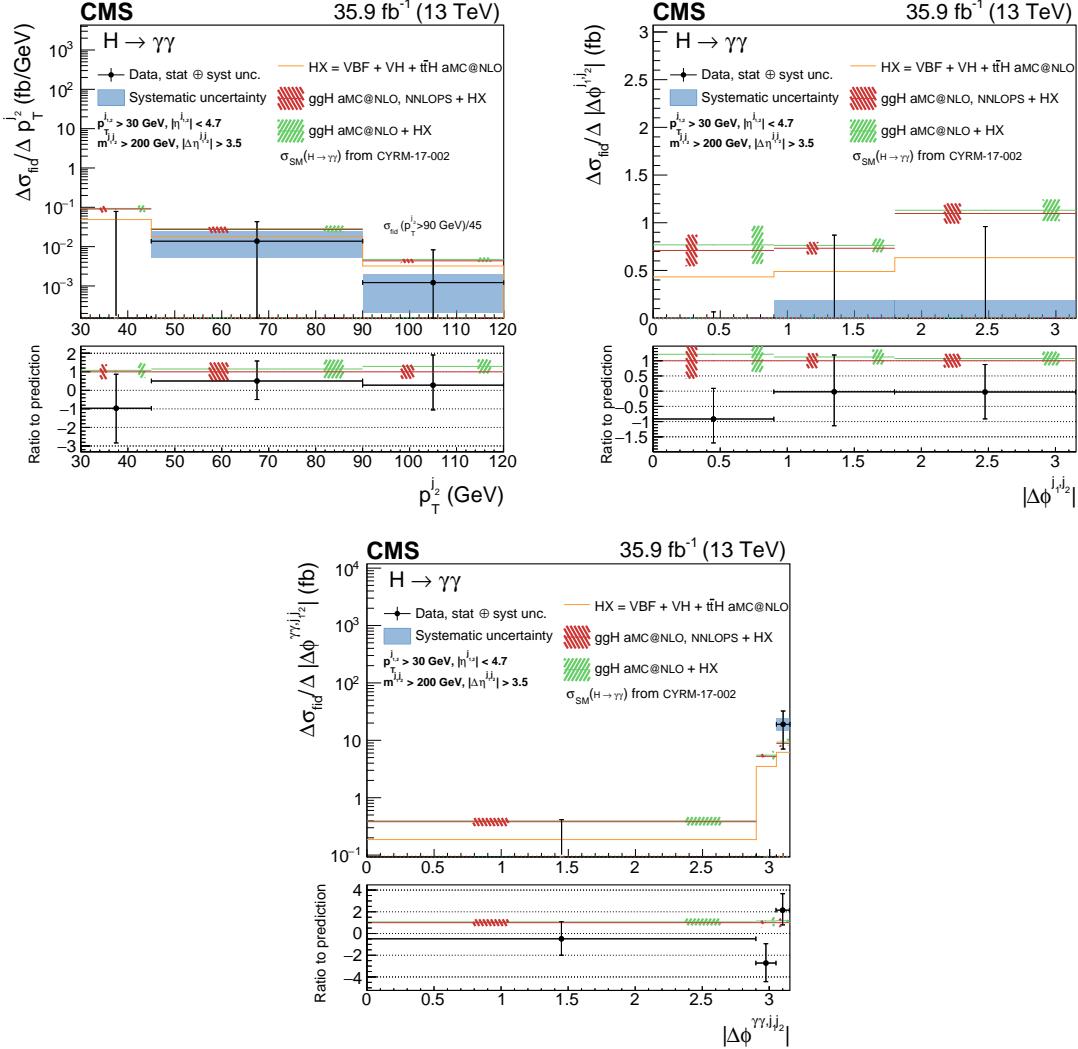


Figure 6.12: The measurement in a VBF-enriched sub-region of the fiducial phase space of the differential cross section for p_T^{j2} , $|\Delta\phi^{j_1 j_2}|$, and $|\Delta\phi^{\gamma\gamma j_1 j_2}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure.

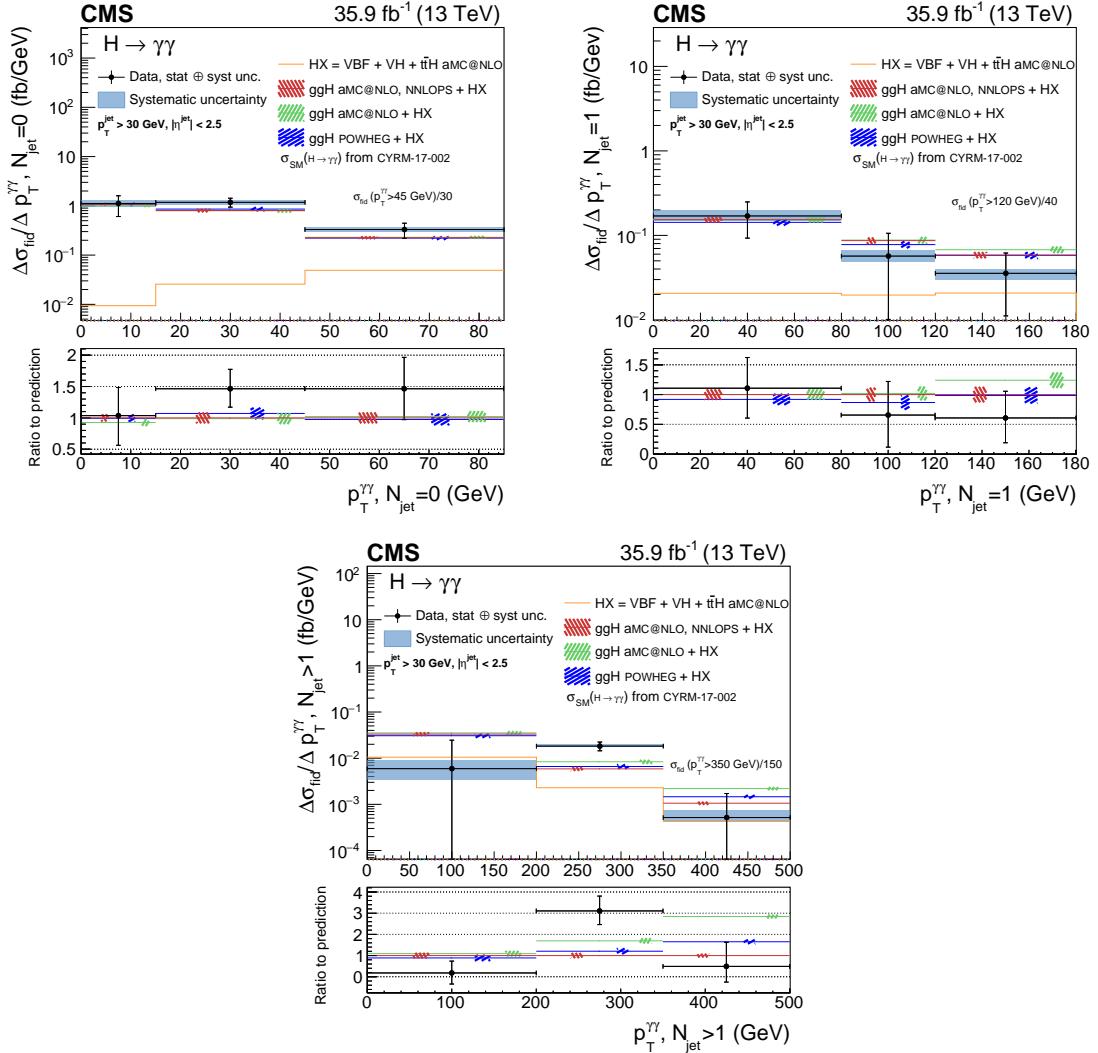


Figure 6.13: The measurement of the differential cross section for $p_T^{\gamma\gamma} \times N_{\text{jet}}$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last, overflow bin of the distribution is indicated in the figure.

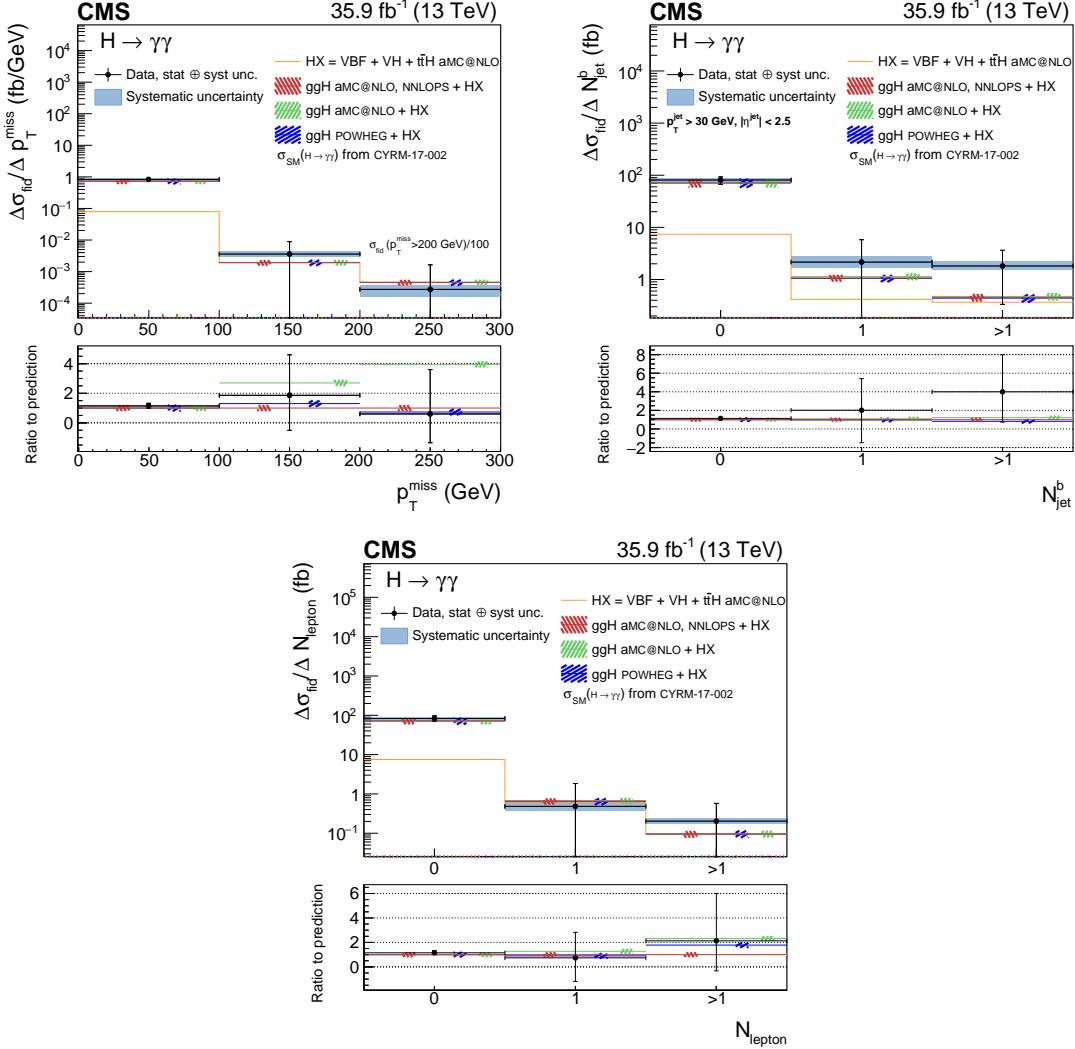


Figure 6.14: The measurement of the differential cross section for p_T^{miss} , N_{jet}^b , and N_{lepton} . The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure.

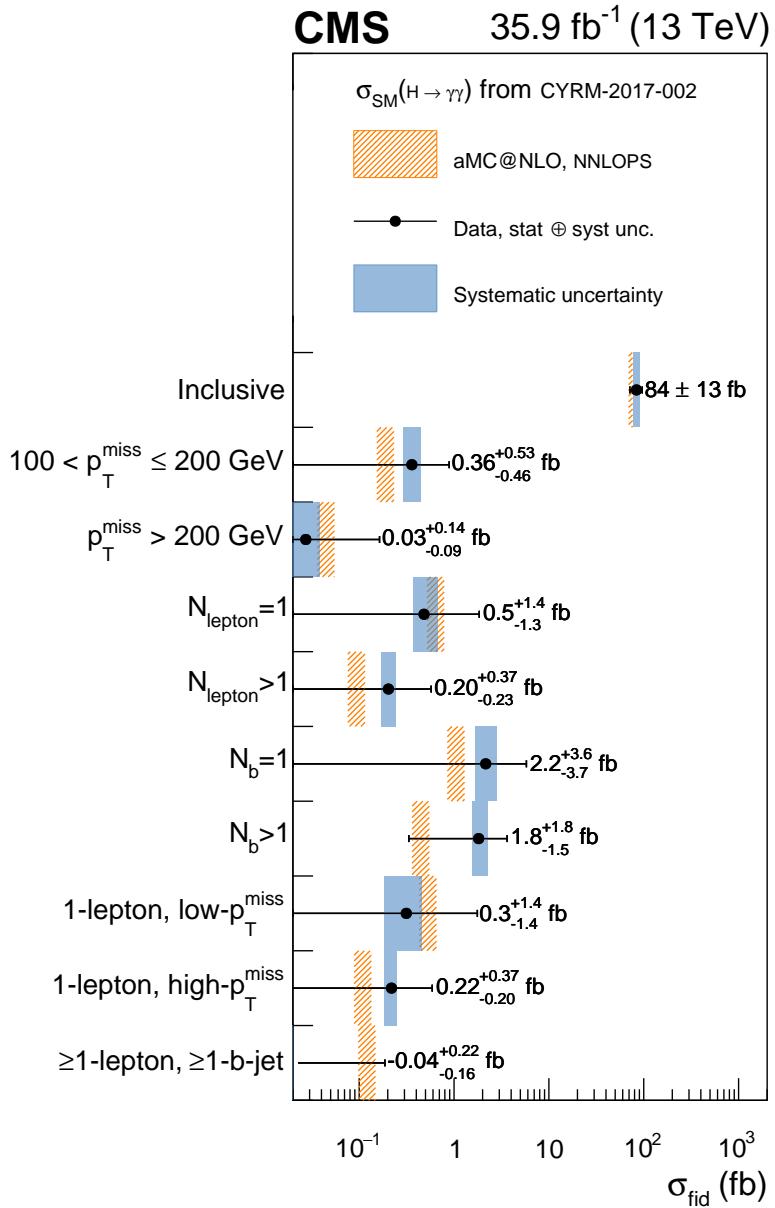


Figure 6.15: The measurement of the differential cross section for different regions of the phase space, listed on the vertical axis. The black points indicate the data, with the black error bars showing the 1 standard deviation total uncertainty, while the blue band indicates the systematic component. The measurements are compared to the theoretical predictions (orange area), normalized to the predictions from Ref. [35].

Part IV

QUALIFICATION OF DETECTOR MODULES FOR THE PHASE 1 UPGRADE OF THE CMS PIXEL DETECTOR

This Part describes the Phase 1 Upgrade of the CMS detector, the detector modules and the tests they underwent prior installation.

SILICON TRACKING DETECTORS

This chapter briefly presents silicon tracking systems in High Energy Physics (HEP) experiments and then concentrates on the role of silicon as a sensitive material.

7.1 TRACKING DETECTORS

The use of tracking detectors provides valuable information on the trajectories of charged particles inside the detector. Such detectors typically have a multi-layer structure, where each layer is equipped with semiconductor devices sensitive to the passage of particles, thus able to provide the measurement of a point of the particle trajectory. By generating a homogeneous magnetic field in the tracking volume, charged particles trajectories are bent, allowing the determination of their momentum via the measurement of the sagitta of the reconstructed track. Silicon is the most common material in HEP used for this purpose, in different geometrical configurations. The use of silicon strips, with a fine sensor segmentation only in one dimension of the sensor plane, while allowing a reduction of costs, requires two consecutive layers, rotated with respect to each other, to provide a good resolution estimate of a point of a trajectory. Such a layout is used in the CMS Tracker detector, as introduced in Section 2.2. As the distance to the interaction point decreases, higher particle fluxes require finer segmentation in both dimensions of the sensors, to cope with the increased hit density. For these reasons, the inner layers are equipped with sensors segmented in *pixels*. High-resolution performance of the inner tracking system, joint with the 3.8 T magnetic field in the CMS volume, enable precise momentum measurement as well as efficient identification of collision (primary) vertices and of vertices from particle decays (secondary).

7.2 INTERACTION OF CHARGED PARTICLES WITH SILICON

The signal generation in the sensor layers is governed by the interaction of charged particles with the material through ionization energy deposited in the material following this interaction. The ionization energy deposit gives rise to the formation of electron-hole pairs, with the average energy needed to create a pair of 3.6 eV in silicon, dictated primarily by the band gap of silicon. The mean rate of energy deposited by ionization is well described by the Bethe-Bloch formula, given in Equation 7.1, for $0.1 < \beta\gamma < 1000$ [95]:

$$\left\langle -\frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (7.1)$$

where W_{max} is the maximum energy transfer in a single collision, I is the mean excitation energy, $K = 4\pi N_A r_e^2 m_e c$, $\delta(\beta\gamma)$ is the density effect correction to ionization energy loss, and A and Z are the atomic and the mass number of the material, respectively.

As can be seen in Figure 7.1, this equation describes a falling energy loss as a function of $\beta\gamma$, until a minimum is reached for $\beta\gamma \approx 3$. A particle with kinematic properties

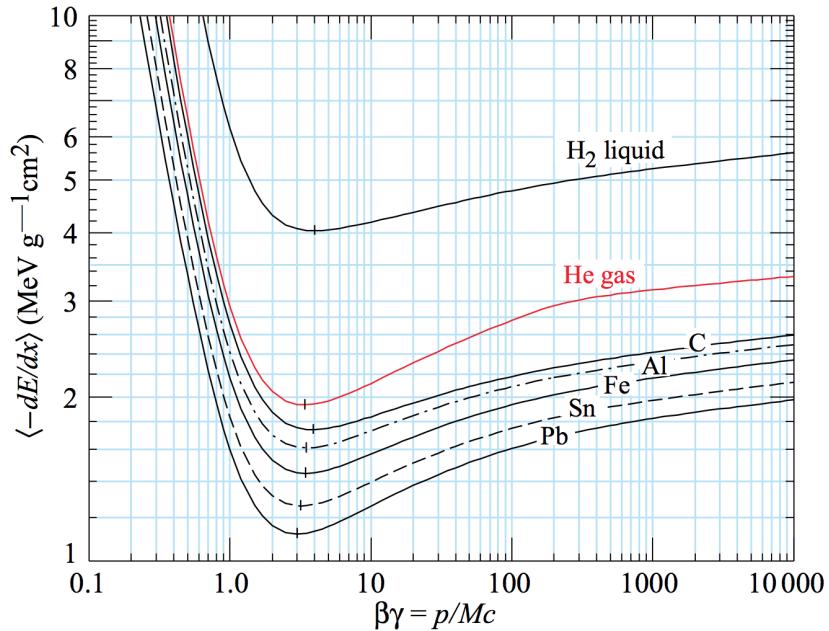


Figure 7.1: The mean rate of energy loss by ionization in matter in different materials. The curves follow the Bethe-Bloch equation, given in 7.1 [95].

and mass placing it in this range is called Minimum Ionizing Particle (MIP). After the minimum, the curve shows a relativistic slow rise at high $\beta\gamma$. This description can be considered universal in terms of $\beta\gamma = p/m$, where p is the momentum of the particle and m its mass, and furthermore it is qualitatively similar in all materials, with the rate of energy loss decreasing with increasing Z .

The Equation 7.1 describes the mean value of the energy loss by ionization. However, the energy depositions in thin layers of material, as those used in silicon detectors ($\sim 50\text{--}300\mu\text{m}$), are subject to large fluctuations, driving the mean of the distribution towards higher values. The fluctuations are caused by rare high-energy-transfer collisions between the incident particle and one of the shell electrons of the absorber [95], which can result in the dislocation of the electron (δ ray) that in turn creates electron-hole pairs. The resulting energy deposition distribution can be approximated by a Landau-Vavilov distribution [70, 101], which has a Gaussian behavior at low energies, whose most probable value corresponds to the mean value estimated by the Bethe-Bloch formula, but with a long tail at high-energy values. The most probable value of the energy deposition for a MIP interacting in a silicon thickness of $285\mu\text{m}$, as used in the CMS Pixel detector, corresponds to the creation of about 24k electron-hole pairs.

7.3 THE P-N JUNCTION AND SIGNAL COLLECTION

The signals created by the ionization of traversing particles are order of magnitudes smaller than the charge carrier densities induced by thermal excitation in silicon, making pure silicon unusable as a sensitive material.

The reduction of the charge carrier densities exploits the properties of silicon when impurities are added to the lattice: the addition of atoms from the Group III of the periodic table, i.e. with one electron less than silicon, leads to the formation of a *p-doped* lattice; similarly, the contamination with atoms of the Group V, i.e. with one electron more than silicon, yields *n-doped* silicon. Silicon with n-doping shows an increased abundance of free negative charge carriers, while p-doped silicon shows a lack of electrons, interpreted as an increased density of positive charge carriers.

The juxtaposition of n-doped and p-doped silicon layers leads to the formation of an electric field, due to the different charge densities in the two materials. Under its effect the charge carriers migrate in the adjacent semiconductor, creating an intermediate zone, across the contact surface between the two semiconductors, electrically neutral due to charge recombination, called depletion zone. As a result of the diffusion, space charge densities with opposite sign due to charged ions accumulate in the two semiconductors, and generate an electric field that goes against the diffusion and eventually stops it. The penetration depth of the depletion zone in each semiconductor depends on the doping concentrations and can be asymmetric. The depletion zone size can then be influenced by applying an external voltage source, called *bias voltage* (V_b). A negative bias voltage, which corresponds to a higher voltage on the n-doped side of the junction, produces a widening of the depletion zone. The depletion zone operates as a capacitor between the two semiconductors, whose capacitance is proportional to $A/\sqrt{V_b}$, where A is the area of the junction.

The presence of a depletion zone creates favorable conditions for the collection of the signal of a passing particle, which can then be isolated from the intrinsic charge carrier densities. Nevertheless, the creation of charge carriers in the depletion zone by thermal excitation is still possible and leads to the presence of a so-called *leakage current* (I_{leak}), which depends on the temperature T as

$$I_{leak} \propto T^2 e^{E_g/2kT}, \quad (7.2)$$

where E_g is the band gap energy (1.12 V in silicon) and k the Boltzmann constant. The I_{leak} decreases by roughly a factor of 2 every 7 K of temperature drop. This dependence allows to significantly decrease the leakage current by means of even moderate cooling of the detector, increasing the signal-to-background ratio.

The charges created by a traversing particle move under the influence of the electric field across the semiconductor junction, generating an induced charge on the electrodes, where a current is measured. With a drift velocity of about 10^7 cm/s, the CMS sensor thickness of $285\mu\text{m}$ allows the signal collection within few ns, well within the LHC bunch crossing of 25 ns. In addition to the charge drift, a random movement due to thermal diffusion is induced, spreading the drift paths of the charge carriers and creating a cloud with a width proportional to the square root of the time of drift, amounting to a few μm . In CMS, a larger lateral deflection is induced by the presence of a magnetic field in the tracking volume, which bends the trajectories of the charge carriers during their drift. For drift paths orthogonal to the magnetic field, a Lorentz force is induced in the lateral direction under the angle θ_L , called *Lorentz angle*, which is proportional to the magnetic field B and the mobility of the carriers. Depending on the magnetic field strength, the sensor pixel pitch, and the sensor thickness, this can cause carriers drifting towards the

electrode of a neighboring pixel with respect to the one where the charge signal was generated, leading to the so-called *charge sharing* effect. This will have an important positive effect in the resolution performance of the detector, as described in Section 7.5.

7.4 RADIATION-INDUCED EFFECTS

High-fluence of particles ($\mathcal{O}(10^{15})$ particles/ cm^2) through the silicon sensor is responsible for radiation damage. Non-ionizing and ionizing radiation are responsible for two distinct damage mechanisms, respectively: displacement damage in the silicon lattice (bulk damage) and accumulation of space charge.

Atoms can be displaced by Non-Ionizing Energy Loss (NIEL) in the sensor, which dislocates them from their location in the lattice. Depending on the kinetic energy transferred to the atom in the collision, the atom can in turn dislocate other atoms, creating clusters of lattice damage. This effectively modifies the charge carrier densities and creates interstitials and impurities in the lattice, leading to the formation of intermediate states between the valence and the conduction band of the silicon and thus increasing the leakage current by favoring the promotion of electrons to the conduction band. As a result, a larger bias voltage is needed to extend the depletion zone to the full sensor (fully depleted sensor). Furthermore, bulk damage causes the *type inversion* of n-doped silicon, by first reducing the effective n-doping concentration and then leading to a p-type doping of the silicon.

Accumulation of space charge occurs at the interfaces between silicon and other materials, mainly insulation/oxide layers, due to ionizing radiation damage, modifying the charge carrier densities.

The concentration of charge carriers varies as a function of time as a result of so-called *annealing* processes: on time scales of days, thermal migration of impurities and interstitials leads to recovery of the lattice defects, reducing the charge carrier densities through beneficial annealing; on longer time scales, the reverse annealing dominates leading to an increase of the charge carrier densities. Both annealing processes can be attenuated by keeping the sensors at temperatures below 0°C. While this approach, together with the use of an increased bias voltage, can partially recover radiation-induced effects, the sensor performance degrades on the long term due to irradiation.

A detailed description of the effects of irradiation on the CMS pixel detector Read Out Chip (ROC) and sensor can be found in [62].

7.5 PIXEL SPATIAL RESOLUTION

Once the signal is collected by electrodes, it is used to estimate the position of the particle hit that originated it. For a sensor segmented in square pixels with pitch p with a binary read out scheme (i.e. without Pulse-Height (PH) information concerning the hit), if a pixel records a signal its geometrical center is taken as a position measurement. The true position of the hit, though, is unknown and has to be assumed to be uniformly distributed on the surface of the pixel. The variance of the measured position with respect to the true one can thus be calculated and found to be $\sigma^2 = p^2/12$, corresponding to a position resolution σ in each direction. This approach to the position measurement thus is limited by the pixel pitch, which in the CMS detector is 100 μm in the $r-\phi$ direction and 150 μm in the z direction.

This estimate can be improved thanks to the charge sharing effect, described in Section 7.3, and to Pulse-Height information. The position estimate in this case is found as the barycenter of the hit cluster, weighting each pixel by the PH magnitude. This approach allows to obtain resolutions of $10 \mu\text{m}$ in the $r-\phi$ direction and $25 \mu\text{m}$ in the z direction, in the barrel section of the CMS pixel detector, to be compared to $100 \mu\text{m}/\sqrt{12} \approx 30 \mu\text{m}$ and $150 \mu\text{m}/\sqrt{12} \approx 45 \mu\text{m}$, respectively.

