SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of

Physics “Giuseppe Occhialini”

PhD program Physics and Astronomy Cycle XXXIII

Curriculum in Subnuclear Physics

Development of silicon pixel sensors for the High Luminosity upgrade of the CMS experiment at LHC and search for Higgs boson pair production in the $b\bar{b}\tau^{+}\tau^{-}$ final state at $\sqrt{s} = 13$ TeV

Surname Zuolo Name Davide

Registration number 797856

Tutor: Dr. Dario Menasce

Supervisor: Prof. Luigi Moroni

Coordinator: Prof. Marta Calvi

ACADEMIC YEAR 2019-2020

Contents

Moi je construis des marionnettes
avec de la ficelle et du papier
Elles sont jolies les mignonnettes
Elles vous diront, elles vous diront
Que je suis leur ami, que je suis leur ami

(Cristophe - Les marionettes)

To all my CERN friends, who stood beside me in my long months in the local area

Sommario

Durante il mio percorso di dottorato mi sono occupato principalmente dello sviluppo di nuovi sensori a pixel in Silicio per l'upgrade dell'esperimento CMS in vista della fase ad alta luminosità del collisionatore LHC (HL – LHC o LHC fase 2) al CERN, un'attività che avevo già intrapreso per il mio lavoro di tesi magistrale. Dopo circa due anni di lavoro sulla caratterizzazione dei nuovi prototipi ho sviluppato un interesse anche per l'impatto del rivelatore a pixel sul programma di fisica dell'esperimento. All'inizio del mio secondo anno di dottorato mi sono messo in cerca di un'analisi in cui le informazioni del rivelatore a pixel fossero largamente sfruttate. Ho scelto di lavorare alla ricerca della produzione di coppie di bosoni di Higgs nello stato finale con 2 b-jets e 2 leptoni tau. Ho scelto questa analisi sia perchè la ricostruzione delle particelle nello stato finale fa ampio uso delle informazioni di tracce e vertici sia perchè l'analisi trarrà grande benificio dell'incremento della statistica previsto ad HL-LHC, che potrebbe portare ad avere la prima evidenza sperimentale di produzione HH.

Lavoro di strumentazione

Un gruppo di lavoro INFN composto da ricercatori appartenenti alle collaborazioni ATLAS e CMS lavora dal 2015 con la Fondazione Bruno Kessler (FBK) di Trento allo sviluppo di nuovi sensori a pixel in Silicio per l'upgrade dell'esperimento CMS in vista delle fase 2 di LHC.

Il nuovo rivelatore a pixel sarà posto ad una distanza di circa 3 cm dal punto di interazione. A questa distanza la fluenza di irraggiamento a cui sarà sottoposto il layer più interno del rivelatore sarà di circa $2 \times 10^{16}\, n_{eq}/cm^2$, dopo 7 anni di operazioni e una luminosità integrata di 2500 fb^{-1}. I sensori planari attualmente installati nell'esperimento non possono essere utilizzati dopo l'esposizione a fluenze superiori ad alcuni $10^{15}\, n_{eq}/cm^2$ a causa della degradazione del rapporto segnale-rumore, che non può essere mitigata in alcun modo. É quindi necessario studiare un nuovo sensore in grado di tollerare fluenze dieci volte superiori a quelle citate.

Il gruppo di lavoro ha sviluppato prototipi di sensori planari sottili e 3D di tipo n-in-p impiegando una tecnologia recente per la realizzazione dei wafer nota come Direct Wafer Bonding che permette di attuare tutti i processi di fabbricazione su una sola superficie del wafer, riducendo notevolmente i costi del processo. Lo spessore attivo dei primi prototipi è di 100 o 130 μm per i sensori planari e 130 μm per quelli 3D mentre la dimensione dei pixel è la stessa dei sensori installati nell'esperimento, $100 \times 150\, \mu m^2$.

Sono state effettuate numerose campagne di prova su fascio per caratterizzare i sensori planari e quelli 3D, prima e dopo l'irraggiamento. I sensori erano tutti collegati al chip di lettura (readout chip, ROC) attualmente utilizzato in CMS (PSI46dig) che ha una resistenza alla radiazione limitata a circa $5 \times 10^{15}\, n_{eq}/cm^2$, molto inferiore alle fluenze di radiazione attese ad HL-LHC. Ciò nonostante i risultati ottenuti sono stati importanti per indirizzare i passi successivi del programma di sviluppo e sono stati pubblicati su una rivista peer-reviewed all'inizio del 2020. I miei contributi alla caratterizzazione dei prototipi sono riportati in questa tesi e includono anche lo sviluppo di algoritmi dedicati.

Durante la seconda metà del 2018 è iniziata la produzione del primo prototipo del ROC di fase-2, chiamato RD53A. Questo ROC ha una tolleranza alla radiazione molto superiore a quella del PSI46dig, può essere operato a soglie inferiori a 1200 elettroni prima

dell'irraggiamento e ha celle di dimensione $50 \times 50\,\mu m^2$. La riduzione della dimensione delle celle è determinata dalla necessità di aumentare il numero di canali di lettura a causa della maggiore molteplicità di tracce attesa ad HL-LHC. In conseguenza di ciò sono stati prodotti sensori planari e 3D con pixel di dimensioni ridotte di un fattore 6 ($50 \times 50\,\mu m^2$ e $25 \times 100\,\mu m^2$). I risultati che ho ottenuto durante le prime campagne di prova su fascio di questi nuovi sensori sono riportati in questa tesi.

L'obiettivo delle prossime campagne è quello di quantificare il cross-talk tra pixel adiacenti nei sensori $25 \times 100\,\mu m^2$, determinato dallo schema con cui le celle di sensore e ROC sono collegate le une alle altre, imposto dalle differenti dimensioni. É importante sottolineare che il cross talk è presente solo nei sensori planari e non in quelli 3D. Questo porta a favorire l'utilizzo dei sensori 3D, almeno nelle regioni più interne del rivelatore, dove le fluenze di radiazione sono più alte.

Lavoro di analisi

La ricerca di coppie di bosoni di Higgs ha un ruolo fondamentale nella caratterizzazione di questa particella in quanto rappresenta il canale migliore per la misura della costante di accoppiamento trilineare λ_{HHH}. La forma del potenziale di Higgs è descritta nel Modello Standard da soli tre parametri: la massa del bosone di Higgs m_H, il valore di aspettazione del vuoto v e la costante di accoppiamento trilineare dell'Higgs λ_{HHH}. L'ultimo di questi parametri è l'unico non ancora misurato. Ogni deviazione dalle previsioni teoriche del Modello Standard porterebbe a cambiamenti importanti nella cinematica e nel rateo di produzione di coppie di bosoni di Higgs, il che rende queste ricerche sensibili a effetti di nuova fisica.

Lo stato finale $b\bar{b}\tau^+\tau^-$ rappresenta uno dei canali più interessanti in questo studio visti il branching ratio elevato e la piccola contaminazione di eventi di fondo. A discapito di ciò la ricostruzione dei leptoni tau è sperimentalmente complessa dato che i prodotti di decadimento possono includere neutrini non rivelabili. Inoltre, è altrettanto complessa la separazione degli eventi di segnali da quelli di fondo.

In questa tesi è riportata l'analisi dei dati raccolti dall'esperimento CMS durante il Run 2 di LHC (2016 + 2017 + 2018), corrispondenti ad una luminosità integrata di 137 fb^{-1}. L'analisi include lo studio di due meccanismi di produzione di coppie di bosoni di Higgs: Gluon-Gluon Fusion (GGF) e Vector Boson Fusion (VBF).

Il processo di revisione interna dell'analisi è iniziato nel mese di Dicembre 2020: in questa fase l'analisi statistica viene condotta senza includere i dati sperimentali (*blind analysis*), per questa ragione sono riportati solamente i risultati attesi e non quelli osservati. Il limite superiore calcolato per la sezione d'urto totale (GGF + VBF) risulta 10.7 fb (pari a 4.5 volte il valore SM), con un livello di confidenza del 95%. Le nuove tecniche impiegate nell'analisi hanno permesso di migliorare notevolmente il limite pubblicato in precedenza, pari a 30 volte il valore SM per il processo GGF, basato sull'analisi dei dati del 2016. Lo studio del processo VBF, condotto per la prima volta in questa analisi, ha permesso di porre un limite di esclusione alla sezione d'urto di produzione pari a 238.2 fb, 138 volte il valore SM. La differenza tra i due risultati è dovuta alla differenza tra le sezioni d'urto dei due processi di produzione, pari a circa un fattore 20.

Dato che la statistica di ogni canale di decadimento della coppia HH è limitata si procederà a combinare i risultati di tutti i canali in modo da ottenere limiti di esclusione più stringenti per tutti i parametri che caratterizzano la produzione di coppie di bosoni di Higgs.

Abstract

My research activity during the PhD has focused mainly on the development of new silicon pixel sensors for the upgrade of the CMS experiment inner tracker in view of the CERN LHC High Luminosity phase (HL-LHC), an activity I started to work on during my master thesis. After a couple of years working on the characterization of the new prototypes I became interested also on the impact of the pixel detector on the physics program of the experiment. At the beginning of my second year of PhD I started to look for a physics analysis where the information of the pixel detector is largely exploited. I chose to work on the search for double Higgs boson production in final states with 2 b-jets and 2 tau leptons since the reconstruction of the b-jets and tau leptons makes large use of vertexes and tracks information. Another reason why I chose this analysis is that it will benefit from the incremented statistics foreseen at HL-LHC, possibly leading to the first experimental evidence of HH production.

Instrumentation work

A joint italian ATLAS-CMS INFN group has been collaborating with the Fondazione Bruno Kessler (FBK) foundry, in Trento, since 2015, to the development of a new silicon pixel sensors for the CMS experiment to be used during the high luminosity phase of the LHC collider at CERN (HL-LHC).

The new pixel detector of the CMS experiment is designed to be located at $\approx 3\ cm$ from the interaction point. At such a close distance, the irradiation fluence, after 2500 fb^{-1} of collected data, i.e. after seven years of operation, is expected to reach $2.0 \times 10^{16}\ n_{eq}/cm^2$. The current planar design of the sensors, no matter which way they are operated, is ultimately limited by the degradation of the signal-to-noise ratio and can be reliably employed, in the best case, only up to few $10^{15}\ n_{eq}/cm^2$. Therefore, a new high-radiation tolerant sensor design, capable of surviving up to ten times such a fluence, needs to be studied.

Our collaboration developed thin planar and 3D pixel sensors prototypes on n-in-p wafers, employing a recent technology called Direct Wafer Bonding (DWB). Using this technology, every fabrication process takes place on one side only of the wafer, with consequent cost savings. The active thickness of the first prototypes is 100 μm or 130 μm for the planar sensors and 130 μm for the 3D ones, while the pixel cells dimensions are the same of the sensors currently installed in CMS, $100 \times 150\ \mu m^2$.

Several test beam campaigns have been carried out in order to characterize planar and 3D sensors coupled to the readout chip (ROC) currently used in CMS (PSI46dig), before and after irradiation. The radiation tolerance of the PSI46dig ROC limited our studies to the irradiation fluence of $5 \times 10^{15}\ n_{eq}/cm^2$, which is much lower than the expected fluence at HL-LHC. Despite this limitation the results we obtained have been important to address the next step of the development program and have been published at the beginning of 2020 on a peer reviewed journal. My contribution to the chracterization of these prototypes are reported in this thesis, together with the software development I worked on.

Since the second half of 2018 the first prototype of the phase 2 ROC, named RD53A, has become available. This ROC can sustain higher radiation fluences with respect to the PSI46dig, can be operated at a threshold lower than 1200 prior irradiation. The

dimensions of the cell have been reduced to $50 \times 50\ \mu m^2$ in order to cope with the higher track multiplicity expected at HL-LHC. The sensors' pixel cell area has also been reduced by a factor 6, leading to $50 \times 50\ \mu m^2$ e $25 \times 100\ \mu m^2$ sensors. The characterization of sensor prototypes bonded to this ROC is reported in this thesis.

The goal of the future test beam campaigns is to quantify the cross talk in the $25 \times 100\ \mu m^2$ pixel cell sensors that is caused by the bonding scheme of the sensor to the ROC imposed by the different cell area. It is important to mention that cross talk effect is present in planar sensors but is absent in 3D sensors. This would advise in favor of the use of 3D sensors in inner layers of the detector, together with the their higher radiation tolerance with respect to the planar sensors.

Physics analysis work

Double Higgs searches play a fundamental role in the characterization of the Higgs boson as they represent the favorite channel to measure the Higgs boson trilinear self coupling λ_{HHH}. Only three parameters shape the Higgs field potential in the Standard Model: the Higgs boson mass (m_H), the vacuum expectation value and the Higgs trilinear coupling (λ_{HHH}). The last one has not been measured yet. Any deviation from the theoretical predictions of the Standard Model would lead to sizable changes in both the kinematics and production rate of HH events, thus making double Higgs searches sensitive to new physics effects.

The $b\bar{b}\tau^+\tau^-$ final state represents one of the most interesting channels to explore double Higgs production, because of the high branching ratio and the relatively small background contamination. At the same time, however, this final state poses some non trivial experimental challenges such as the reconstruction of the tau lepton decays that involves undetectable neutrinos, and the discrimination of signal events from background contributions.

The analysis of the data collected by the CMS experiment during the LHC Run 2 (2016 + 2017 + 2018), corresponding to an integrated luminosity of 137 fb^{-1} is reported in this thesis. The analysis include the study of Gluon Fusion (GGF) and Vector Boson Fusion (VBF) production channels.

The internal review of the analysis started in December 2020: at this stage the statistical analysis is carried out without the inclusion of the observed data (*blind analysis*), for this reason only expected results are reported in this thesis. The 95% Confidence Level upper limit on the total (GGF + VBF) cross section is 10.7 fb (equivalent to 4.5 times the SM value). The new techniques employed in the analysis thus allowed for a major improvement of the previously published result (30 times the SM value for GGF production) based on the analysis of the 2016 dataset only. The studied of the VBF process, carried out for the first time in this analysis, resulted in an exclusion limit on the production cross section of 238.2 fb (138 times the SM values). The difference between these two results is determined by the difference in the cross sections that amounts to a factor 20.

Given the limited statistics of each final state of the HH pair decay the results of all the channels will be combined in order to produce the most stringent exclusion limits for the parameters describing the production of Higgs bosons pairs.

Preamble

The mathematical theory that supports our current description of the Universe is the standard model of particle physics (SM), a quantum field theory formalized between the 1960s and 1970s. It relies on the concept of local gauge invariance under symmetry groups of its Lagrangian L, the function from which the equations of motion of a physical system can be derived. This theory allows the unification of electromagnetic and weak forces, predicts the existence of photon, $W^{\pm}$ and Z bosons, and determines their mutual interactions and those with the matter. The same principle is used to explain the existence of the gluon as the mediator of the strong force and its interactions with the quarks.

The SM has encountered a huge success because of its high predictive power and the large number of experimental confirmations to which it has been subjected. After the discovery of the $W^{\pm}$ and Z bosons at the CERN super proton synchrotron (SPS) and the observation of the top quark at the FNAL Tevatron, extensive verifications of its prediction have been performed, most notably at the CERN large electron-positron collider (LEP). The SM has been capable of correctly describing all the measurements performed until now at the GeV and TeV energy scales.

However, this picture alone cannot account for the experimental observation of massive fermions and W and Z bosons, and predicts unphysical properties for vector boson scattering processes. The cornerstone for the completion of the SM is found in the introduction of the Brout–Englert–Higgs (BEH) mechanism, as called from the name of the physicists that proposed it in 1964. The BEH mechanism postulates the existence of a doublet of complex scalar fields that causes a spontaneous electroweak symmetry breaking (EWSB), thus generating the masses of the vector bosons, providing naturally mass terms for the fermion masses via a Yukawa interaction, and ensuring the unitarity of the theory. The direct consequence of the BEH mechanism is the presence of a scalar boson, called the Higgs boson (H), with a mass m_H that is a free parameter of the model. Following the incorporation of the BEH mechanism in the electroweak model by Weinberg and Salam in 1967 and the proof by 't Hooft and Veltman in 1972 that the theory was renormalizable, the search for the Higgs boson became the main goal of the researches performed at high energy colliders. However, even if the value of m_H could be constrained with theoretical arguments and direct searches at the LEP and Tevatron colliders, its existence could not be directly proved for almost half a century.

The Large Hadron Collider (LHC) was built to give an answer to this question by elucidating the properties of the EWSB mechanism and exploring the physics at the TeV energy scale. It is designed to collide protons at a centre-of-mass energy up to 14 TeV, in four interactions points instrumented with an equal number of detectors. The physics programme of the two general-purpose detectors, A Toroidal LHC Apparatus (ATLAS) and the Compact Muon Solenoid (CMS), is complemented by the measurements performed with the LHC beauty (LHCb) and A Large Ion Collider Experiment (ALICE) detectors.

I carried out my PhD research collaborating in the CMS experiment activities. CMS is instrumented with multiple, concentric layers of detectors to identify and measure the particles produced in the collisions. The interaction point is surrounded by the tracker detector to precisely measure the positions of the primary vertices of the interactions and the trajectory and momentum of the charged particles. The electromagnetic and hadronic calorimeters are located around the tracking system and are designed to absorb electrons, photons, and hadrons within their volume to measure their energy. Muons can traverse the calorimeters and are measured in muon tracking systems located in the outermost part of the detector.

After a 20-years long phase of construction and commissioning the LHC began its high energy operation in March 2010. In July 2012, the ATLAS and CMS collaborations announced the discovery of a new boson with a mass close to 125GeV whose properties where compatible with the ones of the Higgs boson. The Higgs boson is something unique in the SM. It is the only known elementary boson with a spin of 0, a scalar, and whose existence does not arise from a local gauge invariance mechanism. It breaks the degeneracy between the three families of fermions by conferring them their mass with couplings of different strengths, and its observation consequently proves the purely quantum-relativistic nature of the mass. The discovery of the Higgs boson marked a milestone in the history of physics, not only because it completes the SM with its last missing piece, but because it opens up the study of a new sector of the theory, the scalar sector. The exploration of the scalar sector of the SM requires a more precise determination of its couplings and properties, including the measurement of rare production and decay modes. In this context, the Higgs boson self-interactions are of particular interest because they provide information to reconstruct the shape of the scalar potential itself.

The observation of the Higgs boson completes the theoretical description of the SM but raises further questions deeply related to its scalar sector. First, the SM does not provide a mechanism that determines its arbitrary parameters. In particular, it does not explain why three families of fermions exist, what is at the origin of their couplings to the Higgs boson, and why their values range over several order of magnitude. Being responsible for the breaking of the degeneracy between the fermion families, the scalar sector of the SM has a direct role in this context. Second, the mass of the Higgs boson is not protected by any fundamental symmetry of the theory, making it largely sensitive to divergent radiative corrections. These corrections need to be finely tuned to account for the observed Higgs boson mass of about 125 GeV, raising an issue about the naturalness of the theory. Finally, this specific mass value implies an instability of the vacuum, as the evolution of the Higgs self-coupling (and consequently of the shape of the scalar potential) with the energy implies that at higher scales this can become negative. This results in a metastable conditions of our Universe.

These theoretical considerations should be regarded together with the limitations of the SM in explaining the experimental observations at cosmological scales. The SM does not provide a mechanism that is responsible for the observed matter-antimatter asymmetry in the Universe, nor it accounts for the existence of a stable particle that is compatible with the dark matter, of which we observe the gravitational effects.

Although being extremely successful at describing the phenomenology of the experiments at colliders, the SM seems incomplete.

When considered together, these open questions seem to indicate that the SM is only a part of a more extended physics model. The search for physics beyond the SM (BSM) is one of the main reasons that pushed towards an upgrade of the existing LHC. The High

Luminosity LHC (HL-LHC) will feature an increase of the instantaneous luminosity by a factor of five with respect to the original LHC design specifications and the integrated luminosity by a factor of ten. These improvements will allow for a more precise test of the Higgs boson properties and the exploration of its rare decay modes, where deviations from the SM might be found. At the same time, they open the way to direct searches for experimental signatures of BSM physics at the TeV scale.

When I began my master thesis in 2016 the Milano Bicocca CMS group just got involved in the upgrade of the pixel detector in view of the HL-LHC upgrade. I decided to contribute to this project by working on the characterization of the new prototype sensors, before and after irradiation, by means of beam teat campaigns.

Once I became familiar with the operation of the detector, I developed an interest also for its impact on the physics program of the experiment. At the beginning of my second year of PhD I started to look for a physics analysis where the information of the pixel detector is largely exploited. I chose to work on the search for double Higgs boson production in final states with 2 b-jets and 2 tau leptons since the reconstruction of the b-jets and tau leptons makes large use of vertexes and tracks information. Another reason why I chose this analysis is that it will benefit from the incremented statistics foreseen at HL-LHC, possibly leading to the first experimental evidence of HH production.

Double Higgs searches play a fundamental role in the characterization of the Higgs boson as they represent the favourite channel to measure the Higgs boson trilinear self coupling λ_{HHH}. Only three parameters shape the Higgs field potential in the Standard Model: the Higgs boson mass (m_H), the vacuum expectation value and the Higgs trilinear coupling (λ_{HHH}). The last one has not been measured yet. Any deviation from the theoretical predictions of the Standard Model would lead to sizable changes in both the kinematics and production rate of HH events, thus making double Higgs searches sensitive to new physics effects.

The $b\bar{b}\tau^+\tau^-$ final state represents one of the most interesting channels to explore double Higgs production, because of the high branching ratio and the relatively small background contamination. At the same time, however, this final state poses some non trivial experimental challenges such as the reconstruction of the tau lepton decays that involves undetectable neutrinos, and the discrimination of signal events from background contributions.

This thesis is organized in 5 chapters. Chapter 1 describes the LHC accelator complex and the CMS experiment, with special emphasis on the tracking detector. Chapter 2 contains an introduction to the semiconductor detectors in general and on silicon pixel detectors in particular. In Chapter 3 the test beam facilities where the new prototypes have been characterized are described. The characterization of the phase-2 pixel detector prototypes is reported in Chapter 4, together with the software development contribution I gave to the project. The search for Higgs bosons pair production in the $b\bar{b}\tau^+\tau^-$ with the LHC Run 2 data sample (2016 + 2017 + 2018) is described in Chapter 5.

Chapter 1

The CMS experiment at LHC

The European Council for Nuclear Research (CERN) laboratories host the Large Hadron Collider (LHC), designed to accelerate protons inside its 26.7 km long tunnel to a centre-of-mass energy of 14 TeV. The LHC is the largest and most powerful particle accelerator ever built: it collides the particle beams in four interaction points, instrumented with an equal number of detectors. In one of these four points is installed the Compact Muon Solenoid (CMS) experiment, designed to explore the physics at the TeV scale.

1.1 The Large Hadron Collider

The LHC is designed to collide protons at a center-of-mass energy of $\sqrt{s} = 14\,TeV$ with an instantaneous luminosity $L = 10^{34}\,cm^{-2}s^{-1}$, as well as lead ions at a center-of-mass energy of 2.76 GeV per nucleon and $L = 10^{27}\,cm^{-2}s^{-1}$ [1, 2]

1.1.1 Design and Specifications

The LHC is installed in a 26.7 km long tunnel built between 1984 and 1989 to host the CERN Large Electron Positron (LEP) collider. The tunnel is located in the region nearby Geneva and extends across the French and the Swiss borders. In the LHC two separate, counter-rotating particle beam lines are kept in orbit in two magnetic channels thanks to the field generated by superconducting niobium-titanium coils. The particles are steered by a magnetic field of 8.3 T generated by a current of about 11 kA in the 1232 dipole magnets, each measuring 15 m of length and 35 tonnes of weight. The stability of the beam dynamics is ensured by 392 quadrupoles magnets measuring from 5 to 7 m of length, that focus the particles and keep them in a narrow beam. Special quadrupoles are installed in proximity of the collision points to squeeze the beams to increase the collision probability. Superconducting magnets are cooled with superfluid helium-4 and kept to a working temperature of 1.9 K.

The LHC is the last element of an injection chain composed of several smaller particle accelerators, schematically represented in Fig. 1.1. Hydrogen atoms are stripped of their electrons in a duo-plasmatron source and are accelerated to an energy of 50 MeV in the Linear Accelerator (LINAC2), which feeds the Proton Synchrotron Booster (PSB) where protons are accelerated to 1.4 GeV. The beam is then injected into the Proton Synchrotron (PS) for a further acceleration to 25 GeV, and subsequently into the Super Proton Synchrotron (SPS) where protons reach an energy of 450 GeV. The proton beams are finally transferred to the two LHC beam pipes, where the beams are accelerated and

shaped into proton bunches thanks to radio-frequency cavities operated at 400 MHz. Once the protons reach the nominal energy and the beam dynamics is stabilized they are brought to collide in four points along the LHC ring.

An important parameter of particle accelerators is the instantaneous luminosity, $\mathcal{L}$, that depends on the beam properties as [1]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{1.1}$$

N_b is the number of particles in each of the n_b bunches per beam, that revolve in the tunnel with a frequency f_{rev}. The symbol γ_r denotes the relativistic factor. The shape and focus of the beam are described by its transverse emittance ϵ_n and its beta function β^* or focal length, at the collision point. The factor F accounts for the geometric reduction of the instantaneous luminosity, and depends on the beam crossing angle θ_c and on the transverse and longitudinal r.m.s. bunch sizes σ_{xy} and σ_z at the interaction point (i.p.) as:

$$F = \left(1 + \frac{\theta_c \sigma_z}{2\sigma_{xy}}\right)^{-\frac{1}{2}} \tag{1.2}$$

The integrated luminosity $L = \int \mathcal{L}$ is proportional to the total amount of collisions produced. LHC downtimes must be taken into account upon performing the integration, and correspond to the time needed for maintenance, filling, and ramping of the magnetic field. The integrated luminosity is the coefficient of proportionality between the number of events N produced for a specific process and its cross section σ:

$$N = L \times \sigma \tag{1.3}$$

The four collision points of the LHC are instrumented with particle detectors installed in underground caverns. "A Toroidal LHC ApparatuS" (ATLAS) and the "Compact Muon Solenoid" (CMS) experiments are installed in the diametrically opposite Points 1 and 5 of the LHC, where the highest instantaneous luminosity of collision is produced. They are designed as hermetic, multi-purpose detectors that surround the interaction point and measure the products of proton and ion collisions. The "LHC beauty" (LHCb), located at Point 8 is a forward, one-arm spectrometer devoted to the study of CP-violation in beauty and charm sectors. "A Large Ion Collider Experiment" (ALICE) is installed in Point 2 and is built to study heavy ion collisions and quark-gluon plasmas.

The LHC also hosts three smaller size experiments. The "LHC forward" (LHCf) and the "TOTal Elastic and diffractive cross section Measurement" (TOTEM) experiments, located a hundred of meters away on either side from the ATLAS and CMS interaction points, are dedicated to pp interaction cross section measurements and forward diffractive physics. LHCb shares its cavern with the "Monopole and Exotics Detector at the LHC" (MOEDAL) experiment, dedicated to the search for magnetic monopoles.

1.1.2 Operations

The first proton beam circulated in the LHC on September 10th, 2008. An incident occurred a week later due to a faulty electrical connection between two magnets, that caused the release of helium into the tunnel and mechanical damages to the accelerator. Repair works were promptly achieved and the LHC was back to operations in November

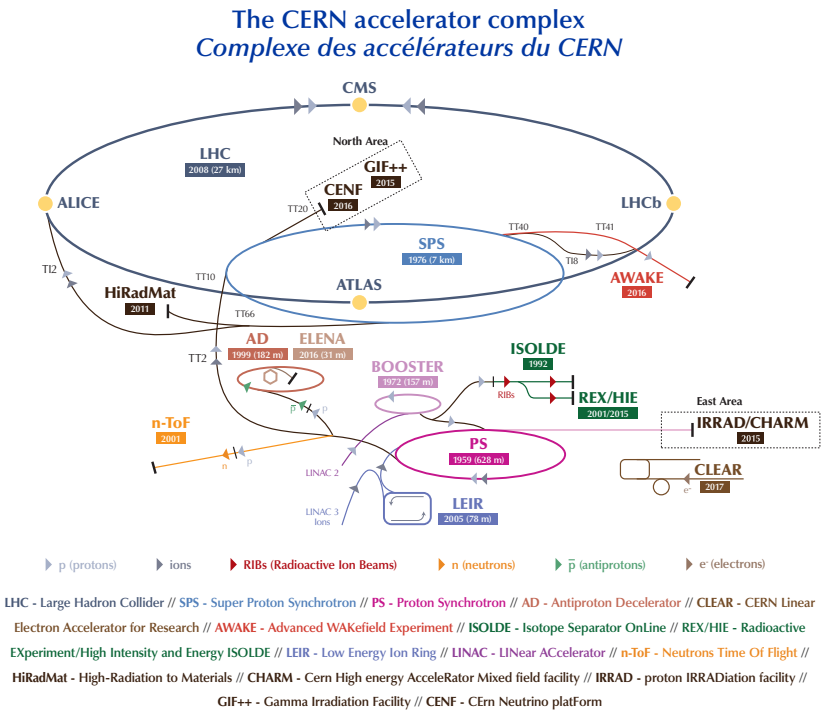


Figure 1.1: Representation of the CERN accelerator complex [3].

2009. After machine commissioning and collisions at lower energy, the first, high energy collisions took place on March 30th, 2010. This moment marked the beginning of the so-called Run I, the data taking era that lasted until 2012. It was decided not to operate the LHC at its design parameters and pp collisions took place at a center-of-mass energy of 7 TeV, soon increased to 8 TeV, with a bunch spacing of 50 ns. About 45 pb^{-1} and 6 fb^{-1} were collected by the CMS experiment at $\sqrt{s} = 7\,TeV$ in 2010 and 2011, respectively, and a larger datasets of 23 fb^{-1} was recorded at $\sqrt{s} = 8 TeV$ in 2012. These data allowed for the discovery of the Higgs boson and for a first measurement of its properties.

The LHC operations halted in 2012 for a two year long shutdown (LS1). Important renovation and consolidation works were performed to push the LHC performance towards the design parameters. In particular, the magnets were trained to withstand higher currents for an increase of the energy per beam to 6.5 TeV. LS1 represented as well the opportunity for the experiments to complete a series of important detector upgrades to cope with the harsher collision conditions. LHC operations restarted in 2015 at a center-of-mass energy of 13 TeV. After a short collision phase with 50 ns bunch spacing, the nominal spacing of 25 ns was reached. Operations in 2015 were focused on the commissioning of the LHC at the new energy, and the instantaneous luminosity was not increased beyond $5 \times 10^{33} cm^{-2} s^{-1}$. In 2016, the LHC was ready to deliver a large dataset to the experiments. The instantaneous luminosity rose up to $1.5 \times 10^{34}\, cm^{-2} s^{-1}$, beyond the original machine design specifications allowing an unprecedented 35.9 fb^{-1} dataset to be recorded over the full year. These numbers refer to the integrated luminosity collected with the CMS detector and validated for a use in physics analyses.

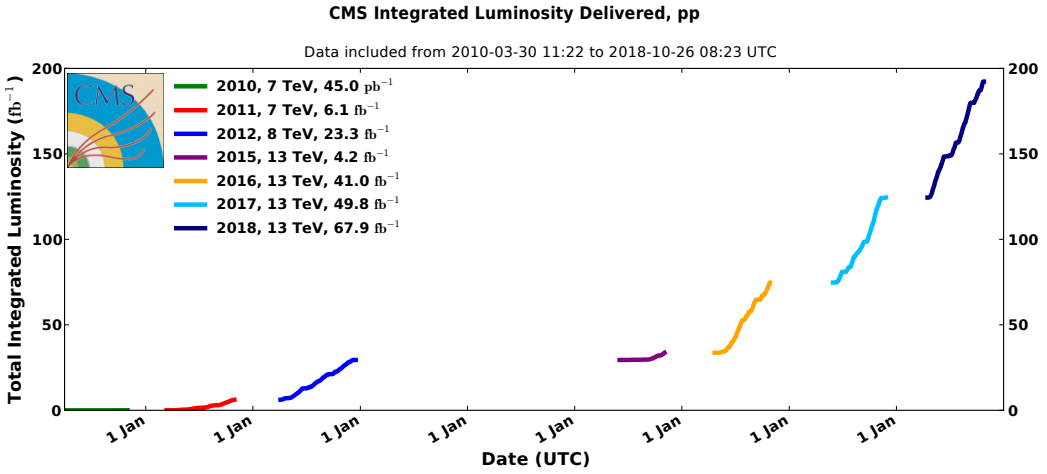
At the end of 2016 an Extended Year End Technical Stop (EYETS) was scheduled to allow the maintenance work of the cryogenic systems, the replacement of a magnet and the maintenance of the PSB and SPS in view of the future works for the HL-LHC.

During this technical stop a new pixel detector was installed in the CMS experiment, as described in Section 1.2.3.

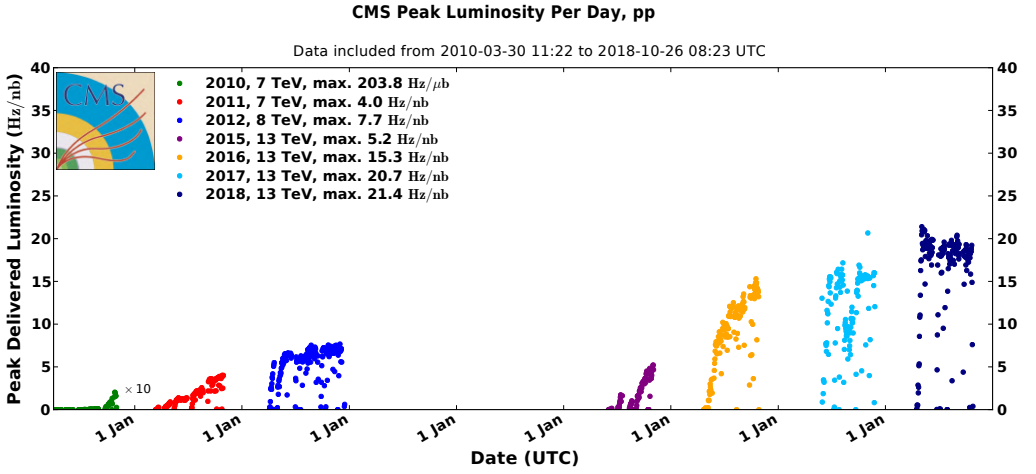
LHC operations continued smoothly during 2017 and 2018 and an integrated luminosity of 45 fb^{-1} and 63 fb^{-1} was collected by the CMS experiment.

At the end of October 2018 the last pp collisions of the Run II took place, the CMS experiment collected an integrated luminosity of approximately 140 fb^{-1}.

The total integrated luminosity and peak instantaneous luminosity of the LHC as a function of the year, as measured by the CMS experiment, are shown in Fig. 1.2.



(a) Integrated luminosity



(b) Peak instantaneous luminosity

Figure 1.2: Total integrated luminosity (a) and peak instantaneous luminosity (b) of the LHC as a function of the year, as measured by the CMS experiment

LHC operations are now halted for a second long shutdown (LS2) devoted to upgrades of the machine injectors in view of the future high luminosity phase. The LHC will restart in 2021 for its Run III, three years of operation at a center-of-mass of 14 TeV and a peak luminosity twice the original machine design. Within the three years of the Run III, experiments are foreseen to record an integrated luminosity of about 300 fb^{-1}.

The third long shutdown (LS3) starting in 2024 will conclude the Phase I of LHC operations that started back in 2008. During 30 months the LHC will undergo profound changes [4].

New niobium-tin superconducting quadrupole magnets, capable of generating a field up to 12 T, will be installed at the ATLAS and CMS interaction points to focus the beams. Compact superconducting cavities (called "crab cavities") will be used to precisely rotate the proton bunches before the collision, reduce the crossing angle and enhancing the factor F in equation 1.2.

These upgrades will increase the instantaneous luminosity by a factor of five with respect to the original design specifications and the integrated luminosity by a factor of ten, producing a huge amount of data that will be used to investigate very rare phenomena at the LHC. The unprecedented conditions of the collisions and, in particular, an average number of simultaneous interactions in one bunch crossing of 140, call for an important upgrade program of the experiments.

A summary of past operations and the baseline future schedule for LHC and High Luminosity LHC (HL-LHC) is shown in Fig. 1.3.

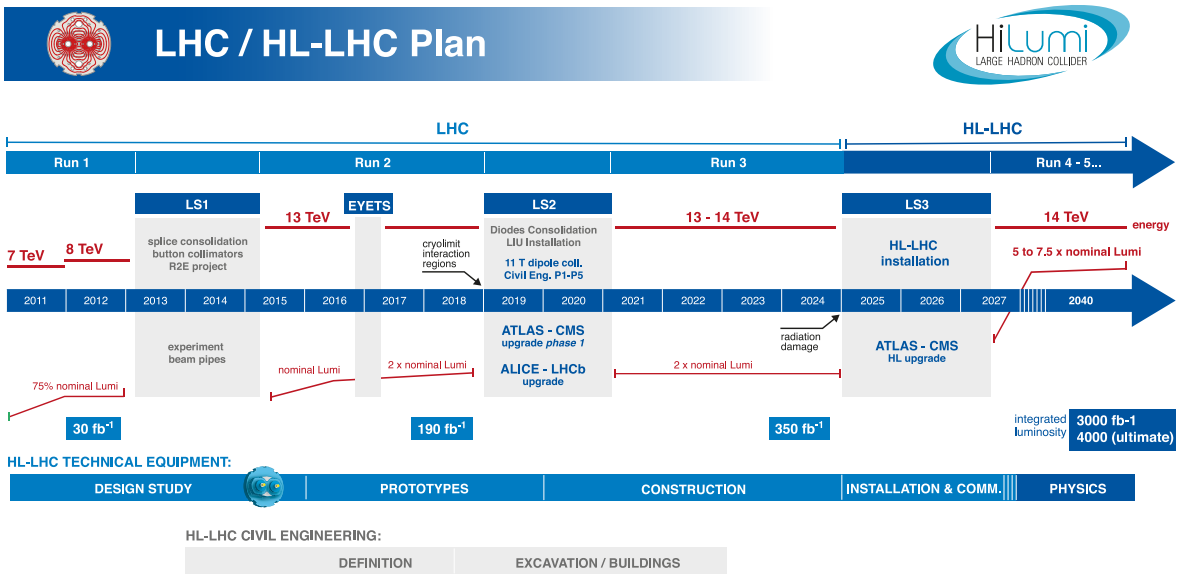


Figure 1.3: Baseline schedule of LHC and HL-LHC operations.

1.2 The CMS Experiment

The CMS detector has been built to explore the physics at the TeV scale in many different signatures and final states. It has been consequently designed as a multi-purpose detector, that hermetically surrounds the interaction point in the underground cavern of Point 5 in Cessy (FR), and is instrumented with several subsystems developed for the identification and measurement of different types of particles. The detector has a cylindrical structure with a diameter of 15 m and a length of 21.5 m, and an overall weight of about 12 500 t.

Collisions take place in the center of the CMS experiment every 25 ns, implying that new particles leave the interaction point before those produced in the previous bunch crossing have even left the external surface of the detector. In addition, multiple proton interactions can take place within each bunch crossing. These two effects are globally denoted as out-of-time and in-time pileup and overlap to the signal of interest represented

by the hard-scatter interaction. These challenging conditions call for a detector design that is highly granular, fast in its response, and resistant to the radiation. At the same time, it must be capable of precisely measuring the energy and the momentum of the final state particles and to identify them.

1.2.1 Coordinate system

A right-handed Cartesian coordinate system is used to describe the detector and the collision products. It is defined with its center in the nominal interaction point, the x axis pointing to the center of the LHC ring, the y axis pointing upwards, and the z axis pointing in the anticlockwise proton beam direction.

Given the cylindrical structure of the detector, a polar system is also used. The azimuthal angle ϕ is defined in the (x,y) or transverse plane as the angle formed with respect to the positive x axis, the radial coordinate in this plane is denoted as r. The polar angle θ is defined in the (r,z) plane as the angle formed with the z axis and usually converted into the pseudorapidity $\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right)$. The spatial separation of two particles can be expressed in terms of their angular distance as $\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$. The projection of the momentum of a particle onto the transverse plane is referred to as the transverse momentum or p_T, and has the advantage to be independent on the Lorentz boost resulting from the initial momentum of the interacting partons along the z axis.

1.2.2 Detector structure

The CMS detector [5] consists of a central section, or "barrel", and two forward regions, or "endcaps", as it can be observed in the schematic representation of Fig. 1.4. The specific boundaries in η between the two regions depend on the subsystem considered.

CMS is instrumented with multiple, concentric layers of detectors to identify and measure the particles produced in the collisions. The interaction point is surrounded by the tracker detector to precisely measure the positions of the primary vertices of the interactions and the trajectory and momentum of the charged particles. The electromagnetic and hadronic calorimeters are located around the tracking system and are designed to absorb electrons, photons, and hadrons within their volume to measure the energy deposited. Muons can traverse the calorimeters and are measured in muon tracking systems located in the outermost part of the detector. The core of the experiment is a niobium-titanium superconducting solenoid of 6 m of diameter. It is operated at a temperature of 4.5 K and generates a 3.8 T magnetic field along the z axis. This strong magnetic field is used to bend the charged particles and measure their transverse momentum with the tracking subdetectors. The tracker and the calorimeters are located inside the solenoid: this poses tight constraints on their size and, in the case of the calorimeters, requires high density materials to contain the incoming particles and their secondary interaction products. The return field of the magnet has an intensity of about 2 T and is used to measure the transverse momentum with the muon detectors located inside the iron structure that surrounds the solenoid. This causes the muons trajectories to be bent in opposite directions in the inner tracker and muon systems. The information from the individual subdetectors are often redundant and can be combined to improve the reconstruction of final state objects.

A description of the subdetectors will be given in the next paragraphs, emphasis will be given to the pixel detector which is the object of this thesis.

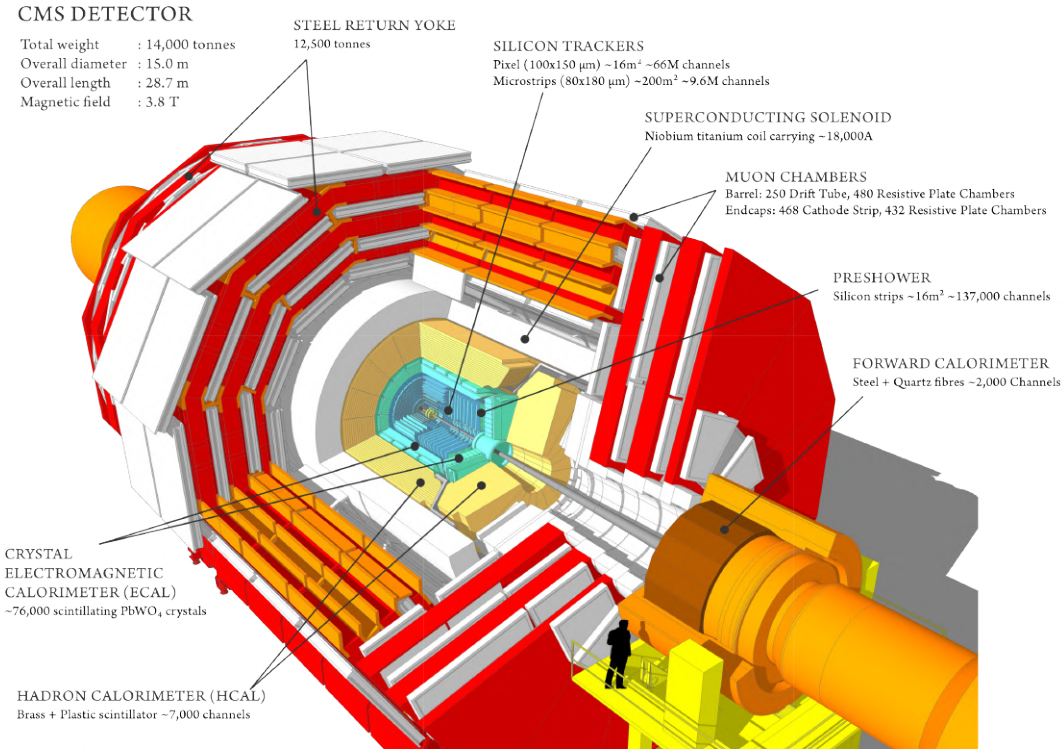


Figure 1.4: Perspective view of the CMS detector layout.

Inner Tracking Systems

The inner tracking system of CMS is designed to provide a precise and efficient measurement of the trajectories of charged particles produced in the LHC collisions, as well as a precise reconstruction of primary and secondary vertices. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m.

The CMS solenoid provides a homogeneous magnetic field of 3.8 T over the full volume of the tracker. At the LHC design luminosity of $1 \times 10^{34}\, cm^{-2}s^{-1}$ there are about 1000 particles emerging from more than 20 overlapping proton-proton interactions and traversing the tracker for each bunch crossing, i.e. every 25 ns. Therefore a detector technology featuring high granularity and fast response is required, such that the trajectories can be identified reliably and attributed to the correct bunch crossing. However, these features imply a high power density of the on-detector electronics which in turn requires efficient cooling. This is in direct conflict with the aim of keeping to the minimum the amount of material in order to limit multiple scattering, bremsstrahlung, photon conversion and nuclear interactions. The intense particle flux will also cause severe radiation damage to sensors and readout electronics. The main challenge in the design of the detector was to develop components able to operate in this harsh environment for an expected lifetime of 10 years. These requirements on granularity, speed and radiation hardness lead to a design entirely based on silicon detector technology.

The LHC physics program requires a robust, efficient and precise reconstruction of the trajectories of charged particles with transverse momentum above 1 GeV in the

pseudorapidity region $-2.5 < \eta < 2.5$. A precise measurement of secondary vertices and impact parameters is necessary for the efficient identification of heavy flavour hadrons which are produced in many of the interesting physics channels. Together with the electromagnetic calorimeter and the muon system the tracker has to identify electrons and muons, respectively. Tau leptons are a signature in several discovery channels and need to be reconstructed in one-prong and three-prong decay topologies. In order to reduce the event rate from the LHC bunch crossing rate of 40 MHz to about 1 kHz which can be permanently stored, tracking information is heavily used in the high level trigger of CMS.

The operating conditions for a tracking system at the LHC are very challenging. As already mentioned, each LHC bunch crossing at design luminosity creates on average about 1000 particles hitting the tracker. This leads to a hit rate density of 1 MHz/mm^2 at a radius of 4 cm, falling to 60 kHz/mm^2 at a radius of 22 cm and 3 kHz/mm^2 at a radius of 115 cm. In order to keep the occupancy at or below 1% pixelated detectors have to be used at radii below 10 cm. For a pixel size of $100 \times 150\,\mu m^2$ in r-phi and z, respectively, which is driven by the desired impact parameter resolution, the occupancy is of the order 10^{-4} per pixel and LHC bunch crossing. At intermediate radii ($20\,cm < r < 55\,cm$) the reduced particle flux allows the use of silicon micro-strip detectors with a typical cell size of $10\,cm \times 80\,\mu m$, leading to an occupancy of up to 2–3% per strip and LHC bunch crossing. In the outer region ($55cm < r < 110cm$) the strip pitch can be further increased. Given the large areas that have to be instrumented in this region, also the strip length has to be increased in order to limit the number of readout channels. However, the strip capacitance scales with its length and therefore the electronics noise is a linear function of the strip length as well. In order to maintain a good signal to noise ratio of well above 10, CMS uses thicker silicon sensors for the outer tracker region (500 μm thickness as opposed to the 320 μm in the inner tracker) with correspondingly higher signal. These thicker sensors would in principle have a higher depletion voltage. Nevertheless the radiation levels in the outer tracker are smaller, so a higher initial resistivity can be chosen such that the initial depletion voltages of thick and thin sensors are in the same range of 100 V to 300 V. In this way cell sizes up to about $25\,cm \times 180\,\mu m$ can be used in the outer region of the tracker, with an occupancy of about 1%. These occupancy-driven design choices for the strip tracker also satisfy the requirements on position resolution.

CMS has been the first experiment using silicon detectors in this outer tracker region. This novel approach was made possible by three key developments:

- sensor fabrication on 6 inches instead of 4 inches wafers reduced the sensor cost to 5–10 CHF/cm^2 and allowed the coverage of the large required surfaces with silicon sensors

- implementation of the front-end readout chip in industry-standard deep sub-micron technology led to large cost savings and to an improved signal-to-noise performance

- automation of module assembly and use of high throughput wire bonding machines

The radiation damage introduced by the high particle flux near the interaction region is a severe design constraint. Three different effects had to be considered in the design of a radiation tolerant silicon tracker: surface damage, bulk damage and single event upsets.

Surface damage is created when the positively charged holes, generated by the passage of a ionizing particle, get trapped in a silicon oxide layer. This is mostly a concern for

the front-end electronics where this additional space charge changes, for instance, the characteristics of MOS structures. Surface damage simply scales with the absorbed dose.

The silicon sensors are mainly affected by bulk damage, i.e. modifications to the silicon crystal lattice which are caused by non-ionizing energy losses (NIEL) and lead to additional energy levels in the band gap. NIEL is a complicated process, depending on particle type and energy, but is found to scale approximately with the fast hadron fluence. The consequences are an increase of the leakage current (linear in fluence), a change in the doping from n to p type with a corresponding change in depletion voltage by a few hundred volts over the lifetime of the tracker and the creation of additional trapping centers which will reduce the signal by roughly 10% after 10 years of LHC running. The design of the silicon sensors and the readout electronics had to take this into account to assure a signal-to-noise ratio of 10:1 or better over the full lifetime of the detector, in order to guarantee a robust hit recognition at an acceptable fake hit rate.

Finally, transient phenomena due to the generation of charge by ionizing particles in the electronic circuitry can change for instance the state of memory cells and therefore disturb or even stop the correct functioning of the readout, this is called a single event upset (SEU).

The detector leakage current has also a dependence on temperature that can become exponential if the temperature is increased, leading to a dangerous positive feedback of the self heating of the silicon sensor, called thermal runaway. This had to be avoided by efficiently coupling the silicon sensors to the cooling system and by operating at a low temperature. A second effect, called reverse annealing, requires to keep the silicon sensors permanently well below 0 $^{\circ}C$ except for short maintenance periods. This effect is caused by the interaction of radiation induced defects in the silicon sensors which can lead to more serious damage and to an even stronger change in depletion voltage with fluence. Experimentally it is found that reverse annealing becomes insignificant for temperatures roughly below 0 $^{\circ}C$. The effect of the irradiation on the silicon sensors will be further described in chapter 2.

The readout chips employed in the CMS tracker are fabricated in standard 0.25 μm CMOS technology which is inherently radiation hard. The lifetime of the tracker is therefore limited by the radiation damage to the silicon sensors. For efficient charge collection they always need to be over-depleted, requiring bias voltages up to 500 V after 10 years of LHC operations. This value represents the limit of the typical high voltage stability of current sensor layouts. All tests have shown that the silicon strip tracker will remain fully operational for 10 years of LHC running. For the pixel detector on the other hand, which has to survive even higher radiation doses, under-depleted operation is possible due to a different sensor layout. Its lifetime ranges from at least 2 years at full LHC luminosity for the innermost layer to more than 10 years for the third layer.

The ultimate position resolution of the pixel and strip sensors is degraded by multiple scattering in the material that is necessary to precisely hold the sensors, to supply the electrical power (in total about 60 kW for the CMS tracker) and to cool the electronics and the silicon sensors. Nuclear interactions of pions and other hadrons in this material reduce significantly the tracking efficiency for these particles. In addition, this material leads to photon conversion and bremsstrahlung which adversely affect the measurement accuracy of the electromagnetic calorimeter. It was therefore a requirement to keep the amount of this material to a minimum.

A schematic drawing of the CMS tracker is shown in Fig. 1.5. At radii of 4.4, 7.3 and 10.2 cm, three cylindrical layers of hybrid pixel detector modules surround the interaction

point. They are complemented by two disks of pixel modules on each side. The pixel detector delivers three high precision space points on each charged particle trajectory. In total the pixel detector covers an area of about 1 m^2 and has 66 million channels.

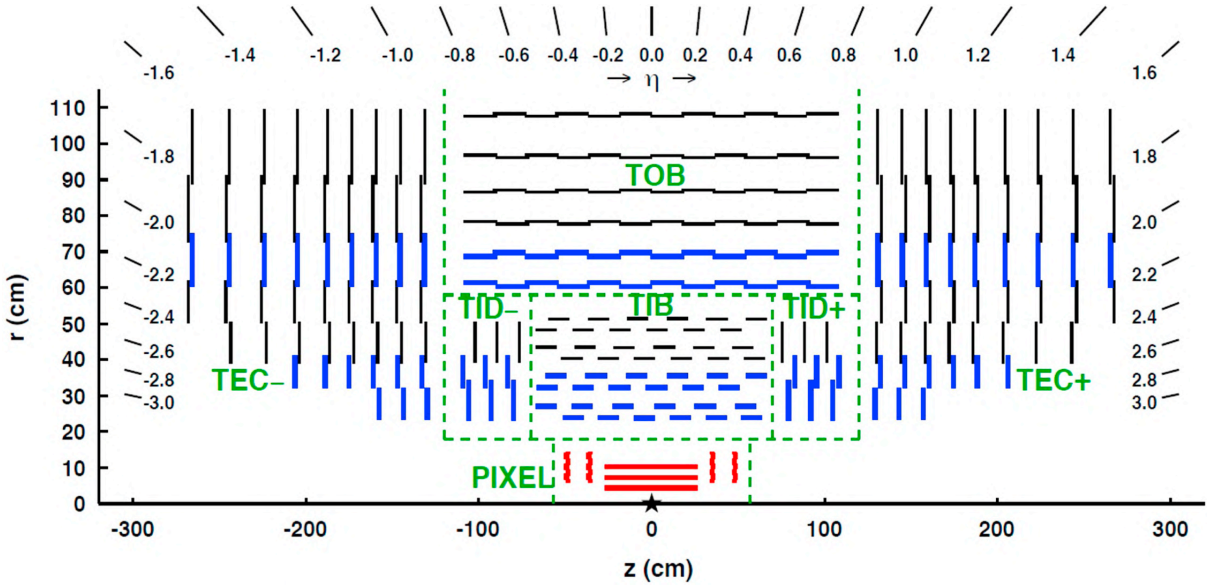


Figure 1.5: Schematic cross section through the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits.

The radial region between 20 cm and 116 cm is occupied by the silicon strip tracker: it is composed of three different subsystems. The Tracker Inner Barrel and Disks (TIB/TID) extend in radius towards 55 cm and are composed of 4 barrel layers, supplemented by 3 disks at each end. TIB/TID delivers up to 4 r-ϕ measurements on a trajectory using 320 μm thick silicon micro-strip sensors with their strips parallel to the beam axis in the barrel and radial on the disks. The strip pitch is 80 μm on layers 1 and 2 and 120 μm on layers 3 and 4 in the TIB, leading to a single point resolution of 23 μm and 35 μm, respectively. In the TID the mean pitch varies between 100 μm and 141 μm. The TIB/TID is surrounded by the Tracker Outer Barrel (TOB). It has an outer radius of 116 cm and consists of 6 barrel layers of 500 μm thick micro-strip sensors with strip pitches of 183 μm on the first 4 layers and 122 μm on layers 5 and 6. It provides 6 more r-ϕ measurements with single point resolution of 53 μm and 35 μm, respectively. The TOB extends in z between $\pm$ 118 cm. Beyond this z range the Tracker EndCaps (TEC+ and TEC- where the sign indicates the location along the z axis) cover the region $124\ cm < \mid z \mid < 282\ cm$ and $22.5\ cm <\mid r \mid < 113.5\ cm$. Each TEC is composed of 9 disks, carrying up to 7 rings of silicon micro-strip detectors (320 μm thick on the inner 4 rings, 500 μm thick on rings 5-7) with radial strips of 97 μm to 184 μm average pitch. Thus, they provide up to 9 phi measurements per trajectory.

In addition, the modules in the first two layers and rings of TIB, TID and TOB as well as rings 1, 2, and 5 of the TECs carry a second micro-strip detector module which is mounted back-to-back with a stereo angle of 100 mrad in order to provide a measurement of the second coordinate (z in the barrel and r on the disks). The achieved single point resolution of this measurement is 230 μm and 530 μm in TIB and TOB, respectively, and varies with pitch in TID and TEC. This tracker layout ensures at least 9 hits in the silicon strip tracker in the full range of $\mid \eta \mid < 2.4$ with at least 4 of them being two-dimensional measurements. The ultimate acceptance of the tracker ends at $\mid \eta \mid < 2.5$. The CMS

silicon strip tracker has a total of 9.3 million channels and 198 m^2 of active silicon area.

The pixel detector is the part of the tracking system that is closest to the interaction region. It contributes precise tracking points in r-ϕ and z and therefore is responsible for a small impact parameter resolution that is important for good secondary vertex reconstruction. With a pixel cell size of $100 \times 150\ \mu m^2$ emphasis has been put on achieving similar track resolution in both r-ϕ and z directions. Through this a 3D vertex reconstruction in space is possible, which is important for secondary vertices with low track multiplicity. The pixel system has a zero-suppressed read out scheme with analog pulse height readout. This improves the position resolution due to charge sharing and helps to separate signal and noise hits as well as to identify large hit clusters from overlapping tracks.

The pixel detector covers the pseudorapidity range $\mid \eta \mid < 2.5$, matching the acceptance of the central tracker. It is essential for the reconstruction of secondary vertices from b and tau decays, and of seed tracks for high level triggering. It consists of three barrel layers (BPix) with two endcap disks (FPix). The 53 cm long BPix layers are located at mean radii of 4.4, 7.3 and 10.2 cm. The FPix disks extending from 6 to 15 cm in radius, are placed on each side at z $\pm$ 34.5 and z $\pm$ 46.5 cm. BPix (FPix) contain 48 million (18 million) pixels covering a total area of 0.78 (0.28) m^2. The arrangement of the 3 barrel layers and the forward pixel disks on each side gives 3 tracking points over almost the full pseudorapidity range.

The proximity to the interaction region also implies a very high track rate and particle fluences that require a radiation tolerant design. For the sensor this led to an n+ pixel on n substrate detector design that allows partial depleted operation even at very high particle fluences. For the barrel layers the drift of the electrons to the collecting pixel implant is perpendicular to the magnetic field of CMS. The resulting Lorentz drift leads to charge spreading of the collected signal charge over more than one pixel. With the analog pulse height being read out a charge interpolation allows to achieve a spatial resolution in the range of 15-20 μm. The forward detectors are tilted at 20 deg in a turbine-like geometry to induce charge sharing. A position resolution of approximately 15 μm in both directions can be achieved with charge-sharing between neighbouring pixels. The reduction in the depletion depth or the increase in bias voltage will lead to a reduction of charge sharing and therefore a degradation of the spatial resolution with radiation damage.

In order to allow a replacement of the innermost layers, the mechanic and the cabling of the pixel system have been designed to allow a yearly access if needed. The pixel system is inserted as the last sub detector of CMS after the silicon strip tracker has been installed and after the central section of the beam pipe has been installed and baked out.

The sensors for the CMS pixel detector adopt the so called n-on-n concept. The pixels consist of n implants introduced into a high resistance n substrate. The rectifying pn junction is placed on the back side of the sensor surrounded by a multi-guard-ring structure. Despite the higher costs due to the double sided processing this concept was chosen as the collection of electrons ensures a high signal charge at moderate bias voltages (< 600 V) after high hadron fluences. Furthermore the double sided processing allows a guard ring scheme keeping all sensor edges at ground potential. The isolation technique applied for the regions between the pixel electrodes was developed in close collaboration with the sensor vendors. Open p-stops were chosen for the disks and moderated p-spray for the barrel. Both types of sensors showed sufficient radiation hardness during an extensive qualification procedure including several test beams.

The disk sensors use the p-stop technique for inter-pixel isolation. To maximize the charge collection efficiency and minimize the pixel capacitance within the design rules of the vendor a width of 8 μm for the p-stop rings and a distance of 12 μm between implants was chosen. Fig. 1.6 shows a photograph of 4 pixel cells. The open ring p-stops, the bump bonding pad and the contact between the aluminium and the implanted collecting electrode are highlighted. The opening on the p-stop rings provides a low resistance path until full depletion is reached to allow IV (current-voltage) characterization of the sensor on wafer and a high resistance path when the sensor is over-depleted (10-20 V over-depletion) to assure interpixel isolation. The process is completely symmetric with five photolithographic steps on each side to minimize the mechanical stress on the silicon substrate and the potential bowing of the diced sensors. The sensors were all fabricated in 2005 on 4 inch wafers. The depletion voltage is 45-50 V and the leakage current is less than 10 nA/cm^2. The 7 different sensor tiles needed to populate a disk blade, ranging from 1×2 readout chips (ROCs) to 2×5 ROCs, are implemented on a single wafer.

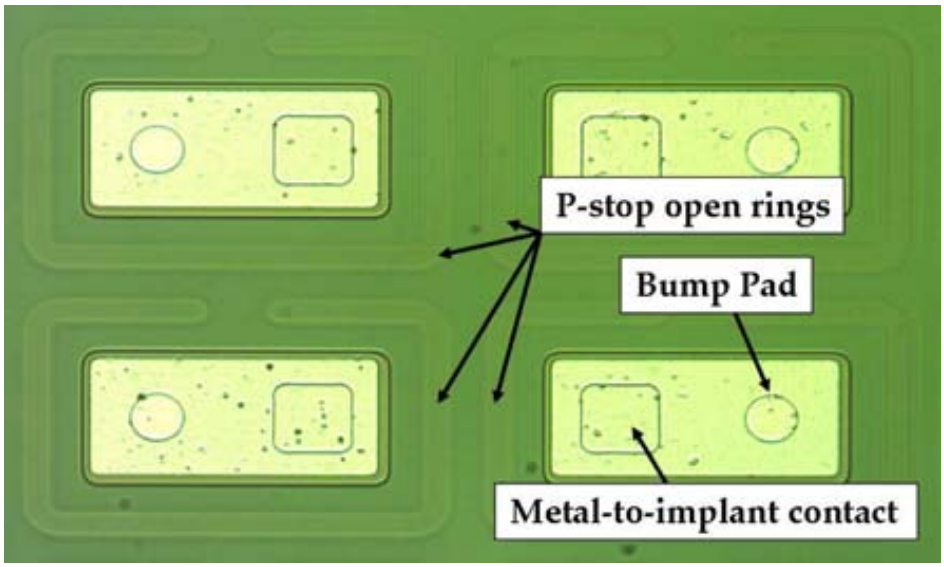


Figure 1.6: Picture of four pixels in the same double column for a pixel disk sensor.

The sensors for the pixel barrel use the moderated p-spray technique for inter-pixel isolation. A photograph of four pixels in a barrel sensor is shown in Fig. 1.7. Most area of a pixel is covered with the collecting electrode formed by the n-implant. The gap between the n-implants is kept small (20 μm) to provide a homogeneous drift field which leads to a relatively high capacitance of the order of 80-100 fF per pixel. In one corner of each pixel the so called bias dot is visible. They provide a high resistance punch-through connection to all pixels which allows on-wafer IV measurements which are important to exclude faulty sensors from the module production. The dark frame around the pixel implants visible in Fig. 1.7 indicates the opening in the nitride covering the thermal oxide. In this region the p-spray dose reaches the full level. The sensor shown in Fig. 1.7 has undergone the bump deposition process. The Indium bumps are visible as roughly 50 μm wide octagons. The sensors are processed on n doped DOFZ silicon with a resistivity of about 3.7 $k\Omega\, cm$ (after processing). This leads to a full depletion voltage of 50-60 V of the 285 μm thick sensors. All wafers for the production of the barrel sensors come from the same silicon ingot to provide the best possible homogeneity of all material parameters. The pixel barrel requires two different sensor geometries, 708 full (2×8 ROCs) and 96

half modules (1×8 ROCs). They were processed in 2005 and 2006 using two different mask sets.

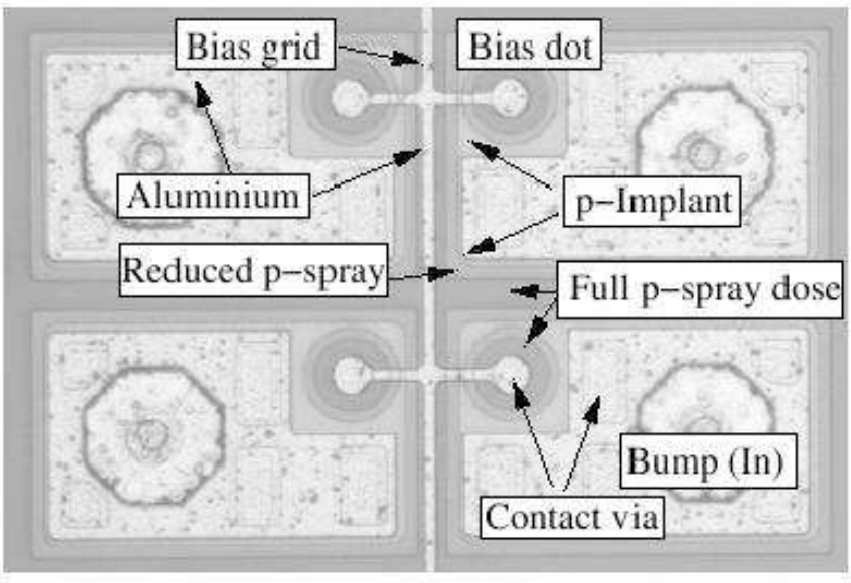


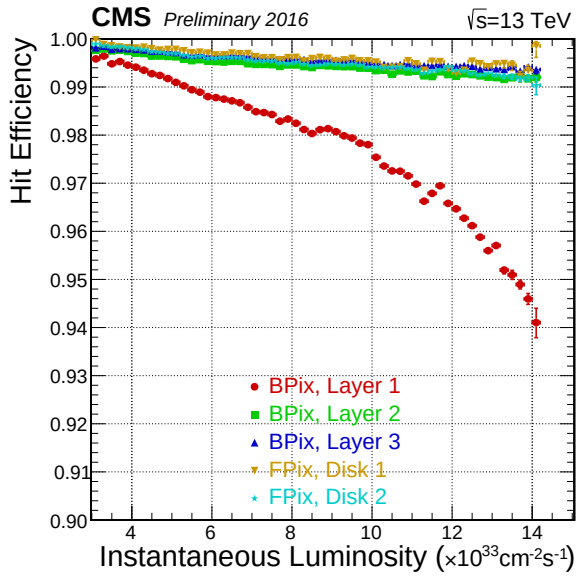
Figure 1.7: Photograph of four pixel cells of a barrel sensor. The Indium bumps are already deposited.

Sensor signals are read out by ROCs bump bonded to the sensors. A ROC is a full custom ASIC fabricated in a commercial 0.25 μm 5 metal layer CMOS process and contains 52×80 pixels [6]. Its main purposes are:

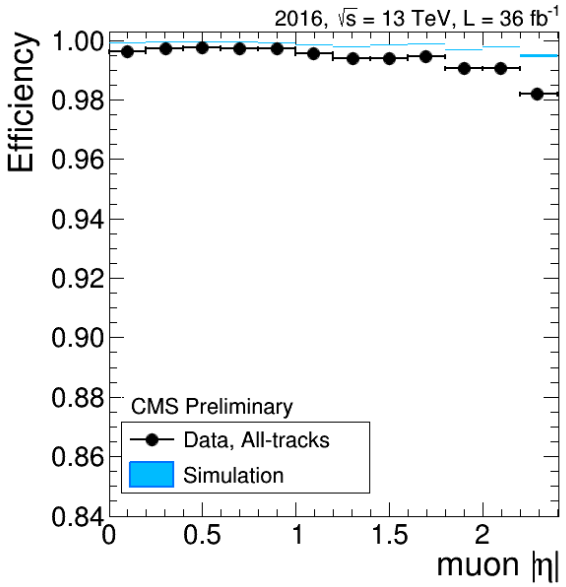
- Amplification and buffering of the charge signal from the sensor
- Zero suppression in the pixel unit cell. Only signals above a certain threshold will be readout. This threshold can be adjusted individually for each pixel by means of four trim bits. The trim bits have a capacitive protection against single event upset (SEU), which has shown to reduce SEUs by 2 orders of magnitude.
- Level 1 trigger verification: hit information without a corresponding L1 trigger is abandoned.

There are a few architecture inherent data loss mechanisms. The particle detection inefficiency has been measured in a high-rate pion beam. It is in fairly good agreement with expectations and reaches 0.8%, 1.2% and 3.8% respectively for the three layers at a luminosity of $1 \times 10^{34}\, cm^{-2}s^{-1}$ and 100 kHz L1 trigger rate.

The performance of the tracker in the first year of the Run II data taking can be summarized by the two plots in Fig. 1.8. In the left plot the hit efficiency of the different layers of the pixel detector as a function of the in instantaneous luminosity is reported. Hit efficiency is defined as the chance to find any clusters within a 500 μm area around an expected hit. The innermost layer is affected by a dynamic inefficiency as anticipated in previous paragraph. In the right plot the tracking efficiency for muons coming from Z boson decays as a function of the absolute pseudorapidity of the probe muon is reported.



(a) Hit efficiency of the different layers of the pixel detector as a function of the in instantaneous luminosity.



(b) Global tracking efficiency measured with the tag and probe method.

Figure 1.8: Performance of the CMS tracker in 2016.

Electromagnetic Calorimeter

The electromagnetic calorimeter, or ECAL, is designed to measure the energy of incident electrons and photons. The CMS ECAL is a homogeneous and highly granular calorimeter constituted of lead tungstate crystals ($PbWO_4$). The energy measurement is based on the conversion of the incident electron or photon to an electromagnetic shower, that interacts with the crystal material producing scintillation light. The crystals are at the same time the dense interacting material and the active scintillating medium, resulting in an excellent energy resolution.

The choice of $PbWO_4$ is motivated by its high density (8.28 g/cm^3), small radiation length ($X_0 = 0.89$ cm) and short Moliere radius (R = 2.2 cm). These parameters ensure an excellent containment of the electromagnetic shower within the crystals, which have a length of approximately 25 X_0. The lead tungstate is radiation hard and about 80% of its scintillation light is produced within 25 ns, making it ideal for the high instantaneous luminosity collisions of the LHC. The disadvantage of this material is the relatively low light yield, corresponding to about 30 photons per MeV of deposited energy, which calls for the usage of photodetectors with internal amplification, as detailed below.

The barrel part of the ECAL is constituted by 61200 crystals with a frontal transverse section of $22 \times 22\ mm^2$ and a length of 23 cm, and ensures the coverage of the region $| \eta | < 1.479$. The two endcaps are each made of 7324 crystals of a frontal transverse section of $28.62 \times 28.62\ mm^2$ and a length of 22 cm, and extend the coverage up to $| \eta | < 3.0$. The layout of the crystals in the ECAL is illustrated in Fig. 1.9. Crystals in the barrel are organized in 36 "supermodules", each covering half a barrel region and an angle of 20° in ϕ, and made of four "modules" where single crystals are mounted in a mechanic support. Crystals in each endcap are disposed in two semi-circular "dees". In both the barrel and the endcaps, crystals are mounted in a quasi-projective geometry, with their axes being tilted of 3° with respect to the direction that points to the nominal

interaction region. This ensures that no particle escapes the ECAL active volumes from the interstices between the crystals.