7.6 HYBRID PIXEL CONCEPT

The pixelated sensor is connected to a readout device in order to extract and process the charge signals deposited in each pixel. The pixels are made into the sensor bulk as implants of highly-doped silicon, whose surface is then metallized to allow connection to the ROC. Each pixel is individually connected to the smallest unit of the readout, the so-called Pixel Unit Cell (PUC). The connection between the pixel implant and the PUC is performed by means of a bump bond, with a diameter of few tens of μm , under high pressure and temperature conditions. This connection allows the readout of the signal from the sensor and its subsequent processing. A sketch of this configuration, called *hybrid pixel detector layout*, is given in Figure 7.2 for a single pixel (a) and for a full ROC (b).

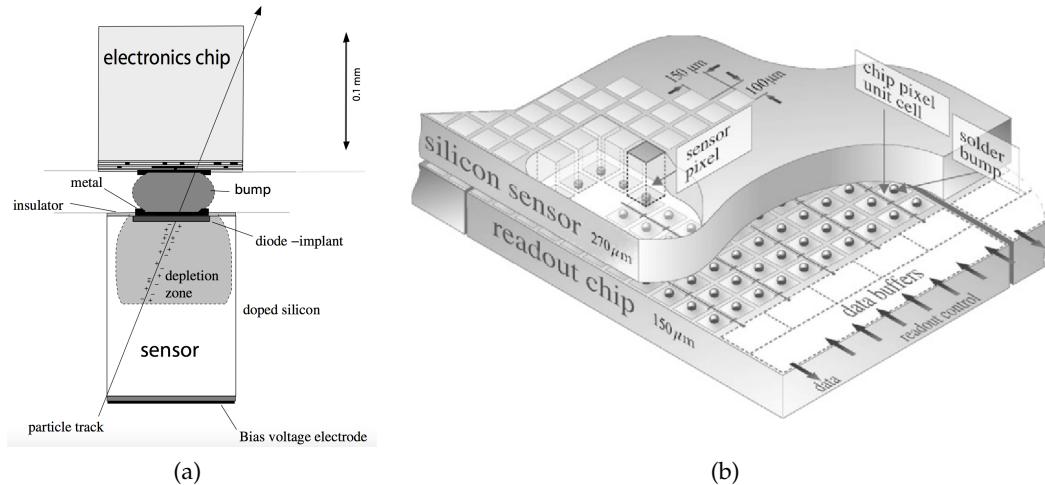


Figure 7.2: The hybrid pixel detector layout, sketched for a single pixel (a) [84] and for a full readout chip (b) [17].

The basic elements of the signal processing chain include a pre-amplifier and a shaper, which increase the signal-to-noise ratio, a comparator, which allows to suppress signals smaller than a given threshold, and an Analog to Digital Converter (ADC) for digitization of the PH information. In Section 8.2, it will be highlighted how the latter is a key difference between the upgrade and the previous CMS pixel detector.

In this chapter, the main motivations and features of the Phase 1 Upgrade project of the CMS pixel detector are discussed.

8.1 MOTIVATION

The pixel detector originally installed in CMS, also referred to as *phase-0* detector, provided excellent performance until it has been decommissioned after the 2016 data-taking. In its place, a new detector, or *phase-1* detector, has been installed. The detector upgrade has been foreseen and designed to cope with the increasing instantaneous luminosity of the LHC, exceeding the design performance of $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ by more than 50% and planning to go even further in the future. The increased instantaneous luminosity results in a higher number of PU collisions per bunch crossing, going from an average of 25 in the design conditions to an average of 50 at $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a bunch spacing of 25 ns [37]. The new detector allows operation in such a severe environment, while recovering the hit inefficiency that the previous detector would suffer, amounting to 4% (15%) at (twice) the LHC design luminosity in the innermost layer [37].

8.2 DESIGN OF THE NEW PIXEL DETECTOR

The new detector, while introducing several and crucial novelties, shares the basic concept and structure with the phase-0 detector. For this reason, the new system is indicated as an evolutionary upgrade of the previous one.

The new system inherits from the former one the $285\mu\text{m}$ -thick sensor and the $100 \times 150\mu\text{m}^2$ pixel size. The ROCs are manufactured with the 250 nm Complementary Metal-Oxide Semiconductor (CMOS) standard and connected to an array of 52×80 pixels. ROCs are arranged in 2×8 arrays which share powering and readout, constituting the detector basic element, the *module*, which will be described in more detail in Section 8.5.

At the same time, the new system benefits from numerous improvements:

- the new system features an additional ($3 \rightarrow 4$) layer in the barrel section (BPix) and an additional ($2 \rightarrow 3$) disk in the each endcap section (FPix), to ensure 4-hit coverage up to $|\eta| = 2.5$, improving on the 3-hit coverage of the previous system. Furthermore, the innermost layer is brought closer to the interaction point, with a radius of 29 mm, compared to the 44 mm of the previous system, improving the primary and secondary vertex identification. The fourth layer of the Bpix extends the pixel coverage much further from the interaction point compared to the previous detector, significantly reducing the gap between the pixel and the tracker systems and decreasing the combinatorics in the track reconstruction. The additional layers lead to an increase in the total number of channels from 66 to 124 millions. A comparison of the geometrical layouts of the phase-0 and phase-1 detectors is given in Figure 8.1;

- a new cooling system, based on an evaporative bi-phase CO₂ process, improves the cooling performance, allowing lower temperatures and thus extending the operational lifetime of the detector, by mitigating of the rise of sensor leakage current after irradiation, as explained in Sections 7.3 and 7.4. The new system replaces the previous mono-phase C₆F₁₄ system. The CO₂ system, together with an optimized layout of services, which have been moved out of the tracking volume, and lightweight supporting structure, provides a similar material budget up to $|\eta| \approx 1.2$ and a reduced material budget at higher pseudorapidities when compared to the phase-0 detector. A comparison of the material budgets between the previous and the new detector is shown in Figure 8.2.
- the ROC, named *psi46digV2.1respin*, benefits from a significantly new design which features on-chip digitization of the PH with an 8-bit ADC, compared to the analog readout of the previous chip. This brings the readout bandwidth to 160 Mbits/s, replacing the previous 40 MHz analog readout. Furthermore, hit inefficiencies are reduced thanks to enlarged buffer sizes for hit and timestamp, which were found to be limiting factors in the performance of the analog ROC. The ROC design and readout scheme are discussed in more detail in Section 8.3. While the *psi46digV2.1respin* ROC is installed in the FPix and the three outer layers of the BPix (L₂, L₃, L₄), the highest expected pixel hit rate in L₂ being 120 MHz/cm², for the innermost layer (L₁) a dedicated ROC, *PROC600*, has been developed to cope with the even higher data flux. The PROC600 ROC achieves hit efficiencies larger than 98% up to rates of 600 MHz/cm², thanks to the readout of 2 × 2 pixel clusters in place of the sequential readout of all pixels used by the *psi46digV2.1respin* ROC described in Section 8.3;
- an improved comparator in the PUC, described in Section 8.3.1, allows the setting of a much lower threshold of about 1750e⁻, compared to the 3500e⁻ of the previous detector, while keeping the noise level at about 130e⁻.

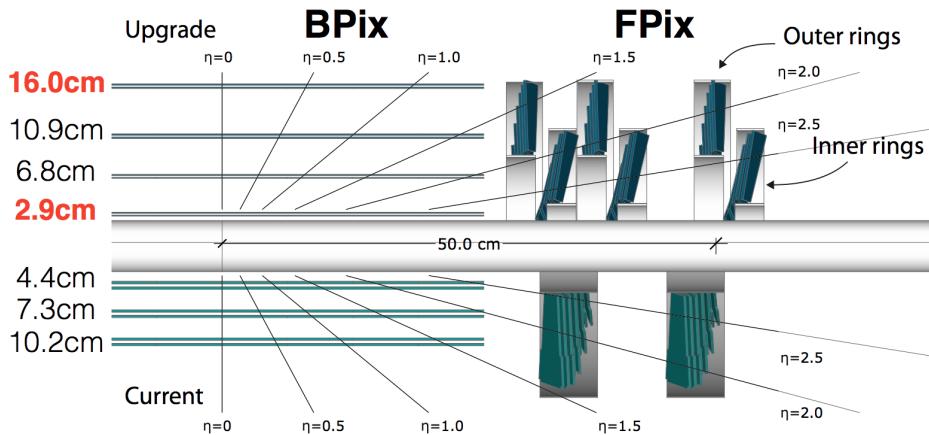


Figure 8.1: The comparison between the geometrical layouts of the previous (bottom) and the new (top) CMS pixel detector [97]. Modified from [37].

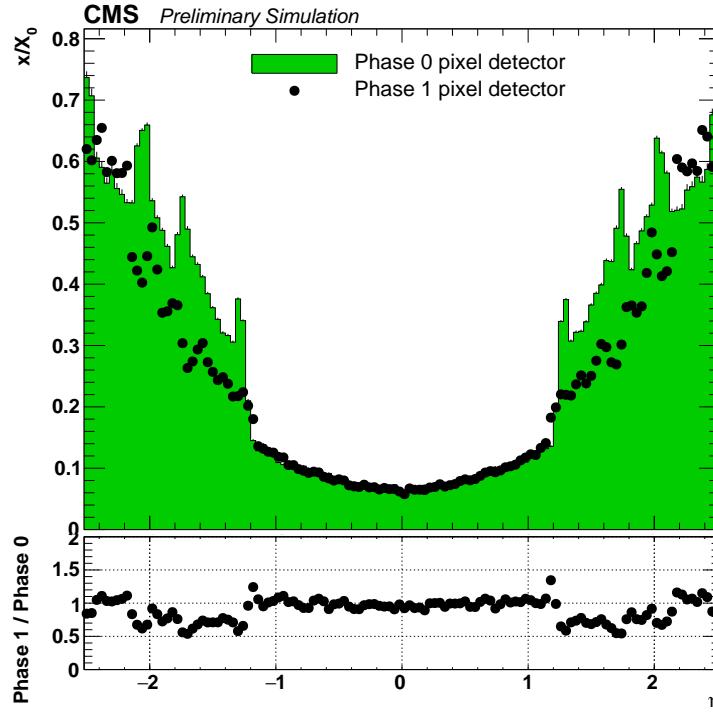


Figure 8.2: The material budget, in units of radiation lengths, is compared between the phase-0 (green histogram) and the phase-1 (black dots) pixel detectors [37].

8.3 THE DIGITAL READOUT CHIP

The readout chip has to meet the main requirements of recording the hit position and charge for all hit pixels, within the LHC bunch crossing of 25 ns and to store them on-chip during the CMS L1 trigger latency of 3.2 μ s, as described in Section 2.2.5. The digital readout chip has been developed starting from its analog predecessor used for the phase-0 pixel detector, the *psi46V2* chip [66]. In this section, its main features and the key elements of its working principles are outlined. A detailed description of the digital ROC design can be found in [64].

Each ROC covers the readout of 4160 pixels, arranged in 80 rows and 26 double-columns. The footprint of the chip is slightly larger than the bare size of the sensor covered, in order to host the so-called *double-column periphery* and the control interface. The double-column periphery holds the data and timestamp buffers for each double column, while the control interface contains several control and supply circuits, including control Digital to Analog Converters (DACs), voltage regulators, and control registers. A schematic of chip footprint is shown in Figure 8.3.

8.3.1 The PUC and the ROC periphery

The smallest independently working component of the ROC is the PUC, which is sketched in the top part of Figure 8.4.

The signal created in the sensor is collected via the bump pad and enters the readout chain. Alternatively, an internal calibration signal (V_{cal}) can be injected through the bump,

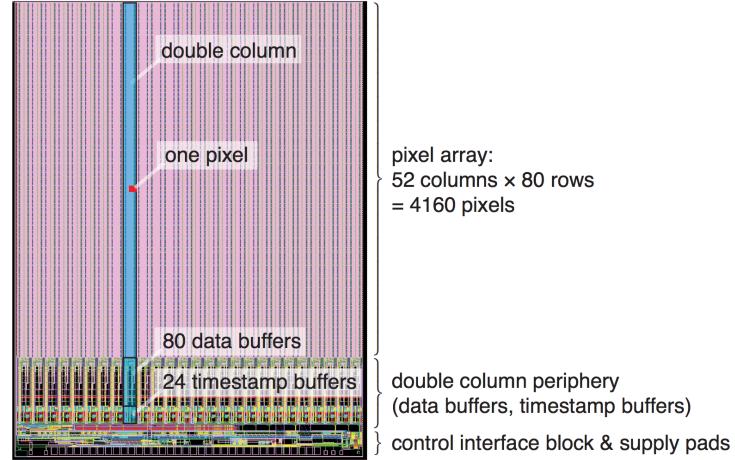


Figure 8.3: The schematics of the digital ROC. [55]

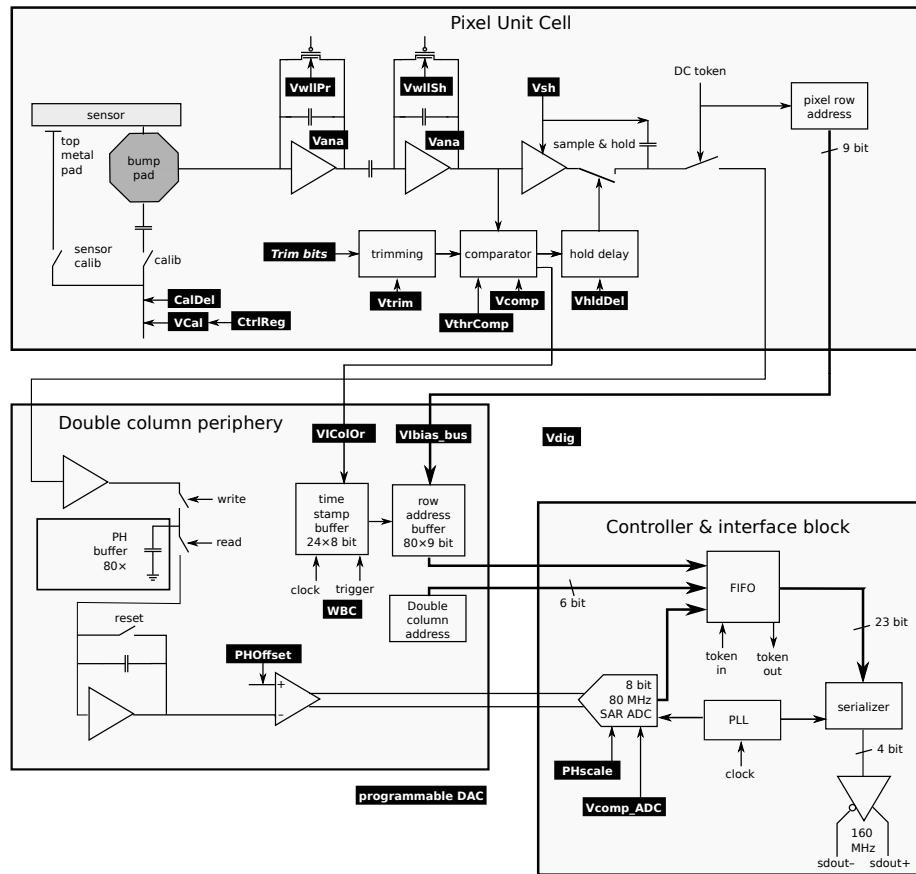


Figure 8.4: The schematics of the PUC (upper part of the figure) and of the double-column periphery and control interface of the ROC (lower part) [55].

allowing the creation of a signal without the need for real hits. The amplitude of the V_{cal} signal is regulated by an 8-bit DAC, which can be set either in low- or high-range mode

via a dedicated register. This allows to extend the range of injectable charges, with a factor 7 in charge occurring between the DAC value in low- and high-range. A V_{cal} unit in the low-range corresponds roughly to an injected charge of $50e^-$. The signal can also be routed, rather than to the bump bond, to the top metal pad and then, via the capacitance of the sensor-ROC gap, to the sensor. This second mechanism of V_{cal} injection is exploited to verify the quality of the bump bonding connection, as explained in Section 9.1. The signal then enters the pre-amplifier and the shaper, which enhance the signal-to-noise ratio and prepare the signal for optimal sampling. Both these circuits are powered by the analog voltage line of the chip, controlled by the V_{ana} DAC, with a nominal working point of 24 mA. In order to operate the detector in zero-suppression mode, the signal is compared to a threshold in the comparator. The setting of the threshold for all PUCs is regulated by the $V_{thrComp}$ DAC. The threshold can be further tuned by a per-pixel 4-bit DAC, called **TrimBits**, allowing the fine adjustment of the threshold for individual pixels. The DAC V_{trim} regulates the size of the **TrimBits** effect on the threshold. The comparator can be switched off, effectively disabling the pixel, by means of a mask bit. Only signals exceeding the threshold are passed to the sample-and-hold circuit, where the signal is sampled, and counted as pixel hits. Once a hit is collected, the pixel becomes unresponsive to further hits and the corresponding double-column is notified with a fast OR signal which is detected by the double-column periphery, regulated by the **VICol0R** DAC. This initiates the double-column drain mechanism: a timestamp is created and stored in the double-column periphery buffer, while a readout token is sent through the 2×80 pixels. Pixels without hits simply pass the token to the next pixel, while pixels with a hit transfer the PH charge into the data buffer, while the token chain is stopped. The column drain requires two clock cycles for the transfer of each pixel hit, plus one clock cycle to synchronize the data transmission with the 40 MHz clock. The readout of the double-columns happens independently, increasing the hit collection efficiency. The bottlenecks for the readout of pixel hits in the phase-0 chip were found to be the size of the timestamp and data buffers, which have been significantly increased in the digital ROC, from 32 to 80 bits and from 12 to 24 bits, respectively.

The PH information is digitized on-chip by an 8-bit ADC. This constitutes an important improvement in the chip design, which allows to replace the 40 MHz analog readout with a 160 Mbit/s digital readout. The output of the PH ADC can be shifted by an amount controlled by the **PhOffset** DAC, while its gain is regulated by the **PhScale** DAC. These two DACs offer control on the PH gain and offset at a ROC level, since only one PH ADC is hosted in the control interface of the ROC. This means that the gain and offset settings have to be carefully tuned to account for pixel-to-pixel variations in PH response and allow optimal PH dynamic range for all of them.

The chip also features a circuit providing an output voltage independent of temperature and supply voltage. This output voltage, called *bandgap* voltage, is used as a reference voltage for DACs and ADC on the chip. The stability of the circuit exploits canceling dependencies in the electrical properties of ohmic components and of the p-n junction [53].

8.4 READOUT FORMAT

The data format of the ROC is shown in Figure 8.5. The start of readout data of the ROC is marked by a 12-bit ROC header, consisting of a 0, eight consecutive 1, and again a 0.

The last two bits of the header are reserved for the transmission of information through a *readback* mechanism, which allows access to on-chip measurements and information through a slow ADC. This information is read out one bit per event via the *readback bit* (D), placed at the end of the ROC header. Since only one bit is transmitted per event, the readout of 16 consecutive events is needed to form a full readback word. The word uses 8 bit for the encoding of the information, while the other 8 are reserved for the ROC address. The second to last bit (S) in the ROC header, immediately preceding the readback bit, is used to mark the start of a new word. Multiple on-chip measurements and information can be obtained through this mechanism, which is discussed in more detail in Section 9.1.8. The ROC header is followed by 6 (9) bits to encode the pixel row (double-column) address, a 0, and 8 bits for PH information, for a total of 24 bits per every pixel hit in the event. The token out signal is raised two clock cycles before the end of the readout, to continue with the readout of the next ROC in the module token chain.

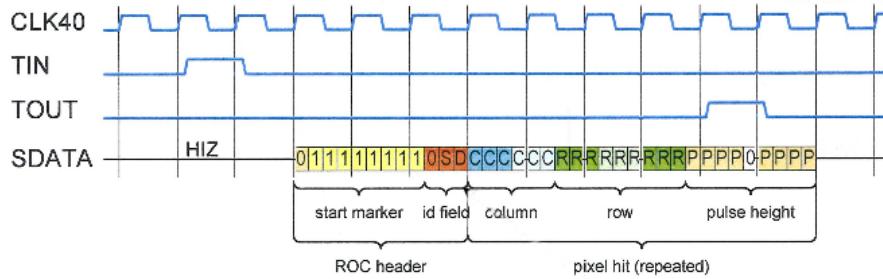


Figure 8.5: The ROC data format. The ROC header contains two trailing bits used to transmit the readback information (D) and to mark the start of a new readback word (S). Modified from [54].

The ROC injects its readout in a 160 Mbit/s data stream. Several of these readout streams are present on each module, regulating the readout of a sub-group of ROCs each. A dedicated chip on the module, the Token Bit Manager (TBM), has multiple cores, each orchestrating the readout and the token chain of a ROC sub-group. The TBM is also in charge of converting each pair of the 160 Mbit/s data streams in 400 Mbit/s data streams, by means of a custom 5-4 bit encoding mechanism, and of transmitting the module readout via the readout cable. The number of readout links for each module and the number and size of the ROCs sub-groups sharing the same readout token varies depending on the position of the module in the detector and the consequent particle rate: modules installed in the third and fourth layer of the Bpix and in the Fpix host a TBM with two cores, each controlling a group of 8 ROCs, leading to a single 400 Mbit/s data stream per module; the L2 modules use two 400 Mbit/s data links, managed by a different type of TBM, as a result of four token chains of four ROCs each; for the even higher rates experienced in the L1, two TBMs are installed on each module, using four readout links for data transmission.

The TBM data are sent via optical readout links to Front-End Drivers (FEDs) (Front-End Drivers), placed off the detector. The FED modules are responsible for decoding the data in order to be used by the HLT system and to be stored. Also the Front-End Controllers (FECs) (Front-End Controllers) are placed off the detector and take care of sending commands and signals, including clock and trigger, and configuring the detector. In the phase-1

upgrade, the FED and FEC systems have been replaced to handle the increased number of channels and to comply with the digital format of the readout. At the same time, a migration from the VME technology to the μ TCA technology has been performed.

8.5 DETECTOR MODULES

The phase-1 pixel detector is built of 1856 modules, of which 1182 constitute the 4 layers of the Bpix and 672 the disks of the FPix. The same type of modules is used everywhere in the detector, except for the innermost layer of the Bpix, where a slightly different layout is used. In this Section, the module design and its components are discussed, mainly referring to the modules used in L2-4 and in the endcaps. Differences in the L1 modules are highlighted where relevant.

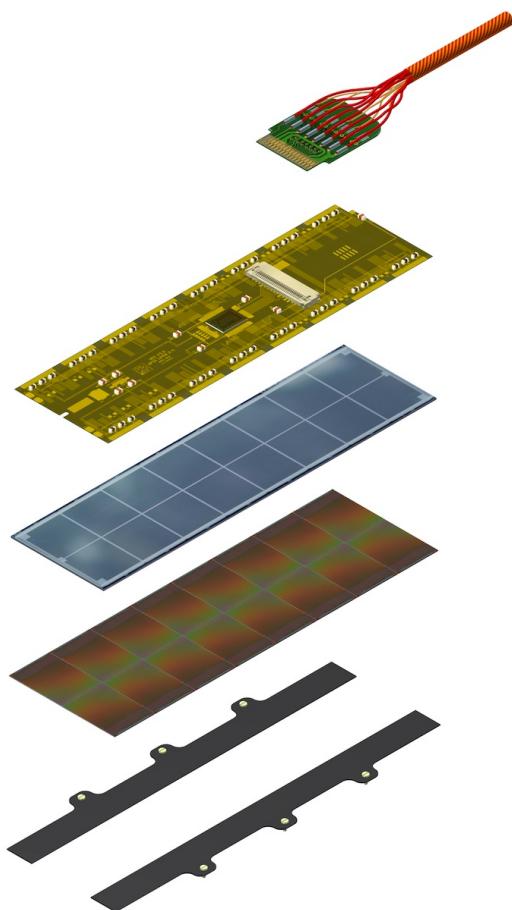


Figure 8.6: Drawing showing separately the components of the modules installed in L2-4 and in the FPix, from top to bottom: the module cable, the HDI with the TBM, the silicon sensor, the ROCs array, and the mounting strips. The FPix modules use a modified mounting system, while L1 modules employ two TBMs and the mounting strips are omitted to reduce the material budget.

An exploded drawing of the L2-4 modules is shown in Figure 8.6. The core feature of the module is the $285\mu\text{m}$ -thick silicon sensor, covering an active area of $16.2 \times 64.8\text{ mm}^2$. The sensor allows the recording of signals induced by the passage of charged particles,

following the principles illustrated in Section 7.2. However, a more complicated structure than the simple pn-junction is used, in order to allow operation of the sensor after irradiation: in an n-doped substrate of the sensor, implants with significantly higher n-doping concentrations (n^+) are installed on one face of the sensor, while the backplane is positively doped. This technology, sketched in Figure 8.7, is referred to as n^+ -in-n silicon sensors. The depletion zone is formed at the pn-junction between the substrate and the backplane. On unirradiated sensors, the depletion zone is extended to the pixel implants by applying a negative bias voltage, allowing the collection of electrons at the pixel implants. After irradiation, though, the n-doped bulk will undergo type-inversion, as discussed in Section 7.4, and convert to p-type. In this configuration, the pn-junction and thus the depletion zone extends from the interface between pixel implants and the bulk and positive bias voltage is applied to extend it and improve charge collection efficiency. It is important to note that simple p-doped implants in an n-doped bulk would not allow the creation of a depletion zone in the proximity of the pixels after type-inversion of the bulk.

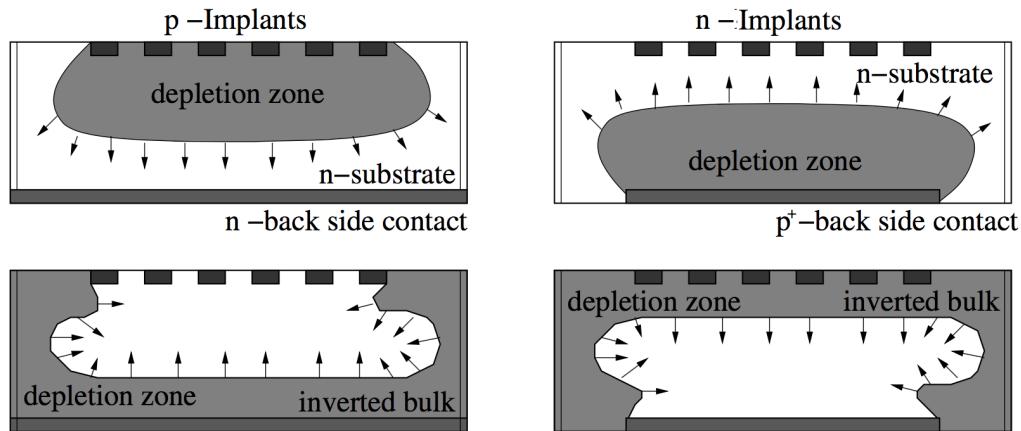


Figure 8.7: The sensor doping structure before (top) and after (bottom) irradiation, for p-type implants in n-substrate (left) and for n^+ -type implants in n-substrate (right). Modified from [84].