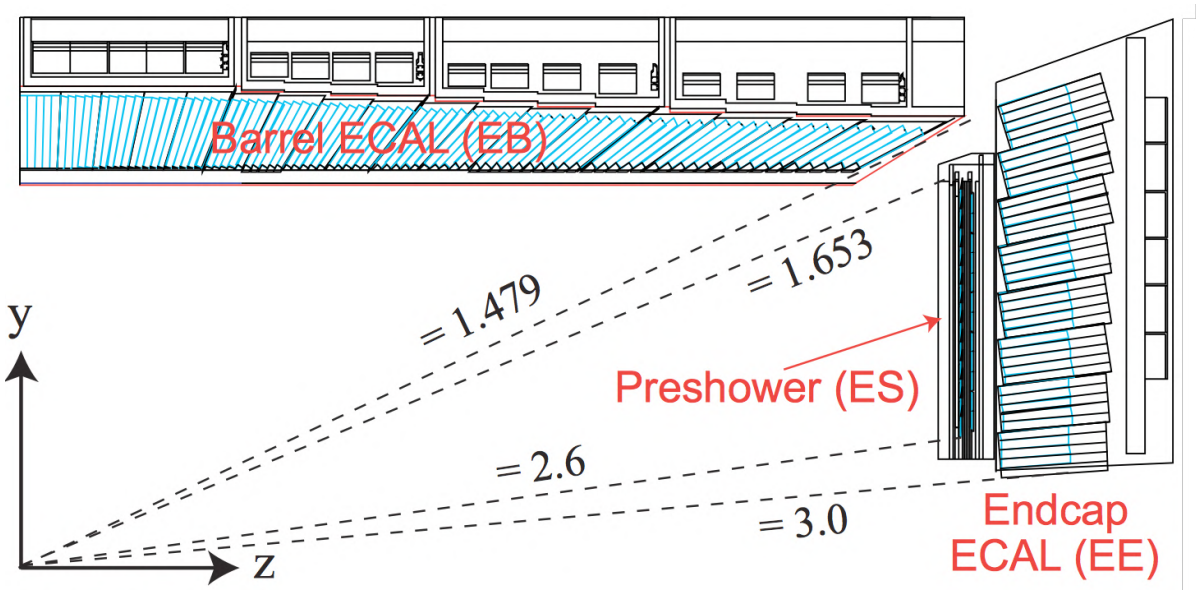


Figure 1.9: Longitudinal view of the ECAL layout, representing one quarter of the detector. The barrel and endcaps sections and the preshower detector are shown.

The crystal scintillation light is read out by detectors that are designed to work in the high magnetic field to which they are exposed and to be resistant to the radiation. The barrel part of the ECAL is instrumented with silicon avalanche photodiodes (APD) while vacuum phototriodes (VPT) are deployed in the endcaps. The signals from these photodetectors are amplified and shaped by the front-end electronics and sampled at a frequency of 40 MHz with a 12 bit analog-to-digital converter. The ECAL is operated at a temperature of 18 °C. The temperature is stabilized within 0.05 °C in EB and 0.1 °C in EE by a cooling system to control the changes in the light yield of the crystals (-2%/°C) and of the photodetectors (-2.3%/°C).

A sampling preshower (ES) is installed in front of the two endcaps to improve the discrimination of single photons from $\pi^0 \to \gamma\gamma$ decays. It is constituted by two layers of lead absorber in which the electromagnetic shower is initiated, followed by a layer of 2 mm-wide silicon strips to measure the deposited energy and the transverse profile of the shower shape.

The large doses of radiation to which the crystals are exposed cause a change in their transparency that is naturally recovered at the operating ECAL temperature. As a consequence the ECAL undergoes cycles of transparency reduction and recovery that correspond to the collisions and refill operations of the LHC. This effect is monitored and corrected with the injection of a 440 nm laser light in each crystal to derive time-dependent correction factors that are applied to the response.

Tracking and calorimeter detectors in CMS provide complementary measurements. The former can identify only charged particles and the precision of its momentum measurement increases as p_T decrease because of the larger curvature of the trajectory. Inversely, the latter can measure both charged and neutral particles with a resolution that increases with the particle energy itself because of the reduced impact of two of the three main effects that determine the resolution of a generic calorimeter. The first one is a

stochastic term that depends on the number n of scintillation photons (or elementary information carriers in general) produced in the interaction as $\sqrt{n}$, where n is in turn proportional to the incident particle energy E. A second term accounts for the noise in the detector and does not depend on E. Finally, a third term is related to detector inhomogeneities, resulting in an error that amounts to a constant fraction of E. The combined effect of these three factors results in a dependence of the calorimetric energy resolution σ on the particle of energy E as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{1.4}$$

where S, N and C denote the stochastic, noise, and constant terms, respectively. In ECAL test beam studies performed with incident electrons [7], the values S = 2.8%, N = 12%, and C = 0.3% were obtained. The ECAL response is calibrated to determine both the absolute energy scale and the channel-to-channel intercalibration, to compensate in particular for the intrinsic crystal light yield variations (≈15%) and the spread in the EE phototriodes (≈25%). The initial calibration derived from laboratory studies and cosmic rays exposures of crystals is now complemented with in-situ measurements based on collision events. As a result, the energy resolution for 45 GeV electrons is of about 2% in the barrel and 2–5% elsewhere, and increases to about 1.5% for electrons in the central part of the detector with little energy radiated by bremsstrahlung. An extensive description of the ECAL performance can be found in [8].

Hadronic Calorimeter

Hadrons typically traverse the ECAL volume releasing an amount of energy well below detection threshold. The hadronic calorimeter, or HCAL, is designed to absorb them within its volume and measure their energy. Compared to electron and photon interactions in the ECAL, hadron energies are intrinsically more difficult to measure from hadron showers induced in the HCAL. Nuclear and hadronic interaction result in non-Poissonian effects in the shower development, where many undetectable particles can also be produced. The presence of π^0 decaying to photon pairs also results in an electromagnetic component of the shower with a different response in the HCAL itself. All these effects limit the intrinsic resolution on hadron energies, which can be improved offline with the usage of the particle flow reconstruction techniques. Despite these limitations, the HCAL is an essential element in the reconstruction of final states containing jets or non-interacting particles such as neutrinos, since the calorimeter hermeticity and geometrical coverage allows for the computation of the imbalance in the transverse momentum sum of the event.

The barrel (HB) and endcap (HE) sections of the HCAL instrument respectively the regions $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$. Both the HB and HE are sampling calorimeters composed of a brass absorber and of active plastic scintillating tiles. The scintillation light is collected by wavelength shifter fibres embedded in the tiles and read out by hybrid photodiodes (HPD). Each readout cell is formed by the addition of a "tower" of scintillating tiles in a spatially localized region, and has a transverse $\Delta\eta \times \Delta\phi$ dimension of about 0.087×0.087 in the HB and of about 0.17×0.17 in the HE. As the HCAL is located between the ECAL and the internal surface of the solenoid, the limited space does not allow for a full containment of the secondary interaction shower. The detector is complemented by a outer hadron calorimeter (HO) located outside the solenoid, which

extends the total interaction depth to about 11 λ_0, where the constant indicates the average interaction length of hadrons in the calorimeter. The energy measurement in the forward region is complemented by the forward hadronic calorimeter (HF), that is located 11.2 m away from the interaction point and measures hadron interactions up to $\mid \eta \mid = 5.2$. Because of the higher radiation levels in the forward region, the HF is composed of steel absorbers and quartz fibres that produce light by Cherenkov effect, which is measured by photomultiplier tubes (PMT). Fibres of two different lengths are installed to estimate the electromagnetic and hadronic components of the shower. The global layout of the HCAL is illustrated in Fig. 1.10 . The overall HCAL performance is dominated by the imperfect containment of the hadronic shower, which results in a resolution sampling term of about 110% and a constant term of 9%, as measured in pion test beams[9].

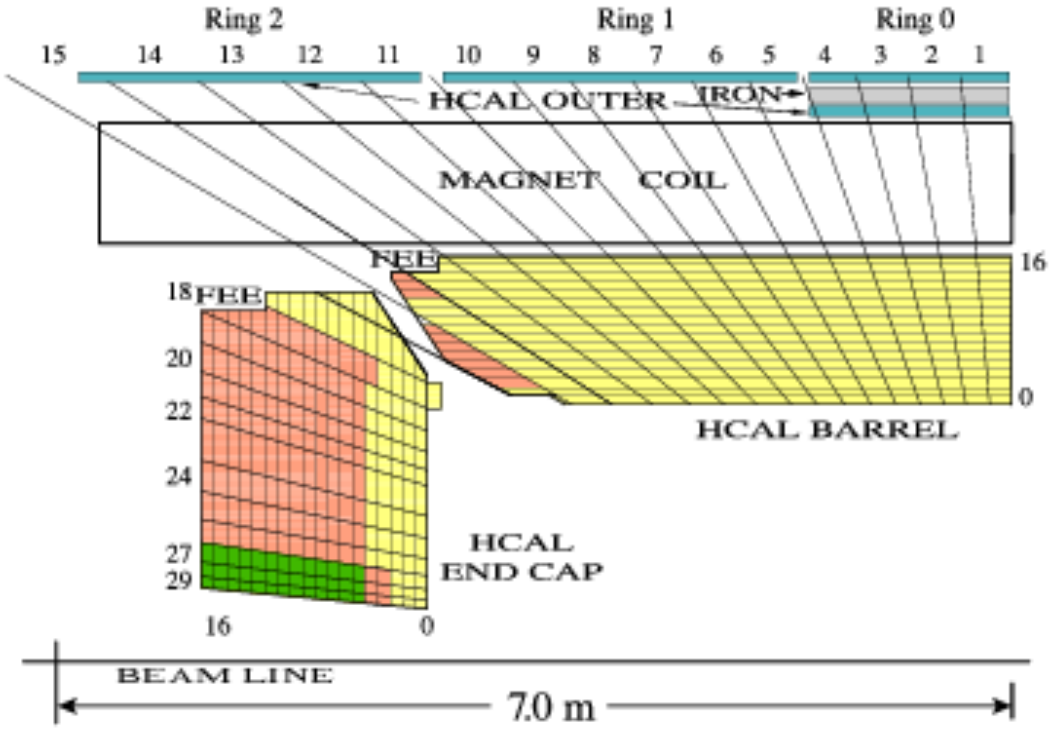


Figure 1.10: Longitudinal view of the HCAL layout. The barrel (HB) and endcaps (HE) detectors located inside the solenoid, the outer detector (HO) outside the solenoid, and the forward calorimeter (HF) are visible.

Muon Detectors

Because of their typical energy, muons produced in collisions at the LHC behave as minimum ionizing particles (MIPs). As a consequence, they traverse the ECAL, the HCAL, and the solenoid volumes without being stopped and are identified and measured in the muon detectors located in the outermost part of CMS. The muon momentum is measured using the return field of the solenoid inside the iron structure in which the muon detectors are embedded, and complements the measurement from the inner tracker previously discussed. CMS is instrumented with three types of muon detectors, chosen accordingly to the expected background rates and uniformity of the magnetic field, as illustrated in Fig. 1.11.

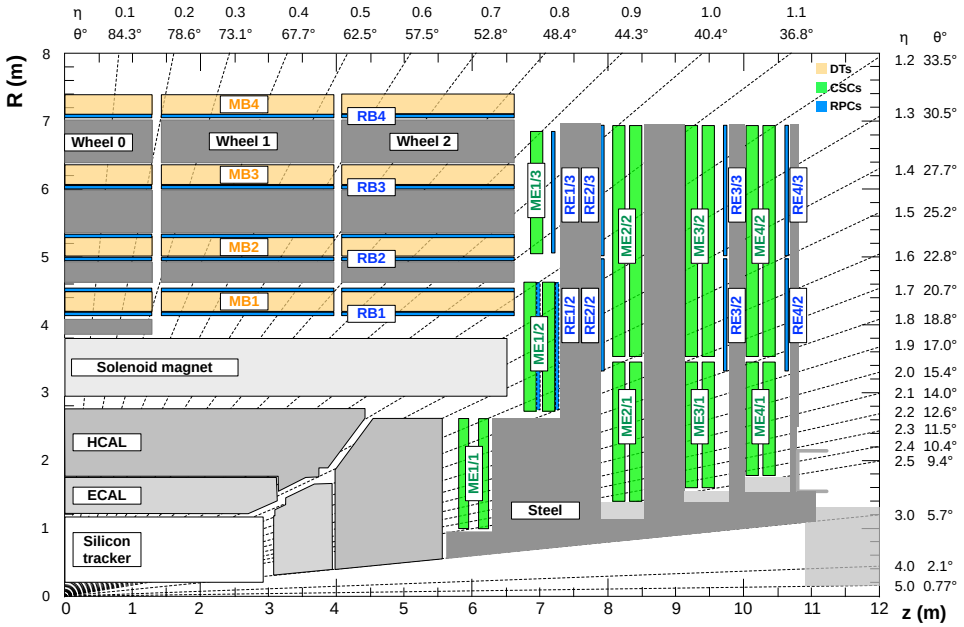


Figure 1.11: Layout of the muon systems of the CMS experiment. A quarter of the detector in its longitudinal view is shown. Orange, green, and blue regions denote respectively the DTs, CSCs, and RPCs subsystems.

In the barrel region, CMS deploys 250 drift tubes (DTs) uniformly distributed in the five barrel sections, or "wheels". Each wheel hosts four concentric rings of DT stations, organized in twelve contiguous sectors. The basic element of the DT detector is a rectangular cell of transverse size $4.2 \times 1.3\, cm^2$, containing an anode wire and filled with a Ar/CO_2 gas mixture. Electrodes on the top and bottom of the cell ensure a constant field and a uniform drift velocity of about 55 $\mu m/s$, while cathodes are placed on the sides of the cell. DT cells are organized in three groups of four elements (three "super-layers", SLs) that together compose a DT chamber. Muon traversing each group ionize the gas, and their position and angle are measured from the time needed by the electrons to drift toward the anode wires. The middle SL is oriented to measure the coordinate position in the z direction while the other two SLs provide a measurement in the (r, ϕ) plane. Each DT cell has a spatial resolution of about 200 μm, resulting in a resolution of 80–120 μm for the global chamber measurement.

Cathode strip chambers detectors (CSCs) are used to instrument the endcap regions of CMS ($0.9 <\mid \eta \mid < 2.1$). The different choice of detector technology is imposed by the higher background rates and the stronger magnetic field. CSCs are detectors designed in a trapezoidal shape and composed of six layers of anode wires interposed between seven segmented cathode plates disposed in the perpendicular direction. CSCs contain a $Ar/CO_2/CF_4$ gas mixture, which is ionized upon the passage of a muon. The resulting signals induced on the wires and on the strips are interpolated and provide a position measurement in the (r, ϕ) plane (anode wires) and along the z direction (strips). The CSC is a fast detector, capable of identifying the bunch crossing of a pp collision, and achieves a spatial resolution of 40-150 μm.

Resistive plate chambers (RPCs) are installed in both the barrel and endcaps and cover the region $\mid \eta \mid < 1.6$. RPCs are formed by two gaps consisting of two resistive Bakelite layers of 2 mm thickness separated by a 2 mm volume filled with a gas mixture. The detector is operated in avalanche mode and, when traversed by a muon, an avalanche

is generated by the high electric field inside the gas volume and is read out by strips located on the outer surface of the gap. Although RPCs have a modest spatial resolution of 0.8–1.2 cm, they have excellent timing properties with a resolution of the order of the ns, allowing for the determination of the pp bunch crossing.

The most recent measurements of the CMS Muon system performance are reported in [10].

1.2.3 The "Phase 1" upgrade of the detector

The Phase 1 upgrade of the CMS detector aimed at an efficient data taking during LHC Run 2 (2015-2018). Many steps were completed during LS1 (2013-2014), the upgrade of the pixel detector took place during the Extended Year End Technical Stop (EYETS) between 2016 and 2017, all the other works were completed by October 2019 during LS2.

Muon System

The baseline luminosity in Run 2 was defined to be $1 \times 10^{34}\, cm^{-2} s^{-1}$. The in-time pileup was foreseen to be right at the edge of the CMS design envelope: this would have posed difficulties for the muon system to trigger on muons with high tranverse momenta, which represent one of the key indicators of interesting electroweak interactions.

The CSC and RPC systems upgrade was driven by considerations of the impact of peak instantaneous luminosity on the muon trigger. A fourth layer of chambers has been added to reduce the accidental trigger rate and to preserve a low p_T threshold for the Level 1 Muon Trigger at high instantaneous luminosity. The CSC layer 1 (ME1/1) electronics has been upgraded with a new "Digital CSC Front End Board" (DCFEB) able to readout every strip separately (they were previously grouped into groups of three). This allowed ME1/1 to continue to contribute effectively to the muon trigger at high instantaneous luminosity so that CMS could retain a four plane coverage in the region $2.1 <\mid \eta \mid < 2.5$. Finally, a new muon trigger primitive electronics has been deployed to deliver the additional muon track segments which will be produced at high lumniosity to the upgraded CSC Trigger Track-Finder. Furthermore an R&D project was started to develop detectors that can extend coverage to the region $1.6 <\mid \eta \mid < 2.1$ or even higher. Possible technologies include RPCs optimzed to handle the high rate or Multi-Pattern Gas Detectors. This R&D project conducted to the design of new detectors that will be installed during the "Phase 2" upgrade, as discussed later.

The work on the DTs was driven by maintenance considerations over the lifetime of the experiment. A supply of front end trigger primitive chips has been produced since they were in short supply due to unexpectedly high mortality. The Sector Collector boards, located at the periphery of the detector where they are exposed to radiation and high magnetic fields and the cooling is marginal, have been moved to the Underground Control Room where the environment is more congenial.

Hadronic Calorimeter

The upgrade of the hadronic calorimeter was directed at handling the harsher conditions foreseen in Run II and providing opportunities to make improvements to the trigger.

The HPDs in the barrel and endcap section of the detector have been replaced with improved photodetectors, the Silicon Photomultipliers (SiPM). SiPMs have higher quantum efficiency and gain and better immunity to magnetic fields than HPDs. Since SiPMs

operate at relatively low voltages they do not produce large pulses from high voltage breakdown that mimic energetic showers like HPDs do. These features of the SiPMs together with their low cost and compact size compared to HPDs enable several major changes to the HCAL. These include the implementation of depth segmentation, which has advantages in coping with higher luminosities and compensating for radiation damage to the scintillators, and the use of timing information to clean up backgrounds.

The photomultipliers of the Forward Hadron Calorimeter have been replaced with new photomultipliers that have thinner glass windows and metal envelopes to reduce the amount of Cherenkov light generated by charged particles passing through the glass. The Cherenkov light from the glass creates signals with large pulse height that can be misinterpreted by the trigger as energetic particles. The new PMTs also have 4-way segmented anodes that provide additional rejection of these spurious signals. These PMTs also have higher quantum efficiency so the resolution of the HF will improve, and HF will last longer under irradiation.

Pixel Detector

The "Phase 0" CMS pixel detector was conceived over 15 years ago and designed for a maximum luminosity of $1 \times 10^{34}\, cm^{-2}s^{-1}$. Following the Phase 1 upgrade of the LHC the peak luminosity reached $2 \times 10^{34}\, cm^{-2}s^{-1}$ before the LS2. The Phase 0 pixel system was not able to sustain such extreme operating conditions due to large data losses in the read out chip (ROC) and was replaced in the 2016 EYETS. The modular design of CMS allows good access to the pixel system, which can be extracted relatively easily, independently of the beam pipe or the strip tracker.

The most severe limitation of the Phase 0 detector was the ROC, which was just adequate at the LHC design luminosity of $1 \times 10^3 4\, cm^{-2}s^{-1}$. At this luminosity, with a bunch spacing time of 25 ns, the dynamic inefficiency of the first layer amounts to 4%. The dynamic inefficiency increases exponentially with increasing luminosity. At $2 \times 10^{34}\, cm^{-2}s^{-1}$ bunch spacing the ROCs in the inner region will suffer an inefficiency of 15%, leading to a major degradation of the overall level of tracking performance. The new ROC installed in the first layer is called PROC600, a detailed description can be found in [11]. Furthermore, the three hit coverage of the detector was not completely hermetic leading to 10-15% inefficiencies at $\mid \eta \mid < 1.6$ and larger track seeding inefficiencies in the region $1.5 < \mid \eta \mid < 2.5$. This limits the efficiency of HLT tracking triggers and slows the full tracking algorithm. The situation could degrade even further at higher luminosities. As anticipated, the radiation hardness of the detector was not sufficient for operation up to the end of Phase 1, when the foreseen integrated luminosity will be around 300 fb^{-1}. Although the detector was constructed using the most radiation resistant technology known at the time of its fabrication, radiation damage will degrade its performance and necessitate replacement of the inner regions. Finally, the detector contains significant passive material that degrades tracking and calorimetric measurements due to multiple scattering, photon conversions and nuclear interactions.

The Phase 1 CMS pixel detector consists of 4 barrel layers and 3 disks in each endcap, as showed in Fig. 1.12. The 4 barrel layers are of equal length and are placed at radii of 3.9, 6.8, 10.9, and 16.0 cm. The three endcap disks are placed on each side of the central barrel detector, with a radial coverage ranging from 4.5 to 16.1 cm. The location of the first disk along the beam line is at 29.1 cm from the interaction point, the second and third disks are located at 39.6 cm and 51.6 cm from the interaction point.

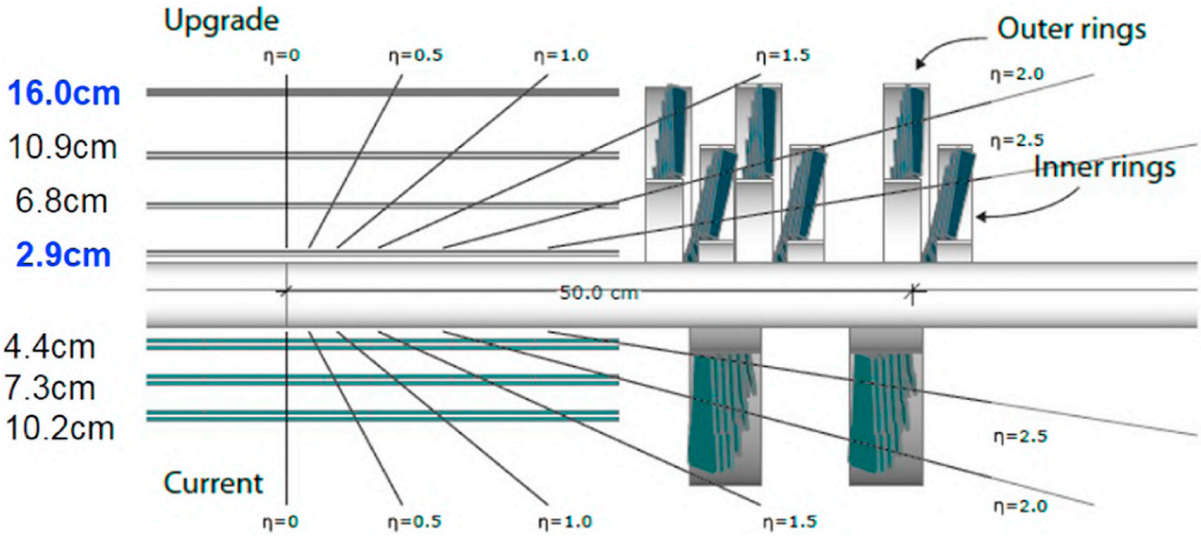
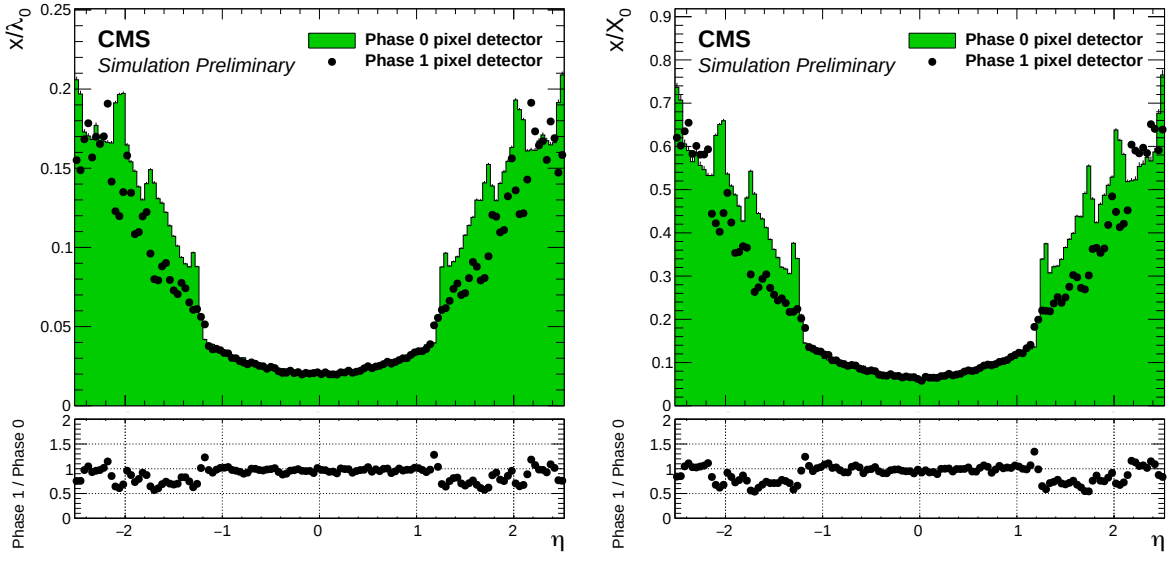


Figure 1.12: Layout of the Phase 0 and Phase 1 CMS Pixel Detector.

In the new design there is only one type of module with 16 ROCs in a 2×8 arrangement. They are mounted on ultra-lightweight support structures integrated with the cooling distribution system. Two-phase CO2 cooling system replaced the Phase 0 single phase C_6F_{14} system resulting in significant material reduction. Further material reduction has been achieved by using longer twisted pair or light-weight flex-cables to carry the signals to the optical hybrid boards. These boards, as well as the port cards and cooling manifolds, has been moved out of tracking region. The reduction in the material budget is shown in Fig. 1.13 .



(a) Thickness in terms of hadronic interaction lengths.

(b) Thickness in terms of electromagnetic interaction lengths.

Figure 1.13: Material budget of the Phase 0 and Phase 1 CMS pixel detector

The outer and inner parts of the detector has been designed such that they allow the inner layers and rings to be easily replaced after radiation damage. For FPIX, this requires each half-disk to be divided into an inner and outer ring. Similar to the Phase 0 detector,

the blades in the forward disks are rotated by 20° in a turbine like geometry to induce charge sharing. The separation of each half disk into an inner and outer assembly allowed the optimization of the orientation and tilting to obtain the best position resolution in both radial and ϕ directions.

The implementation of the Phase 1 pixel detector largely improved all aspects of CMS tracking:

- The addition of the extra layer dramatically improved the efficiency and resolution of pixel-only tracks. Pixel tracks are a crucial part of the HLT and they are also used to seed the full tracking, leading to an increase of the efficiency and a decrease of the fake rate for full tracks.

- The decrease in the amount of material and the increase in the number of measurement points improve the resolution of all track parameters. In particular, the resolution of longitudinal and transverse impact parameters are significantly improved.

- The efficiency and resolution enhancements lead to much improved primary and secondary vertexing. Vertexing is essential to associate the final state particles with the correct primary vertex in the high pile-up LHC environment. Secondary vertexing plays a key role in b-tagging and the search for various long-lived exotic states.

- The improvements in tracking efficiency, fake rate, parameter resolution, and vertexing all contribute to significant improvements in the b-tagging performance of the tracker, as is showed in Fig. 1.14 . The performance of the DeepCSV and DeepFlavour b jet identification algorithms is reported as the probability for non-b jets to be misidentified as b jet, as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated ttbar events using jets within $\mid \eta \mid < 2.5$ and with $p_T > 30 GeV$, b jets from gluon splitting to a pair of b quarks are considered as b jets. For comparison, the performance of DeepCSV with the 2016 detector (Phase 0) is also shown. The absolute performance in this figure serves as an illustration since the b jet identification efficiency depends on the event topology and on the amount of b jets from gluon splitting in the sample. An extensive description of the b-tagging algorithms used by CMS is reported in [12].

The performance of the upgraded tracker is summarized by the two plots in Fig. 1.15. In the left plot the hit efficiency of the different layers of the pixel detector as a function of the in instantaneous luminosity is reported. Hit efficiency is defined as the chance to find any clusters within a 500 micron area around an expected hit. The dynamic inefficiency in the first pixel layer is strongly reduced. The behavior of the new ROC has been further investigated and the submission of a new version of the chip was already done in order to rebuild the layer 1 during LS2. In the right plot the tracking efficiency for muons coming from Z boson decays as a function of the absolute pseudorapidity of the probe muon is reported: the global efficiency is higher than in 2016.

1.2.4 The "Phase 2" upgrade of the detector

The brightness of the beams and the new focusing/crossing scheme at the interaction point that will be used in the HL-LHC will enable the accelerator to potentially deliver

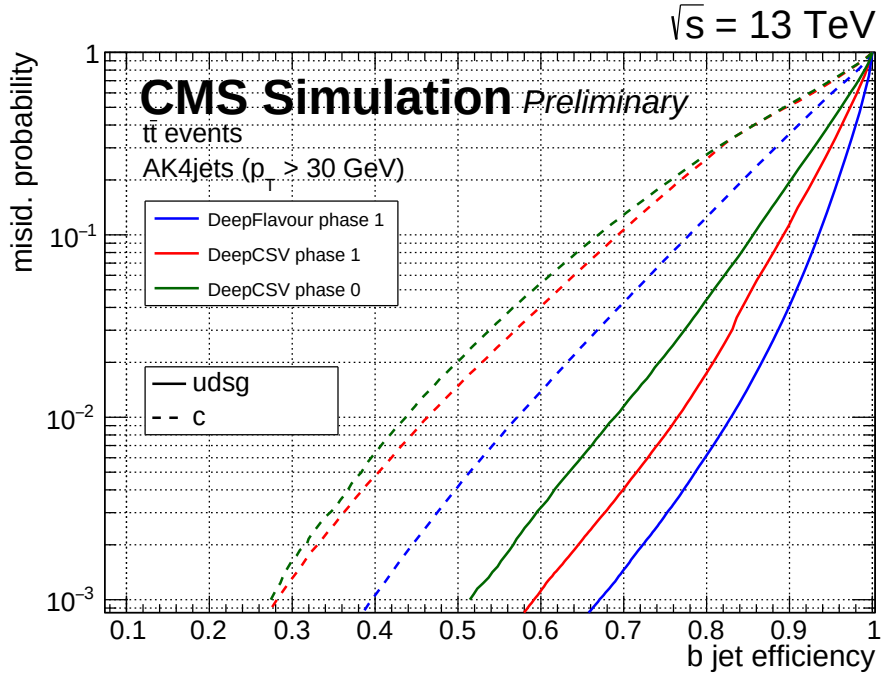
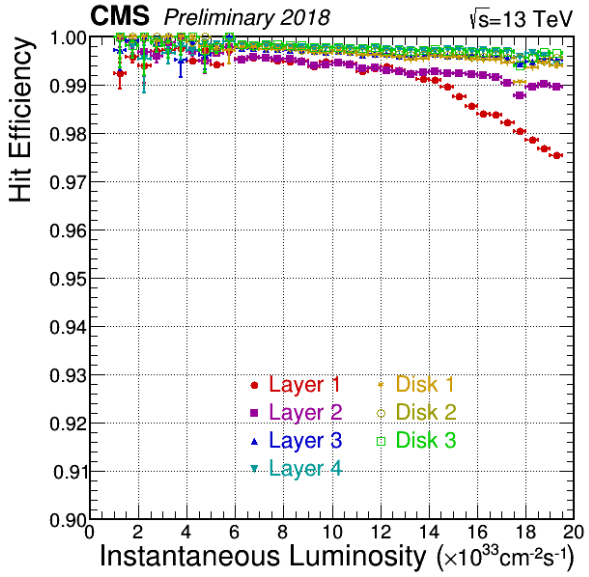
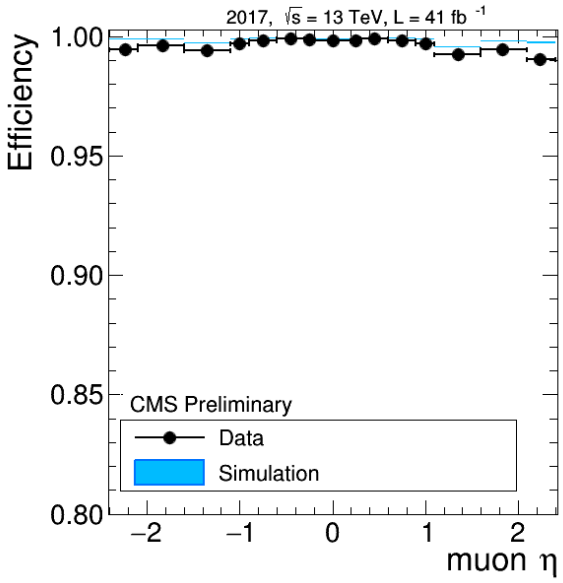


Figure 1.14: Performance of b-tagging algorithms currently in use in CMS.



(a) Hit efficiency of the different layers of the pixel detector as a function of the in instantaneous luminosity.



(b) Global tracking efficiency measured with the tag and probe method.

Figure 1.15: Performance of the CMS tracker in 2017.

a luminosity of $2 \times 10^{35}\ cm^{-2}s^{-1}$ at the beginning of each fill. This would increase the interaction rate and collision PU beyond the capabilities of the existing detector and trigger technologies. It is therefore proposed to maintain a lower but stable instantaneous luminosity by continuously tuning the beam focus and crossing profile throughout the duration of beam fills in a process referred to as luminosity levelling. The nominal scenario is to operate at a leveled luminosity $5 \times 10^{34}\, cm^{-2}s^{-1}$, corresponding to a mean pileup of 140 interactions per beam crossing. The so-called "ultimate" scenario foresees an increase of the instantaneous luminosity to $7.5 \times 10^{34}\, cm^{-2}s^{-1}$, leading to an integrated luminosity of 4000 fb^{-1} at the end of operation. The primary goal of the Phase II upgrade program is therefore to maintain the excellent performance of the Phase I detector under these challenging conditions throughout the extended operation of HL-LHC.

A significant effort has been expended to understand the effect of radiation damage: the exposure of test components to radiation levels matching anticipated HL-LHC doses showed that the tracker and the endcap calorimeters must be replaced for Phase II.

With these required changes, the performance issues associated with high PU, that are also the most important in the inner and forward detector regions, can be addressed. Pile-up mitigation in CMS heavily relies upon particle-flow event reconstruction. To this end, the tracker granularity can be increased to maintain the excellent tracking efficiency to enable the determination of the original p-p collision points for all charged particles. New endcap calorimeter configurations will also provide the opportunity to optimize segmentation and improve energy resolution, particularly for jets.

The ability to ensure efficient event selection for data acquisition is a key prerequisite to fully benefit from increased luminosity. The precise study of the relatively low-mass Higgs boson discovered in 2012, and the search for new particles occurring in cascade decays will require continued use of low transverse momentum, p_T, trigger thresholds. To achieve this, the trigger electronics must be upgraded. A sufficient reduction in trigger rate can only be accomplished by improving p_T resolution to obtain lower rates without loss of efficiency, and by mitigating the effect of the combinatorial backgrounds arising from PU. A new approach is therefore required, namely the introduction of tracking information at L1, providing the capability to implement trigger algorithms similar to that of the current HLT, including the use of precise momentum measurements. Facilitating tracking in the L1 trigger is an important driver of the design of the Phase II Tracker. The upgraded L1 "track trigger" will require a new hardware architecture to incorporate the tracking information. While the addition of track information in the L1 trigger provides significant gains in rate reduction with good efficiency, it will nevertheless be necessary to increase the trigger accept rate in order to maintain the required efficiency for all of the important physics channels. This is particularly the case for triggers involving hadrons and photons, for which the sensitivity to PU is higher and the track trigger is somewhat less efficient.

The measurement of processes with small production cross-sections and/or decay branching ratios is a major goal of the HL-LHC physics program. This requires specific upgrades in the forward regions of the detector to maximize the physics acceptance over the largest solid angle possible. To ensure proper trigger performance within the present coverage, the muon system will be completed with new chambers. The new endcap calorimeter configuration offers the opportunity to extend the muon coverage with a tagging station up to $\mid \eta \mid \approx 3$ or more, with significant acceptance gain for mult-muon final states. To mitigate PU effects in jet identification and energy measurement, the tracker will be extended up to $\mid \eta \mid \approx 4$, thereby also covering the peak production region

of jets accompanying Vector Boson Fusion (VBF) and Vector Boson Scattering (VBS) processes, which are among the highest priorities of the physics program. With this extension, measurements of total energy and missing energy will be greatly improved, and b-tagging acceptance will be increased.

As the luminosity integrated over the Phase II operation period will not be limited by the accelerator performance but by the ability of the detector to sustain high PU, the upgrades of the readout electronics will be designed with some margin to allow efficient data taking up to a PU of 200. It is expected that the sustainable luminosity limit will be driven by the performance of subdetectors that are not going to be replaced for Phase II.

Calorimeter Endcaps

The electromagnetic and hadronic endcap calorimeters will suffer significant radiation damage by LS3, and so must be replaced. The replacement is called "High Granularity Calorimeter" (HGC) and has electromagnetic and hadronic sections with excellent transverse and longitudinal segmentation. It will provide detailed three dimensional images of showers.

The longitudinal cross section of one half of one endcap of the HGCAL design is shown in Fig. 1.16. It consists of a 52 layer sampling calorimeter with 28 layers constituting a front electromagnetic section (CE-E) of 25 electromagnetic radiation lengths (X_0) and 1.3 hadronic interaction lengths (λ_0) thickness and 24 layers composing a rear hadronic section (CE-H) of 8.5 λ_0 thickness. Copper, tungsten, and lead are used as absorber materials in the CE-E, while stainless steel is used in the CE-H. Detector planes equipped with either silicon sensors or scintillator tiles with silicon photomultiplier (SiPM) readout are used. The silicon sensors will account for $\approx$ 55% of the total active area, which is 1100 m^2. The silicon is divided into $\approx$ 6 million channels with an area of either $\approx$ 0.5 or $\approx$ 1 cm^2. The CE-H design foresees $\approx$ 400,000 scintillator+SiPM channels.

The total ionizing dose absorbed by the HGCAL active materials after the full HL-LHC run period is expected to be between 10^1 and 10^6 Gy depending on η and depth. Within this harsh radiation environment, sensitivity to the energy deposit of a single minimum ionizing particle (MIP) during the complete HL-LHC run period is required for calibration. Silicon sensors will be placed in the HGCAL wherever the predicted signal-to-noise (S/N) ratio for single MIP in a scintillator+SiPM channel would be less than 5 after 3000 fb^{-1} of integrated luminosity. In this scheme, the minimum MIP S/N predicted in the silicon channels after 3000 fb^{-1} is between 2.2 and 4.7. The entire calorimeter will be operated at -30 °C to reduce the silicon sensor and SiPM leakage currents after irradiation. The S/N values are quoted for this temperature.

More detailed information can be found in the Techincal Deisgn Report released in 2017 [13].

Muon Endcaps

The muon system in the region $1.5 <| \eta |< 2.4$ currently consists of four stations of Cathode Strip Chambers (CSC). It is the only region of the muon detector that lacks redundant coverage despite the fact that it is a challenging region for muons in terms of backgrounds and momentum resolution. To maintain good L1 muon trigger acceptance in this region it is therefore proposed to enhance these four stations with additional chambers that make use of new detector technologies with higher rate capability. The

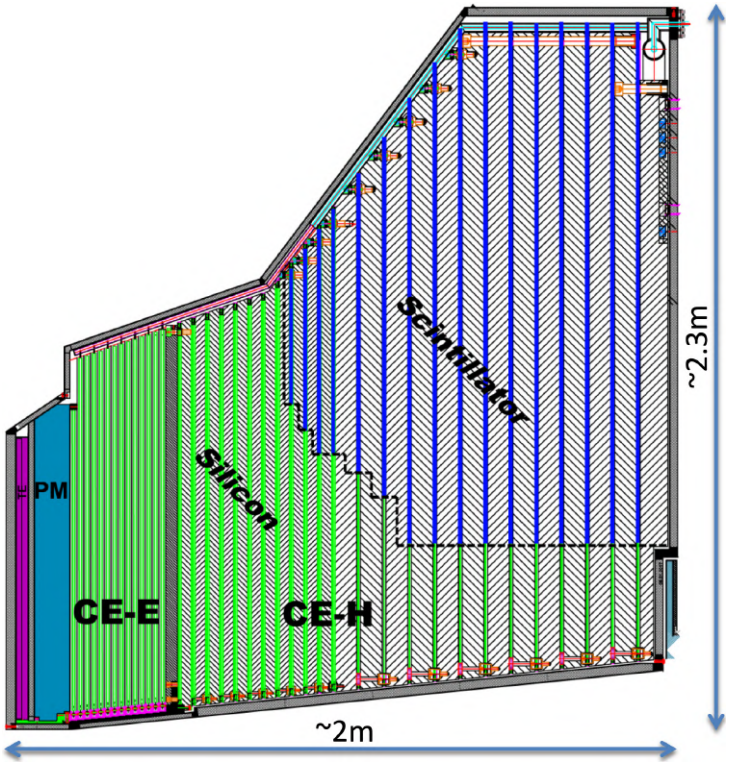


Figure 1.16: Longitudinal cross section of one half of one endcap of the proposed HGCAL design.

two first stations are in a region where the magnetic field is still reasonably high and so will use Gas Electron Multiplier (GEM) chambers for good position resolution in order to improve momentum resolution for the standalone muon trigger and matching with tracks in the global muon trigger. The two last stations will use low-resistivity Resistive Plate Chambers (RPC) with lower granularity but good timing resolution to mitigate background effects. The upgraded layout of the Muon System is showed in Fig. 1.17. More detailed information can be found in the Techincal Deisgn Report released in 2017 [14].

Tracker

The present strip tracker was designed to operate with high efficiency at an instantaneous luminosity of $1 \times 10^{34}\ cm^{-2}s^{-1}$, with an average pileup of 20-30 collisions per bunch crossing and up to an integrated luminosity of 500 fb^{-1}. The tracker is indeed performing very well at current instantaneous luminosities that are well above the design value. Performance will however degrade due to radiation damage beyond 500 fb^{-1}.

The original pixel detector has already been replaced with a new device, the "Phase-1" pixel detector, during the extended year-end technical stop (EYETS) 2016/2017. As the instantaneous luminosity exceeded the original design value and is projected to increase further prior to LS3, this upgrade was needed to address dynamic inefficiencies in the readout chip at high rates.

Before the start of the HL-LHC both the strip tracker and the Phase 1 pixel detector will have to be replaced due to the significant damage and performance degradation they would suffer during operation at the HL-LHC, and to cope with the more demanding operational conditions.

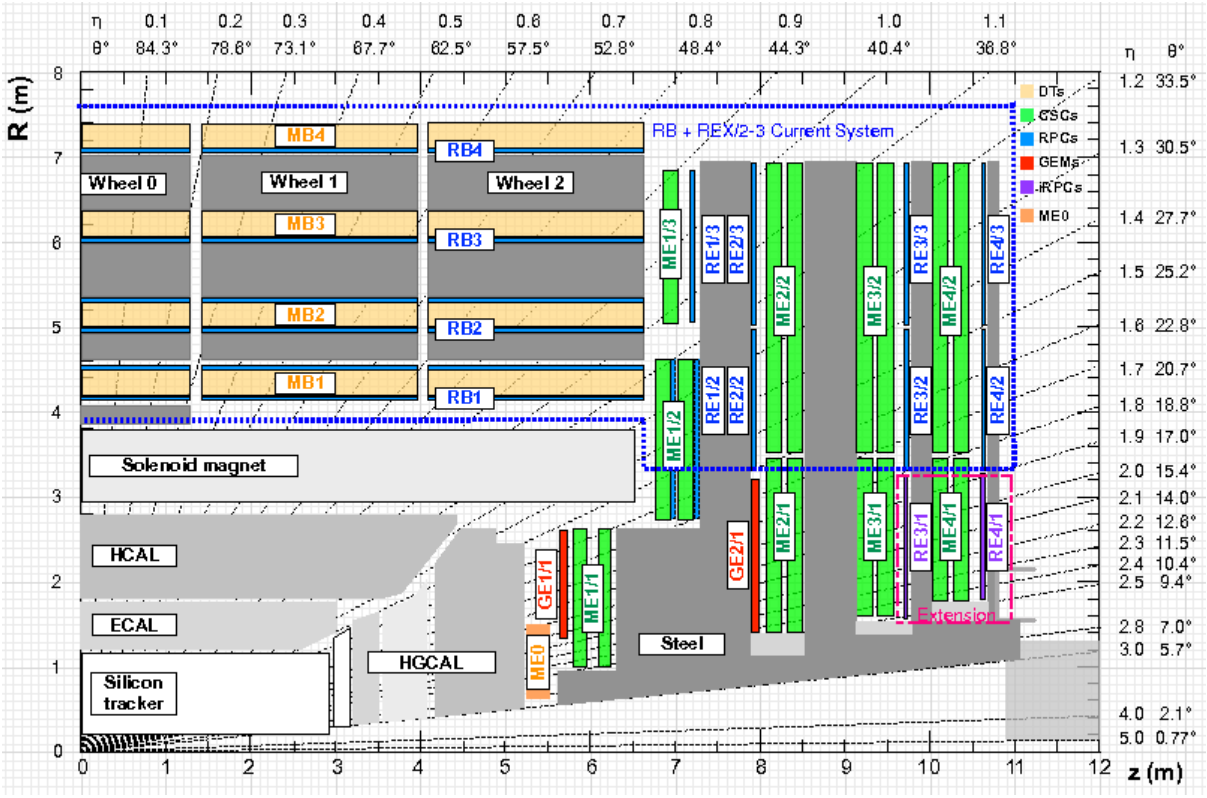


Figure 1.17: Proposed layout of the muon systems of the CMS experiment for Phase II. A quarter of the detector in its longitudinal view is shown.

Accumulated radiation damage in the pixel sensors reduces the charge collection efficiency as well as the Lorentz angle, leading initially to decreased charge sharing among neighbouring pixels and hence to deteriorated spatial resolution, and possibly to reduced hit efficiency.

For the present strip tracker, the most prominent changes of detector properties with irradiation are the increase of the sensor depletion voltage and of the leakage current. The latter can be mitigated, up to a certain point, by lowering the operating temperature of the cooling system, while the former cannot. It has been demonstrated that basically all double-sided strip modules cannot be operated anymore at the nominal cooling temperature already after 1000 fb^{-1}.

Studies of the expected performance of the current tracking system as a function of integrated luminosity have shown unacceptable degradation beyond about 1000 fb^{-1}, including the deterioration of tracking and b tagging performance and a worsening of the impact parameter resolution. The physics program would also be affected by limitations in readout bandwidth and trigger latency. The tracker must therefore be entirely replaced for the Phase 2 physics program.

The Phase 2 tracker will consist of an Inner Tracker (IT) based on silicon pixel modules and an Outer Tracker (OT) made of silicon modules with strip and macro-pixel sensors. The main requirements for the tracker upgrade can be summarized as follows.

- *Radiation tolerance.* The upgraded tracker must be fully efficient up to a target integrated luminosity of 4000 fb^{-1}. This requirement must be fulfilled without any maintenance intervention for the Outer Tracker. For the Inner Tracker, where pixel detector modules are deployed, it is envisaged to keep the present concept of

accessibility, allowing the extraction of the Inner Tracker during regular shutdowns and offering the option to replace modules and other elements as they accumulate substantial radiation damage. A replacement of the IT inner layer is currently foreseen during LS5. Detailed FLUKA [15, 16] simulations have been performed to estimate the radiation exposure of the different detector regions, which is about one order of magnitude higher compared to the requirements that were used for the design of the existing tracker, reaching a 1 MeV neutron equivalent fluence of $1.9 \times 10^{16}\ n_{eq}/cm^2$ in the innermost regions of the Inner Tracker. The particle fluence depends primarily on r, while the variation with z is very moderate (Fig. 1.18).

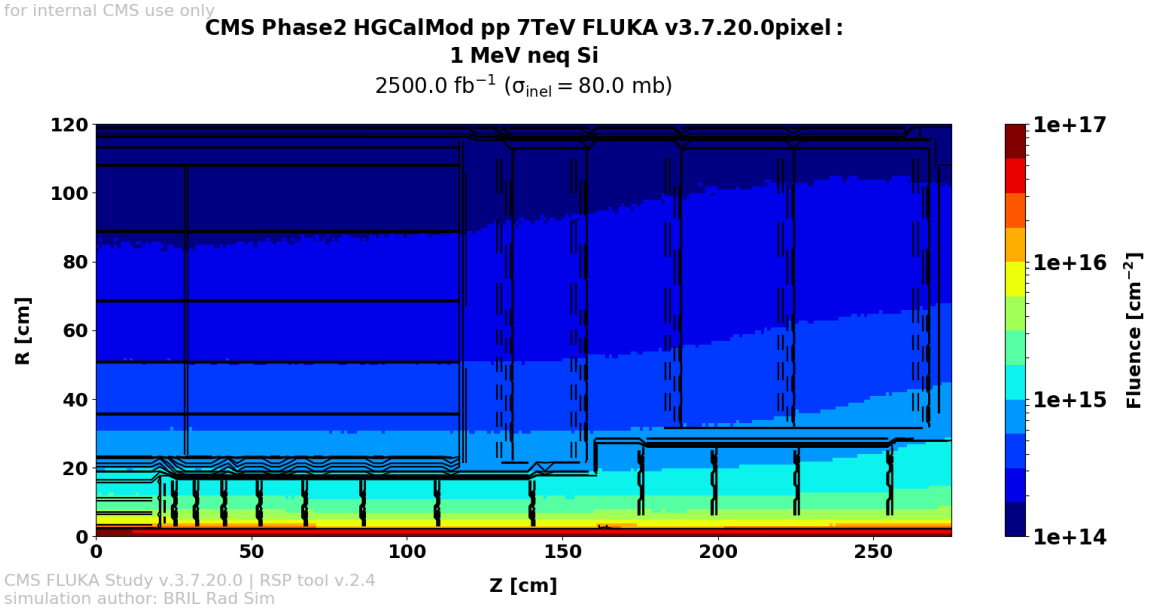


Figure 1.18: Integrated particle fluence in 1 MeV neutron equivalent per cm^2, for the Phase 2 tracker. The estimates shown correspond to a total integrated luminosity of 2500 fb^{-1} of pp collisions at $\sqrt{s} = 14$ TeV.

- *Increased granularity.* In order to ensure efficient tracking performance with a high level of pileup, the channel occupancy must be kept at around or below the per cent level (per mille level) in the Outer Tracker (Inner Tracker), which requires a high channel density. Target values of 140 and 200 collisions per bunch crossing are used to benchmark the performance of the detector.

- *Improved two-track separation.* The present tracker has limited track finding performance in highly energetic jets, due to hit merging in the pixel detector. In order to optimally exploit the large amounts of collision data that will be taken during high luminosity operation, two-track separation needs to be improved.

- *Reduced material in the tracking volume.* The exploitation of the high luminosity will greatly benefit from a lighter tracker. The performance of the current tracker is affected by the amount of material, which also influences the performance of the calorimeters and of the overall event reconstruction in CMS.

- *Robust pattern recognition.* Track finding under high pileup conditions becomes increasingly more difficult and time consuming. The design of the upgraded tracker

should enable fast and efficient track finding, notably at the HLT.

- *Contribution to the level-1 trigger.* The selection of interesting physics events at the first trigger stage becomes extremely challenging at high luminosity, not only because of the rate increase, but also because selection algorithms become inefficient in high pileup conditions. The CMS trigger will operate with substantially increased latency and output rate, and the tracker has to comply with those. In addition, in order to preserve and possibly enhance the performance in a wide spectrum of physics channels, CMS has chosen to use tracking information in the L1 event selection, anticipating part of the reconstruction presently performed in the HLT.

- *Extended tracking acceptance.* The overall CMS physics capabilities will greatly benefit from an extended acceptance of the tracker and calorimeters in the forward region. The upgraded tracking system will provide efficient tracking up to about $| \eta | = 4$.

One quarter of the Phase 2 tracker layout can be seen in Fig. 1.19. Figure 1.20 shows the average number of active layers that are traversed by particles originating from the luminous region, for the complete tracker as well as for the Inner Tracker and the Outer Tracker separately.

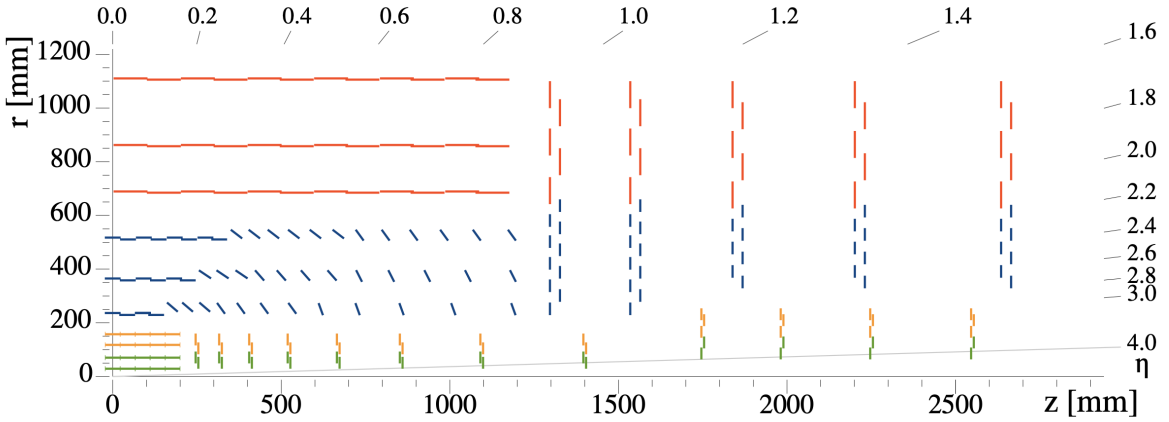


Figure 1.19: Sketch of one quarter of the tracker layout in r-z view. In the Inner Tracker the green lines correspond to pixel modules made of two readout chips and the yellow lines to pixel modules with four readout chips. In the Outer Tracker the blue and red lines represent the two types of modules described in the text.

The number of layers has been optimized to ensure robust tracking, i.e. basically unaffected performance when one detecting layer is lost in some parts of the rapidity acceptance. The six layers of the Outer Tracker are the minimum requirement to ensure robust track finding at the L1 trigger in the rapidity acceptance of $| \eta | < 2.4$. In the central region, the four layers of the Inner Tracker (the same number as already implemented in the Phase 1 upgrade) are the key for the pixel-based track seeding, which ensures good track finding performance with affordable computing time down to very low transverse momentum. Preliminary studies indicate that with the same number of layers as in the Phase 1 detector good performance is preserved also at the expected HL-LHC pileup levels, thanks to the smaller pixel size. In the forward part, the number of detection layers deployed ensures that particles originating from the luminous region traverse on

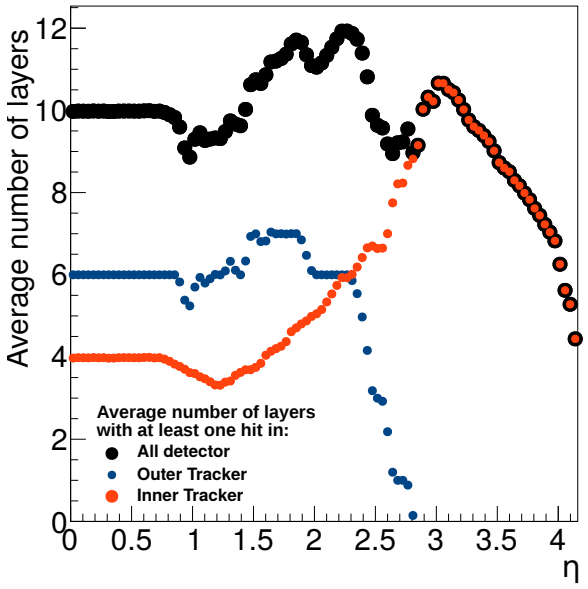


Figure 1.20: Average number of module layers traversed by particles, including both the Inner Tracker (red) and the Outer Tracker (blue) modules, as well as the complete tracker (black). Particle trajectories are approximated by straight lines, using a flat distribution of primary vertices within $\mid x_0 \mid < 70$ mm, multiple scattering is not included.

average more than eight layers of active detector modules up to $\mid \eta \mid \approx 3.5$, and more than six up to $\mid \eta \mid \approx 4.0$, as shown in Fig. 1.20, providing robust performance over the whole rapidity acceptance.

The enhancement of the trigger performance involves both a higher output rate of interesting events and an improved discriminating power of the event selection, which is more challenging in a high pileup environment. Improved discriminating power will be achieved by using more information in the trigger decision, with a longer latency available for its processing. The use of tracking information in the L1 trigger will improve the transverse momentum resolution of various objects at L1 (e.g. jets), will allow the exploitation of information on track isolation, and will contribute to the mitigation of pileup. CMS plans to enhance the first level trigger rate from presently 100 kHz to 750 kHz and to increase the latency from the present value of 3.2 μs to 12.5 μs. The front-end electronics and the L1 trigger track reconstruction need to comply with these new requirements.

The necessity of providing tracking information to the L1 trigger is a main driver for the design of the Outer Tracker, including its module concept. The use of tracking information in the L1 trigger implies that the tracker has to send out self-selected information at every bunch crossing. Such functionality relies upon local data reduction in the front-end electronics, in order to limit the volume of data that has to be sent out at 40 MHz. This is achieved with modules that are capable of rejecting signals from particles below a certain pT threshold, referred to as "p_T modules". Tracks from charged particles are bent in the transverse plane by the 3.8 T field of the CMS magnet, with the bending angle depending on the p_T of the particle. The modules are composed of two single-sided closely-spaced sensors readout by a common set of front-end ASICs that correlate the signals in the two sensors and select the hit pairs (referred to as "stubs") compatible with particles above the chosen p_T threshold (Fig. 1.21 (a)). A threshold of

around 2 GeV corresponds to a data volume reduction of roughly one order of magnitude, which is sufficient to enable transmission of the stubs at 40 MHz, while all other signals are stored in the front-end pipelines and read out when a trigger signal is received. The pT threshold depends on the acceptance window, which can be tuned to a certain level by programming the respective setting in the readout chip. For the p_T modules a few different values of sensor spacing are used, optimized to achieve the desired p_T filtering in different regions of the detector (Fig. 1.21 (b) and (c)). For a pitch of about 100 μm between silicon strips (or macro-pixels, as detailed below) in the transverse plane, sufficient p_T resolution can be achieved down to a radius of about 200 mm in a barrel geometry, thanks to the 3.8 T magnetic field of CMS. The concept is therefore applicable in the Outer Tracker, and limited in angular acceptance to about $|\eta| < 2.4$.

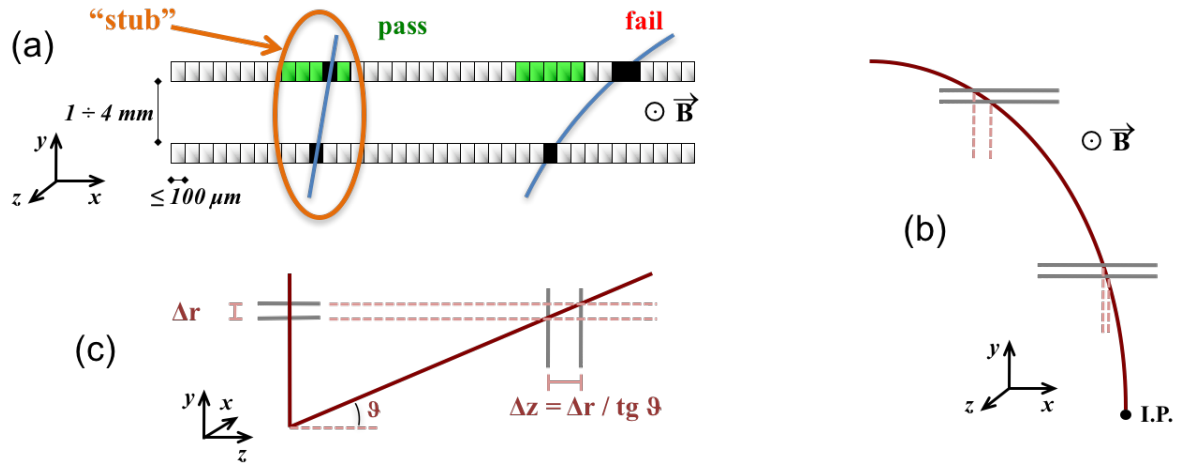


Figure 1.21: Illustration of the p_T module concept. (a) Correlation of signals in closely-spaced sensors enables rejection of low-p_T particles; the channels shown in green represent the selection window to define an accepted stub. (b) The same transverse momentum corresponds to a larger distance between the two signals at large radii for a given sensor spacing. (c) For the endcap discs, a larger spacing between the sensors is needed to achieve the same discriminating power as in the barrel at the same radius.

The Outer Tracker is populated with p_T modules, implementing the L1 trigger functionality. The p_T module concept relies on the fact that the strips of the top and bottom sensors of a module are parallel to each other. With the strip direction being parallel to the z axis in the barrel and nearly radial in the endcaps, this prevents the concept of stereo strips to be used to measure the z coordinate (r coordinate) in the barrel (endcaps). For this reason two versions of p_T modules have been realized: modules with two strip sensors (2-strip or 2S modules) and modules with a strip and a macro-pixel sensor (pixel-strip or PS modules). The strips in the 2S modules have a length of about 5 cm, while those in the PS modules are about 2.4 cm long. In PS modules one of the two sensors is segmented into macro-pixels of about 1.5 mm length, providing the z(r) coordinate measurement in the barrel (endcaps). The PS modules are deployed in the first three layers of the Outer Tracker, in the radial region of 200-600 mm, i.e. down to radii at which the stub p_T resolution remains acceptable and the data reduction effective. The 2S modules are deployed in the outermost three layers, in the radial region above 600 mm. In the endcaps the modules are arranged in rings on disc-like structures, with the rings

at low radii, up to about 700 mm, equipped with PS modules, while 2S modules are used at larger radii. The precision on the z coordinates provided by the three PS barrel layers constrain the origin of the trigger tracks to a portion of the luminous region of about 1 mm, which is sufficiently precise to partially discriminate particles coming from different vertices.

The p_T module concept implies that both the top and the bottom silicon sensors of a module must be connected to the readout electronics that performs stub finding. In order to implement the connectivity between the upper and lower sensors with reliable and affordable technologies, the two halves of each module are read out independently by front-end hybrids on the two ends, which prevents communication between the sensor halves and thus the reconstruction of stubs when particles cross the module near the center with a large incident angle. In a flat barrel layout such an effect translates into a geometrical inefficiency of stub finding, which is larger than 30% at the edge of the first barrel layer. To overcome this limitation, CMS has developed an innovative layout where the first three barrel layers, that are populated with PS modules, feature progressively tilted modules, nearly perpendicular to incident particles over the entire barrel length (Fig. 1.19). In the three outer layers of the barrel the effect of stub finding inefficiency is much less severe because of the smaller incidence angles (the incidence angle is measured with respect to the sensor normal), the smaller sensor spacing at those radii, and the double length of the 2S modules along z.

The Phase 2 Inner Tracker (IT) is designed to maintain or improve the tracking and vertexing capabilities under the high pileup conditions of the HL-LHC. The Inner Tracker has to cope with a ionizing radiation dose of up to 1 Grad and a hadron fluence of up to $1.9 \times 10^{16}\, n_{eq}/cm^2$ after 2500 fb^{-1} of integrated luminosity. This benchmark was chosen assuming a replacement of the IT inner layer during LS5.

Thin planar n-in-p type silicon sensors with an active thickness of 150 μm, segmented into pixel sizes of $25 \times 100\,\mu m^2$ (with the long side pointing along z in the barrel and along r in the endcaps) or $50 \times 50\ \mu m^2$, are expected to allow for a good detector resolution that is relatively stable with respect to radiation damage. The resulting reduction in the pixel area by a factor of six compared to the Phase 0 and Phase 1 pixel detectors will enable to achieve low occupancy and improved track separation in dense environments like high p_T jets.

An alternative option that is being actively pursued is the possibility to use 3D silicon sensors, offering intrinsically higher radiation resistance because of the shorter charge collection distance. Since the production process is more expensive and thus not suitable for large volumes, the use of 3D sensors could be limited to the regions of highest particle fluences.

For the readout chip, the envisaged small cell size can be achieved with the use of 65 nm CMOS technology and an architecture where a group of channels (referred to as pixel region) shares digital electronics for buffering, control, and data formatting. Such a pixel readout chip is being developed within RD53, a joint ATLAS-CMS collaboration [17].

The baseline Inner Tracker layout is shown in Fig. 1.22 .The detector comprises a barrel part with four layers (referred to as Tracker Barrel Pixel Detector, TBPX), eight small double-discs per side (referred to as Tracker Forward Pixel Detector, TFPX) and four large double-discs per side (referred to as Tracker Endcap Pixel Detector, TEPX). In the TBPX the pixel modules are arranged in "ladders". In each layer, neighbouring ladders are mounted staggered in radius, so that r-ϕ overlap between ladders is achieved. The modules on a ladder do not overlap in z. A projective gap at $\eta = 0$ is avoided

by mounting an odd number of modules along z, and by splitting the barrel mechanics in z into slightly asymmetric half. In TFPX and TEPX the modules are arranged in concentric rings. Each double-disc is physically made of two discs, which facilitates to mount modules onto four planes, with overlaps in r as well as r-ϕ. Each disc is split into two halves, and these D-shaped structures are referred to as "dees". The TEPX will provide the required luminosity measurement capability by an appropriate implementation of the readout architecture. In total, the pixel detector will have an active surface of approximately 4.9 m^2.

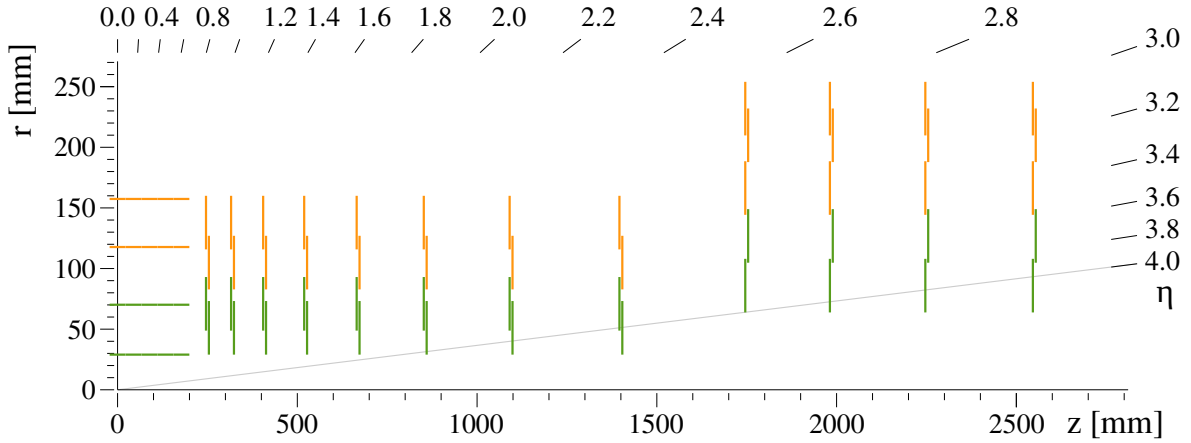


Figure 1.22: Sketch of one quarter of the pixel detector layout in the r-z view. Green lines correspond to modules made of two readout chips and orange lines represent larger modules with four chips.

Such a compact and complex detector, albeit large in terms of active surface, poses challenges for the implementation of the services. A total power of about 50 kW needs to be fed into the active volume and a serial powering approach has been identified to minimize the material of the cables carrying the current. The approximate 1 W/cm^2 of power dissipated by the ROCs will be removed by a network of low mass cooling pipes fed by the common CO_2 cooling system. Bidirectional data transfer is implemented using low mass electrical links to connect the front-end to Low-power Gigabit Transceivers (LpGBTs, the same as will be used in the Outer Tracker) located on the IT service cylinder, while the LpGBTs are connected via optical fibres with the back-end electronics in the service cavern. Power, cooling, and data transmission services are carried on a cylindrical shell enclosing the pixel detector. The detector is designed to be installable after the Outer Tracker and the beampipe are already in place, thus enabling the possibility to replace degraded parts over an Extended Technical Stop.

Figure 1.23 shows one of the four Inner Tracker structures, as it will look at installation time. The TBPX layers and the TFPX and TEPX dees are visible. The barrel, forward, and endcap elements are supported by half-cylinders that also hold their corresponding services, appropriately routed on the outer surface of the structures.

1.2.5 Physics Object Identification and Reconstruction

The raw detector information is combined and used to reconstruct "physics objects", that constitute the input of all the data analyses. A global event reconstruction is performed to identify elementary objects: charged and neutral hadrons, electrons, photons, and

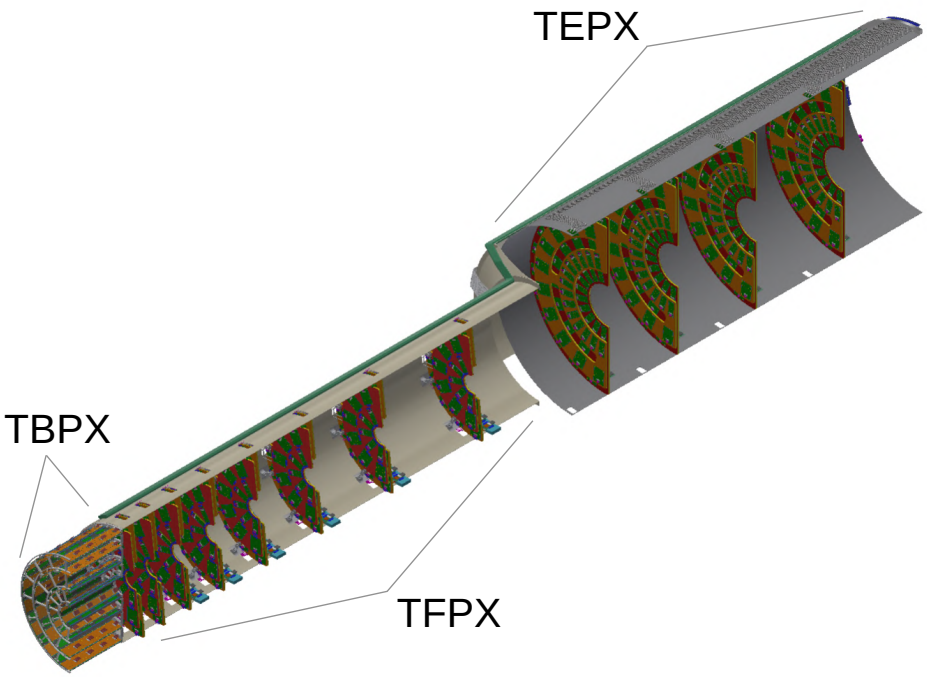


Figure 1.23: Perspective view of one quarter of the Inner Tracker, showing the TBPX ladders and TFPX and TEPX dees inside the supporting structures. The pixel modules are shown as orange elements in TBPX and as green elements in TFPX and TBPX. The dees are depicted as red and orange surfaces.

muons. These are subsequently combined to reconstruct more complex objects such as τ leptons, jets and missing transverse momentum.