MODULE PRODUCTION

This chapter presents the procedures developed to test the modules produced for the Phase 1 Upgrade of the CMS Pixel detector. An overview of the module production is given, with a focus on the modules assembled for the second layer construction.

9.1 MODULE FULLTEST

A suite of calibration and testing procedures is set up to characterize modules and assess their quality and suitability for installation. The *Fulltest* routine is composed by several dedicated sub-routines, aimed at verifying a specific functionality or at calibrating a given parameter. The algorithms are designed to be robust against anomalous module behavior and to be fast and reliable, in order to be employed in the testing of the mass production of modules for the detector construction. In the following, all sub-routines are reviewed, following the order in which they are executed. More details about individual tests can be found in [44].

9.1.1 *Pretest*

The *Pretest* is meant to verify basic functionalities of the module and the ROCs and to put them in an operable state. The programmability of the ROC is tested by changing the V_{ana} DAC value between the default value and zero and by checking if any change is observed in the analog current. If this check fails, it is assumed that no successful communication can be established, the Fulltest is interrupted, and the module discarded. During the Pretest, the value of the V_{ana} DAC is tuned such that the analog current drawn by each ROC is 24 mA, corresponding to the optimal operational value of the PUC pre-amplifier. Then, the threshold of the comparator, controlled by the $V_{thrComp}$ (for which higher DAC values correspond to lower threshold settings), and the delay of the calibration pulse with respect to the clock signal, $CalDel$, are optimized, by choosing the central value in the region where these two parameters provide full efficiency for a randomly chosen pixel. The typical hit efficiency plot as a function of ($V_{thrComp}, CalDel$) is shown in Figure 9.1. The shape of the distribution, often referred to as *tornado-figure*, is due to the fact that larger-amplitude signals are faster to cross the threshold, while smaller-amplitude signals are slower. In the same way, this effect corresponds to the full-efficiency window moving to smaller values of $CalDel$ for increasing settings of the threshold, i.e. lower values of $V_{thrComp}$. First, the optimal value of $V_{thrComp}$ is chosen to be 50 DAC units below the upper edge of the efficiency figure. Then, the central value of the efficiency plateau for that $V_{thrComp}$ value is chosen as the $CalDel$ setting. It should be noted that, for $V_{thrComp}$ settings higher than the upper edge of the efficiency figure, the efficiency is zero irrespective of the $CalDel$ setting. This is due to the fact that for such low thresholds the pixel is constantly firing on electronic noise and becomes unresponsive.

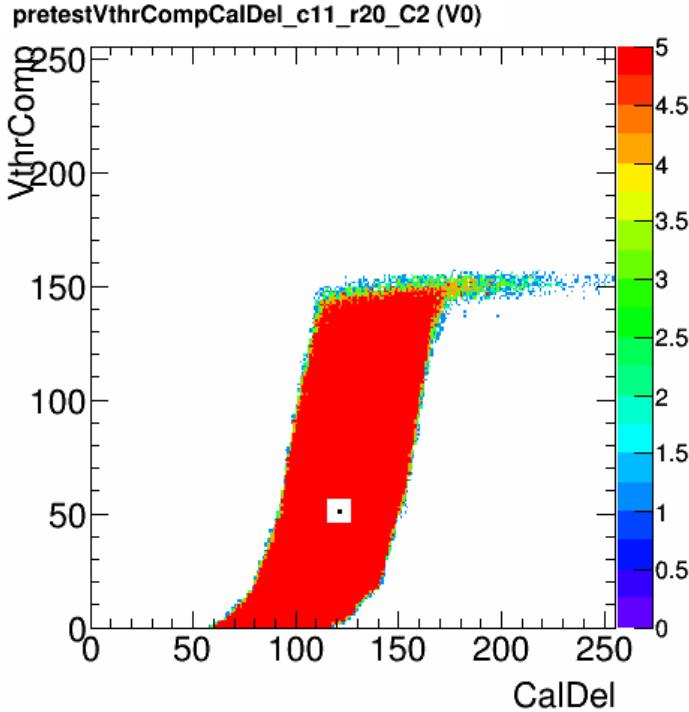


Figure 9.1: The hit efficiency map as a function of V_{thrComp} and CalDel DAC settings. The white square with the black dot indicates the optimal point chosen during the Pretest.

9.1.2 PixelAlive

Once the optimal operational parameters are set for each ROC, the *PixelAlive* test measures the hit efficiency for all pixels, by sending 10 calibration signals to each pixel and counting how many hits are recorded. Also the masking mechanism is tested for each pixel. This mechanism allows to disable pixels by switching off their comparator. If, when masked, a pixel shows an efficiency larger than zero, its masking bit is considered to be faulty. Furthermore, the pixel address format is checked for errors. Pixels that do not show a 100% efficiency, whose masking mechanism is malfunctioning or have address decoding errors are identified and considered defective.

9.1.3 Noise via S-curve

The noise for each pixel is measured in the *S-curve* test, with the goal of identifying noisy pixels, so that they can be masked and excluded from the readout. The method used in this test and described below is rather general and it is used in other tests described later. The test measures the efficiency curve as a function of V_{cal} . In absence of noise, such curve would be a step function, assuming only the following values: 0 for signals smaller than the pixel threshold and 1 for signals above the threshold. The presence of noise, though, smears the sharp transition, giving to the curve a shape similar to that of the letter *S*, hence the name of the test. By fitting the curves with an error function, the width parameter of such function is taken as the measurement of the pixel noise. Noisy

pixels ($> 400e^-$) and pixels with an anomalous low noise ($< 50e^-$) are identified and considered defective.

9.1.4 *BumpBonding*

The quality of the bump-bonding connection between the silicon sensor and the ROC is tested. The *BumpBonding* test takes advantage of the possibility to send calibration signals to the *top metal pad* shown in the PUC diagram in Figure 8.4, instead of passing directly through the *bump pad*. The signal then passes into the silicon sensor, via the capacitance of the sensor-ROC gap, and can then be read out if the bump bond is connected. The identification of anomalous, faulty or missing bump bonds relies on the measurement of the threshold setting needed to see a signal of fixed amplitude through the mechanism described above. For pixels with good bump bonds, these measurements have a Gaussian distribution. Pixels showing thresholds more than 5 standard deviations away from the mean of the distributions are considered to have a defect in the bump bond connection.

9.1.5 *Trimming*

The threshold of the pixels is set and unified across the chip by the *Trimming* test. This step is important in order to have a defined threshold, in units of the internal calibration signal V_{cal} , and a uniform level of zero-suppressions for all pixels.

The threshold is influenced by the settings of two ROC DACs: $V_{thrComp}$ and V_{trim} . The former sets a baseline threshold, with higher values corresponding to lower charge thresholds, while the latter sets the scale of impact of the `TrimBits` setting on the threshold. The `TrimBits` allow a per-pixel tuning of the threshold via a 4-bit DAC: the effect of the `TrimBits` is maximum, i.e. the threshold is maximally reduced, when they are set to 0. The effective action of the `TrimBits` is multiplied by the value of V_{trim} . The effect of the `TrimBits` can be turned off by setting them to 1111 = 15 or, for all pixels, by choosing $V_{trim} = 0$.

The Trimming algorithm first optimizes the setting of the $V_{thrComp}$ DAC. In order to do so, V_{trim} is set to zero and a calibration signal corresponding to the target threshold (by default, $V_{cal} = 35$, low-range) is injected and the S-curves for all pixels are recorded with respect to $V_{thrComp}$. The result of this S-curve measurement is used to identify the pixel with the lowest $V_{thrComp}$ turn-on value. This is the most sensitive pixel, which keeps firing for the target threshold signal for the lowest $V_{thrComp}$ (highest threshold setting). Care is taken to verify that $V_{thrComp}$ chosen is at least 10 DAC units away from the noise level of the *tornado-figure* for this pixel. In these conditions, all pixels will have a threshold higher than the target threshold.

The algorithm proceeds with the optimization of the V_{trim} DAC, with the goal of setting the ground for an optimal use of the `TrimBits` dynamic range. For this purpose, V_{trim} is set to the lowest value that still allows the least sensitive pixel to be trimmed to the target threshold. Such pixel is identified by recording S-curves with respect to V_{cal} and choosing the pixel with the highest turn-on point. With the `TrimBits` set to zero for all pixels, a two-dimensional efficiency scan as a function of V_{cal} and V_{trim} is performed for this pixel. An example of such scan is shown in Figure 9.2. The value of V_{trim} is iteratively lowered until the V_{cal} efficiency turn-on occurs for the V_{cal} value of the target threshold. Being

this the least sensitive pixel and being the effect of the TrimBits set to their maximum, with this value of V_{trim} all other pixels will reach the target threshold with higher values (lower effect) of the TrimBits. The value of the TrimBits is then optimized for each pixel using a binary search over the full DAC range.

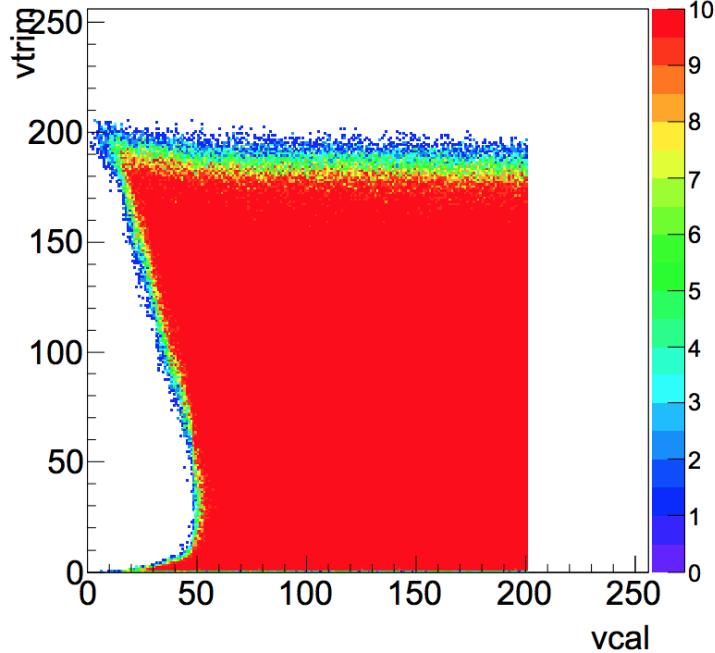


Figure 9.2: The efficiency scan with respect to V_{cal} and V_{trim} DAC values, for the least sensitive pixel, used to optimize the V_{trim} setting.

The programmability of the TrimBits is verified by sequentially enabling them and recording at each step the V_{cal} threshold with the S-curve method. If the threshold is found unchanged with respect to a previous configuration, the corresponding Trimbit is considered non programmable. In case the defective bit is the least significant one, the pixel is not considered to be defective, since it can still be trimmed successfully. Pixels with problems in the more significant TrimBits are considered defective.

9.1.6 Pulse-Height optimization

The PH response as a function of signals of different amplitude is optimized with the goal of obtaining an appropriate dynamic range for all pixels. The optimal configuration for an individual pixel corresponds to a PH response saturating the 8-bit ADC readout for the largest signal of interest, i.e. the typical MIP energy deposit for detector operation, and, simultaneously, starting with non-zero values at threshold. An example of this configuration is illustrated in Figure 9.3. A PH response starting with zero-values, typically for signals larger than the threshold, has to be avoided, since in this configuration signals passing the threshold, but with amplitude smaller than the first non-zero-PH signal, are assigned a PH value of 0, being actually lost.

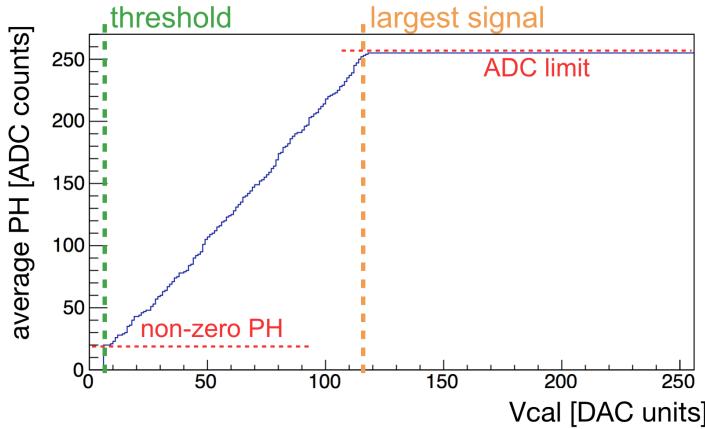


Figure 9.3: Illustration of the full exploitation of the dynamic range of the PH ADC for one pixel, as function of the signal charge in units of the V_{cal} DAC. The pixel responds with non-zero PH from the set threshold (green vertical line) and saturates the ADC for the largest signal of interest (orange vertical line).

The PH information is digitized on chip, as described in Section 8.3.1, and the PH response is manipulated by two ROC-level DACs: the PhOffset DAC, which controls a constant offset in the PH measurement, and the PhScale, which controls the reciprocal of the gain. These two ROC parameters are optimized aiming at providing ideal PH configuration for all pixels, taking into account pixel-to-pixel variations of their inherent gain and inherent offset. This corresponds to tuning the PH ADC such that it provides sufficiently large values for low-inherent-gain and low-inherent-offset pixels when sampled near the threshold and, at the same time, such that the PH response of high-inherent-gain and high-inherent-offset pixels for the largest signals expected does not saturate the 8-bit ADC range. These conditions provide the constraints for the optimization of the two DAC values.

The algorithm developed to carry out the optimization first identifies the lowest-response pixel and the highest-response pixel. In order to do so, the PhOffset and PhScale DACs are set to default intermediate values and PH is measured for all pixels for signals of $V_{cal} = 60$, low-range, and $V_{cal} = 255$, high-range, in order to identify the lowest-response and the highest-response pixel, respectively. The value of $V_{cal} = 255$, high-range, is chosen to probe the PH with the largest V_{cal} signal available, whereas the choice $V_{cal} = 60$, low-range, allows the measurement of the PH for small signals, while being safely far from the threshold of $V_{cal} = 35$. Examples of these PH measurements are shown in Figure 9.4. The range of the color axis provides the size of pixel-to-pixel variations that have to be accommodated during the optimization. The initial PhScale setting is varied until the PH measurement for the highest-(lowest-) response pixel is below 255 (above 0) ADC counts. The distribution of the PH measurement is used to find the sought pixels by selecting the highest (lowest) 98-th percentile pixel as the highest-(lowest-)response pixel, in order to exclude outliers. For the lowest-response pixel, also the minimum V_{cal} at which the PH response exceeds 10 ADC counts is recorded and stored for use later in the optimization. The search is restricted by excluding pixels that have hit efficiency below 100%, in order to avoid using defective or anomalous pixels. Furthermore, the search is first performed

in a fiducial region of the chip, excluding pixels within 3 columns and 5 rows of the edge of the ROC. This allows the exclusion of edge and corner pixels, which have different sensor sizes, and of the areas where imperfect bump bonding are observed to concentrate. If the search in the fiducial region of the chip is not successful, it is repeated on the full chip.

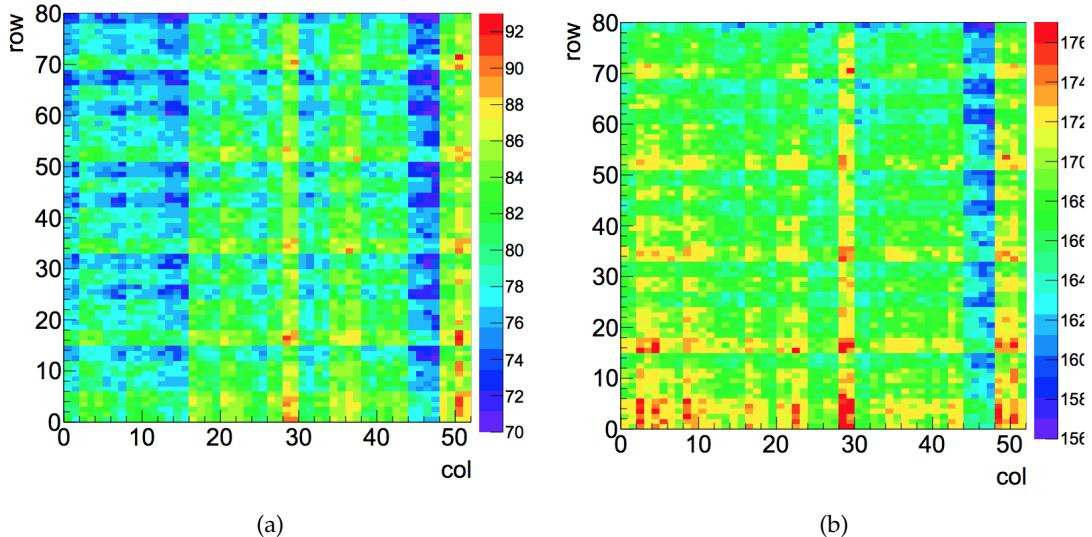


Figure 9.4: PH measurement for all pixels of a ROC at $V_{cal} = 60$, low-range (a) and $V_{cal} = 255$, high-range, (b). The maps are used to identify the lowest-response and highest-response pixel, respectively.

An important aspect of the search is that it is particularly fast: in fact, no full PH response curve as a function of V_{cal} is recorded, whose recording requires about 600 ms for each pixel and thus about 40 minutes for a full ROC; instead, the sampling of the PH is performed at chosen values of V_{cal} for all pixels simultaneously, requiring only few seconds. The following steps of the algorithm involve only operations on the two identified pixels, reducing the time consumption of the optimization.

The algorithm performs the optimization of the DACs based on two-dimensional scans of the PH measurements as functions of the PhScale and PhOffset DACs, using the two extremal pixels as benchmarks. For the lowest-response pixel, the scan is performed sending signals at the minimum V_{cal} recorded in the previous step. For the highest-response pixel, a user-defined value of V_{cal} is used. During module production testing, this value has been set to 100, corresponding to about $35ke^-$, larger by about $10ke^-$ than the most probable total signal charge for a MIP traversing the sensor. This choice allows to restrict the ADC dynamic range to the range of signals charges of interest, thus improving the PH resolution. The optimization, in fact, avoids using part of the dynamic range to record PH for signals up to the saturation level of the pre-amplifier, typically around $60ke^-$ and much larger than the signals observed in the detector. The two-dimensional scans, shown in Figure 9.5, represent the most time consuming step of the algorithm, requiring about 4 minutes each.

The optimal values of the PhScale and PhOffset DACs are chosen as those that simultaneously satisfy the following conditions:

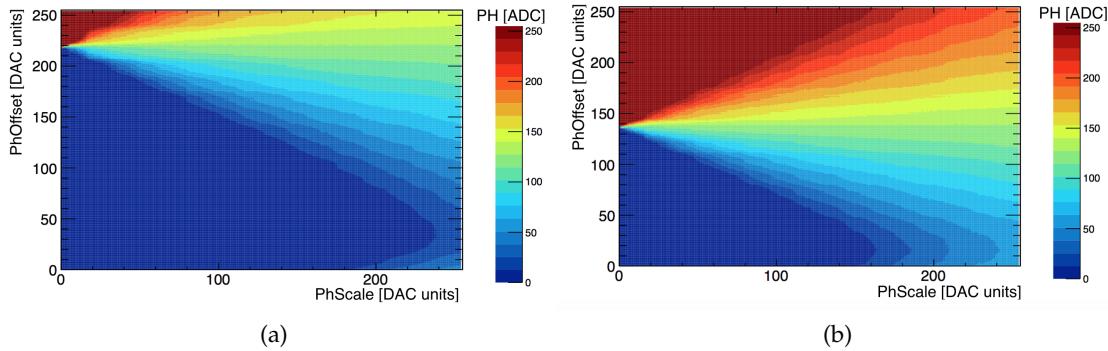


Figure 9.5: PH response as a function of (PhScale,PhOffset) for the lowest-response (a) and highest-response (b) pixel. The scans are performed injecting signals close to the threshold of the lowest-response pixel for (a), and the largest signal of interest, here $V_{cal} = 100$, high-range, for (b).

- above-zero PH response for the the lowest-response pixel. A minimum PH of 20 ADC counts is requested, to provide a sufficiently large safety margin from zero for the solution of the optimization. If no solution is found, the minimum requirement is relaxed to > 0 ;
- close-to-saturation, but smaller than 255, PH response for the highest-response pixel. A tolerance of 3 ADC counts within the maximum value is given, to widen the range of possible solutions.

In case of multiple pairs of values satisfying the constraint, the one with the lowest PhScale is chosen and, if multiple values of PhOffset are possible for this value of PhScale, the lowest PhOffset is chosen.

The result of applying the two constraints described above to the distributions shown in Figure 9.5 is given in Figure 9.6. It is relevant to notice that the definition of optimal values given above correspond to only very few combinations of (PhScale,PhOffset), among the 65536 possible ones, providing the desired PH behaviour of the ROC.

The result of the optimization is validated by measuring the PH response of all pixels for the chosen values of PhScale and PhOffset at three different values of V_{cal} : low, mid-range and high V_{cal} . These maps are used to check the percentage of pixels with non-optimal PH dynamic range after optimization.

9.1.7 GainPedestal

The *GainPedestal* test samples the PH response as a function of the V_{cal} setting for all pixels. The sampling is then fitted and the fit result stored to be used in other characterization procedures, briefly summarized in Section 9.2.4, whose detailed description is beyond the scope of this thesis. The sampling of the PH is performed in steps of 10 V_{cal} units, in the low range, and for 5 sampling points in the high range. This sparse sampling allows to keep the duration of the procedure under control (< 15 minutes), while allowing a good

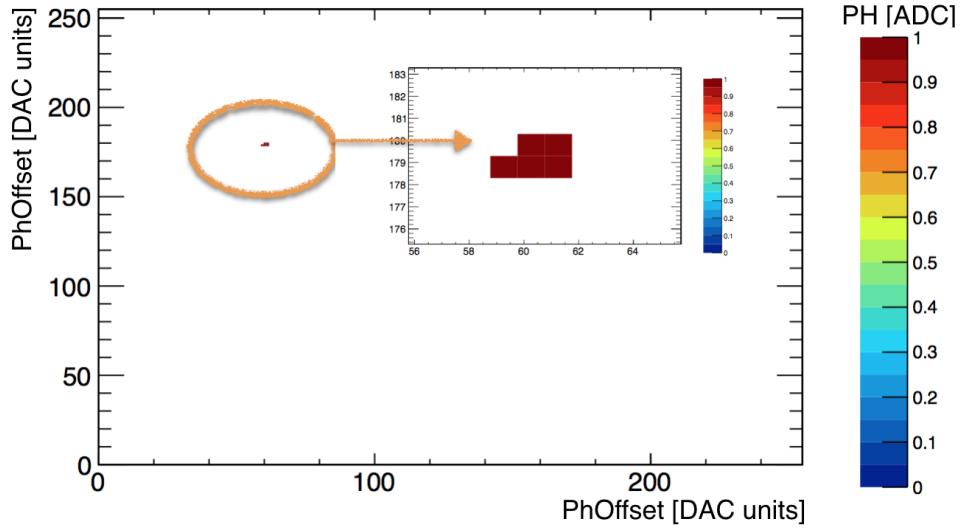


Figure 9.6: The combinations of (PhScale,PhOffset) which satisfy the optimization requirements described in the text.

accuracy of the fit. The measurements are then fitted with a hyperbolic tangent function, parametrized as in equation 9.1, over the whole sampling range.

$$\text{PH} = p_3 + p_2 \tanh(p_0 \cdot V_{cal} - p_1) \quad (9.1)$$

The parameter p_1 is associated with the linearity of the response of the PH and it is used in the identification of pixels with anomalous PH curves. The PH distributions are also fitted with a linear function in a restricted range to extract the values of gain and pedestal for each pixel.

9.1.8 Readback mechanism and calibration

The digital ROC offers a readback mechanism that allows access to on-chip measurements and information through a slow ADC, as introduced in Section 8.4.

Several information can be read out through this mechanism, depending on which value the *readback register* is set: these include digital data and analog on-chip measurements of currents and voltages. The full list of measurements accessible and their respective register value are given in Table 9.1.

The analog readback measurements need to be calibrated using external voltage and current references to be converted from ADC units to physical units. These measurements represent an important tool to provide on-chip voltage and current measurements, which can not be otherwise accessed in the final detector system due to the powering scheme. The laboratory setup in which the Fulltest is performed, described in Section 9.2.1, allows powering and communication with each individual module via a Digital Test Board (DTB), contrary to the FED and FEC system described in Section 8.4. For this reason, the readback ADC calibration cannot be performed once the modules are installed in the detector, while it is possible to use the DTB to calibrate individually the analog readback response of each chip.

	Register	Function
Digital	0	last data
	1	last address
	2	last pixel column
	3	last pixel row
Analog	8	unregulated digital voltage
	9	unregulated analog voltage
	10	regulated analog voltage
	11	bandgap reference voltage
	12	analog current

Table 9.1: The readback functions corresponding to different settings of the readback register.

For this purpose, a calibration procedure is put in place. The readback measurements of the unregulated digital and analog voltages are calibrated by comparing the voltage supplied to the chip via the DTB to the measurement performed by the readback mechanism. The calibration coefficients are extracted through a linear fit. The calibration can be applied to single-chip units or to full modules. In the latter case, voltage drops occurring between the DTB and the chip have to be taken into account. In order to do so, the average resistance R_C of the module cable has been measured and used to correct the voltage supplied to the module. Furthermore, a difference Δ_{GND} of about 180 mV is observed between the ground levels of the module adaptor card and of the HDI, which is used to further correct the estimate of the chip voltage supply. The calibration function thus can be written as:

$$RB(V, I) = p_0 + p_1 \cdot (V - R_C \cdot I - \Delta_{GND}) \quad (9.2)$$

where $RB(V, I)$ is the readback ADC response for a voltage V and a current I measured with the DTB and p_0 and p_1 are the calibration coefficients determined by the fit. The function can then be inverted to obtain a measurement of the voltage in physical units. The number of points used in the calibration has been tuned to reduce the execution time to about 40 s, without affecting the accuracy.

The accuracy of the calibrations for both the unregulated analog and digital voltage has been verified using as a benchmark the bandgap reference voltage V_{bg} on a sample of 16 chips. Its value, indeed, can be measured via the readback, as shown in Table 9.1, and directly on-chip, using a probe needle to access the dedicated spy pad on the chip periphery. The calibration of the unregulated digital voltage provides measurements of V_{bg} with an accuracy of 0.01 V, while the calibration of the unregulated analog voltage, though achieving the same accuracy, shows a systematic underestimation of V_{bg} of about 0.01 V. On modules, a residual effect is caused by the finite resistance of the HDI producing an additional voltage drop. This effect has the largest impact for the ROC furthest away from the landing position of the module cable, causing an inaccuracy in the measurement of V_{bg} of about 0.01 V.

The readback analog current calibration is carried out in two steps: first, the non-linear relation between the analog current I_{ana} and the V_{ana} DAC setting is measured with the DTB and described with a second order polynomial, as:

$$I_{ana}(V_{ana}) = q_0 + q_1 \cdot V_{ana} + q_2 \cdot V_{ana}^2 \quad (9.3)$$

Then, the linear relation between the readback analog current ADC measurement and V_{ana} is described as:

$$RB(V_{ana}) = r_0 + r_1 \cdot V_{ana} \quad (9.4)$$

and the parameters q_i and r_i are extracted through fits. Combining the two equations and solving for I_{ana} , the calibration function is obtained as:

$$I_{ana}(RB) = q_0 + q_1 \cdot \left(\frac{RB - r_0}{r_1} \right) + q_2 \cdot \left(\frac{RB - r_0}{r_1} \right)^2 \quad (9.5)$$

Examples of the distributions described by Equations 9.3 and 9.4 are shown in Figure 9.7.

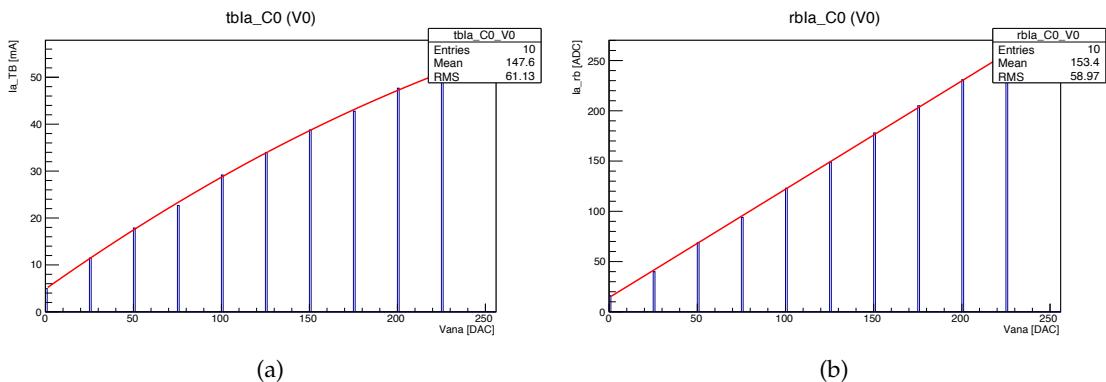


Figure 9.7: The measurement of the analog current I_{ana} with the DTB (a) and with the readback ADC (b) as a function of the V_{ana} DAC.

The calibration has to be carried out sequentially for each ROC of the module, given their connection in series, and takes about 7 minutes per module.

9.1.9 IV curve

Finally, the quality of the sensor is tested. The most important parameter of the sensor is its leakage current, which will account for a significant fraction of the power consumption of the detector after irradiation. While, as described in Section 7.3, this current can be reduced by lowering the operational temperature of the detector, sensor with very large leakage currents have to be discarded, indicating a faulty or damaged sensor. The sensors are tested by measuring the leakage current as a function of the bias voltage, recording the so-called *IV curve*. While the nominal operational bias is -150 V, the scan extends from 0 to -600 V, since larger voltages are needed as the irradiation dose received by the sensor increases to fully deplete it. The IV curve has typically an initial sharp rise, which then

continues approximatively linearly for bias voltages larger than -100 V . The current then sharply increases after a breakdown voltage: for this reason, during the test a compliance current of $100\mu\text{A}$ is set to avoid damaging the sensor. The IV curve is recorded at 17°C and the values of the leakage current at -150 V and the slope of the curve between -100 and -150 V are used to assess its quality. Furthermore, the IV curve is measured also at -20°C and compared to the one obtained at 17°C to evaluate if the scaling of the current with the temperature is well described by the Equation 7.2.

9.2 MODULE QUALIFICATION

All produced modules undergo a series of tests aimed at calibrating their parameters and at assessing their quality and suitability for detector installation. The results of the tests are combined to assign a grade (A, B or C) to the module. Grade-A modules represent cases where virtually no or very minor flaws are observed in the module and its sensor. In grade-B modules, minor defects or anomalies are observed, which do not impact the operation and performance of the modules in the final system. Modules graded A or B are thus suitable for installation in the final detector. Modules which show major defects are graded C and discarded for detector construction. The details of the grading system are provided in the following of this Section.

9.2.1 Qualification setup

The module qualification is carried out in a temperature- and humidity-controlled environment by means of a *cooling box*, built and commissioned for this purpose. The cooling box features 4 high-power active cooling Peltier elements, mounted in a copper block through which pre-cooled water flows, removing the excess heat from the Peltier elements. The cooling plate is designed to host up to 4 modules. The presence of a lid allows the closure of the qualification volume and enables the control of the humidity, via a flux of dry air. In this way, it is possible to control via software that the temperature never falls below the dew point, protecting the modules from damage. The cooling box can be controlled via a JUMO500 interface. A picture of the cooling plate with four modules is shown in Figure 9.8(a).

The module is connected, via a cable-adaptor card, to a DTB, which can send and receive signals and powering and where the firmware for module control and readout resides. The DTB also forwards the bias voltage from a Keithley power supply to the module. The DTB is connected to a Personal Computer (PC), where the qualification software runs, containing the Fulltest. Furthermore, a client-server controller software is in charge of operating the different parts of the test setup (cooling box via JUMO500 interface, DTBs, power supply, testing software) and dictates the sequence of operations of the qualification. A picture of the full qualification setup is shown in Figure 9.8(b), while a diagram of the setup is given in Figure 9.9.

9.2.2 Configuration

The qualification of a module comprises the following steps:

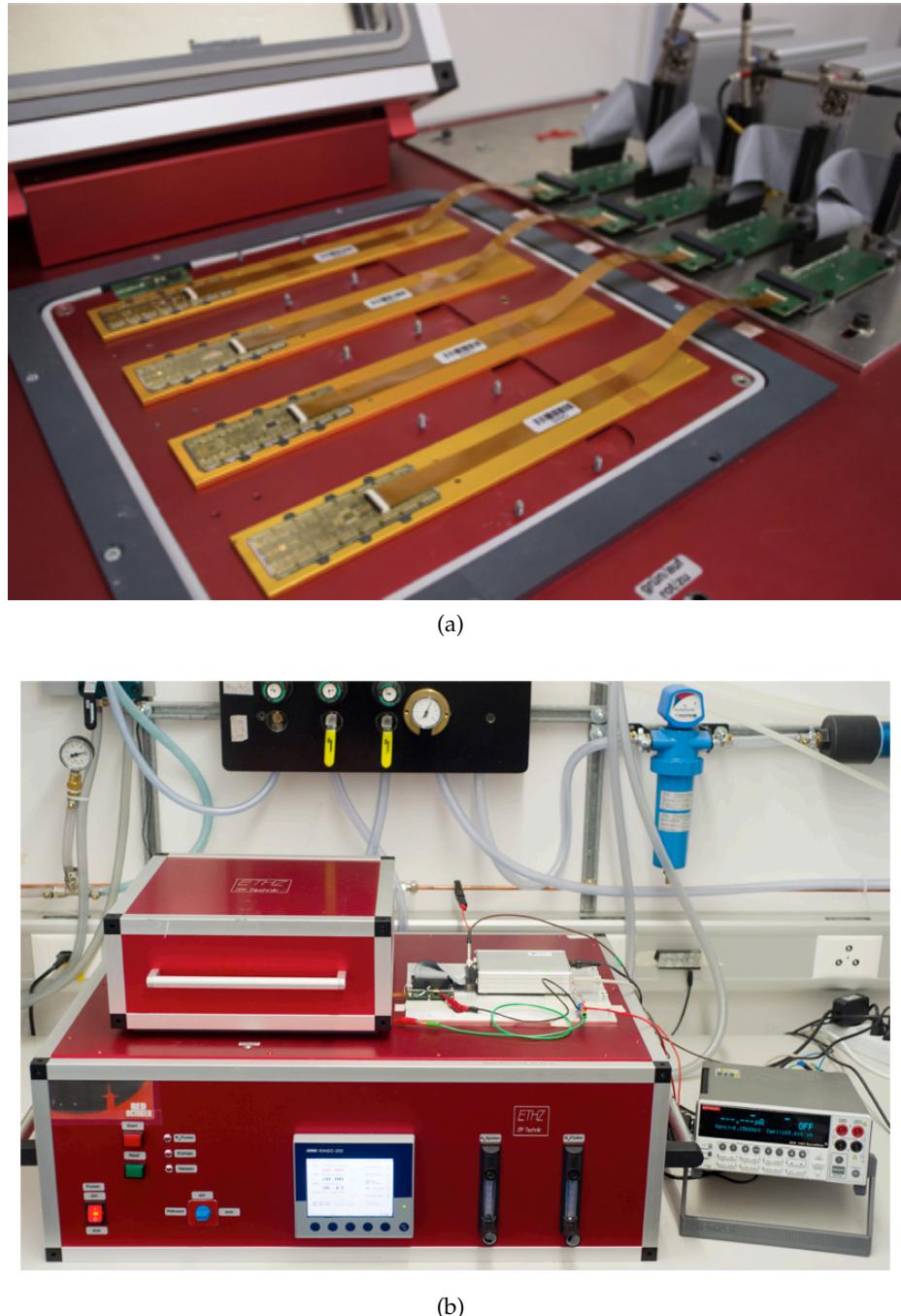


Figure 9.8: (a) The plate of the cooling box where modules are placed for temperature-controlled qualification. The Peltier elements are installed right below the plate and provide the necessary cooling power. In the background, the cable-adapter cards and the DTBs to which the module is connected.
 (b) The cooling box, one DTB, and the Keithley power supply.

- at 17°C, a quick measurement of the leakage current is performed at the nominal bias voltage of -150 V . If it exceeds $30\mu\text{A}$, the module is graded C and the qualifi-

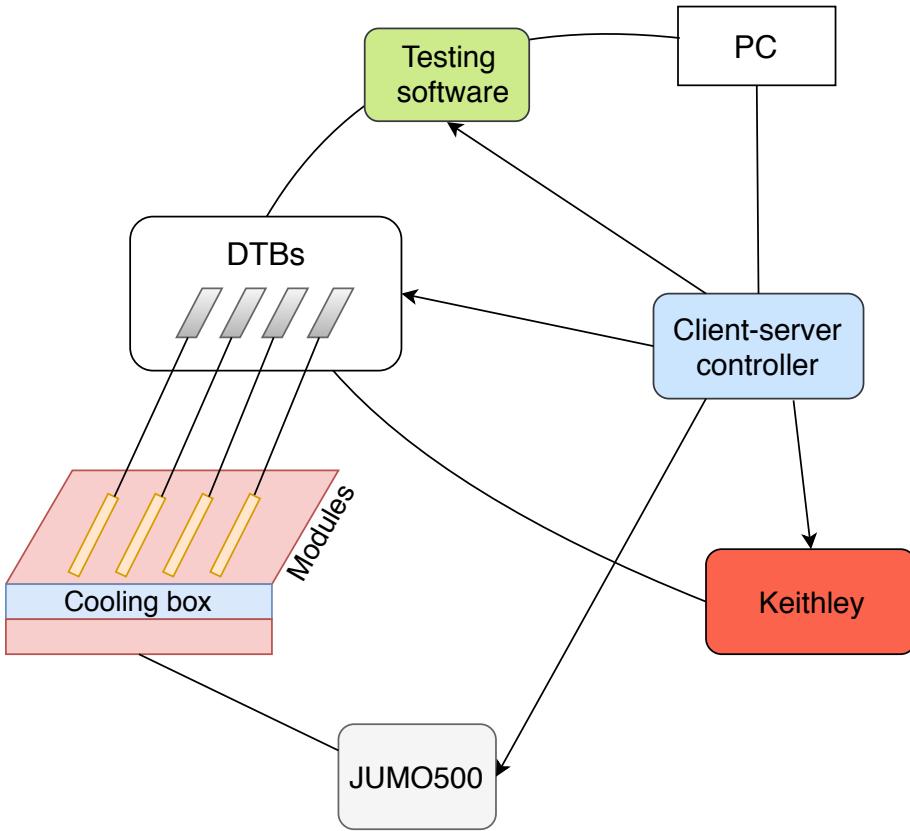


Figure 9.9: The diagram of the qualification setup. The client-server controller, running on a PC, controls both software instances (the testing software) as well as hardware components of the setup: the cooling box, via the JUMO500 interface; the DTBs; the Keithley power supply, providing the bias voltage to the modules via the DTBs.

cation interrupted, while if it is larger than $5\mu\text{A}$ the module is assigned a B grade at this stage and proceeds to the qualification;

- the module is cooled down to -20°C and it undergoes a first Fulltest. If errors that prevent the completion of the Fulltest are encountered, the qualification is stopped and the module is graded C;
- a thermal stress test is performed, simulating the conditions which modules will encounter in the full system. The module is warmed up to 17°C and cooled down to -25°C ten consecutive times. A second Fulltest at -20°C is then performed, in order to see if mechanical stress due to temperature variations caused damage in the functionalities of the module. The sensor is characterized at -20°C by recording the IV curve, described in Section 9.1.
- the Fulltest+IV curve sequence is repeated at 17°C . Testing the module at two different temperatures allows to perform a linear interpolation as a function of the temperature of its calibration and configuration parameters, as well as to verify how closely the leakage current follows the expected behavior as a function of the temperature described by equation 7.2.

The whole qualification lasts about 8 hours and is performed in parallel for 4 modules. An example of the temperature diagram as a function of time is shown in Figure 9.10, where the different routines of the qualification can be identified.

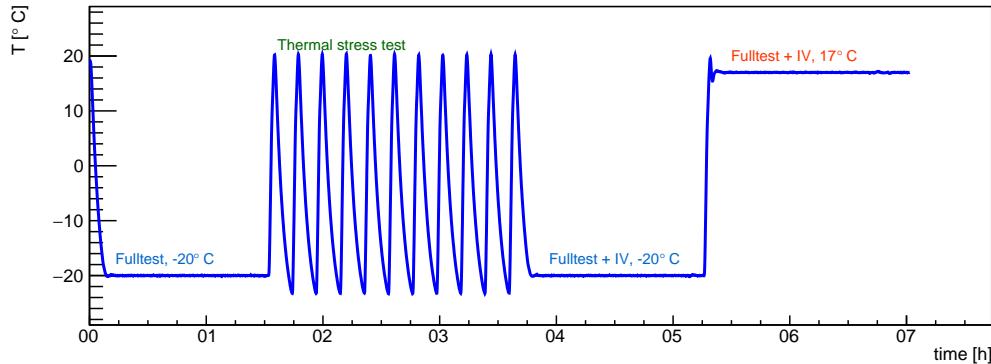


Figure 9.10: Example of a temperature diagram as a function of time during module qualification. Each Fulltest requires about 70 minutes and 10 more minutes are needed for an IV curve recording. Each thermal cycle between 17°C and –25°C takes about 15 minutes, leading to a total qualification time around 8 hours.

9.2.3 Grading scheme

The results of the Fulltests and of the sensor tests are combined to assign a grade to the module. In each Fulltest, the grading is assessed at the ROC level, as described in Section 9.2.3.1, and the module receives the grade of the worst-scoring ROC. This grade defines the overall quality of the electrical functionalities of the module. For each module, also the sensor is graded, according to the measurements performed in the IV tests. The quantities considered and the corresponding grades are discussed in Section 9.2.3.2. A suite of tests under X-ray irradiation is performed and a corresponding grading scheme is defined, as discussed in Section 9.2.4.

9.2.3.1 ROC grading

The goal of the ROC grading scheme is to select chips that offer the best performance in terms of resolution and hit efficiency. For this purpose, among the several properties and parameters of the PUC, those with the largest impact on such performance have been identified and chosen as benchmarks for the ROC grading. Simulation studies have been carried out during the module production of the previous CMS Pixel detector [34] and the grading scheme presented here builds on the experience from this study. However, the vast difference in the chip design and readout demand an ad-hoc investigation on the performance of the digital chip, which is presented in Section 9.3.7.

A charged particle traversing the Pixel detector typically deposits charge in more than one pixel, due to the angle of the particle trajectory and the Lorentz angle, leading to charge sharing among pixels. A clustering algorithm groups the single pixel deposits, by identifying clusters as groups of adjacent pixels with signals exceeding the set threshold. Charge sharing has a direct impact of the threshold and noise levels of pixels on hit

efficiency and resolution: on the one hand, a threshold as low as possible is desirable, since too high thresholds can cause the loss of pixels in the cluster that collect only a fraction of the initial hit charge; on the other hand, too low thresholds lead to pixels firing on noise and polluting the cluster reconstruction. Furthermore, a low-threshold setting extends the lifetime of the detector, by allowing to efficiently separate smaller and smaller signals from the increasing noise as the sensor suffers radiation damage. Once the target value of the threshold for all pixels is set (nominally at $1750e^-$), the uniformity of the threshold among different pixels and the pixels noise levels represent the most important parameters to control a clear separation between proper signal hits and noise. For this reason, the threshold width around the target value and the mean pixel noise are taken as grading parameters. Furthermore, the non uniformity of the PH among pixels has an impact on the resolution, since a per-pixel granularity for the PH gain and pedestal is not a sustainable payload at reconstruction time and only average and coarser information can be used. The chips are thus also graded based on the uniformity of their PH gain values, using the width of the gain distribution for all the pixels of the chip divided by its mean (*relative gain width*) and the spread of the pedestal values as grading parameters.

Hit efficiency is directly affected by the amount of dead or defective pixels in a ROC, which enters the grading scheme as shown in Table 9.2. On top of inefficient pixels, identified in the PixelAlive test, other pixel defects can affect the efficiency and quality of the readout:

- threshold defects, i.e. threshold after trimming more than $10 V_{cal}$ units away from the target;
- large ($> 400e^-$) or anomalous noise ($< 10e^-$);
- PH gain defect (larger than 5 or smaller than 1);
- PH linearity (larger than 7 or smaller than 0);
- TrimBits defect, i.e. if any bit of the TrimBits DAC, except the least significant one, is found defective in the Trimming test;
- address decoding problem (misplaced or out-of-range column or row);
- defective bump bonding connection.

Defects in the pixel masking mechanism are treated differently than all other pixel defects: if a single pixel can not be masked and becomes noisy or malfunctioning at some point in the detector lifetime, this impairs the readout of the full chip. For this reason, a single mask defect causes the whole chip to be graded C and the module to be discarded.

All other defects are summed together, avoiding double counting of pixels showing more than one defect, and the total fraction of defective pixels enters the grading system of the chip.

The set of parameters considered for the evaluation of the chip quality and the thresholds corresponding to A, B, and C grades are summarized in Table 9.2.

9.2.3.2 Sensor grading

The sensor is graded with the goal of ensuring efficient charge collection and keeping the detector operating within the power consumption constraints, after the sensor suffered

	A	B	C
Mean Noise	$< 200e^-$	$< 300e^-$	$> 300e^-$
Threshold width	$< 200e^-$	$< 400e^-$	$> 400e^-$
Width of relative PH gains	$< 0.1\%$	$< 0.2\%$	$> 0.2\%$
PH pedestal spread	$< 2500e^-$	$< 5000e^-$	$> 5000e^-$
Fraction of defective pixels	$< 1\%$	$< 4\%$	$> 4\%$

Table 9.2: The list of parameters entering the ROC grading and the corresponding thresholds for A, B, and C grades.

radiation damage. The main types of damage caused by ionizing and non-ionizing radiation in the sensor, described in Section 7.4, are accumulation of space charge and bulk damage. The results of the damage in the operation of the detector are increased leakage current, higher bias voltage for full depletion, and lower charge collection efficiency. Based on this, stringent criteria on the leakage current of unirradiated sensors and on the shape of the IV curves are set. Limits on the leakage current measured at 17°C are set at 2 (10) μA for modules to be graded A (B), based on the studies of sensors after irradiation. Furthermore, the IV curve for unirradiated sensors is expected to be slowly linearly rising between about -100 V and the breakdown voltage, expected to happen at bias voltages larger than -600 V . More steeply-rising curves correspond to sensors drawing larger currents when operated at higher bias voltages after irradiation and typically identify modules which suffered sensor damage and which are more prone to early current breakdown. A limit on the *slope* of the IV curve, defined as $I_{\text{leak}}(V = -150\text{V})/I_{\text{leak}}(V = -100\text{V})$ is set at 2, with modules exceeding this limit being graded C, while no grade B is foreseen based on this parameter. The grading of the sensor is summarized in Table 9.3.

	A	B	C
$I_{\text{leak}}(T = 17^\circ\text{C})$	$< 2\mu\text{A}$	$< 10\mu\text{A}$	$> 10\mu\text{A}$
$I_{\text{leak}}(V = -150\text{V})/I_{\text{leak}}(V = -100\text{V})$	< 2	-	> 2

Table 9.3: The list of parameters entering the sensor grading and the corresponding thresholds for A, B, and C grades.

9.2.4 X-ray tests and grading

In addition to the cooling box qualification, each module undergoes a second suite of tests in an environment where high-rate X-ray beams can be shone onto the modules. The purposes of the tests under X-ray radiation are multiple:

- measure the pixel noise and the readout efficiency for internal calibration signals in the high-occupancy background of the X-ray hits;
- verify the quality of the bump bonds with a procedure that, differently from what described in Section 9.1, relies on the injection of the V_{cal} signal through the sensor;

- identify defects in the double-column readout mechanism and errors in the readout data that appear only when stressing the readout with large volume of hits generated by X-rays;
- isolate *hot pixels*, whose threshold anomalously varies under intense readout activity and that require masking.

Pixels with defective bump bonds, noisy pixels and hot pixels contribute to the total of defective pixels, used to grade the module. The results of these tests are combined in a grading scheme, summarized in Table 9.4. More details about the X-ray test stand and the test routines can be found in [62, 85].

	A	B	C
Efficiency at 120MHz/cm ²	> 98%	> 95%	< 95%
Column uniformity problems	0	-	≥ 1
Readout uniformity problems	0	-	≥ 1
Fraction of defective pixels	< 1%	< 4%	> 4%

Table 9.4: The list of parameters entering the grading of the tests under X-ray radiation and the corresponding thresholds for A, B, and C grades.

9.3 QUALIFICATION RESULTS FOR MODULE PRODUCTION

The module qualification described in the previous section has been designed to tune and assess the quality of the modules produced to build the 4 layers of the barrel section and the 3 disks of each endcap section of the upgraded detector. While the qualification scheme described has been used for all modules, regardless of their final positioning in the detector, the discussion of the results will focus on the modules produced to populate the second innermost layer (L2), which were tested at the ETHZ laboratories. While also modules of the innermost layer (L1) were tested at ETHZ, the L1 ROC design was still under development at the time the studies and the qualifications described in this thesis were carried out. The more severe irradiation conditions and high readout flux of the L1 demanded the development of a more stringent grading scheme, whose description is beyond the scope of this thesis.

In Figure 9.11, the cumulative number of modules per grade is given as a function of time during the production phase. In green, modules with grades A or B are shown, while modules with grade C are indicated in red. The production reached completion with 264 modules evaluated to be fit for installation (191 A-grade modules and 73 B-grade modules), accounting for the full L2 demand and about 20% of spare modules. With 62 C-grade modules, the production yield is 81%.

9.3.1 ROC functionality

The parameters defining the ROC grading scheme based on the results of the qualifications are reported for the whole L2 production in Figures 9.12-9.15. In each Figure, the range

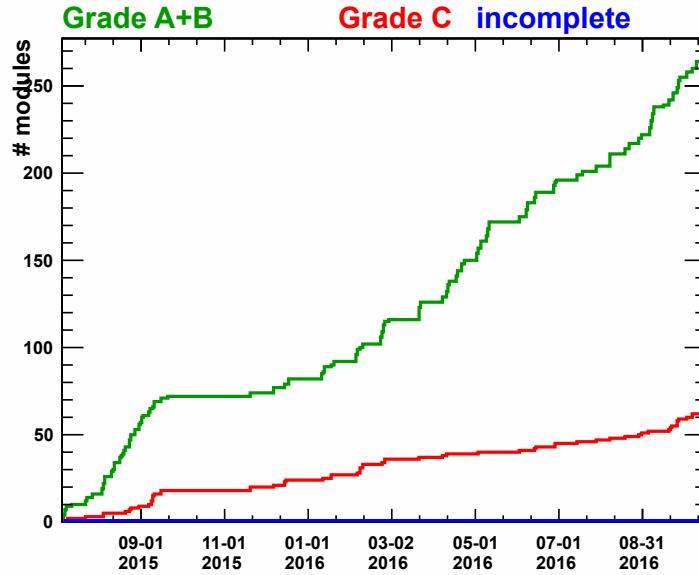


Figure 9.11: The cumulative number of modules with grades A or B (green) and grade C (red) as a function of time, during the production of L2 modules.

of values for a parameter corresponding to a given grade is shown in different colors, identifying the A-grade region (green shade), the B-grade region (blue shade), and the C-grade region (red shade).

The mean noise per ROC is shown in Figure 9.12 as measured in the three Fulltests: first Fulltest at -20°C (a), second Fulltest at -20°C (b), Fulltest at 17°C (c). The second Fulltest at -20°C and the Fulltest at 17°C are performed after the thermal stress test, following the sequence described in Section 9.2.2. In each Figure, the distribution of A-grade modules is shown in green, B-grade modules in yellow, and C-grade modules in red.

In Figure 9.13, the width of the threshold distribution after Trimming is shown for the three Fulltests.

The distributions of the widths of the relative gain, defined in Section 9.2.3.1 and measured in the GainPedestal test, are presented in Figure 9.14, for the three different Fulltests.

The spread of the ROC pedestal distribution, as well measured in the GainPedestal test, is shown in Figure 9.15.

The plots in Figures 9.12–9.15 show that the distributions of almost all chips fall in the grade-A regions, with a very minor population of outliers entering the B-grade region. The fact that the bulk of the distribution is entirely contained in the A-grade region denotes the high quality and reproducibility of the ROC production. It can also be noted how no ROC falls in the C-grade regions of the four grading parameters, indicating the high performance and electrical functionality of the chips produced. This corresponds to no module being discarded based on this part of the grading scheme. Furthermore, by comparing the distributions relative to the Fulltests at -20°C before and after the thermal cycling, no sign of deterioration of the chip performance is observed.