Global Event Reconstruction

The particle flow (PF) algorithm [18] is designed to exploit the redundant measurements from the CMS subsystems and reconstruct physics objects from raw detector data, performing a global event reconstruction.

The philosophy of this approach is illustrated in Fig. 1.24 , where the typical signatures of different particles in the CMS detector are compared. The trajectory of charged particles, or tracks, are reconstructed from the hits in the tracker systems, and matched to deposits in the ECAL only (electrons) or in the HCAL as well (charged hadrons). The absence of a track in front of a calorimetric deposit reveals the passage of a photon or a neutral hadron. Finally, the presence of a track in the muon systems identifies the interaction of a muon.

This simplified picture is complicated by earlier interactions of particles within the tracker, representing up to 2 radiation lengths of material (at $\mid \eta \mid \approx 4$) in front of the calorimeters. This results in a probability of photon conversion or of bremsstrahlung emission from electron of about 85%, and a probability of hadron nuclear interaction of about 20%. Similarly, muon can undergo multiple scattering before reaching the muon detectors, with a subsequent degradation of the momentum resolution. To overcome these problems, advanced specific algorithms have been developed to reconstruct the key elements of the PF algorithm, namely tracks and energy clusters.

Tracks must be reconstructed with an efficiency as close to 100% as possible. This is

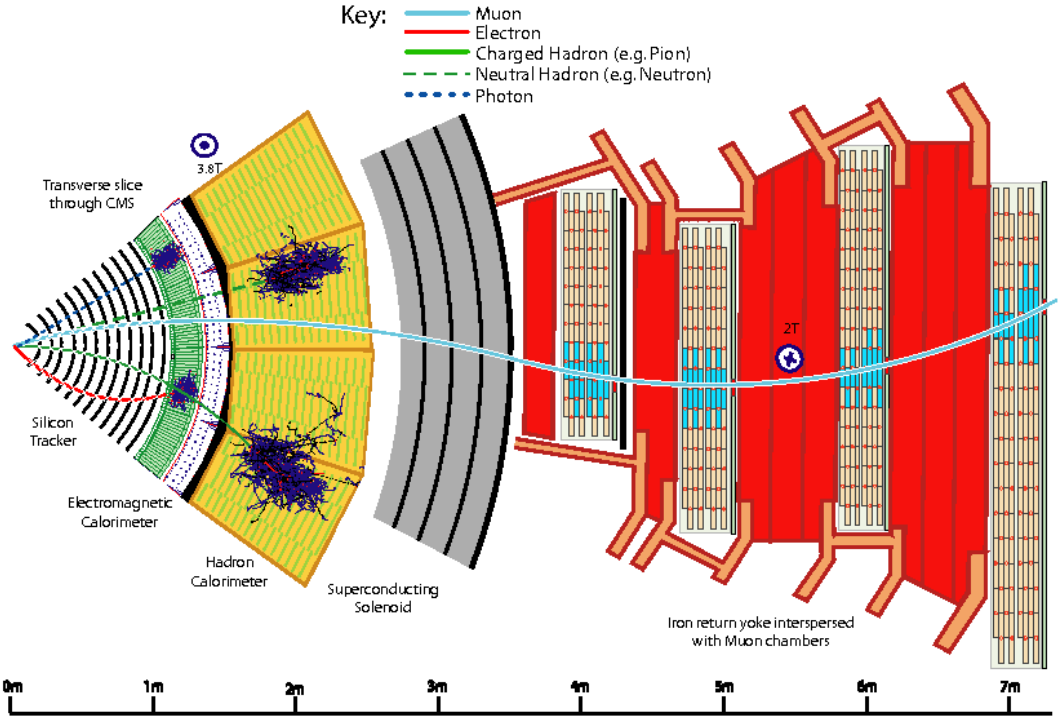


Figure 1.24: Illustration of the experimental signature of final state particles in the CMS detector.

especially important when they originate from charged hadrons in a jet, as the complementary measurement based solely on the calorimeters is not fully efficient and suffers from a direction bias and energy degradation. At the same time, a low rate of erroneously reconstructed tracks from random hit association must be achieved to ensure a good description of the event. An iterative tracking procedure [19] fulfils these requirements by initially applying strict quality criteria on track seeding and reconstruction, removing hits unambiguously assigned, and progressively loosening the quality criteria to increase the efficiency.

The energy deposits in the CMS calorimeters are grouped together with the PF clustering algorithm. The clustering algorithm is operated separately in the preshower, ECAL, and HCAL subdetectors. It identifies local maxima of energy, or "seeds", regroups the neighbouring energy deposits that satisfy topological and energy criteria, and selects the PF clusters among the groups of deposits.

The individual PF elements can be associated, or "linked", to create "PF blocks". The PF identification proceeds by analysing the structure and properties of these blocks to identify the final physics objects by linking one or more tracks to one or more clusters.

Once all tracks have been assigned to a candidate cluster, the remaining clusters are considered as photons in case of ECAL deposits and as neutral hadrons in case of HCAL deposits. After the association of all the PF elements, the redundant information from the CMS subsystems is combined to estimate their momenta. The total calorimetric energy is the linear sum of the calibrated ECAL and HCAL energies. In case this is not found in agreement, within the expected uncertainties, with the associated track momentum, an overlap between a charged and neutral candidate is assumed and the energy of the latter is estimated as the difference of the two measurements. A more detailed description of

the linking and reconstruction of PF candidates can be found in [18].

Muon Reconstruction

Muons leave a very clean signature in the CMS detector thanks to their interactions in the muon spectrometers. As a consequence, muon tracks are reconstructed with dedicated algorithms that are independent from the iterative PF tracking discussed above, and are based on a Kalman filter method that accounts for the muon energy loss in the detector materials. Three muon reconstruction algorithms are defined and exploit differently the subdetectors information [10]:

- **Standalone muons** are built by exploiting information from muon subdetectors to gather all CSC, DT, and RPC information along a muon trajectory using a Kalman-filter technique. Reconstruction starts from seeds made up of groups of DT or CSC segments.

- **Tracker muons** are built "inside-out" by propagating tracker tracks to the muon system with loose matching to DT or CSC segments. Each tracker track with transverse momentum p_T ¿ 0.5 GeV and a total momentum p ¿ 2.5 GeV is extrapolated to the muon system. If at least one muon segment matches the extrapolated track, the tracker track qualifies as a tracker muon track. The track-to-segment matching is performed in a local (x,y) coordinate system defined in a plane transverse to the beam axis, where x is the better-measured coordinate (in the R-ϕ plane) and y is the coordinate orthogonal to it. The extrapolated track and the segment are matched either if the absolute value of the difference between their positions in the x coordinate is smaller than 3 cm, or if the ratio of this distance to its uncertainty (pull) is smaller than 4.

- **Global muons** are built "outside-in" by matching standalone-muon tracks with tracker tracks. The matching is done by comparing parameters of the two tracks propagated onto a common surface. A combined fit is performed with the Kalman filter using information from both the tracker track and standalone-muon track.

Given the high reconstruction efficiency in both tracker and muon systems, about 99% of muons are reconstructed either as tracker and global muons, and those candidates that share the same inner tracks are merged into a single object. As standalone-only muons have a worse momentum resolution and show a high contamination from cosmic rays background, they are rarely used for physics analyses. Muon charge and momentum assignments are computed solely from the tracker measurement for muons of $p_T < 200$ GeV, as multiple scattering effects degrade the measurement of the muon detectors. The global track curvature is instead used for muon with $p_t > 200$ GeV, if the charge-to-momentum ratio agrees within two standard deviations from the tracker only measurement. The muon transverse momentum resolution thus achieved ranges between 1 to 6%, depending on the η coordinate, for muons with $p_t < 100$ GeV , and is better than 10% for central muons of $p_T = 1$ TeV.

Electron Reconstruction

Electron reconstruction is complicated by their interaction in the tracker material before they reach the ECAL. Tracker algorithms must take into account the non-Gaussian energy

loss while clustering algorithms must collect the bremsstrahlung photon energy deposits that can be located away from the electron interaction point in the ECAL. The electron reconstruction algorithm [20] addresses these two effects with a dedicated tracking and an advanced energy clustering.

The clustering algorithm regroups PF ECAL clusters in "superclusters". This procedure identifies a seed cluster and gathers together the energy deposits associated to bremsstrahlung photons. Preshower energy clusters in the endcaps are also taken into account. The supercluster aggregation depends on the cluster E_T and exploits the correlation between their η and ϕ positions, preferring clusters that are spread along the ϕ direction because of the magnetic field.

Tracks are reconstructed with a Gaussian sum filter (GSF) method. In contrast to the Kalman filter, the GSF method accounts for the large bremsstrahlung energy emissions by approximating the radiated energy loss probability with a sum of Gaussian distributions. The GSF tracking is initiated by two complementary algorithms. An ECAL seeding procedure makes an estimate for the track position starting from the PF superclusters, while a tracker-seeding relies on the general charged particle iterative tracks and looks for a correspondence with a PF supercluster.

GSF tracks and PF superclusters are associated into an electron candidate if they satisfy some loose requirements on their qualities and matching. They are subsequently used to estimate the electron charge and its momentum, the latter being computed from a combination of GSF track curvature and supercluster total energy. To improve the resolution, electrons are classified in five categories depending on their quality and bremsstrahlung properties. The momentum resolution for electrons produced in Z boson decays ranges between 1.7 and 4.5% depending on the electron category and position in the detector.

Tau Reconstruction

The branching fraction of the decays of the τ lepton are reported in Fig. 1.25 .

		τ Decay Mode	Branching Fraction (%)
Leptonic		$\tau^{\pm} \to e^{\pm} + \overline{\nu}_e + \nu_\tau$	17.84 ± 0.04
		$\tau^{\pm} \to \mu^{\pm} + \overline{\nu}_\mu + \nu_\tau$	17.41 ± 0.04
Hadronic	One-prong	$\tau^{\pm} \to \pi^{\pm} + (\geq 0\ \pi^0) + \nu_\tau$	49.46 ± 0.10
		$\tau^{\pm} \to \pi^{\pm} + \nu_\tau$	10.83 ± 0.06
		$\tau^{\pm} \to \rho^{\pm}(\to \pi^{\pm} + \pi^0) + \nu_\tau$	25.52 ± 0.09
		$\tau^{\pm} \to a_1(\to \pi^{\pm} + 2\pi^0) + \nu_\tau$	9.30 ± 0.11
		$\tau^{\pm} \to \pi^{\pm} + 3\pi^0 + \nu_\tau$	1.05 ± 0.07
		$\tau^{\pm} \to h^{\pm} + 4\pi^0 + \nu_\tau$	0.11 ± 0.04
Hadronic	Three-prong	$\tau^{\pm} \to \pi^{\pm} + \pi^{\mp} + \pi^{\pm} + (\geq 0\pi^0) + \nu_\tau$	14.57 ± 0.07
		$\tau^{\pm} \to \pi^{\pm} + \pi^{\mp} + \pi^{\pm} + \nu_\tau$	8.99 ± 0.06
		$\tau^{\pm} \to \pi^{\pm} + \pi^{\mp} + \pi^{\pm} + \pi^0 + \nu_\tau$	2.70 ± 0.08

Figure 1.25: Branching fraction of the decays of a τ lepton.

Fully leptonic decays to a electron or a muon in association with neutrinos are reconstructed by the correspondent object algorithms detailed above. Semileptonic decays

to hadrons and a neutrino (indicated in this thesis with τ_h) result in small and collimated hadron jet that requires a specific reconstruction algorithm. The decay can occur through an intermediate $\rho(770)$ or $a_1(1260)$ resonance, and result in different multiplicities of charged and neutral hadrons, usually pions. Decays containing one and three charged hadrons are generally referred to as one and three prong decays, respectively.

The "Hadron Plus Strips" (HPS) algorithm [21] is used in CMS for the τ_h reconstruction. The HPS reconstruction is initiated by PF jets that are formed as detailed in the next paragraph. The algorithm analyses the PF candidates composing each jet to verify their compatibility with a τ_h object. The contribution from neutral pions in $\pi^0 \to \gamma\gamma$ can appear either directly as photon PF candidates, or as electron candidates clustered inside the jet, because of the large $\gamma \to e^+e^-$ conversion probability. Photon and electron PF candidates of $p_T > 0.5$ GeV are thus clustered into "strips" with an iterative procedure. Electron and photon candidates within a clustering region around the strip are added to the strip itself, which position is recomputed as a p_T-weighted average. The strip creation ends when no candidates are found within the clustering region. A dynamic strip reconstruction defines the $\Delta\eta$ and $\Delta\phi$ clustering window sizes as functions of the strip p_T itself, to ensure an optimal collection of the energy and minimize the impact of background.

The strips and the charged hadrons in the jet are combined together to reconstruct any of these decay modes:

- $h^\pm$, single charged hadron with no strips
- $h^\pm\pi^0$, single charged hadron with one strip
- $h^\pm\pi^0\pi^0$, single charged hadron with two strips
- $h^\pm h^\mp h^\pm$, three charged hadrons
- $h^\pm h^\mp h^\pm\pi^0$, three charged hadrons with one strip

Quality and invariant mass criteria are applied on top of all the valid decay mode hypotheses to verify their compatibility with a τ decay. The selections applied depend on the decay mode considered and on the e/γ candidates clustered into the strip. In general, it is verified that tracks originate from the same vertex for multi-prong decays, that the combined invariant mass is compatible with the one of the intermediate meson resonance, and that the total electric charge is ± 1. Decay mode hypotheses are also discarded if they include additional charged hadrons or strips outside of a signal cone centred on the τ_h momentum axis and with an aperture of $\Delta R = 3\,GeV/p_T$ (bounded to $0.05 < \Delta R < 0.1$). In case multiple decay mode hypotheses are satisfied, the one with the largest p_T is retained, resulting in an unique association of a τ_h candidate to a jet.

Jet Reconstruction

As quarks and gluons undergo an hadronization process, the estimation of their initial momentum requires the recollection and measurement of the hadronization products. Jets are thus reconstructed by clustering the PF candidates with the anti-kT algorithm [22]. The algorithm iteratively combines PF candidates that are close to each other according to a metric, that is defined to produce jets of an approximate conic shape clustered around the hardest particles in the event. The size of the jet cone is determined by the distance

parameter R at which the algorithm is operated. Both the values R = 0.4 and R = 0.8 are used in CMS. The anti-kT algorithm is resilient against infrared and collinear effects, i.e. it is not affected by soft radiation or collinear parton splitting.

The jet four momentum is computed as the vector sum of the clustered PF candidates four momenta, and a set of corrections are applied to calibrate the jet response using the information of generated particles in a simulation. These corrections of the jet energy scale take into account the contribution from pileup in the event, non-linearities in the detector response to hadrons, and residual differences between the data and the simulation used for the method. They are validated using dijet, multijet, γ+jets and leptonic Z+jets events [23]. Typical jets resolutions achieved are of about 15–20% for at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV.

Missing Transverse Momentum Reconstruction

The presence of undetected final state particles such as neutrinos can be indirectly inferred from the imbalance of the total transverse momentum vector sum. The negative projection of this vector onto the transverse plane is denoted as missing transverse momentum $\vec{p}_T^{miss}$. This value is usually referred to as "MET", an acronym that means "Missing E_T".

The $\vec{p}_T^{miss}$ vector is reconstructed with the PF algorithm [24] as the negative vectorial sum of the transverse momenta of the PF candidates reconstructed in the event. As inefficiencies of the tracking algorithm and nonlinearities of the energy response of the calorimeters for hadronic particles can introduce a bias in the $\vec{p}_T^{miss}$ determination, a correction is applied by propagating to the $\vec{p}_T^{miss}$ sum the jet energy corrections introduced. In particular, the corrected $\vec{p}_T^{miss}$ miss vector is estimate as:

$$\vec{p}_T^{miss,corr} = \vec{p}_T^{miss} - \sum_{jets} (\vec{p}_T^{corr} - \vec{p}_T) \tag{1.5}$$

i.e. taking into account the difference between the initial jet $\vec{p}_T$ and its corrected value $\vec{p}_T^{corr}$.

1.2.6 Trigger System

Proton-proton collisions occur in the centre of the CMS detector every 25 ns, and generate a huge amount of information in the detector, corresponding to about 70 terabytes of data every second. No technology exists nowadays to readout, store and analyse such volumes of data. However, the large majority of the collisions result in low-energy proton-proton interactions that are not interesting to pursue the physics programme of CMS. The role of the trigger system of the CMS experiment is to identify and select the interesting collision events, thus reducing the acquisition rate by a factor of about 10^5. The trigger is at the interface between the "online" data taking and the "offline" data analysis, and must at the same satisfy the technical constraints of the former and ensure a high efficiency for the latter.

The CMS experiment adopted a two-tiered approach in the development of its trigger system. The Level 1 (L1) trigger is composed of custom hardware that processes the information from calorimeters and muon systems only, with reduced granularity. It has a fixed latency (i.e. the time available for data processing) of 3.2 μs, in which the event accept decision is made and the event rate is reduced down to about 100 kHz.

Following this first selection, the high level trigger (HLT) can access the complete detector information at the full granularity to perform an event reconstruction that is similar to the one performed offline. The HLT is implemented in a farm of commercial processors, where sophisticated algorithms running on its 22 000 CPU cores produce a decision in an average time of about 220 μs and further reduce the trigger rate below 1 kHz. The events thus selected are recorded on the tapes of the CERN Tier 0 and become available for subsequent offline analysis. Every algorithm used by L1 Trigger or HLT is assigned an adjustable factor f , or "prescale", that reduces the trigger rate of 1/f by retaining only one accept decision every f occurrences. The set of prescale values is changed during the data taking runs as a function of the instantaneous luminosity $\mathcal{L}$ to maintain a constant trigger rate when $\mathcal{L}$ is reduced and consequently maximize the signal acceptance. Events that satisfy all the requirements of one HLT "paths" are directed to a corresponding data stream for their storage on tape. These streams include events for physics analyses and detector calibration, alignment and monitoring, and differ by the amount of detector information stored.

Object reconstruction in the L1 trigger is performed separately using the inputs from the calorimeter and the muon subdetectors. The former are organised into trigger towers (TT), calorimeter readout units that are combined into objects representing jets, electrons, photons, and τ_h, and used to compute energy sums. As no information from the tracking detector is available, electron and photons result in a similar experimental signature and are both reconstructed as an e/γ object. Similarly, hits in the DT, CSC, and RPC subdetectors are combined to reconstruct muon tracks.

The HLT implements an online object reconstruction and selection that is a streamlined version of the offline reconstruction algorithms. HLT object reconstruction is usually performed only locally around the L1 seed objects, reducing the time needed to read the raw detector information and to process it. Selections on variables that discriminate the signal of interest from the background are applied as early as possible to optimize the processing time, and priority is given to the least time-consuming algorithmic steps. With these optimizations, HLT reconstruction follows the PF approach of reconstructing PF candidates with simplified clustering and track reconstruction algorithms. Jets are formed by clustering together these PF candidates with the anti-kT algorithm. The presence of secondary displaced vertices inside the jet is used to determine whether the jet is compatible with the hadronization of a b quark. Muons are initially built from patterns of CSC and DT segments, subsequently combined to inner tracks locally reconstructed and globally fitted into a muon track. Isolation criteria based on tracks around the muon candidates and calorimetric information are used to reduce the trigger rate. Electron reconstruction closely follows the offline algorithm detailed in this section and makes use of ECAL superclusters locally reconstructed around L1 e/γ seed and matched to inner tracks reconstructed with a GSF tracking algorithm adapted to HLT timing constraints. Pileup-resilient isolation criteria, based on the reconstructed PF candidates, can be applied to reduce the trigger rate. The reconstruction of τ_h objects at HLT is also similar to the HPS algorithm detailed here. The HLT algorithm considers up to 3 charged PF candidates clustered inside the PF jet and builds e/γ strips. Timing constraints do not presently allow for evaluating the τ_h decay mode from all the possible combinations of charged tracks and strips as done in the HPS algorithm. Consequently, HLT τ_h reconstruction has a larger efficiency with respect to the offline algorithm but a background contamination of about one order of magnitude larger.

Chapter 2

Solid State Silicon Detectors

Solid state detectors based on semiconductor materials represent the state of the art of tracking devices in high energy physics experiments. With respect to gas devices, which are less expensive and therefore employed in the outer parts of the detectors at LHC, solid state detectors represent a very compact solution which guarantees higher granularity and radiation hardness with faster timing. They are therefore particularly suited to be operated close to the interaction points where the high particle multiplicity per event and the integrated radiation dose demands high tracking performance and radiation tolerance.

In this chapter the properties of semiconductor detectors and in particular those of silicon pixel detectors are described. Section 2.1 presents the general characteristics of semiconductors and the way they are employed for particle detection. In Section 2.2, the effects of irradiation on silicon sensors is discussed. Finally, Section 2.3 focuses on pixel detectors and their design and production for high energy physics applications.

2.1 General properties

The periodic structure of the crystalline lattice in solid materials defines discrete energy levels in which electrons are confined. The highest energetic level which is fully filled with the outer-shell electrons constituting the covalent bondings of the atoms within the crystal, is called valence band, while the next higher energetic level, which may be empty or partially filled, is called conduction band. Depending on the material, these bands may be overlapped or separated by forbidden energy gaps. In the former case the electrons are free to move in the crystalline lattice and the solid is called conductor. In the latter case, instead, the electrons need to acquire enough energy to cross the forbidden gap and reach the conduction band where they can migrate throughout the crystal. Such a material is classified as insulator or semiconductor, depending on the width of the gap between the valence and the conduction band. The probability for an electron to occupy an energy level E at thermodynamic equilibrium is given by the Fermi-Dirac distribution:

$$f_e(E) = \frac{1}{1 + \exp\frac{E - E_f}{k_B T}} \tag{2.1}$$

where k_B is the Boltzmann constant, T is the absolute temperature and E_f is the Fermi energy, which in an intrinsic material, i.e. an ideal pure material without impurities, lies between the energies of the valence (E_V) and conduction (E_C) bands, in the middle of the band gap:

$$E_f \approx \frac{E_C + E_V}{2} \tag{2.2}$$

Semiconductors are usually defined by a band gap energy E_g of less than 3 eV which allows electrons to have a non negligible probability of occupying a state in the conduction band at room temperature. Electrons lifted from the valence band to the conduction band are free to migrate throughout the crystal. A corresponding number of vacancies (or holes) is generated in the valence band: this missing electrons represent a net positive charge that can move in the lattice as well. The resulting resistivity of the material is determined by the concentration of electrons in the conduction band (n) and holes in the valence band (p), and by their mobilities (μ_e and μ_h) as:

$$\rho = \frac{1}{e\,(n\mu_e + p\mu_h)} \tag{2.3}$$

Since the detector technology discussed in this thesis is based on silicon, the following sections are mainly focused on the treatment of this material. Silicon is the most common semiconductor material used for charged particle detection since it is available with high purity and offers a good energy resolution with reduced noise at moderate temperatures. It is characterized by a crystalline lattice with a diamond structure and an indirect band gap of E_g(300 K)=1.12 eV. Commonly available silicon has a non negligible amount of impurities in the crystal which compromises its use as particle detector as it is. A bare silicon wafer can be doped artificially, to change its properties, introducing additional impurities which increase either n or p. In the first case the dopants, called donor, introduce an energy level in the band gap that is close to the conduction band while in the second case they are called acceptors and they introduce an energy level close to the valence band. The most commonly used dopants are boron and phosphorous. Boron atoms has one bonding electron less than silicon atoms so boron doping results in a so-called p-type silicon wafer. Phosphorous atoms provide one additional bonding electron with respect to silicon atoms: the doped wafer will be called n-type. In doped silicon, the energy required to move an electron from the valence band to the new acceptor level or from the new donor level to the conduction band is much smaller than E_g and a high current is easily induced by thermal excitation.

2.1.1 The pn-junction

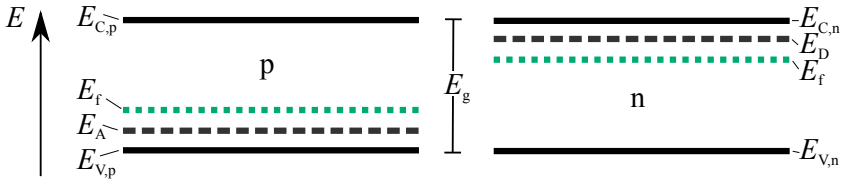
For particle detection a junction of p-type and n-type silicon is used. This is obtained from a single crystal doped with donors and acceptors on two different sides. At the junction of the two regions, due to their different concentration, electrons and holes diffuse from the zones of high concentration to those of low concentration where they recombine with the opposite charged carriers and create a so-called depleted region. As an effect of the recombination a space charge is produced in each side of the depleted region, whose sign is opposed to the one of the free charge. This fixed charges generate an electric field that oppose to the diffusion movements up to the equilibrium. The potential across the junction can be found solving the Poisson equation:

$$\nabla^2\phi = -\frac{\rho}{\epsilon} \tag{2.4}$$

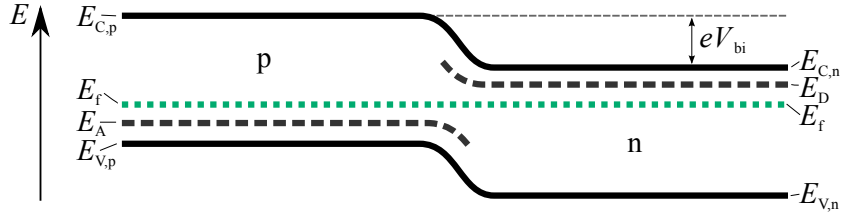
where ϵ is the dielectric constant of the medium and ρ is the space charge density. At the junction between the two doping concentrations, the band structure is therefore

modified as shown in Fig. 2.1. In this configuration the depleted region exhibits a very high resistivity compared to the doped areas and can be used for detection of ionizing particles, but only with very poor performance, due to the small built-in voltage V_{bi} of about 0.7 V that is spontaneously created across the junction. Applying a reversed bias voltage V_b to the junction, i.e. a voltage polarized inversely with respect to the intrinsic electric field at the junction, further charge carriers are removed from the doped regions extending the width of the depleted region d_a according to:

$$d_a = \sqrt{\frac{2\epsilon(V_b + V_{bi})}{e}\left(\frac{1}{N_D} + \frac{1}{N_A}\right)} \tag{2.5}$$



(a) p-type and n-type silicon



(b) pn-junction

Figure 2.1: Energy levels in the band structure of p-type and n-type silicon before (a) and after (b) the formation of the pn-junction. The energies of the valence (E_V) and the conduction (E_C) bands are drawn together with the Fermi energy (E_f). The symbols E_D and E_A indicate the energy levels introduced by donor and acceptor dopings, respectively. The widths of the energy gap and the energy given by the built-in voltage are indicated by E_g and eV_{bi}, respectively.

where N_D and N_A are the dopant concentrations of donors and acceptors, respectively. Since the pn-junction is usually obtained with a predominant doping concentration and $V_b >> V_{bi}$, Eq. 2.5 can be simplified as:

$$d_a \approx \sqrt{\frac{2\epsilon V_b}{eN}} \tag{2.6}$$

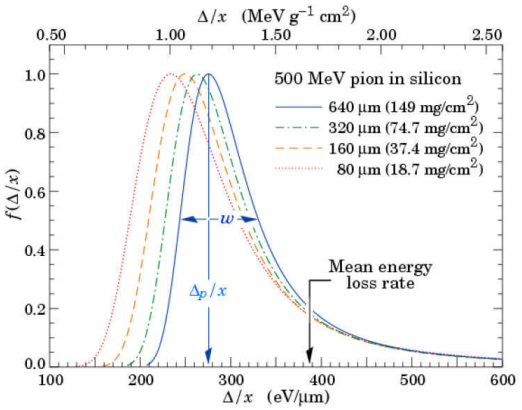
where N represent the dopant concentration of the side of the junction that has the lower dopant level. The minimum voltage required to extend the depleted region over the entire detector thickness d_b is called full depletion voltage V_{fd}. When $V_b > V_{fd}$ the electric field in the depleted region increases by $\frac{V_b - V_{fd}}{d_b}$ and the detector is said to be over-depleted. Since the depleted region is the area where the charge produced by the transit of a ionizing particle can be detected, the maximum signal for a particle traversing the sensor through its thickness is obtained when the full depletion is reached.

2.1.2 Charge generation

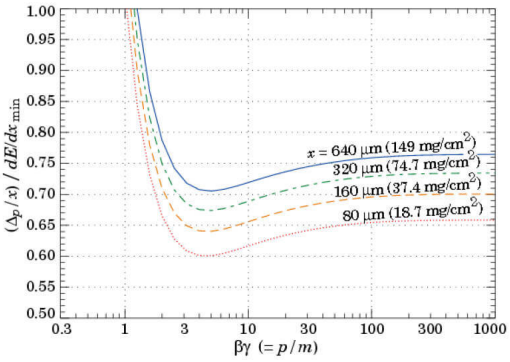
The average number of electron-hole pairs produced in the depleted region of a pn junction is proportional to the energy transferred to the lattice divided by the average energy necessary to create an electron-hole pair (3.65 eV at 300K in silicon). Charged particles interact in the depleted region creating electron-hole pairs along their entire path with a mean energy loss per length transferred to the material described by the Bethe-Bloch equation. Highly ionising particles, such as alpha from radioactive decays, are stopped in the material losing most of their energy at the end of their path. Particles that reach their minimum energy loss, so-called Minimum Ionising Particles (MIPs), can pass through the full active thickness of the silicon detector. In high energy physics experiments particles measured with tracking devices can be considered as MIPs. The energy loss of charged particles passing through detectors of moderate thickness is subject to large fluctuations and can be described by a Landau-Vavilov distribution. The Most Probable Value (MPV) of this distributions can be computed as:

$$\Delta_p = \xi \left[\log \frac{2m_e\beta^2\gamma^2}{I} + \log \frac{\xi}{I} + 0.200 - \beta^2 + \delta(\beta\gamma) \right] \tag{2.7}$$

where m_e is the electron mass, I is the mean excitation energy of the material, $\delta(\beta\gamma)$ is a density effect correction to the ionisation energy loss and ξ is proportional to $\frac{d_a}{\beta^2}$. The MPV of the Landau-Vavilov distribution therefore depends on the active thickness d_a and the particle energy as shown in Fig. 2.2. The long tail of the Landau-Vavilov distribution is given by rare but not negligible high energy transfer events in which a delta ray or a gamma ray is generated. The most probable energy loss is less sensitive to fluctuations and tails of the distribution, it is therefore used to characterize the charge collected by a detector.



(a) Landau-Vavilov distributions obtained for 500 MeV pions crossing different silicon thicknesses calculated from Eq. 2.7



(b) Most probable energy loss in different silicon thicknesses, scaled to the mean loss of a minimum ionizing particle (388 $eV/\mu m$), as a function of the incident particle energy.

Figure 2.2: Landau-Vavilov distribution dependence on incident particle energy and detector thickness

2.1.3 Signal formation

When a reverse bias voltage is applied to the pn-junction the electric field E in the detector increases and the electron-hole pairs created in the depleted region drift towards the positive and negative electrodes, respectively, with a velocity given by:

$$v_{e,h} = \mu_{e,h} E \tag{2.8}$$

The movement of the charge in the electric field induces a current i on the electrodes which is described by the Shockley-Ramo theorem [25, 26]:

$$i = e\vec{v} \cdot \vec{E_w} \tag{2.9}$$

Here E_w is the weighting field which is defined as the negative gradient of the weighting potential ϕ_w obtained solving the Laplace equation $\nabla^2 \phi_w = 0$ with specific boundary conditions. These imply a unit voltage on the electrode for which the charge is calculated and a 0 voltage for all the other electrodes. The resulting charge signal Q is obtained integrating this current over the charge collection time $t_c = [t_1; t_2]$ for a charge moving from x_1 to x_2:

$$Q = \int_{t_1}^{t_2} i(t)dt = e[\phi_w(\vec{x_1}) - \phi_w(\vec{x_2})] = -e\Delta\phi_w \tag{2.10}$$

In the most simple case of a pad detector, in which the electrodes dimensions are larger than the bulk thickness, the weighting potential is a linear function of the distance from the electrode z, as illustrated in Fig. 2.3.

$$\phi_w = V_w \left(1 - \frac{z}{d}\right) \tag{2.11}$$

where V_w is 1 for the electrode of interest and 0 for the other.

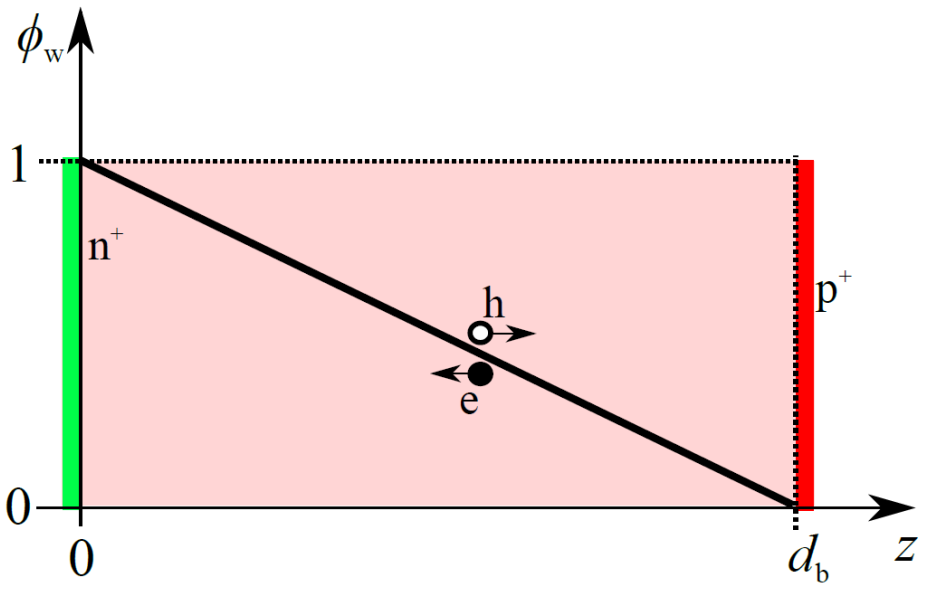


Figure 2.3: Weighting potential as a function of the interaction depth z for a pad detector of thickness d_b.

The total charge Q is given by the sum of the charges collected at the two electrodes:

$$Q = ne\left(1 - \left(1 - \frac{z}{d}\right)\right) - ne\left(0 - \left(1 - \frac{z}{d}\right)\right) = ne \tag{2.12}$$

where n is the number of electron-hole pairs generated. The first term in Eq. 2.12 is the contribution of electrons and the second is the contribution of holes, hence the signal induced by each charge carrier separately depends on the interaction depth, but the total signal depends only on n. As shown in Fig. 2.4, this linear approximation of the weighting potential does not hold if the dimensions of the electrodes become similar or even smaller than the thickness of the detector, as it is generally the case in pixel detectors. For these configurations the region further away from the electrode, where the weighting potential approaches zero, gets larger and, as a result, most of the signal is induced by the charges drifting towards the electrode in the last part of the their path. As opposite, the contribution of charges drifting away from the electrode becomes negligible. Nevertheless, when all charge carriers reach the electrodes, the integral of the induced current is always equal to the number of electron-hole pairs generated.

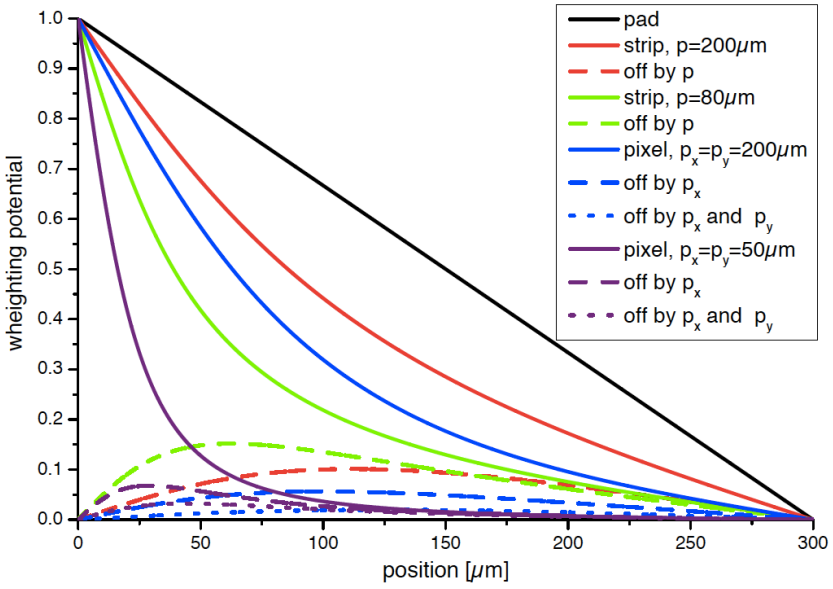


Figure 2.4: Weighting potential for different electrode dimensions as a function of the distance from the electrodes for a 300 μm thick sensor. The symbols p_x and p_y indicate the x and y pitches of the pixels and p indicates the pitch of the strips. The weighting potential is calculated in the center of the electrode (solid lines) and between the electrodes (dashed lines). Image taken from [27].

During their drift, electrons and holes are also subject to random movements which leads to a diffusion of the charge cloud. After a drift time t_d, the lateral diffusion of the minority carriers σ_D is given by:

$$\sigma_D = \sqrt{Dt_d} \tag{2.13}$$

where D is a diffusion constant which depends on the temperature T and on the charge carrier mobility as:

$$D = \mu_{e,h} \frac{k_B T}{e} \tag{2.14}$$

According to Eq.2.8, t_d is proportional to $1/\mu_{e,h}$ hence electrons and holes are subject to the same diffusion. For segmented semiconductors, due to charge diffusion, the induced signal may be spread over several electrodes. In a pixel detector this effect can be used to improve the spatial resolution but at the same time it may lead to inefficiencies if the electronics is not able to deal with the smaller signal induced in each pixel.

2.1.4 Leakage current

When the pn-junction is polarised a leakage current is also observed. This current has a surface and a bulk component. The latter is due to electron-hole pairs thermally generated in the depleted zone and is therefore directly proportional to the depleted volume V. Hence, given Eq. 2.6, the contribution of this component rises with $\sqrt{V_b}$ and saturates when V_{fd} is reached. In addition, the bulk current I_b has a strong dependence on the temperature T according to:

$$I_b \propto T^2 \exp\left(-\frac{E_g}{2k_B T}\right) \tag{2.15}$$

which could lead to an increase of the leakage current even after full depletion, due to the self heating of the sensor.

When the depleted region reaches the wafer surface an additional contribution to the leakage current emerges. This surface current is usually small compared to the bulk current and does not depend on V_b. If, increasing the bias voltage, the electric field at the pn-junction reaches critical values, the charge carriers can gain enough energy to create further electron-hole pairs leading to an avalanche multiplication that causes an exponential increase of the leakage current. This phenomenon is called breakdown and defines the operational voltage limit of a sensor. An early breakdown can be due the quality of the processing at the $Si - SiO_2$ interface and edge defects from the dicing inducing high electric field regions. The overall leakage current also contributes to the noise of the detector and can become a critical issue after the sensor irradiation.

2.2 Radiation damage

Particles penetrating the silicon sensor can also undergo non-ionizing energy loss through scattering with the atoms of the crystalline lattice. A damage can be produced by these interactions both in the silicon bulk and in the interface with the SiO_2 layer, deposited on the sensor surface for protection. The resulting defects created in the sensor are classified as bulk defects or surface defects, respectively.

Surface defects create a positive charge density on the sensor surface which attracts electrons leading to a compensation of the doping between the pixel implants. This has to be taken into account in the sensor design to avoid that damages in the oxide influence the detector performance, resulting in a decreased inter-pixel resistance.

Bulk defects are instead the main cause of performance deterioration for silicon detectors in high energy physics experiments. These are produced by highly energetic particles which interact with the nuclei of the silicon atoms. If the colliding particle transfers an energy larger than 25 eV to the nucleus this can be displaced from its original position

in the lattice creating a defect which can be charged and hence change the electrical properties of the material. As a result of this displacement a vacancy is left in the crystal lattice and the recoiling atom can either move to an interstitial lattice position or travel in the crystal displacing other atoms and creating further point-like defects. In case the transferred energy exceeds 2 keV, the atoms lose most of their energy in a localized area at the end of their path in the lattice, leading to so-called cluster defects. The formation of these complex defects produces new energy states located deeply in the band gap which act as generation and recombination centers [28].

2.2.1 The NIEL scaling hypothesis

Numerous observations have led to the result that damage effects by energetic particles in the bulk of any material can be described as being proportional to the so called displacement damage cross section D. This quantity is equivalent to the Non Ionizing Energy Loss (NIEL) and hence the proportionality between the NIEL-value and the resulting damage effects is referred to as the NIEL-scaling hypothesis [29]. D is normally quantified in $[MeVmb]$, whereas the NIEL-value is given in $\left[keVcm^2/g\right]$. The D or NIEL value is depending on the particle type and energy. According to an ASTM standard, the displacement damage cross section for 1 MeV neutrons is set as a normalizing value: $D_n(1MeV) = 95\,MeVmb$. On the basis of the NIEL scaling the damage efficiency of any particle with a given kinetic energy E can then be described by the hardness factor k, defined as

$$k_{particle}\,(E) = \frac{D_{particle}}{D_N\,(1MeV)} \tag{2.16}$$

2.2.2 Changes to silicon properties

Bulk defects in reversely biased silicon sensors lead to three main effects: change in doping concentration, leakage current increase and charge trapping.

Doping concentration Most of the bulk defects induced by irradiation behave as acceptors and can change the initial effective doping concentration $N_{eff,0} = N_D - N_A$ of the silicon. The resulting effective doping concentration N_{eff} can be expressed as a function of the fluence according to the following parameterization [28] :

$$N_{eff}(\Phi) = N_D \exp^{-c\Phi} - N_A - b\Phi \tag{2.17}$$

where c is the donor removal constant and b is the acceptor creation rate. While in p-type silicon the exponential term is negligible, since $N_A >> N_D$, in n-type silicon the acceptor-like defects can compensate the donor concentration up to the inversion of the bulk into an effective p-type. For n-in-n pixel sensors this inversion can be observed for fluences between 1 and 3 $\times 10^{13}\,n_{eq}/cm^2$ as shown in Fig. 2.5.

As a consequence of the growth of the acceptor concentration the full depletion voltage increases with the fluence and can be calculated substituting N with $N_{eff}(\Phi)$ in Eq. 2.6 as:

$$V_{fd}(\Phi) \approx \frac{e}{2\epsilon} N_{eff}(\Phi) d_b^2 \tag{2.18}$$

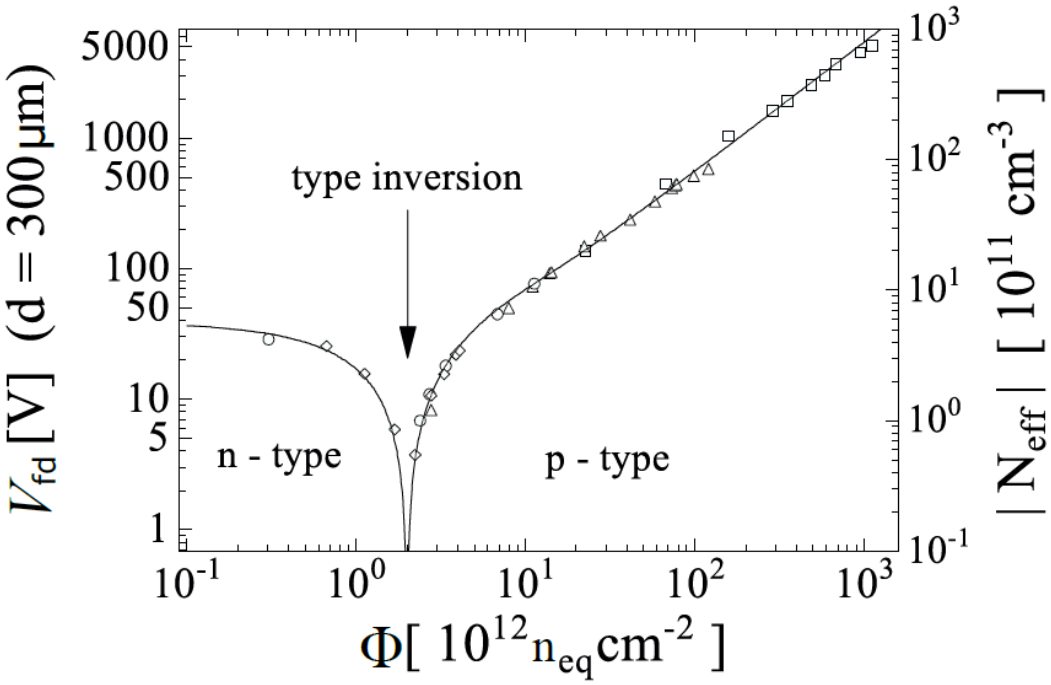


Figure 2.5: Effective doping concentration and full depletion voltage as a function of the irradiation fluence [28].

Leakage current increase Defects which create energy levels close to the middle of the band gap act as generation centers increasing the probability for electrons to be excited into the conduction band. This leads to an increase of the leakage current ΔI generated in the active volume of the semiconductor V which is proportional to the fluence:

$$\frac{\Delta I(\Phi)}{V} = \alpha\Phi \tag{2.19}$$

Here α is the current-related damage rate, which is independent of the initial resistivity of the silicon, the dopant concentrations and the type of irradiation, as can be seen in Fig. 2.6. As discussed above the leakage current in the sensor depends also strongly on the temperature. Hence, after high irradiation doses, the increase of the leakage current can lead to the heating of the sensor and a consequent thermal runaway that would destroy the junction. To avoid this effect irradiated detectors need to be properly cooled.

Trapping Charged defects in the silicon bulk constitute trapping centers which can absorb charge carriers during the collection time, thus reducing the signal in the detector. The collected charge Q becomes a function of the fluence and can be calculated by:

$$Q(\Phi) \approx Q_0 e^{-\frac{t_c}{\tau}} \tag{2.20}$$

where $\frac{1}{\tau} = \beta_T \Phi$, Q_0 is the charge collected before irradiation, t_c is the collection time and τ is the trapping time. The latter is a linear function of the fluence with a coefficient β_T that depends on the charge carrier type and also on the particle type (neutron or charge hadrons) used for the irradiation[30]. Even if the effective trapping times for electrons and holes are similar, the collection time for holes is almost three times as large due to their different mobility ($\mu_e = 1350\,cm^2/Vs$ and $\mu_h = 480\,cm^2/Vs$ at

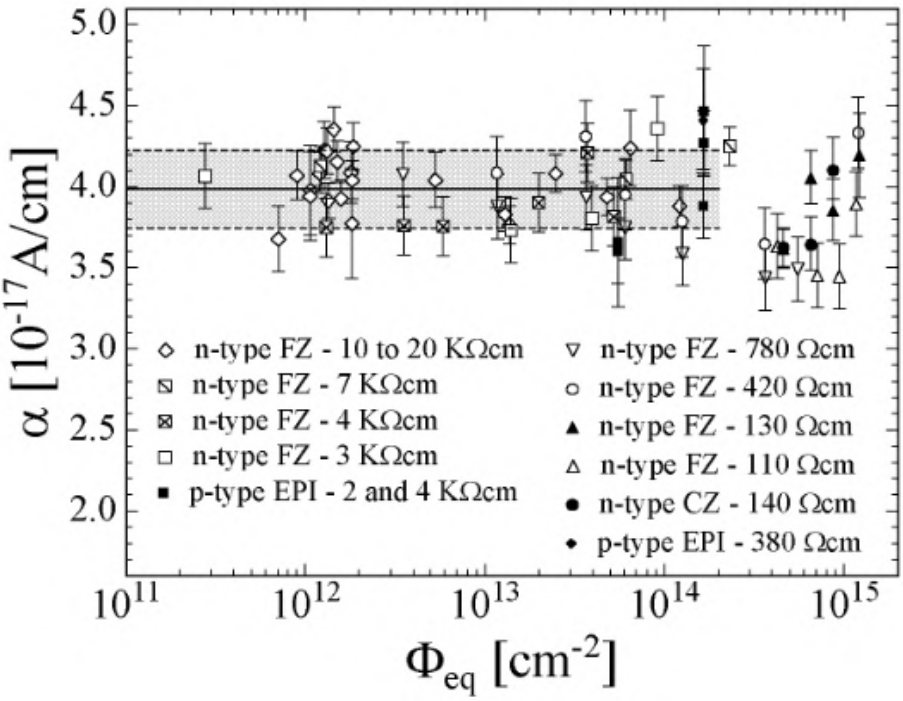


Figure 2.6: Measurement of the α parameter for different sensors wafers at different irradiation fluences [28].

300 K). Hence silicon pixel detectors employing electron collecting electrodes have better performance after high irradiation. Trapping becomes the dominant effect for fluences above 10^{15} n_{eq}/cm^2 and therefore is expected to be the main limiting factor for pixel sensors at the HL-LHC reducing the Charge Collection Effciency (CCE). The CCE is defined as the ratio of the collected charge Q to the collected charge Q_0 measured at full depletion before irradiation: CCE $= \frac{Q}{Q_0}$.

2.2.3 Annealing

The previously described point-like and cluster defects created in the silicon bulk by radiation are not static, they can move freely in the lattice reacting with other defects or impurities and recombining or forming new defect structures. The evolution in time of the defects, known as annealing, has a strong dependence on the temperature.

The effect of an annealing on the damage rate constant α for different temperatures, shown in Fig. 2.7, is always beneficial and leads to a consequent decrease of the leakage current.

On the other hand, the change of the effective doping concentration with the annealing time, illustrated in Fig. 2.8, is characterized by an initial short-term beneficial annealing, of the order of few days, and a subsequent long term reverse annealing. These two phases have a direct effect on V_{fd} and therefore on the charge collection for particle detection. In particular, for p-type sensors (or n-type sensors after that type inversion has occurred), V_{fd} decreases during the beneficial annealing and increases again during the reverse annealing. Due to this behaviour, irradiated sensors are usually intentionally kept at room temperature (or higher temperatures) for short and monitored periods of time, but they have to be maintained at temperatures lower than 0 °C both during storage

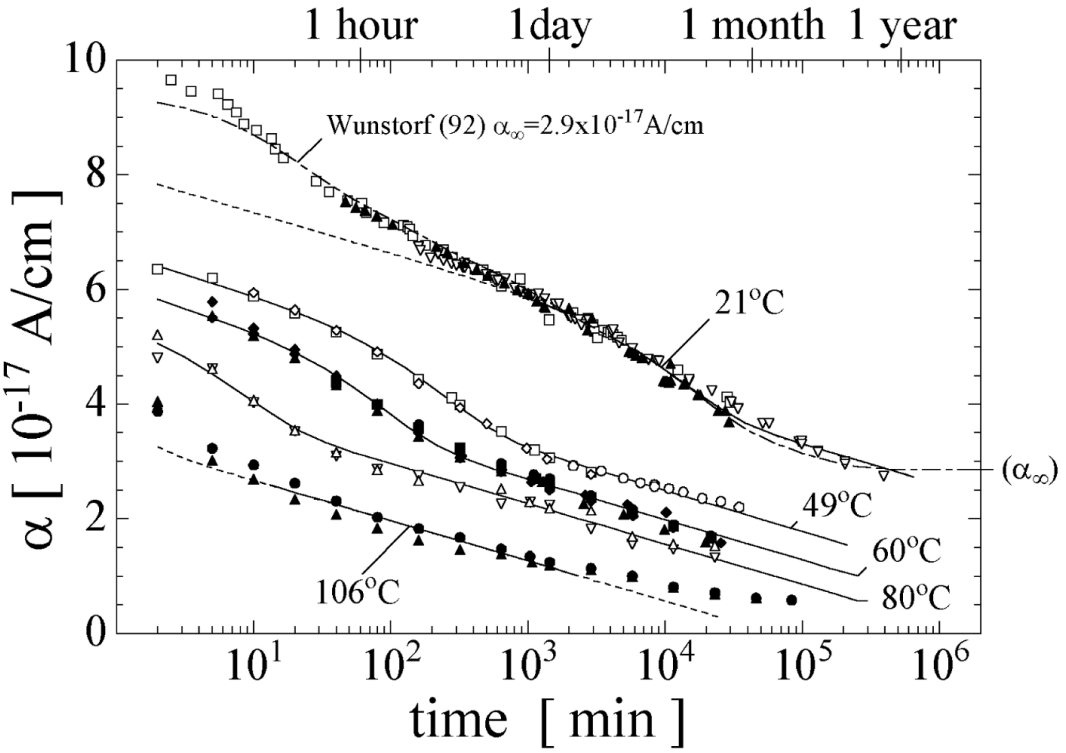


Figure 2.7: Evolution of the damage rate constant α at different annealing temperatures [28].

and operations to avoid a further degradation of their performance.

2.3 Pixel detector

To obtain an unambiguous two-dimensional information of the particle impact point, the electrodes of semiconductor sensors are segmented in several cell structures. If both dimensions of the cells are small (i.e. around 500 μm or less), these are usually called pixels and the sensor is called a pixel sensor. Planar pixel sensors are tipically produced from an initial lightly doped silicon bulk material with two subsequent heavily doping implantations (n+ and p+) on the two opposite sides, as illustrated in Fig. 2.9. First the polished surface of the silicon wafer is covered with a thin layer of SiO_2 and a photolithography mask is used to partially etch this oxide for the subsequent n+ implantation defining the pixel electrodes. The backside is then implanted with Boron to form an ohmic contact. The implanted ions undergo an annealing step at high temperatures, around 1000 °C, to be electrically activated. Successively, silicon nitride and LTO (Low Temperature SiO_2) are deposited on the front side and contacts are formed through these layers to the implants. An aluminium layer is deposited and patterned on the front side, and a passivation layer (usually polyimide or low temperature SiO_2) is applied to isolate and protect the sensor surface.

3D detectors were proposed by Parker and Kenney in 1997 to solve the problem of charge loss in gallium arsenide detectors [31]. In this new configuration the p+ and n+ electrodes are processed inside the silicon bulk, instead of being implanted on the

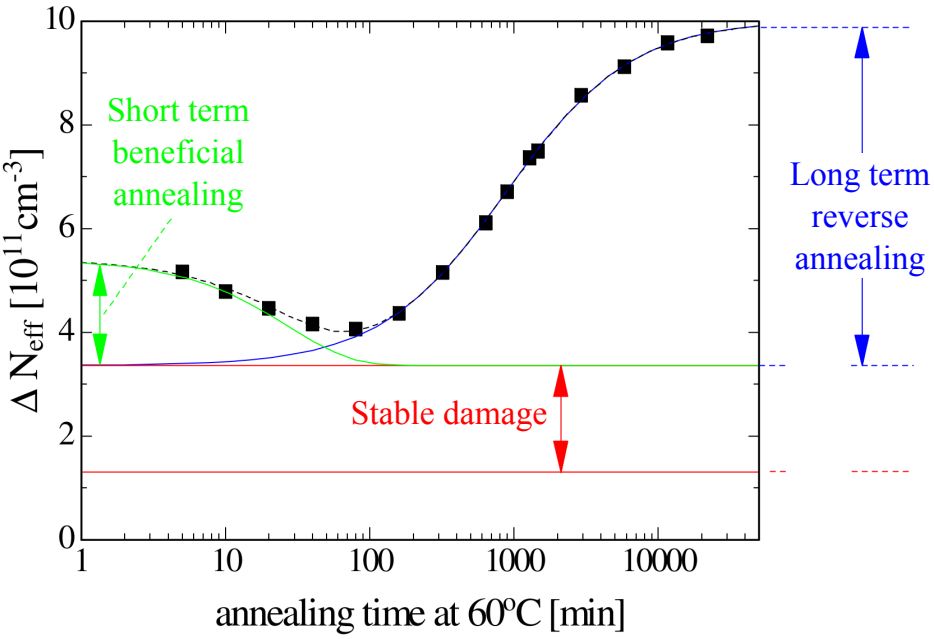
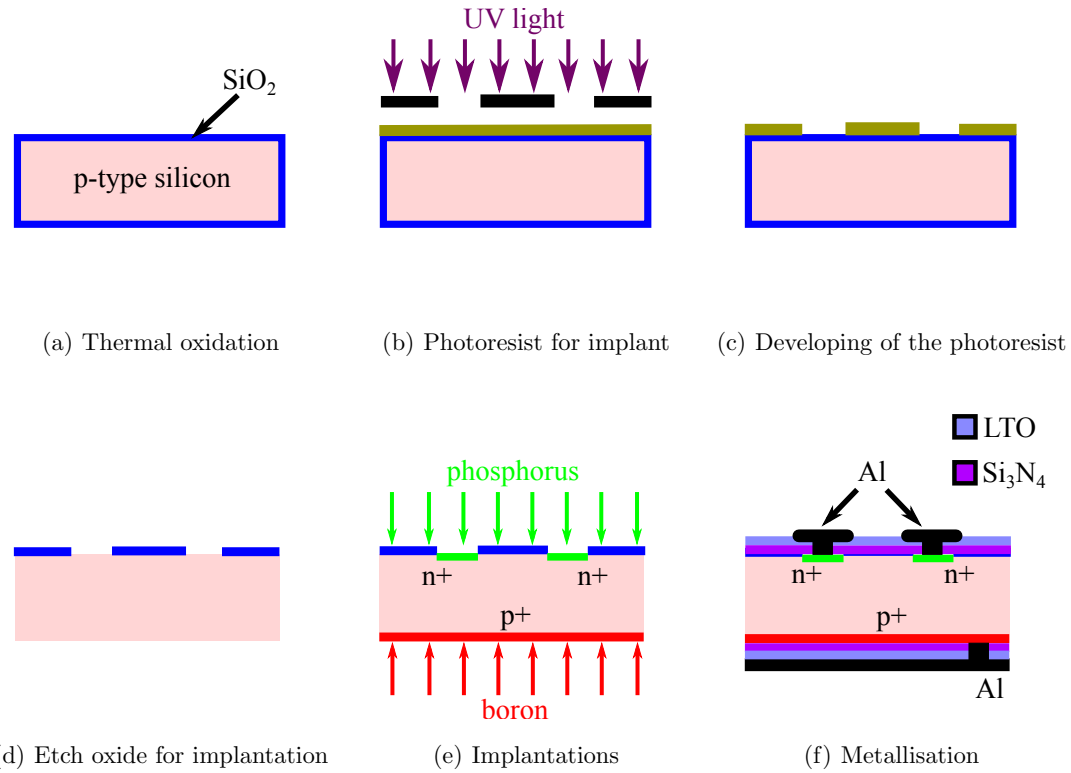


Figure 2.8: Effect of the annealing at 60 °C on the effective doping concentration. On top of a stable damage effect (red), the beneficial annealing is shown in green and the reverse annealing in blue.



(a) Thermal oxidation (b) Photoresist for implant (c) Developing of the photoresist

(d) Etch oxide for implantation (e) Implantations (f) Metallisation

Figure 2.9: Main production steps for a planar n-in-p silicon pixel sensor.

wafer surface. The advantages of 3D design, compared with the traditional planar one are shown schematically in Fig. 2.10. Since the electric field is parallel, rather than orthogonal, to the detector surface, the charge collection distance can be several times shorter, the collection time considerably faster and the voltage needed to extend the electric field across the junction an order of magnitude smaller, for the same amount of generated charge, typically 24 000 electron-hole pairs for a minimum ionising particle in 300 μm thick sensors. This technology has many potential applications such as extreme radiation environments, luminosity monitors and medical and biological imaging. The fabrication of 3D structures depends on the capability of etching narrow holes with high precision in the silicon bulk. One possible way is by using deep reactive ion etching (DRIE). DRIE has been developed for micro-electro-mechanical systems (MEMS). It allows to drill micro-holes in silicon with a thickness:diameter ratio as large as 20:1. These holes are filled afterwards with polysilicon doped either with boron or phosphorus which are then diffused into the silicon crystal in order to create the detector electrodes. Once the electrodes are filled, aluminium can be deposited either as a bump contact on the top of each electrode, which can then be readout alone or as a connecting microstrip to several electrodes of the same type. 3D detectors are currently employed in the CMS PPS detector [32] and in the ATLAS IBL [33] layer of the inner tracking detector.

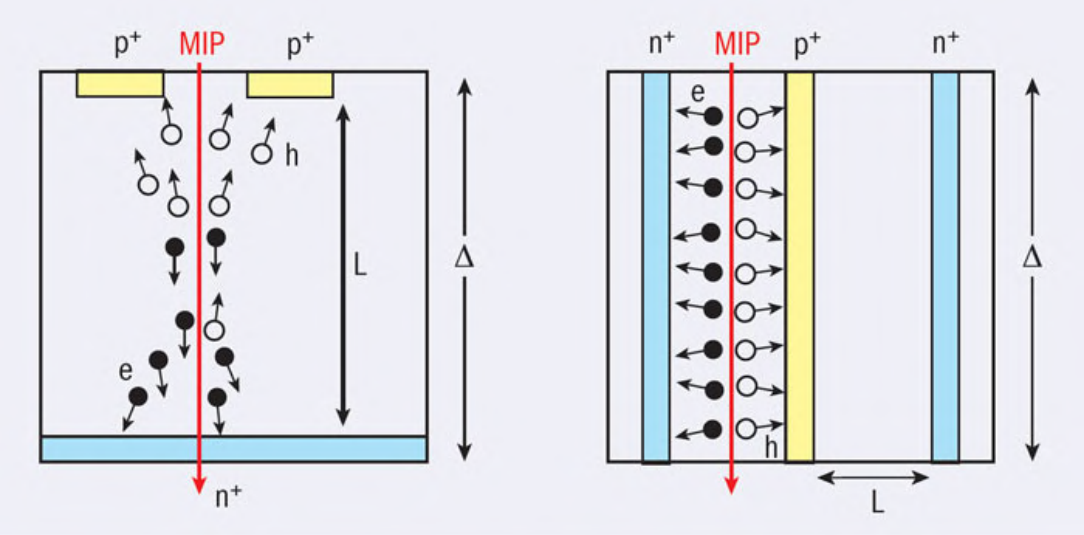
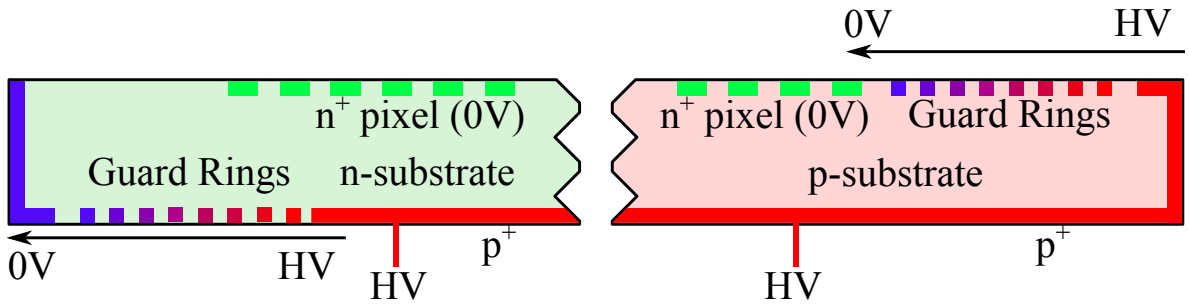


Figure 2.10: Schematic of charge collection in planar and 3D pixel sensors.

2.3.1 Sensor concepts

Four combinations of substrate and electrode dopings are possible, which can be suitable for different applications. For pixel sensors employing p+ pixel implants, the signal is mainly generated by the hole drift, due to the higher weighting field close to the electrodes. Since the mobility of holes in silicon is three times lower than for electrons, these pixel sensor designs are not suitable for high radiation fluences when trapping effects becomes significant. Hence, in high energy physics experiments, where radiation hardness is a fundamental requirement, the pixel sensor technology has to profit from the higher charge mobility of the electrons using n+ pixel implants. Two configurations, illustrated in Fig. 2.11, are thus possible: the so-called n-in-n technology, where n+ doped pixels are implanted in an n type silicon substrate, which is the sensor technology presently

used for planar pixel sensors in CMS; or the n-in-p technology where n+ doped pixels are implanted in a p type silicon substrate, which is the sensor technology that will be employed for the CMS Phase 2 upgrade.



(a) n-in-n (b) n-in-p

Figure 2.11: Comparison of the n-in-n (a) and n-in-p (b) planar sensor technologies.

The n-in-n sensors In this configuration the pn-junction is located on the backside. Thus, additional processing steps to pattern the backside are necessary to implement protecting structures around the p+ implantation. Before irradiation the depletion starts from the backside: pixel implants are shorted together by the conductive bulk until the depletion region reaches the front side. In these conditions, the detector cannot be operated partially depleted. Moreover at full depletion the electric field is lower closer to the segmented side where, instead, the weighting field is higher and the movement of the electrons induces most of the signal. However, since the bulk is n-type, after high radiation fluence this is subject to type inversion and therefore the junction moves to the pixel side and the sensor can be operated even strongly under-depleted.

The n-in-p sensors In n-in-p pixel sensors the pn-junction is located on the pixel implant side and therefore only one single-sided structured processing of the wafer is necessary. This makes this kind of technology cost effective and therefore particularly interesting when large areas, of the order of 1 m^2 or more, need to be covered. Furthermore the p-type bulk is not subject to type inversion and the depletion always starts from the pixel implant side.

2.4 Additional structures

The final pixel sensor is completed by a series of different structures which are described in the following.

Guard rings The cutting edge of the sensor is conductive due to the damage induced by the mechanical cutting procedure. If the depleted region extends up to this cutting edge, the high electric field causes a sudden increase of the leakage current that leads to the breakdown of the junction. Hence, the edges of the sensors have to be protected with dedicated structures called guard rings. These are additional ring implantations around the active area, at the edges of the sensor surface, implemented to ensure a smooth drop of the potential from the pn-junction to the cutting edge. The number and the shape of

the guard rings is crucial for the breakdown properties of the sensor. An example of a guard ring structure is reported in Fig. 2.12.

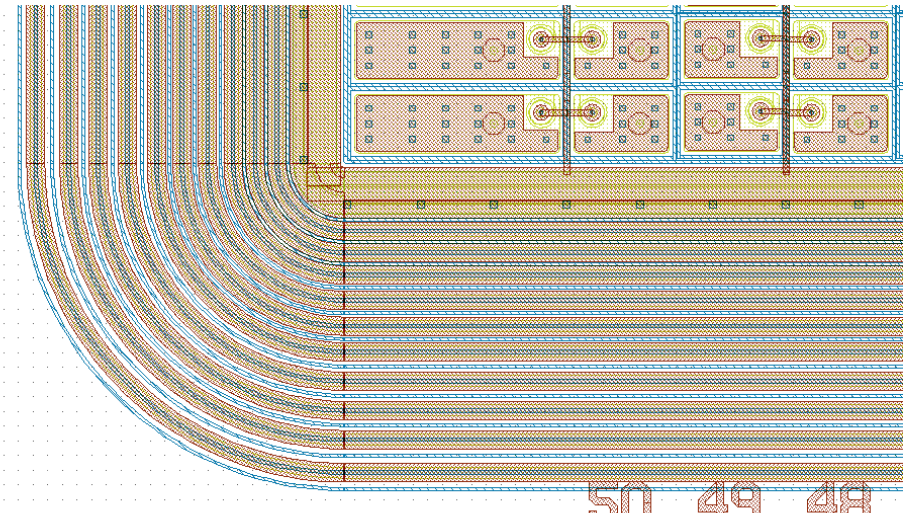


Figure 2.12: Example of guard rings implemented in one of the sensors described in this thesis.

While in n-in-n sensors the guard rings are placed on the backside and the entire edge is therefore kept at ground potential, in n-in-p sensors these are on the pixel implant surface and thus the edge is at the backside potential. In the latter case, for hybrid pixel detectors, the maximum bias voltage that can be applied to the sensor is limited by the risk of sparks between the sensor edge and chip surface that could damage permanently the detector. Possible remedies employ additional insulation layers between sensor and chip.

Biasing structures In experiments like CMS for which a large number of silicon modules is required, a high production yield is a fundamental requirement to limit the costs. To optimize the module production process and select the best sensors, dedicated biasing structures are implemented. These structures allow to measure the electrical properties of the sensors before interconnection to discard the problematic ones. For the sensors described in this thesis, the punch through mechanism is used. All pixels are connected through a common implant around the active area of the sensor, called bias ring, and a metal bias rail, as shown in Fig. 2.13. The connection between the bias rail and the pixel implant is realized with a separate circular implantation (bias dot) located within the pixel cell. When the depleted zone encloses both the grounded bias dot and the floating pixel implant, the potential of the latter follows that of the grounded one. This effect, called punch through, is used twofold: to ensure a common potential for possibly disconnected pixel cells and, at the same time, to ground all pixels with just one connection allowing the characterization of the full sensor before interconnection. Since neither the bias dot nor other bias structures are connected to the readout, the signal induced in their proximity is partially lost. To maximize the signal, the size of these structures as well as the distance of the bias dot to the pixel implant has to be reduced as much as possible within the limitations of the production process. Especially after irradiation, due to the decrease of the CCE, these structures can represent a critical area for the inefficiency of the detector.

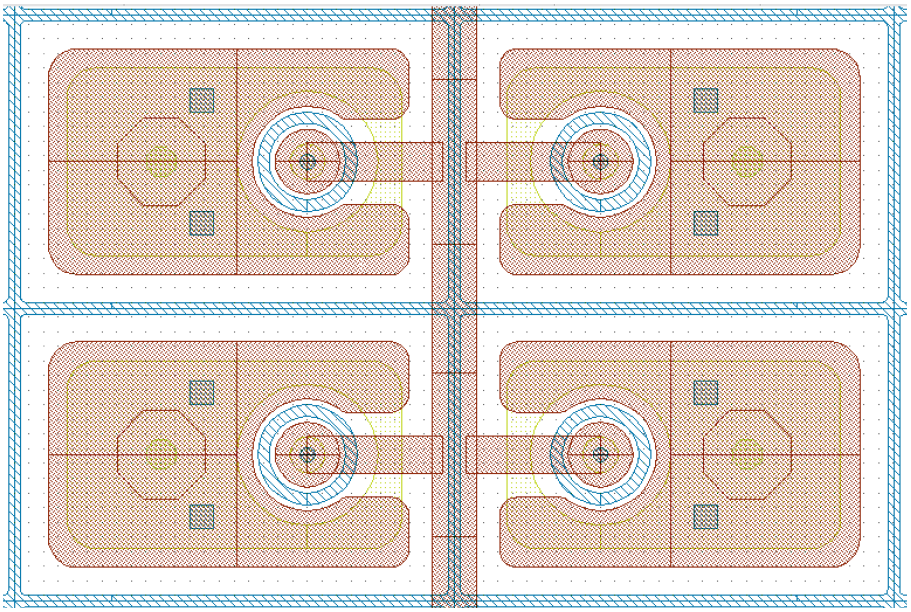
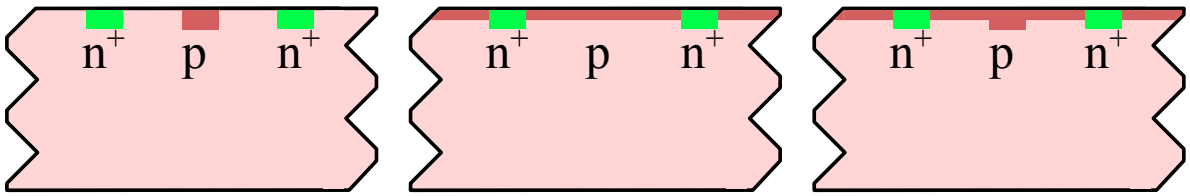


Figure 2.13: Example of biasing structures implemented in one of the sensors described in this thesis. The bias dot and rail are visible in the same red color of the pixel implant. The blue structures represents the p-stop pixel isolation.