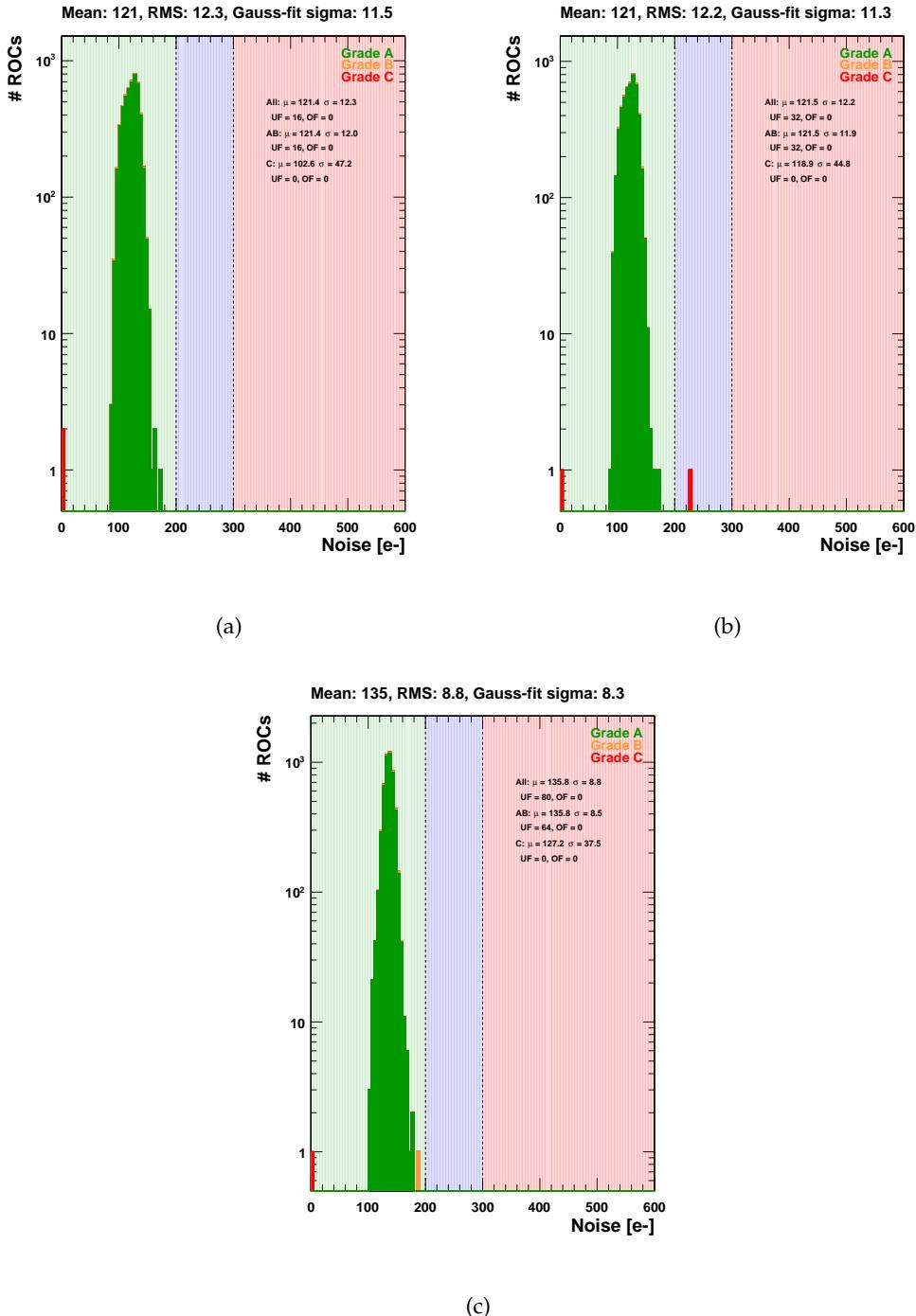


Figure 9.12: The mean noise per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red).

9.3.2 Pixel defects

The number of pixel defects per ROC is defined by measuring specific characterizing properties of the pixels and verifying if they fall inside the range of expected values, as defined in Section 9.2.3.1.

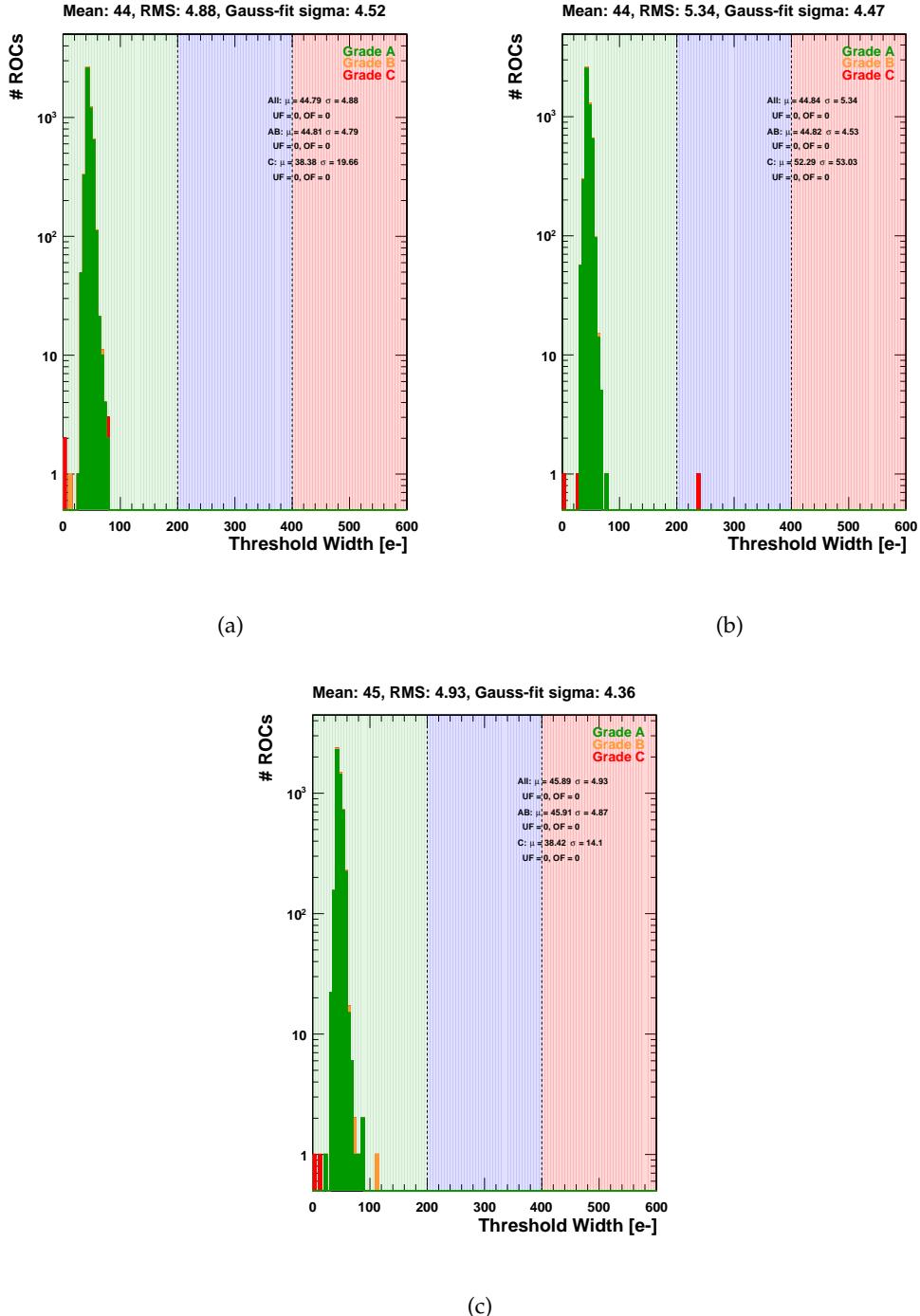


Figure 9.13: The width of the threshold distribution after the Trimming procedure per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red).

The number of pixels with a defective threshold setting is defined as those pixels whose final threshold is more than $10 V_{\text{cal}}$ units away from the target of 35. The distributions in

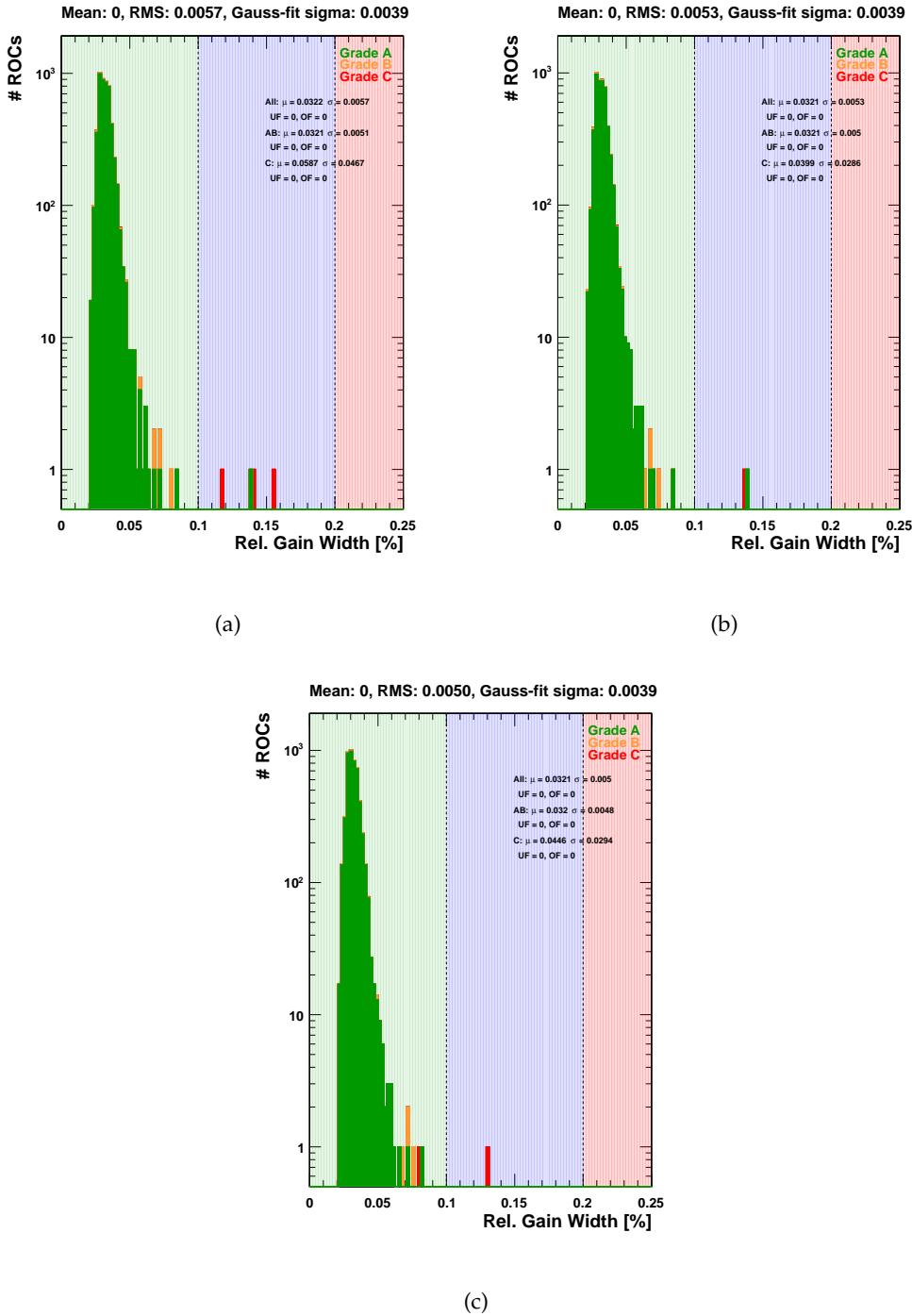


Figure 9.14: The width of the relative gain distribution per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red).

Figure 9.16 shows the per-pixel threshold in the two Fulltests at -20°C and in the Fulltest at 17°C . The distributions are shown separately for modules with final grade A (green),

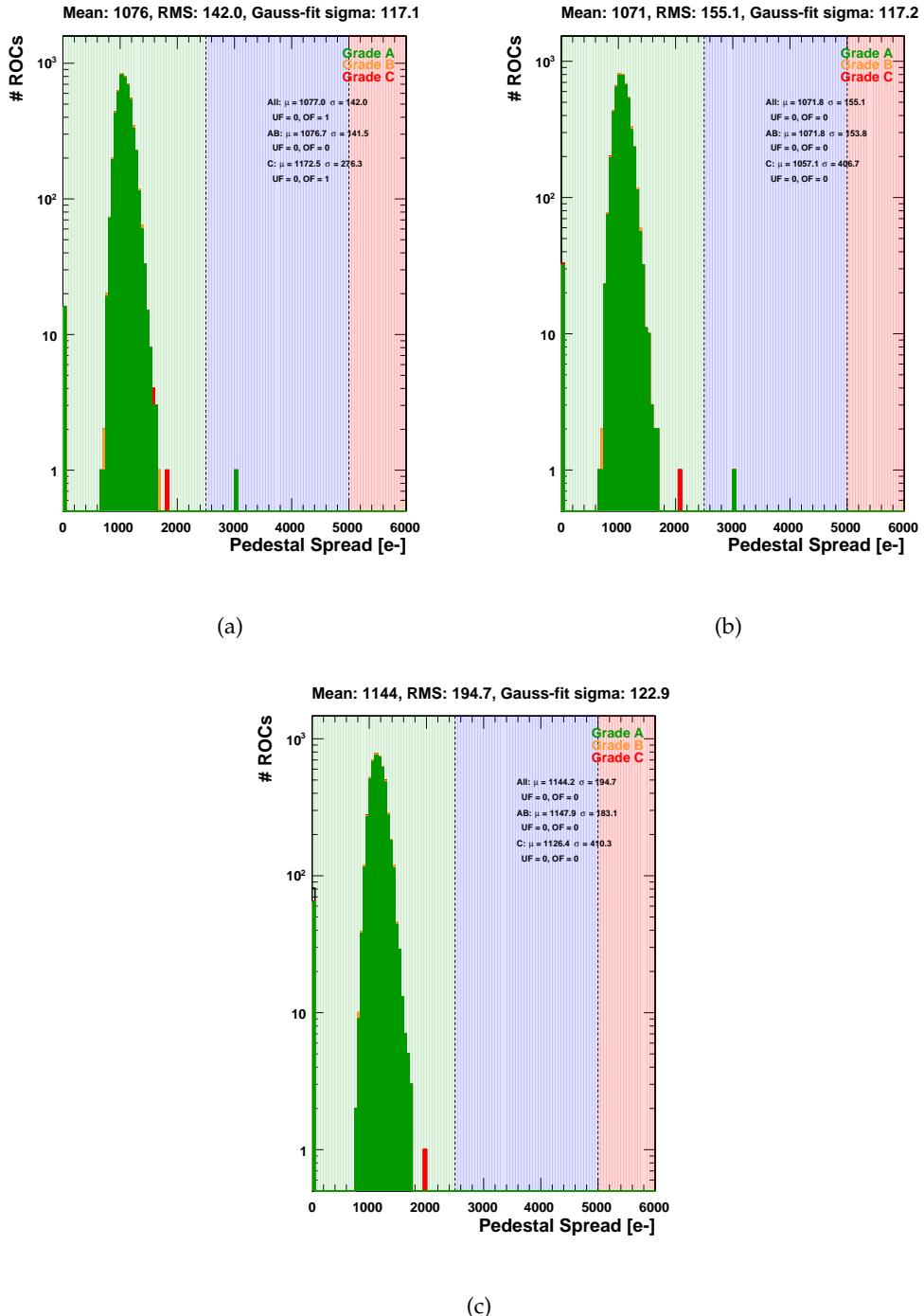


Figure 9.15: The spread of the pedestal distribution per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red).

B (yellow), and C (red). The boundaries in V_{cal} units for a pixel to be considered defective are indicated by the vertical dashed lines.

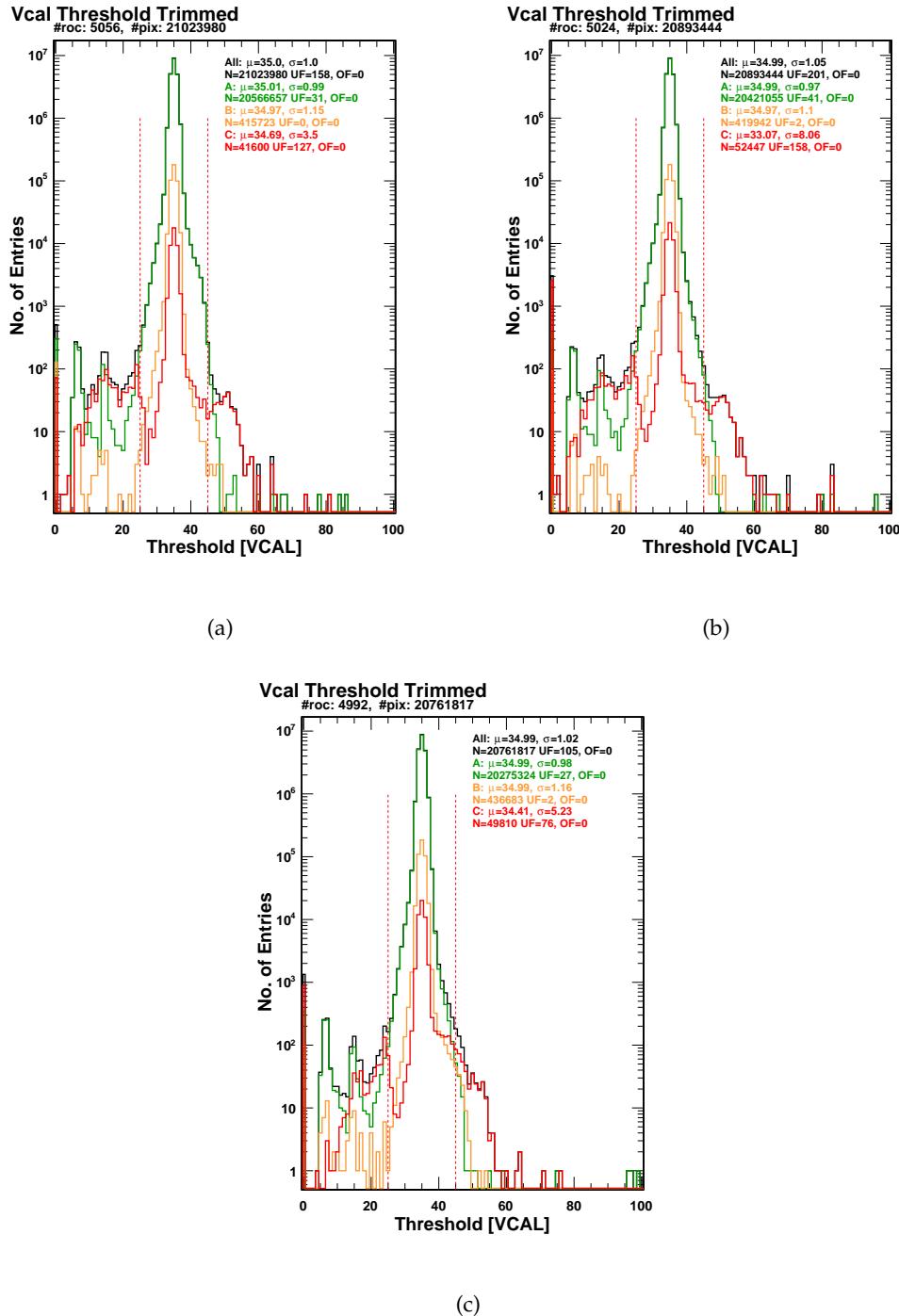


Figure 9.16: The distribution of the pixel threshold, in units of V_{cal} , as measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for pixels in modules with final grade A (green), B (yellow), and C (red).

The noise per pixel is measured in the S-curves test and required to be smaller than 400e^- for non-defective pixels. Furthermore, the noise is required to be larger than 10e^- ,

to exclude pixels with anomalous S-curve shapes which lead to fit failure. The distribution of the pixel noise in the three Fulltest is shown in Figure 9.17, where the same color scheme described for Figure 9.16 applies.

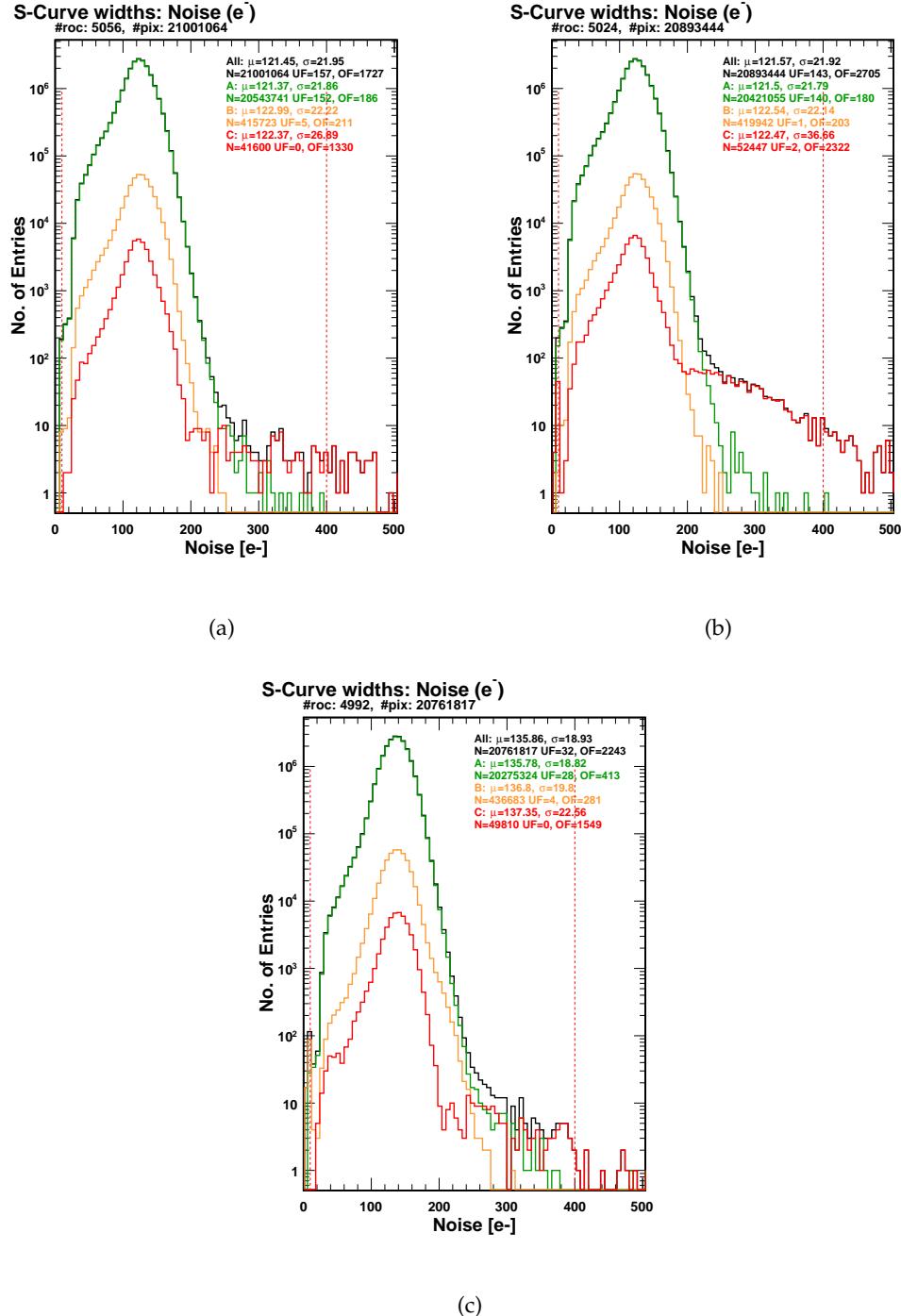


Figure 9.17: The distribution of the pixel noise as measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for pixels in modules with final grade A (green), B (yellow), and C (red).

In Figure 9.18, the distributions of the pixel gain are shown, as measured in the Gain-Pedestal test in the three Fulltests. Pixels with a large (> 5) gain are counted as defective, since their behavior significantly deviates from the mean and cannot be accounted for during the operation of the detector, when only average information on the gain can be used. Also pixels with a gain smaller than 1 are considered defective, indicating an anomalous shape of the PH curve as a function of V_{cal} .

As noted for the chip quality, also the number of pixel defects and distributions of the corresponding grading parameters are not observed to worsen after thermal stress. As described in Section 9.2.3.1, the number of pixels showing the defects described above is added to the number of pixels dead, inefficient, with defective TrimBits, or with address decoding errors. Based on these pixel defects, 2 modules are graded C and rejected for installation.

The main source of module discard among the pixel defect evaluations is found to be the number of defective bump bonding connections, which is responsible for the C-grading of 4 additional modules. These modules consistently show defects localized in the corners of each ROC, while the center of the chip either shows diffused defects or no defects at all, due to a momentary misconfiguration of the bump bonding procedure, which was subsequently fixed. The modules map of the bump bonding defects, recorded at 17°C, for these 4 four modules are given in Figure 9.19.

No defective bits have been observed in the pixel masking mechanism, which allows to disable the readout of each pixel individually, thus no module is downgraded based on this parameter.

9.3.3 Sensor quality

The grading of the sensor is based on the leakage current at the nominal bias voltage of -150 V and on the slope of the IV-curve, as described in Section 9.2.3.2. The distributions of these two parameters are shown in Figure 9.20, separately for modules with final grade A (green), B (yellow), and C (red).

Since only six modules showed an IV-slope larger than 2, such cases are inspected individually. Since the excess for all these modules is small, having an IV-slope smaller than 3, and based on irradiation studies of the sensors showing a decrease of the slope after irradiation, these modules are upgrade to grade *B*, provided that their leakage current at -150 V and 17°C is smaller than $2\mu\text{A}$.

Figure 9.21 shows the overlay of the IV-curves at 17°C for all tested modules. The clear vertical stacking of grade-A (green), grade-B (yellow), and grade-C (red) modules show how the leakage current is the driving parameter in the grading of the sensor.

The quality of the sensor is the major reason for module failure and rejection for installation, accounting for 25 modules.

9.3.4 Functionality under X-ray

The module tests performed under X-ray radiation and the corresponding grading, described in Section 9.2.4, account for the C-grading of 15 modules. In all these cases, the reason for downgrading the module is found to be the presence of errors in the double column readout mechanism, when stressed under high-rate X-ray radiation.