Interpixel isolation The positive charges of the oxide attract electrons which accumulate between the pixel implants. When n+ implants are used for the segmented side of the sensor, this electron layer creates a short of all pixels. The three available solutions to this problem are shown in Fig.2.14.



(a) p-stop (b) Homogeneous p-spray (c) Moderated p-spray

Figure 2.14: Schematics of the inter-pixel isolation techniques for n-in-p sensors. The p-spray, represented in red in (b, c), is implanted on the entire sensor surface including the n+ implants. For the moderated p-spray approach shown in (c), a narrow region with a higher p-spray dose is created in the center between two pixel implants by etching of the silicon nitride layer.

The first consists of applying a further p+ implantation between the n+ pixels (p-stop). This technique requires an additional photolithographic step in the production with a critical alignment procedure needed to avoid the superposition of the two implants. Moreover, after irradiation, since the positive charges in the oxide increment, also the potential difference between p-stop and the pixel increases, resulting in a lower breakdown voltage. The inter-pixel p-stop is represented by blue lines if Fig. 2.12.

The second solution consists of a low dose of p-implantation, called p-spray isolation. The dose is chosen to be low to avoid high electric fields close to the n+ implants. This solution has several advantages with respect to the p-stop approach: it allows for lower

pixel dimensions and it does not require a second mask with the consequent alignment problems.

The third solution combines a moderated dose of p-spray with the p-stop approach. Here a region with a higher p-spray dose is created by etching the silicon nitride layer in a narrow line at the center between the pixel implants, thus allowing more boron ions to reach the silicon bulk.

Chapter 3

Fermilab Test Beam Facility

In this chapter a description of the Fermilab test beam facility, where most of the detectors object of this thesis have been tested, will be given. The software development of the facility received a strong contribution from the INFN section of Milano Bicocca and I personally contributed to this development. I also gave a contribution to the data analysis of the detectors tested at DESY in 2018 and 2019 so a brief description of the DESY test beam facility will be given at the end of the chapter.

3.1 Introduction

The Fermilab Test Beam Facility (FTBF) at the Fermi National Accelerator Laboratory provides beam in a multitude of particle types and a range of energies with which users can test their detectors. The beam is resonantly extracted in a slow spill for each Main Injector cycle delivering a single 4.2 second long spill per minute. The primary beam consists of high energy protons (120 GeV) at variable intensities between 1 and 300 kHz. This beam can also be targeted to create secondary particle beams of pions, muons or electrons with energies down to about 1 GeV. Users have access to the facility instrumentation to measure the position and energy of the incident beam. Four pre-installed scintillation counters give the rough beam position, a lead glass calorimeter measures the beam energy with a precision of $\approx 3\%$, two time of flight detectors can be set up for particle identification. A silicon pixel telescope can be used to measure the positions of the impact points of the beam particles with a precision of less than 8 μm. This telescope has been replaced in the first half of 2018 by a silicon strip telescope that provides a better position resolution (less than 5 μm). The telescopes will be described in this chapter together with the data acquisition (DAQ) system and alignment software.

3.2 The pixel telescope

A silicon pixel telescope has been built to provide precision tracking information to any test beam experiment at the FTBF in need of knowing the particle impact point on their detector under test (DUTs) with accurate precision. The telescope is placed along the FTBF beam line and consists of eight detector planes, each made of modules left over from the CMS Forward Pixel detector production, mounted on a carbon fiber frame.

Four of the eight telescope planes are composed of six (2×3) PSI46V2 Read Out Chips (ROCs) [6], while the remaining four planes are equipped with eight (2×4) ROCs.

Each ROC reads an array of 52×80 pixel cells, where each pixel cell has dimensions $100 \times 150 \times 285\,\mu m^3$ (except for the pixels at the edge columns and upper row which have dimensions $100 \times 300 \times 285\ \mu m^3$ and $200 \times 150 \times 285\ \mu m^3$, respectively), for a squared active area of $0.81 \times 0.81\ cm^2$. The active area of a 2×4 plane is then $1.62 \times 3.24\ cm^2$ while the 2×3 plane active area is $1.62 \times 2.43\,cm^2$. The eight planes are arranged in two stations, between the two lies a third station where DUTs can be connected.

Each plane has a carbon fiber layer with the modules glued on it. Heat is passively dissipated through the carbon fiber and hence no cooling is required. For each station, the planes are grouped into two pairs and each pair of planes is screwed together to a light aluminum mechanical support, then mounted on the mechanical structure of the telescope. The mechanical structure is of modular design: the frame consists of three basic cells, one for each station, built with carbon fiber tubes, with dimension $17.0 \times 17.0 \times 34.0\,cm^3$. The frame is covered by a Mylar anti-static layer which also serves the secondary purpose of keeping the detectors dark.

A three-dimensional schematic view of the telescope is presented in Fig. 3.1, where the laboratory coordinate system is also indicated. In this reference frame the Z axis is along the beam direction with +Z pointing downstream, the Y axis is perpendicular to the beam with +Y pointing upwards, and the X axis is the horizontal axis with +X given by the right-hand rule. The coordinate system origin is placed between the two stations so that each plane of the upstream (downstream) station has negative (positive) Z coordinate.

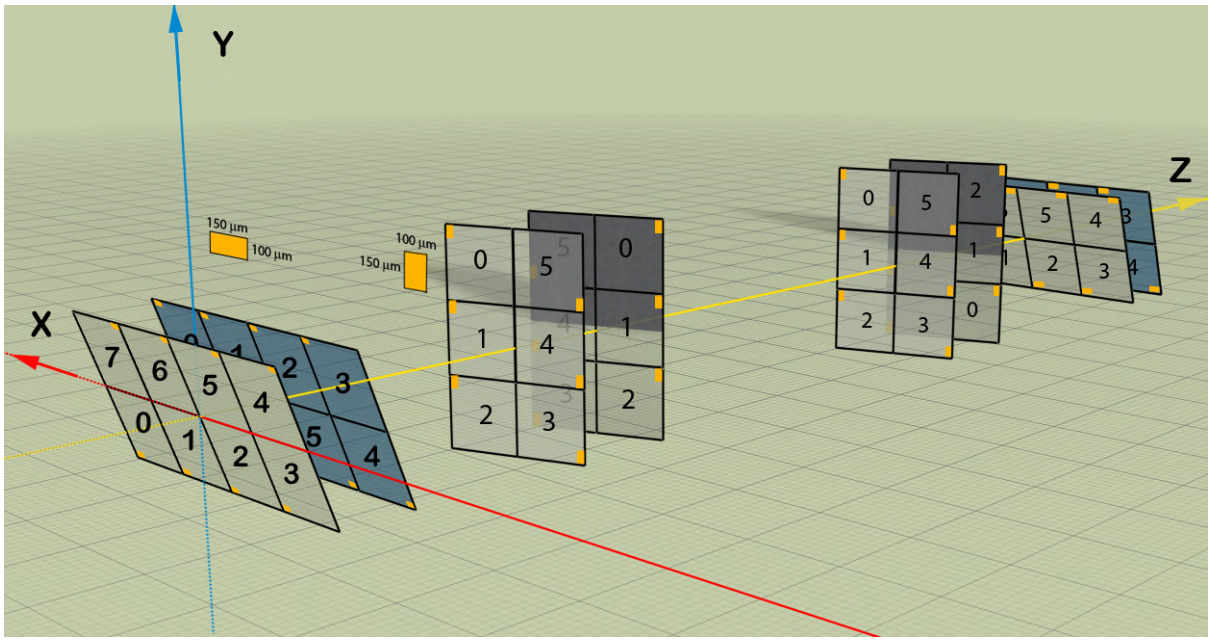


Figure 3.1: Three-dimensional schematic view of the pixel telescope.

In order to exploit the improvement in the spatial resolution when there is charge sharing between adjacent pixels, the planes are tilted by 25 degrees as follows:

- Four planes (2×4 detectors) tilted around the X axis with the long pixel side oriented in the X direction;
- Four planes (2×3 detectors) tilted around the Y axis with the long pixel side oriented in the Y direction.

In the first case, the most precise measurement is in the Y coordinate, while in the second case the most precise measurement is in the X coordinate. This geometry give an overlap active area of $\approx 1.6 \times 1.6\, cm^2$.

Beam particles are triggered by a coincidence generated from three scintillation counters placed behind the telescope. The trigger signal opens a small time window in which the data acquisition system collects and sends data from the detectors to the computer. Data from each ROC are tagged with the trigger count in order to build events starting from time-correlated hits from all the ROCs. An event may contain hit data associated with one or more particle tracks passing through the telescope. To keep the ROCs synchronized with the particle beam, the accelerator clock signal is fed into one of the station and then redistributed to the other stations through SATA cables. Since the CMS ROC has been designed to run at a frequency of 40 MHz, while the Main Injector accelerator frequency is 53 MHz, the clock that is distributed to the stations runs at 26.5 MHz, half of the Main Injector frequency, allowing the detector to work properly and remain synchronized with the beam.

3.3 The Data Acquisition System (DAQ)

3.3.1 Hardware

The Data Acquisition (DAQ) hardware is based on the CAPTAN system developed at Fermilab [34]. The CAPTAN (Compact And Programmable daTa Acquisition Node) is a flexible and versatile data acquisition system designed to meet the readout and control demands of a variety of pixel and strip detectors for high energy physics applications.

The system consists of three CAPTAN nodes, one for each station. The node is a stack of boards with different functions connected via a vertical bus for high-speed data exchange and features a Gigabit Ethernet Link (GEL) for high-speed communication through the network.

In the present case, two boards are stacked in each CAPTAN node (Fig. 3.2):

- The Node Processing and Control Board (NPCB), provided with a Virtex-4 FPGA.
- The Data Conversion Board (DCB), provided with a 12-bit MAX1438 Analogue to Digital Converter.

Each NPCB is connected to a gigabit Ethernet router and the router is, in turn, connected to the computer placed in the FTBF control room through an Ethernet cable. This computer runs the Graphical User Interface (GUI), which controls the entire telescope system.

The master CAPTAN node, which is connected to the DUTs, receives the accelerator clock and the trigger from the scintillators and redistributes them to the other two nodes after having reduced the clock to half the frequency (26.5 MHz). In each node, data from ROCs are received by the DCB, digitalized by the ADC and then sent to the FPGA through the vertical bus. The formatted data are then transferred to the control room PC via gigabit Ethernet. The diagram in Fig. 3.3 shows the full readout system schematically.

At the control room PC, the data from each CAPTAN node are stored in separate directories. Files from different directories are then sequentially scanned and all temporally correlated (marked with the same the trigger count) data from every pixel plane

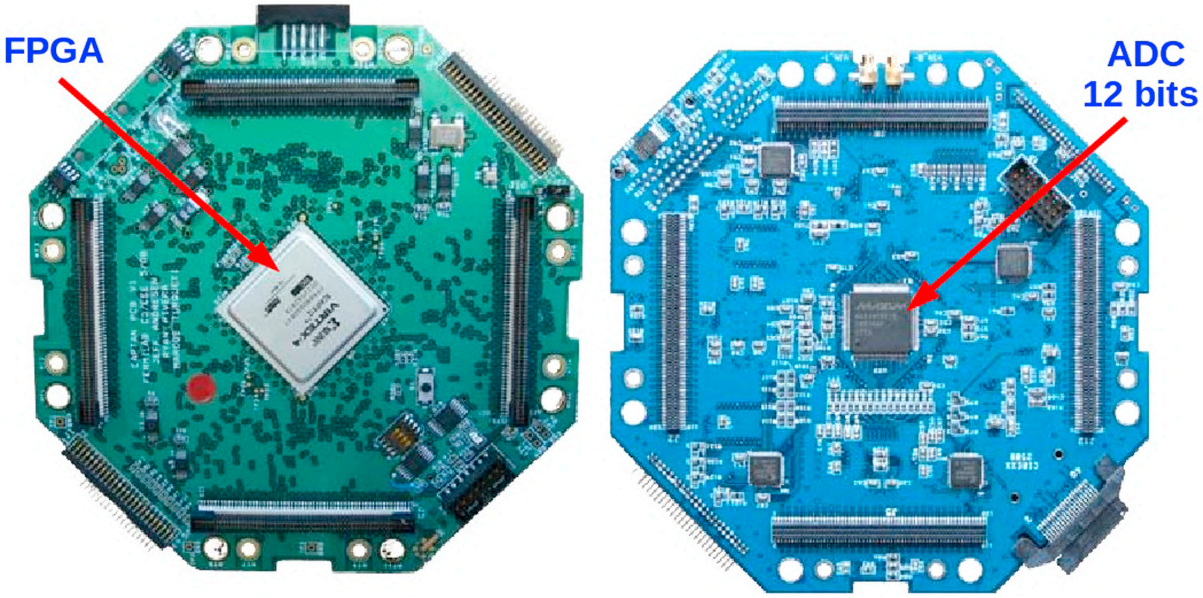


Figure 3.2: Pictures of the primary boards of the CAPTAN stack used for telescope data acquisition. Left: Node Processing and Control Board. Right: Data Conversion Board.

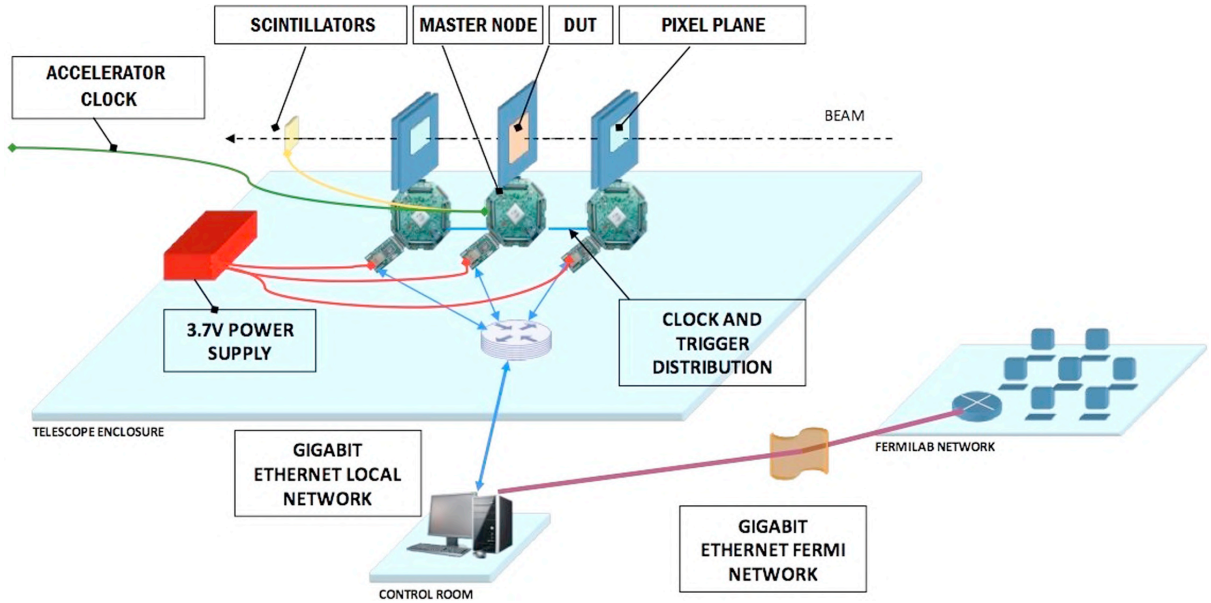


Figure 3.3: Schematic of the telescope readout system based on CAPTAN hardware.

are pulled from the file and merged to form an event. A single binary file containing the merged events is then created for each run. Each pixel data is stored in a 64 bits word (8 bytes), so the typical size of a raw data event with clusters made of two pixels hit on all eight planes is only about 128 bytes.

3.3.2 Software

A complete software solution for interfacing with the CAPTAN system has been designed for Microsoft Windows using Microsoft Visual C++ 2008 [35].

The diagram in Fig. 3.4 shows the CAPTAN software topology schematically.

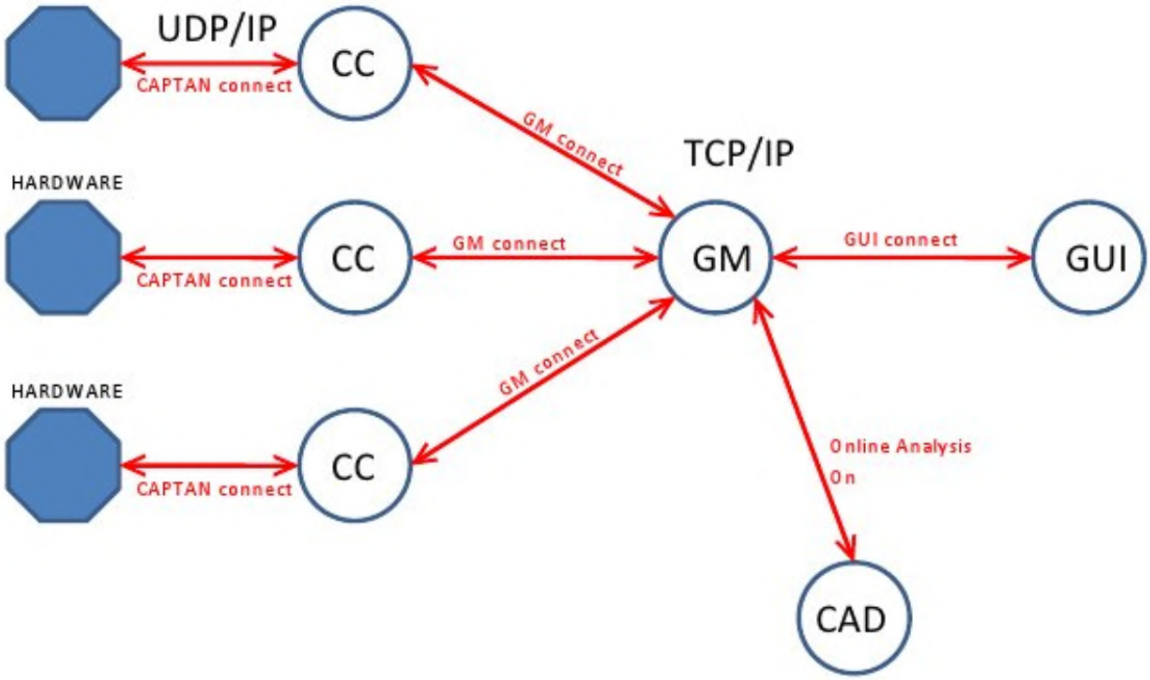


Figure 3.4: Schematic of the telescope readout sotware elements.

The building blocks of the software are the Global Master (GM), the CAPTAN Controller (CC), a Graphical User Interface (GUI) and the CAPTAN Analysis and Display (CAD). The GM is the server for the entire system. It forwards commands from the user to the CAPTAN system and it sends data from the system to the user for interpretation. The CC provides the basic connection between a CAPTAN stack and the GM. There is a CC for each of the three stations: it receives messages from the GM destined to its CAPTAN stack and sends the data that are returned from the stack to the disk or to the GM for delivery to the GUI. The GUI initiates all the write and read processes to and from the CAPTAN stack. It allows the user to configure the readout chips, trigger and clock system, run calibration procedures, as well as to start or stop the data acquisition. The final block of the software architecture is represented by the CAD that allows the user to immediately visualize the telescope merged data in three dimensions, as can be seen in Fig. 3.5. An eventual displacement of the telescope with respect to the beam direction will be visible in this representation and could be easily corrected since the telescope is seated on a X-Y motion table that can be controlled remotely from the control room.

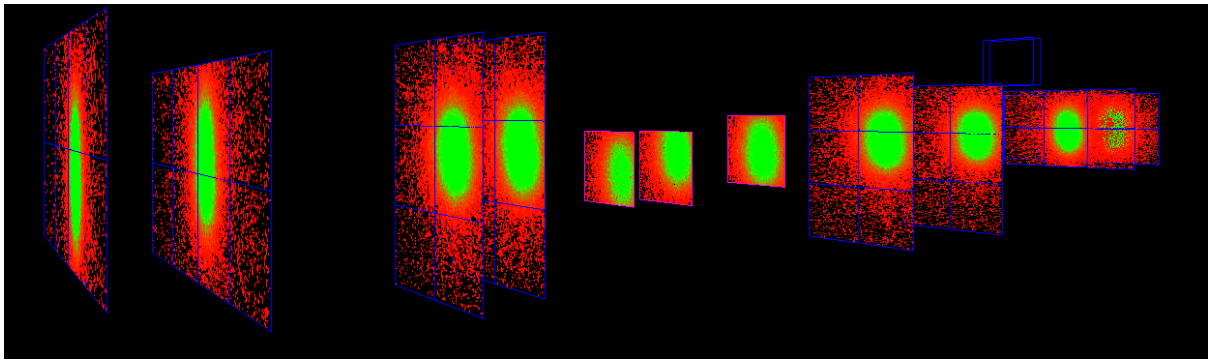


Figure 3.5: Three-dimensional data visualization with the CAPTAN software.

3.4 Telescope tracking and alignment

The main goal of the test beam experiments is to probe the performance of the DUTs with a set of well reconstructed beam tracks. For this purpose the telescope has been designed to achieve an optimal resolution on the coordinates of the track impact point on the DUTs, which are placed at the center of the telescope. The track reconstruction and the telescope alignment are performed by means of the C++ Monicelli [1] application, developed by the INFN Milano-Bicocca group. This software package provides the user with an appropriate iterative procedure to converge toward the optimal alignment of the telescope. All the operations can be accomplished in steps through a GUI that makes the software user-friendly. For debugging, the GUI allows the user to perform the individual steps of the alignment, change critical parameters interactively and finally monitor the partial results by means of specific histograms and distributions. The software is also equipped with efficient interactive tools allowing the user to browse, examine, print and save these distributions in real-time. Fig. 3.6 shows a snapshot of GUI and its components.

The next paragraphs describe the procedure adopted to reconstruct beam tracks and align the telescope.

3.4.1 Track reconstruction

The track reconstruction code implemented in Monicelli was developed for the pixel telescope, it has been adapted to the strip telescope as described later on. It performs a simple and fast straight line fit to the coordinates of the arrays of hits which were preselected by a pattern recognition based on their alignment. In case of clusters of adjacent hits, the coordinates are linearly interpolated over the expected charge-sharing width.

In the first step of the track reconstruction process, Monicelli reads the merged binary data file together with an XML file describing the position and composition of each telescope plane for that particular set of collected data. This file is editable by the user by means of an easy editor included in the package. The geometrical details of each plane are set specifying its space coordinates and rotation angles in the laboratory frame, as well as the number and orientation of ROCs and, for each ROC, the number of columns and rows together with their pitch.

In this stage the information contained in the binary file is decoded: for each event, the ADC value of every pixel hit is associated with the new row and column values

[1] Link to Git repository

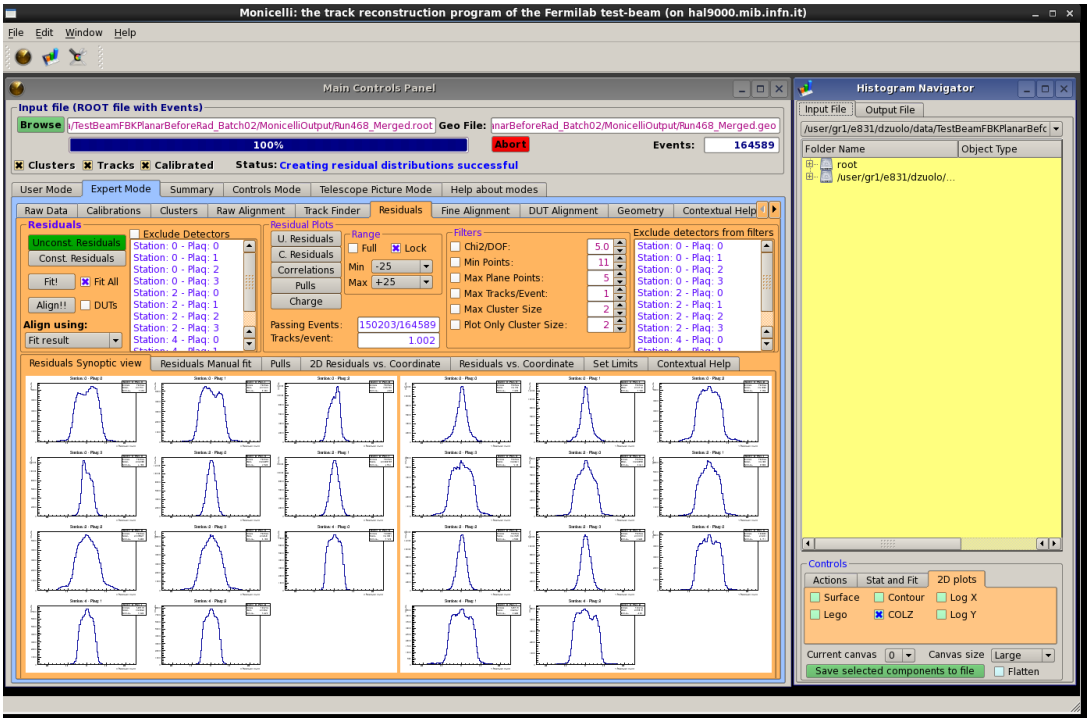


Figure 3.6: Monicelli Graphical User Interface. Inside the desktop style GUI, the main panel on the left controls the operation of the software while the histogram browser on the right is used to plot the control histograms created after each step of the track reconstruction and telescope alignment.

corresponding to the position on the detector according to the telescope information provided by the XML file. Two output files are created on disk: one with decoded events data and another with the associated geometry. In addition, the two-dimensional distributions of the hits accumulated by each pixel plane in that Run is generated as shown in Fig. 3.7. These histograms allow users to cross-check the geometry information provided to the program as well as to check that the beam is roughly centered on all the telescope planes.

The next step in the reconstruction process consists in finding the clusters of adjacent fired pixels in order to attribute to each of them the proper weighted coordinate. Only clusters of one pixel, two adjacent pixels on a row or a column and four adjacent pixels (2×2) are considered for track reconstruction. The others are not used because very rare and, in some cases, resulting from complex processes or detector pathologies, which do not allow for a reliable coordinate interpolation (two tracks merging, delta-rays, noisy or dead pixels, etc.). The cluster coordinates are expressed as local coordinates (to distinguish them from the laboratory system coordinates X and Y defined in Fig. 3.1) of the particular detector they belong to: the x-axis is, by definition, parallel to the pixel rows, the y-axis to the pixel columns.

For single hits, the coordinates are simply those of the pixel center: x $= x_c$ e y $= y_c$.

For two adjacent hits on a column, the y-coordinate is linearly interpolated over the expected charge-sharing width, W, whereas the x-coordinate is the one of the two pixel centers, x_c. W represents the lateral extent of the charge cloud generated by the incoming particle. Depending on the angle α between the y-axis and the plane normal to the beam, W can range from a typically diffusive value of 20 μm ($\alpha \approx 0°$) to a geometry-dominated

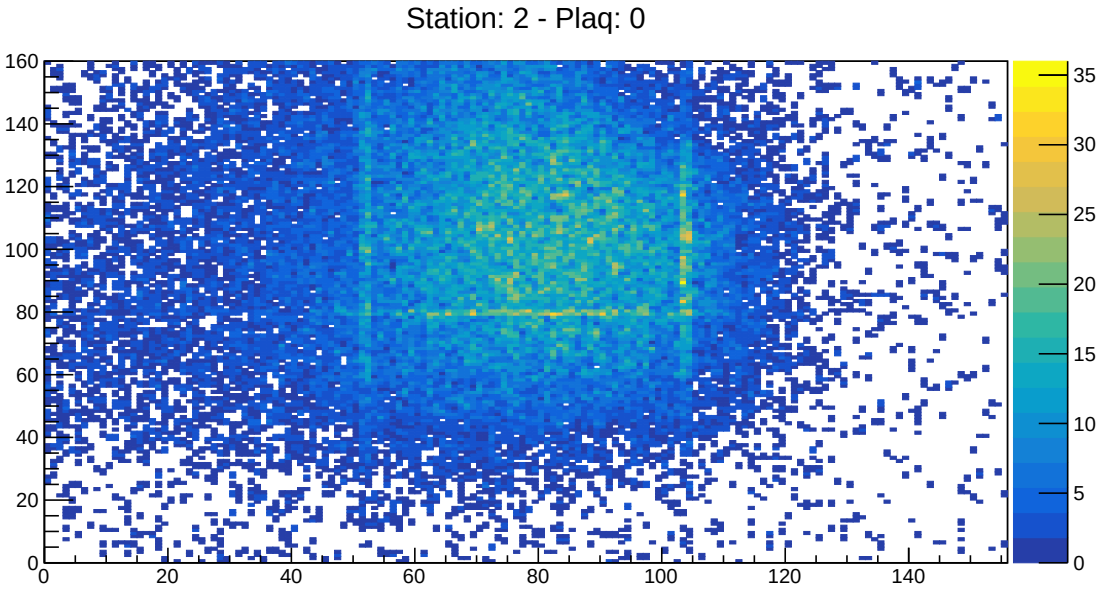


Figure 3.7: Beam spot on one of the telescope detectors.

value of $t \cdot tan(\alpha)$ ($\alpha > 10°$), where t = 285 μm is the thickness of the pixel sensor. In the case of the pixel telescope ($\alpha \approx 25°$) the resulting coordinates are:

$$x = x_c$$

$$y = y_d - \frac{W(25°)}{2} + W(25°)\frac{Q_L}{Q_L + Q_R}$$

where y_d is the coordinate of the divide of the two pixels, $Q_L(Q_R)$ is the charge collected by the pixel on the left(on the right) and W(25°) = 130 μm.

Recalling that the angle β between the x-axis and the plane normal to the beam is $\approx$ 0° the resulting coordinates for two adjacent hits on a row are:

$$x = x_d - \frac{W(0°)}{2} + W(0°)\frac{Q_L}{Q_L + Q_R}$$

$$y = y_c$$

where W(0°) = 20 μm.

The same argument can be extended to treat 2×2 clusters applying a redefinition of Q_L and Q_R:

$$Q_L^x = Q_{i,j} + Q_{i+1,j}$$

$$Q_R^x = Q_{i,j+1} + Q_{i+1,j+1}$$

for x-interpolation, and

$$Q_L^y = Q_{i,j} + Q_{i,j+1}$$

$$Q_R^y = Q_{i+1,j} + Q_{i+1,j+1}$$

for y-interpolation, where i and j are respectively the row and column index of the bottom-left pixel of the cluster.

The resulting coordinates are:

$$x = x_d - \frac{W(0°)}{2} + W(0°)\frac{Q_L^x}{Q_L^x + Q_R^x}$$

$$y = y_d - \frac{W(25°)}{2} + W(25°)\frac{Q_L^y}{Q_L^y + Q_R^y}$$

The errors attributed to the coordinates were calibrated during the commissioning phase of the telescope. They were initially estimated on the basis of what one would a priori expect for the different cluster topologies. Once the alignment was performed they have been refined applying an iterative procedure that aims at having Gaussian unconstrained pulls distribution on each telescope plane, with nearly unitary RMS. The unconstrained (or unbiased) residuals and pulls of a track on a plane are calculated excluding the hits on that particular plane from the track fit.

The measured errors are:

$$\sigma_x = \frac{150}{\sqrt{12}}\ \mu m = 43.3\ \mu m$$

$$\sigma_y = \frac{100(1 - \tan 25°)}{\sqrt{12}}\ \mu m = 15.4\ \mu m$$

for single hits,

$$\sigma_x = \frac{150}{\sqrt{12}}\ \mu m = 43.3\ \mu m$$

$$\sigma_y = 7.8\sqrt{1 + (\tan 25°)^2}\ \mu m = 8.6\ \mu m$$

for double hits along a column, and

$$\sigma_x = 7.8\ \mu m$$

$$\sigma_y = \frac{100}{\sqrt{12}}\ \mu m = 28.9\ \mu m$$

for double hits along a row.

For 2×2 clusters the errors are assumed to be

$$\sigma_x = 7.8\ \mu m$$

$$\sigma_y = 7.8\sqrt{1 + (\tan 25°)^2}\ \mu m = 8.6\ \mu m$$

Once the clusters have been reconstructed, the next operation consists of finding the track candidates.

This is done trying to join with a straight line any combination of cluster hits on the first and the last plane of the telescope and looking for clusters on the intermediate planes. The nearest cluster to the intersection point of the line with each plane, within an adjustable window, is selected. A minimum of 6 hits is required to define a track candidate and the same cluster hit can be included in multiple track candidates. The position of the clusters associated to each track candidate are fitted to a straight line using the least square method.

When all track candidates in an event are reconstructed and fitted, they are sorted by total number of hits and χ^2/DoF. All hits belonging to the track with the highest number of hits and the lowest χ^2/DoF that have been associated to other track candidates are removed from them and those tracks are refitted and resorted again. This operation is repeated until all the hits are assigned to a single track candidate and typically results in the reconstruction of a single track for each event.

3.4.2 Alignment procedure

This paragraph describes the strategy that has been successfully put in place to align the pixel telescope. This procedure resulted to be effective also for the alignment of the strip telescope, as described later on.

1. The alignment requires an initial set of track candidates to start with. In turn, this implies that a preliminary raw alignment of the telescope detectors has to be performed. This first-order approximation is obtained in Monicelli by a relative alignment of the beam profiles on each detector. To this extent, the X and Y projections of the beam spot are fitted with a Gaussian function to obtain the space coordinates of the beam spot centers on each telescope detector. The position of each plane is then modified in such a way that its center coincides with the center of the fitted beam spot.

2. An initial suitable sample of tracks is then found through a "road search", as described in the previous section, performed with large enough tolerances (usually 1000 μm) and without any other cut.

3. A finer alignment is obtained by looking at the mean values of the X and Y residuals on each plane (at this stage they are typically off from zero by a few hundred microns). The X and Y positions of each plane are adjusted in order to compensate the offset from zero of the residual distribution mean value.

4. A new "road search" is now performed with a narrower window (usually 250 μm wide). The quality of the track fit is improved after this preliminary alignment: the peak of the χ^2/DoF distribution is typically below 20.

5. Before moving to the next steps some additional cuts are applied for a further "road search", requiring for instance:

   - $\chi^2/\mathrm{DoF} < 20$
   - at least 8 hits per track
   - no more than 1 hit per plane

   A further look at X and Y residual distributions will give, at this stage, a residual detector translation of a few microns in both directions and the core of the χ^2/DoF distribution will be below 10.

6. Step 3 and 4 can be iterated multiple times. This can be necessary, for instance, when the planes of the telescope are moved to accommodate larger DUTs.

7. At this stage the alignment is precise enough to proceed with the fine alignment iterative algorithm that exploits a least-squares minimization to compute the first order roto-translational corrections. A sample made of tracks with 8 associated hits, each of maximum cluster size 2, is selected to compute and minimize the x and y (local coordinates) unconstrained residuals on each telescope detector. In this first iterative minimization run the rotations are kept fixed. Furthermore, if the angular dispersion of the beam tracks is tiny (≈ 0.1 mrad), even the Z positions of all the detectors have to be fixed since the fit would be insensitive to any shift

in Z (the derivatives of χ^2 with respect to these Z positions would be very close to zero). At the end of each iteration, a new "road search" is performed, using the updated geometry. Typically, 5 iterations are enough to converge to stable values of the parameters. The number of iterations can be set by the user (5 iterations by default) as well as the χ^2/DoF cut (χ^2/DoF < 10 by default). At the end of this process, pull, unconstrained residual and correlation distributions are produced to check the alignment progress.

8. Finally, a second iterative minimization run is executed releasing also the rotations while keeping fixed the Z positions of all the telescope planes. Usually, as for step 7, 5 iterations are set with a χ^2/DoF cut of 10.

3.4.3 Alignment results

In this paragraph some distributions produced after the alignment of the pixel telescope are described. Such distributions can be produced at every step of the procedure giving the user the possibility of adjusting some parameters.

Fig. 3.8 shows the x and y unconstrained residual distributions (local coordinates) for one of the telescope planes. As shown in the figure, the residuals distributions show a large non-Gaussian shape along the non-tilted coordinate (pitch = 150 μm) and a narrow shape along the tilted one (pitch = 100 μm). The former is determined by the most probable single-hit events, resulting in distributions with RMS of about $\frac{150}{\sqrt{12}}$ $\mu m \approx 43$ μm. In the second case the distribution is dominated by double hit events since the tilt of the sensor increases the charge sharing between adjacent pixels. The distribution is consequently much more similar to a Gaussian with an RMS of about 20 μm.

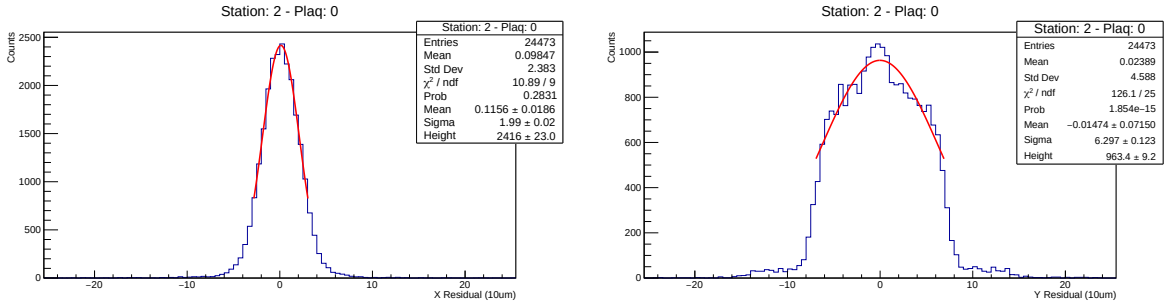


Figure 3.8: The x and y residuals distributions for one telescope plane after a complete alignment.

Fig. 3.9 shows the x and y pull distributions for one telescope detector. The pull ($p_{x,i}$, $p_{y,i}$) of a track on a detector i is defined as the unconstrained residual normalized to its error. The error is the square root of the sum in quadrature of the error associated to the measured hit coordinates ($x_{m,i}$, $y_{m,i}$) and the error on the impact point ($x_{p,i}$, $y_{p,i}$) of the track fit obtained excluding the hit on that plane:

$$p_{x,i} = \frac{x_{m,i} - x_{p,i}}{\sqrt{\sigma^2_{x_{m,i}} + \sigma^2_{x_{p,i}}}}$$

$$p_{y,i} = \frac{y_{m,i} - y_{p,i}}{\sqrt{\sigma^2_{y_{m,i}} + \sigma^2_{y_{p,i}}}}$$

The quasi unitarity of the x and y pull distributions confirms the correct estimation of the measurement errors and the good quality of the fit.

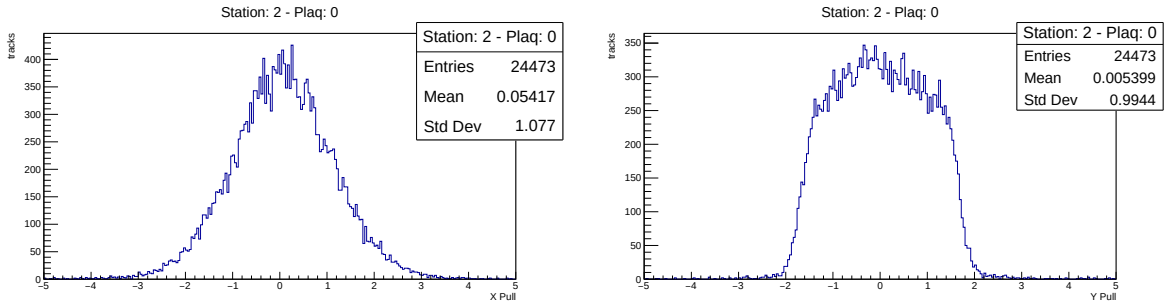


Figure 3.9: The x and y pull distributions for one telescope plane after a complete alignment.

The accuracy of the alignment is further investigated studying the plots of Fig. 3.10, showing the correlations between the unconstrained residuals and the impact point coordinates on the detector. If the telescope is properly aligned the core of the distribution is centered around the value 0 for the residual and is uniform along the coordinate. A residual correlation between track residual and coordinate indicates a residual angular misalignment.

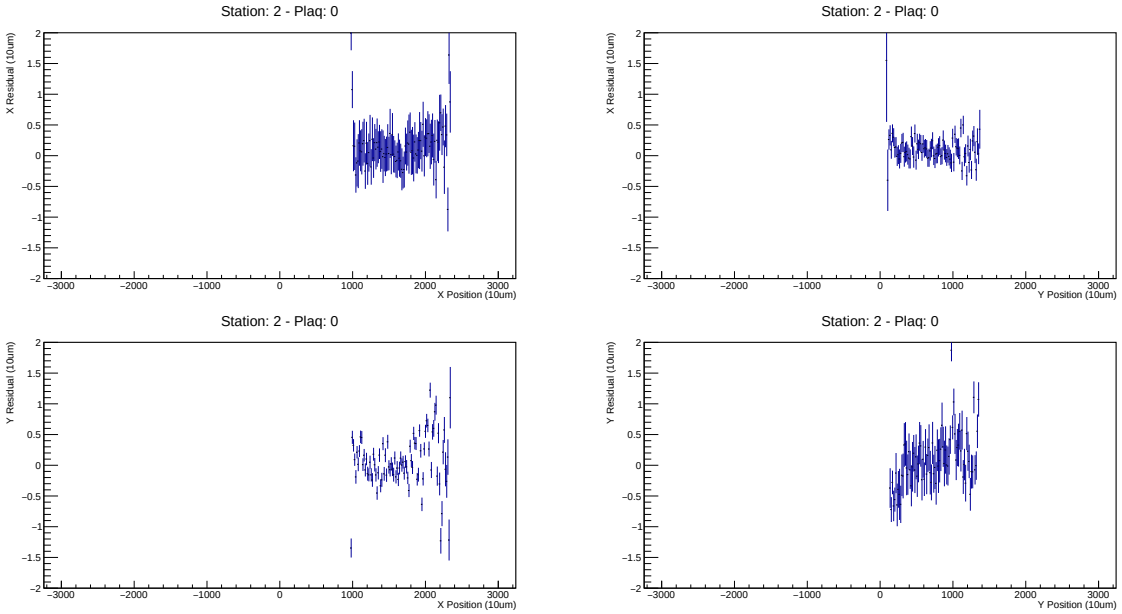


Figure 3.10: Plots of the correlation between unconstrained residuals and impact point coordinate for one of the telescope detectors after a complete alignment.

The fitted tracks χ^2/DoF distribution is showed in Fig. 3.11. Once the alignment of the telescope is performed the peak of the distribution is around 1, indicating a good quality of the fit.

Fig. 3.12 reports the tracks X and Y slope distributions showing a very small angular dispersion ($\approx 10^{-4}$ rad), which results in a low resolving power for the determination of Z position corrections.

Finally, the track extrapolation errors at the DUT Z position (Z $\approx$ 0) is reported in Fig. 3.13 for tracks with associated hits on each telescope plane. A silicon 3D pixel

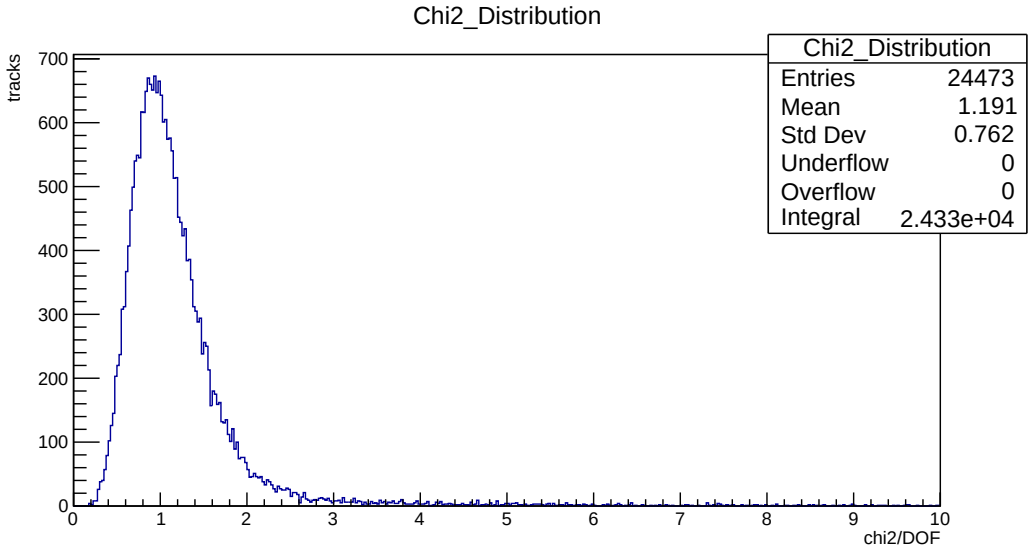


Figure 3.11: Fitted tracls χ^2/DoF distribution after a complete alignment.

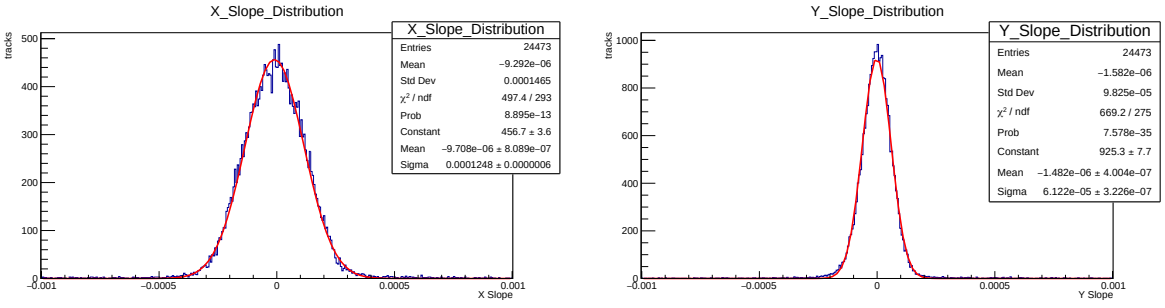


Figure 3.12: Distributions of the tracks X and Y slopes after a complete alignment.

detector was used to estimate the telescope resolution since the diffusive charge sharing between adjacent pixels is much reduced with respect to the planar case. The sensor (Sintef 2E) had a $100 \times 150 \mu m^2$ pixel cell with two central columnar signal-electrodes (10 μm radius) and six columnar field-electrodes (7 μm radius) on the periphery. The sensor thickness was 220 μm.

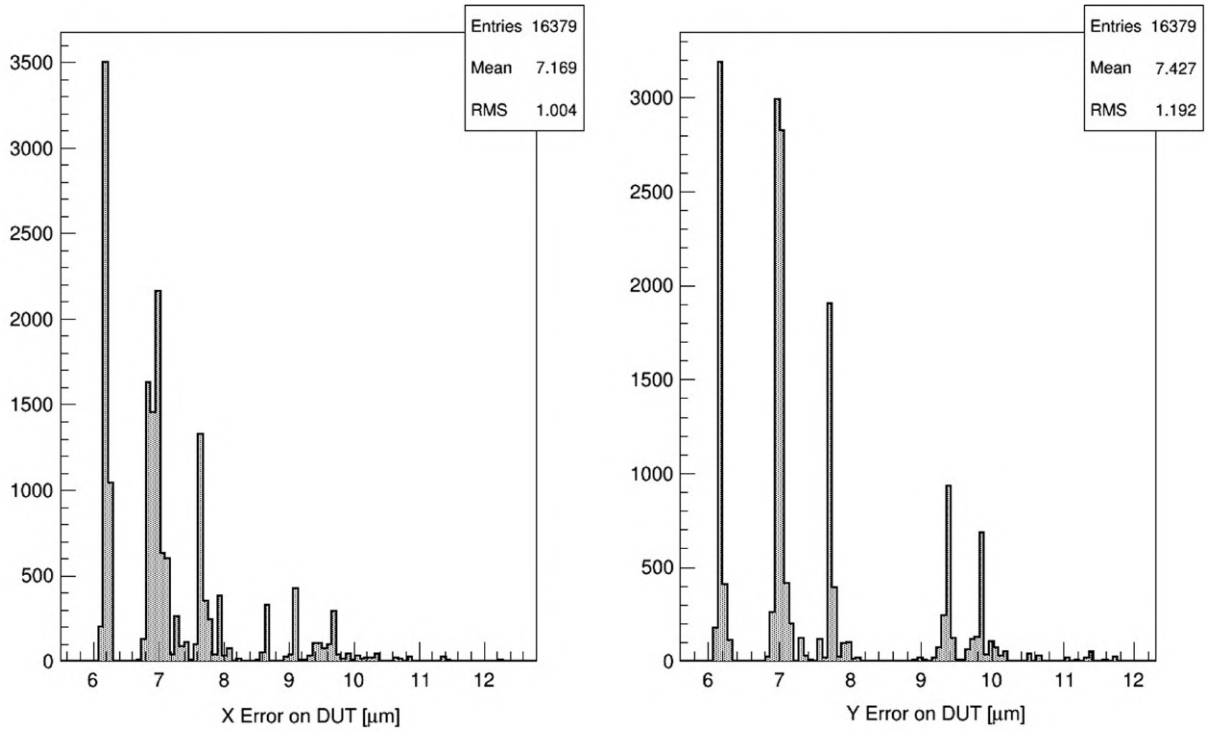


Figure 3.13: Distributions of the track impact point errors at the DUT Z position (Z $\approx$ 0) after a complete alignment. The discrete peaks result from different combinations of single and double hits on the telescope planes.

It turns out that the best achievable telescope resolution on the DUT is as small as 6.2 μm in both X and Y coordinates, and that the bulk of tracks give resolutions better than 8 μm. Excluding tracks with extrapolation error larger than 8 μm, as it is typically done for data analysis, the average error at DUT is approximately 6.9 μm on both projections (6.87 μm in X and 6.94 μm in Y). This error will be called Pure Telescope extrapolation error. To obtain the Total Expected extrapolation error (at 120 GeV proton energy), the error due to the Multiple Coulomb Scattering (MCS) in the DUT must be added in quadrature. For three or less DUTs placed in the standard configuration around the telescope center at Z_1 = +5 cm, Z_2 = -5 cm and Z_3 = +6 cm, the Total Expected extrapolation error is given by

$$\sigma_x = \sqrt{6.87^2 + 1.37^2 \times L_{DUT_1} + 1.95^2 \times L_{DUT_2} + 2.01^2 \times L_{DUT_3}}\ \mu m$$
$$\sigma_y = \sqrt{6.94^2 + 1.37^2 \times L_{DUT_1} + 2.68^2 \times L_{DUT_2} + 2.74^2 \times L_{DUT_3}}\ \mu m$$

where L_{DUT_n} is the thickness of the nth DUT in percent of a radiation length and the numerical factors are the width of the Gaussisn scattering angle at the different DUTs z-positions. The errors in the previous formulae are negligible ($< 0.1\ \mu m$) and hence omitted.

Fig. 3.14 shows the residual distribution on the DUT. Normally incident telescope tracks with associated double hits on the DUT are considered to produce these histograms. The coordinate attributed to the doublet is exactly the divide between the two adjacent pixels: this is done to minimize the contribution of the error attributed to the coordinate to the RMS of the distribution.

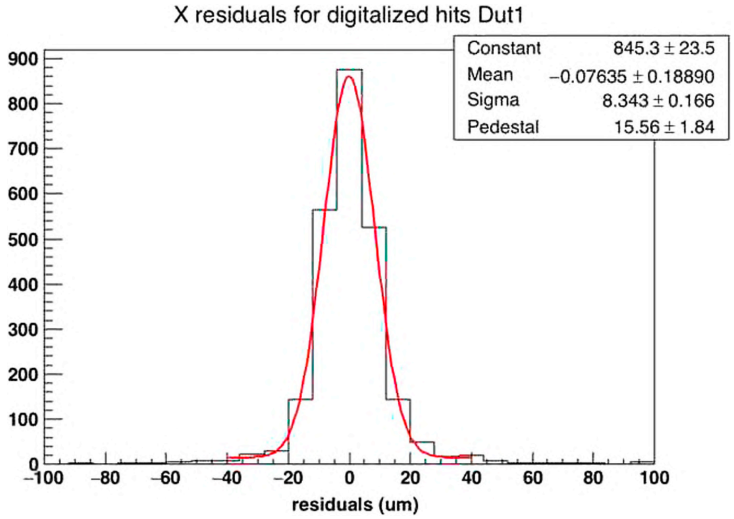


Figure 3.14: Track residuals of two adjacent hits along X on the 3D Si-pixel detector.

The measured RMS of the residual, $8.34 \pm 0.17 \mu m$, results from the sum in quadrature of the actual track extrapolation error and that of the coordinate measured on the DUT. From a study of the correlation between the telescope track impact point and the charge asymmetry of the two pixels of the doublet on the DUT, it is found that the effective region of sharing is concentrated around the divide of the doublet with RMS $2.38 \pm 0.60\ \mu m$. Unfolding this error from the measured residual, the actual track extrapolation error results $7.99 \pm 0.25\ \mu m$ and it is fully consistent with a Total Expected value of 7.86 μm, resulting from the sum in quadrature of the Pure Telescope resolution, 6.87 μm, and a multiple Coulomb scattering error of 3.82 μm, due to the simultaneous presence of three DUTs (1.5% of a radiation length each).

On this basis it is possible to conclude that the actual track extrapolation error at the DUT is well matched by the Total Expected extrapolation error resulting from the formulae in this section and that the pixel telescope allows for a real track extrapolation error at the DUT typically better than 8 μm both in X and Y.

As an example, in Fig. 3.15 the residuals distributions for x and y coordinates of a DUT are shown. The DUT is a planar sensor with $100 \times 150 \mu m^2$ pixel cells and 130 μm active thickness. Only cluster of size 2 on the DUT are considered and the measured coordinate is assigned taking into account its correlation with the charge asymmetry, in order to get a more precise measurement.

3.5 The strip telescope

During the first half of 2018 the pixel telescope has been replaced by a strip telescope. The new telescope is still composed by an upstream and downstream station, each of them consists of three couples of planes. Planes in a couple are rotated by 90 degrees relative to each other: in this way three (X,Y) impact points are measured by each station. The

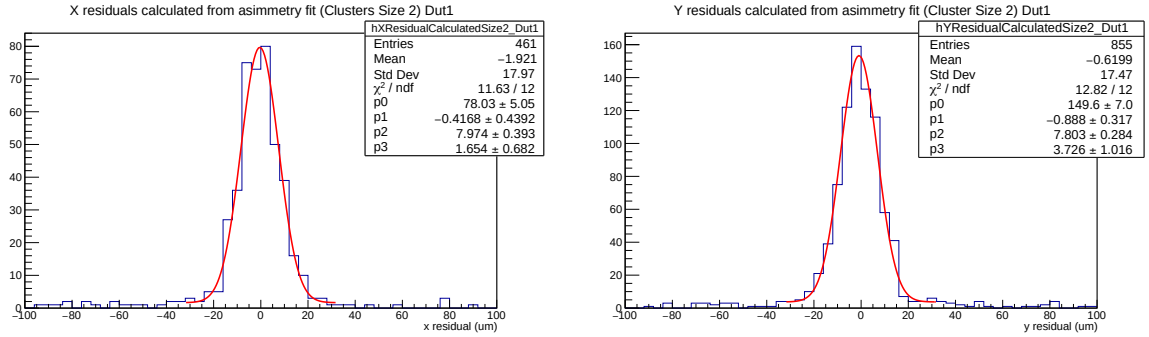


Figure 3.15: The x and y residuals distributions for a generic planar DUT with $100 \times 150\,\mu m^2$ pixel cells and 130 μm active thickness.

overlap area of a pair of sensors is $3.8 \times 3.8\,cm^2$ which is approximately six times the active area of a plane of the old telescope. Each couple is mounted on a support which is tilted by 15 degrees around the Y axis, increasing the charge sharing between adjacent pixels along the X axis. A picture of the telescope is showed in Fig. 3.16.

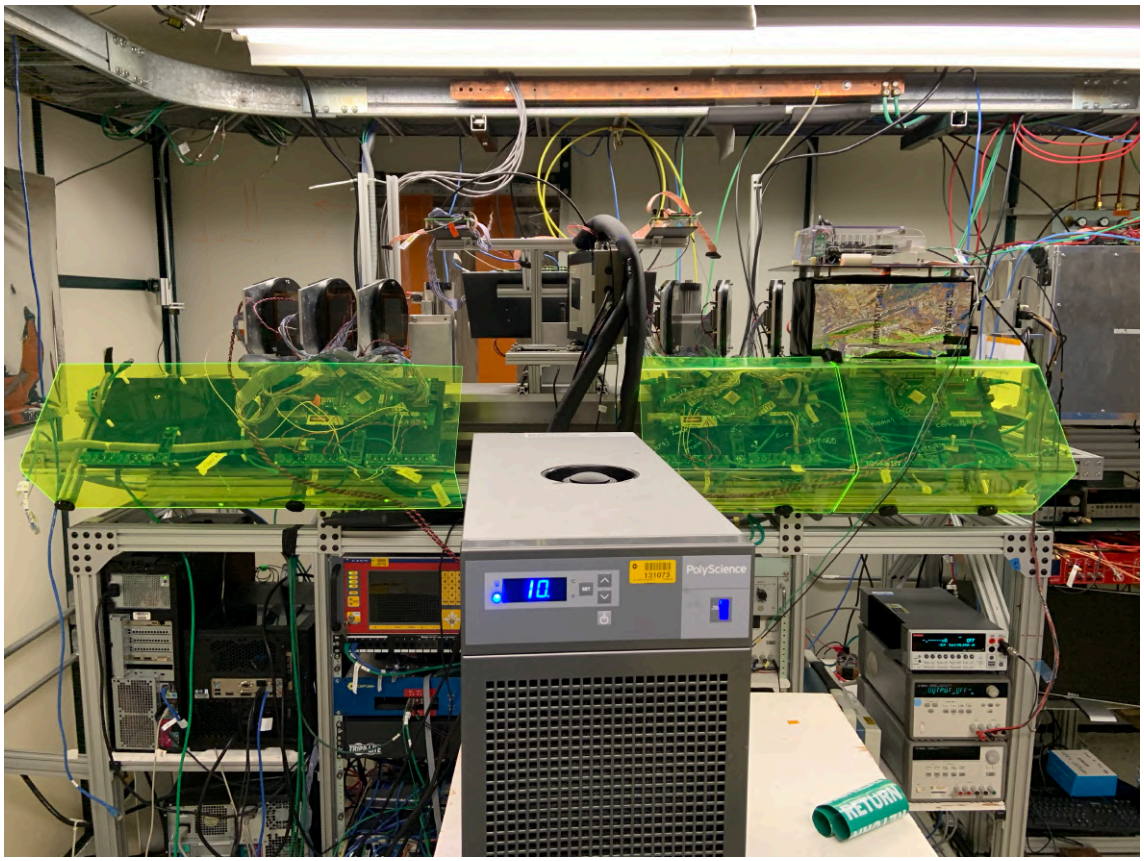


Figure 3.16: Picture of the strip telescope in the FTBF experimental area.

The strip sensors employed were originally produced for the D0 Run IIb upgrade, they have an active thickness of 320 μm and a pitch of 30 μm. This is a sensible reduction with respect to the dimensions of the pixel cells of the old telescope ($100 \times 150\,\mu m$) and will result in a strong improvement of the resolution, as described later. Each sensor is made of 639 strips that are capacitively read-out so the effective pitch is 60 μm.

The employed read-out chip is called "FSSR2" [36] and was developed for the proposed

BTeV experiment. It has 128 analog inputs, so each sensor requires 5 ROCs. The main feature of this ROC is a data driven read-out: no external trigger is required and all data that are above threshold are time-stamped with a running 8 bit counter called "bunch counter" (BCO) and transmitted. In order to correlate these data with a trigger another counter, synchronous with the BCO counter of the FSSR2 chip, has been implemented in the CAPTAN FPGA. This counter is used to extend to 48 bits the native FSSR2 8 bits counter but also to time-stamp the trigger coming from the scintillators. Since the trigger is always registered in the FPGA with a fixed delay, it is possible to correlate the particle arrival and the FSSR2 BCO with the constant delayed trigger counter latched in the FPGA.

One of the stations of the pixel telescope has been kept as last station of the telescope and is employed to minimize the number of fake tracks by requiring that each track reconstructed by the strip telescope is associated to a hit in a pixel plane.

The clustering and track reconstruction algorithms had to be revisited since they were designed to deal with two-dimensional hits.

We decided to refer to the measured coordinate of the strip as "x" and to use the center of the plane as "y" coordinate. For double hits we implemented a very simple algorithm to determine the coordinate:

$$x = x_L + \frac{Q_R}{Q_R + Q_L} x_L$$

The error associated to the y coordinate has been set to $3800/\sqrt{12}\ \mu m$ where 3800 μm is the length of the overlap area of two strip planes in tens of microns.

The error associated to the x coordinate for single hits has been set to $30/\sqrt{12} = 8.66 \mu m$ while for double hits it has been determined using the same iterative produced applied for the pixel telescope and resulted to be

$$x_{err} = 30/\sqrt{12} = 8.66\ \mu m$$

for tilted planes, and

$$x_{err} = 0.7(30/\sqrt{12}) = 6.06\ \mu m$$

for non-tilted planes.

The existing algorithm for track reconstruction turned out to be inadequate due to the higher impact of the MCS of the strip telescope planes. Nevertheless we decided to keep it to determine a first approximation of the track parameters. This has been possible since each plane in a pair measures the x or y coordinate and the missing coordinate can be taken from the other plane measurement.

A new algorithm based on the Kalman filter approach has been developed and will be described in the following.

The coordinate system used to describe the sensor orientation has the unit vector $\hat{u}$ parallel to the strips and directed away from the ROCs. The unit vector $\hat{v}$ is in the plane of the sensor, perpendicular to the strips and pointing in the direction of increasing strip number. The unit vector $\hat{n}$ is normal to the surface of the sensor as shown in Fig.3.17. The nominal center of the sensor is located on strip 320.

A generic track can be parametrized by the equations:

$$x(z) = x_0 + \alpha z$$

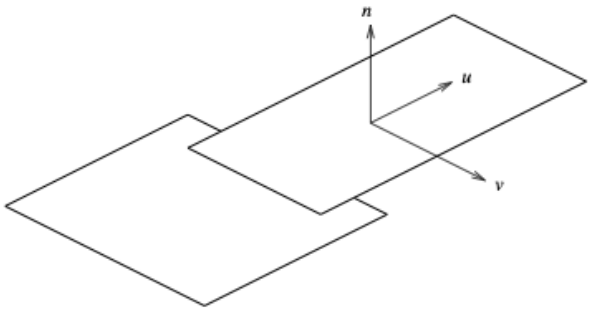


Figure 3.17: Coordinate system used to define the position and orientation of a sensor.

$$y(z) = y_0 + \beta z$$

and its state at plane k can be represented by the vector:

$$\vec{x} = (x_k,\ y_k,\ \alpha_k,\ \beta_k)$$

where x_k and y_k are the intercept along the x and y axis and α_k and β_k are the slopes.

The position of each sensor can be defined by a displacement vector $\vec{p}_k$ and the unit vectors $\hat{u}_k$, $\hat{v}_k$, and $\hat{n}_k$. The intersection point of a track with a sensor can be determined by solving the equation:

$$\hat{n}_k \cdot \big(\vec{p}_k - \vec{x}(z)\big) = 0$$

for z and calculating the position of the intersect strip as:

$$s_k = \hat{v}_k \cdot \big(\vec{p}_k - \vec{x}(z)\big)$$

In order to be able to linearize this equation the slopes of the tracks must be very close to 0 (i.e. tracks parallel to the z axis): this is exactly the first step of the alignment procedure described earlier and can be carried out using the old track reconstruction algorithm.

The linearized equation is:

$$s_k = \vec{H}_k \cdot \vec{x}_k - d_k$$

where $\vec{H}_k$ and d_k contain the information on the plane position and orientation in space.

The covariance matrix for the state vector which includes information of all planes before k is written $\mathbf{C}_{k-1}$ and the predicted covariance matrix for the state vector extrapolated to plane k is written:

$$\mathbf{C}_k^{k-1} = \mathbf{C}_{k-1} + \mathbf{Q}_{k-1}$$

where $\mathbf{Q}_{k-1}$ is the contribution of the MCS of plane k-1 to the covariance matrix.

The initial covariance matrix comes from a fit to the aligned cluster hits using the old algorithm while $\mathbf{Q}_{k-1}$ can be written as:

$$\begin{bmatrix} z_{k-1}^2\rho_{k-1}^2 & 0 & -z_{k-1}^2\rho_{k-1}^2 & 0 \\ 0 & z_{k-1}^2\rho_{k-1}^2 & 0 & -z_{k-1}^2\rho_{k-1}^2 \\ -z_{k-1}^2\rho_{k-1}^2 & 0 & \rho_{k-1}^2 & 0 \\ 0 & -z_{k-1}^2\rho_{k-1}^2 & 0 & \rho_{k-1}^2 \end{bmatrix}$$

where the parameter ρ_{k-1} is the width of the Gaussian scattering angle measured for the telescope planes and a generic DUT.

The residual of a measurement of a cluster position m_k with covariance matrix $\mathbf{V}_k = \left[\sigma_k^2\right]$ can be written as:

$$r_k^{k-1} = m_k - \vec{H}_k \cdot \vec{x}_k + d_k$$

i.e. the difference between the measured and the predicted coordinate.

The correction to the state vector after incorporating the measurement on plane k

$$\Delta\vec{x}_k = \mathbf{K}_k r_k^{k-1}$$

depends on the Kalman filter gain matrix $\mathbf{K}_k$ which can be computed as:

$$\mathbf{K}_k = \frac{\mathbf{C}_k^{k-1}\vec{H}_k^T}{e}$$

where

$$e = \vec{H}_k \mathbf{C}_k^{k-1} \vec{H}_k^T + \sigma_k^2$$

which represents the sum in quadrature of the track extrapolation error at plane k with the error attributed to the measured coordinate m_k. It is evident that hits whose position is measured with a better resolution (i.e. small e) have a large contribution to the K matrix and hence to the change in the track state vector, as expected.

The new covariance matrix for the state vector can be computed as:

$$\mathbf{C}_k = \mathbf{C}_k^{k-1} - \frac{\mathbf{C}_k^{k-1}\mathbf{B}_k\mathbf{C}_k^{k-1,T}}{e}$$

where

$$\mathbf{B}_k = \vec{H}_k \vec{H}_k^T$$

Once the track state vector and covariance matrix have been updated for every plane going from the first to the last, a smoothing of these quantities is performed considering the contribution of each plane in reverse order.

The last state vector $\vec{x}_n$ provides the initial estimate for the smoothed parameters. Smoothed parameters at the other planes can be calculated as:

$$\vec{x}_k^n = \vec{x}_k + \mathbf{A}_k \left(\vec{x}_{k+1}^n - \vec{x}_{k+1}^k\right)$$

where

$$\mathbf{A}_k = \mathbf{C}_k \left(\mathbf{C}_{k+1}^k\right)^{-1}$$

.

The matrices $\mathbf{C}_k$ and $\mathbf{C}_{k+1}^k$ have been computed in the previous filtering stage, as the state vectors $\vec{x}_k$ and $\vec{x}_{k+1}^k$.

The covariance matrix can be updated as:

$$\mathbf{C}_k^n = \mathbf{C}_k + \mathbf{A}_k \left(\mathbf{C}_{k+1}^n - \mathbf{C}_{k+1}^k\right) \mathbf{A}_k^T$$

In both filtering and smoothing stages the residuals of the track, the error associated to the predicted impact point and the contribution to the χ^2 can be computed for each plane as follow:

$$r_k = m_k - \vec{H}_k \cdot \vec{x}_k + d_k$$
$$e_k = \sigma_k^2 - \vec{H}_k \mathbf{C}_k \vec{H}_k^T$$
$$\chi_k^2 = \frac{r_k * r_k}{e}$$

This algorithm was successfully extended to include the pixel planes: their MCS contribution is computed in the filtering phase and they are included in the smoothing phase to compute their χ^2 contribution in order to decide if a track is genuine or fake. [2]

The alignment algorithm was not modified since it resulted to be effective for the alignment of the pixel+strip telescope. Nevertheless the alignment procedure had to be revisited in order to guarantee a fast enough convergence:

- The alpha angles of the strip planes (rotations along the x-axis) must be set to 0 and fixed during the alignment: this is necessary because none of the planes is tilted along the x-asis so there is no sensitivity to this angle
- The beta angles of the strip planes (rotations along the y-axis) must be set to -15: the two orthogonal planes of a pair are mounted on the same support which is tilted, the value will be corrected only for one of the planes which is sensible to this angle
- The first step of the alignment is now done fixing all the parameters for the first and last pair of strip planes
- In a second step the parameters of the first or last pair can be released
- In a third step the positions of the other pair can be released but the angles must remain fixed: this necessary because the orthogonality of the planes of each pair is not explicitly imposed in the alignment code

The performance of this new telescope is showed in Fig. 3.18 where the x and y residuals distributions for a $50 \times 50\ \mu m^2$ pixel cells planar sensor with 100 μm active thickness are reported.

The x and y resolutions are different because the planes of the strip telescope are tilted around the Y-axis only. In this case the resolution of the telescope could not be measured directly because the RMS of these residuals distributions are so small that none of the available sensors can provide a lateral charge diffusion negligible compared to it, as was the case for the pixel telescope using a 3D sensor.

The resolution of the telescope can then be assumed to be $< 5\,\mu m$ in the x-coordinate and $< 6\ \mu m$ in the y-coordinate: this is a considerable improvement with respect to the resolution of 8 μm measured for the pixel telescope.

[2] I want to thank Matthew Jones (Purdue University) for his contribution to the development of the algorithm

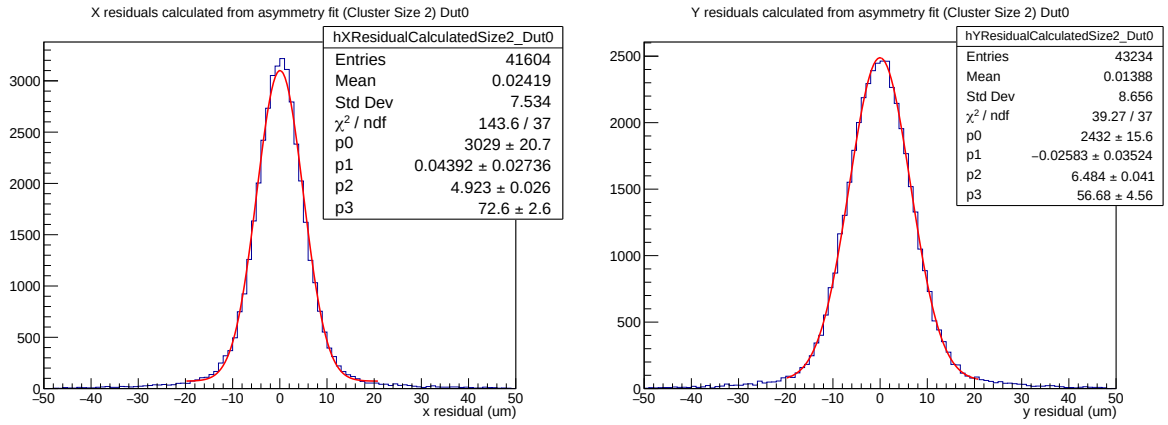


Figure 3.18: The x and y residuals distributions for a generic planar DUT with $50\times50\mu m^2$ pixel cells and 100 μm active thickness.