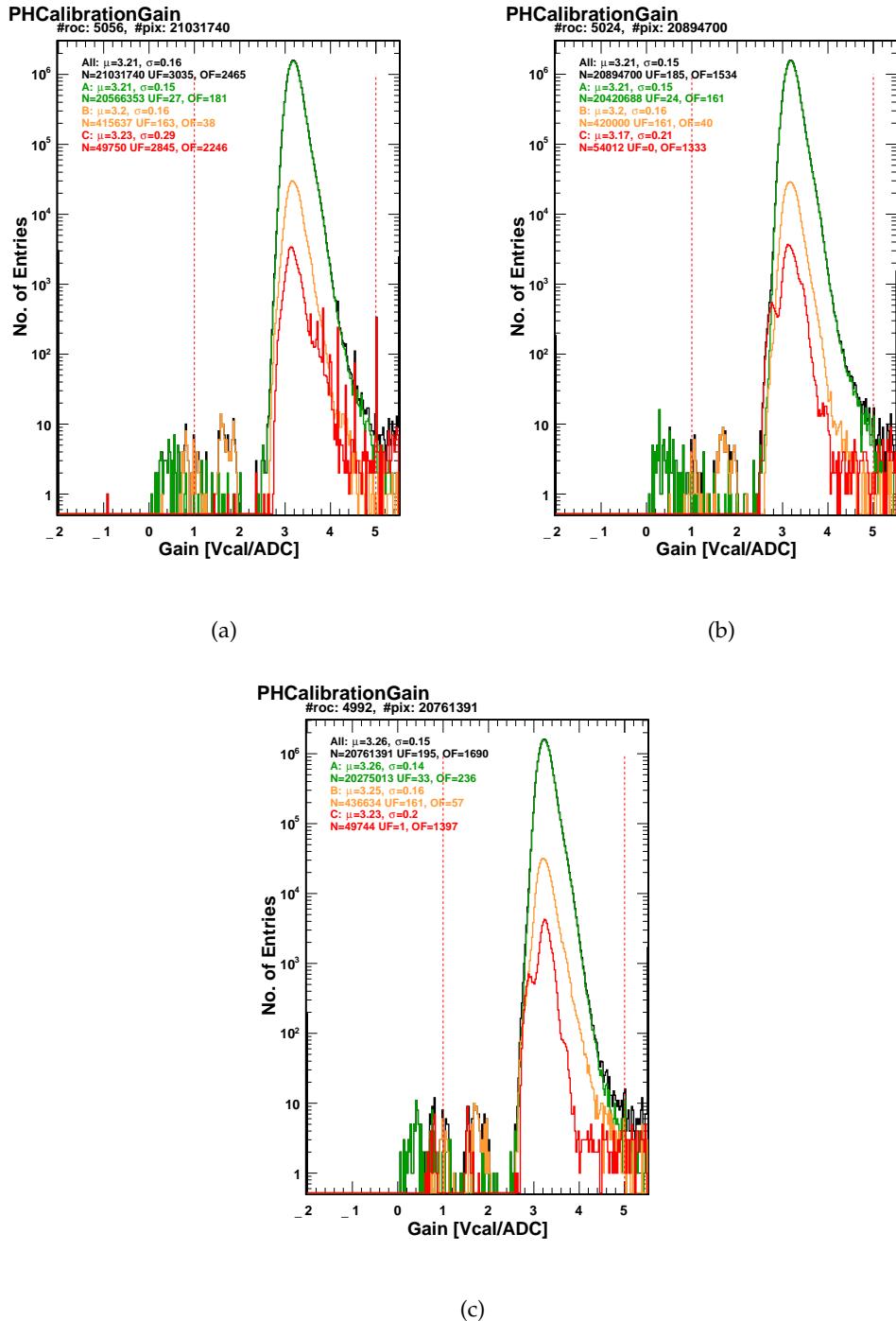


Figure 9.18: The distribution of the pixel PH gain as measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for pixels in modules with final grade A (green), B (yellow), and C (red).

9.3.5 Other failure modes

Despite the thorough set of tests modules undergo, it is possible that atypical module defects, caused by unexpected behavior, might go undetected by the grading scheme.

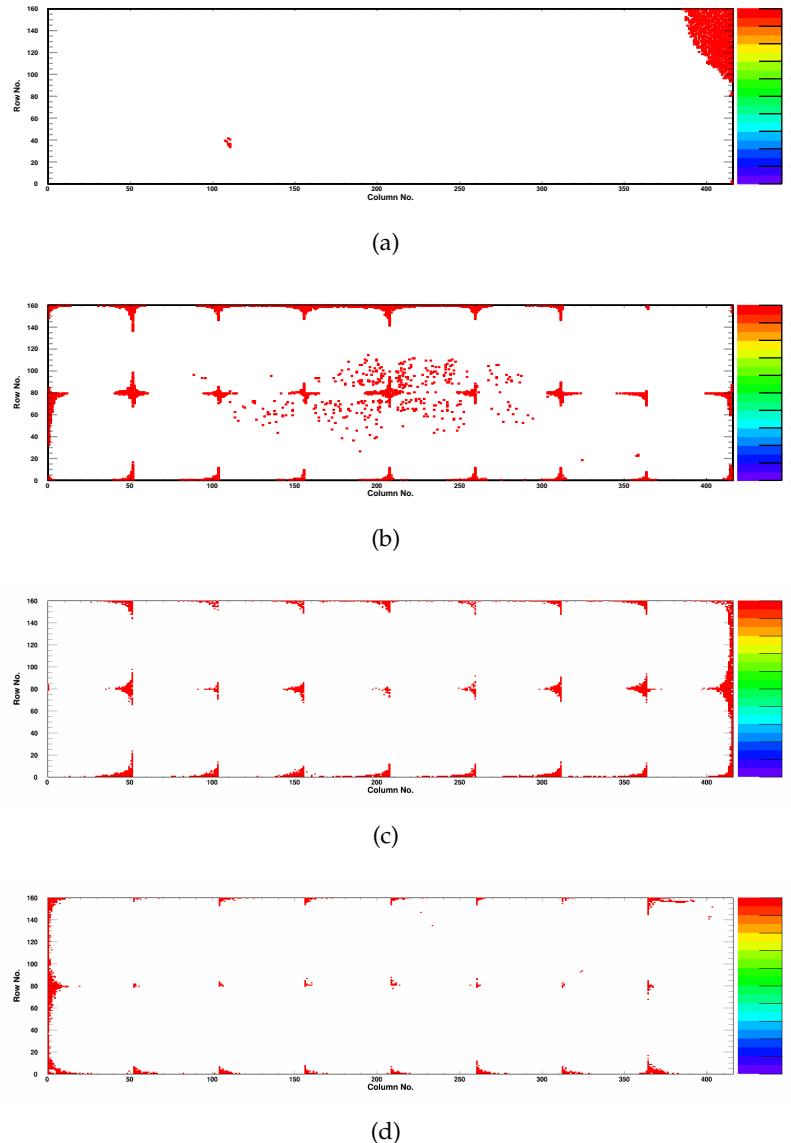


Figure 9.19: The module map of bump bonding defects observed at 17°C for the 4 modules with at least one ROC exceeding the 4% threshold of defective pixels because of bump bonding problems. The modules are consequently graded C.

Nevertheless, such defects will cause failure or anomalous results in one or more parts of the qualification. For this reason, the qualifications whose results are irregular or suspicious are set aside and inspected individually for further understanding. In this process, a few modules, produced consecutively, are found showing troublesome communication with only the two ROCs that are placed further away from the landing position of the power and signal cable on the HDI. The resistance of the power lines of the HDIs used for these modules is measured on a probe-station and found to be anomalously high, thus limiting the voltage supply to the ROCs as their distance from the cable increases. Given the impossibility to reliably operate these modules, a total of 8 are given a C grade and

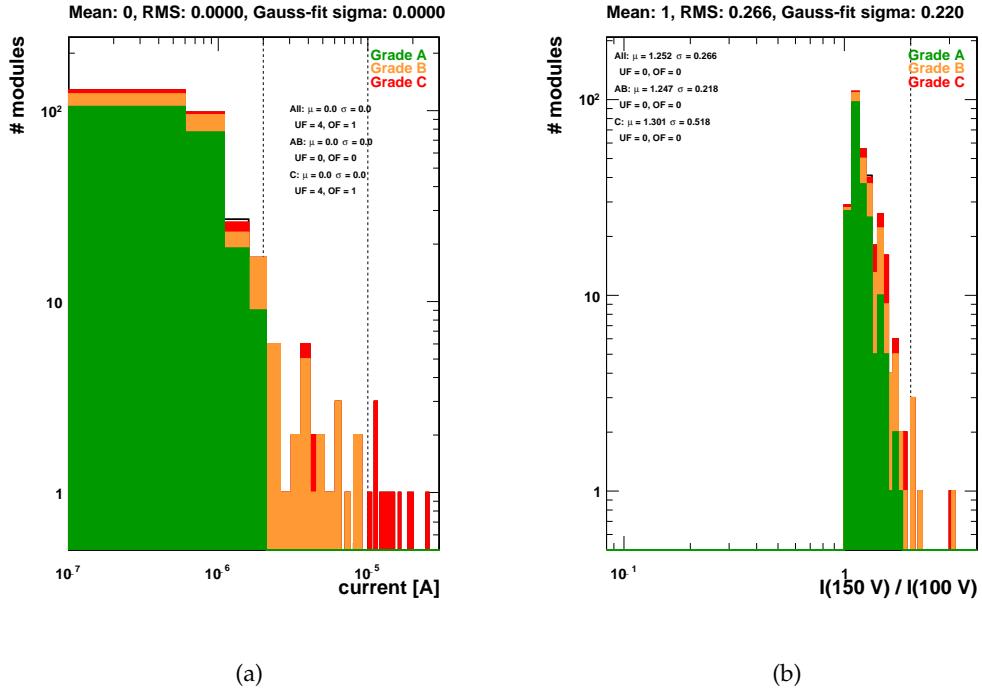


Figure 9.20: The distribution of the sensor leakage current measured for a bias voltage of -150 V at 17°C (a) and of the slope of the IV curve, defined as the ratio of the leakage currents at -150 and -100 V . The different colors identify modules with different final grade: green for A, yellow for B, and red for C.

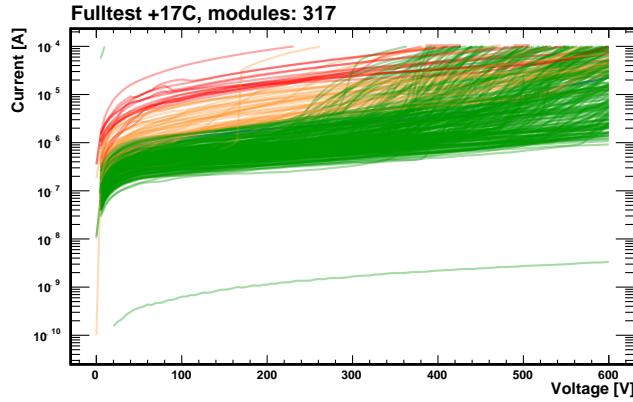


Figure 9.21: The overlay of the IV-curves of all tested modules, separately for grade-A (green), grade-B (yellow), and grade C (red) modules.

discarded for this reason. The quick identification of the issue prevented the production of more modules with HDIs from the same faulty batch.

9.3.6 Summary of module failures

A total of 62 modules in the whole production are discarded for installation out of the 326 produced, providing a production yield of 81%. The reasons for these modules failing the tests are described in detail in the Sections above and are summarized in Figure 9.22. The sensor quality, responsible for the grade C of 25 modules, is found to be the main cause of module failure. Errors in the double-column readout mechanism, when tested under high-rate X-ray radiation, account for the exclusion of 15 modules. 8 modules are rejected due to high resistance of the powering lines on their HDIs, while 6 modules show at least one ROC with more than 4% defective pixels. The non-programmability of at least one ROC causes 3 modules to be excluded and 3 more modules have been manually downgraded due to specific reasons.

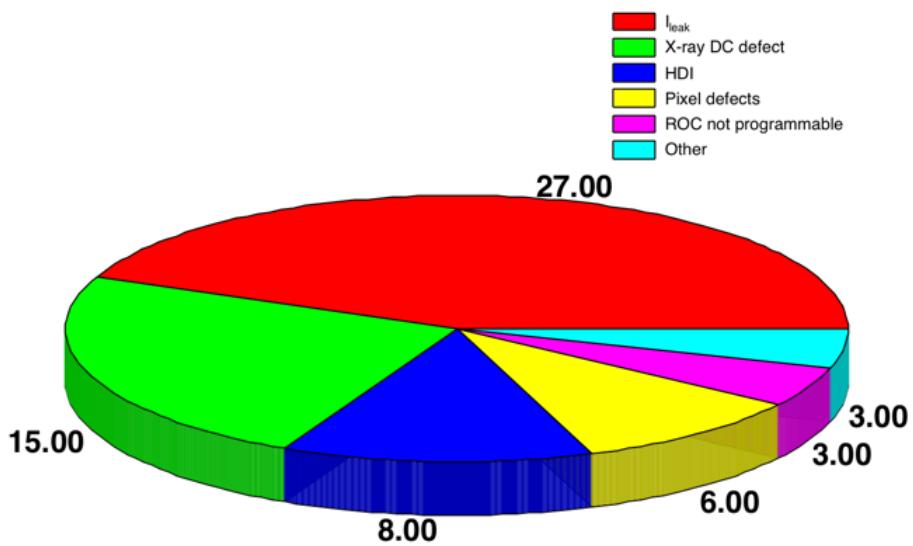


Figure 9.22: The summary of the failure modes responsible for the rejection of 62 modules for the construction of the second layer.

9.3.7 Potential impact on resolution of C-grade modules

As described in Section 9.2.3.1, the ROC grading scheme expands on the studies performed for the analog chip production, adapting it to the different features and behaviors of the new chip. Furthermore, the improved features of the digital chip require the tightening of some grading parameters, in order to restrict the grade-A and grade-B modules to the core of the distributions, selecting the best-quality chips. Nevertheless, it is important to estimate the impact of using grade-C chips on the detector performance and, in particular, on the position resolution. For this purpose, the passage of 10k muons with a momentum of 10 GeV and $|\eta| < 0.1$ through the pixel detector is simulated under different conditions, reproducing degradation corresponding to each of the grading crite-

ria of the ROC. The position of the reconstructed charge cluster is compared to the true position of the particle, separately for each layer and for each direction (x, y) in the local coordinate system on the layer. The difference between the reconstructed and the true position, called *residual*, is computed and the RMS of the residuals distribution is taken as an estimator of the position resolution along a given direction on a given detector layer. The simulation employs the geometry of the phase-0 detector, since at the time the study has been conducted the simulation of the phase-1 geometry had not been finalized. This approach, though, evaluates the position resolution layer by layer, and it does not require the full reconstruction of the muon track, which would be highly impacted by the different number of layers between the phase-0 and the phase-1 detectors. Furthermore, since the simulation entails phase-0 modules, care is taken to tune the module configuration and response to reproduce the digital chip properties, i.e. setting the pixel threshold, the threshold spread, the pixel noise to the mean values measured in the module qualifications. The impact of the different geometry is thus limited only to the different distances of the layers from the interaction point and it is considered to have a negligible effect on the interpretation of the simulation results.

In Figure 9.23, the impact on the mean pixel noise on the position resolution is shown. No degradation of the resolution is observed for noise values smaller than $500e^-$, with boundaries for A (B) grade placed at 200 (300) e^- . The resolution worsens only at high values of noise ($1000e^-$), where chips are graded C. The steeper trend observed for the resolution in the X direction is explained by charge sharing, which causes the individual pixel charge to be smaller in the X direction and thus more affected by larger noise.

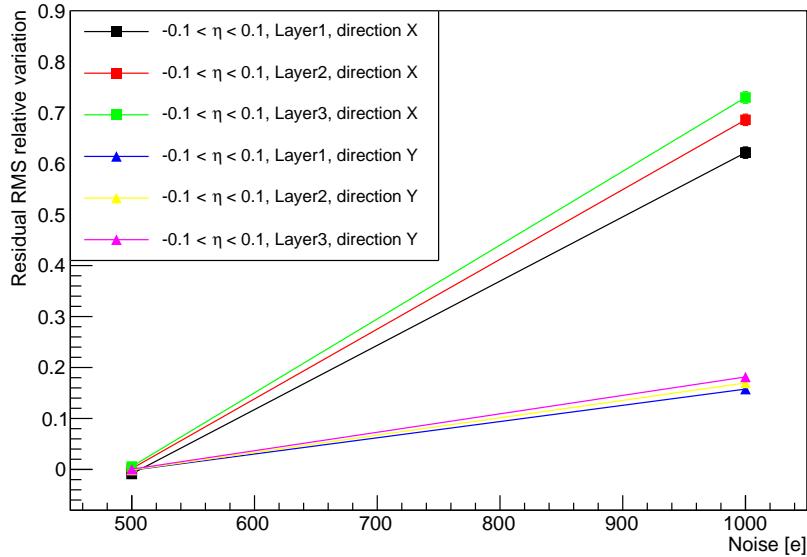


Figure 9.23: The impact of the pixel noise on the position resolution of the Pixel detector.

The effect of the width of the threshold distribution is shown in Figure 9.24, indicating that A- and B-grade modules do not impact the resolution significantly.

Similarly, Figures 9.25 and 9.26 report the resolution degradation as a function of the width of the relative PH gains and of the PH pedestal spread. The resolution is affected

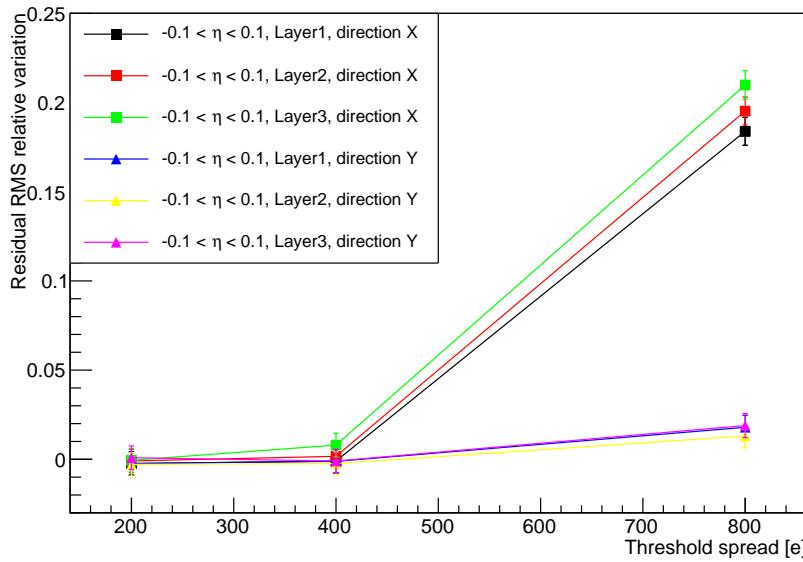


Figure 9.24: The impact of the threshold spread on the position resolution of the Pixel detector.

only by significant variations of the gain width ($> 2\%$), much larger than the criteria chosen even for C-grade modules ($> 0.2\%$). A sizable worsening of the resolution in the X direction is observed starting for pedestal spreads of $2500e^-$, while the Y direction remains unaffected, due to charge sharing in the X direction which makes the reconstructed charge more sensitive to approximations caused by the use of mean values for the pedestal. A-grade modules (gain widths smaller than 0.1, pedestal spread) are found to cause no impact on the resolution.

The effect of the fraction of defective pixels is studied by simulating the random inefficiency of a chosen fraction of pixels. The impact on the resolution is shown in Figure 9.27, where a worsening is observed for C-grade modules (4% of defective pixels) in the X direction. Since the average cluster size in the Y direction is of 1 pixel, the inefficiency in this direction has a negligible impact on the resolution: if the cluster size is larger than one pixel in the X direction, the information on the Y position can be preserved without loss of resolution if at least one of the pixels in the cluster records a signal; if the dimension of the cluster in X is 1 pixel or all the pixels in the cluster are inefficient, the cluster is lost completely and is not considered.

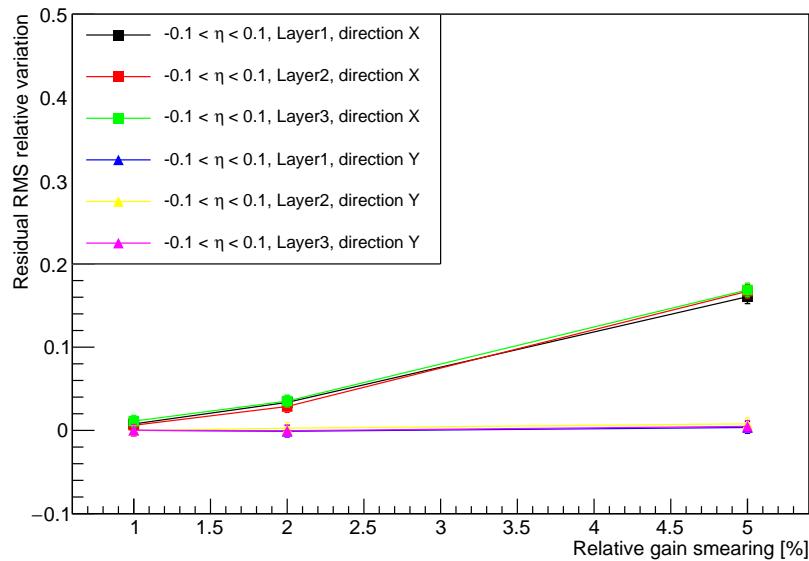


Figure 9.25: The impact of the width of the relative PH gains on the position resolution of the Pixel detector.

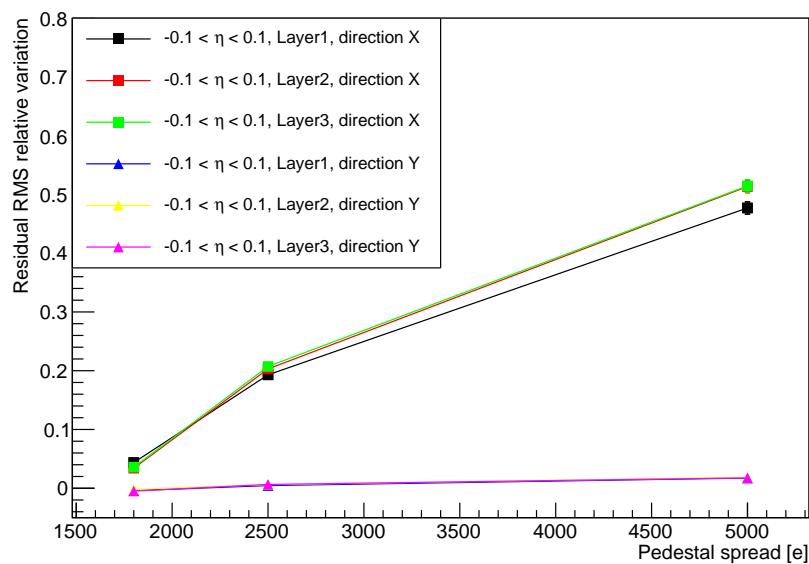


Figure 9.26: The impact of the PH pedestal spread on the position resolution of the Pixel detector.

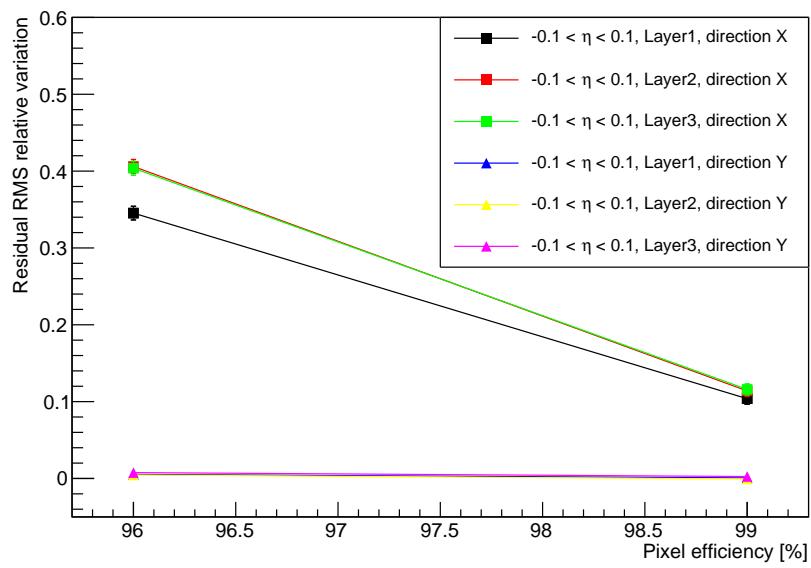


Figure 9.27: The impact of the fraction of defective pixels on the position resolution of the Pixel detector.

SUMMARY

This thesis presented the work performed to characterize the Higgs boson in its diphoton channel exploiting the data set of 35.9 fb^{-1} recorded by CMS in 2016. The production of the Higgs boson is explored first inclusively, by measuring the signal strength modifiers with respect to the SM predictions. The overall signal strength is measured to be $1.18^{+0.17}_{-0.14} = 1.18^{+0.12}_{-0.11} (\text{stat})^{+0.09}_{-0.07} (\text{syst})^{+0.07}_{-0.06} (\text{theo})$, being thus compatible with the SM value of 1. Signal strengths are also measured for individual production mechanisms and found to agree with the SM expectations.

The Higgs decay to a diphoton pair is studied in more detail by performing differential measurements of the production cross section with respect to a large number of observables, describing the production and decay kinematic features, as well as the properties of associated particles. The cross section is also measured inclusively and in regions of the phase space enriched in specific production modes (ttH , VH). The total cross section in the fiducial volume defined for the analysis presented in this thesis is found to be $\hat{\sigma}_{\text{fiducial}} = 84 \pm 11 \text{ (stat)} \pm 7 \text{ (syst)} \text{ fb} = 84 \pm 13 \text{ (stat+syst)} \text{ fb}$, in agreement with the theoretical prediction of $\sigma_{\text{fiducial}}^{\text{theory}} = 73 \pm 4 \text{ fb}$. The measurement of differential distributions are also found to overall agree with the SM calculations. These measurements represent the most precise and comprehensive study of the $H \rightarrow \gamma\gamma$ channel properties and cross sections performed by the CMS collaboration so far.

The installation in early 2017 of a new pixel detector has been a major upgrade to the CMS detector. A crucial aspect of the construction of a new system is the thorough investigation and precise calibration of its components: this thesis presented the qualification procedures designed for the production of the detector modules, together with the rationale of the quality criteria chosen to select modules for installation. The procedures implemented have been used to qualify more than 1200 modules, required to assemble the new detector. This thesis focused on the qualification results of the modules produced to constitute the second layer of the central region of the detector, for which 264 modules have been produced with a production yield of 81%.

*Ogni cura si doni al diletto,
e s'acorra nel magico tetto:
tra la folla de' creduli ognuno
s'abbandoni e folleggi con me.*

— A.Somma, G.Verdi

ACKNOWLEDGMENTS

Finally, I would like to express my gratitude to the people who contributed in one way or another to shape the work of the past four years that converged in this thesis.

First, I am grateful to Professors Rainer Wallny and Günther Dissertori, for letting me be part of their research groups and for their guidance through the journey of my PhD. They allowed me to pursue my personal ideas and interests, while constantly supporting and directing my work towards the final goals.

I would also like to thank Mauro Donegà, who incessantly kept a close eye on my steps throughout the work of these years, for giving me invaluable advice and for pushing me towards higher and higher objectives. A special thanks to Pasquale Musella, for being an incessant source of ideas and knowledge to whom I could turn to with any kind of problem. Thanks to Mauro and Pasquale both, for their incredible patience in relentlessly answering my questions, for the countless discussions and consultations, which have been indispensable to accomplish the results presented here, and for facing even the most serious situations with a smile and a tip of humor.

I wish to thank Andrey Starodumov, for guiding the work on the pixel detector and to whom I am indebted for sharing his knowledge and experience in several occasions.

I am grateful to the members of the research groups I have worked with, especially the $H \rightarrow \gamma\gamma$ analysis group and the ETH and PSI Pixel groups, who contributed in numerous ways to improve my work.

I had the pleasure to share offices and enjoy uncounted coffee breaks with so many interesting colleagues, first in Zurich and then at CERN. I am thankful to the PhD students I had the chance to meet along the way, of which many proved to be priceless companions in this long and dangerous journey. While naming everyone is impossible, a special mention goes to Jan, Simon, Philip, Felix, Pirmin, Maren, David, Myriam, Leonora, Nadya, Thomas K., and Thomas R., many of which turned into friends over this time. Furthermore, these years would not have been the same without all the laughter and the good memories, for which I thank Francesco M., Francesco P., Marc, Mario, Andrea Carlo, Paolo, Livia, Marco, Federico, jellysim, Giulia, Pier. I thank again Andrea Carlo and Paolo for the dozens of hikes together, which helped me keeping the right mental balance and improved my well-being and that of those around me. A special acknowledgment to Giuseppe, with whom I often shared the burdens ever since that first semester in Rome in 2008 and who has always been there to listen to my rants and encourage me.

I am lucky to have a big and special family (Paolo, Germana, Astrid, Ruben, Egle) that supports me and taught me to be open-minded and curious ever since I was little. I thank especially my sister Egle, for teaching me a lesson every day with her infinite strength.

Last, but surely not least, I thank Martina, for having her by my side made it possible to live this adventure and innumerable others.

LIST OF FIGURES

- Figure 2.1 The CERN accelerator complex [77]. 6
- Figure 2.2 Integrated luminosity per year delivered to CMS in proton-proton collisions [31]. 7
- Figure 2.3 An exploded view of the CMS detector [29]. 8
- Figure 2.4 Layout of the CMS ECAL [19]. 10
- Figure 3.1 The relative photon energy resolution for photons with $R_9 > 0.94$ as a function of the pseudorapidity of the photon. The resolution is measured in $Z \rightarrow e^+e^-$ events in data (black points) and simulation (red squares). The module boundaries in the EB are indicated by the vertical dashed lines, while the grey area corresponds to the transition area between EB and EE [67]. 17
- Figure 3.2 The invariant mass distributions for $Z \rightarrow e^+e^-$ is compared between data and simulation events, where electrons are reconstructed as photons. The distributions are shown in two particular bins where the corrections are derived, corresponding to events with both electrons in EB (left), and the remainder of the events (right), and with $R_9 > 0.94$. The distribution in simulation is normalized to the integral in data for $87 < m_{e^+e^-} < 93$ GeV [27]. 18
- Figure 4.1 The elementary particles of the SM of Particle Physics. Quarks and leptons are organized in three *flavor* families, each containing an up-type and a down-type quark, and an electrically charged lepton and a neutrino. The mediators of the forces are gauge bosons: the diagrams indicate the particles subject to the interaction mediated by a given boson. Adapted from [80]. 24
- Figure 4.2 The production cross section for the SM Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV is shown as a function of the mass of the Higgs boson, separately for different production mechanisms [35]. 27
- Figure 4.3 The Feynman diagrams for the four main production mechanisms in proton-proton collisions at the LHC: ggH (a), VBF (b), VH (c), and ttH (d, e, f). 29
- Figure 4.4 The Martin-Stirling-Thorne-Watt parametrization of the PDFs as a function of the momentum fraction of the proton, for two different energy scales [75]. 31
- Figure 5.1 The production cross section, for different production mechanisms in proton-proton collisions, for a Higgs boson with a mass of 125 GeV as a function of the center-of-mass energy [35]. 34
- Figure 5.2 (Left) Validation of the vertex identification BDT in data and simulation for $Z \rightarrow \mu^+\mu^-$ events, where the muon tracks are removed to mimic a diphoton event. (Right) Vertex finding efficiency on signal events as a function of the diphoton p_T , compared to the average vertex probability BDT score. 37

- Figure 5.3 (Left) The photon identification BDT score is compared between data (points) and simulation (histograms), separately for signal (red) and background (blue) events. The BDT output for the lowest-scoring photon of the pair is shown. (Right) The same distribution is shown for $Z \rightarrow e^+e^-$ events in data and simulation, where electrons are reconstructed as photons. The red hashed region shows the systematic uncertainty assigned to the shape of the discriminator. 39
- Figure 5.4 Data-simulation comparisons for BDT input variables (photon kinematic variables, mass resolution estimates under correct and incorrect vertex hypotheses, cosine of the angle in ϕ between the two photons, vertex probability). For every variable, the comparison between the cross-section-weighted mixture of simulated (stacked histograms) samples with data (black points), normalized to the area, is shown, together with the ratio plot. 41
- Figure 5.5 Distribution for simulated signal events of the BDT output. Different colors correspond to different bins of signal efficiency with respect to a selection performed on the BDT output. 43
- Figure 5.6 Distributions for simulated signal events of input variables (leading photon kinematic variables, mass resolution estimates under correct and incorrect vertex hypotheses, cosine of the angle in ϕ between the two photons). The different colors identify different bins of signal efficiency with respect to a selection performed on the BDT output. It can be seen that events classified in efficiency quantiles corresponding to higher BDT scores have higher p_T , more central pseudorapidities and higher photon identification BDT scores for both photons, narrower mass resolution, and photons closer in the ϕ angle. 44
- Figure 5.7 The diphoton mass distributions, normalized to unity in every signal efficiency bin, for background (left) and signal (right) events. 45
- Figure 5.8 The classifier 45
- Figure 5.9 Shape of the BDT output variable split for different background (left) and signal (right) samples. Background events with two prompt photons tend to have higher BDT scores. VBF, VH, and ttH processes tend to accumulate to higher BDT scores compared to gluons fusion. 46
- Figure 5.10 The score of the diphoton BDT for events with two photons in data (black points) and simulated signal (red shades) and background (histograms) events. The distribution has been transformed to have a flat distribution between 0 and 1 for the sum of the simulated signal processes. The vertical dashed lines represent the boundaries of the categories defined for the analysis, described in Section 5.9, and the grey shaded area indicates the events discarded by the categorization. 47

- Figure 5.11 The BDT input variables (photon kinematic variables, mass resolution estimates under correct and incorrect vertex hypotheses, cosine of the angle in ϕ between the two photons, vertex probability) are compared between data (black points) and simulation (filled histogram) on electrons reconstructed as photons from Z boson decay. 49
- Figure 5.12 Data-simulation comparison for diphoton BDT output for electrons reconstructed as photons from the decay of the Z boson. The red band indicates the impact on the BDT output variable of the systematic uncertainties on the photon identification BDT variable and on σ_E/E . 50
- Figure 5.13 Expected significance as a function of the number of analysis categories. The estimate of the expected significance is approximated since it is extracted from a simplified signal+background fit. It can be seen how the relative gain in expected significance becomes of the order of 1% after $n = 4$. 51
- Figure 5.14 Signal shape for the Untagged 1 category as a function of m_H (left) and evaluated at $m_H = 125$ GeV (right). The points indicated the simulated signal events and the blue lines the fitted models. 55
- Figure 5.15 The different sources of systematic uncertainties and their impact on the signal strengths, inclusively and for the VBF and $t\bar{t}H$ production mechanisms. The solid (empty) bars show the observed (expected) results. 58
- Figure 5.16 Number of expected signal events in each category, with the breakdown of the relative contributions of individual production mechanisms. 59
- Figure 5.17 The signal-plus-background model fits and data for the four Untagged categories. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit. 60
- Figure 5.18 The signal-plus-background model fits and data for the VBF and $t\bar{t}H$ categories. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit. 61
- Figure 5.19 The signal-plus-background model fits and data for the VH categories. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit. 62
- Figure 5.20 The signal-plus-background model fits and data, summer over all the categories of the analysis. The green (yellow) bands indicate the one (two) standard deviation uncertainty bands in the background component of the fit. 63
- Figure 5.21 The likelihood scan for the signal strength where the value of the SM Higgs boson mass is profiled in the fit. 63

- Figure 5.22 Signal strength modifiers measured for each process (black points), with the SM Higgs boson mass profiled, compared to the overall signal strength (green band) and to the SM expectation (dashed red line). 64
- Figure 5.23 The best fit value (black cross) of the two-dimensional likelihood scan for $\mu_{\text{ggH}, t\bar{t}\text{H}}$ versus $\mu_{\text{VBF}, \text{VH}}$ is compared to the SM expectation (red diamond). The solid (dashed) contour shows the 68% (95%) confidence region. 64
- Figure 5.24 The best fit values (black crosses) of the two-dimensional likelihood scan for κ_f versus κ_V , on the left, and κ_g versus κ_γ , on the right, are compared to their respective SM predictions (red diamonds). The solid (dashed) contour shows the 68% (95%) confidence region. 65
- Figure 6.1 The profile correlation distribution of the (decorrelated) mass resolution estimator as a function of the di-photon mass is shown in the plot on the left (right), for events where both photons are reconstructed in EB. The points (bars) correspond to the mean (RMS) of the mass resolution estimator in a given mass bin. 71
- Figure 6.2 The profile correlation distribution of the (decorrelated) mass resolution estimator as a function of the di-photon mass is shown in the plot on the left (right), for events where at least one photon is reconstructed outside EB. The points (bars) correspond to the mean (RMS) of the mass resolution estimator in a given mass bin. 71
- Figure 6.3 Distribution of $\sigma_m/m|_{\text{decorr}}$ in simulated diphoton background events. Each of the stacked histograms corresponds to different exclusive selections in pseudorapidities and R_9 of the two photons. 72
- Figure 6.4 The decorrelated mass resolution estimator $\sigma_m/m|_{\text{decorr}}$ for data and simulated $Z \rightarrow e^+e^-$ events. The distributions are shown separately for events with both electrons in EB (left) and the remainder of the events (right). The shaded area represents the region of values of $\sigma_m/m|_{\text{decorr}}$ where events are not selected for the analysis. 73
- Figure 6.5 The response matrices for $p_T^{\gamma\gamma}$ (a) and N_{jet} (b). The matrices are derived separately for categories 0, 1, and 2 of $\sigma_m/m|_{\text{decorr}}$, shown from left to right. The numbers in the cells of the matrices indicate the efficiency of a cell in selecting events in the corresponding generator-level bin, across all $\sigma_m/m|_{\text{decorr}}$ categories. 79
- Figure 6.6 The diphoton mass spectrum in data (black points), together with the best signal-plus-background fit (red lines), for each $\sigma_m/m|_{\text{decorr}}$ category employed for the measurement of the inclusive fiducial cross section. The green (yellow) band indicates the one (two) standard deviation uncertainty in the background component. 82
- Figure 6.7 Likelihood scan (black curve) for the fiducial cross section measurement, where the value of the SM Higgs boson mass is profiled in the fit. The measurement is compared to the theoretical prediction (vertical red line), with which it is in agreement within uncertainties. 83

- Figure 6.8 The measurement of the differential cross section for $p_T^{\gamma\gamma}$, N_{jet} , $|y^{\gamma\gamma}|$, and $|\cos(\theta^*)|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure. 85
- Figure 6.9 The measurement of the differential cross section for $p_T^{j_1}$, $|y^{j_1}|$, $|\Delta\phi^{\gamma\gamma,j_1}|$, and $|\Delta y^{\gamma\gamma,j_1}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure. 86
- Figure 6.10 The measurement of the differential cross section for $p_T^{j_2}$, $|y^{j_2}|$, $|\Delta\phi^{j_1,j_2}|$, and $|\Delta\phi^{\gamma\gamma,j_1,j_2}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure. 87
- Figure 6.11 The measurement of the differential cross section for $|\bar{\eta}_{j_1,j_2} - \eta_{\gamma\gamma}|$, m_{j_1,j_2} , and $|\Delta\eta_{j_1,j_2}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure. 88
- Figure 6.12 The measurement in a VBF-enriched sub-region of the fiducial phase space of the differential cross section for $p_T^{j_2}$, $|\Delta\phi^{j_1,j_2}|$, and $|\Delta\phi^{\gamma\gamma,j_1,j_2}|$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure. 89

- Figure 6.13 The measurement of the differential cross section for $p_T^{\gamma\gamma} \times N_{\text{jet}}$. The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last, overflow bin of the distribution is indicated in the figure. 90
- Figure 6.14 The measurement of the differential cross section for $p_T^{\text{miss}}, N_{\text{jet}}^b$, and N_{lepton} . The black points indicate the data, with the error bars showing 1 standard deviation total uncertainty. The blue band indicates the systematic component of the total uncertainty. The measurements are compared to the theoretical predictions, normalized to the predictions from Ref. [35] and obtained with different Monte Carlo programs. The normalization of the cross section in the last bin of the distribution, when this is an overflow bin, is indicated in the figure. 91
- Figure 6.15 The measurement of the differential cross section for different regions of the phase space, listed on the vertical axis. The black points indicate the data, with the black error bars showing the 1 standard deviation total uncertainty, while the blue band indicates the systematic component. The measurements are compared to the theoretical predictions (orange area), normalized to the predictions from Ref. [35]. 92
- Figure 7.1 The mean rate of energy loss by ionization in matter in different materials. The curves follow the Bethe-Bloch equation, given in 7.1 [95]. 96
- Figure 7.2 The hybrid pixel detector layout, sketched for a single pixel (a) [84] and for a full readout chip (b) [17]. 99
- Figure 8.1 The comparison between the geometrical layouts of the previous (bottom) and the new (top) CMS pixel detector [97]. Modified from [37]. 102
- Figure 8.2 The material budget, in units of radiation lengths, is compared between the phase-0 (green histogram) and the phase-1 (black dots) pixel detectors [37]. 103
- Figure 8.3 The schematics of the digital ROC. [55] 104
- Figure 8.4 The schematics of the PUC (upper part of the figure) and of the double-column periphery and control interface of the ROC (lower part) [55]. 104
- Figure 8.5 The ROC data format. The ROC header contains two trailing bits used to transmit the readback information (D) and to mark the start of a new readback word (S). Modified from [54]. 106
- Figure 8.6 Drawing showing separately the components of the modules installed in L2-4 and in the Fpix, from top to bottom: the module cable, the HDI with the TBM, the silicon sensor, the ROCs array, and the mounting strips. The FPix modules use a modified mounting system, while L1 modules employ two TBMs and the mounting strips are omitted to reduce the material budget. 107

- Figure 8.7 The sensor doping structure before (top) and after (bottom) irradiation, for p-type implants in n-substrate (left) and for n+-type implants in n-substrate (right). Modified from [84]. 108
- Figure 9.1 The hit efficiency map as a function of $V_{thrComp}$ and $CalDel$ DAC settings. The white square with the black dot indicates the optimal point chosen during the Pretest. 110
- Figure 9.2 The efficiency scan with respect to V_{cal} and V_{trim} DAC values, for the least sensitive pixel, used to optimize the V_{trim} setting. 112
- Figure 9.3 Illustration of the full exploitation of the dynamic range of the PH ADC for one pixel, as function of the signal charge in units of the V_{cal} DAC. The pixel responds with non-zero PH from the set threshold (green vertical line) and saturates the ADC for the largest signal of interest (orange vertical line). 113
- Figure 9.4 PH measurement for all pixels of a ROC at $V_{cal} = 60$, low-range (a) and $V_{cal} = 255$, high-range, (b). The maps are used to identify the lowest-response and highest-response pixel, respectively. 114
- Figure 9.5 PH response as a function of ($PhScale, PhOffset$) for the lowest-response (a) and highest-response (b) pixel. The scans are performed injecting signals close to the threshold of the lowest-response pixel for (a), and the largest signal of interest, here $V_{cal} = 100$, high-range, for (b). 115
- Figure 9.6 The combinations of ($PhScale, PhOffset$) which satisfy the optimization requirements described in the text. 116
- Figure 9.7 The measurement of the analog current I_{ana} with the DTB (a) and with the readback ADC (b) as a function of the V_{ana} DAC. 118
- Figure 9.8 (a) The plate of the cooling box where modules are placed for temperature-controlled qualification. The Peltier elements are installed right below the plate and provide the necessary cooling power. In the background, the cable-adapter cards and the DTBs to which is module is connected.
 (b) The cooling box, one DTB, and the Keithley power supply. 120
- Figure 9.9 The diagram of the qualification setup. The client-server controller, running on a PC, controls both software instances (the testing software) as well as hardware components of the setup: the cooling box, via the JUMO500 interface; the DTBs; the Keithley power supply, providing the bias voltage to the modules via the DTBs. 121
- Figure 9.10 Example of a temperature diagram as a function of time during module qualification. Each Fulltest requires about 70 minutes and 10 more minutes are needed for an IV curve recording. Each thermal cycle between 17°C and -25°C takes about 15 minutes, leading to a total qualification time around 8 hours. 122
- Figure 9.11 The cumulative number of modules with grades A or B (green) and grade C (red) as a function of time, during the production of L2 modules. 126

- Figure 9.12 The mean noise per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red). 127
- Figure 9.13 The width of the threshold distribution after the Trimming procedure per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red). 128
- Figure 9.14 The width of the relative gain distribution per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red). 129
- Figure 9.15 The spread of the pedestal distribution per ROC measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for ROCs in modules with final grade A (green), B (yellow), and C (red). 130
- Figure 9.16 The distribution of the pixel threshold, in units of V_{cal} , as measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for pixels in modules with final grade A (green), B (yellow), and C (red). 131
- Figure 9.17 The distribution of the pixel noise as measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for pixels in modules with final grade A (green), B (yellow), and C (red). 132
- Figure 9.18 The distribution of the pixel PH gain as measured in the Fulltests at -20°C , before (a) and after (b) the thermal stress test, and at 17°C (c). The distributions are shown for pixels in modules with final grade A (green), B (yellow), and C (red). 134
- Figure 9.19 The module map of bump bonding defects observed at 17°C for the 4 modules with at least one ROC exceeding the 4% threshold of defective pixels because of bump bonding problems. The modules are consequently graded C. 135
- Figure 9.20 The distribution of the sensor leakage current measured for a bias voltage of -150 V at 17°C (a) and of the slope of the IV curve, defined as the ratio of the leakage currents at -150 and -100 V . The different colors identify modules with different final grade: green for A, yellow for B, and red for C. 136
- Figure 9.21 The overlay of the IV-curves of all tested modules, separately for grade-A (green), grade-B (yellow), and grade C (red) modules. 136
- Figure 9.22 The summary of the failure modes responsible for the rejection of 62 modules for the construction of the second layer. 137
- Figure 9.23 The impact of the pixel noise on the position resolution of the Pixel detector. 138
- Figure 9.24 The impact of the threshold spread on the position resolution of the Pixel detector. 139

Figure 9.25	The impact of the width of the relative PH gains on the position resolution of the Pixel detector.	140
Figure 9.26	The impact of the PH pedestal spread on the position resolution of the Pixel detector.	140
Figure 9.27	The impact of the fraction of defective pixels on the position resolution of the Pixel detector.	141

LIST OF TABLES

Table 5.1	Summary of preselection requirements on shower shape variables and isolation quantities.	38
Table 6.1	The differential observables studied with the corresponding bins chosen, grouped by the region of fiducial phase space where the measurements are performed.	78
Table 9.1	The readback functions corresponding to different settings of the readback register.	117
Table 9.2	The list of parameters entering the ROC grading and the corresponding thresholds for A, B, and C grades.	124
Table 9.3	The list of parameters entering the sensor grading and the corresponding thresholds for A, B, and C grades.	124
Table 9.4	The list of parameters entering the grading of the tests under X-ray radiation and the corresponding thresholds for A, B, and C grades.	125

LISTINGS

ACRONYMS

ALICE	A Large Ion Collider Experiment
ADC	Analog to Digital Converter
APD	Avalanche Photo Diode
ATLAS	A Toroidal LHC ApparatuS

BDT	Boosted Decision Tree
BSM	Beyond the Standard Model
CDF	Cumulative Distribution Function
CERN	Conseil Européen pour la Recherche Nucléaire
CMOS	Complementary Metal-Oxide Semiconductor
CMS	Compact Muon Solenoid
CPU	Central Processing Unit
CSC	Cathode Strip Chambers
CSV	Combined Secondary Vertex
DAC	Digital to Analog Converter
DT	Drift Tubes
DTB	Digital Test Board
EB	ECAL Barrel
EE	ECAL Endcap
ECAL	Electromagnetic Calorimeter
ETHZ	Eidgenössische Technische Hochschule Zürich
FEC	Front-End Controller
FED	Front-End Driver
FWHM	full width at half maximum
HB	HCAL Barrel
HE	HCAL Endcap
HDI	High Density Interconnect
HF	HCAL Forward
HLT	High Level Trigger
HO	HCAL Outer
HCAL	Hadronic Calorimeter
HEP	High Energy Physics
HPD	Hybrid Photo Diode
L1	Level-1

LEP	Large Electron Positron
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LO	Leading Order
LS ₁	Long Shutdown 1
LS ₂	Long Shutdown 2
MC	Monte Carlo
MIP	Minimum Ionizing Particle
MVA	Multi-Variate Analysis
NIEL	Non-Ionizing Energy Loss
NLO	Next to Leading Order
NNLO	Next-to-Next to Leading Order
N ₃ LO	Next-to-Next-to-Next to Leading Order
PC	Personal Computer
PDF	Parton Distribution Function
PF	Particle Flow
PH	Pulse-Height
pQCD	perturbative Quantum Chromo Dynamics
PS	Proton Synchrotron
PU	Pile Up
PSB	Proton Synchrotron Booster
PUC	Pixel Unit Cell
QCD	Quantum Chromo Dynamics
QED	Quantum Electro Dynamics
QFT	Quantum Field Theory
ROC	Receiver Operating Characteristic
ROC	Read Out Chip
RPC	Resistive Plate Chambers
SC	Super Cluster

SiPM	Silicon PhotoMultiplier
SM	Standard Model
SPS	Super Proton Synchrotron
TBM	Token Bit Manager
T&P	Tag and Probe
TMVA	Toolkit for Multivariate Data Analysis with ROOT
UE	Underlying Event
VPT	Vacuum Photo-Triode

BIBLIOGRAPHY

- [1] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, and Aben. "Combined measurement of the Higgs Boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments." In: *Phys. Rev. Lett.* 114 (2015), p. 191803. doi: [10.1103/PhysRevLett.114.191803](https://doi.org/10.1103/PhysRevLett.114.191803). URL: <http://link.aps.org/doi/10.1103/PhysRevLett.114.191803>.
- [2] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys. Lett. B* 716 (2012), p. 1. doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [3] Georges Aad et al. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments." In: *Phys. Rev. Lett.* 114 (2015), p. 191803. doi: [10.1103/PhysRevLett.114.191803](https://doi.org/10.1103/PhysRevLett.114.191803).
- [4] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV." In: *JHEP* 08 (2016), p. 045. doi: [10.1007/JHEP08\(2016\)045](https://doi.org/10.1007/JHEP08(2016)045).
- [5] S. Agostinelli et al. "GEANT4 - a simulation toolkit." In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. doi: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [6] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. "NLO Higgs boson production via gluon fusion matched with shower in POWHEG." In: *JHEP* 0904 (2009), p. 002. doi: [10.1088/1126-6708/2009/04/002](https://doi.org/10.1088/1126-6708/2009/04/002).
- [7] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." In: *JHEP* 07 (2014), p. 079. doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [8] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, Achilleas Lazopoulos, and Bernhard Mistlberger. "High precision determination of the gluon fusion Higgs boson cross-section at the LHC." In: *Journal of High Energy Physics* 2016.5 (2016), p. 58. ISSN: 1029-8479. doi: [10.1007/JHEP05\(2016\)058](https://doi.org/10.1007/JHEP05(2016)058). URL: [https://doi.org/10.1007/JHEP05\(2016\)058](https://doi.org/10.1007/JHEP05(2016)058).
- [9] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties." In: (2013). Ed. by S Heinemeyer, C Mariotti, G Passarino, and R Tanaka.
- [10] Ryan Atkin. "Review of jet reconstruction algorithms." In: *Journal of Physics: Conference Series* 645.1 (2015), p. 012008. URL: <http://stacks.iop.org/1742-6596/645/i=1/a=012008>.
- [11] ATLAS and CMS Collaborations, LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005. CERN, 2011. URL: <http://cdsweb.cern.ch/record/1379837>.

- [12] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. “Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM.” In: *Journal of High Energy Physics* 2012.2 (2012), p. 88. ISSN: 1029-8479. DOI: [10.1007/JHEP02\(2012\)088](https://doi.org/10.1007/JHEP02(2012)088). URL: [https://doi.org/10.1007/JHEP02\(2012\)088](https://doi.org/10.1007/JHEP02(2012)088).
- [13] Richard D. Ball et al. “Parton distributions for the LHC Run II.” In: *JHEP* 04 (2015), p. 040. DOI: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040).
- [14] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II.” In: *J. Phys. G* 43 (2016), p. 023001. DOI: [10.1088/0954-3899/43/2/023001](https://doi.org/10.1088/0954-3899/43/2/023001).
- [15] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays.” In: *Phys. Rev. Lett.* 10 (1963), pp. 531–533. DOI: [10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [16] Stefano Carrazza, Stefano Forte, Zahari Kassabov, Jose Ignacio Latorre, and Juan Rojo. “An Unbiased Hessian Representation for Monte Carlo PDFs.” In: *Eur. Phys. J.* C75.8 (2015), p. 369. DOI: [10.1140/epjc/s10052-015-3590-7](https://doi.org/10.1140/epjc/s10052-015-3590-7).
- [17] G. B. Cerati et al. “Radiation tolerance of the CMS forward pixel detector.” In: *Nucl. Instrum. Meth. A* 600 (2009), pp. 408–416. DOI: [10.1016/j.nima.2008.11.114](https://doi.org/10.1016/j.nima.2008.11.114).
- [18] S. Chatrchyan et al. “The CMS experiment at the CERN LHC.” In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [19] S. Chatrchyan et al. “The CMS experiment at the CERN LHC.” In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [20] Serguei Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.” In: *Phys. Lett. B* 716 (2012), p. 30. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [21] Serguei Chatrchyan et al. “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV.” In: *JINST* 8 (2013), Po9009. DOI: [10.1088/1748-0221/8/09/P09009](https://doi.org/10.1088/1748-0221/8/09/P09009).
- [22] Serguei Chatrchyan et al. “Identification of b-quark jets with the CMS experiment.” In: *JINST* 8 (2013), Po4013. DOI: [10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013).
- [23] Serguei Chatrchyan et al. “Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV.” In: *JHEP* 06 (2013), p. 081. DOI: [10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081).
- [24] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.” In: *Phys. Lett. B* 716 (2012), p. 30. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [25] CMS Collaboration. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV.” Submitted to *J. Inst.* 2012.
- [26] CMS Collaboration. “Measurement of inclusive and differential Higgs boson production cross sections in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: (2018). submitted to *JHEP*. arXiv: [1807.03825 \[hep-ex\]](https://arxiv.org/abs/1807.03825).
- [27] CMS Collaboration. “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: (2018). Submitted to *JHEP*. arXiv: [1804.02716 \[hep-ex\]](https://arxiv.org/abs/1804.02716).