3.6 Data analysis

The final step of the analysis aims at producing distributions describing the charge collection properties, detection efficiency and position resolution of the DUTs through a second program, called "Chewie" [3], developed by a group of the INFN section of Milano Bicocca as well.

This program converts the files produced by Monicelli in ROOT ntuples where each entry correspond to a reconstructed track.

Complementary information on the planes of the telescope and on the DUTs and the selections that will be applied while producing the plots are contained in an XML configuration file processed by the program. In this file it is also possible to define the DUT region to be analyzed in terms of rows and columns.

The selections required to obtain a sample of well reconstructed beam tracks are the following:

- hit on each telescope plane
- no more than 5 hits on each telescope plane
- χ^2/DoF of track < 5
- one track per event
- track impact point inside the detector area

The program is composed by three C++ classes named "Efficiency", "Charge" and "Resolution" that produce all the distributions needed to study the performance of a DUT and can be adapted by the user to different types of detectors. The analysis is very rapid as it is carried out with a multi-threaded process and consists only of filling histograms and performing fits.

There is a fourth C++ class, called "Window", whose execution is automatic, which defines the region of the DUT to be analyzed in terms of rows and columns. This very useful in case of non-uniformly irradiated DUTs and contributes to further speed up the analysis since only tracks with impact point on the DUT inside that region are considered.

[3] Link to Git repository

Figure 3.19 shows the GUI of this program, which is also completed by a navigator for histogram parsing.

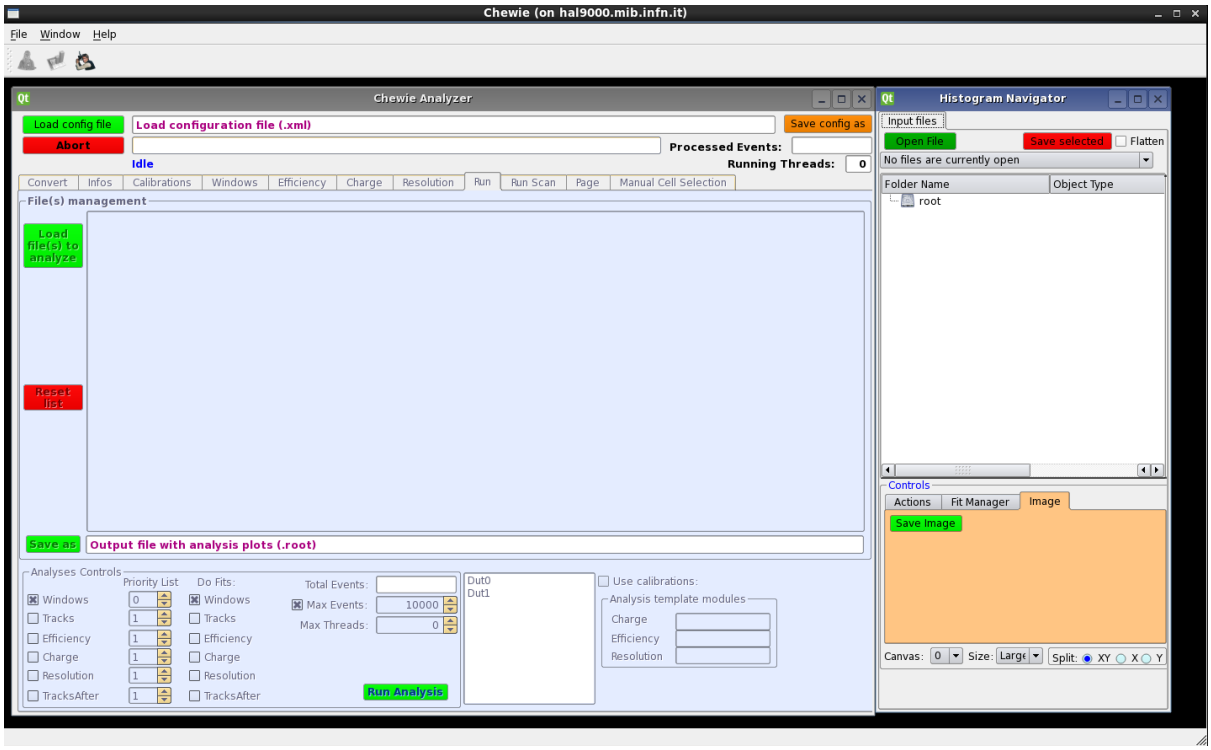


Figure 3.19: Graphical User interface of the Chewie program.

3.7 DESY test beam facility

The facility offers three independent beam lines with electron or positron particles with selectable momenta from 1 to 6 GeV/c.

The test beams are generated by a double conversion instead of using a direct extraction of the primary beam in DESY synchrotron. Initially bremsstrahlung photons are generated by a fiber target positioned in the synchrotron beam orbit. These photons hit the secondary target generating electron/positron pairs. Depending on the polarity and strength of the magnetic field of the following dipole magnet, the test beam particles reaching the test beam areas are electrons or positrons with a certain momentum. There are three independent beam lines, called TB21, TB22 and TB24, named after the positions of the primary targets located behind the quadrupoles QF21, QF22 and QF24 respectively. A schematic view of the beam generation and the three beam lines is shown in Fig. 3.20.

The DESY Test Beam Facility is equipped with EUDET-type pixel beam telescopes which allow to track the test beam particles[37]. These kinds of test beam trackers were originally developed within the EUDET project in order to meet most user requirements in terms of easy integration of the DUTs, precise spatial resolution and suitable event rates.

Each beam telescope is composed of two telescope arms incorporating each three planes. Each plane consists of MIMOSA26 monolithic active pixel silicon sensors [38]. The cell size is $18.4\,\mu m \times 18.4\,\mu m$ and pixels are arranged in 1152 columns and 576 rows,

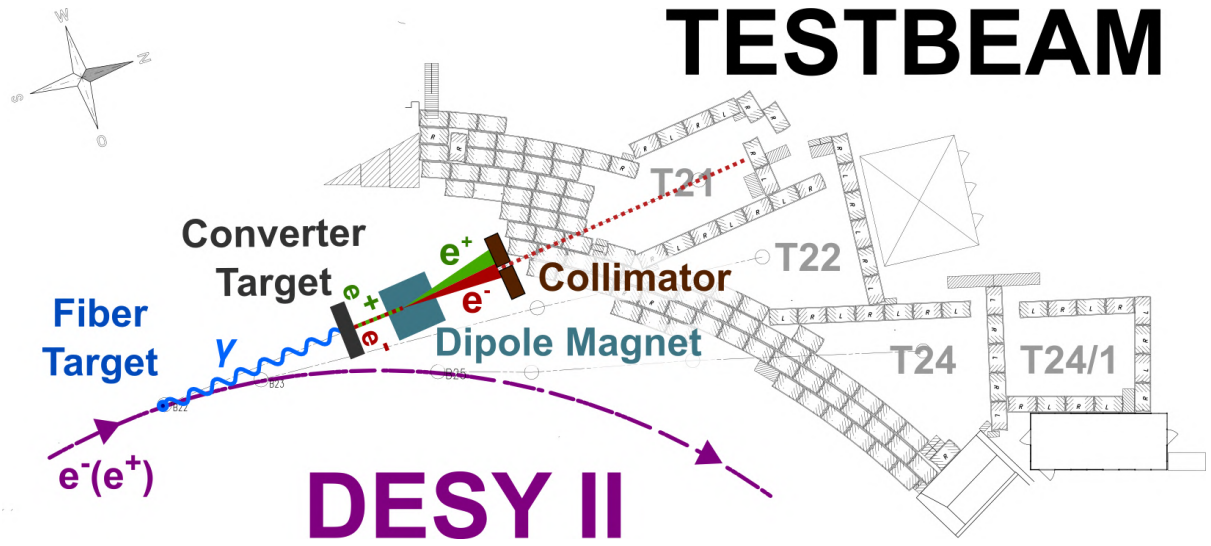


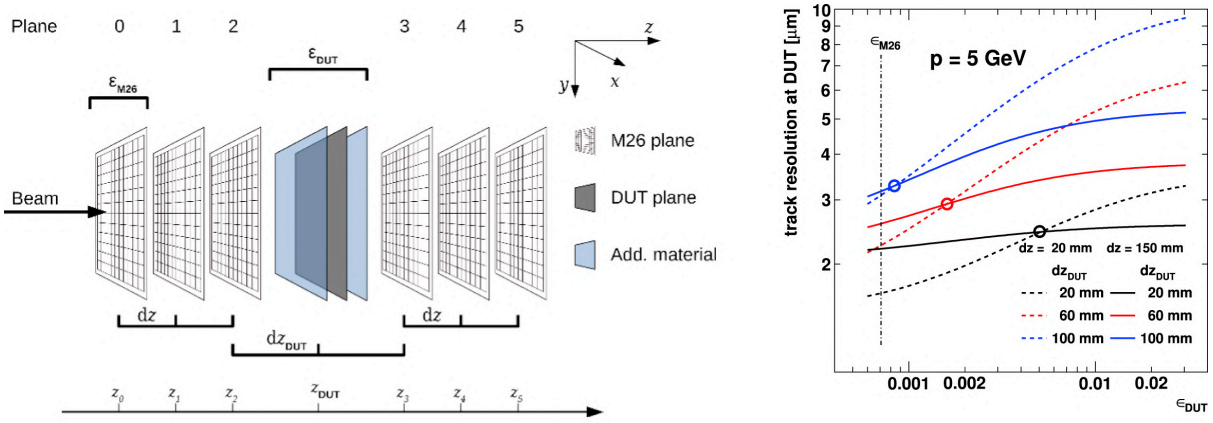
Figure 3.20: A schematic view of the test beam generation at the DESY Test Beam Facility, here for beamline TB21. Bremsstrahlung photons generated in the primary target travel through the tunnel and hit the secondary target generating electron/positron pairs. The dipole magnet selects particles according to their momentum and charge and the particle beam can be further collimated before entering the test beam areas.

which results in an active area of about 21.2 mm × 10.6 mm. Pixel states are continuously read out in a rolling shutter by buffering line by line, the on-chip digitization provides a binary pixel information, and the output data stream is zero-suppressed. Therefore, the integration time is 115.2 μs per frame. Each MIMOSA26 sensor has an active thickness of 50 μm and is shielded from environmental light using 25 μm thick Kapton foil on each side. This keeps the material budget as low as possible in order to achieve a high track resolution at 1–6 GeV/c. The intrinsic resolution of a sensor was measured to be (3.24 ± 0.09) μm. The best track resolution is estimated to (1.83 ± 0.03) μm using an equidistant plane spacing of 20 mm at a 5 GeV/c test beam. The realistic track resolution depends on the beam momentum, the plane spacing and the material budget of the DUT (Fig. 3.21).

A EUDET Trigger Logic Unit (TLU) provides timestamp information on a particle passage through four trigger devices in coincidence. Two trigger devices are located in front of the first telescope plane and two of them behind the last plane. Each trigger device is built up by a 3 mm thick and 2 cm × 1 cm scintillator matching the MIMOSA26 sensor area and attached to a photomultiplier tube. The TLU and the DAQ of the MIMOSA26 sensors communicate in a handshake mode, so that if the TLU asserts a trigger, the MIMOSA26 DAQ raises a busy signal during readout of the frame. Additional sensors or DUTs can be integrated in the same way or in a no-handshake mode, only accepting triggers.

The entire hardware is integrated in the EUDAQ data acquisition framework which merges data streams of all components as event-based data. EUDAQ version 1 works for synchronous DAQ systems requiring one event per trigger. Thus, the trigger rate is limited by the slowest device. Running only the telescope without any DUT results in an event rate of 2.0 kHz at a test beam particle rate of about 10kHz. Users can integrate the DAQ of their DUT by writing a EUDAQ component which matches a defined interface.

DUTs are mechanically integrated between the two telescope arms at a x-, y-, ϕ-stage



(a) Configuration (b) Resolution

Figure 3.21: (a): Sketch of the standard telescope geometry and definition of important parameters. (b): The calculated track resolutions at the DUT for two geometries are shown at a particle momentum of 5 GeV/c. ϵ_{DUT} is defined as the DUT thickness normalized to its radiation length.

system providing a micron precision. This allows a geometrical scan of the DUT response in respect of the particle tracks, which is larger than the 2 cm $\times$ 1 cm active area of the telescope.

A complete description of the DESY Test Beam Facility can be found in[39].

Chapter 4

The new sensors

In this chapter test beam studies on planar and 3D sensors, before and after irradiation, will be presented. They will be grouped on the basis of the readout chip (ROC) bonded to the sensors.

4.1 Introduction

The purpose of the research project to which I contributed during my PhD is the realization of a new ultra-radiation-resistant Silicon pixel sensor for the CMS experiment in view of the future high luminosity phase of the LHC collider (HL- LHC).

One of the most critical geometrical parameters in the development of these sensors is the distance between the electrodes that generate the electric field for charge collection. It is well known that in order to operate these sensors at high irradiation fluences, the input of the pre-amplifier should be connected to the electrode which collects electrons (the faster carriers). Furthermore, in order to keep the bias voltage as low as possible while preserving the largest part of the signal, the distance between opposite sign electrodes should not exceed a few times the electrons' mean-free-path at saturation velocity. To achieve this goal, it was decided to develop a new type of thin silicon sensors, both planar and 3D.

The best choice is a n+ on p sensor which avoids type-inversion of the bulk and is less expensive than a n+ on n since it allows for a single-sided process, with both the pixel implants and the guard-rings on the same side.

In the considered HL-LHC scenario, at the highest fluence ($2.3 \times 10^{16} n_{eq}/cm^2$ for the first layer of the pixel detector) the expected electron lifetime becomes ≈ 0.3 ns and the mean-free-path, at saturation velocity, $\approx 30\ \mu m$ [40] (the mean-free-path of the holes is shorter, hence their contribution to the signal is even smaller).

Two different technological solutions are available: planar sensors, where the electrodes are parallel to the sensor surface, and 3D sensors, where the electrodes are orthogonal to the sensor surface. In the first case the distance between the electrodes is fixed by the sensor's active layer thickness, in the second case it is limited by the layout and the technological process used to build the sensor.

To keep the pixel occupancy at per mille level at the expected HL-LHC peak luminosity of $7.5 \times 10^{34}\, cm^{-2} s^{-1}$, and to improve the spatial resolution, the foreseen pixel cell size is of $25 \times 100\ \mu m^2$ or $50 \times 50\ \mu m^2$.

A joint ATLAS-CMS INFN group is collaborating with the Fondazione Bruno Kessler (FBK) foundry (Trento, Italy), to develop thin planar and 3D silicon pixel sensors on 6"

n+ on p Float Zone (FZ) wafers. Both developments employ a recent technology, called Direct Wafer Bonding (DWB), where a high resistivity p-type wafer is directly bonded with a low resistivity p+ wafer. The former constitutes the active part of the sensor, while the latter provides mechanical support and Ohmic contact for the other wafer[1].

The dimensions and granularity of the prototype sensors from the first productions were designed to be compatible with the readout chip currently employed in CMS, namely the PSI46 digital chip [41], which consists of a matrix of 80 rows and 52 columns of $100 \times 150\,\mu m^2$ pixels.

The thin planar sensors (Fig. 4.1) are produced with two nominal active layer thicknesses, 100 μm and 130 μm. The measured thickness is about 10 μm smaller than the nominal one due to Boron diffusion from the underlying p-type layer [42]. Some sensors have been fabricated with bias punch-through structures, as shown in Fig. 4.2, to investigate their impact on performance. Several variants of pixel isolation techniques are implemented. They can be p-spray only or p-spray with p-stop. In addition, the p-spray concentration can be Low, Medium and High, while the p-stop rings can be Open or Closed. To properly characterize the isolation performance of these structures, the sensors should be irradiated at a much higher fluence than those tolerated by the PSI46dig chip ($5 \times 10^{15}\ n_{eq}/cm^2$).

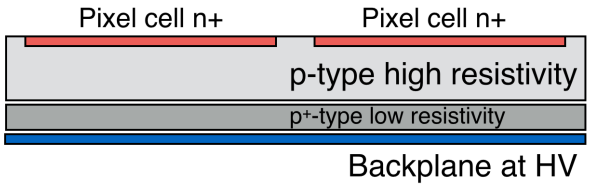


Figure 4.1: Sketch showing the cross section of a thin planar n+ on p sensor. The thickness of the high resistivity layer (i.e. active layer) can be 100 or 130 μm, while that of the low resistivity layer can range from 185 to 50 μm after thinning. The two layers are bonded together with the Direct Wafer Bonding technique.

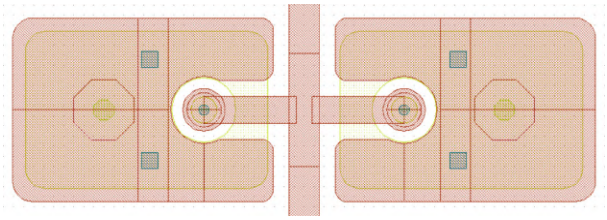


Figure 4.2: Drawing of the bias punch-through structure. The bias line runs between two adjacent pixel columns and biases the punch-through dots of the nearby pixel cells.

The 3D silicon sensors are made with a single-sided process, optimized by FBK [42] and sketched in Fig. 4.3. Two types of columnar electrodes are implemented: p+ Ohmic columns, which terminate in the underlying layer (i.e. the low resistivity one) in order to be biased, and n+ junction columns, which end $\approx 20\ \mu m$ before the low resistivity layer. The nominal column diameter is $\approx 5\ \mu m$ for both junction and Ohmic columns. Sensor modules are produced with three different pixel sizes and different numbers of junction/Ohmic columns [43]. Standard sensors, i.e. those with $100 \times 150\,\mu m^2$ pixel size

[1]Wafers produced by IceMos Technology, Belfast, UK.

and therefore fully compatible with the readout chip, come in two flavors: with three and two junction columns, Fig. 4.3. By convention, these two types of pixel cells are called 3E and 2E, respectively, where E stands for readout Electrodes.

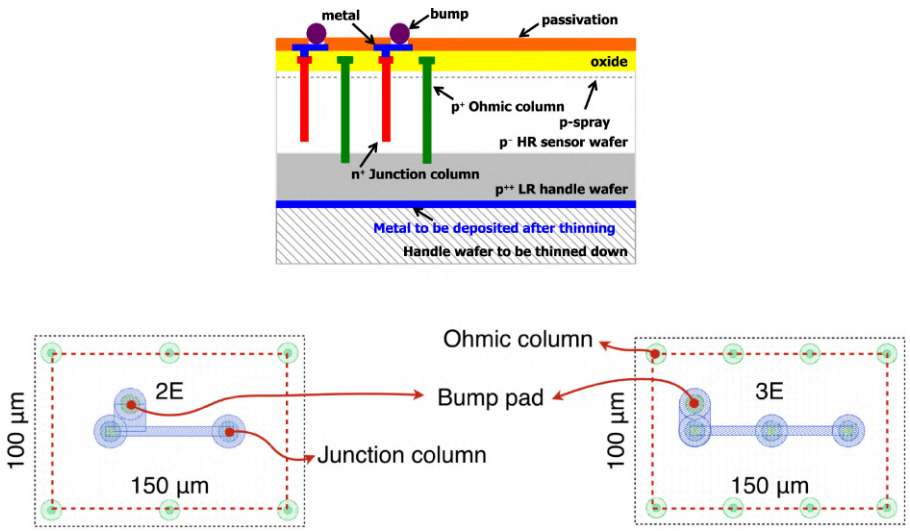


Figure 4.3: The top figure sketches out the production process of the 3D silicon pixel sensors. The high resistivity layer (i.e. active layer) thickness is 130 μm, while the low resistivity layer thickness is $\approx$ 500 μm. The two layers are bonded together with the Direct Wafer Bonding technique. The layouts of the 100 $\times$ 150 μm^2 pixel cells with two (2E, left) and three (3E, right) junction columns are shown at the bottom. The Ohmic columns are in light green at the periphery of the pixel cells.

The small-pitch sensors, instead, can have 25 $\times$ 100 μm^2 or 50 $\times$ 50 μm^2 pixel size: squared pixels can be of type 1E only, while rectangular pixels can be of type 1E or 2E. In each cell the bump pad is typically located next to one of the junction columns except for a variant of the 25 $\times$ 100 μm^2 sensor where it is placed on top of a junction column. This is to free space between the junction and Ohmic columns thus reducing the risk of short circuits. While standard sensors can be fully read out, only one sixth of the pixels of the other sensors can be read out with the PSI46 digital chip. In addition, the small-pitch sensors require additional circuits to match the readout chip input pad and to bias pixels that are not read out (Fig. 4.4).

Fig. 4.5 shows a detail of a single cell of the 3D small pitch sensors: it is evident that the distance between electrodes is much reduced with respect to the thin planar sensors.

Sensors coming from the most recent productions were designed to be bonded to the RD53A [17] ROC: this is the first prototype of pixel ROC for HL-LHC operations. It has been designed by a CERN ATLAS-CMS collaboration named RD53 to demonstrate in a large format IC the suitability of the chosen 65 nm CMOS technology for HL-LHC upgrades of ATLAS and CMS, including radiation tolerance, stable low threshold operation, and high hit and trigger rate capabilities. RD53A is not intended to be a final production IC for use by the experiments, and contains design variations for testing purposes, making the pixel matrix non-uniform. Specifically three different front-end schemes are available on this chip, they are called Differential, Linear and Synchronous. The RD53A pixel matrix is 400 pixels wide by 192 pixel tall and the cell size is $50 \times 50 \mu m^2$. The power and bias distribution have been designed for a larger number of rows, up to 384, giving the possibility to bond to this ROC also the 25 $\times$ 100 μm^2 sensors. The

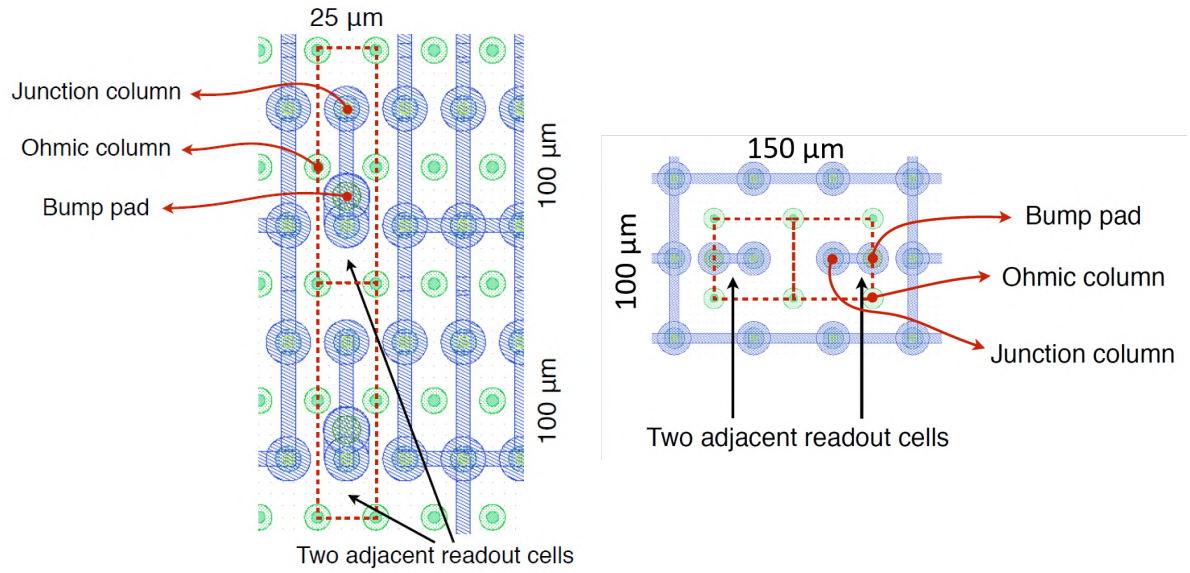


Figure 4.4: Layout of the $25 \times 100\,\mu m^2$ (left figure) and $50 \times 50\,\mu m^2$ (right figure) pixel cells. Only the cells pointed by the black arrows are readout. The blue layer represents a metallization grid need to short-circuit the pixels that are not readout. The Ohmic columns are shown in light green at the periphery of the cells.

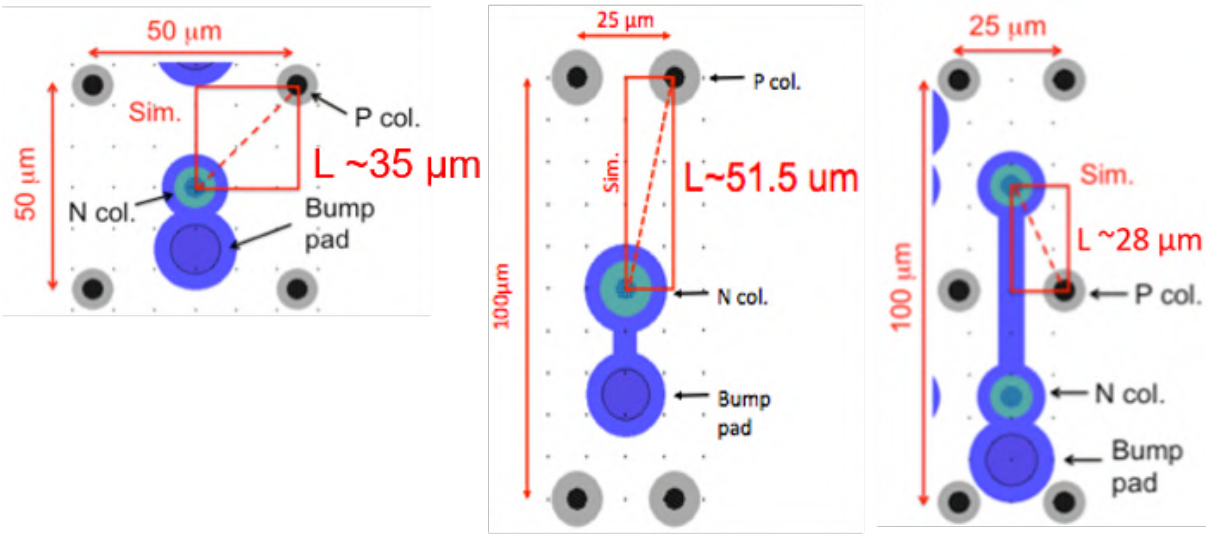


Figure 4.5: Detail of a single cell of the three different designs of small pitch 3D sensor described in the text.

bonding schemes for $25 \times 100\,\mu m^2$ and $50 \times 50\,\mu m^2$ sensors are showed in Fig. 4.6. This ROC allowed to irradiate sensors up to the fluence of $10^{16}\ n_{eq}/cm^2$ and to readout each cell of $25 \times 100\,\mu m^2$ and $50 \times 50\,\mu m^2$ sensors.

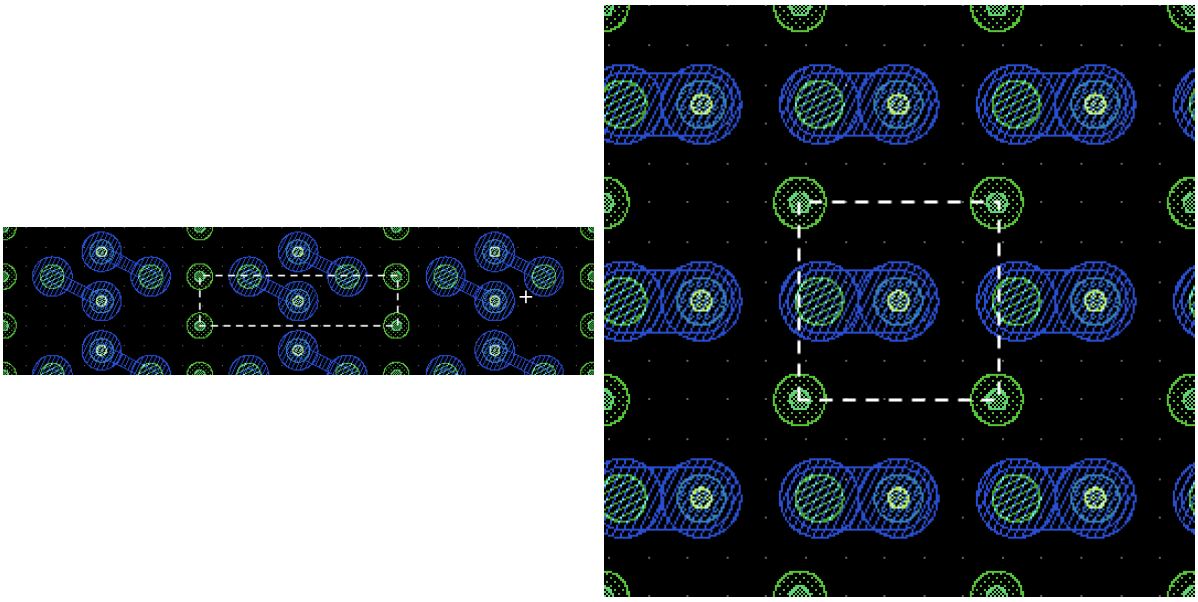


Figure 4.6: Layout of the pixel cells of a $25 \times 100\,\mu m^2$ 1E (left figure) and $50 \times 50\,\mu m^2$ (right figure) sensors bonded to the RD53A ROC. A single cell is highlighted.

The employed ROCs perform zero-suppressed readout using a comparator. A global threshold can be programmed for the whole chip. In order to compensate for local transistor mismatches each pixel has a 4-bit DAC to trim the threshold. Furthermore, a mask bit allows to disable noisy pixels. In the PSI46dig ROC the signal from the sensor is processed by a two stage charge sensitive pre-amplifier/shaper system. An 8-bit successive approximation ADC digitizes the analog pulse height information in the ROC periphery. In the RD53A ROC a single amplification stage has been chosen due to power consumption and area constraints. The signal from the CSA is fed to a current comparator that, combined with a time-over-threshold(ToT) counter, is exploited for time-to-digital conversion.

Both planar and 3D sensors from the two productions have been characterized on beam before and after irradiation. In this thesis studies based on data acquired at the Fermilab Test Beam Facility using 120 GeV protons and DESY Test Beam Facility using 5.2 GeV electrons will be reported. It is not guaranteed that the same sensor was tested before and after irradiation and there may be several reasons for this:

- The detector was not irradiated (given the limited number of irradiation slots a choice had to be made)
- The detector was sent to an irradiation facility before any test on beam
- The calibration of the ROC after irradiation was unreliable (PSI46dig only)

Three different facilities were used for irradiation:

- Los Alamos National Laboratory (800 MeV protons)

- CERN PS IRRAD (24 GeV protons)
- KIT (25 MeV protons)

This chapter is organized in the following way:

- Section 1: Planar sensors bonded to the PSI46dig ROC, before irradiation
- Section 2: Planar sensors bonded to the PSI46dig ROC, after irradiation
- Section 3: 3D sensors bonded to the PSI46dig ROC. before and after irradiation
- Section 4: 3D sensors bonded to the RD53A ROC, before and after irradiation

4.2 Planar sensors before irradiation

In this first section the performance of planar sensors before irradiation will be described: to this extent five sensors have been selected among the tested ones. Their main characteristics are highlighted in Fig. 4.7.

ID Chip	Active Thickness (µm)	P-stop (around cell)	n.GR/P-stop between GR	Punch-Through
43B	100	yes	10/yes	yes
45B	100	no	1/yes	yes
23C	130	no	1/yes	yes
42C	130	yes	10/yes	yes
53B	130	yes	10/yes	yes

Figure 4.7: Summary table containing the relevant characteristics of the planar sensors tested before irradiation.

4.2.1 Detection efficiency

The detection efficiency is computed as the ratio between the number of incident tracks associated to a hit on the DUT and the total number of incident tracks. In order to avoid a critical dependence of the efficiency on the resolution of the impact point a fiducial region must be defined. Only tracks pointing to a pixel cell surrounded by 8 working pixel cells are considered in the efficiency computation: in this way the periphery cells and the dead cells are excluded. A pixel cell is considered working if it registered at least one hit during the run. This information is stored in the histogram reported in Fig. 4.8 for one of the sensors.

The detection efficiency of the five planar sensors at a bias voltage of 40 V resulted to be:

- 43B: 99.88%
- 45B: 98.38%
- 23C: 99.84%

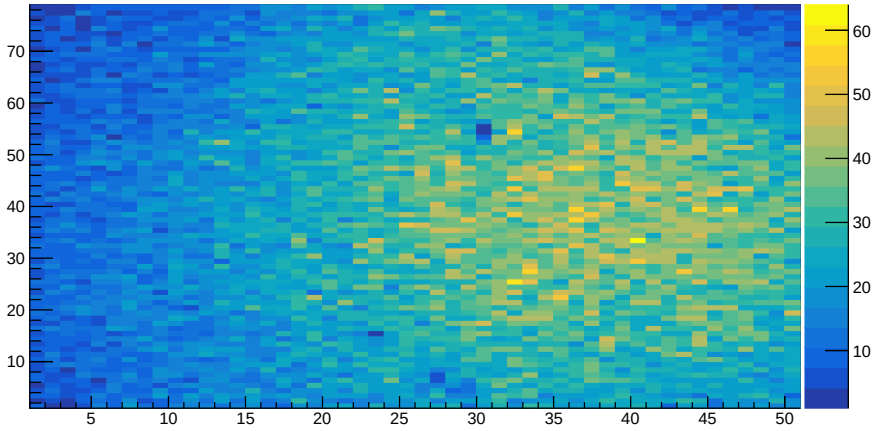


Figure 4.8: Number of hits registered by each pixel cell of the 43B sensor during a run acquired at 40 V bias voltage.

- 42C: 99.71%
- 53B: 99.90%

The difference between the 43B and 45B sensors which have the same active thickness (100 μm) is clearly visible by comparing Fig. 4.9 and Fig. 4.10 which reports the efficiency map of the sensor and the efficiency as a function of the track impact point coordinates on the pixel cell.

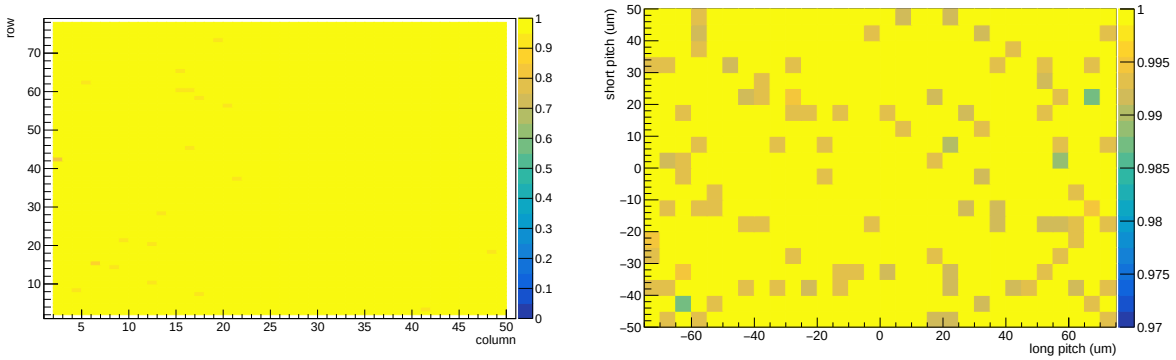


Figure 4.9: Efficiency map of the sensor (left) and efficiency as a function of the track impact point coordinates on the pixel cell (right) for the 43B sensor at a bias voltage of 40 V.

It is evident that on the pixel cells of the 45B sensor there are regions where the efficiency is strongly reduced. These regions can be investigated by separating the cells belonging to even and odd columns and by comparing the histograms in Fig. 4.11 with the design of the pixel cells in Fig. 4.12.

The efficiency results to be reduced in the region of the punch through structures. This effect is not observed in the 43B sensor, as can be seen in Fig. 4.13 and Fig. 4.14.

The reason for such difference is the different threshold applied to the sensors: it was set to a higher value for the 45B sensor. As a consequence the amount of charge collected

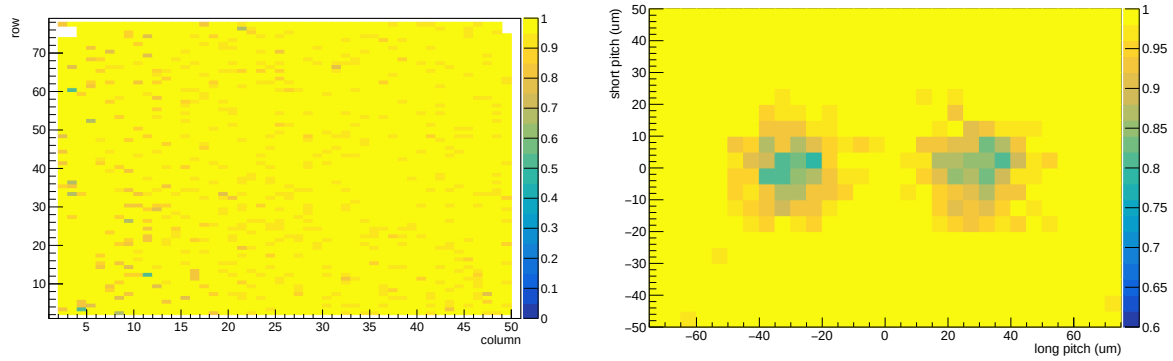


Figure 4.10: Efficiency map of the sensor (left) and efficiency as a function of the track impact point coordinates on the pixel cell (right) for the 45B sensor at a bias voltage of 40 V

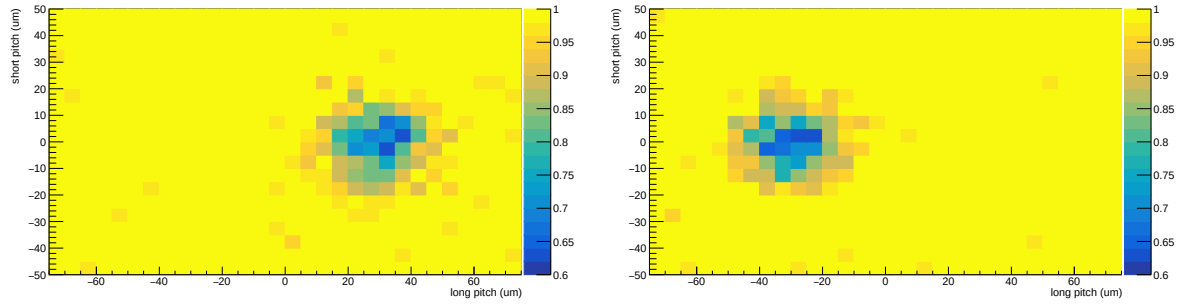


Figure 4.11: Efficiency as a function of the track impact point coordinates on the pixel cells of even (right) and odd (left) columns of the 45B sensor at a bias voltage of 40 V.

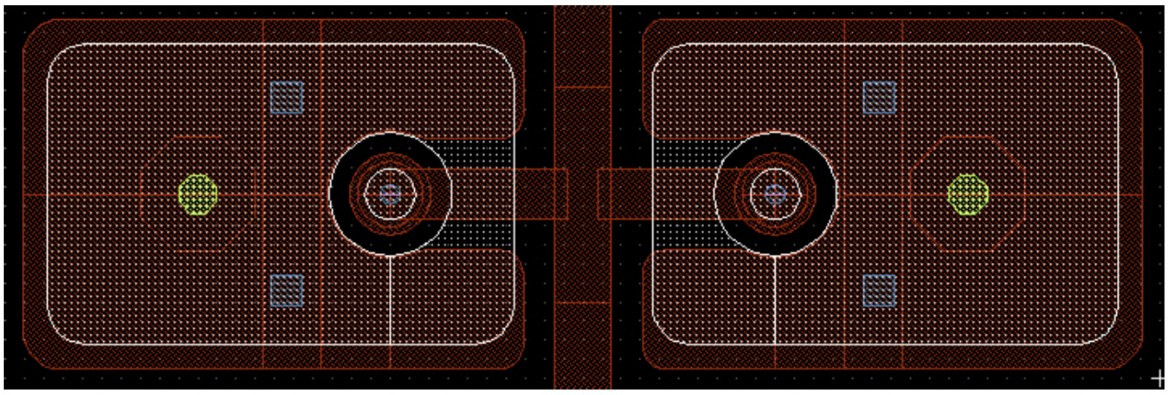


Figure 4.12: Schematic representation of two neighboring pixel cells of the 45B sensor.

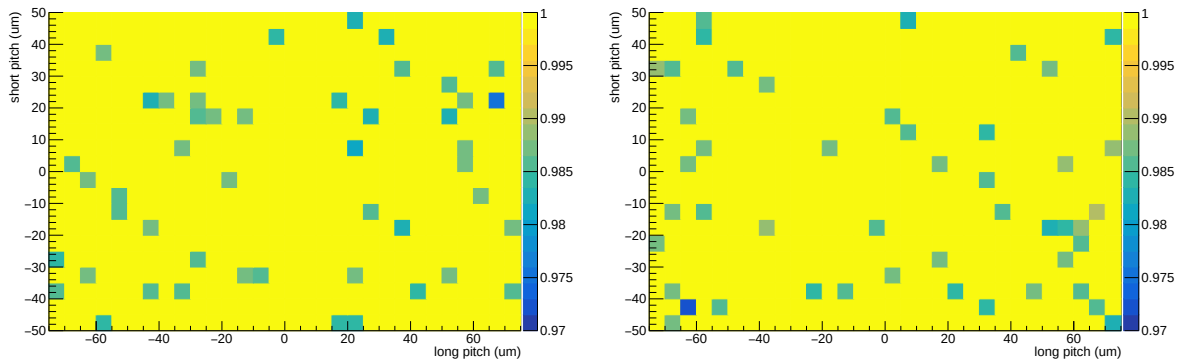


Figure 4.13: Efficiency as a function of the track impact point coordinates on the pixel cells of even (right) and odd (left) columns for the 43B sensor at a bias voltage of 40 V.

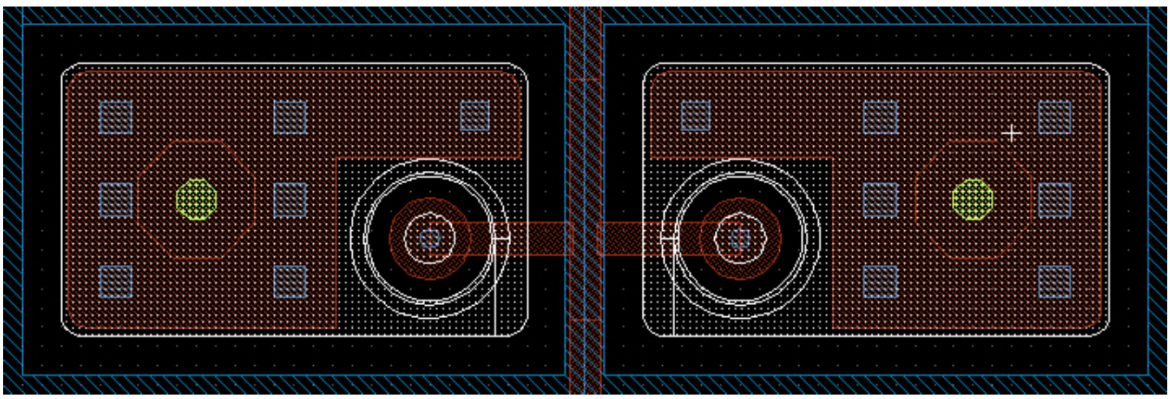


Figure 4.14: Schematic representation of two neighboring pixel cells of the 43B sensor

in the punch through region was not sufficient to overcome this threshold and resulted in a local inefficiency.

These efficiency losses are mitigated when the sensor is tilted. Fig. 4.15 and Fig. 4.16 reports the same plot of Fig. 4.11 for tracks impacting the sensor at an angle of 10 and 20 degrees, respectively. The detection efficiency resulted to be 98.88% at 10 degrees and 99.71% at 20 degrees, compatible with the ones of the sensors where the losses are not visible at normale incidence.

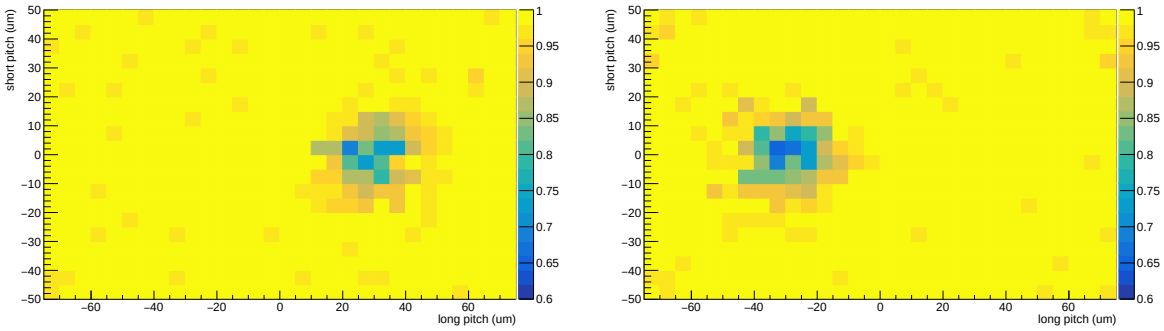


Figure 4.15: Efficiency as a function of the track impact point coordinates on the pixel cells of even (right) and odd (left) columns for the 45B sensor at a bias voltage of 40 V and a tilt of 10 deg.

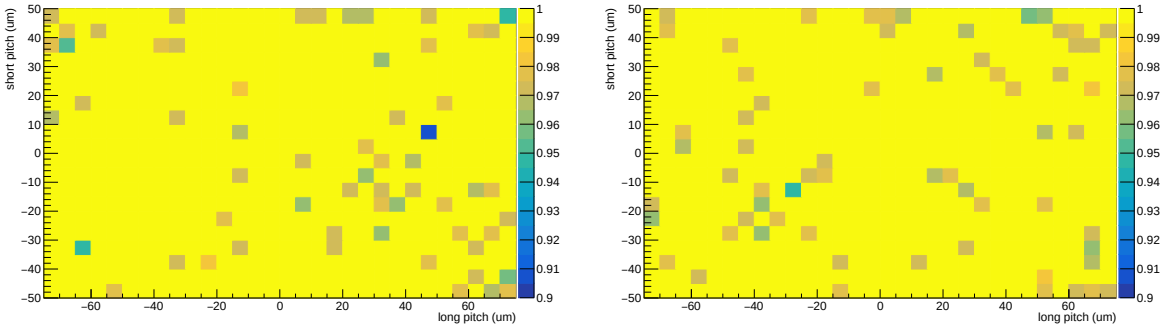


Figure 4.16: Efficiency as a function of the track impact point coordinates on the pixel cells of even (right) and odd (left) columns for the 45B sensor at a bias voltage of 40 V and a tilt of 20 deg.

The detection efficiency can be studied more in detail considering two neighboring cells along a row or a column. A fiducial window must be defined in order to get rid of tracks that could possibly generate a hit on a third cell due to charge sharing. Tracks are selected according to the following criteria:

- associated cluster of size 1 and 2 only
- associated cluster of size 2 along a row or a column
- impact point distant more than 20 μm from the periphery of the pixel cell in the y(x) direction for pixel adjacent along the y(x) direction

The efficiency as a function of the distance of the track impact point from the divide between the two pixel cells is reported in Fig. 4.17 for the 43B sensor at a bias voltage of 40V. Histograms in blue are obtained selecting the tracks with an associated hit on the DUT in the pixel pointed by the track. This results in a drop of the efficiency near the divide which is determined by the the charge sharing between adjacent cells and by the resolution of the track impact point. Once this request is relaxed the hit associated to the track can be located in any of the two neighboring cells and the efficiency becomes uniform along the coordinate (red histograms).

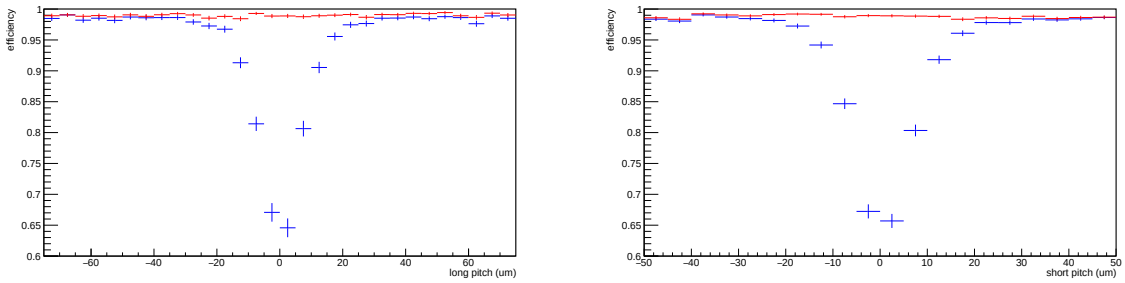


Figure 4.17: Efficiency as a function of the track impact point position on the pixel cell in X and Y directions for the 43B sensor at a bias voltage of 40 V. A detailed description can be found in the text.

In conclusion: the efficiency of the sensors with punch through structures is greater than 98% and becomes compatible with one of the sensors without punch through (greater than 99%) when the sensors are tilted by 20 degrees, as expected from geometrical considerations. Sensors produced for the RD53A ROC are equipped with smaller punch through structures shared between adjacent pixel cells in order to minimize the losses.

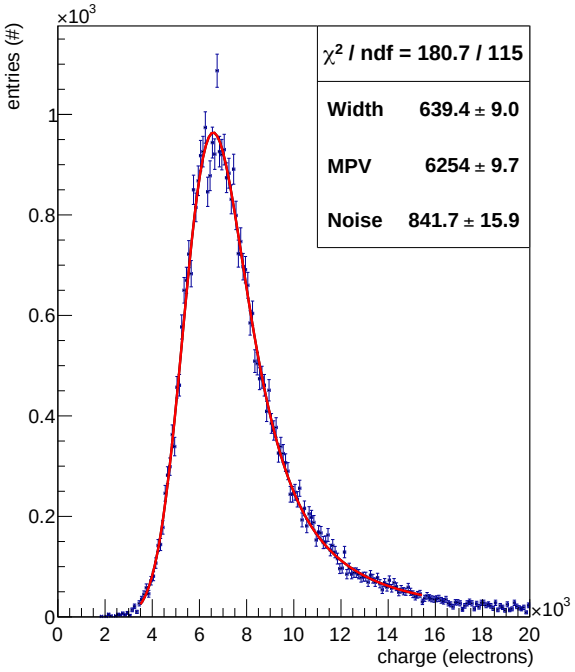
4.2.2 Charge collection properties

The distributions of the charge collected by the sensors are reported in Fig. 4.18: they are obtained considering only tracks pointing to the central region of a pixel cell in order to get rid of signal losses due to charge sharing between adjacent cells. The following selections are applied to the tracks:

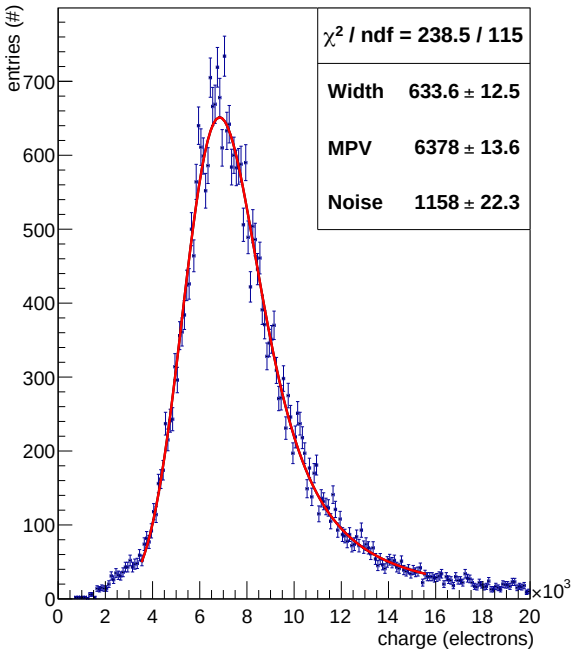
- associated cluster on the DUT must have size 1
- impact point on the pixel cell must be far from the periphery of the cell by more than 20 μm. This value corresponds to 2.5 times the error associated to the track impact point coordinates ($\approx$ 8 μm for the pixel telescope)

The fit function in red is a Landau convoluted with a Gaussian which takes into account the electronic noise and other dispersive effects like the inter-calibration errors. The spectra reported in the figure are the result of a superimposition of the signals collected in all the pixel cells and hence require a precise calibration of the gains of the ROC ADCs. This is a common practice in the study of the charge collected by segmented detectors where it is not possible to consider a single cell for evident statistic reasons.

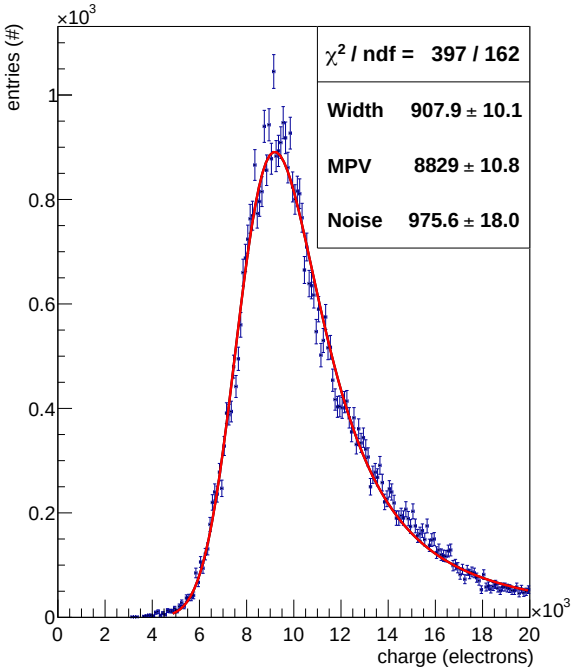
The bias voltage has been chosen according to Fig. 4.19 where the Most Probable Value (MPV) of the collected charge is reported as function of the bias voltage. The error



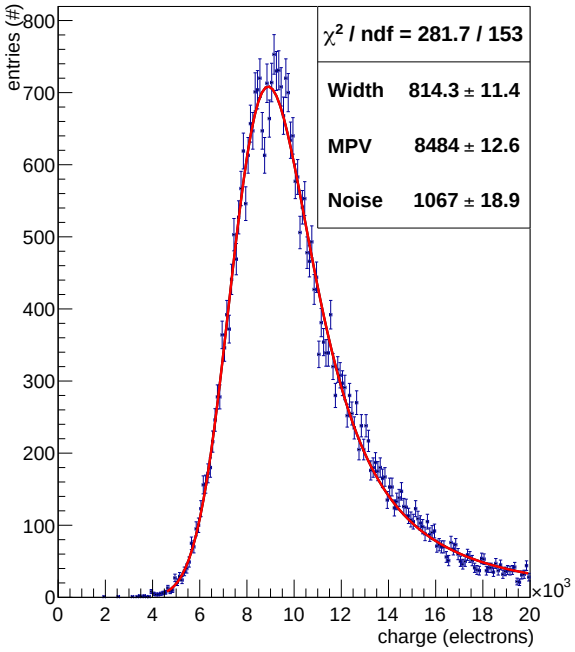
(a) 43B - Vbias = 70 V



(b) 45B - Vbias = 70 V



(c) 23C - Vbias = 60 V



(d) 42C - Vbias = 60 V

Figure 4.18: Distribution of the charge collected by the four planar sensors before irradiation.

bars represents the systematic error associated to the pixel inter-calibration which has been measured to be 5% of MPV. This error is evaluated imposing statistical consistency ($\chi^2/N = 1$) of a sample of N MPVs measured in each quadrant of a sensor and for M sensors of the same thickness (N = 4 × M). The statistical error is negligible since several thousands of tracks are used to produce the distributions. The expected MPVs for the signal released by an orthogonally incident Minimum Ionizing Particle (MIP) are about 6350 e- and 8740 e- for 100 μm and 130 μm thick sensors, respectively [44]. They are also reported in Fig. 4.19 as dashed lines. The collected charge is compatible with expectations and it is worth noting that the it reaches a plateau at lower bias voltages in the sensors with p-stop implantation around the pixel cells.

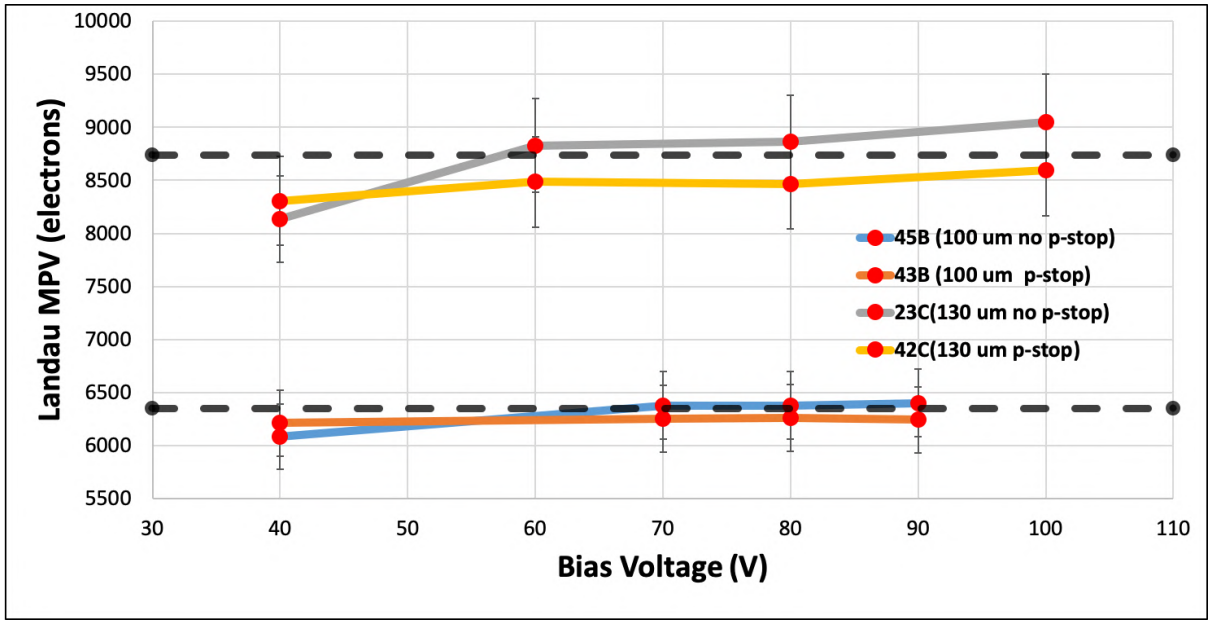


Figure 4.19: MPV of the collected charge as a function of the bias voltage for the four planar sensors.

For one of the sensors such bias voltage scan was carried out starting from lower values, as can be seen in Fig. 4.20, in order to determine a reasonable operating point for the other sensors.

The collected charge of the 43B and 45B sensors as a function of the track impact point position on the pixel cells are reported in Fig. 4.21 and Fig. 4.22 for a single cell and four adjacent pixel cells sharing a corner, respectively. In the periphery, and more clearly in the corners, of the first plots a loss in the collected charge is observed. This is more evident in the second plots and is determined by the charge sharing among the four pixel cells with a common corner.

To investigate the effect of the punch through structures the collected charge must be displayed as a function of the track impact position in pixel cells belonging to even and odd columns. These plots are shown in Fig. 4.23 and a loss can be observed in the punch through regions.

To investigate the charge sharing properties of the sensors, the case of two adjacent cells along the same row or column can be analyzed in detail. As in the case of the histograms of Fig. 4.17 the periphery of the cells must be excluded in order to get rid of contributions from other cells, which may fall below the threshold. The results of these studies are shown in Fig. 4.24, which reports the average charge collected by the single

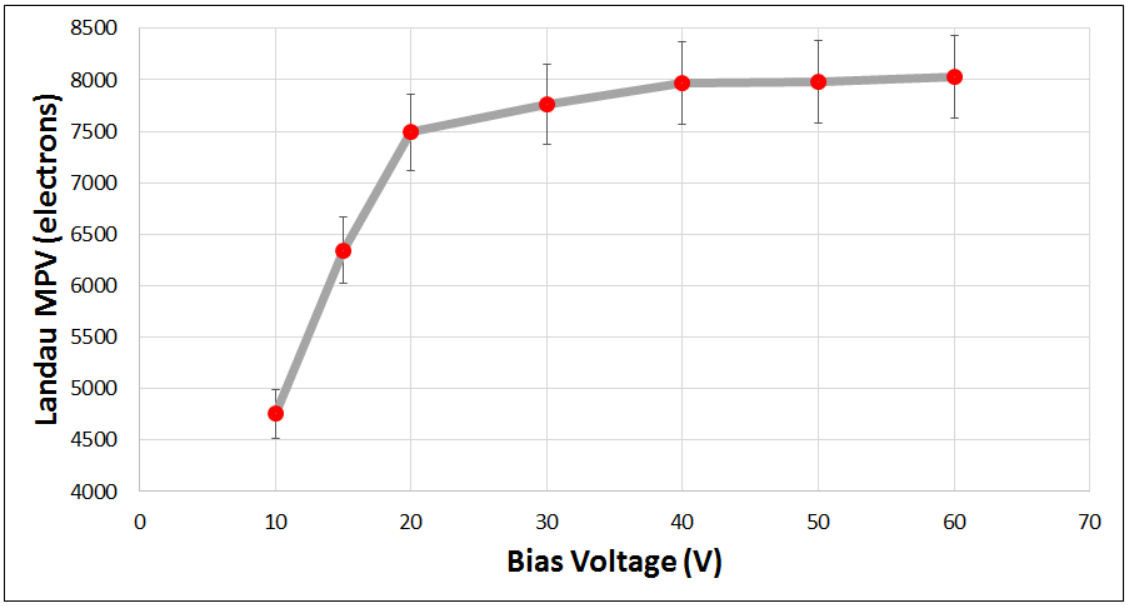


Figure 4.20: MPV of the collected charge as a function of the bias voltage for the 53B planar sensor.

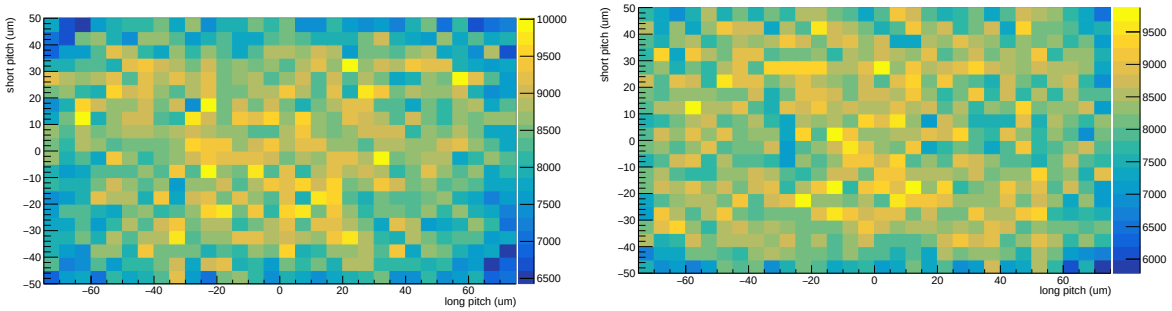


Figure 4.21: Collected charge as a function of the track impact point position on the pixel cells of the 43B and 45B sensors at bias voltage of 40 V.

cell pointed by the track, or by both cells, as a function of the distance of the track impact point from the divide between the two cells. The 43B at a bias voltage of 40 V has been chosen for sake of simplicity.

The two blue histograms in Fig. 4.24 show the average charge collected by the cell pointed by the track: it can be seen that, due to the resolution on the track impact point and/or the charge division, the charge decreases near 0, which corresponds to the divide between two adjacent cells. If the charge collected by the adjacent cell is also taken into account (red histograms), this effect disappears and the value measured at the divide becomes comparable with the one measured at the center of the cell (first and last points in the histograms) where the probability of charge sharing between contiguous cells is negligible.

In order to measure, with a good approximation, the extension of the area in which there is charge sharing between two adjacent pixel cells it is necessary to measure the distribution of the charge asymmetry of the two cells as a function of the distance from

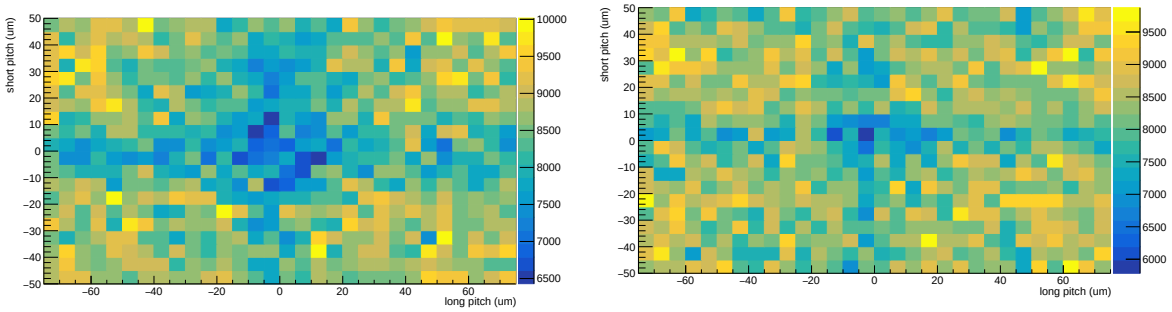


Figure 4.22: Collected charge as a function of the track impact point position on four adjacent pixel cells of the 43B and 45B sensors at bias voltage of 40 V.

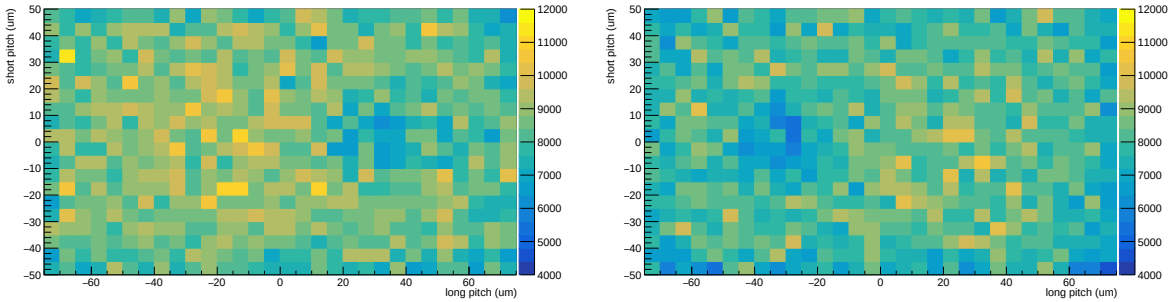


Figure 4.23: Collected charge as a function of the track impact position on the pixel cells of even (right) and odd (left) columns for the 45B sensor at a bias voltage of 40 V.

the divide. The charge asymmetry is defined as $\frac{C_L - C_R}{C_L + C_R}$, where C_L and C_R are the charges collected by the cell to the left and right of the divide, respectively. The selection applied are the same of the previous case but only clusters of size 2 for which the sum of the charges of the two pixels is less than 13600 electrons, that is 1.7 times the relative MPV, are considered. This specific cut on the collected charge is imposed to reduce the contamination due to the emission of δ rays.

The resulting distributions are shown in Fig. 4.25. The two correlation plot in the upper row represent the signed distance of the track impact impact point from the divide and the corresponding asymmetry value for each event while the two histograms in the lower row represent the average value of the distances for each asymmetry bin. Similarly to the previous case, the distributions on the left of the figure are for two adjacent cells on the same row, while those on the right are relative to two cells on the same column.

The histograms of the average of the distances as a function of the asymmetry can be fitted with a linear function to determine the correlation coefficient between asymmetry and distance. This coefficient can be used to predict the distance of the track impact point from the divide once the asymmetry is known. Such linear fit can generally be applied, with good approximation, only in the asymmetry region from -0.5 to 0.5 for tracks orthogonal to the detector. When the detector is tilted the distribution becomes substantially linear over the entire asymmetry range. In both cases the determination of the track impact point position from the asymmetry value results much more precise than other algorithms. Furthermore, even if the linear fit is a good approximation in a region limited both in asymmetry and in total charge, it includes the great majority of cases.

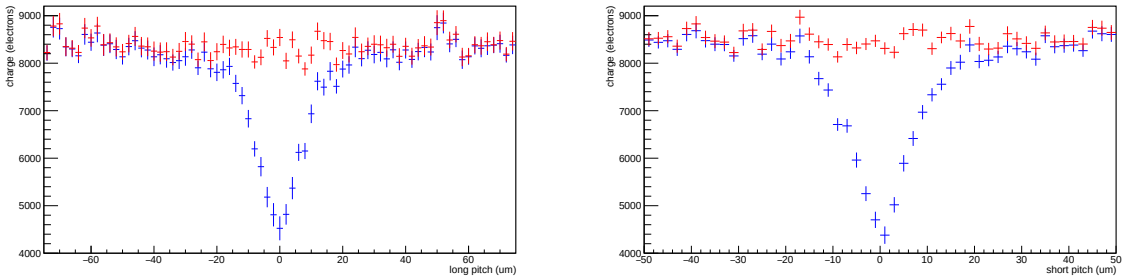


Figure 4.24: Collected charge as a function of the track impact point position on the pixel cell in X and Y directions for the 43B sensor at a bias voltage of 40 V. A detailed description can be found in the text.

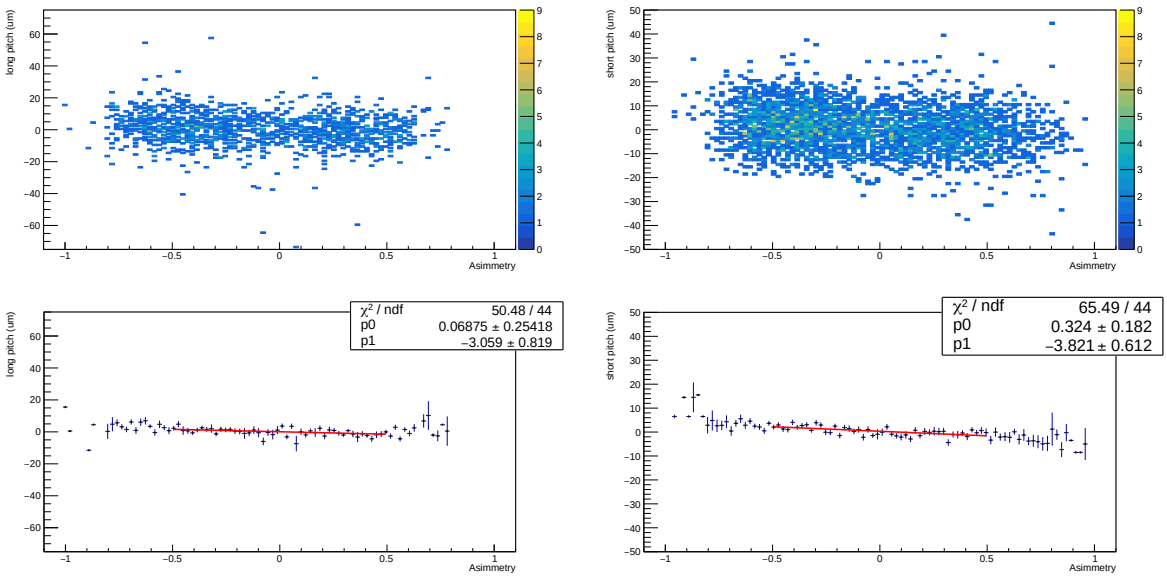


Figure 4.25: Distance of the track impact point from the divide of two pixel cells as a function of the charge asymmetry in X and Y directions for the 53B sensor at a bias voltage of 40 V.