- [28] CMS *Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2257069>.
- [29] CMS *Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. There is an error on cover due to a technical problem for some items. Geneva: CERN, 2006. URL: <https://cds.cern.ch/record/922757>.
- [30] CMS Collaboration. “CMS Luminosity - Public Results.” 2017. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [31] CMS Collaboration. *CMS Luminosity Public Results*. 2018 (accessed August 20, 2018). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2017_proton_proton_13_TeV_collis.
- [32] John C. Collins and Davison E. Soper. “Angular Distribution of Dileptons in High-Energy Hadron Collisions.” In: *Phys. Rev.* D16 (1977), p. 2219. DOI: [10.1103/PhysRevD.16.2219](https://doi.org/10.1103/PhysRevD.16.2219).
- [33] G. Cowan. “A survey of unfolding methods for particle physics.” In: *Conf. Proc. C0203181* (2002). [,248(2002)], pp. 248–257.
- [34] Susanna Cucciarelli, Universitat Basel, Ivan-Kresimir fur, Basel, Switzerland/ D Kotlinski, PSI, Villigen, and Switzerland/. “Pixel Hit Reconstruction.” In: (2004). URL: <https://cds.cern.ch/record/1364681>.
- [35] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector.” 2016. URL: <https://arxiv.org/abs/1610.07922>.
- [36] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. “Handling uncertainties in background shapes: the discrete profiling method.” In: *JINST* 10 (2015), Po4015. DOI: [10.1088/1748-0221/10/04/P04015](https://doi.org/10.1088/1748-0221/10/04/P04015).
- [37] A Dominguez et al. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CERN-LHCC-2012-016. CMS-TDR-11. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. 2012. URL: <https://cds.cern.ch/record/1481838>.
- [38] *Electron and photon performance in CMS with the full 2016 data sample*. Tech. rep. CMS-DP-2017-004. CERN, 2017. URL: <https://cds.cern.ch/record/2255497>.
- [39] *Electron and Photon performance using data collected by CMS at $\sqrt{s} = 13$ TeV and 25 ns*. Tech. rep. CMS-DP-2015-067. CERN, 2015. URL: <https://cds.cern.ch/record/2118397>.
- [40] Victor Daniel Elvira. *Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment*. Tech. rep. CMS-NOTE-2004-020. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/800406>.
- [41] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons.” In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [42] Lyndon Evans and Philip Bryant. “LHC Machine.” In: *JINST* 3 (2008), So8001. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).

- [43] Marco Farina, Maxim Perelstein, and Nicolas Rey-Le Lorier. "Higgs Couplings and Naturalness." In: *Phys. Rev. D* 90.1 (2014), p. 015014. DOI: [10.1103/PhysRevD.90.015014](https://doi.org/10.1103/PhysRevD.90.015014). arXiv: [1305.6068 \[hep-ph\]](https://arxiv.org/abs/1305.6068).
- [44] *FPix module testing reference guide*. Tech. rep. CMS-doc-12690. CMS, 2015. URL: <https://cms-docdb.cern.ch/cgi-bin/DocDB>ShowDocument?docid=12690>.
- [45] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO." In: *JHEP* 12 (2012), p. 061. DOI: [10.1007/JHEP12\(2012\)061](https://doi.org/10.1007/JHEP12(2012)061).
- [46] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics." In: *Eur. Phys. J. C* 71 (2011), p. 1. DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0).
- [47] John Erthal Gaiser. "Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ' ." PhD thesis. SLAC, 1982. URL: <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-255.html>.
- [48] Shireen Gangal and Frank J. Tackmann. "Next-to-leading-order uncertainties in Higgs+2 jets from gluon fusion." In: *Phys. Rev. D* 87 (2013), p. 093008. DOI: [10.1103/PhysRevD.87.093008](https://doi.org/10.1103/PhysRevD.87.093008).
- [49] Sheldon L. Glashow. "The renormalizability of vector meson interactions." In: *Nuclear Physics* 10 (1959), pp. 107–117. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8). URL: <http://www.sciencedirect.com/science/article/pii/0029558259901968>.
- [50] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. "Event generation with SHERPA 1.1." In: *JHEP* 02 (2009), p. 007. DOI: [10.1088/1126-6708/2009/02/007](https://doi.org/10.1088/1126-6708/2009/02/007).
- [51] J. Goldstone. "Field theories with « Superconductor » solutions." In: *Il Nuovo Cimento (1955-1965)* 19.1 (1961), pp. 154–164. ISSN: 1827-6121. DOI: [10.1007/BF02812722](https://doi.org/10.1007/BF02812722). URL: <https://doi.org/10.1007/BF02812722>.
- [52] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. "Broken Symmetries." In: *Phys. Rev.* 127 (1962), pp. 965–970. DOI: [10.1103/PhysRev.127.965](https://doi.org/10.1103/PhysRev.127.965). URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- [53] V. Gromov, Anne J. Annema, R. Kluit, J.L. Visschers, and P. Timmer. "A Radiation hard bandgap reference circuit in a standard 0.13 μ m CMOS Technology." In: *IEEE transactions on nuclear science* 54.07CH37910C/6 (2007). DOI: [10.1109/TNS.2007.910170](https://doi.org/10.1109/TNS.2007.910170).
- [54] CMS Pixel group. *Digital ROC operation*. 2015 (accessed August 20, 2018). URL: <http://cms.web.psi.ch/phase1/psi46dig/index.html>.
- [55] CMS Pixel group. *Digital ROC documentation*. 2016 (accessed August 20, 2018). URL: <https://twiki.cern.ch/twiki/bin/view/CMS/Psi46dig>.
- [56] F. Halzen and Alan D. Martin. *Quarks And Leptons: An Introductory Course In Modern Particle Physics*. 1984. ISBN: 0471887412, 9780471887416.
- [57] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "MINLO: Multi-Scale Improved NLO." In: *JHEP* 1210 (2012), p. 155. DOI: [10.1007/JHEP10\(2012\)155](https://doi.org/10.1007/JHEP10(2012)155).
- [58] Keith Hamilton, Paolo Nason, Emanuele Re, and Giulia Zanderighi. "NNLOPS simulation of Higgs boson production." In: *JHEP* 10 (2013), p. 222. DOI: [10.1007/JHEP10\(2013\)222](https://doi.org/10.1007/JHEP10(2013)222).

- [59] Werner Herr and B Muratori. "Concept of luminosity." In: (2006). URL: <https://cds.cern.ch/record/941318>.
- [60] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [61] Gerard 't Hooft. "Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking." In: *NATO Sci. Ser. B* 59 (1980), pp. 135–157. DOI: [10.1007/978-1-4684-7571-5_9](https://doi.org/10.1007/978-1-4684-7571-5_9).
- [62] Jan Hendrik Hoss. "Search for Supersymmetry with Multiple Charged Leptons at $\sqrt{s} = 13$ TeV with CMS and Radiation Tolerance of the Readout Chip for the Phase I Upgrade of the Pixel Detector." PhD thesis. Zurich, ETH, 2017. DOI: [10.3929/ethz-b-000182698](https://doi.org/10.3929/ethz-b-000182698). URL: <https://www.research-collection.ethz.ch/handle/20.500.11850/182698>.
- [63] Thomas Junk. "Confidence level computation for combining searches with small statistics." In: *Nucl. Instrum. Meth. A* 434 (1999), p. 435. DOI: [10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2).
- [64] H. C. Kaestli. "Frontend electronics development for the CMS pixel detector upgrade." In: *Nucl. Instrum. Meth. A* 731 (2013), pp. 88–91. DOI: [10.1016/j.nima.2013.05.056](https://doi.org/10.1016/j.nima.2013.05.056).
- [65] Adam Kardos, Paolo Nason, and Carlo Oleari. "Three-jet production in POWHEG." In: *JHEP* 04 (2014), p. 043. DOI: [10.1007/JHEP04\(2014\)043](https://doi.org/10.1007/JHEP04(2014)043).
- [66] H. Chr. Kastli, M. Barbero, W. Erdmann, Ch. Hormann, R. Horisberger, D. Kotlinski, and B. Meier. "Design and performance of the CMS pixel detector readout chip." In: *Nucl. Instrum. Meth. A* 565 (2006), pp. 188–194. DOI: [10.1016/j.nima.2006.05.038](https://doi.org/10.1016/j.nima.2006.05.038).
- [67] Vardan Khachatryan et al. "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV." In: *JINST* 10 (2015), Po8010. DOI: [10.1088/1748-0221/10/08/P08010](https://doi.org/10.1088/1748-0221/10/08/P08010).
- [68] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV." In: *JINST* 12 (2017), Po2014. DOI: [10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014).
- [69] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction." In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652). URL: <http://dx.doi.org/10.1143/PTP.49.652>.
- [70] L. Landau. "On the energy loss of fast particles by ionization." In: *J. Phys.(USSR)* 8 (1944), pp. 201–205.
- [71] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*. Tech. rep. CERN-2013-004. CERN, 2013. DOI: [10.5170/CERN-2013-004](https://doi.org/10.5170/CERN-2013-004).
- [72] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. Tech. rep. CERN-2017-002-M. CERN, 2017. DOI: [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002). URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.

- [73] Xiaohui Liu and Frank Petriello. “Reducing theoretical uncertainties for exclusive Higgs-boson plus one-jet production at the LHC.” In: *Phys. Rev. D* 87 (2013), p. 094027. DOI: [10.1103/PhysRevD.87.094027](https://doi.org/10.1103/PhysRevD.87.094027).
- [74] M. Anfreville et al. “Laser monitoring system for the CMS lead tungstate crystal calorimeter.” In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 594.2 (2008), pp. 292–320. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2008.01.104>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900208001599>.
- [75] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. “Parton distributions for the LHC.” In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. doi: [10.1140/epjc/s10052-009-1072-5](https://doi.org/10.1140/epjc/s10052-009-1072-5).
- [76] “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV.” In: *Journal of High Energy Physics* 2016.8 (2016), p. 45. ISSN: 1029-8479. DOI: [10.1007/JHEP08\(2016\)045](https://doi.org/10.1007/JHEP08(2016)045). URL: [http://dx.doi.org/10.1007/JHEP08\(2016\)045](http://dx.doi.org/10.1007/JHEP08(2016)045).
- [77] Esma Mobs. “The CERN accelerator complex. Complexe des accélérateurs du CERN.” In: (2016). General Photo. URL: <https://cds.cern.ch/record/2197559>.
- [78] “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV.” In: *Journal of Instrumentation* 10.06 (2015), Po6005. URL: <http://stacks.iop.org/1748-0221/10/i=06/a=P06005>.
- [79] *Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector*. Tech. rep. CMS-PAS-JME-16-004. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2205284>.
- [80] Andrew Purcell. “Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules.” In: BUL-NA-2012-269. 35/2012 (2012), p. 10. URL: <https://cds.cern.ch/record/1473657>.
- [81] R. Boughezal, X. Liu, F. Petriello, F. J. Tackmann, and J. R. Walsh. “Combining resummed Higgs predictions across jet bins.” In: *Phys. Rev. D* 89 (2014), p. 074044. DOI: [10.1103/PhysRevD.89.074044](https://doi.org/10.1103/PhysRevD.89.074044).
- [82] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. “Probing color singlet exchange in Z + two jet events at the CERN LHC.” In: *Phys. Rev. D* 54 (1996), pp. 6680–6689. DOI: [10.1103/PhysRevD.54.6680](https://doi.org/10.1103/PhysRevD.54.6680).
- [83] Alexander L. Read. “Presentation of search results: The CL(s) technique.” In: *J. Phys. G* 28 (2002), p. 2693. DOI: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [84] L. Rossi, P. Fischer, T. Rohe, and N. Wermes. *Pixel Detectors*. Particle Acceleration and Detection. Berlin: Springer-Verlag, 2006. ISBN: 9783540283324, 9783540283331. DOI: [10.1007/3-540-28333-1](https://doi.org/10.1007/3-540-28333-1). URL: <http://www.springer.com/book/3-540-28333-1>.
- [85] Marco Rossini. “Module Prototype Qualification for the CMS Pixel Detector Upgrade.” PhD thesis. Zurich, ETH, 2015. DOI: [10.3929/ethz-a-010594693](https://doi.org/10.3929/ethz-a-010594693). URL: <https://www.research-collection.ethz.ch/handle/20.500.11850/155450>.

- [86] Abdus Salam and J. C. Ward. "Weak and electromagnetic interactions." In: *Il Nuovo Cimento (1955-1965)* 11.4 (1959), pp. 568–577. ISSN: 1827-6121. DOI: [10.1007/BF02726525](https://doi.org/10.1007/BF02726525). URL: <https://doi.org/10.1007/BF02726525>.
- [87] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector." In: *JINST* 12 (2017), P10003. DOI: [10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003).
- [88] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV." In: *Submitted to JINST* (2017).
- [89] Albert M Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV." Submitted to *JINST*. 2017.
- [90] Albert M Sirunyan et al. "Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV." In: *Phys. Lett. B* 776 (2018), p. 355. DOI: [10.1016/j.physletb.2017.11.043](https://doi.org/10.1016/j.physletb.2017.11.043).
- [91] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A brief introduction to PYTHIA 8.1." In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036).
- [92] P Speckmayer, A Hocker, J Stelzer, and H Voss. "The toolkit for multivariate data analysis: TMVA 4." In: *J. Phys.: Conf. Ser.* 219 (2010), p. 032057. URL: <https://cds.cern.ch/record/1270192>.
- [93] Iain W. Stewart and Frank J. Tackmann. "Theory uncertainties for Higgs and other searches using jet bins." In: *Phys. Rev. D* 85 (2012), p. 034011. DOI: [10.1103/PhysRevD.85.034011](https://doi.org/10.1103/PhysRevD.85.034011).
- [94] Iain W. Stewart, Frank J. Tackmann, Jonathan R. Walsh, and Saba Zuberi. "Jet p_T resummation in Higgs production at $NNLL' + NNLO$." In: *Phys. Rev. D* 89 (2014), p. 054001. DOI: [10.1103/PhysRevD.89.054001](https://doi.org/10.1103/PhysRevD.89.054001).
- [95] M. Tanabashi et al. "Particle Data Group." In: *Phys. Rev. D* 030001 (2018), p. 98.
- [96] V.R. Tavolaro. "Measurements of Higgs boson production and properties in the diphoton decay channel using the CMS detector." In: *Proceedings of 38th International Conference on High Energy Physics* (2016). PoS(ICHEP2016)385.
- [97] V.R. Tavolaro. "The Phase1 CMS Pixel detector upgrade." In: *Journal of Instrumentation* 11.12 (2016), p. C12010. URL: <http://stacks.iop.org/1748-0221/11/i=12/a=C12010>.
- [98] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/349375>.
- [99] by The Operations group, Stéphane Fartoukh (BE-ABP), and Stefania Pandolfi. "LHC Report: record luminosity, well done LHC." In: (2017). URL: <http://cds.cern.ch/record/2295012>.
- [100] *Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s} = 13$ TeV in pp collisions at CMS*. Tech. rep. CMS-PAS-HIG-16-020. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2205275>.
- [101] P. V. Vavilov. "Ionization losses of high-energy heavy particles." In: *Sov. Phys. JETP* 5 (1957). [Zh. Eksp. Teor. Fiz. 32, 920(1957)], pp. 749–751.

- [102] Steven Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (1967), p. 1264. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [103] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. "Experimental Test of Parity Conservation in Beta Decay." In: *Phys. Rev.* 105 (1957), pp. 1413–1415. DOI: [10.1103/PhysRev.105.1413](https://doi.org/10.1103/PhysRev.105.1413). URL: <https://link.aps.org/doi/10.1103/PhysRev.105.1413>.
- [104] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance." In: *Phys. Rev.* 96 (1954), pp. 191–195. DOI: [10.1103/PhysRev.96.191](https://doi.org/10.1103/PhysRev.96.191). URL: <https://link.aps.org/doi/10.1103/PhysRev.96.191>.

Vittorio Raoul Tavolaro

Curriculum Vitae

Personal Information

Gender Male
Date of birth 2 April 1989
Place of birth Locarno, Switzerland
Nationalities Italian, Swiss
Address 3bis Route de Drize, 1227, Carouge, Switzerland
Telephone +41 787839489, +39 3280950809
Email vittorio.tavolaro@cern.ch, vittorio.raoul.tavolaro@gmail.com

Professional history

Current position Ph.D. student in Particle Physics, Eidgenössische Technische Hochschule Zürich (ETHZ), I am completing my Ph.D in Particle Physics at ETHZ, in collaboration with the CMS experiment since Apr 2014 at **CERN**. The main topics of my thesis include silicon pixel detector construction and the study of the Higgs boson properties in the diphoton decay channel. I am also considerably involved in teaching activities. Advisors: Profs. Rainer Wallny and Günther Dissertori.

Mar 2013 - **Scholarship holder at Accademia Nazionale dei Lincei**, I was awarded the "Enrico Persico" scholarship by the Academia, based on my academic record and my master thesis project.

Feb 2013 - **INFN associate**, I carried out my master thesis in the INFN-Roma I group, in collaboration Sep 2013 with the CMS experiment. Supervisor: Dr. Paolo Meridiani.

Jun 2012 - **CERN Summer Student**, I was chosen to be part of the CERN Summer Student Program, Sep 2012 working in the Higgs Physics Analysis Group of the CMS experiment. Supervisor: Dr. Alessio Bonato.

Scientific activity

Sep 2015 - **Higgs boson physics**, I have a major role in the study of the Higgs boson properties in its present diphoton decay channel. I have strongly contributed to the study and understanding of several aspects of the analysis, built on the basis of the Run1 developments, guiding its evolution towards precision measurements using LHC Run2 data. My contributions range from analysis optimisation to results interpretation. The effort lead to the standalone observation of the Higgs boson in the diphoton channel in 13TeV data in summer 2016 [1]. The latest and most precise results are described in [2], which has been recently submitted to JHEP.

I am also the principal author of the differential measurements of fiducial cross sections for the production of the Higgs boson in the diphoton channel with 13 TeV data, contributing at all levels of the analysis. The latest results are described in [3], recently submitted to JHEP. This publication will represent one of the most comprehensive and precise studies of the Higgs boson fiducial cross sections. I am the contact person in the CMS experiment for this analysis and the editor of the paper.

As part of my CERN Summer Student Project in 2012, I have performed feasibility studies for the measurement of the Higgs boson production in the $ZZ^* \rightarrow 2l2q$ channel.

- Mar 2014 - **CMS Pixel detector Phase I Upgrade**, A new pixel silicon detector has been installed in the CMS detector in Dec 2016. I had several responsibilities in the construction of its core components, the detector modules [4]. I contributed to the verification of the readout chip design and characterisation of module prototypes; I am also a principal author of the design and implementation of the calibration and test procedures modules undergo prior to installation in the final system. This process allowed identifying and discarding sub-standard components and, at the same time, optimising single module's performance and capabilities in order to achieve improved detector operation. I am also one of the main authors of the design and construction of dedicated test benches for module testing. I was responsible for the testing of Pixel detector modules for the Swiss Consortium (ETHZ, UZH, PSI) (2015-2016).
- Feb 2013 - **SM Physics**, In the context of my master thesis, I have developed a new approach for the measurement of the isolated photon differential cross section at the LHC using the CMS detector. A first measurement using 8 TeV data is described in my thesis [7] .

Education

- Mar 2014 - **Ph.D. in Particle Physics**, *I am expecting to obtain a Ph.D in Particle Physics at ETHZ in 2018. Advs: Profs. R. Wallny and G. Dissertori.*
- Sep 2011 - **Master degree in Physics**, *La Sapienza University of Rome*, Master degree in Nuclear and Subnuclear Physics, final grade 110/110 cum laude. Grade point average: 30.0/30. Thesis: *Measurement of the differential cross section for the production of isolated photons in pp collisions at $\sqrt{s} = 7$ TeV at LHC using the CMS detector*. Supvr.: Dr. P. Meridiani.
- Oct 2012 - **Erasmus Program: Master 2 Recherche en Physique**, *Spécialité Noyaux, particules, astroparticules et cosmologie, Pierre and Marie Curie University, Paris VI.*
- Feb 2013
- 2008 - 2011 **Bachelor degree in Physics**, *La Sapienza University of Rome*, 110/110 cum laude. Thesis: Optimisation processes through genetic algorithms. Supvr.: Prof. G. Organtini .
- Sep 2008 - **Collegio Universitario "Lamaro Pozzani"**, *Scholarship holder, funded by the Federazione Nazionale dei Cavalieri del Lavoro.*
The College offers students of particular merit a broader education in foreign languages and business fundamentals. It organises cultural activities and study trips (Iran, China, Turkey Cuba).
- 2008 **Maturità classica**, *Liceo Statale "G.Palmieri", Lecce, Italy. Final grade 100/100 cum laude.*

Honours and Awards

- May 2014 **Laureato Eccellente Sapienza 2012/2013**, *La Sapienza University of Rome.*
The price allows the best 12 students of the Master in Physics to follow dedicated series of in-depth lectures.
- Mar 2013 - **Enrico Persico fellowship, by Accademia Nazionale dei Lincei.**
- Oct 2013 *Four students in Physics every year are awarded this fellowship by the Accademia.*
- February 2013 **ARAP prize for students**, *Associazione Romana per le AstroParticelle (ARAP).*
The prize is bestowed upon the best students in Astrophysics and Physics of Elementary Particles.

Presentations and schools

- Jul 2017 **ICHEP2018**, *Fiducial inclusive and differential Higgs boson cross sections at CMS*, Parallel Talk, Seoul, South Korea.
- Jul 2017 **EPS HEP2017**, *Measurement of the Higgs boson differential and fiducial cross sections at CMS*, Parallel Talk, Venice, Italy.
- Sep 2016 **PIXEL2016**, *The Phase1 CMS Pixel detector upgrade*, Invited Talk, Sestri Levante, Italy.
- Aug 2016 **ICHEP2016**, *Measurements of Higgs boson production and properties in the di-photon decay channel using the CMS detector*, Parallel Talk, Chicago, IL.
- Aug 2016 **ICHEP2016**, *The phase 1 upgrade of the CMS pixel detector: qualification of barrel pixel detector modules*, Poster, Chicago, IL.
- Sep 2015 **Joint ETHZ/UZH/PSI PhD seminars**, *Qualification of pixel detector modules for CMS Upgrade Phase I*, Plenary Talk, PSI, Switzerland.

- Jul 2014 **Swiss Physical Society meeting 2014**, CMS Upgrade Phase I: pixel modules testing, Parallel Talk , Fribourg, Switzerland.
- Jul 2014 **Swiss Physical Society meeting 2014**, CMS pixel phase 1 upgrade, Parallel Talk, Fribourg, Switzerland.
- Jan 2015 **CMSDAS 2015**, CMS Data Analysis School, Bari, Italy.
- Aug 2014 **HCPSS 2014**, Hadron Collider Physics Summer School, Fermi National Accelerator Laboratory, Batavia, IL.

Publications and reports

- [1] [CMS Collaboration] V.R. Tavolaro et al. *Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s} = 13$ TeV in pp collisions at CMS*. Tech. rep. CMS-PAS-HIG-16-020. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2205275>.
- [2] [CMS Collaboration], V.R. Tavolaro et al. "Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV". Submitted to JHEP. 2018.
- [3] [CMS Collaboration], V.R. Tavolaro et al. "Measurement of inclusive and differential Higgs boson production cross sections in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV". Submitted to JHEP. 2018.
- [4] V.R. Tavolaro et al. "Test beam performance measurements for the Phase I upgrade of the CMS pixel detector". In: *Journal of Instrumentation* 12.05 (2017), P05022. URL: <http://stacks.iop.org/1748-0221/12/i=05/a=P05022>.
- [5] V.R. Tavolaro. "Measurements of Higgs boson production and properties in the di-photon decay channel using the CMS detector". In: *Proceedings of 38th International Conference on High Energy Physics* (2016). PoS(ICHEP2016)385.
- [6] V.R. Tavolaro. "The Phase1 CMS Pixel detector upgrade". In: *Journal of Instrumentation* 11.12 (2016), p. C12010. URL: <http://stacks.iop.org/1748-0221/11/i=12/a=C12010>.
- [7] V.R. Tavolaro. *Misura nell'esperimento CMS a LHC della sezione d'urto di produzione di fotoni isolati in collisioni pp all'energia del centro di massa di 8 TeV*. 2013. URL: <http://www.roma1.infn.it/cms/tesi/tavolaro.pdf>.

I am co-author of more than 320 scientific publications as a member of the CMS collaboration, corresponding to an index h_{HEP} of 49.

Teaching and supervision

- Sep 2014 - **Teaching assistant at ETH Zürich**, Courses:
 present Nuclear and Particle Physics (3rd year undergraduate students)
 Physics II (1st year undergraduate students)
 Physics I (1st year undergraduate students)
 Physics for Engineers (2nd year undergraduate students)
 Laboratory courses (2nd year undergraduate students).
- Mar 2015 - **Master student supervisor, semester project**, Title: *Grading of pixel digital modules based on cooling box tests*, author A. Battaglioni.
 Sep 2015 on cooling box tests, author A. Battaglioni.
- Apr 2016 - **CMS guide**, I guide groups of visitors from the general public to tours of CMS detector.
 present

Languages

- Italian **Mother-tongue**.
- English **Fluent**.
- French **Good**.
- German **Basic**.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both L^AT_EX and LyX:

<https://bitbucket.org/amiede/classicthesis/>

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>

Thank you very much for your feedback and contribution.