At this point it is useful to introduce a brief digression on how the coordinate to be assigned to a size two cluster is calculated. The charge deposited by a particle that crosses two adjacent pixel cells is directly proportional to the length of the path taken in each cell. Fig. 4.26 shows that this length can be correlated with that of the green segments AB (d_L) and CD (d_R), which represent the distance from the divide of the crossing points of the track in the two cells. Indeed, we have that

$$d_R : Q_R \;=\; d_L : Q_L$$

and

$$d_R + d_L \;=\; w \cdot \tan\theta$$

where w is the active thickness of the sensor. The x coordinate can be expressed as

$$x \;=\; \frac{(x_d - d_L) + (x_d + d_R)}{2}$$

where x_d is the coordinate of the divide. Using this geometric analogy it is possible to express the charge asymmetry, previously defined, as:

$$\frac{2(x - x_d)}{w \cdot \tan\theta}$$

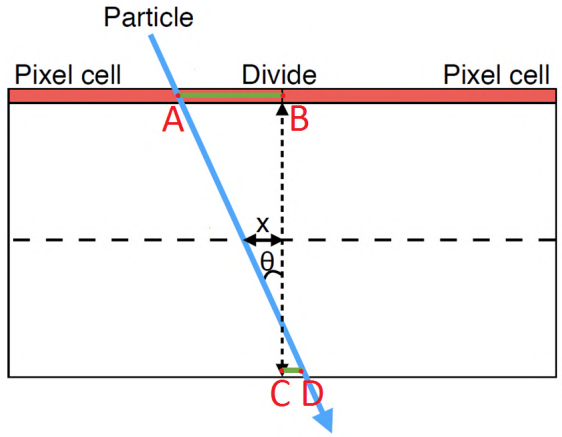


Figure 4.26: Schematic representation of a particle crossing two adjacent pixel cells.

It can be deduced that the slope of the line used to fit the asymmetry plots is equivalent to

$$(w/2) \cdot \tan\theta$$

This means that for a detector with an active thickness of 130 μm tilted by 20 degrees the slope should be 23.7 μm per unit of asymmetry: the result shown in Fig. 4.27 is therefore extremely accurate since the parameter p1, i.e. the slope, is 23.3 μm. The same figure also shows the histogram in the X direction which is compatible with the one obtained in a position orthogonal to the beam.

4.2.3 Resolution

The spatial resolution of the detectors can be estimated by measuring the residuals of the coordinates attributed to the hit or cluster of hits with respect to those of the track impact point. Fig. 4.28 shows the residuals distributions for the sensor 53B.

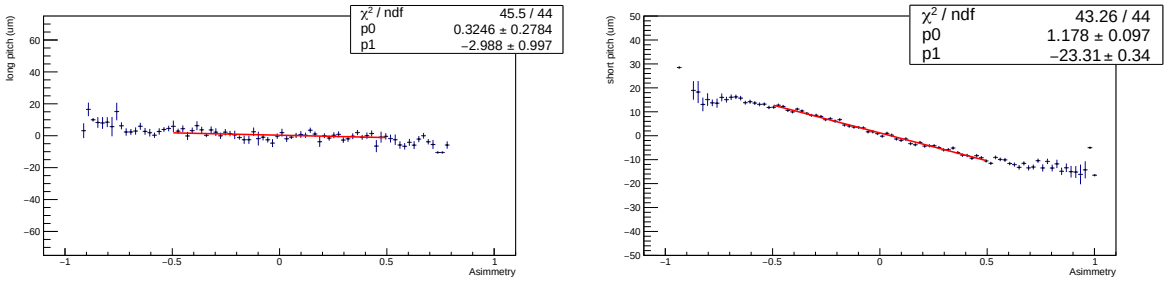


Figure 4.27: Distance of the track impact point from the divide of two pixel cells as a function of the charge asymmetry in X and Y directions for the 53B sensor at a bias voltage of 40 V tilted by 20 degrees.

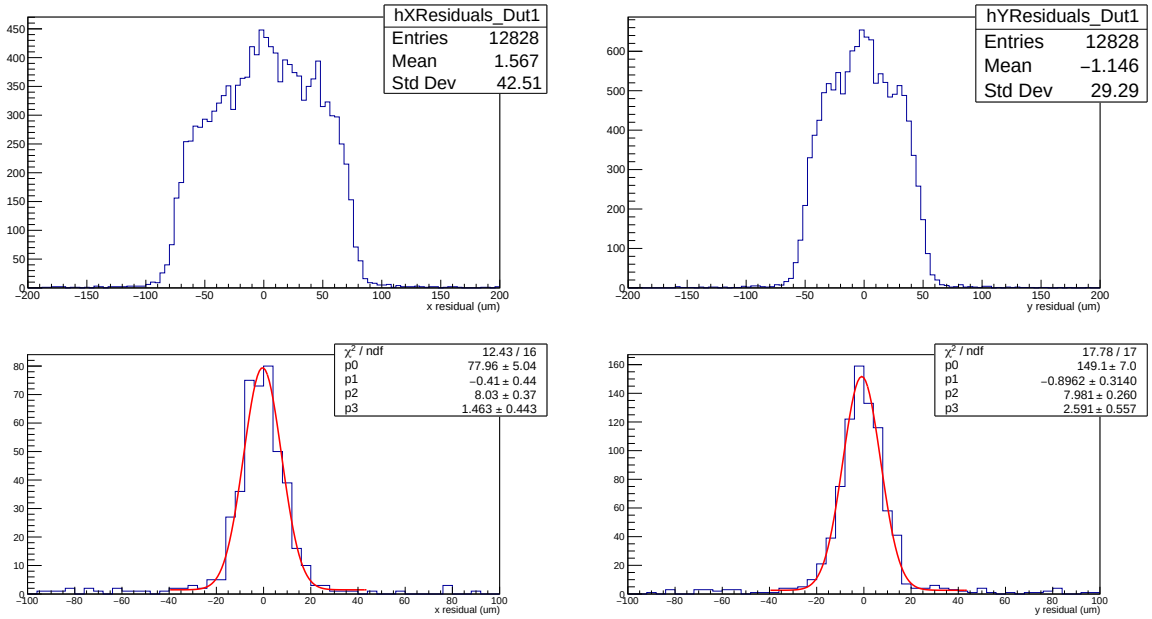


Figure 4.28: Residuals distributions for the 53B sensor at a bias voltage of 40 V. Upper row: clusters with size from 1 to 4. Lower row: size 2 clusters only with associated coordinate extrapolated from the asymmetry fit.

Histograms in the upper row represent the residuals of all the clusters considered on the DUT, from size 1 to 4: clearly the distributions are dominated by single hits which, for tracks orthogonal to the DUT, dominate the statistics (see Fig. 4.29). Consequently the RMS of the residuals distribution in X is substantially the one expected for single hits of 150 μm pitch, that is $150\,\mu m/\sqrt{12} = 43.3\,\mu m$, while the one in Y is very close to $100\,\mu m/\sqrt{12} = 28.9\,\mu m$.

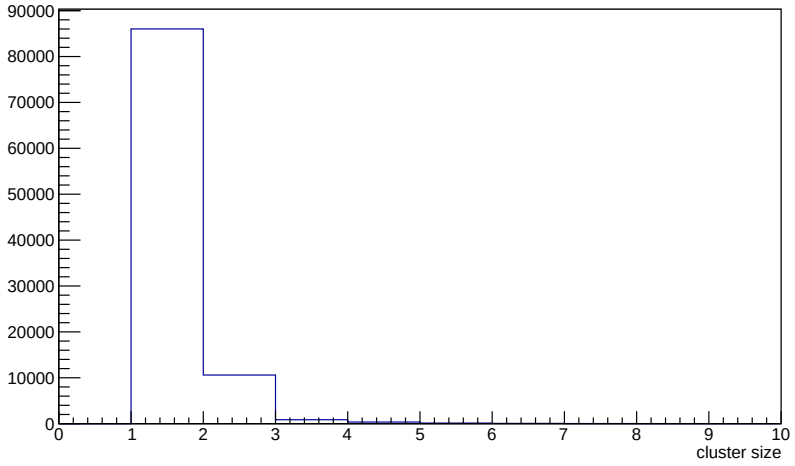


Figure 4.29: Cluster size distribution for sensor 53B at a bias voltage of 40 V.

Histograms in the lower row represent the distributions of the residuals for clusters of size 2 only. These residuals are obtained by attributing to the clusters the coordinate deduced from the existing correlation with the asymmetry.

When the modules are tilted the number of clusters with size 2 increases since there is a larger charge sharing between adjacent pixels along either rows or columns. This can be seen in Fig. 4.30 which reports the cluster size distribution for sensor 53B at a bias voltage of 40 V, tilted by 20 degrees.

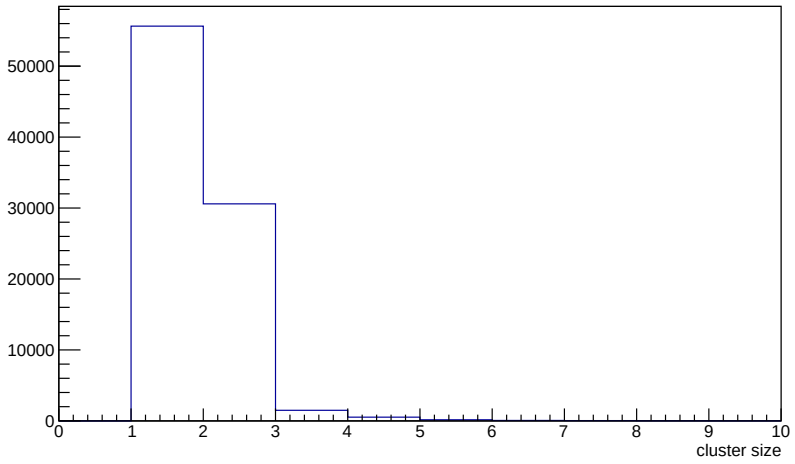


Figure 4.30: Cluster size distribution for sensor 53B at a bias voltage of 40 V and a tilt of 20 degrees.

As a consequence the correlation between charge asymmetry and cluster coordinate

is modified, as shown in Fig. 4.27, and the linear approximation becomes valid in a wider range than -0.5,0.5. The residuals distribution are also modified as shown in Fig. 4.31. The resolution in Y is worse than in the orthogonal position, but the linear approximation applied to the asymmetry plot is valid in a wider range; consequently the coordinate of the cluster, the residual and the resolution are calculated with greater precision in a wider range.

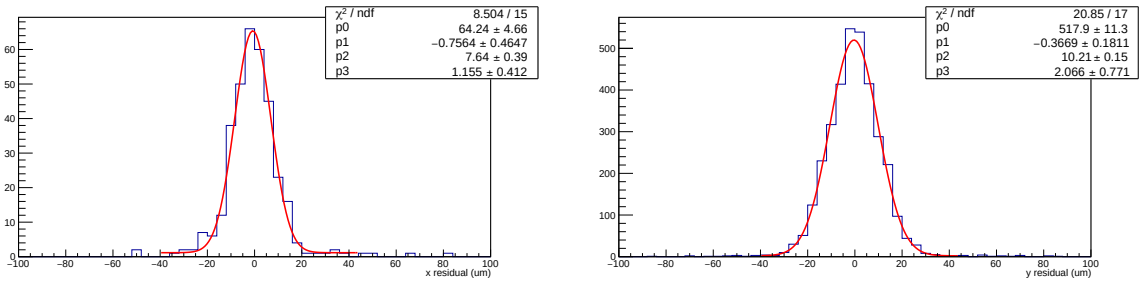


Figure 4.31: Residuals distributions for the 53B sensor at a bias voltage of 40 V and 20 degrees tilt. Only clusters of size 2 are considere and the coordinate is extrapolated from the asymmetry fit as explained in the text.

4.3 Planar sensors after irradiation

Irradiated planar sensors are all bonded to the PSI46dig ROC. Such ROC was designed to be operated up to irradiation fluences of $\approx 5 \times 10^{15}\, n_{eq}/cm^2$. Nevertheless some of the samples irradiated at a fluence of 1, 3 or 5 $\times 10^{15}\, n_{eq}/cm^2$ showed a digital behavior of the ROC analog output during the calibration step: this made the measurements of the collected charge impossible. Four sensors among the ones surviving the irradiation have been chosen, their characteristics are reported in Fig. 4.32.

ID Chip	Active Thickness (µm)	P-stop (around cell)	n.GR/P-stop between GR	Punch-Through	Irradation Fluence (neq/cm2)
62D	100	yes	1/no	no	3,1E+15
31D	100	no	1/no	yes	5,0E+15
53B	130	yes	10/yes	yes	1,2E+15
23C	130	no	1/yes	yes	3,1E+15

Figure 4.32: Summary table containing the relevant characteristics of the planar sensors tested after irradiation.

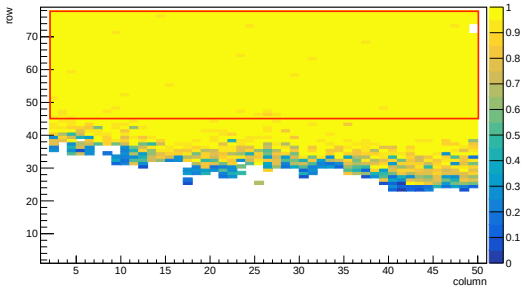
In some cases the irradiation fluence turned out to be not uniform on the sensor surface, as can be seen in Fig. 4.33 where the efficiency map of three sensors are reported. Some of the pixels have been exposed to a fluence such different from the others that the charge signal amplitude is heavily modified. As a consequence the time needed to reach the readout threshold is considerably longer or shorter and the signal cannot be acquired since the readout time window is fixed. This is the reason for the presence of white regions in the histograms. In order to focus on the regions with the highest irradiation fluence only pixels in the red boxes are considered.

The choice for the 31D sensor may seem peculiar but it finds its justification in the aforementioned non uniformity of the charge signal, as showed in Fig. 4.34. It seems that the peak of the irradiation fluence was higher than $\approx 5 \times 10^{15}\, n_{eq}/cm^2$ hence ROC cells in the upper right corner of the detector were exposed to a higher irradiation fluence than the others and were damaged making the analog readout of the collected charge impossible.

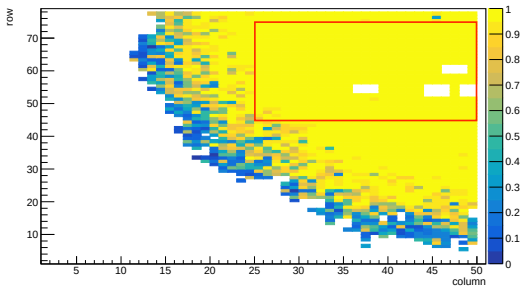
Fig. 4.35 shows the measured particle detection efficiency as a function of bias voltage for the three sensors. The sensors have punch-through structures except the one at the intermediate fluence. The detection efficiency of the sensor irradiated to the highest fluence, i.e. $\approx 5 \times 10^{15}\, n_{eq}/cm^2$, results heavily compromised by the bias punch-through structure and, to a lower extent, by the relatively high threshold set in order to mitigate the noise hit rate (around 2600 electrons).

This is clearly illustrated in Fig. 4.36, which shows the maps of the particle detection efficiency within the pixel cells of the same sensor before the irradiation at a bias voltages of 40 V (first plot) and 150 V (second plot), and, once irradiated, at a bias of 650 V (third plot). The inefficiency, initially limited only to the punch-through dot region, at higher bias voltage starts to affect also the region of the bias grid and, finally, at much higher bias voltage and after the irradiation, it extends to the whole area of the bias grid. This is probably due to the lack of p-stop isolation around the pixel cell implant.

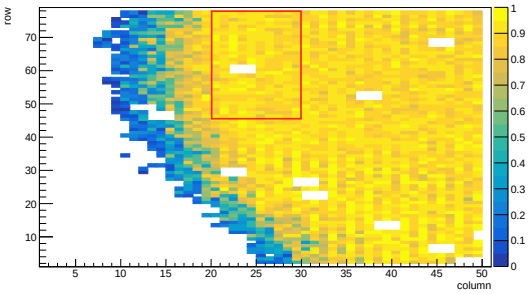
Fig. 4.37 shows the maps of detection efficiency within the pixel cells of the other two sensors at the highest bias voltages, namely 500 V for that irradiated to $\approx 1.2 \times 10^{15}\, n_{eq}/cm^2$ and 800 V for that irradiated to $\approx 3.1 \times 10^{15}\, n_{eq}/cm^2$.



(a) 53B - Fluence $\approx 1.2 \times 10^{15}\ n_{eq}/cm^2$ - Vbias = 500 V



(b) 62D - Fluence $\approx 3,1 \times 10^{15}\ n_{eq}/cm^2$ -Vbias = 800 V



(c) 31D - Fluence $\approx 5 \times 10^{15}\ n_{eq}/cm^2$ - Vbias = 600 V

Figure 4.33: Efficiency maps of three irradiated sensors. Pixels selected in the analysis are contoured by a red box.

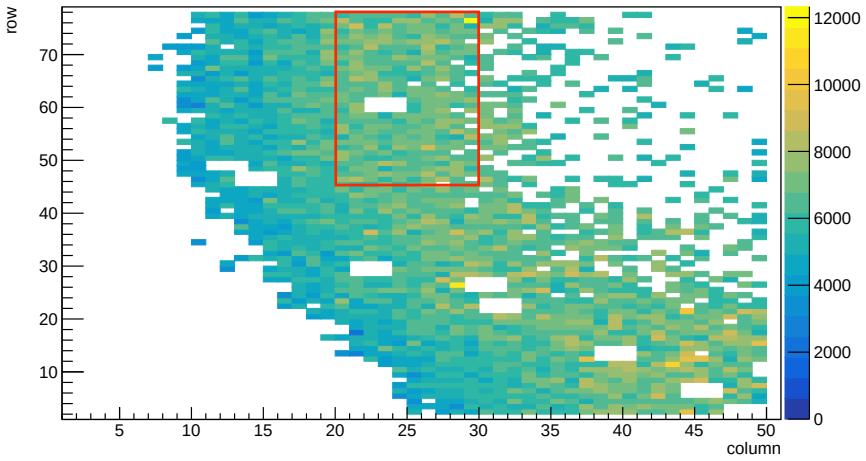


Figure 4.34: Charge map of the 31D sensor at a bias voltage of 600 V. Pixels selected in the analysis are contoured by a red box.

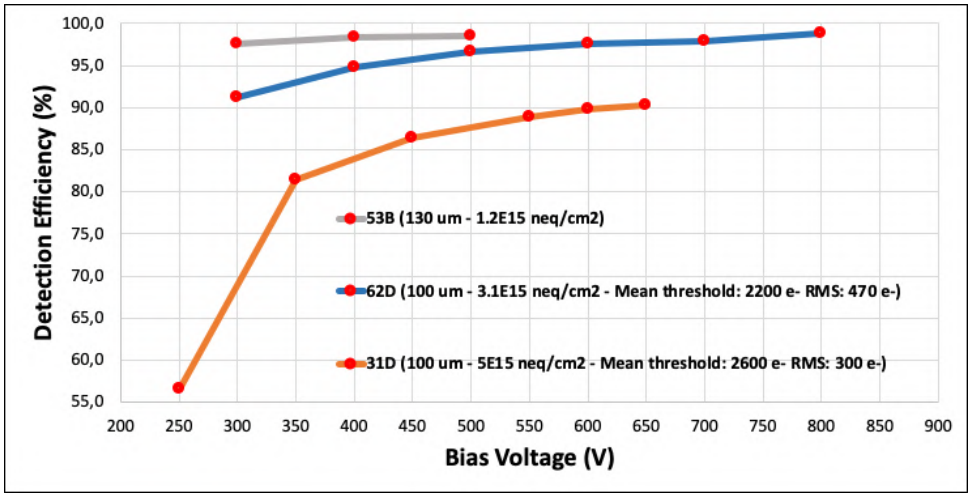


Figure 4.35: Particle detection efficiency as a function of the bias voltage for irradiated thin planar sensors. All sensors have bias punch-through structures except the one at the intermediate fluence.

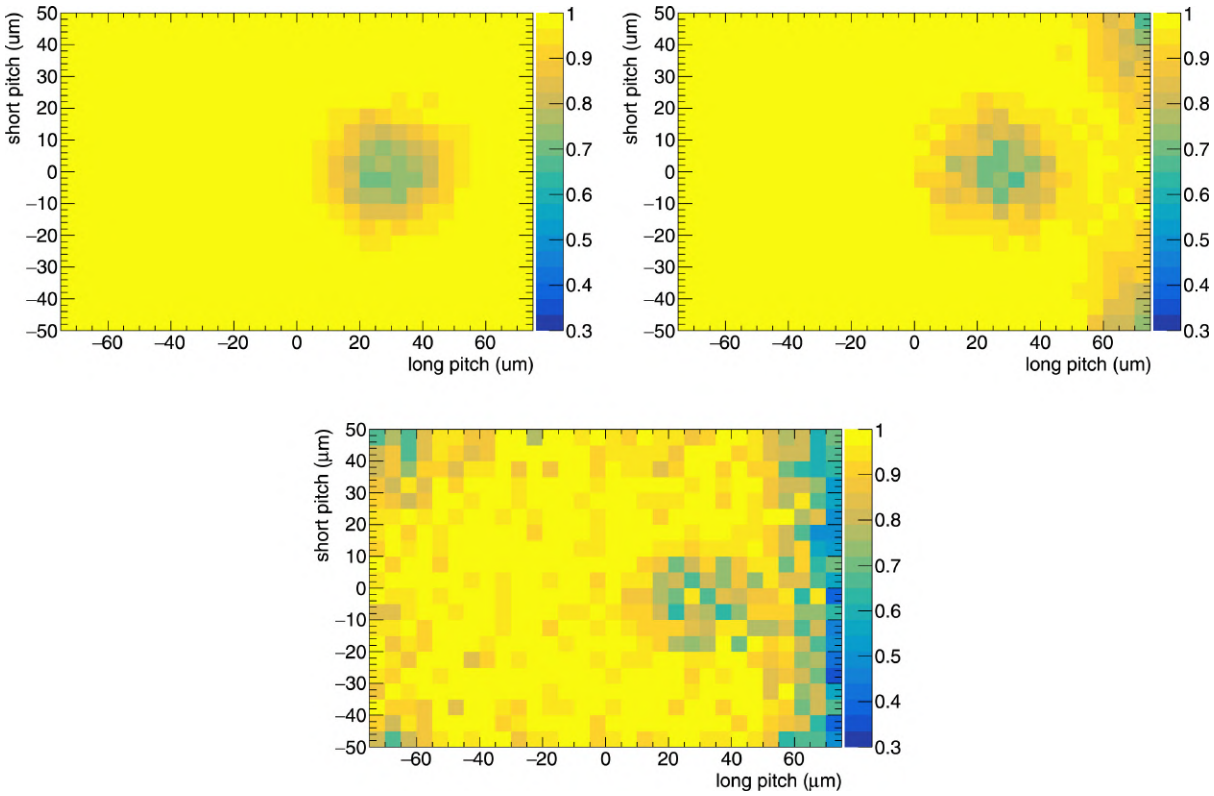
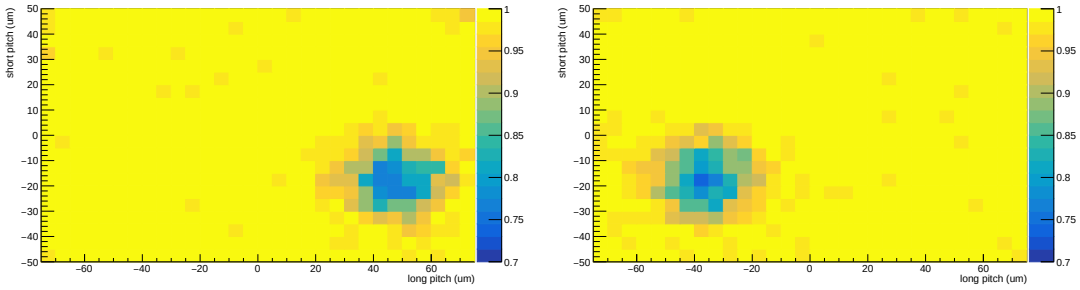
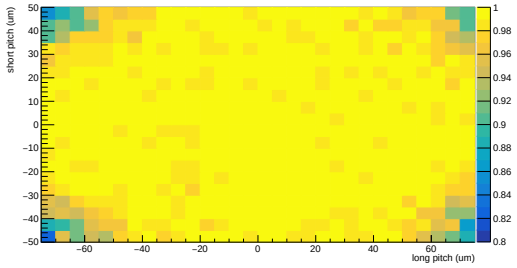


Figure 4.36: Maps of the particle detection efficiency within the pixel cells of the most irradiated sensor of Fig. 4.35: same sensor before the irradiation at bias voltages of 40 V (first plot) and 150 V (second plot), and, once irradiated up to $\approx 5\times10^{15} n_{eq}/cm^2$ at a bias of 650 V (third plot). The detection efficiency is clearly affected by the punch-through and its bias grid.



(a) 53B - Even columns pixels

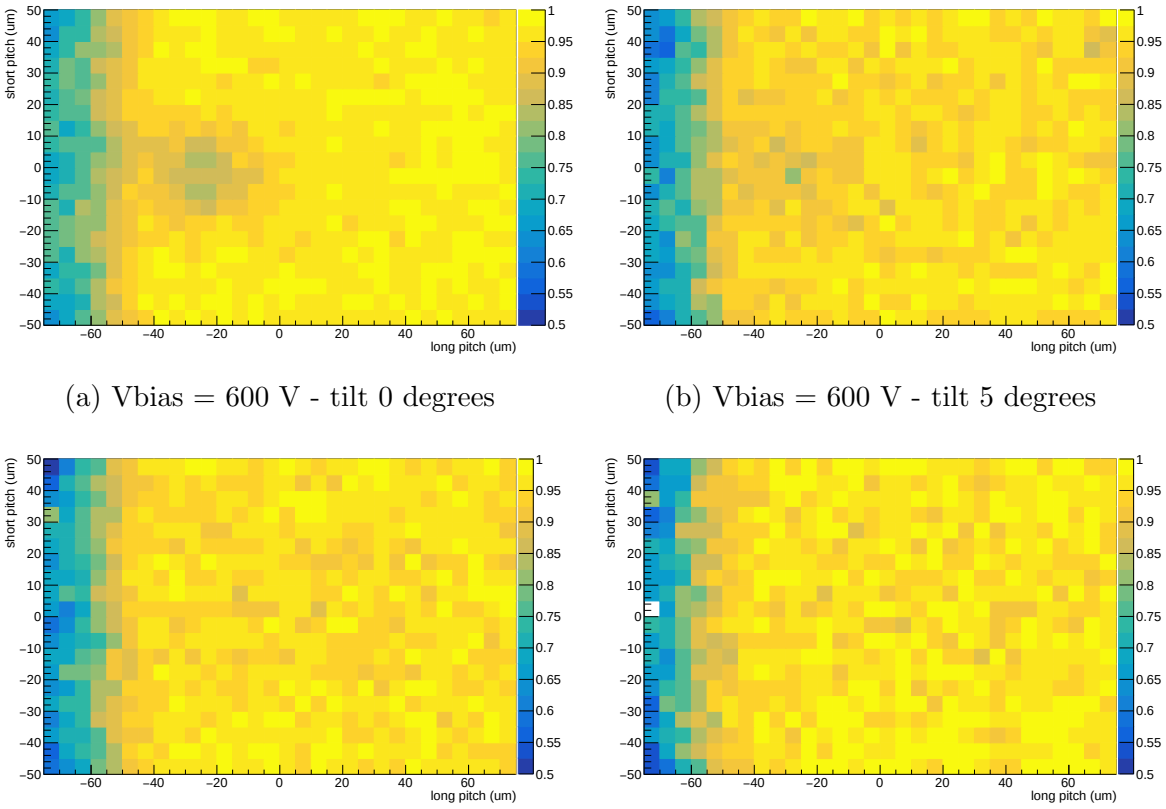
(b) 53B - Odd columns pixels



(c) 62D - All pixels

Figure 4.37: Maps of the particle detection efficiency within the pixel cells of the sensors irradiated at the smallest and the intermediate fluence (see Fig. 4.35) at a bias of 500 V and 800 V respectively. The sensor irradiated at intermediate fluence does not have punch-through structures so a cumulative efficiency for odd and even columns pixels is reported.

It is important to mention that the efficiency loss in the region of the punch through dot is strongly reduced once the sensor is tilted as shown in Fig. 4.38 for sensor 23C irradiated to a fluence of $\approx 3.1 \times 10^{15}\, n_{eq}/cm^2$: a tilt of 10 degrees is sufficient to recover full efficiency. The efficiency degradation in the punch-through bias grid region cannot be wiped out due to the large extension of this region.



(a) Vbias = 600 V - tilt 0 degrees

(b) Vbias = 600 V - tilt 5 degrees

(c) Vbias = 600 V - tilt 10 degrees

(d) Vbias = 600 V - tilt 20 degrees

Figure 4.38: Maps of the particle detection efficiency within the pixel cells of the sensor 23C irradiated up to a fluence of $\approx 3.1 \times 10^{15}\, n_{eq}/cm^2$ at a bias voltage of 600 V and different tilt angles, as specified in the captions.

In order to compare the real performance of the three irradiated sensors independently of the problems caused by the punch-through structures, the measurements should be limited only to the half-cells without punch-through structures. Doing this, the particle detection efficiency of the two sensors with punch-through notably improves as shown in Fig. 4.39. The detection efficiency of the most irradiated sensor, for bias voltages above 500 V, approaches that of the sensor at intermediate fluence, while that of the least irradiated sensor practically reaches 100%.

The corresponding (i.e. as measured only on the half-cells without punch-through structures) MPV of the collected charge as a function of bias voltage are shown in Fig. 4.40. They are corrected to account for the main factors affecting the calibration circuit of the readout chip when operated at high radiation doses, i.e. the variation of the bandgap reference voltage [45] and the change of the amplifier gain due to high leakage currents in the sensor (of the order of hundreds of μA). At the maximum fluence of $\approx 5 \times 10^{15} n_{eq}/cm^2$ and 650 V bias voltage, the 100 μm thick sensor reaches more than 90% charge collection efficiency (charge collection efficiency is the ratio between the measured MPV in electrons

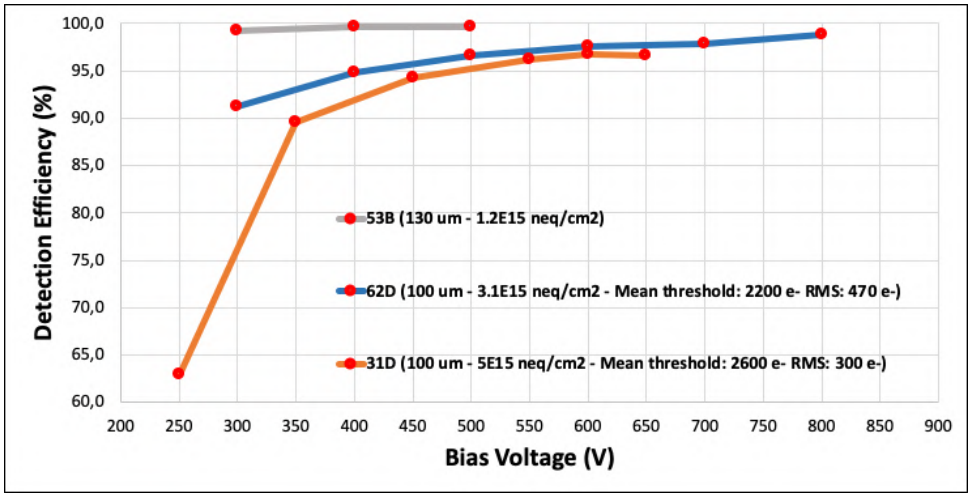


Figure 4.39: Particle detection efficiency as a function of the bias voltage for irradiated thin planar sensors. The measurements are limited only to the half-cells without punch-through structures. All sensors have bias punch-through structures except the one at the intermediate fluence.

and the expected one in absence of carrier trapping, i.e. 6350 e^- for the thinner sensors and 8740 e^- for the others). At the intermediate fluence of $\approx$ 3.1×10^{15} n_{eq}/cm^2 the 100 μm thick sensor reaches full charge collection efficiency at about 600 V, while, at the same bias voltage and less than half the fluence, the thicker sensor is still losing about 20% of the charge.

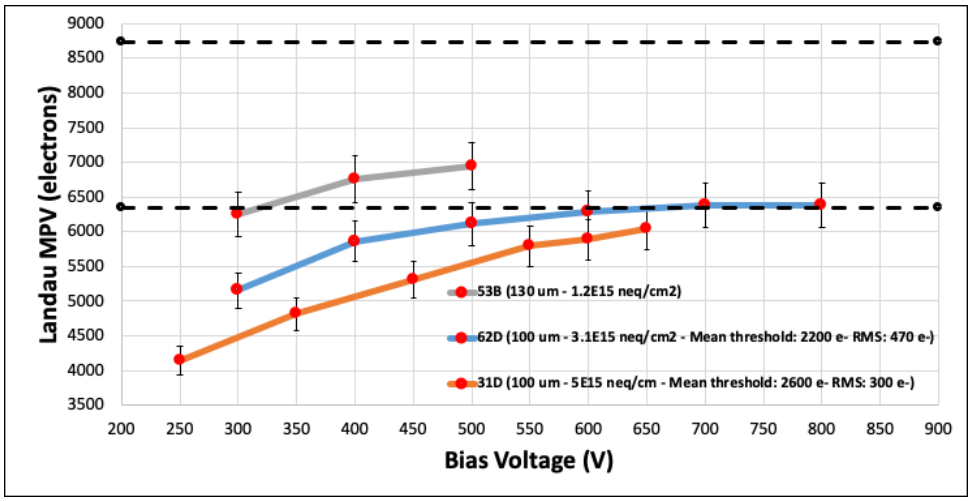
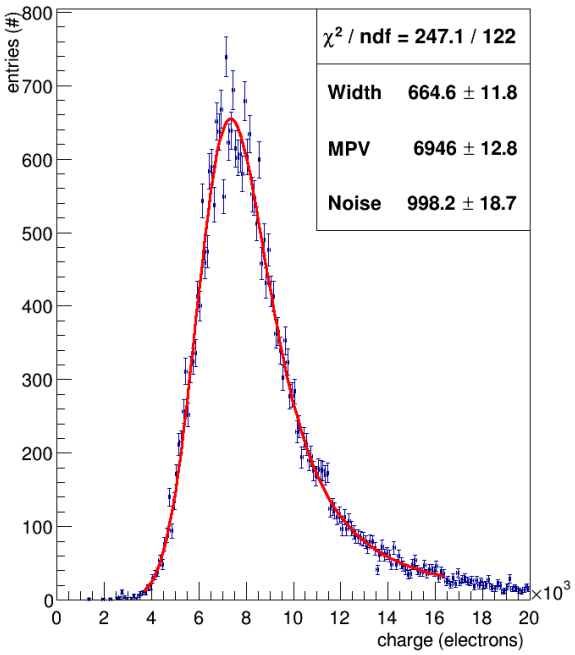
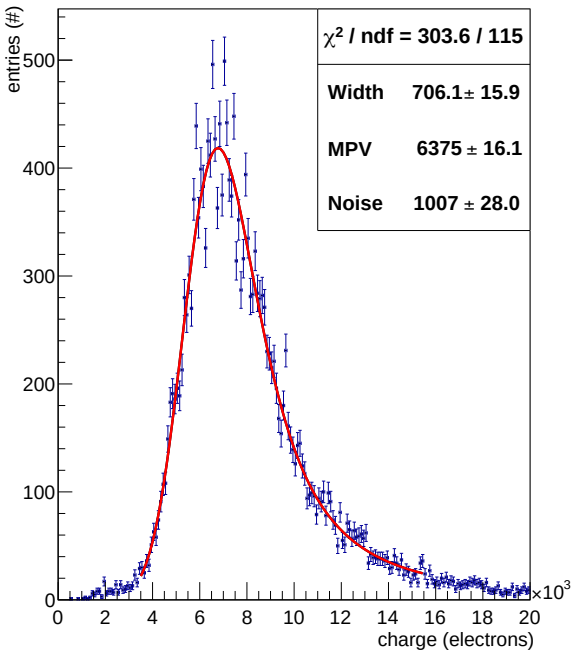


Figure 4.40: Collected charge MPV as a function of the bias voltage for irradiated thin planar sensors. For the two sensors with punch-through, i.e. the least and the most irradiated one, it is measured only on the half-cells without punch-through structures. Dashed lines denote their expected values before irradiation.

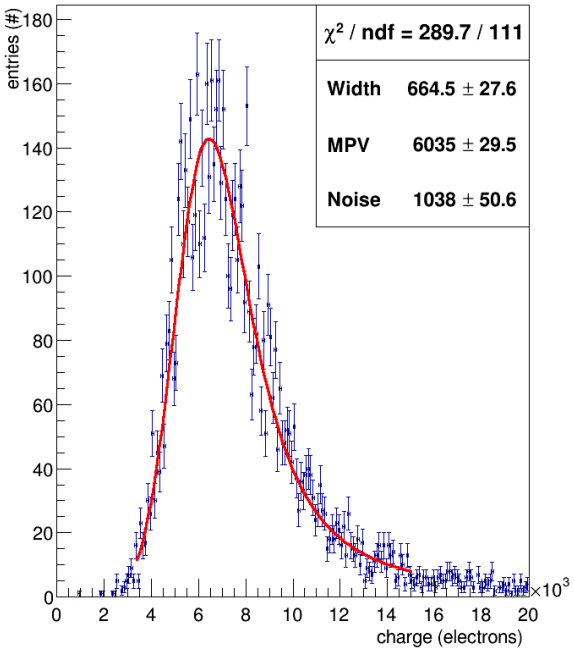
The measured MIP signal spectra, at the highest bias voltages applied to each sensor, are shown in Fig. 4.41.



(a) 53B - Vbias = 500 V



(b) 62D - Vbias = 800 V



(c) 31D - Vbias = 650 V

Figure 4.41: MIP signal spectra at the highest bias voltages applied to each sensor of Fig. 4.40: sensor irradiated up to $\approx 1.2 \times 10^{15}\ n_{eq}/cm^2$ (first), sensor irradiated up to $\approx 3.1 \times 10^{15}\ n_{eq}/cm^2$ (second), and sensor irradiated up to $\approx 5 \times 10^{15}\ n_{eq}/cm^2$ (third). The values of the applied bias voltages are 500 V, 800 V, and 650 V, respectively. Superimposed on each spectrum is its best fit performed with a convolution of a Landau and a Gaussian functions.

4.4 3D sensors bonded to PSI46dig ROC

As described in the introduction, 3D sensors have been produced with standard, i.e. $100 \times 150\ \mu m^2$, and reduced, i.e. $50 \times 50\ \mu m^2$ and $25 \times 100\ \mu m^2$, pixel cell dimensions. In this section the performance of standard pitch sensors before and after irradiation will be presented, together with some studies on small pitch sensor before irradiation. The characteristics of the sensors showed in this section can be found in Fig. 4.42.

A more detailed description of the performance of small pitch sensors will be given in the next section since the introduction of the RD53A ROC made possible to read-out all the pixels of these sensors.

ID Chip	Active Thickness (µm)	Pitch (µm^2)	Number of electrodes	Irradation Fluence (neq/cm2)
w76-1	130	100 x 150	2E	3,1E+15
w76-2	130	100 x 150	3E	5,0E+15
w76-14	130	100 x 150	2E	5,0E+15
w79_x1_y4	130	25 x 100	2E	Not irradiated
w91_x1_y6	130	50 x 50	1E	Not irradiated

Figure 4.42: Summary table containing the relevant characteristics of the 3D sensors presented in this section.

The particle detection efficiency as a function of the bias voltage for the standard pitch sensors can be found in Fig. 4.43.

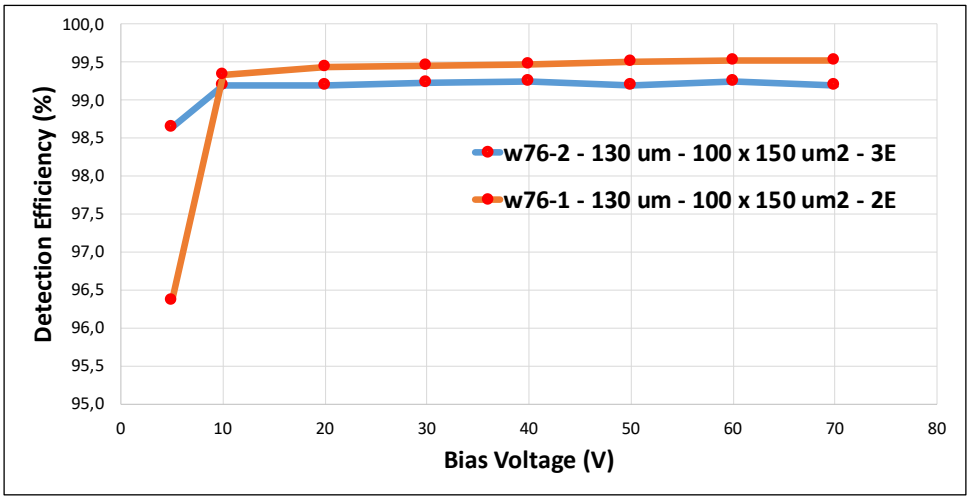
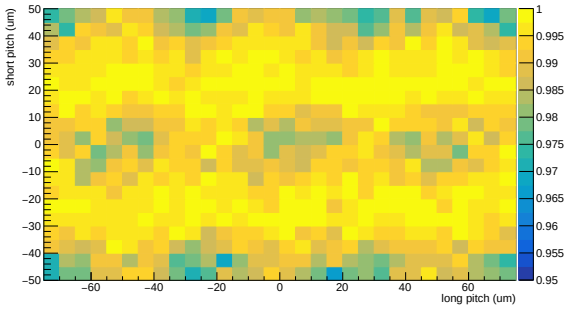


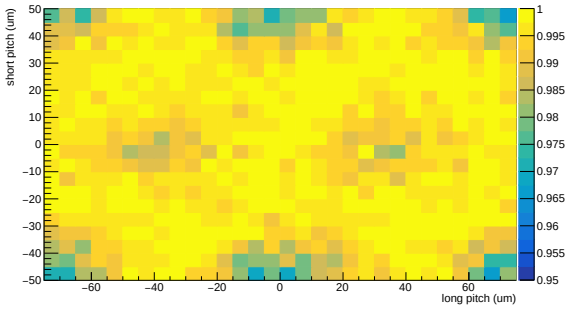
Figure 4.43: Particle detection efficiency as a function of the bias voltage for the standard pitch 3D sensors.

It is worth noting that both sensors reach a full detection efficiency at a bias voltage of only 10 V and that the sensor with a greater number of electrodes (11 for 3E, 8 for 2E) has a lower efficiency: the efficiency losses in the regions of the electrodes are evident in the in-pixel detection efficiency maps showed in Fig. 4.44.

It is important to mention that such efficiency losses are completely recovered once the sensors are tilted, even by a small angle as 5 degrees, since the column diameter is very small, $\approx 5\ \mu m$. The efficiency as a function of the tilt angle is showed in Fig. 4.45.



(a) w76-2 - 3E - Vbias = 40 V



(b) w76-1 - 2E - Vbias = 40 V

Figure 4.44: Maps of the particle detection efficiency within the pixel cells of the 3D standard pitch sensors at a bias voltage of 40 V.

Angle (degree)	Efficiency 3E (%)	Efficiency 2E (%)
0	99.27	99.45
5	99.77	99.85
10	99.88	99.87

Figure 4.45: Efficiency of standard pitch 3D sensors as a function of the angle of the incident particle at 30 V bias voltage.

The collected charge MPV as a function of the bias voltage, measured before irradiation, is shown in Fig. 4.46 for the standard pitch 3D sensors and in Fig. 4.47 for the small pitch ones. The collected charge is compatible, within the uncertainties, with that of planar sensors with the same active layer thickness.

The corresponding MIP signal spectra measured before irradiation are reported in Fig. 4.48 for the standard pitch 3D sensors at a bias voltage of 70 V and in Fig. 4.49 for the small pitch 3D sensors at a bias voltage of 50 V. The small ridge on the left-hand side of the two distributions of Fig. 4.49 is mainly due to the charge-sharing effects with the surrounding pixel cells which are not read out.

Unfortunately, it is not possible to show the collected charge MPV as a function of the bias voltage for irradiated 3D sensors because of unreliable calibration of the readout chip due to radiation damage. Probably during the irradiation of these 3D detectors, the center of the proton beam was slightly offset with respect to the center of the sensor, hence damaging the voltage regulator which is responsible for the dynamics of the analog to digital converter, possibly resulting in a binary readout. Nevertheless, their detection efficiency as a function of the bias voltage has been measured and is reported in Fig. 4.50 for orthogonally incident particles. At a bias voltage of 200 V the sensor irradiated to the highest fluence ($5 \times 10^{15}\ n_{eq}/cm^2$) reaches 96% efficiency at a threshold of about 3000 electrons. This performance is very good given the rather high threshold set. By lowering the threshold to about 2000 electrons it should be possible recover the full detection efficiency.

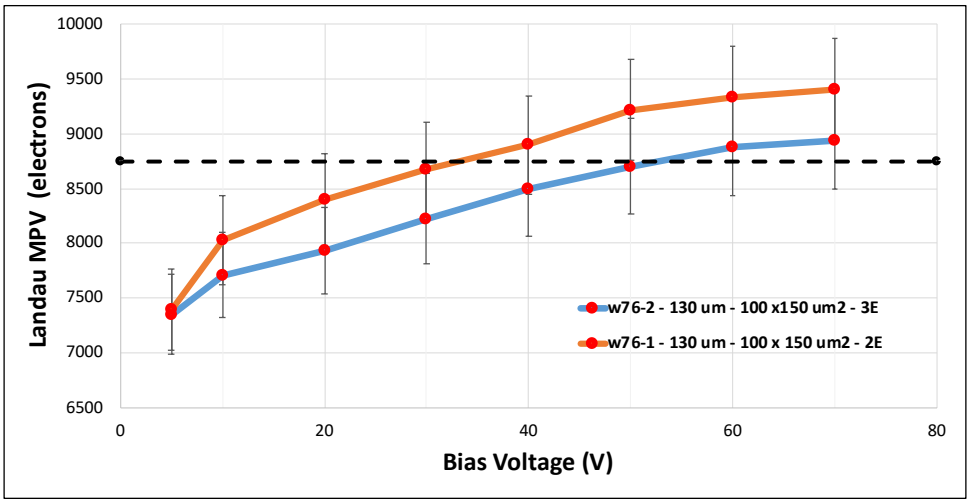


Figure 4.46: Collected charge MPV as a function of the bias voltage for two non-irradiated 3D sensors, one with two junction columns (2E), the other with three (3E). The sensor thickness is 130 μm and the pixel size is $100 \times 150\ \mu m^2$. The horizontal dashed line indicates the expected value.

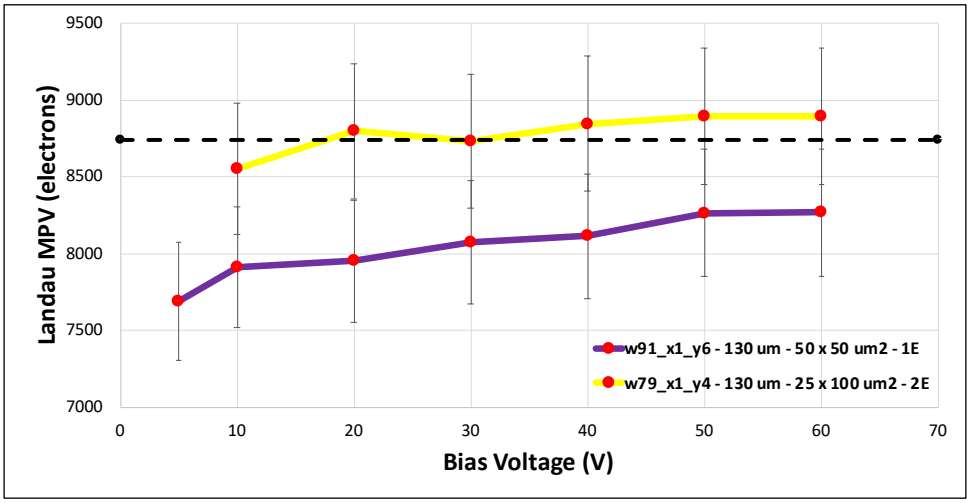
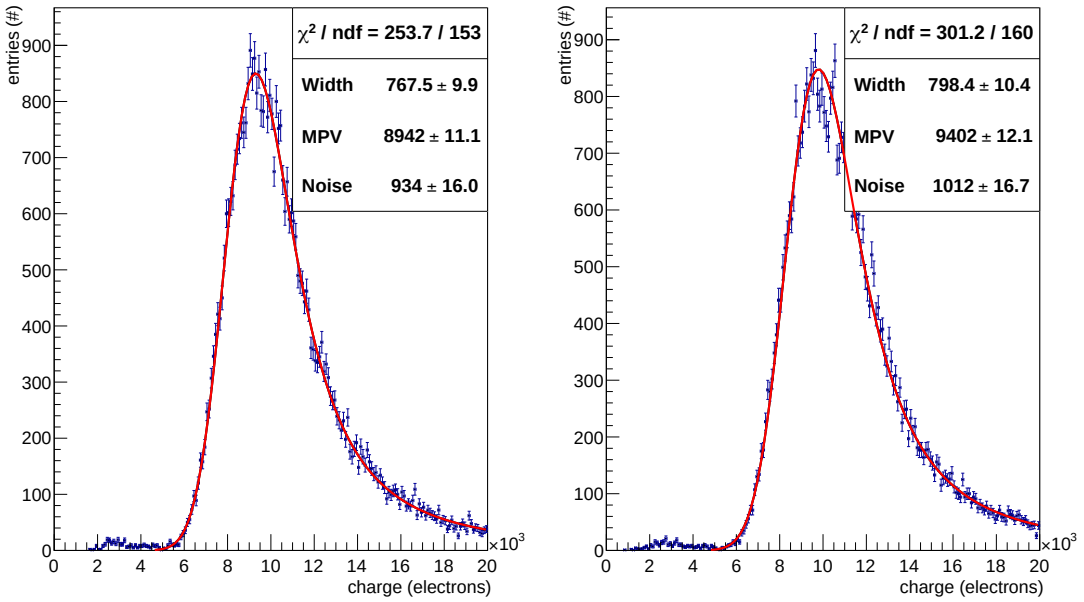


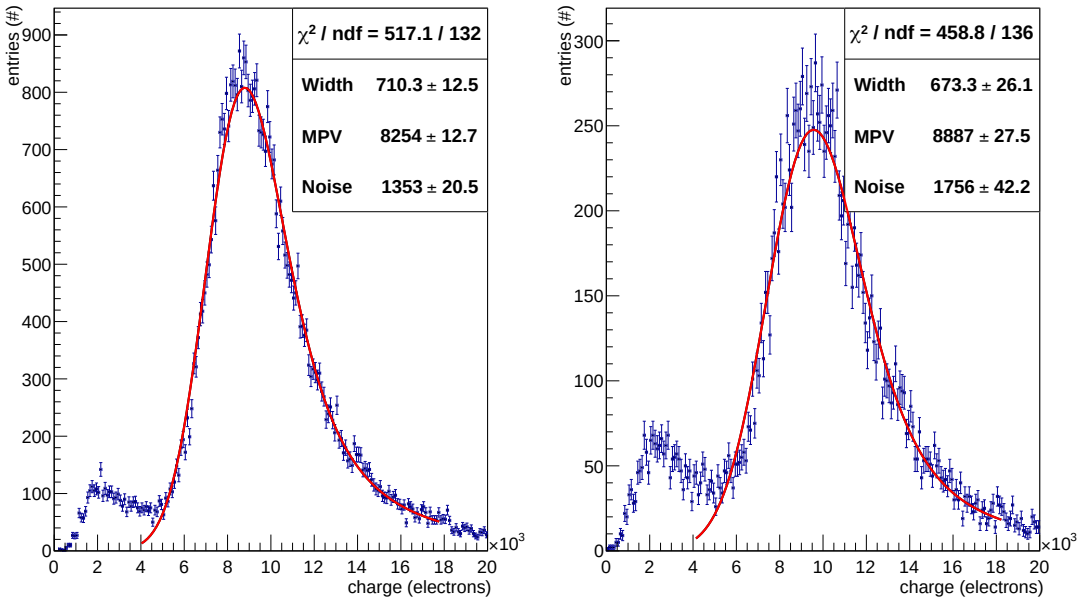
Figure 4.47: Collected charge MPV as a function of the bias voltage for two non-irradiated 3D sensors with small pitch. The sensor thickness is 130 μm and the pixel size are $25 \times 100\ \mu m^2$ and $50 \times 50\ \mu m^2$. The horizontal dashed line indicates the expected value.



(a) w76-2 - 3E - Vbias = 40 V

(b) w76-1 - 2E - Vbias = 40 V

Figure 4.48: The MIP signal spectra measured with $100 \times 150\,\mu m^2$ 3D sensors before irradiation. The plot on the left-hand side refers to a three junction electrode sensor (3E), while plot on the right-hand side refers to a two junction electrode sensor (2E). The plots correspond to the two measurements of Fig. 4.46 at a bias voltage of 70 V.



(a) w91_x1_y6 - $50 \times 50\,\mu m^2$ 1E - Vbias = 50 V

(b) w79_x1_y4 - $25 \times 100\,\mu m^2$ 2E - Vbias = 50 V

Figure 4.49: The MIP signal spectra measured, before irradiation, with a $50 \times 50\,\mu m^2$ 3D, type 1E, sensor (left plot) and with a $25 \times 100\,\mu m^2$ 3D, type 2E, sensor (right plot). The spectra correspond to the two measurements reported in Fig. 4.47 at 50 V bias voltage.

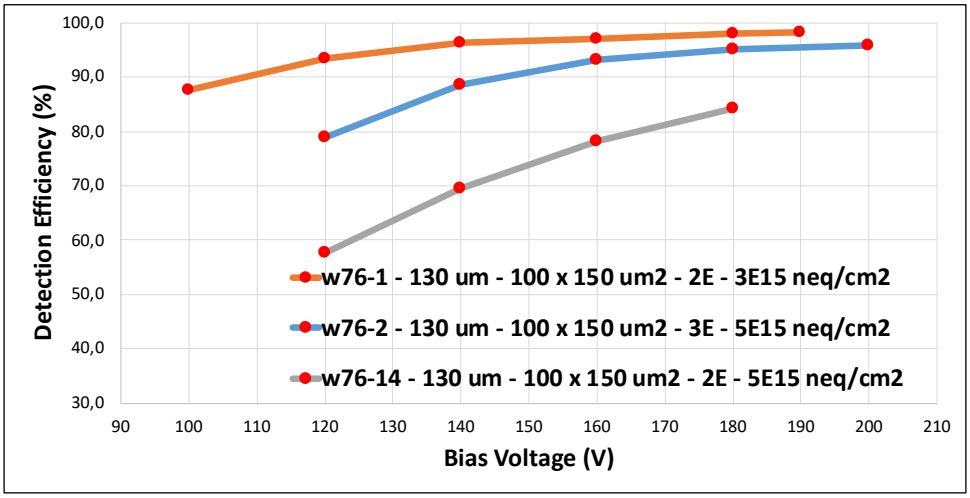


Figure 4.50: Particle detection efficiency as a function of the bias voltage for irradiated 3D sensors. The threshold of the sensors irradiated to the highest fluence was set to ≈ 3000 electrons, while for the other sensor it was set to ≈ 2000 electrons.

4.5 3D sensors bonded to RD53A ROC

Some 3D sensors bonded to the RD53A ROC have been tested on beam at Fermilab and DESY facilities. A description of these sensors can be found in Fig. 4.51.

ID Chip	Alias	Active Thickness (µm)	Pitch (µm^2)	Number of electrodes	Irradiation Fluence (neq/cm2)	Test beam facility	Threshold(e-)
w79_x3_y3	1	130	50 x 50	1E	Not irradiated	Desy	900
w79_x2_y1	2	130	25 x 100	1E	Not irradiated	Desy	900
w30_x3_y1	3	150	50 x 50	1E	Not irradiated	Desy	940
w3_x3_y4	4	130	50 x 50	1E	1,0E+16	Desy	950\1150
w91_x2_y3	5	130	50 x 50	1E	1,0E+16	Fermilab	1500

Figure 4.51: Summary table containing the relevant characteristics of the 3D sensors presented in this section.

It is important to mention that the thresholds applied to irradiated and not irradiated sensors tested at DESY are quite lower than the ones applied to sensors bonded to the PSI46dig ROC: this is particularly beneficial for the irradiated sensors, where the collected charge is lower. A higher threshold was applied to the sensor tested at Fermilab since it has been the first one put on beam.

Complementary information have been collected at the two test beam facilities since the DAQ system at Fermilab allows for a calibration of the ADC output of the ROC from TimeOverThreshold (ToT) to electrons while the track reconstruction at DESY is more precise and allows for more detailed measurements at different angles. It is worth noting that $25 \times 100\,\mu m^2$ sensors have been mounted on a rotating support in such a way that the effect of the charge sharing along the short pitch direction could be investigated.

The DESY telescope resolution can be estimated from the RMS of the distribution of the difference of the track impact points coordinates at the center of the telescope predicted by the upstream and downstream arms, showed in Fig. 4.52. A student-t

function is used for the fit. The track residual at the DUT position is computed as the difference between the measured coordinate and the mean value of the track impact points coordinates predicted by the upstream and downstream arms. This way of computing the residuals implies that the tracking error at the DUT is half the RMS of the distributions in the figure. These errors should be estimated every time the telescope arms are moved. Estimating the errors for different runs taken with the same geometry proved that the values remain constant.

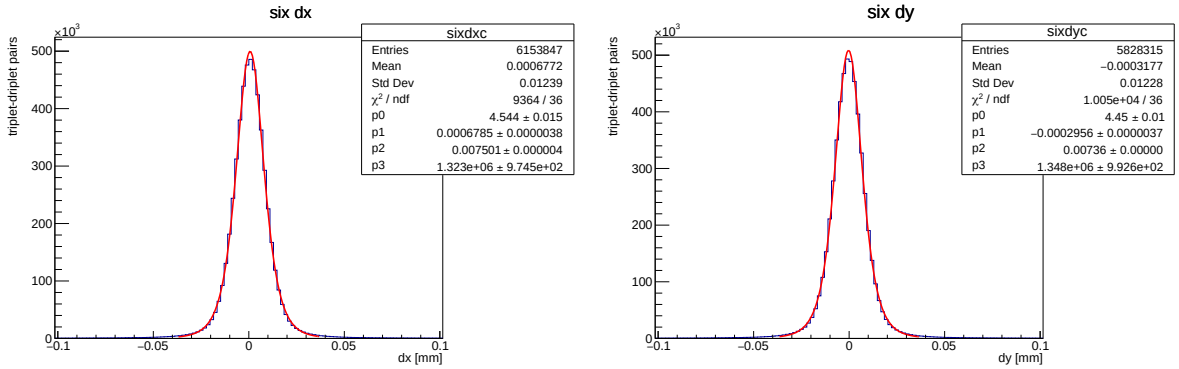


Figure 4.52: Distributions of the difference of the track impact points coordinates at the center of the telescope predicted by the upstream and downstream arms.

Such estimation cannot be applied when irradiated modules are tested: the presence of the cold box used to control the temperature of the sensors cause such a high scattering of the electrons that tracks reconstructed by the downstream arm cannot be used.

The estimated resolution at a bias voltage of 30 V as a function of the rotation angle around the x-axis can be found in Fig. 4.53. The minimum should be located at

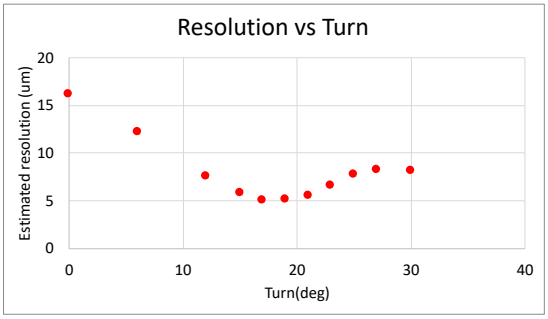
$$\arctan(sensorPitch/sensorThickness)$$

which means 21 deg for sensor w79_x3_y3 ($50 \times 50\,\mu m^2$, 130 μm active thickness), 11 deg for sensor w79_x2_y1 ($25 \times 100\,\mu m^2$, 130 μm active thickness) and 18 deg for sensor w30_x3_y1 ($50 \times 50\,\mu m^2$, 150 μm active thickness). Measurements are in good agreement with such predictions: the two $50 \times 50\,\mu m^2$ sensors show a similar trend with an optimal resolution below 5 μm while the $25 \times 100\,\mu m^2$ sensor shows an optimal resolution of 3 μm at the minimum.

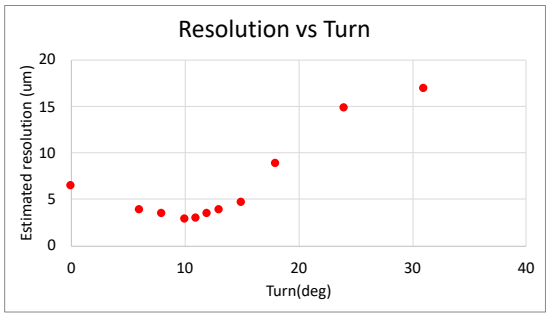
Distributions of the residuals along the tilted coordinate at the predicted angle for the resolution minimum can be found in Fig. 4.54 .

The precision achieved in the tracks reconstruction allows for a detailed measurement of the in-pixel efficiency as can be seen in Fig. 4.55. The detection efficiency as a function of the track impact point coordinates is reported for the three sensors: plots on the left refer to normal incidence data taking while plots on the right refer to the smallest tilt angle applied. Efficiency losses in the regions of the peripheral columnar electrodes are well evident for normal incident tracks but are wiped out once the sensor is tilted.

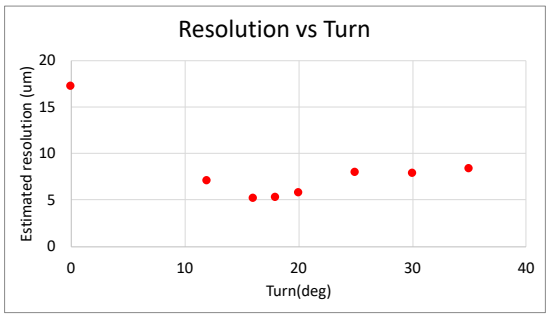
Another feature that can be measured with great precision is the distribution of the cluster size on the pixel cell as can be seen in Fig. 4.56. The cluster size as a function of the track impact point coordinates is reported for the three sensors: once again plots on the left refer to normal incidence data taking while plots on the right refer to the smallest tilting angle applied. It is evident that once the sensors are tilted the cluster size increases in the direction of the x-axis, as expected.



(a) Sensor 1 ($50 \times 50\ \mu m^2$, 130 μm active thickness)

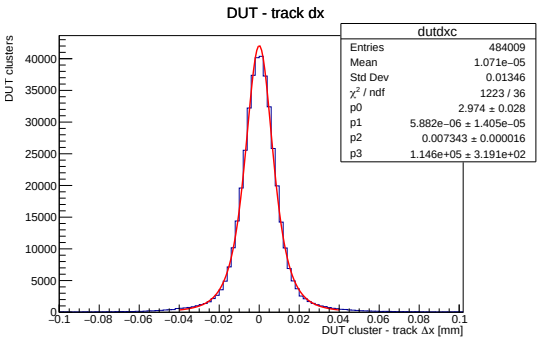


(b) Sensor 2 ($25 \times 100\ \mu m^2$, 130 μm active thickness)

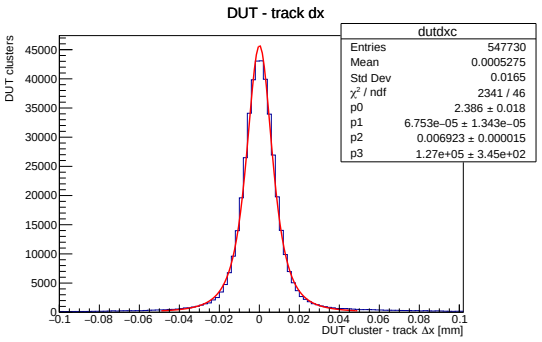


(c) Sensor 3 ($50 \times 50\ \mu m^2$, 150 μm active thickness)

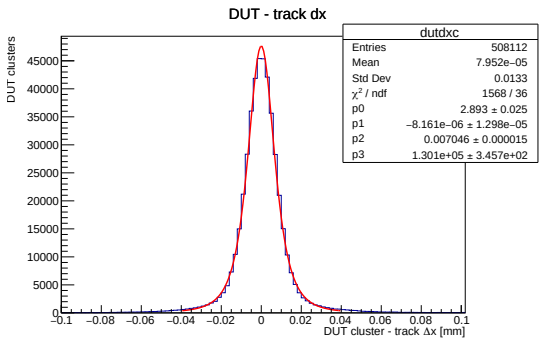
Figure 4.53: Estimated resolution as a function of the rotation angle around the x-axis for the three not irradiated sensors tested at DESY.



(a) Sensor 1 - 21 deg

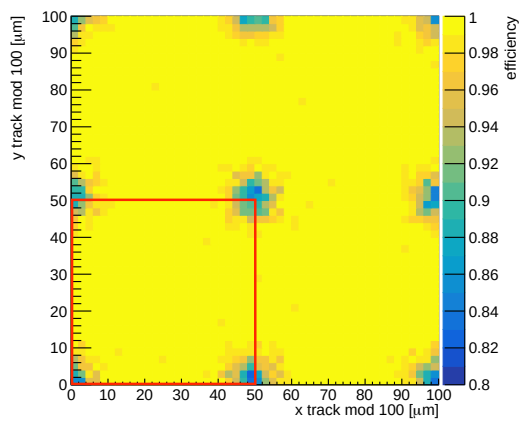


(b) Sensor 2 - 11 deg

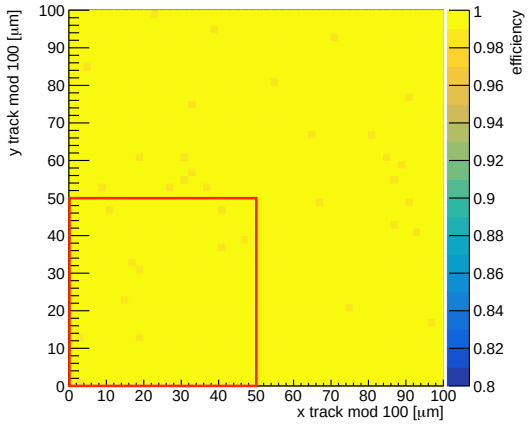


(c) Sensor 3 - 18 deg

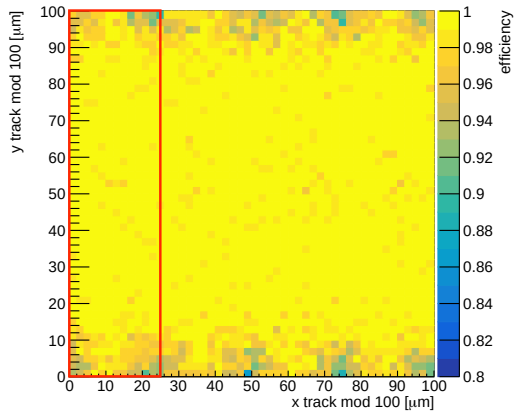
Figure 4.54: Distributions of the residuals along the tilted coordinate for the three not irradiated sensors tested at DESY.



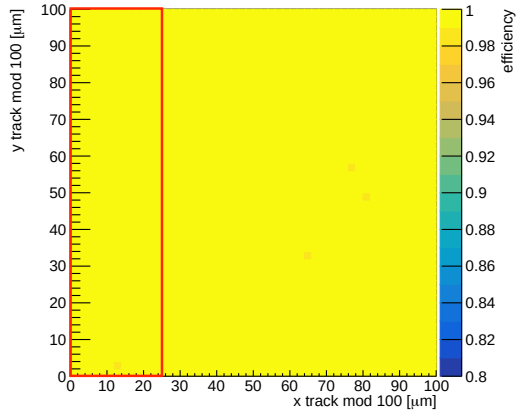
(a) Sensor 1 - Tilt: 0 deg - Efficiency: 99.37%



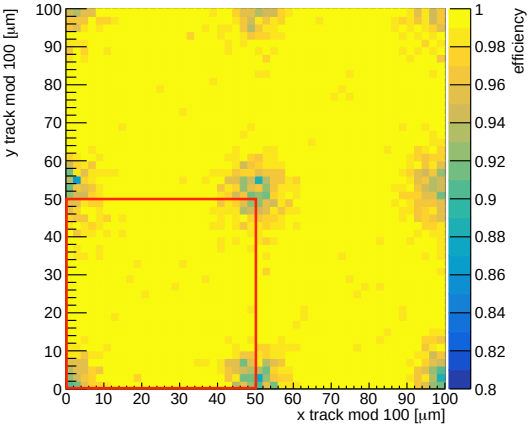
(b) Sensor 1 - Tilt: 6 deg - Efficiency: 99.98%



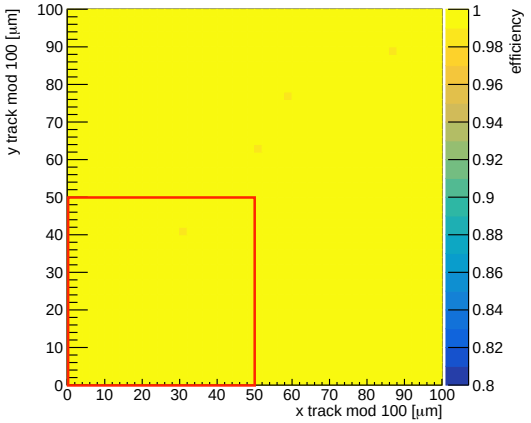
(c) Sensor 2 - Tilt: 0 deg - Efficiency: 99.17%



(d) Sensor 2 - Tilt: 6 deg - Efficiency: 99.95%

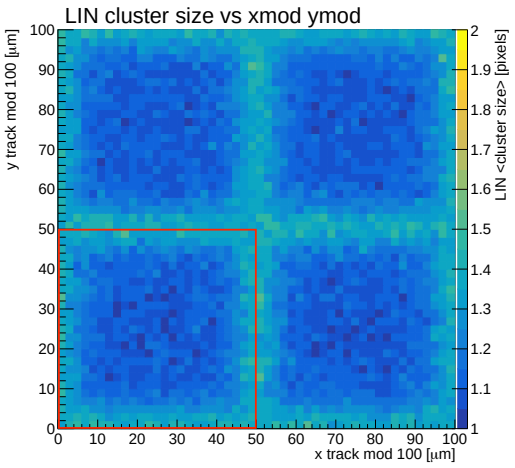


(e) Sensor 3 - Tilt: 0 deg - Efficiency: 99.29%

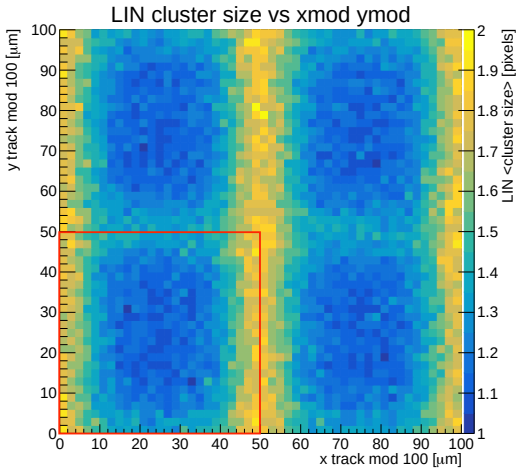


(f) Sensor 3 - Tilt: 12 deg - Efficiency: 99.97%

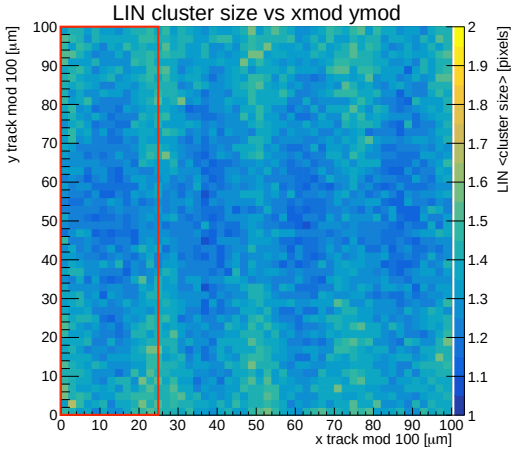
Figure 4.55: Distributions of the in-pixel efficiency on four adjacent pixel cells for the three not irradiated sensors tested at DESY at bias voltage of 30 V.



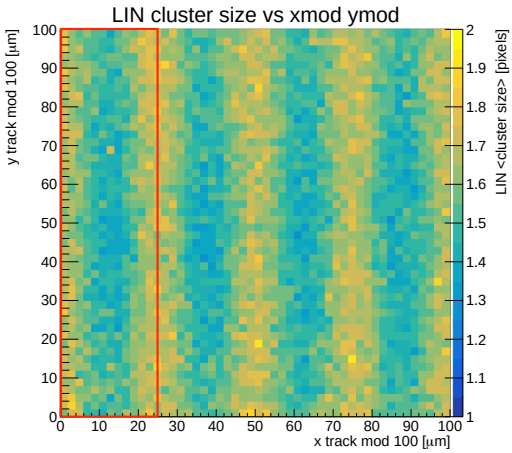
(a) Sensor 1 - Tilt: 0 deg



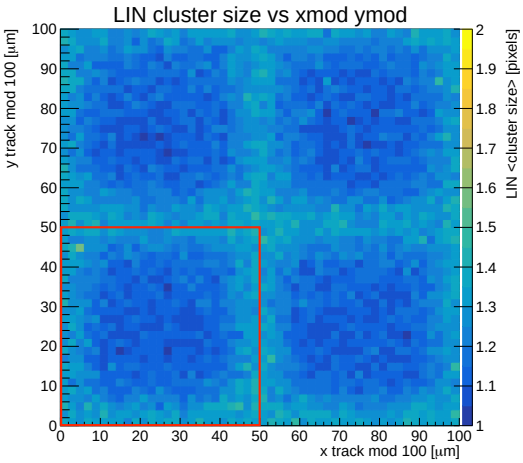
(b) Sensor 1 - Tilt: 6 deg



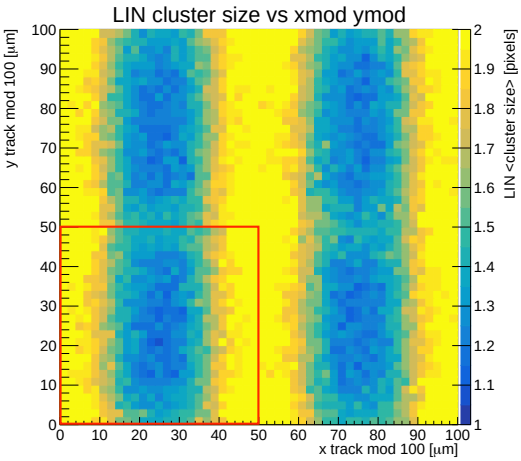
(c) Sensor 2 - Tilt: 0 deg



(d) Sensor 2 - Tilt: 6 deg



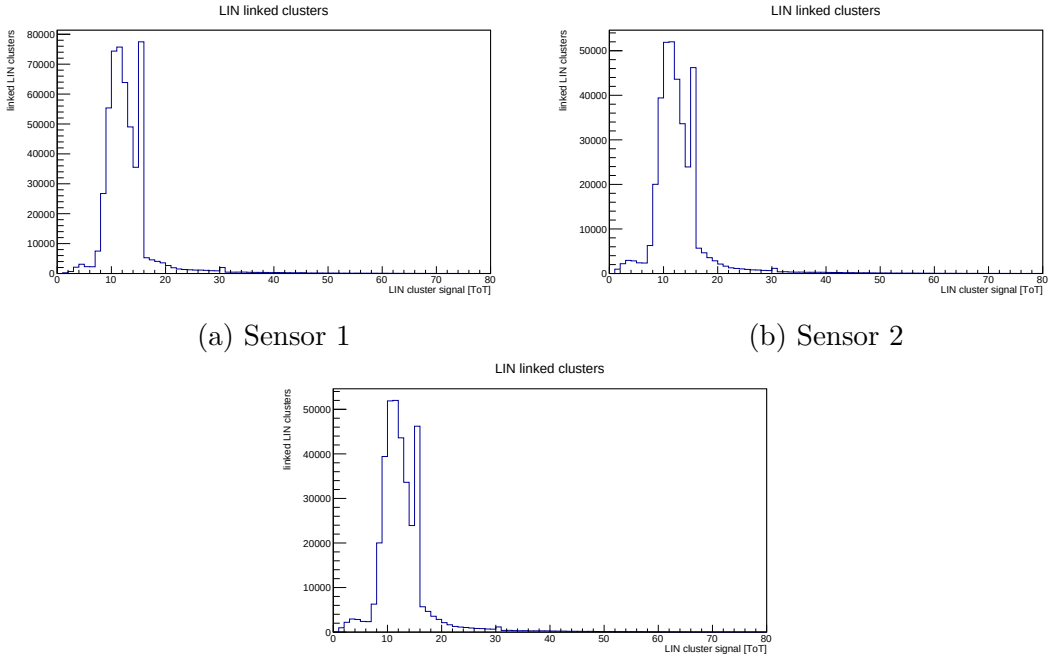
(e) Sensor 3 - Tilt: 0 deg



(f) Sensor 3 - Tilt: 12 deg

Figure 4.56: Distributions of the in-pixel cluster size on four adjacent pixel cells for the three not irradiated sensors tested at DESY at bias voltage of 30 V.

The distributions of the charge collected by the three sensors can be found in Fig.4.57. The Landau shape of the distributions is a confirmation that the sensors were collecting Minimum Ionizing Particles(MIPs) and not just noise. The peak at the ToT value 15 is caused by the configuration of the readout: pixels are readout in groups of four, a ToT value between 0 and 14 is assigned to the ones that actually collected the charge, the ToT value 15 is assigned to the others.



(a) Sensor 1

(b) Sensor 2

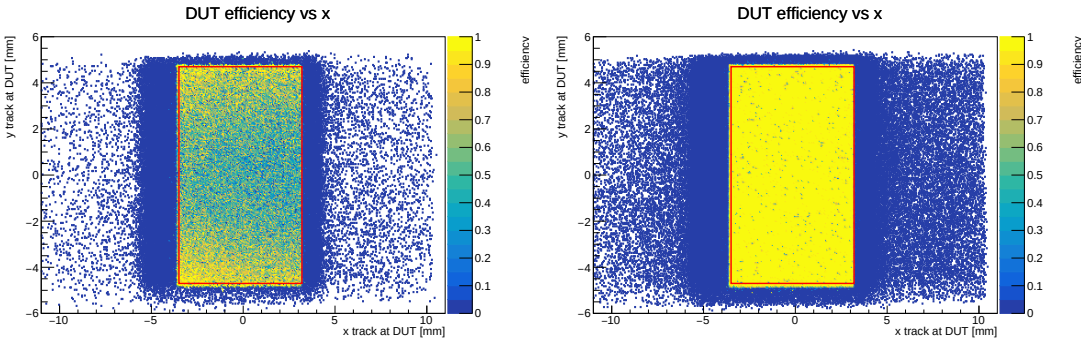
(c) Sensor 3

Figure 4.57: Distributions of the charge collected by the three not irradiated sensors tested at DESY at bias voltage of 30 V.

Fig. 4.58 shows the detection efficiency maps of the irradiated sensors tested at DESY at two different bias voltages, 28 V and 146 V. The red box defines the fiducial window of the sensor used in the data analysis. The sensor was irradiated at the CERN PS irradiation facility and, as can be seen in the left figure, the irradiation beam was centered on the region of the linear front-end of the ROC. The efficiency becomes more and more uniform as the bias voltage is raised, as can be seen in the right figure.

The detection efficiency of the sensor as a function of the applied bias voltage is reported in Fig. 4.59. A "low" and a "high" irradiation regions have been defined according to the coordinates of the track impact point, as specified in the figure. It can be seen that at lower bias voltages the central region, exposed to a higher irradiation fluence, has lower efficiency than the peripheral regions, where the irradiation fluence was lower. Once full depletion of the sensor is reached the efficiency becomes uniform across the detector surface.

The detection efficiency has also been measured as a function of the rotation angle around the x-axis, as reported in Fig. 4.60. It is important to mention that for the measurement at normal incidence the applied threshold was 1150 e^-, while for the other measurements the applied threshold was 950 e^-. The applied bias voltage is 146 V for all the measurement points. Despite the reduced charge collection a rotation of 7 degrees is enough to get a detection efficiency of 99.5%. The in-pixel detection efficiency maps



(a) Bias voltage = 28 V (b) Bias voltage = 146 V

Figure 4.58: Detection efficiency maps of the irradiated sensor tested at DESY at two different bias voltages, before and after full depletion of the Silicon bulk.

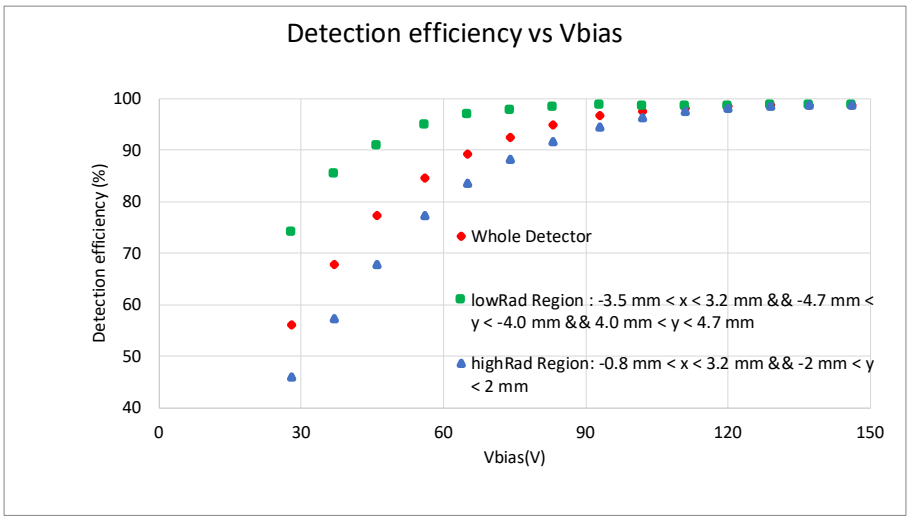


Figure 4.59: Detection efficiency of the irradiated sensor tested at DESY as a function of the applied bias voltage for two different regions of the sensor and the whole sensor.

at normal incidence and 7 degrees rotation are reported in Fig. 4.61.

Also in the case of irradiated sensors the cluster size increases when the sensor is tilted, as can be seen in Fig. 4.62. The average value in the pixel cell is lower due to the lower collected charge.

The sensor tested at Fermilab was irradiated at the CERN irradiation facility but in a different set with respect to the one tested at DESY: in this case the irradiation beam spot was centered between the linear and differential front-ends of the ROC as can be seen in the efficiency map of the left side of Fig. 4.63, which was acquired at a low bias voltage (59 V). Once the sensor reached full depletion the efficiency became uniform, as can be seen from the map on the right side of the figure.

As a consequence of this non uniform irradiation only the pixels enclosed in the red boxes in the figure have been considered in the analysis as that should be the zone which received the highest irradiation fluence, close to $1 \times 10^{16}\, n_{eq}/cm^2$.

The charge collected by this sensor could be measured in electrons thanks to a proper calibration procedure of the ROC available through the FTBF DAQ. The distribution is reported in Fig. 4.64. The MPV of 5602 electrons corresponds to 64% of the one of a non irradiated sensor: this represents an encouraging result given the rather high

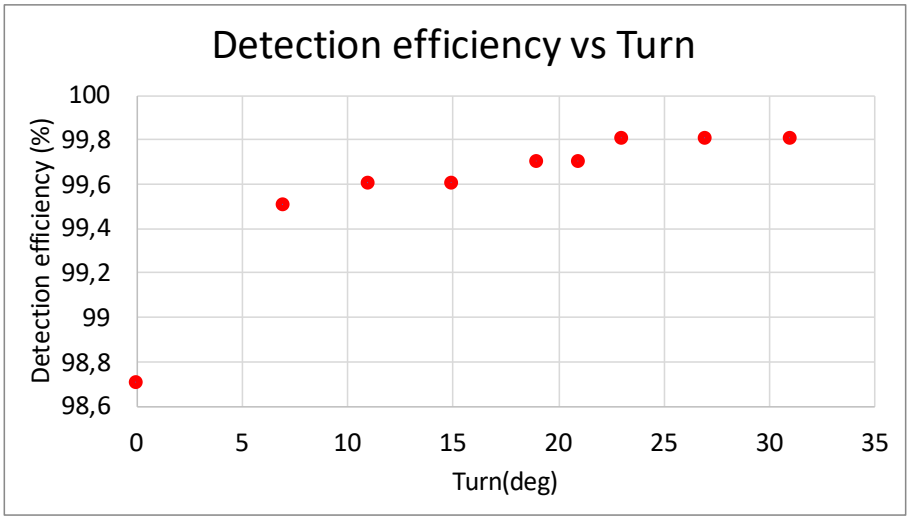
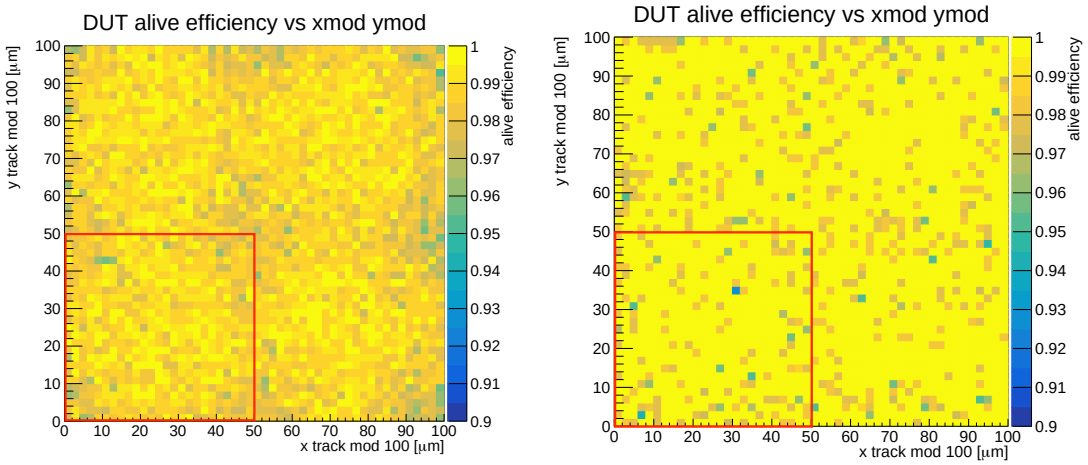


Figure 4.60: Detection efficiency of the irradiated sensor tested at DESY as a function of the rotation angle around the x-axis. The applied bias voltage is 146 V.

threshold that was set (1500 electrons) and the fact that a fluence of $1 \times 10^{16}\ n_{eq}/cm^2$ could represent the highest value reached before the replacement of the first layer of the pixel detector.

The representation of the collected charge as a function of the track impact point position on one and four adjacent pixel cells is reported in Fig. 4.65: losses are visible in the regions of the columnar electrodes even if they are less evident due to the low statistics available.

Finally, the detection efficiency and collected charge MPV as a function of the bias voltage are reported in Fig. 4.66, no sign of saturation is present indicating that the performance of the detector could be further improved by lowering the threshold.



(a) Tilt = 0 deg - Detection efficiency = 98.73% - Threshold = 1150 e-

(b) Tilt = 7 deg - Detection efficiency = 99.51% - Threshold = 950 e-

Figure 4.61: Distributions of the in-pixel efficiency on four adjacent pixel cells for the irradiated sensor tested at DESY at bias voltage of 146 V.

4.6 Conclusions

Several pixel sensor prototypes have been tested on beam and characterized since the beginning of the R& D project in 2015.

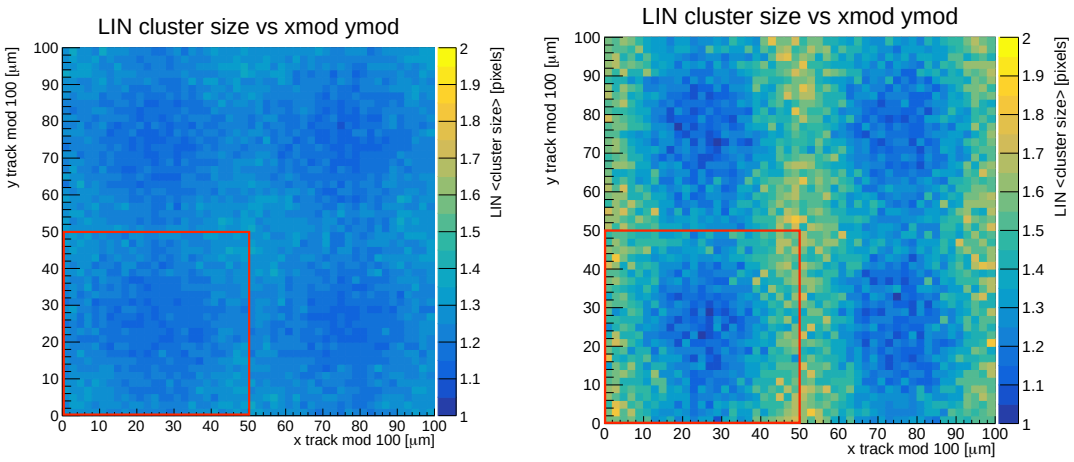
The first batches of planar and 3D sensors were designed for bonding to the readout chip (ROC) currently in use in the CMS pixel detector, PSI46dig. The radiation tolerance of this ROC limited the irradiation of the samples to a fluence of $5 \times 10^{15}\, n_{eq}/cm^2$.

It was demonstrated that planar sensors without punch-through structures, irradiated at a fluence of $3 \times 10^{15}\, n_{eq}/cm^2$, reach full charge collection efficiency and 97.5% detection efficiency at a bias voltage of 600 V. Efficiency losses in the sensors with punch through structures were seen to increase after irradiation since higher thresholds and bias voltages were required to operate the sensors: in particular the affected area extended from the bias dot to the whole bias grid. The design of the next batches of sensors was modified in order to have a single bias dot in common among four pixel cells, in order to mitigate efficiency losses.

Test beam measurements of unirradiated 3D sensors showed performances compatible with the ones of planar sensors with the same active thickness. This is a remarkable result since the Direct Wafer Bonding technique, in combination with a single-sided implantation process, was applied for the first time in the production of 3D sensors. It was also demonstrated that sensors with smaller pixel cells ($25 \times 100\, \mu m^2$ and $50 \times 50\, \mu m^2$) could be produced with the same fabrication processes with an acceptable yield. Detection efficiency measurements of irradiated 3D sensors showed that the bias voltage needed to reach the same value of the planar sensors are a factor three smaller. This would advise in favor of the use of 3D sensors in the inner layers of the phase-2 detector, together with their higher radiation tolerance.

The results on this first batches of sensors have also been published on a peer reviewed journal [46].

In the second half of 2018 the first prototype of the phase-2 ROC, named RD53A, became available. This ROC features a higher radiation tolerance than the PSI46dig, can



(a) Tilt = 0 deg - Threshold = 1150 e-

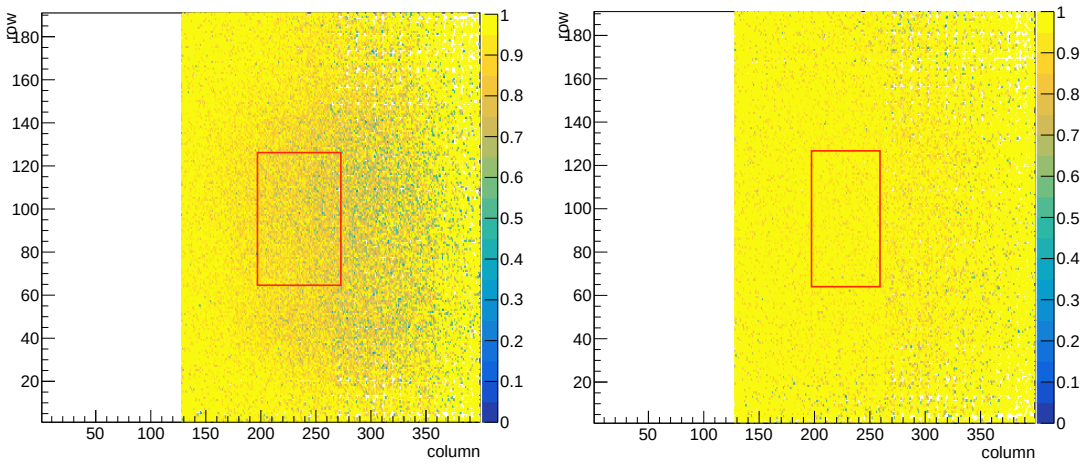
(b) Tilt = 7 deg - Threshold = 950 e-

Figure 4.62: Distributions of the in-pixel cluster size on four adjacent pixel cells of the irradiated sensor tested at DESY at bias voltage of 146 V.

be operated at a lower threshold and has $50\times50\mu m^2$ cells. Test beam measurements of 3D sensors bonded to this ROC demonstrated an excellent resolution: 5 μm for $50 \times 50\,\mu m^2$ pixel cells sensors and 2.5 μm for the $25 \times 100\,\mu m^2$ ones.

Test stand measurements showed the presence of cross-talk between adjacent cells in $25 \times 100\,\mu m^2$ planar sensors. This is a consequence of the particular bump-bonding scheme adopted since the sensors and ROC cells have different pitches. Several design solutions have been developed in order to reduce the entity of this effect and a less than 10% cross-talk has been achieved. It is important to mention that no cross-talk have been observed in $25 \times 100\,\mu m^2$ 3D sensors.

In the first half of 2021 a decision will be made regarding the employment of 3D sensors in the inner layers of the detector and the dimensions of the pixel cells. Such decision will be based both on further test beam campaigns that will be carried out and on simulations of the track reconstruction performance within the CMS data analysis framework.



(a) Vbias = 59 V (b) Vbias = 128 V

Figure 4.63: Efficiency maps of sensor w91_x2y3 at two different bias voltages, before and after reaching full depletion of the bulk.

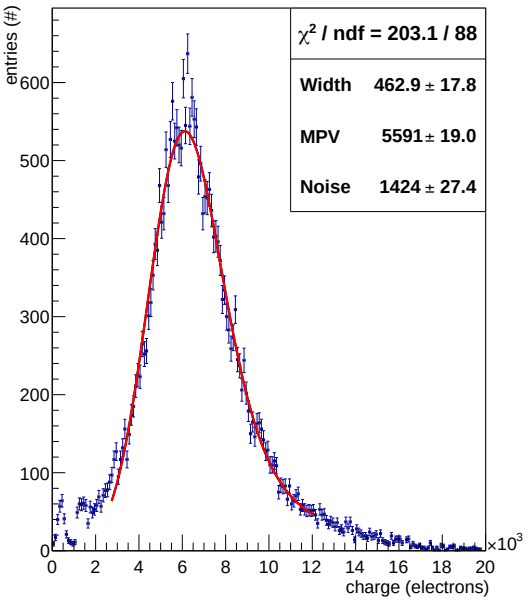
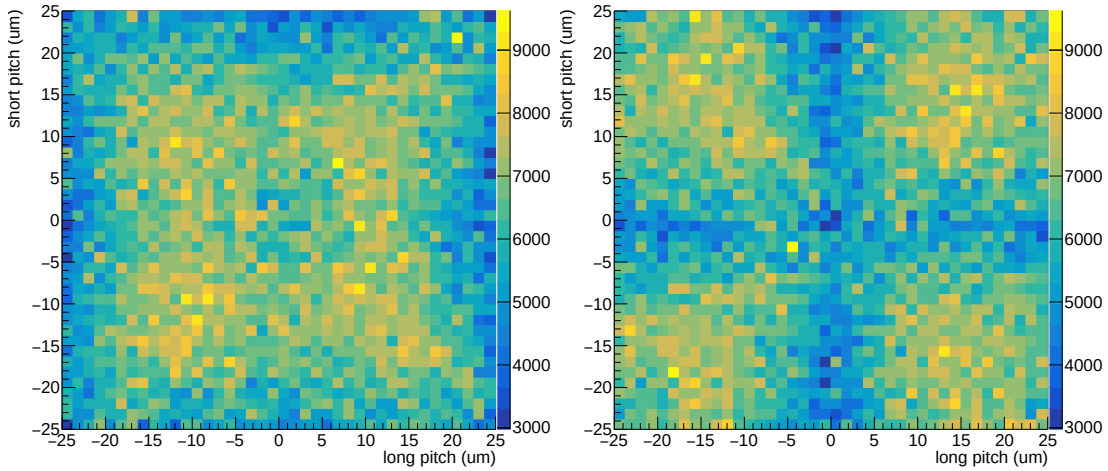


Figure 4.64: Collected charge spectrum of sensor w91_x2y3 at a bias voltage of 128 V.



(a) Single cell (b) 4 adjacent cells

Figure 4.65: Maps of the charge collected by sensor w91_x2y3 at a bias voltage of 128 V.

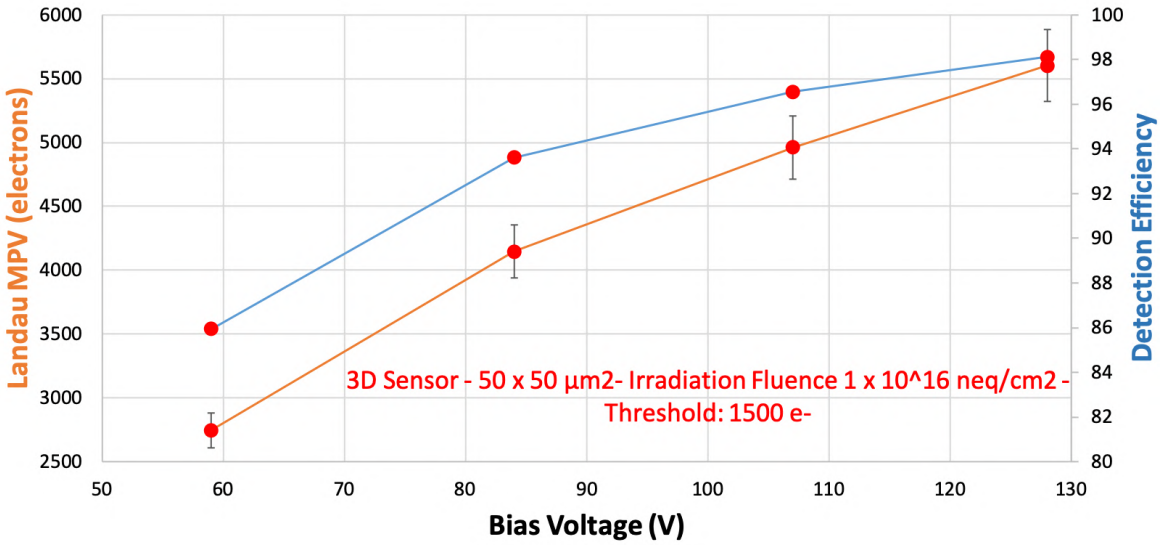


Figure 4.66: Collected charge MPV and decetion efficiency of sensor w91_x2y3 as a function of the bias voltage.

Chapter 5

Search for the $HH \to b\bar{b}\tau^+\tau^-$ decay

5.1 Phenomelogy Of Higgs Boson Pair Production

The Standard Model of particle physics (SM) is a renormalizable quantum field theory that describes phenomena at the subnuclear scale. It provides a unified description of the strong, weak, and electromagnetic forces, and incorporates a minimal scalar sector that is at the origin of the spontaneous breaking of the electroweak symmetry and of the masses of the fermions. The SM has undergone several verifications at collider experiments, and received further confirmation with the discovery of the Higgs boson. Despite the excellent agreement with direct experimental tests performed so far, some observations from the subnuclear to the astrophysical scales, as well as theoretical considerations, suggest that it is incomplete and that a broader theory exists beyond its current formulation. This physics beyond the SM (BSM) is possibly connected to the scalar sector, and the Higgs boson discovery opens new ways to its exploration. Being intimately related to the nature of the scalar sector, the production of Higgs boson pairs (HH) at the LHC is a powerful tool to investigate this sector.

This section discusses the importance of the study of HH production in the context of both SM and BSM physics. After introducing the SM gauge structure and its scalar sector, with a focus on the Higgs boson properties and couplings, HH production in the SM is reviewed. This process is subsequently discussed in the context of BSM models. Finally, the phenomenology in collider experiments is discussed.

5.1.1 The Standard Model of Particle Physics

The gauge sector and the scalar sector are the main elements of the SM. They are closely interconnected but rely on different theoretical concepts of gauge invariance and spontaneous symmetry breaking. They are discussed separately in what follows, with a focus on the properties of the Higgs boson.

The mathematical formulation of the SM is based on the local gauge invariance of its Lagrangian under the gauge group $SU(3)_C \times SU(2)_L \times SU(1)_Y$ to explain the strong, weak, and electromagnetic interactions. In particular, the $SU(3)_C$ invariance results in the existence of "gluons" (g) as the mediators of the strong force, which is described by the quantum chromodynamics (QCD). The $SU(2)_L \times SU(1)_Y$ symmetry explains the origin of weak and electromagnetic forces, mediated by the $W^\pm$ and Z bosons and the photon (γ), respectively.

Matter is described in the SM by fermion fields of spin-$\frac{1}{2}$, which interactions are

mediated by spin-1 boson fields. Experimental observations show that twelve physical fermion fields exist, six "quark" fields and six "lepton" fields. They are organized in three families, made up of two quarks of electric charge $+\frac{2}{3}$ and $-\frac{1}{3}$ and and two leptons of electric charge -1 and 0. Fermions in one family and their counterparts in the others have identical properties and only differ in their mass. This is related to their coupling strength to the scalar field, that is described in the next section. To each fermion corresponds an antiparticle with identical properties but opposite quantum numbers.

Quarks Quarks are subject to all the three forces and, in particular, are the only fermions to possess a "colour" charge to which QCD owes its name. The first family of quarks is composed of the up (u) and down (d) quarks, with a mass of a few MeV. The former has a positive electric charge of $\frac{2}{3}$ while the latter has a negative electric charge of $\frac{1}{3}$. Being the lightest quarks, they are stable and compose the ordinary matter. Their counterparts in the second family are the charm (c) and the strange (s) quarks, of masses of about 1.28 GeV and 95 MeV respectively. Finally, the third family is composed of the top (t) and bottom (b) quarks, which masses are about 173 and 4.2 GeV respectively.

Because of the QCD colour confinement properties, quarks do not exist as free states but can be experimentally observed only as bound states. Collider experiments thus detect "mesons", that are composed by a quark-antiquark pair, or "baryons", composed by three quarks. The proton and neutron composing the ordinary matter are stable examples of the latter. Mesons and baryons are collectively denoted as hadrons. The creation of hadrons from a single quark produced in a collider experiment is a complex process that goes under the name of "hadronization". As its timescales, related to the QCD energy scales, are of the order of $10^{-24}\,s$, hard scatter and hadronization phenomena can be treated separately thanks to a factorization of their effects. The top quark represents an exception in this sense, as its lifetime is so short ($\approx 0.5 \times 10^{-24}\,s$) that it decays before bound states can be formed. Quark flavour is conserved in electromagnetic and strong interactions but not in weak ones, as quark mass eigenstates do not correspond to the weak interaction eigenstates. Their mixing is described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix.

Leptons Leptons have no colour charge and are subject only to the electromagnetic and weak forces. The charged leptons of the three families are respectively denoted as the electron (e), muon (μ) and tau lepton (τ). The electron is stable, being the lightest lepton with a mass of 511 keV. The muon has a mass of 105.7 MeV and a lifetime of 2.2 μs, that is sufficiently long to consider it as a stable particle at an LHC experiment given the detector size and the momentum of the particles produced in the collisions. Finally, the tau lepton has a mass of 1.8 GeV and a lifetime of $2.9 \times 10^{-13}\,s$, that is instead short enough for observing it only through its decay products. Most notably, the tau is the only lepton that has a sufficient mass to decay to leptons and hadrons.

To each lepton corresponds a neutrino, respectively denoted as ν_e, ν_μ, and ν_τ. Being electrically neutral, neutrinos interact with the matter only via the weak force and consequently they are not directly detectable at collider experiments. The observation of their flavour oscillations prove that neutrino masses are not zero as predicted by the SM: this has been the first observation of BSM physics. The mixing of weak and mass eigenstates is represented by the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix.

Strong Interaction QCD is built on the local gauge invariance under the $SU(3)_C$ group, the subscript denoting the relation to the colour charge arising from this symmetry. The free-field Dirac Lagrangian density of a massless spin fermion (the quark fields in this case) is:

$$\mathcal{L} = \bar{\psi(x)} \left(i\gamma^\mu \partial_\mu \right) \psi(x) \tag{5.1}$$

where ψ is the fermion field at the space-time coordinate x and γ^μ are the Dirac matrices. The following discussion on QCD is valid in presence of a $m\psi\bar{\psi}$ mass term. The reason to consider massless fermions is explained in the context of the electroweak interactions described in the next section. The fermion field transforms in the following way under the $SU(3)_C$ group:

$$\psi(x) \rightarrow e^{ig\frac{\lambda^a}{2}\theta_a(x)}\psi(x) \tag{5.2}$$

where $\frac{\lambda^a}{2}$ are the eight Gell-Mann matrices that generate the group. it is important to mention that the derivatives $\partial_\mu\psi(x)$ do not transform in the same way. Hence, the only way for the Lagrangian density 5.1 to be invariant under the transformation 5.2 is to redefine the derivative as covariant derivative:

$$D_\mu = \partial_\mu - igA^a_\mu(x)\frac{\lambda^a}{2} \tag{5.3}$$

where the gauge vector fields $A^a_\mu(x)$ correspond to the eight gluons that mediate the strong force.

To satisfy the local gauge invariance of the Lagrangian, the gluon fields must transform as:

$$A^a_\mu \rightarrow A^a_\mu + \partial_\mu\theta^a + gf^{abc}A^c_\mu\theta^c \tag{5.4}$$

The f^{abc} symbols denote the structure constants of the group that are defined from the commutation rules $\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = if^{abc}\frac{\lambda^a}{2}$. The introduction of the vector fields ensure that the covariant derivative transforms in the same way as the fermion fields and that, consequently, the Lagrangian density is invariant under the local gauge transformation. The Lagrangian density can be completed with a kinetic term for the gluon fields in the form:

$$-\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a \tag{5.5}$$

where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^a_\mu A^c_\nu \tag{5.6}$$

The QCD lagrangian becomes then:

$$\mathcal{L}_{QCD} = \bar{\psi}\left(i\gamma^\mu\partial_\mu\right)\psi - g\bar{\psi}\gamma^\mu\frac{\lambda_a}{2}\psi A^a_\mu - \frac{1}{4}F^{\mu\nu}_a F^a_{\mu\nu} \tag{5.7}$$

with a summation over all quark fields being implied.

The first term is the same as in the original Lagrangian and represents the free-field propagation of the quark. The second one originates from the introduction of the covariant derivative and represents the interaction of the quark with the vector field A_μ. The strength of the interaction is parametrized by the constant g, usually redefined as

the strong coupling constant $\alpha_s = \frac{g^2}{4}\pi$. The third term has been introduced as the kinetic term of the vector field.

The generators of the $SU(3)_C$ group do not commute and therefore the structure constants f^{abc} are not all zero. As a consequence, the $gf^{abc}A^b_\mu A^c_\nu$ term in Eq. 5.6, when inserted in the kinetic term of Eq. 5.7, result in cubic and quartic self-interactions of the gluon fields. Such interactions between the force mediators are a general property of non-abelian gauge theories.

Requiring the local gauge invariance led to the introduction of gauge bosons (the gluons) and to the description of their interactions with the fermion fields (the quarks). Choosing the $SU(3)_C$ group implies the presence of eight generators, the eight gluons, that are mathematically described by the adjoint representation of the group (**8**), and differ by the colour charge that they carry. Quarks and anti-quark are instead described in the simplest non-trivial representations of $SU(3)_C$, **3** and $\bar{\mathbf{3}}$ which explains the three colour quantum numbers of quarks. It is important to remark that explicit mass terms in the form $A^a_\mu A^\mu_a$ for the gauge bosons in Eq. 5.7 would break the gauge invariance of the Lagrangian.

Electroweak Interaction Electroweak interactions are explained in the SM with the same local gauge invariance mechanism as strong interactions by imposing a symmetry under the $SU(2)_L \times U(1)_Y$ group. Experimental observations show that parity is violated by weak interactions: this is accounted for in the theoretical description by assigning different interactions to fermions of opposite chiralities. The left and right chiral components of a field are defined from the $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ matrix. This is used to define left and right chirality projection operators as $\frac{1-\gamma^5}{2}$ and $\frac{1+\gamma^5}{2}$, respectively. In the limit of a massless particle, the chirality corresponds to the helicity, that is defined as the normalized projection of the spin vector onto the spatial momentum vector.

The $SU(2)_L$ gauge group is a non-abelian group to which the weak isospin quantum number (I_3) is associated. The gauge invariance under this group results in the presence of 3 gauge fields W^i_μ (i=1,2,3). Fermion fields of left chirality are represented by $SU(2)_L$ doublets while fermions of right chirality are $SU(2)_L$ singlets and do not interact with the W^i_μ fields.

The $U(1)_Y$ gauge group is abelian and is associated to the weak hypercharge Y. A single gauge field, denoted as B_μ, results from the $U(1)_Y$ local gauge invariance, and interacts with both ψ_L and ψ_R. The Gell-Mann–Nishijima formula determines the relation with the electric charge:

$$Q = I_3 + \frac{Y}{2} \tag{5.8}$$

Fields can therefore be represented as one doublet and two singlets:

$$\begin{aligned} \Psi_L = \frac{1-\gamma^5}{2}\begin{pmatrix}\psi\\ \psi'\end{pmatrix} = \begin{pmatrix}\psi_L\\ \psi'_L\end{pmatrix} \\ \psi_R = \frac{1+\gamma^5}{2}\psi \\ \psi'_R = \frac{1+\gamma^5}{2}\psi' \end{aligned} \tag{5.9}$$

The fields ψ and ψ' represent either the neutrino and charged lepton fields or the up– and down–type quark fields.

Under this notation, the Lagrangian can be written as a sum of three terms:

$$\mathcal{L} = i\bar{\Psi}_L \not{D} \Psi_L + i\bar{\psi}_R \not{D} \psi_R + i\bar{\psi}'_R \not{D} \psi'_R \tag{5.10}$$

The covariant derivative must be defined in this case as

$$D_\mu = \partial_\mu - igW^i_\mu T_i - ig' \frac{Y_\psi}{2} B^\mu \tag{5.11}$$

with $T_i = \frac{\sigma_i}{2}$, where σ_i are the Pauli matrices, generators of the $SU(2)_L$ group.

In this case two couplings, g and g', are present to account for the interaction with the two gauge fields of the $SU(2)_L$ and $U(1)_Y$ groups.

The introduction of the covariant derivative and of the following redefinition of the gauge fields:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}} \left(W^1_\mu \mp i W^2_\mu \right) \\ \sigma^\pm_\mu &= \frac{1}{2} \left(\sigma^1_\mu \mp i \sigma^2_\mu \right) \end{aligned} \tag{5.12}$$

allows to write the lagrangian as a sum of three terms:

$$\mathcal{L} = \mathcal{L}_{kin} + \mathcal{L}_{CC} + \mathcal{L}_{NC} \tag{5.13}$$

The explicit definition of these terms is the following:

$$\begin{aligned} \mathcal{L}_{\text{kin}} &= i\bar{\Psi}_L \not{\partial} \Psi_L + i\bar{\psi}_R \not{\partial} \psi_R + i\bar{\psi}'_R \not{\partial} \psi'_R \\ \mathcal{L}_{\text{CC}} &= g W^1_\mu \bar{\Psi}_L \gamma^\mu \frac{\sigma_1}{2} \Psi_L + g W^2_\mu \bar{\Psi}_L \gamma^\mu \frac{\sigma_2}{2} \Psi_L \\ &= \frac{g}{\sqrt{2}} W^+_\mu \bar{\Psi}_L \gamma^\mu \sigma^+ \Psi_L + \frac{g}{\sqrt{2}} W^-_\mu \bar{\Psi}_L \gamma^\mu \sigma^- \Psi_L \\ &= \frac{g}{\sqrt{2}} W^+ \bar{\psi}_L \gamma^\mu \psi'_L + \frac{g}{\sqrt{2}} W^- \bar{\psi}'_L \gamma^\mu \psi_L \\ \mathcal{L}_{\text{NC}} &= \frac{g}{\sqrt{2}} W^3_\mu \left[\bar{\psi}_L \gamma^\mu \psi_L - \bar{\psi}'_L \gamma^\mu \psi'_L \right] \\ &\quad + \frac{g'}{\sqrt{2}} B_\mu \left[Y_{\Psi_L} (\bar{\psi}_L \gamma^\mu \psi_L + \bar{\psi}'_L \gamma^\mu \psi'_L) + Y_{\psi_R} \bar{\psi}_R \gamma^\mu \psi_R + Y_{\psi'_R} \bar{\psi}'_R \gamma^\mu \psi'_R \right] \end{aligned}$$

The suffix "CC" stands for "charged current": this term describes the interaction of the fermions (fields ψ_L and ψ'_L) with the charged weak bosons (fields $W^\pm_\mu$).

The suffix "NC" stands for "neutral current": in order to better understand the meaning of this term the fields B_μ and W^3_μ must be expressed as combination of the physical fields Z_μ(the Z boson field) and A_μ(the photon field) by means of the Weinberg angle θ_μ:

$$\begin{aligned} B_\mu &= A_\mu \cos\theta_w - Z_\mu \sin\theta_w \\ W^3_\mu &= A_\mu \sin\theta_w + Z_\mu \sin\theta_w \end{aligned} \tag{5.14}$$

Inserting this parametrization in the last term of the langrangin the neutral current interaction with the photon field A_μ is determined by the coupling

$$g \sin\theta_w I_3 + g' \cos\theta_w \frac{Y}{2} \tag{5.15}$$

The unification of the weak and electromagnetic forces is completed if this expression is required to be equal to the coupling constant of the photon field eQ. Y_{ψ_L} can be arbitrarily set to -1 as the hypercharge only appears multiplied by g'. The equation can be solved by using $Q = 0$ for the neutrino fields and $Q = -1$ for the lepton fields: this leads to equation 5.8.

As already observed for the strong interactions, explicit mass terms of the gauge fields would break the gauge invariance. Direct fermion mass terms are also not allowed, because they are not invariant under the gauge transformation: the left and right chiralities of the fields transform differently under $SU(2)_L \times U(1)_Y$.

This formulation of the SM is of extraordinary beauty and elegance. Matter fields are completely described in terms of quantum numbers, and their interactions follow from the application of a symmetry principle to the Lagrangian density. Differences in the representation of the fields under a specific group completely determine the phenomenology that we observe experimentally. The electromagnetic and weak force are jointly described, and all the relevant forces at the subnuclear scale are explained from a common symmetry principle. However, the theory requires both the fermions and the gauge bosons to be massless, as any explicit mass term would violate the gauge invariance itself. This is in clear contrast with the experimental observation of massive weak bosons and fermions. A simple addition "a posteriori" of the mass terms produces a theory that is not renormalizable, and results in unhphysical predictions for scattering of longitudinally polarized vector bosons. The solution needed to ensure the unitarity of the theory and to explain bosons and fermions masses is provided by the Brout-Englert-Higgs mechanism, a natural way of breaking the $SU(2)_L \times U(1)_Y$ symmetry to $U(1)_{em}$ without explicitly violating the local gauge invariance.

5.1.2 The Brout-Englert-Higgs Mechanism

The Brout-Englert-Higgs (BEH) mechanism was proposed in 1964 independently by physicists Englert and Brout[47], Higgs[48], and also by Guralnik, Hagen, and Kibble[49] as a solution to generate the gauge boson masses and explain the fermion masses in the Standard Model. The mechanism is based on the concept of spontaneous symmetry breaking, a phenomenon that is often observed in Nature whenever individual ground states of a system do not satisfy the symmetries of the system itself.

In the BEH mechanism, spontaneous symmetry breaking is realized through the introduction of a complex scalar doublet of fields:

$$\Phi = \begin{pmatrix} \phi_0 \\ \phi_+ \end{pmatrix} \tag{5.16}$$

The field must be scalar in order to satisfy space isotropy and its expectation value on the vacuum must be constant to satisfy space homogeneity.

The proposed Φ field has hypercharge -1 and its covariant derivative is defined as:

$$D_\mu = \partial_\mu - igW_\mu^i \frac{\sigma_i}{2} - i\frac{g'}{2} B_\mu \tag{5.17}$$

The BEH lagrangian can be written as:

$$\mathcal{L} = \left(D_\mu \Phi\right)^\dagger \left(D^\mu \Phi\right) - V\left(\Phi^\dagger \Phi\right) \tag{5.18}$$

where the potential $V\left(\Phi^\dagger \Phi\right)$ has the form:

$$V\left(\Phi^\dagger \Phi\right) = -\mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right) \tag{5.19}$$

All the doublets that satisfy the condition:

$$\left|\Phi^2\right| = \frac{\mu^2}{2\lambda} = \frac{v^2}{2} \tag{5.20}$$

are minima of these potential, and are connected through gauge transformations that change the phase of the field Φ but not its modulus. The quantity v is called the vacuum expectation value (VEV) of the scalar potential. Once a specific ground state is chosen, the symmetry is explicitly broken but the Lagrangian is still gauge invariant with all the important consequences for the existence of gauge interactions.

If the symmetry is spontaneously broken to the ground state that is parallel to the ϕ_0 component of the doublet, it can be shown that this specific ground state is still invariant under the $U(1)_{em}$ symmetry group. As a consequence, the field expansion around this minimum is written as:

$$\Phi(x) = \frac{1}{\sqrt{2}} exp\left[\frac{i\sigma_i \theta^i(x)}{v}\right] \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{5.21}$$

This corresponds to the presence of a scalar real massive field H and of three massless fields θ^i. The latter are expected as consequence of the Goldstone theorem[50] that states that the spontaneous breaking of a continuous symmetry generates as many massless bosons (the Goldstone boson) as broken generators of the symmetry. However, such massless bosons are not observed in Nature. They can be removed with an $SU(2)_L$ transformation that consists in the choice of a specific gauge called "unitary gauge":

$$\Phi(x) \rightarrow \Phi(x)' = exp\left[\frac{-i\sigma_i \theta^i(x)}{v}\right] \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{5.22}$$

After this transformation, only the real scalar field $H(x)$ remains and its quanta correspond to a new physical massive particle, the Higgs boson (H).

The BEH lagrangian can be written in a more explicit way by introducing the expression of the covariant derivative and applying the unitary gauge:

The first line represents the evolution of the scalar Higgs field that has a mass $m_H^2 = 2\lambda v^2 = 2\mu^2$. It is a free parameter of the theory, directly related to the parameter μ of the scalar potential.

The second line represents the mass terms of the weak bosons (those that multiply the constant term), of mass:

$$\begin{aligned} m_W^2 &= \frac{g^2 v^2}{4} \\ m_Z^2 = \frac{(g^2 + g^{2\prime})v^2}{4} &= \frac{m_W^2}{\cos\theta_w^2} \end{aligned} \tag{5.23}$$

$$\mathcal{L}_{\text{BEH}} = \frac{1}{2}\partial^{\mu} H \partial_{\mu} H - \frac{1}{2}\left(2\lambda v^2\right) H^2 + \left[\left(\frac{gv}{2}\right)^2 W^{\mu+} W^{-}_{\mu} + \frac{1}{2}\frac{(g^2+g'^2)v^2}{4} Z^{\mu} Z_{\mu}\right]\left(1+\frac{H}{v}\right)^2 + \lambda v H^3 + \frac{\lambda}{4} H^4 - \frac{\lambda}{4} v^4$$

It can be observed that the Goldstone bosons, removed with the transformation of the unitary gauge, are absorbed as additional degrees of freedom of the $W^{\pm}$ and Z bosons, corresponding to their longitudinal polarizations: the mechanism confers mass to the weak bosons. The second line also describes the interactions of the weak bosons with the Higgs field. There are a HWW and a HZZ interactions from the $\frac{2H}{v}$ term and a HHWW and a HHZZ interaction from the $\frac{H^2}{v^2}$ term.

The third line shows that cubic and quartic self-interactions of the Higgs boson are predicted. The BEH potential can be rewritten in terms of a trilinear and a quadrilinear coupling as:

$$V(H) = \frac{1}{2} m_H^2 H^2 + \lambda_{HHH} v H^3 + \frac{1}{4}\lambda_{HHHH} H^4 - \frac{\lambda}{4} v^4 \tag{5.24}$$

with the two couplings constants defined as:

$$\lambda_{HHH} = \lambda_{HHHH} = \lambda = \frac{m_H^2}{2v^2} \tag{5.25}$$

An important remark is that both Higgs boson self-couplings are directly related to the parameters of the scalar potential and are entirely determined from the Higgs boson mass and the VEV. Their measurement thus represents a test of the validity and coherence of the SM. In a wider perspective, the Higgs boson self-couplings have no equal in the SM: in contrast to the weak boson self-interactions, that have a gauge nature, the Higgs boson self-interactions are purely related to the scalar sector of the theory. They are responsible for the mass of the Higgs boson itself, as shown in the mass term of the Lagrangian. Their experimental determination is thus crucial to reconstruct the Higgs boson potential and explore the nature of the EWSB.

There are at this point two free parameters of the BEH mechanism: the VEV v and the Higgs boson mass m_H. The first corresponds to the energy scale of the electroweak symmetry breaking and can be computed from the Fermi constant G_F that is precisely determined from the muon lifetime:

$$v = \sqrt{\frac{1}{\sqrt{2} G_F}} \approx 246\, GeV \tag{5.26}$$

The mass of the Higgs boson m_H has to be determined experimentally.

Fermions have been assumed to be massless until now. Mass terms are generated by the Higgs field itself through a Yukawa interaction that couples the left and right chiral fields.

Once the EWSB mechanism is taken into account the Yukawa lagrangian for the fermions can be written as:

$$\mathcal{L}_{Yukawa} = -\sum_f m_f \left(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L\right)\left(1 + \frac{H}{v}\right) \tag{5.27}$$

where the sum runs on both up– and down–type fermions and the mass terms are:

$$m_f = y_f \frac{v}{\sqrt{2}} \tag{5.28}$$

Fermion masses are thus explained in the SM as the interaction of the fermion fields with the Higgs field, which changes the chirality of the fermions. The strengths of the interactions are directly related to the fermion masses, and are free parameters of the theory. An important remark is that the SM does not explain the origin of these couplings and, consequently, the hierarchy of the three fermion families.

In conclusion, the BEH mechanism solves the aforementioned problems of the electroweak theory of the SM. Upon breaking the electroweak symmetry, the scalar field generates Goldstone bosons that are absorbed as degree of freedoms of the vector boson fields, which become massive. The Higgs boson contributions to the quantum loops in the scattering of longitudinally polarized vector bosons regularizes the process and ensure its unitarity at the TeV scale and beyond. Finally, the Higgs boson couples the left and right chiral components of the fermion fields in a Yukawa interaction, determining the fermion masses with a purely quantum-relativistic mechanism. Finally, the theory obtained by incorporating the BEH mechanism in the electroweak theory is renormalizable, as demonstrated by 't Hooft and Veltman[51].

5.1.3 Higgs Boson Pair Production in and beyond the SM

Two production mechanisms of Higgs boson pairs are considered in this thesis: Gluon Fusion (GF) and Vector Boson Fusion (VBF).

The first mechanism is the one with the highest cross section at proton-proton colliders. It involves either the production of a Higgs boson pair through the trilinear Higgs boson self-coupling, or the radiation of two on-shell Higgs bosons from a heavy quark loop. The cross section consequently depends on λ_{HHH} and on the top quark Yukawa couplings y_t. The contribution from b quarks is smaller than 1% at leading order and can be neglected given the current accuracy of the theoretical computations and the experimental sensitivity. The corresponding Feynman diagram is report in Fig. 5.1.

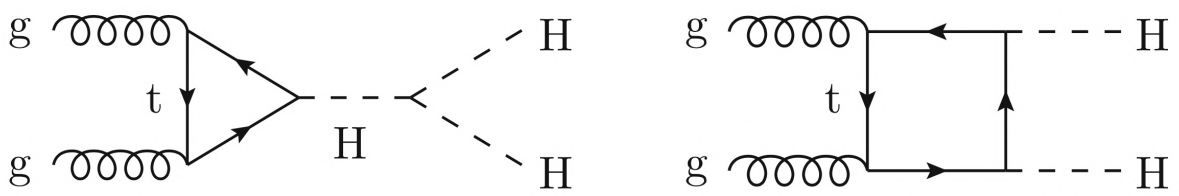


Figure 5.1: Feyman diagrams for HH production via Gluon Fusion.

The VBF process, in addition to the Higgs boson trilinear coupling, also depends on the quadrilinear coupling of a Higgs boson pair to a vector boson pair as well as on the single Higgs boson coupling to vector bosons. Despite its cross section being one order of

magnitude smaller than the gluon fusion one, the two additional final state jets provide a clean signature that can be used to discriminate signal events from background. The corresponding Feynman diagram is reported in Fig. 5.2.

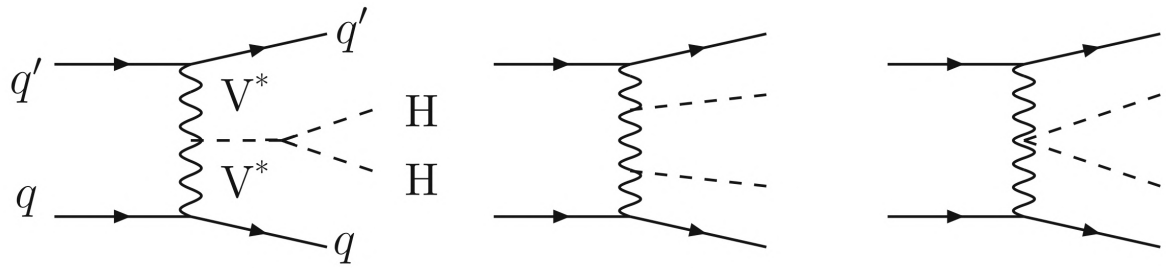


Figure 5.2: Feyman diagrams for HH production via Vector Boson Fusion.

Higgs boson pairs can be produced also in association with a pair of top quarks or vector bosons or with a single top quark. These production mechanisms are very rare, as can be seen in Fig. 5.3, where the cross sections of the five mechanisms are reported as a function of the center-of-mass energy.

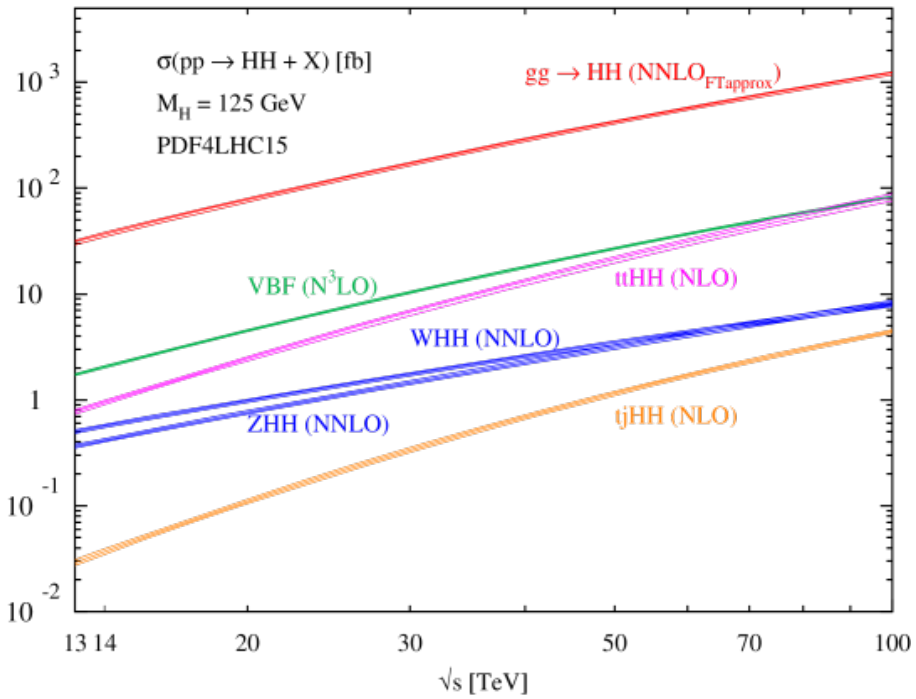


Figure 5.3: Total cross section for HH production in pp collisions for the production modes described in the text. The size of the bands shows the total uncertainties originating from the scale dependence and the PDF+α_S uncertainties.

An important property of the gluon fusion production channel should be highlighted at this point. The two production diagrams discussed before have amplitudes that are about the same order of magnitude, but interfere destructively. Combined with the restricted phase space of production of two Higgs bosons, this results in the small cross section discussed above. However, the destructive interference makes HH production extremely sensitive to physics beyond the SM (BSM). BSM physics contributions might alter the destructive interference and produce large modifications that can be probed with the current LHC data.

If the scale of BSM physics is at the LHC reach, new states can be directly produced and subsequently decay to a HH pair. The experimental signature of this resonant production mechanism is an enhancement of σ_{HH} at a specific value of m_{HH}, corresponding to mass of the resonance. If instead the scale of BSM physics is significantly higher than the LHC centre-of-mass collision energy, its effects could still be observed as a nonresonant enhancement of the production cross section, due to either new particles in the quantum loops or to anomalous Higgs boson couplings.

The value of λ_{HHH} is completely determined in the SM once the values of v and m_H are known. However, several BSM models predict a modification of the trilinear Higgs boson self-coupling, modifying the properties of HH production. This can provide the first hints at the LHC of the presence of BSM physics, and serve as an important criterion to discriminate between alternative models. In this context, a parametric approach is adopted and consist in considering the λ_{HHH} value as a free parameter. Deviations from the SM prediction are quantified with the ratio $k_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}}$. This coupling rescaling approach is usually referred to as k–framework.

The modification of the value of λ_{HHH} has a profound impact on the HH production cross section. The most recent calculation predicts the following dependence:

$$\sigma = 70.3874 - 50.4111 \times k_\lambda + 11.0595 \times k_\lambda^2 \ [fb] \tag{5.29}$$

Some values of the cross section for special values of k_λ are reported in Fig. 5.4.

c_{hhh}	-1	0	1	2	2.4	3	5
σ [fb]	$131.9^{+2.5\%}_{-6.7\%}$	$70.38^{+2.4\%}_{-6.1\%}$	$31.05^{+2.2\%}_{-5.0\%}$	$13.81^{+2.1\%}_{-4.9\%}$	$13.1^{+2.3\%}_{-5.1\%}$	$18.67^{+2.7\%}_{-7.3\%}$	$94.82^{+4.9\%}_{-8.8\%}$
$\sigma/\sigma^{\mathrm{SM}}$	4.25	2.27	1	0.445	0.422	0.601	3.05
$\sigma/\sigma_{\mathrm{NLO}}$	1.13	1.13	1.12	1.11	1.12	1.15	1.16

Figure 5.4: Inclusive GF cross-sections for Higgs boson pair production at 13 TeV, for different values of the Higgs self-coupling modifier k_λ (c_{hhh} in the table), obtained for m_H = 125 GeV with the central scale $\mu_0 = \mu_R = \mu_F = \frac{M_{HH}}{2}$ at NNLO [52].

Variations of the k_λ value not only affect the cross section but also the HH pair kinematics, as illustrated in Fig. 5.5 for the m_{HH} distribution under a few representative values of k_λ. The SM case, corresponding to the green curve, shows a broad peak around m_{HH} = 400 GeV. Its shape is the result of the interference of the two "triangle" and "box" diagrams described previously. The former, that involves λ_{HHH}, has an important role to populate the low m_{HH} region (black curve, $k_\lambda = 20$) while the latter significantly contributes to the high m_{HH} tail (red curve,$k_\lambda = 0$). The interference effects strongly influence the m_{HH} distribution as λ_{HHH} is modified. A soft m_{HH} spectrum is observed for $k_\lambda = 5$ (yellow curve), while in correspondence of the maximal interference for k_λ= 2.45 (blue curve) a characteristic double peak structure is observed. These effect have important consequences for the experimental searches, that are sensitive to anomalous λ_{HHH} couplings through both the total HH production cross section and the kinematic distribution of HH events.

5.1.4 Searches for Higgs Boson Pair Production at the LHC

Measuring the production of Higgs boson pairs at the LHC requires to reconstruct their decay products in the detector and to discriminate them from a large background. The

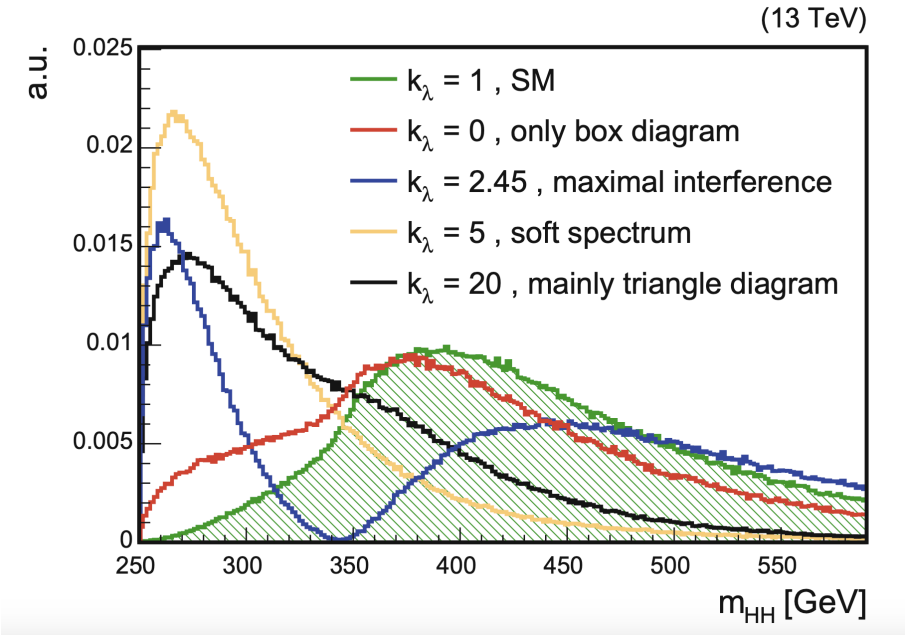


Figure 5.5: Comparison of the m_{HH} distribution for different values of k_λ.

choice of the decay channel of the HH system is crucial in this sense and determines a different trade-off between the branching fraction and the background contamination. Higgs boson pair production, at least in the context of the SM, is characterized by tiny cross sections, so that decay channels with a sizeable branching fraction are preferred. The decay branching fractions for some selected HH final states are shown in Fig. 5.6.

The sensitivity to HH production at the LHC is driven by three main channels:

- $HH \to b\bar{b}\gamma\gamma$ is a very pure final state but suffers from a small branching fraction. The clean signature of the photon pairs results in a high signal selection efficiency and provides a powerful tool to separate signal events from the background through the use of the photon pair invariant mass.

- $HH \to b\bar{b}\tau^+\tau^-$ represents an optimal compromise between the branching fraction and the background contamination. Nevertheless the reconstruction of the τ lepton hadronic decays poses non-trivial issues that must be solved with proper offline identification algorithms.

- $HH \to b\bar{b}b\bar{b}$ is characterized by the highest branching fraction but is affected by a copious multijet background. It has consequently little sensitivity to low m_{HH} values but can profit from the large signal yields to probe regions up to $m_{HH} \approx 3\,TeV$.

Many other final states can be studied at the LHC to improve the sensitivity of experiments to HH production. Those listed above represent nevertheless the decay channels that are considered the most sensitive and which combination can ensure the largest coverage of the possible HH topologies. Experimental challenges are very different depending on the final state considered. The exploration of HH production in its $b\bar{b}b\bar{b}$ final state crucially relies on the capability to identify jets from b quarks and to reject instrumental background from the misidentification of gluon or light flavour quark jets. Inversely, final states such as $b\bar{b}\gamma\gamma$ are mostly affected from irreducible backgrounds, that

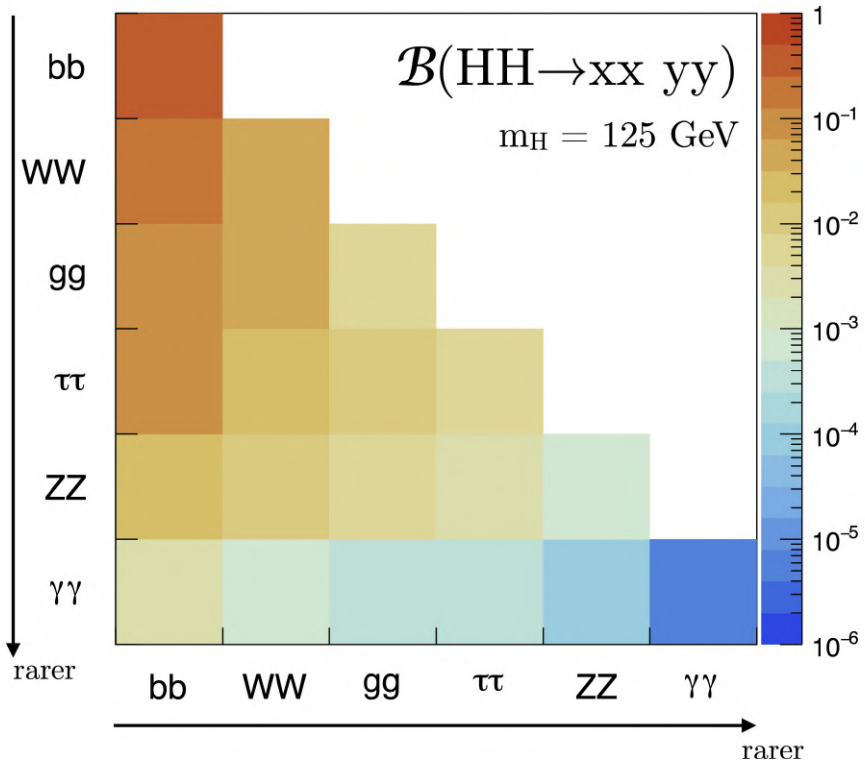


Figure 5.6: Branching fractions for the decay of a HH pair to a selected group of final states. The decays of the two Higgs bosons are indicated in the two axes of the figure.

can be statistically suppressed only by exploiting the kinematic properties of the selected events.

The $b\bar{b}\tau^+\tau^-$ final state, that is the topic of this thesis, represents an intermediate and particularly interesting situation. As the tau lepton is unstable and can decay to either leptons or hadrons in associations to neutrinos, the searches must exploit several final states. Neutrinos from τ decays do not allow for a complete reconstruction of the event, and final states where the tau leptons decay to hadrons and neutrinos must be distinguished from instrumental backgrounds caused by the misidentification of a quark or gluon jet. At the same time, irreducible background contamination also affects this decay channel and calls for the usage of the event kinematic properties to reduce it. For these reasons, the $b\bar{b}\tau^+\tau^-$ decay channel is probably one of the most challenging at the LHC, but the effort is rewarded by one of the best sensitivities over several resonant and nonresonant HH signal hypotheses, as pointed out in several phenomenological studies.

5.2 Trigger requirements

Events are recorded and stored using a set of HLT triggers, or "paths", that require the presence of specific objects in the final state: the trigger paths used in this analysis are tuned to look for the decay products of the tau leptons. The decay of a tau lepton pair can happen in six different channels, reported in Table 5.1 together with the relative branching fractions.

The $b\bar{b}\tau^+\tau^-$ search is performed exclusively in three final states: $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$,

Decay mode	Branching fraction(%)
$\tau_\mu\tau_h$	22.5
$\tau_e\tau_h$	23.1
$\tau_h\tau_h$	42.0
$\tau_e\tau_\mu$	6.2
$\tau_\mu\tau_\mu$	3.0
$\tau_e\tau_e$	3.2

Table 5.1: Decay modes and relative branching fraction of a $\tau\tau$ pair.

which in total cover about 88% of the decays. Fully leptonic channels are neglected in this search due to their smaller branching fractions and the huge contamination of background events coming from the Drell-Yan processes $Z/\gamma^* \to \mu^+\mu^-/e^+e^-$.

The HLT trigger paths are listed in Tables 5.2, 5.3 and 5.4.

Channel	HLT path name	runs
$\mu\tau_h$	HLT_IsoMu22	all runs
	HLT_IsoMu22_eta2p1	all runs
	HLT_IsoTkMu22	all runs
	HLT_IsoTkMu22_eta2p1	all runs
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20	all runs
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1	all runs
$e\tau_h$	HLT_Ele25_eta2p1_WPTight_Gsf	all runs
$\tau_h\tau_h$	HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_Reg	runs B to G
	HLT_DoubleMediumCombinedIsoPFTau35_Trk1_eta2p1_Reg	run H

Table 5.2: Trigger paths used in the $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$ channels for 2016 data and Monte Carlo simulation.

Whenever multiple trigger paths cover the same runs they are combined in a logical OR to decide if an event can be accepted or not.

Muons reconstruction at the HLT level starts from L1 trigger candidates and the isolation of such candidates can be computed using either information from ECAL and HCAL deposit or the tracks information: this is the main difference between the paths defined for 2016.

In the $e\tau_h$ final state, the electron required by the HLT path is reconstructed with a similar approach to the offline strategy and its isolation is computed from the scalar sum of the energy clusters and tracks in a cone of size $\Delta R < 0.3$ around the candidate.

Finally, in the $\tau_h\tau_h$ channel, two τ_h objects are required at trigger level. The candidates are built from charged hadrons and π_0 candidates in an approach similar, but simplified due to timing constraints, to the offline HPS algorithm, detailed in Section 1.2.5. In 2018 an online version of the HPS algorithm has been developed and deployed starting from run 317509, as marked in table 5.4. It is important to mention that in 2017 a trigger was introduced to collect events containing a pair of tau leptons coming from the decay of a Higgs boson produced via Vector Boson Fusion (VBF): this trigger has a lower p_T threshold for the tau candidates (20 GeV) and specific requirements on the jets p_T and invariant mass.

In 2017 and 2018 the single object triggers (muons and electrons), differ with respect to the ones of 2016 by an increase in the p_T thresholds of the objects in order to reduce the

Channel	HLT path name	runs
$\mu\tau_h$	HLT_IsoMu24	all runs
	HLT_IsoMu27	all runs
	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1	all runs
$e\tau_h$	HLT_Ele32_WPTight_Gsf_L1DoubleEG	all runs
	HLT_Ele32_WPTight_Gsf	all runs
	HLT_Ele35_WPTight_Gsf	all runs
	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau30_eta2p1_CrossL1	all runs
$\tau_h\tau_h$	HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg	all runs
	HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg	all runs
	HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg	all runs
	HLT_VBF_DoubleLooseChargedIsoPFTau20_Trk1_eta2p1_Reg	run D to F

Table 5.3: Trigger paths used in the $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$ channels for 2017 data and Monte Carlo simulation.

Channel	HLT path name	runs
$\mu\tau_h$	HLT_IsoMu24	all runs
	HLT_IsoMu27	all runs
	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1	run $<$ 317509
	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTauHPS27_eta2p1_CrossL1	run $\geq$ 317509
$e\tau_h$	HLT_Ele32_WPTight_Gsf	all runs
	HLT_Ele35_WPTight_Gsf	all runs
	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau30_eta2p1_CrossL1	run $<$ 317509
	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTauHPS30_eta2p1_CrossL1	run $\geq$ 317509
$\tau_h\tau_h$	HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg	run $<$ 317509
	HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg	run $<$ 317509
	HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg	run $<$ 317509
	HLT_DoubleMediumChargedIsoPFTauHPS35_Trk1_eta2p1_Reg	run $\geq$ 317509
	HLT_VBF_DoubleLooseChargedIsoPFTau20_Trk1_eta2p1	run $<$ 317509
	HLT_VBF_DoubleLooseChargedIsoPFTauHPS20_Trk1_eta2p1	run $\geq$ 317509

Table 5.4: Trigger paths used in the $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$ channels for 2018 data and Monte Carlo simulation. Since multiple paths cover the same runs, the logical OR of all these paths is used.

trigger rate and cope with the higher instantaneous luminosity reached by the machine. The loss of acceptance due to the tighter selections was mitigated by the introduction of the so-called "cross-lepton" triggers, which save events only if both an isolated lepton, electron or muon, and one τ_h object are found.

MC simulated events are required to pass the same trigger selections. To account for systematic differences in the data and MC efficiencies, "scale factors" (SFs) computed with a tag and probe technique, using $Z/\gamma^* \to \mu^+\mu^-/e^+e^-$, and applied in MC events to the selected leptons. The tag and probe technique uses tight trigger, reconstruction and identification selections to identify one "tag" lepton, while it exploits the kinematics of the $Z \to l^+l^-$ decay to identify the other "probe" lepton, without directly applying trigger criteria on it. The probe leptons are thus unbiased with respect to the trigger requirements and can be used to compute the trigger efficiency itself.

For the $\tau_h\tau_h$ final state, the trigger efficiencies and SFs are measured using $Z \to \tau\tau \to \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ events selected with a tag and probe technique and cover the logical OR of the different trigger paths used.

For the $\mu\tau_h$ and $e\tau_h$ final states, the SFs must take into account the efficiency of the logical OR between single and cross triggers: assuming the efficiencies of the two legs to be independent, the efficiency of the logic OR can be factorized and easily computed from the single objects efficiencies. The resulting event by event SF formula is:

$$SF = \frac{Eff_{DATA}}{Eff_{MC}}$$

where the efficiency for both MC and data is:

$$Eff = passSingle \times \epsilon_L + passCross \times \epsilon_l\,\epsilon_\tau - passSingle \times passCross \times min(\epsilon_l, \epsilon_{tau}) \times \epsilon_\tau$$

where
ϵ_L = single lepton trigger efficiency
ϵ_l = cross lepton trigger efficiency for the τ_e or τ_μ leg
ϵ_τ = cross lepton trigger efficiency for the τ_h leg

As an example the efficiency and SF for the tau triggers employed in 2018 are reported in Fig. 5.7.

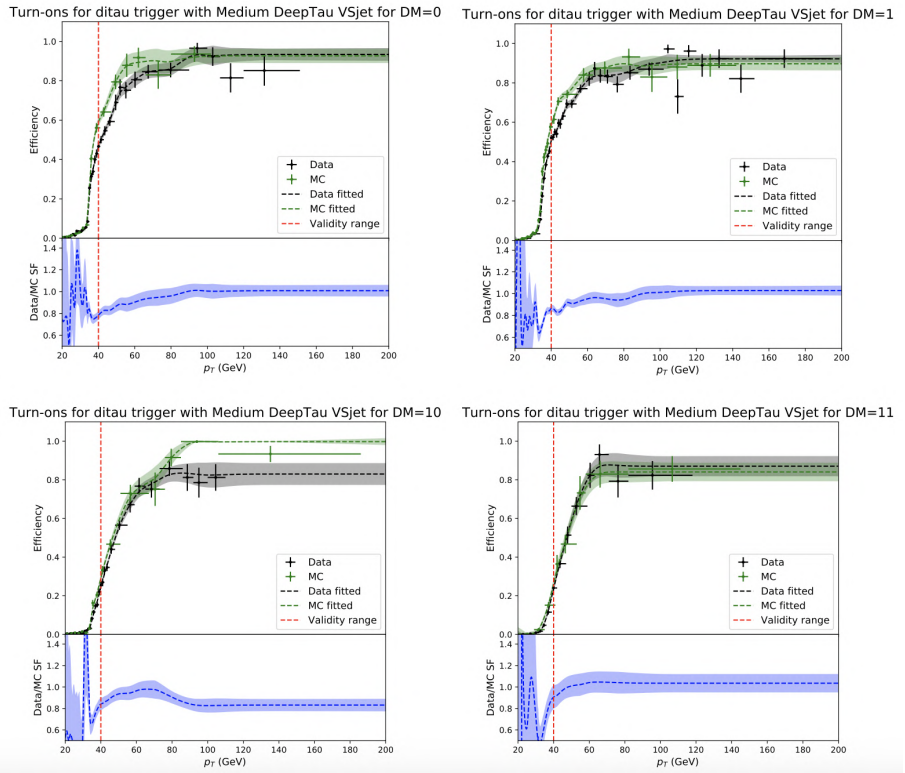


Figure 5.7: Efficiency and Data/MC Scale Factors computed for the tau triggers employed in 2018 using the Tag And Probe method.

5.3 Object selections

Quality criteria are applied to the reconstructed muons, electrons, τ_h objects, jets and missing transverse momentum, in order to optimize the selection of real $b\bar{b}\tau^+\tau^-$ events.

This Section describes the specific choices of the final state objects as well as the corrections applied to the Monte Carlo simulated events to eliminate any possible discrepancy. Most of these criteria and corrections are made available by CMS groups taking care of the reconstruction of each particular object, referred to as "Physics Object Groups", POGs.

5.3.1 Primary vertex selection and pile-up treatment

During the LHC Run 2 there were an avarage of 23, during 2016, and 32, during 2017 and 2018, proton-proton interactions per bunch crossing. Interaction vertices are identified using tracks clustered with the deterministic annealing algorithm [53, 54]. The vertex with the highest sum of p_T^2 of the associated tracks that passes certain quality criteria is selected as the primary vertex. The other vertices in the event are considered to be originated from the pileup collisions (PU).

The distribution of the number of interaction vertices in the Monte Carlo events does not match exactly the one of the data. In order to improve the agreement between data and simulations a weight must be computed and applied to each event. These weights come from a bin-by-bin ratio between the data and Monte Carlo distributions of the number of PU vertices. Data distribution is computed assuming a total cross section for proton-proton interactions of 69.2 mb.

5.3.2 Electrons

The standard CMS electron reconstruction algorithm is used for this analysis [20]. Electron candidates are reconstructed from clusters of energy deposits in the ECAL, which are then matched to tracks in the inner silicon tracker refitted by a Gaussian-Sum Filter (GSF).

The isolation of electron candidates is defined as the sum of the transverse momenta of Particle Flow particles inside a $\Delta R < 0.3$ cone around the electron relative to the transverse momentum p_T^ℓ of the electron. Three isolation components are considered separately for the electron identification: the scalar sums of the transverse momenta of charged hadrons, neutral hadrons and photons, originating from the primary vertex.

The electron identification uses a multivariate approach (MVA) that has been updated and improved for the Run2 analysis. The discriminator is based on a Boosted Decision Tree (BDT) that combines observables sensitive to the amount of bremsstrahlung along the electron trajectory, the three PF isolation components, the energy density within the isolation cone, the geometrical and momentum matching between the electron trajectory and associated clusters, shower-shape observables, and electron conversion variables. The General Purpose MVA electron identification is used, which is tuned for best performance and is trained on all electrons, regardless if they pass the trigger requirements or not. This MVA is tuned for electrons with $p_T > 10$ GeV and defines three categories based on electron eta: two for the barrel and one for the endcap. For each MVA category, a tight working point with 80% of signal efficiency provided by the E/gamma POG is used. In order to reject events with an extra electron, a tight working point with 90% of signal efficiency is required for veto electrons.

A correction factor is applied to the MC simulation to take into account differences with respect to data in identification efficiencies of electrons. These scale factors are

derived from $Z \to ee$ events selected with a tag and probe technique and are provided by the E/gamma POG.

The efficiency and SF of the MVA identification algorithm computed on 2017 data are reported in Fig. 5.8 for the working point used in this analysis.

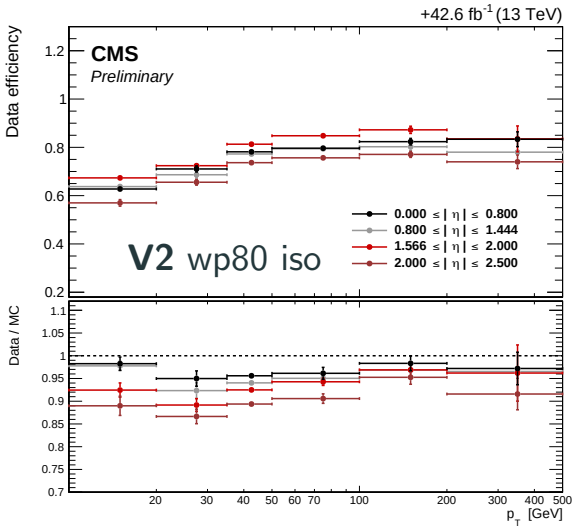


Figure 5.8: Efficiency and Data/MC Scale Factors computed on 2017 data for the MVA identification working point used in this analysis.

5.3.3 Muons

Muons are required to be reconstructed by the Tracker or the Global muon reconstruction algorithm [10]. The Muon POG provides three recommended identification selections for the particle-flow muons: *loose*, *medium* and *tight*. In this analysis the signal muon candidates are required to pass the tight identification criteria while, in order to reject events with an extra muon, the medium or tight identification is required for veto muons.

The isolation of muon candidates is defined as the sum of the transverse momenta of PF particles inside a $\Delta R < 0.4$ cone around the muon relative to the transverse momentum p_T^ℓ of the muon:

$$\mathrm{I}_{rel}^{\ell} = \left(\sum p_T^{\mathrm{charged}} + \max\left[0, \sum p_T^{\mathrm{neutral-had}} + \sum p_T^{\gamma} - \frac{1}{2}\sum p_T^{\mathrm{PU}}\right]\right) / p_T^{\ell}, \qquad (5.30)$$

where $\sum p_T^{\mathrm{charged}}$, $\sum p_T^{\mathrm{neutral-had}}$, and $\sum p_T^{\gamma}$ are the scalar sums of the transverse momenta of charged hadrons neutral hadrons and photons originating from the primary vertex, respectively; $\sum p_T^{\mathrm{PU}}$ is the sum of transverse momenta of charged hadrons not originating from the primary vertex. In this analysis, the signal muon candidates are required to have $\mathrm{I}_{rel}^{\ell} < 0.15$, while for veto muons $\mathrm{I}_{rel}^{\ell} < 0.3$ is required.

A correction factor is applied to the MC simulation to take into account differences with respect to data in the isolation and identification efficiencies of muons. These scale factors are derived from $Z \to \mu\mu$ events selected with a tag and probe technique and are provided by the Muon POG.

The efficiency and SF of the identification and isolation working point used in this analysis are reported in Fig. 5.9.

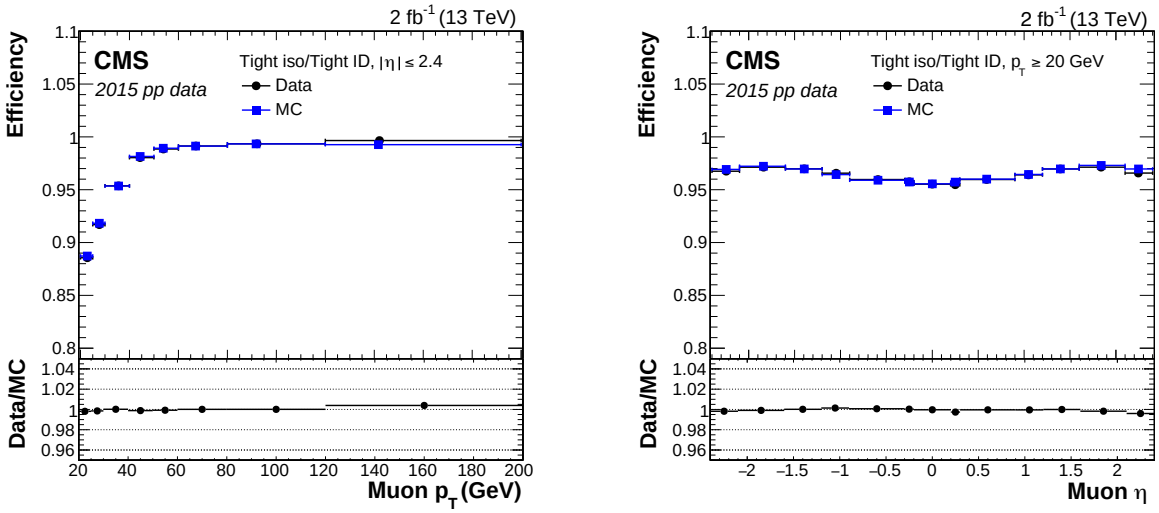


Figure 5.9: Tag-and-probe efficiency for the tight PF isolation working point on top of the tight ID (left) versus p_T for muons in the acceptance of the muon spectrometer, and (right) versus pseudorapidity for muons with $p_T > 20GeV$, for 2015 data (circles), simulation (squares), and the ratio (bottom inset). The statistical uncertainties are smaller than the symbols used to display the measurements.

5.3.4 Hadronic taus

Decays of τ leptons into hadrons and a neutrino are reconstructed by the Hadrons Plus Strips (HPS) identification algorithm [21]. The constituents of the jets are analyzed in order to identify individual tau hadronic decay modes (τ_h). τ_h candidates are required to be reconstructed in either one of the decay modes: Single Hadron, Hadron plus one Strip, Hadron plus two Strips or Three Hadrons. For this analysis, taus are required to pass the `decayModeFindingNewDMs` discriminator, that associates to each candidate the most probable decay mode.

In order to correctly identify hadronic τ decays from from quark and gluon jets, and from electrons and muons, the *DeepTau* algorithm is used [55]. In this analysis the version denoted with *2017v2p1* is used.

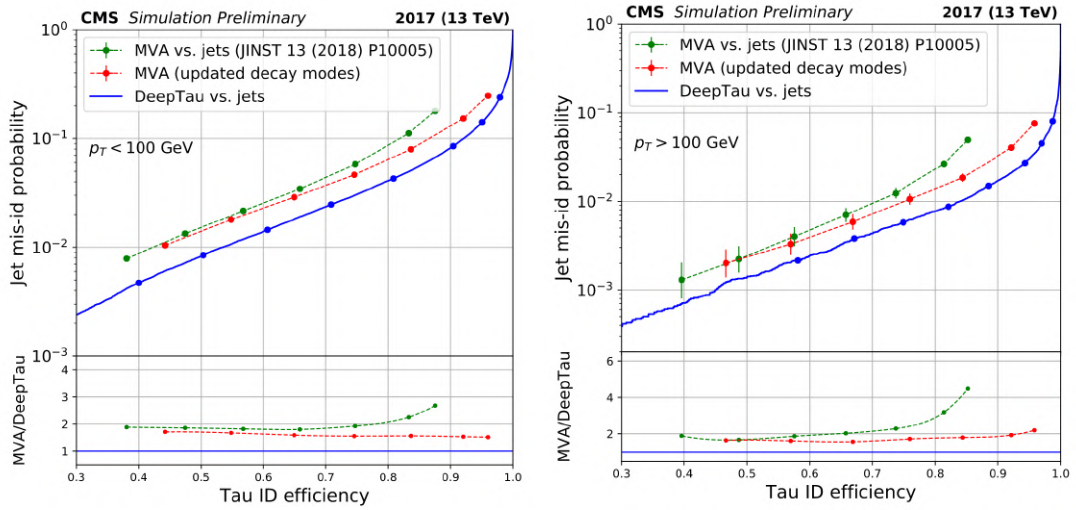
Three different discriminators are defined within the *DeepTau* algorithm:

- *DeepTauVSjets* - Working Points: *VVVLoose*, *VVLoose*, *VLoose*, *Loose*, *Medium*, *Tight*, *VTight* and *VVTight*
- *DeepTauVSelectrons* - Working Points: *VVVLoose*, *VVLoose*, *VLoose*, *Loose*, *Medium*, *Tight*, *VTight* and *VVTight*
- *DeepTauVSmuons* - Working Points: *VLoose*, *Loose*, *Medium* and *Tight*

The identification of τ_h candidates in this analysis is performed using the *Medium* working point of the *DeepTauVSjets* discriminator. For a complete list of the selections and working points applied, see Section 5.4.1. In addition, as discussed in Section 5.5,

the *VVLoose* working point is used to define the QCD-enriched sideband regions with a relaxed tau isolation.

The performance of the *DeepTau* discriminator compared to other available discriminators is reported in Fig. 5.10.



(a) $p_T < 100$ GeV (b) $p_T > 100$ GeV

Figure 5.10: Comparison of different algorithms jet to τ misidentification probability as function of τ identification efficiency for tau leptons with $p_T < 100$ GeV (left) and $p_T > 100$ GeV (right) [55].

The scale factors related to the hadronically decaying taus for the DeepTau algorithm are applied following the recommendations of the TauPOG:

- For genuine taus, the scale factors are provided binned as function of the transverse momentum of the tau in the range from 20 GeV to 1 TeV.
- For genuine electrons misidentified as taus, the scale factors are provided split into two eta regions, barrel and endcap.
- For genuine muons misidentified as taus, the scale factors are provided in η bins.

5.3.5 Jets

Jets are reconstructed using the "anti-k_t" algorithm [56] with a distance parameter $R = 0.4$ (AK4 jets) and $R = 0.8$ (AK8 jets).

Since tracking information is only available in the central region of the CMS detector and the b-tagging process heavily relies on tracking information, all b-jet candidates are required to have $|\eta| < 2.4$ and $p_T > 20$. A more detailed description of jets coming from b quark, identified as b-jets, follows in the next section.

In order to include VBF jet candidates, the $|\eta|$ selection must be loosened from 2.4 to 4.7. In addition, VBF jet candidates are required to have $p_T > 30$ GeV.

All jets are required to pass the tight working point of the particle-flow jet identification, while jets with $p_T < 50$ GeV are also required to pass the loose working point of the Pileup jet discriminator. Both these algorithms are developed by the JetMET POG.

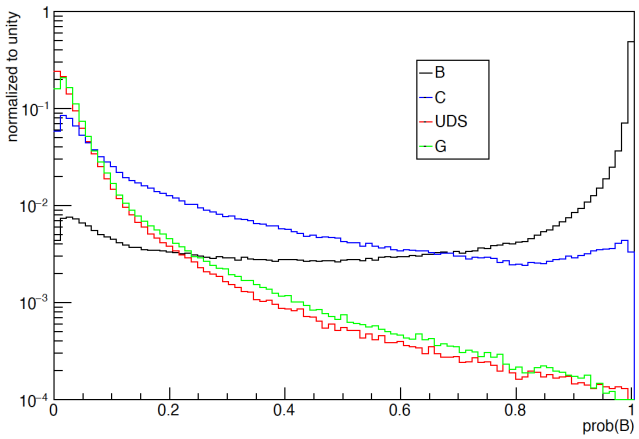


Figure 5.11: Distribution of the DeepFlavour *bb+b+lepb* discriminator value for jets of different flavours in $t\bar{t}$ events.

The usage of AK8 jets allows a more efficient reconstruction of the decays when the two b quarks are close to each other (*boosted regime*). In this case, a jet substructure technique is used to identify the hadronization products of the two b quarks inside the reconstructed jet and the invariant mass of AK8 jets is estimated using the Soft Drop algorithm [57]. In addition for AK8 jets the "Pileup Per Particle Identification" (PUPPI) method is required to take into account the pileup mitigation [58].

5.3.6 Identification of b-jets

Jets originating from b quarks are identified using the DeepFlavour algorithm [59]. In order to separate b-jets from other jets, the DeepFlavour algorithm combines secondary vertex properties, track-based variables and Particle Flow jet constituents (neutral and charged candidates) in a deep neural network.

The strategy of the DeepFlavour algorithm is a multi-classification of jets into six different categories. Three of these categories are summed together to define a single discriminator used to tag b-jets in physics analyses (as shown in Figure 5.11):

- *bb* - two ore more B hadrons
- *b* - exactly one B hadron
- *lepb* - exactly one B hadron with leptonic decay

Thresholds on the DeepFlavour discriminator value corresponding to *loose*, *medium* and *tight* working points are provided by the b Tag and Vertexing POG and are described in Table 5.5.

To account for discrepancies in the b tag performance in MC the whole b-tagging discriminant distribution in MC is corrected to match that in data.

For each MC event with a given jet configuration, the weight ω is computed as:

$$\omega = \prod_{i}^{N_{jets}} SF\left(D, p_T, \eta\right)$$

Year	Working Point	Discriminator score
2016	Loose	0.0614
	Medium	0.3093
	Tight	0.7221
2017	Loose	0.0521
	Medium	0.3033
	Tight	0.7489
2018	Loose	0.0494
	Medium	0.2770
	Tight	0.7264

Table 5.5: DeepFlavour scores defining the b-tagging working points.

where the scale factors SF are provided by the BTV POG as a funtion of the discriminator score, transverse momentum and pseudorapidity of the i-th jet.

The event weights extracted with this method are supposed to change the shape of the b-tagging discriminant. Before applying any b-tag selection criteria expected event yields should be preserved, this means that the number of events (i.e. the sum of event weights) before and after applying b-tag weights should be identical. In order to ensure this the sum of event weights before and after applying b-tag event weights, without requiring any b-tag selection, is computed. The ratio $r = \sum \omega_{\text{before}} / \sum \omega_{\text{after}}$ represents a phase space extrapolation and is multiplied to the b-tag event weight. After selecting events to have a certain b-tag multiplicity, expected event yields should and are allowed to change (since b-tag event weights and b-tagging discriminant values are correlated).

5.3.7 Missing transverse momentum

For each event, the missing transverse momentum $p_{T,miss}$ is reconstructed with the particle-flow MET algorithm [60].

For this analysis, we applied filters, as recommended by the JetMET POG, to ensure a good quality of the reconstructed MET. Events where primary hard interaction vertex is not of good quality, are rejected by `Flag_goodVertices` filter. Beam halo filter `Flag_globalSuperTightHalo2016Filter`, is used to reduce the non negligible probability of the high energy halo muons to interact in the calorimeters creating cluster of up to several hundreds of GeV. To reject events with high HCAL or ECAL noise, four filters are applied: `Flag_HBHENoiseFilter`, `Flag_HBHENoiseIsoFilter`, `ecalBadCalibFilterUpdate` and `eeBadScFilter` . Events with problematic dead cell trigger primitive energy recovery are removed by the `Flag_EcalDeadCellTriggerPrimitiveFilter`. Events where a large unphysical $p_{T,miss}$ is erroneously reconstructed due to the presence of extra muons, are rejected by `Flag_BadPFMuonFilter`.

5.3.8 Additional corrections

L1 ECAL prefiring weight

In 2016 and 2017, the gradual timing shift of ECAL signals was not properly propagated to L1 trigger primitives (TP) resulting in a significant fraction of high η TP being mistakenly associated to the previous bunch crossing. Since Level 1 trigger rules forbid two

consecutive bunch crossings to fire, an unpleasant consequence of this (in addition to not finding the TP in the bunch-crossing 0) is that events can self veto if a significant amount of ECAL energy is found in the region $2 < \eta < 3$.

Since this effect is not described by the simulation it is necessary to compute the probability for an event not to prefire, and apply it as a weight to the simulated events.

Trigger prescale weight

If the instantaneous luminosity at the beginning of an LHC fill exceed a certain threshold some trigger paths may be "prescaled": in order to reduce the rate of accepted events only one out of the n events firing the trigger are streamed out. In order to account for this reduced rate the events that fired only prescaled paths are corrected with the weights reported in table 5.6. The VBF trigger was active only in a fraction of the 2017 data taking runs hence a correction factor must be applied also to the events that fired only this trigger.

Year	HLT path name	Prescale weight
2016	HLT_IsoMu22	0.795221971
	HLT_IsoMu22_eta2p1	0.923782353
	HLT_IsoTkMu22	0.795221971
	HLT_IsoTkMu22_eta2p1	0.923782353
2017	HLT_IsoMu24	0.91613901
	HLT_VBF_DoubleLooseChargedIsoPFTau20_Trk1_eta2p1_Reg	0.65308574

Table 5.6: Weights applied to events firing prescaled HLT paths in all three years.

Custom SF for the fully hadronic channel in 2017

When all the recommended corrections are applied a satisfactory agreement between data and Monte Carlo simulation can be observed in the $\mu\tau_h$ and $e\tau_h$ channels for 2017. Instead, in the $\tau_h\tau_h$ channel, a disagreement at the level of 20% is still present (see Figure 5.12) in the region where the Drell-Yan background contribution is dominant.

A region enriched with Drell-Yan events is defined by applying the regular $\tau_h\tau_h$ channels selections with an additional cut on the angular separation between the two taus: $\Delta R(\tau_h, \tau_h) < 2$. In order to increase the statistic in this sideband region the request of having at least two b-jet canidates ($p_T > 20\ GeV$ and $|\eta| < 2.4$) is dropped. This region is relatively pure in genuine hadronically decaying taus and is used to derive alternative scale factors. For these events all other recommended corrections are applied.

Four different categories are defined by requiring that both the tau legs have the same decay mode: a correction $\mathrm{SF_{DM}}$ is extrapolated within each of these categories, i.e. for each of the decay modes of the tau leptons considered in this analysis.

The simulated events are split in events with one, two and zero genuine hadronic tau leptons; the global event yield of simulated events in these categories should match the number of data events, through the variation of the multiplicative factors $\mathrm{SF^2_{DM}}$, for events with two real tau leptons, and $\mathrm{SF_{DM}}$ for events with one real tau lepton. As the QCD estimation (Section 5.5.1) is affected by the changes in the event yield of the simulated backgrounds in the sidebands, the measurement is performed through a simultaneous fit in the four ABCD regions.

The values of the scale factors thus obtained are listed in 5.7, and the corresponding uncertainties are obtained directly from the fit described in the previous paragraph. Their effect, once applied, is visible in Figure 5.13.

DM	SF	Error
0	1.078	$-0.036/+0.034$
1	1.112	$-0.023/+0.023$
10	0.984	$-0.067/+0.063$
10	0.759	$-0.259/+0.178$

Table 5.7: Custom scale factors derived for the $\tau_h\tau_h$ channel in 2017.

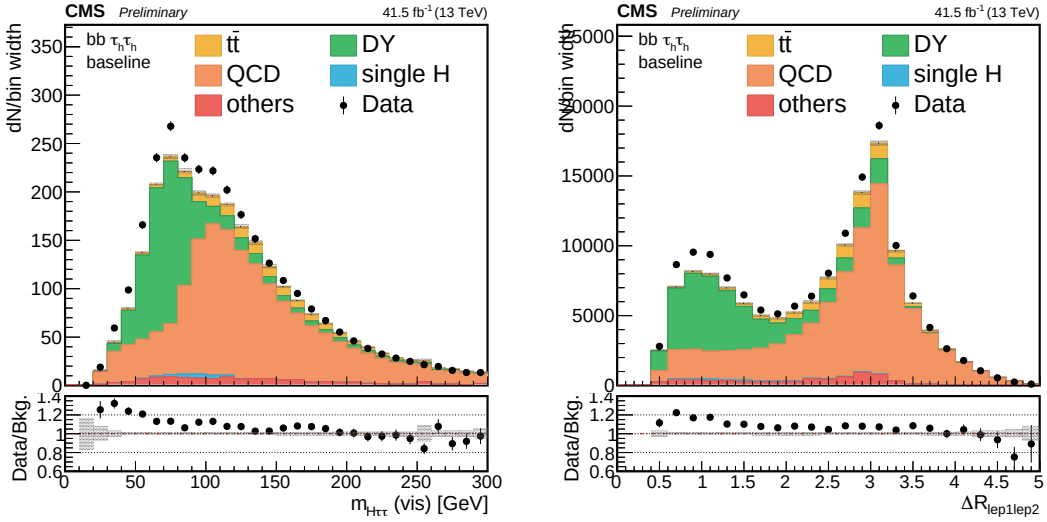


Figure 5.12: Distributions of the invariant mass of the visible decay products of the $\tau\tau$ pair (left) and of the spatial separation ΔR between the two tau candidates (right). Events are shown for the $\tau_h\tau_h$ channel in 2017. A clear disagreement between data and MC simulation is visible in the low mass region where the main background contribution is the Drell-Yan process. The shaded band in the plots only shows the statistical uncertainty.

5.4 Analysis Flow

The analysis flow is articulated in three main steps.

The first step consists in the selection of the candidates used to reconstruct the $\tau\tau$ pair from the decay of one Higgs boson in the event. At this same step, the final state of the $\tau\tau$ system is assessed according to the decay products of the tau leptons (see Section 5.4.1).

The second step is the identification and selection of the VBF jets and the $b\bar{b}$ pair candidate from the decay of the second Higgs boson, thus the categorisation of the events in resolved 1jet-1b, resolved 2b, boosted and VBF categories (see Section 5.4.2). Events in the VBF category are further splitted in five sub-categories by mean of a multiclass deep neural network, as described in Section 5.4.5.

A different deep neural network, described in Section 5.4.6, is used to discriminate signal from background events in each category.

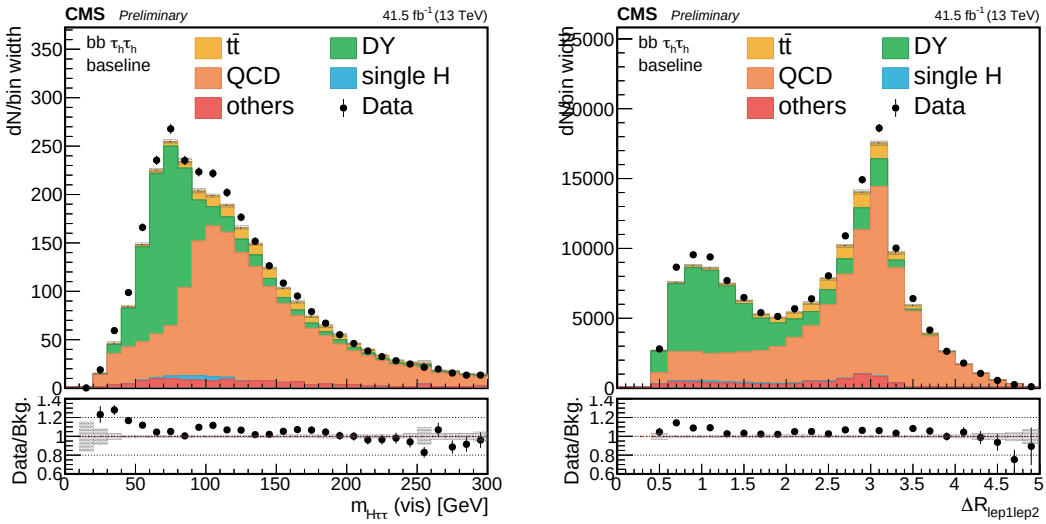


Figure 5.13: Distributions of the invariant mass of the visible decay products of the $\tau\tau$ pair (left) and of the spatial separation ΔR between the two tau candidates (right). Events are shown for the $\tau_h\tau_h$ channel in 2017. The shaded band in the plots only shows the statistical uncertainty. The custom scale factors are applied on top of all other recommended corrections: the disagreement between data and MC simulation is drastically reduced and only remains in a very low mass region which will be anyway excluded from the analysis by the invariant mass selections defined in Section 5.4.6.

5.4.1 $\tau\tau$ candidate selection

This step of the analysis is aimed at identifying the visible decay products of one of the 125 GeV Higgs boson decaying into a τ pair.

Selected signal events are required to have at least one τ candidate that decayed hadronically and has been reconstructed by the HPS algorithm. For these events, a first loop is performed over the offline objects looking for muon and electron candidates passing the baseline selection criteria.

An event is classified as $\mu\tau_h$ if a muon is found, otherwise it is classified as $e\tau_h$ if an electron is found, otherwise it is classified as $\tau_h\tau_h$ if a second hadronic τ is present. In the $\mu\tau_h$ and $e\tau_h$ final states, legs are ordered inside each pair by assigning to the leptonic leg (μ, e) the first position. In the $\tau_h\tau_h$ final states both legs permutations are built and compared as described below.

After the pair type has been assessed, all the pairs of the same type are sorted according to the following algorithm. Pairs are at first sorted according to the isolation of their first leg. If the two first legs have the same isolation, the highest first leg p_T is used to order the pair. If also the p_T is the same (i.e., the pairs share the same first leg) the pair with the most isolated second leg is preferred, and, if there is still ambiguity, priority is given to the pair with the highest second leg p_T .

After all pairs have been sorted, the first pair that passes the baseline selections is chosen.

This strategy has been chosen because it maximizes the purity of the event and it also removes any possible event overlap between the three different final states.

A third lepton veto is applied to all events, discarding events where an additional electron or muon, besides the 2 leptons used to build the $\tau\tau$ pair, is present. A description of the selections applied to identify additional electrons and muons is given in Section 5.4.1.

A subsequent check that the event is firing the trigger path associated to the selected final states and that the selected offline leptons are geometrically matched to the online ones is performed.

Each reconstructed offline lepton is required to pass a p_T threshold depending on the HLT trigger path fired by the event:

$$p_T^{offline} \geq p_T^{HLT} \; + \; threshold$$

where $p_T^{offline}$ is the transverse momentum of the offline selected lepton, p_T^{HLT} is the p_T threshold applied at trigger level and *threshold* is a fixed number depending on the lepton type: $1\,GeV$ for muons and electrons, and $5\,GeV$ for taus. The thresholds are chosen to be conservative with respect to the trigger turn-on curves and accordingly to the recommendations of the Tau POG.

$\mu\tau_h$ channel

Events in the $\mu\tau_h$ channel are selected by requiring:

- A muon of $\left|\eta_\mu\right| < 2.1$ passing tight particle–flow muon identification criteria plus the relative isolation requirement $I^{\mu}_{rel} < 0.15$. The reconstructed muon production vertex must be close to the main primary vertex within a distance $\Delta_{xy} < 0.045\,mm$ and $\Delta_z < 0.2\,mm$.
  The p_T threshold on the muon depends on the HLT trigger path fired:
  - 2016 - $p_T > 23$ (20) GeV if it fired a single-muon (cross lepton) trigger.
  - 2017 - $p_T > 25$ (21) GeV if it fired a single-muon (cross lepton) trigger.
  - 2018 - $p_T > 25$ (21) GeV if it fired a single-muon (cross lepton) trigger.
- A hadronic tau of $\left|\eta_{\tau_h}\right| < 2.3$ (2.1 for events firing only the cross lepton trigger), using `decayModeFindingNewDMs` for the decay mode reconstruction, and passing the Tight DeepTauVSmu and the VLoose DeepTauVSe discriminators. The isolation requirement on the hadronic τ is the Medium working point of the DeepTauVSjet algorithm. The same requirements on the reconstructed vertex Δ_z as in the case of the muon is applied.
  The p_T threshold on the hadronic tau depends on the HLT trigger path fired:
  - 2016 - $p_T > 20$ (25) GeV if it fired a single-muon (cross lepton) trigger.
  - 2017 - $p_T > 20$ (32) GeV if it fired a single-muon (cross lepton) trigger.
  - 2018 - $p_T > 20$ (32) GeV if it fired a single-muon (cross lepton) trigger.
- Muon and hadronic tau are required to have opposite electric charge and a spatial separation $\Delta R(\mu, tau_h) > 0.5$.
- In case multiple combinations of muon plus τ_h exist in an event, the best pair is chosen with the algorithm decribed at the beginning of this Section.
- The event is required to pass any of the $\mu\tau_h$ triggers described in Section 5.2 and the offline leptons to match the HLT ones.

$e\tau_h$ channel

Events in the $e\tau_h$ channel are selected by requiring:

- An electron of $|\eta_e| < 2.1$ passing tight MVA identification criteria (80% efficiency WP, `mvaEleID-Fall17-iso-V2-wp80`). The reconstructed electron production vertex must be close to the main primary vertex within a distance $\Delta_{xy} < 0.045\,mm$ and $\Delta_z < 0.2\,mm$.
  The p_T threshold on the electron depends on the HLT trigger path fired:
  - 2016 - $p_T > 26$ GeV if it fired a single-electron trigger.
  - 2017 - $p_T > 33$ (25) GeV if it fired a single-electron (cross lepton) trigger.
  - 2018 - $p_T > 33$ (25) GeV if it fired a single-electron (cross lepton) trigger.
- A hadronic tau of $|\eta_{\tau_h}| < 2.3$ (2.1 for events firing only the cross lepton trigger), using `decayModeFindingNewDMs` for the decay mode reconstruction, and passing the Tight DeepTauVSmu and the Tight DeepTauVSe discriminators. The isolation requirement on the hadronic τ is the Medium working point of the DeepTauVSjet algorithm. The same requirements on the reconstructed vertex Δ_z as in the case of the electron isolated applied.
  The p_T threshold on the hadronic tau depends on the HLT trigger path fired:
  - 2016 - $p_T > 20$ GeV if it fired a single-electron trigger.
  - 2017 - $p_T > 20$ (35) GeV if it fired a single-electron (cross lepton) trigger.
  - 2018 - $p_T > 20$ (35) GeV if it fired a single-electron (cross lepton) trigger.
- Electron and hadronic tau are required to have opposite electric charge and a spatial separation $\Delta R(e, \tau_h) > 0.5$.
- In case multiple combinations of electron plus τ_h exist in an event, the best pair is chosen with the algorithm decribed at the beginning of this Section.
- The event is required to pass any of the $e\tau_h$ triggers described in Section 5.2 and the offline leptons to match the HLT ones.

$\tau_h\tau_h$ channel

Events in the $\tau_h\tau_h$ channel are selected by requiring:

- Two hadronic τ with $|\eta_{\tau_h}| < 2.1$, using `decayModeFindingNewDMs` for the decay mode reconstruction, and passing the VLoose DeepTauVSmu and the VVLoose DeepTauVSe discriminators. The isolation requirement on the hadronic τ is the Medium working point of the DeepTauVSjet algorithm. The vertex requirement $\Delta_z < 0.2\,mm$ is applied. The p_T threshold on the hadronic taus depends on the HLT trigger path fired:
  - 2016 - $p_T > 40$ GeV if it fired a di-tau trigger.
  - 2017 - $p_T > 40$ (25) GeV if it fired a di-tau (VBF+$H \to \tau_h\tau_h$) trigger.
  - 2018 - $p_T > 40$ (25) GeV if it fired a di-tau (VBF+$H \to \tau_h\tau_h$) trigger.

- The two hadronic τ are required to have opposite electric charge and a spatial separation $\Delta R(\tau_h, \tau_h) > 0.5$.
- The two hadronic τ are ordered by decreasing p_T inside the pair (i.e. $p_T\tau_1 > p_T\tau_2$). In case multiple pair satisfy the previous requirements, the pair with the most isolated τ_1 is preferred. In case the two isolation are equal, the pair with the highest $p_T\tau_1$ is chosen. If also this requirement does not allow to select a pair, the one with the most isolated τ_2 is chosen.
- The event is required to pass any of the $\tau_h\tau_h$ triggers described in Section 5.2 and the offline leptons to match the online ones.

Third lepton veto

In the analysis events are rejected if they contain an electron or a muon, in addition to the signal electron (muon) for the $e\tau_h$ ($\mu\tau_h$) channel, that passes the following selection:

- An electron of $|\eta_e| < 2.5$ and $p_T > 10$ GeV. The electron passes the loose MVA iso-identification criteria (`mvaEleID-Fall17-iso-V2-wp90`) OR the logical AND between the loose MVA non-iso-identification criteria (`mvaEleID-Fall17-noIso-V2-wp90`) and the relative isolation requirement $I^e_{rel} < 0.3$ The reconstructed electron production vertex must be close to the main primary vertex within a distance $\Delta_{xy} < 0.045\,mm$ and $\Delta_z < 0.2\,mm$.
- A muon of $|\eta_\mu| < 2.4$, $p_T > 10$ GeV and passing the tight or medium particle–flow muon identification criteria, plus the relative isolation requirement $I^\mu_{rel} < 0.3$. The reconstructed muon production vertex must be close to the main primary vertex within a distance $\Delta_{xy} < 0.045\,mm$ and $\Delta_z < 0.2\,mm$

5.4.2 b and VBF jets selection

This step of the analysis is aimed at selecting the two jets from the decay in a $b\bar{b}$ pair of the second 125 GeV Higgs boson and the two jets associated to VBF production of the Higgs bosons pair.

The two b-jet candidates must have $p_T > 20\,GeV$ and $|\eta| < 2.4$, in order to be within the tracker acceptance. This is necessary since b-tagging algorithms use tracks and vertices information. The distance between each jet and both selected τ candidates must be $\Delta R > 0.5$. As shown in Figure 5.14, often the second jet by DeepFlavour score does not fulfill the minimal b-tag requirement (medium WP) in the gluon-gluon fusion signal sample.

In order to minimize the jet mistagging probability, a new algorithm (HH-BTag) based on deep neural networks has been developed to select the b-jets [1].

The performance of the different algorithms is compared using the purity of the selection, defined in equation 5.31. Jets are ordered according to the score of each tagger and only the two jets with the highest score are selected: if they match with the generated objects the event is added to the numerator.

$$\text{purity} = \frac{\#\text{events with 2 jets selected correctly}}{\#\text{total number of events}}. \tag{5.31}$$

[1] The algorithm is fully described in the CMS internal document AN-2019/283

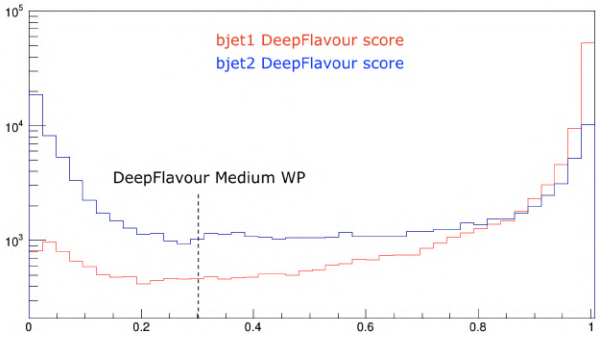
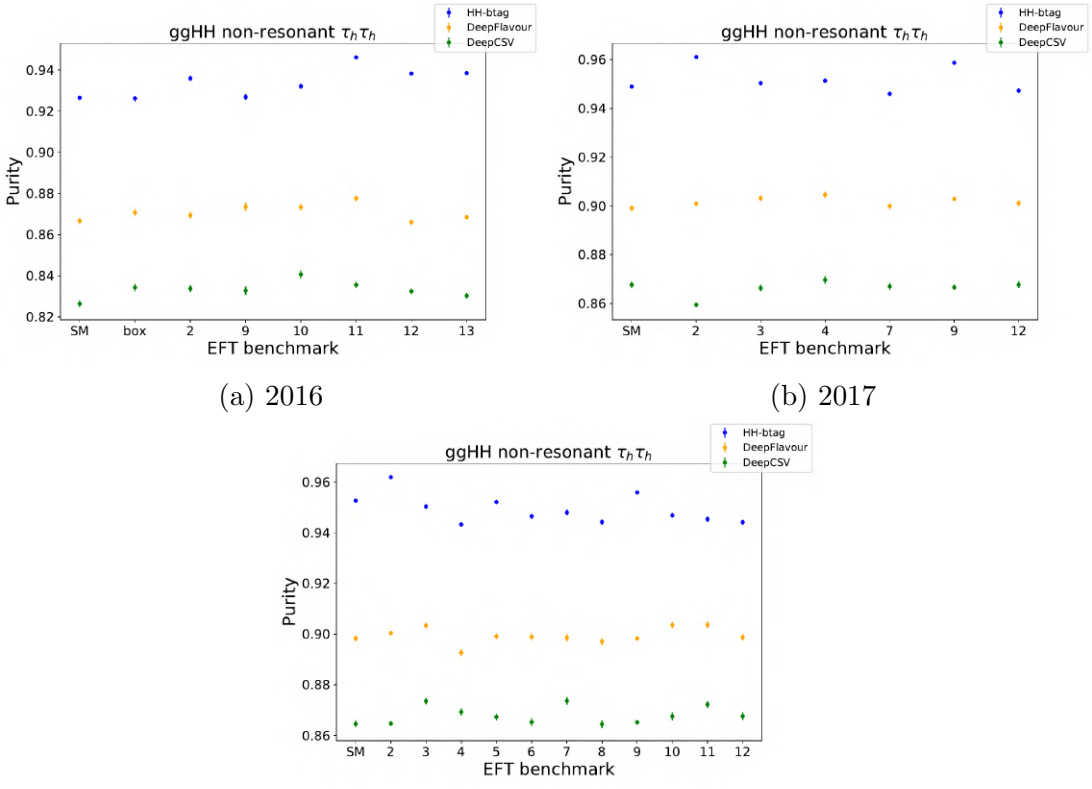


Figure 5.14: Distribution of the DeepFlavour score for the first two jets ordered by DeepFlavour score for a gluon-gluon fusion signal sample produced with the 2017 conditions.

In figures 5.15 and 5.16 the performance of the HH-BTag NN is compared to one of the other available discriminators (DeepFlavour and DeepCSV) considering non-resonant di-Higgs production via GF and VBF, in the $\tau_h\tau_h$ channel, for each year. The use of HH-BTag increases the purity by 5% with respect to DeepFlavour, reaching values of 98%.



(a) 2016

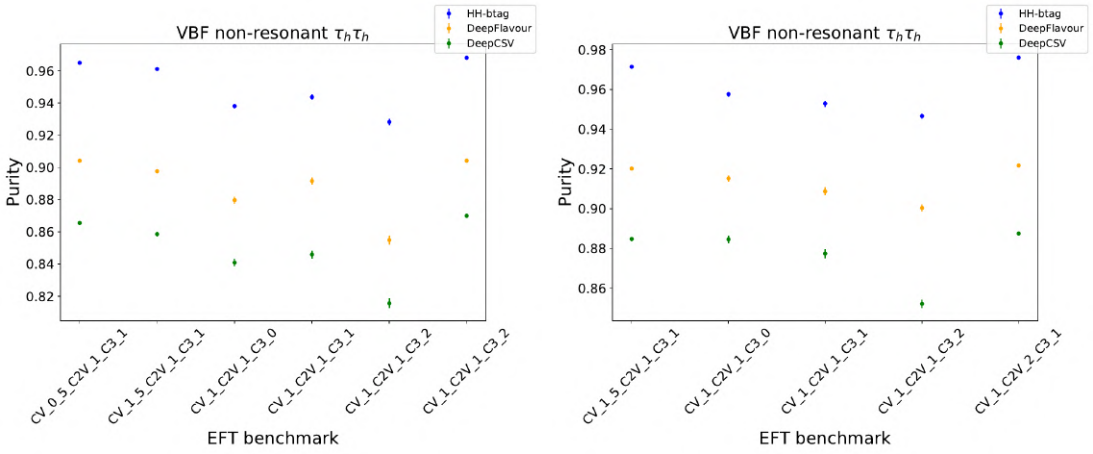
(b) 2017

(c) 2018

Figure 5.15: Purity distributions for the GF di-Higgs production in the $\tau_h\tau_h$ channel, using the HH-Btag, DeepFlavour and DeepCSV discriminators

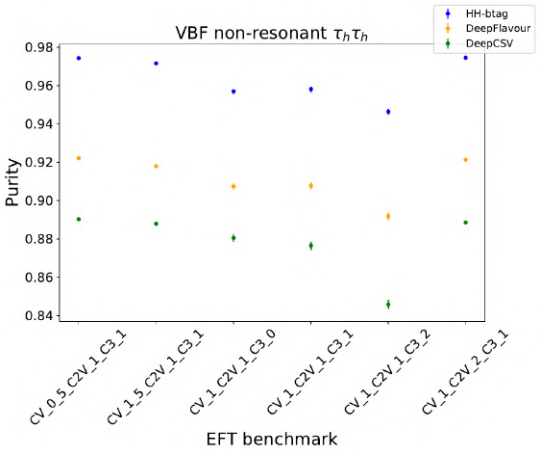
Every b-jet candidate is assigned a score by the HH-BTag algorithm: the two with the highest scores are taken to be the two b-jets originating from the decay of the Higgs boson.

Events are then divided into resolved or boosted categories in order to increase the analysis sensitivity for high m_H values by profiting of the different regime for $b\bar{b}$ pair pro-



(a) 2016

(b) 2017



(c) 2018

Figure 5.16: Purity distributions for the VBF di-Higgs production in the $\tau_h\tau_h$ channel, using the HH-Btag, DeepFlavour and DeepCSV discriminators

duction. Three different regimes for $b\bar{b}$ pair production, depending on the b-jet separation, are taken into account:

- $\Delta R(b,\bar{b}) > 0.8$: jets are reconstructed as separated object with AK4 (resolved jet);
- $0.4 < \Delta R(b,\bar{b}) < 0.8$: jet are reconstructed both as separated as a merged jet (fatjet) with AK8;
- $\Delta R(b,\bar{b}) < 0.4$: jet are merged and reconstructed only as a fatjet.

Since the last scenario ($\Delta R\,(b,\bar{b}) < 0.4$) is never reached for non-resonant signals, which are the types of signals studied in this analysis, events are separated in just 2 categories: the *boosted* and *resolved* categories.

Events where a fatjet with $m_{fatjet} > 30\,GeV$ exists, it is composed by at least two subjets, and the distance between the two subjets and the previously selected AK4 jets is $\Delta R < 0.4$ fall in the boosted category, while all other events are assigned to the resolved category. As an additional requirement for events falling in the boosted category, both jets are required to pass the b-tag *Loose* working point, otherwise the event is discarded.

Based on the b-tagging, a further categorisation is applied to events in the resolved category:

- **resolved 2jet–1tag (res1b)** one of the jet passes the *Medium* working point, the other one does not.
- **resolved 2jet–2tag (res2b)** both jets pass the *Medium* working point. This is the most signal-sensitive category.

A second scan of the jet collection is subsequently performed excluding the selected b-jets and the jets with $p_T < 30\,GeV$ and $|\eta| > 4.7$. The invariant mass of every pair of jets passing this selection is computed and the pair with the highest value is selected as the VBF-jets candidate pair. Only pairs having $M_{jj} > 500\,GeV$ and $\Delta\eta_{jj} > 3$ are considered in the analysis. Those values have been chosen in order to find a compromise between background events rejection and signal events preservation. The effect of these selections can be seen in Fig. 5.17 where the jet pair invariant mass versus angular separation is reported for two MC samples: VBF di-Higgs SM production and $t\bar{t}$. The selections of the VBF category are completed by the requirement of having at least one b-jet candidate passing the *Medium* working point of DeepFlavour.

5.4.3 HH invariant mass selection

The goal of the previous steps was to determine whether in each event were present suitable $H \to b\bar{b}$ and $H \to \tau^+\tau^-$ candidates and, in this case, to identify the most likely candidate Higgs bosons. Nevertheless, the events selected after these steps are still expected to be background dominated.

The selection described below profits from the kinematics of the $HH \to b\bar{b}\tau^+\tau^-$ decay to reduce the background contributions and only select events that are compatible with the HH decay hypothesis.

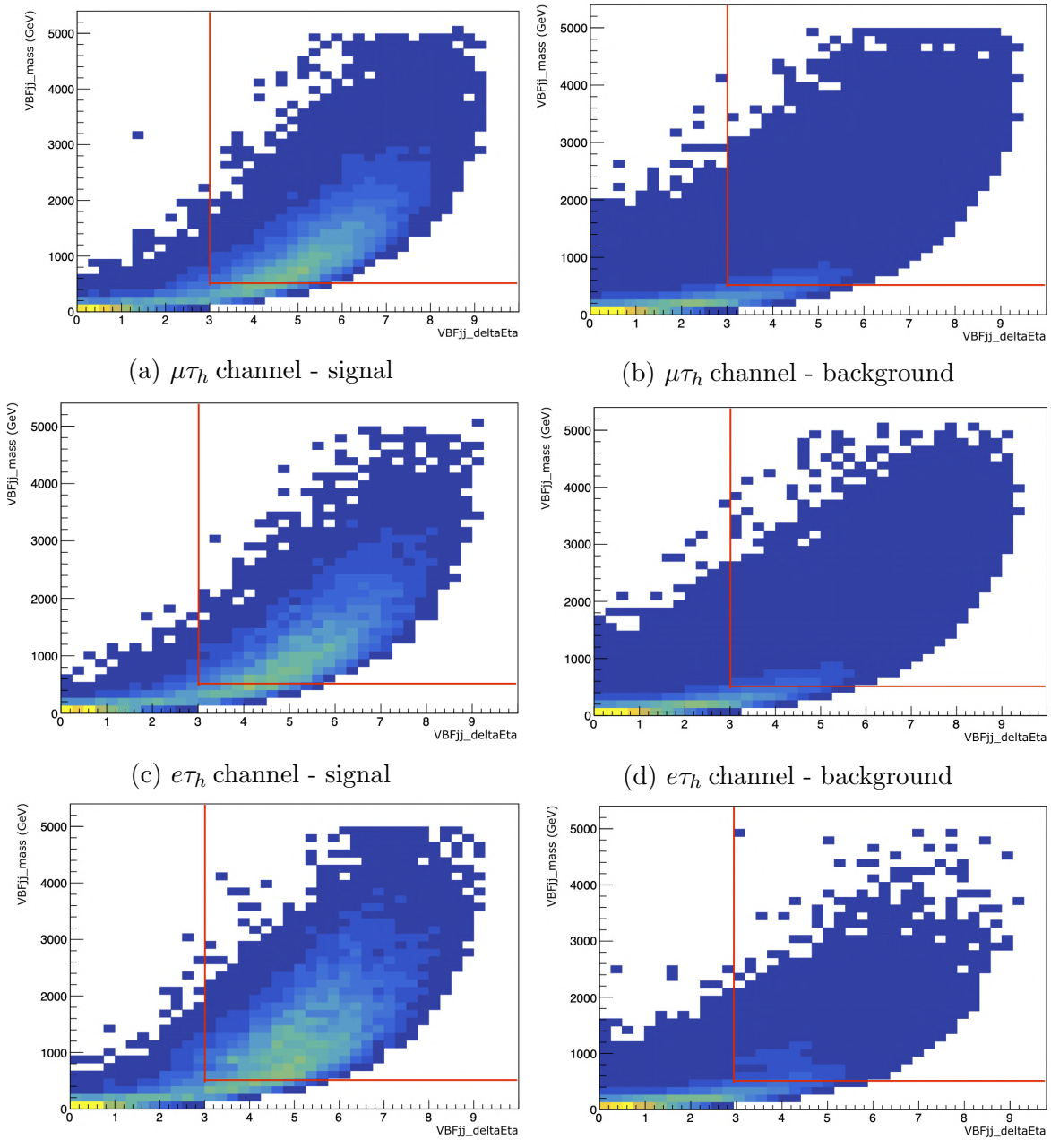
The invariant mass of the $\tau\tau$ pair is reconstructed using the SVfit algorithm [61] that is based on a likelihood function which quantifies the level of compatibility between a Higgs mass hypothesis and the measured momenta of the visible tau decay products plus the missing transverse energy reconstructed in the event. The SVfit algorithm improves the resolution on the invariant mass of the $\tau\tau$ pair compared with the visible mass, therefore allowing a better signal to background discrimination.

To define the signal region in the resolved categories of the analysis an elliptical cut is imposed on the SVfit mass ($m_{\tau\tau}$) and on the reconstructed invariant mass of the bb pair (m_{bb}).

An "elliptical" selection around the SM Higgs boson reconstructed masses, already used in [62], leads to a region with higher S/B ratio, compared to the previous analyses [63], where a "squared" selection defined as $80GeV < m_{\tau\tau}^{SVfit}, m_{bb} < 160GeV$ was used.

The mass cuts were defined using a random search for the offsets and resolutions, of both the $m_{\tau\tau}$ and the m_{bb} distributions, that minimized the acceptance of weighted background events for an acceptance of weighted signal events above 90%:

- Randomly sample the 4 parameters (2 offsets and 2 resolutions).
- Compute the weighted acceptances.
- If signal acceptance is lower than 90% reject the parameters.
- If background acceptance is lower than the current lowest background acceptance: select the new parameters.



(a) $\mu\tau_h$ channel - signal

(b) $\mu\tau_h$ channel - background

(c) $e\tau_h$ channel - signal

(d) $e\tau_h$ channel - background

(e) $\tau_h\tau_h$ channel - signal

(f) $\tau_h\tau_h$ channel - background

Figure 5.17: 2D distributions of $(VBFjj_{\Delta\eta}, VBFjj_{mass})$ for the VBF SM signal (a, c, e) and for the $t\bar{t}$ background (b, d, f), for the 2018 MC simulation. Only events belonging to the VBF category described in the text are considered.

This procedure is repeated more than 20000 times and the final HH invariant mass selection used in the 1b and 2b resolved categories is:

$$\frac{(\mathrm{m}_{\tau\tau} - 129\,GeV)^2}{(53\,GeV)^2} + \frac{(\mathrm{m}_{bb} - 169\,GeV)^2}{(145\,GeV)^2} < 1 \tag{5.32}$$

while in the boosted category, because of the different kinematics, the selection is:

$$\frac{(\mathrm{m}_{\tau\tau} - 128\,GeV)^2}{(60\,GeV)^2} + \frac{(\mathrm{m}_{bb} - 159\,GeV)^2}{(94\,GeV)^2} < 1 \tag{5.33}$$

It is important to mention that the mass cut is not applied with the principle aim of improving the S/B ratio, but rather to remove significantly outlying background events in regions where no signal overlap is expected. The analysis is designed to leave all discrimination up to the DNNs, hence the high target for the signal efficiency, with only a reasonable attempt at reducing the background acceptance. Studies of the limit were performed as a function of signal efficiency of the mass cut, in which the DNNs were retrained each time. The results showed that a more discriminant mass-cut resulted in a poorer limit, i.e. the DNNs are able to separate signal from background in a better way than the basic mass cut can.

As an example, the 2D $(m_{bb}, m_{\tau\tau}^{\mathrm{SVfit}})$ signal (Gluon Fusion SM di-Higgs production) and background event distributions are shown in Figure 5.18. In the plots the combinations of all $\tau\tau$ decay channels for the different years are shown requiring that both b-jet candidates pass the Medium WP, but before the invariant mass cuts are applied. The region defined by the elliptical selection is also shown superimposed on the plot.

5.4.4 Event categorization and signal extraction

As described in Section 5.4.2, three categories are defined to better identify the Gluon Gluon Fusion events: **res1b**, **res2b** and **boosted**.

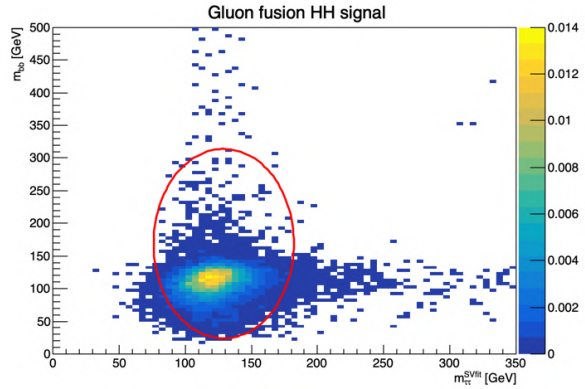
A fourth category, **VBF**, is defined to isolate signal events produced through the Vector Boson Fusion mechanism. Since a substantial fraction of GGF events may be included in the VBF category, a multi-classification approach represents the most suitable option to discriminate genuine VBF events. The multi-classification Deep Neural Network is described in Section 5.4.5.

The Deep Neural Network developed to separate signal from background events in each category is described in Section 5.4.6.

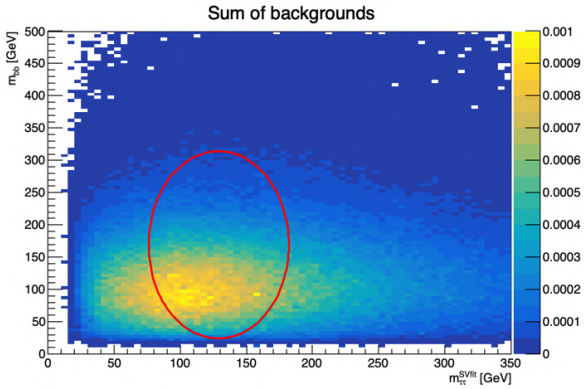
Fig. 5.19 describes schematically the way the events are split into the different categories.

5.4.5 Multi-class categorization for VBF category

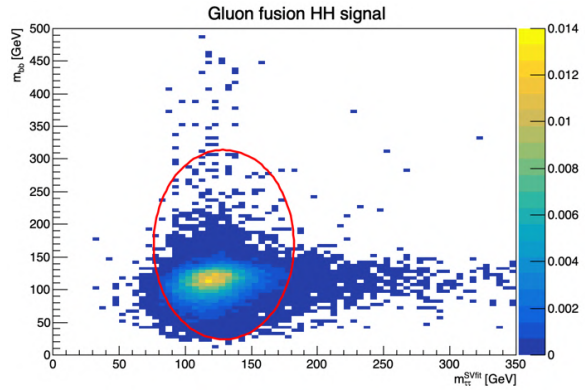
In the multi-class categorization strategy, machine learning techniques are used to assign probability estimates for an event to belong to categories associated to any of the relevant physics processes under consideration. The following section first explains the overall categorization strategy and the multi-class classification approach. Thereafter, the network architecture is described alongside a listing of input feature distributions. The section closes with results of the process-based categorization scheme in terms of ROC curves of the trained networks to assess potential overtraining, DNN output discriminant distributions, and confusion matrices of the classification process.



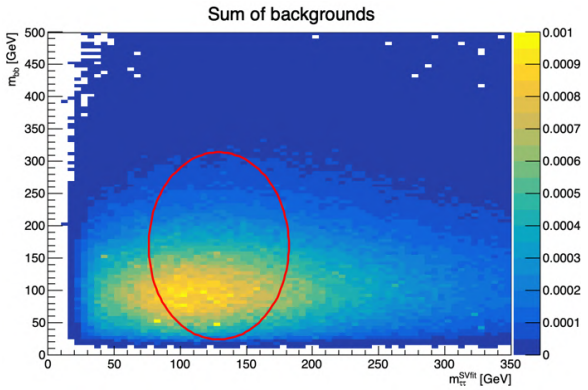
(a) $\mu\tau_h$ channel - signal



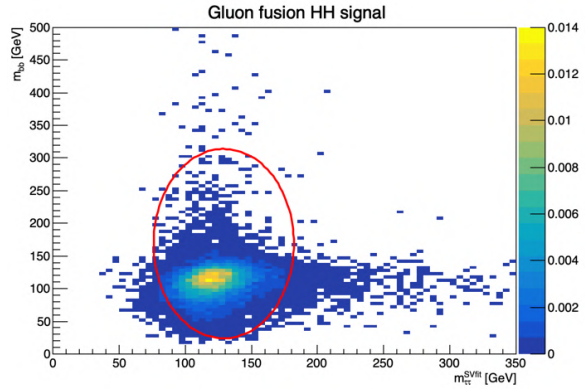
(b) $\mu\tau_h$ channel - background



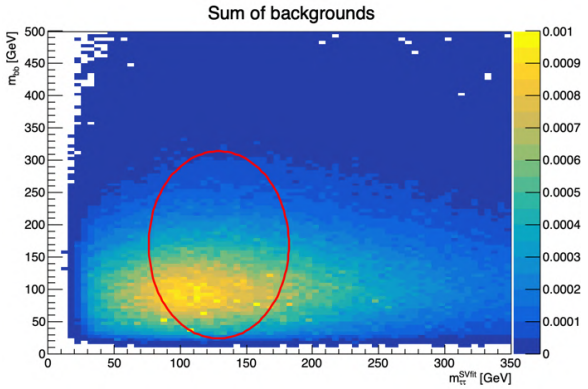
(c) $e\tau_h$ channel - signal



(d) $e\tau_h$ channel - background



(e) $\tau_h\tau_h$ channel - signal



(f) $\tau_h\tau_h$ channel - background

Figure 5.18: 2D distributions of $(m_{bb}, m_{\tau\tau}^{\text{SVfit}})$ for the Gluon Fusion SM signal (a, c, e) and for the sum of MC backgrounds (b, d, f). Plots show the normalized process yields after the $\tau\tau$ and bb candidates selections and before the invariant mass requirements for 2016 (a, b), 2017 (c, d), and 2018 (e, f). The presence of two b-tagged jets passing the medium WP is required; the three $\tau\tau$ decay channles considered in the analysis are merged together. The red line shows the region selected by the resolved elliptical cut.

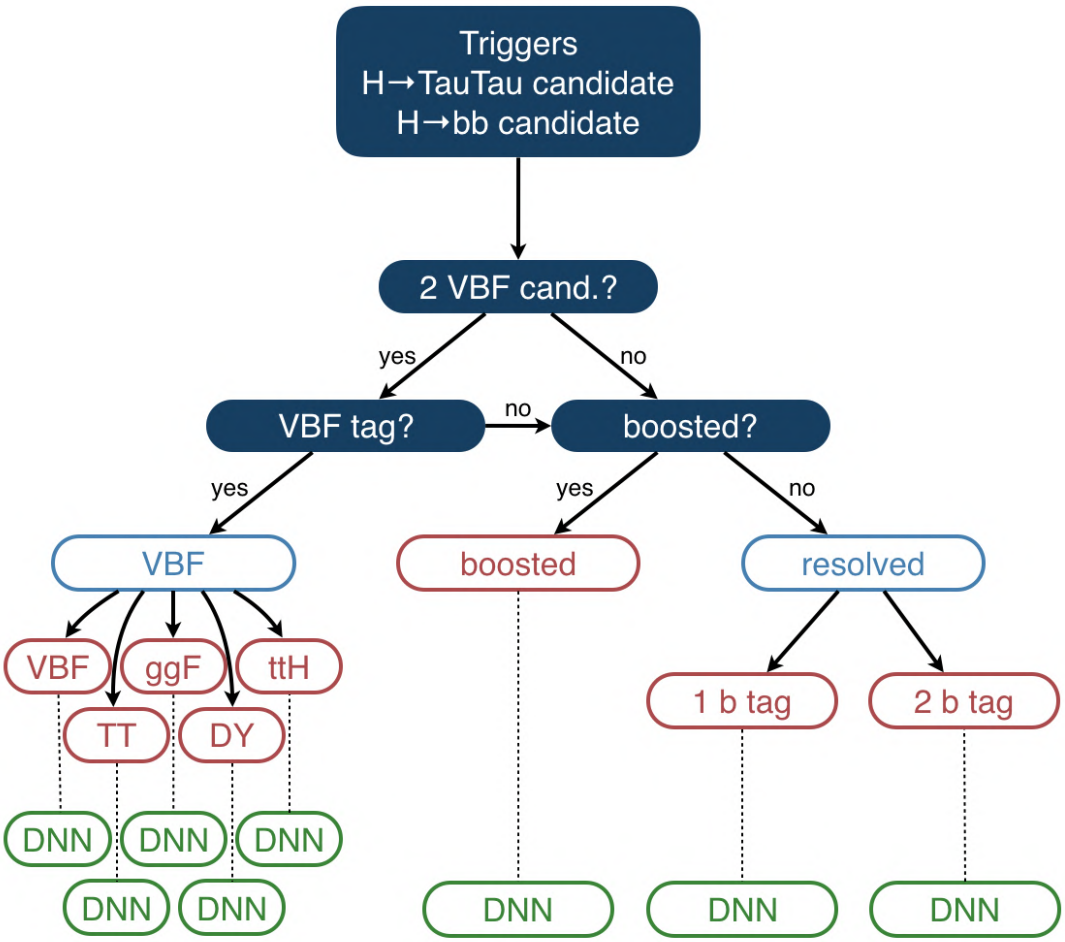


Figure 5.19: Description scheme of event categorization and discriminating variables used in each category. Events in the VBF category are split accordingly to the "most probable process" predicted by the multiclass neural network. In the green boxes, "DNN" represents the Gluon Fusion optimized DNN score which is used for signal extraction both in the GGF and in the VBF categories.

Categorization Strategy

In the context of the small amount of expected signal events compared to all backgrounds, signal extraction processes usually entail a leverage effect, which implies a crucial dependence of signal strength modifiers (μ) on the systematic uncertainties of the statistical fit model. First, any residual modeling discrepancy between data and simulation is compensated during the fit primarily by adjusting central nuisance parameter values. Second, the determination of each nuisance parameter is intrinsically subject to a measurement uncertainty. For that reason, large posterior uncertainties of nuisance parameters cause undesired, high variances on the measurement of μ. The sensitivity of the analysis is therefore directly connected to the ability to measure and constrain particular nuisance parameters.

This can be achieved by creating categories that are solely enriched with one of the relevant physics processes. By doing so, interference between involved nuisance parameters is mitigated, rendering their measurement in the fitting procedure less perturbed such that stronger model constraints can be deduced, thus leading to an enhanced analysis sensitivity.

Unlike binary classification that distinguishes between two classes of events, i.e., usually *one* signal process and *all* background processes combined, multi-class classification has the advantage that different physics processes can be treated with equal importance. This is especially desirable in scenarios where multiple signal processes with different topologies are to be examined, or where several background processes, affected by large systematic uncertainties, contribute significantly to the overall expected background.

The search for HH production in the $b\bar{b}\tau^+\tau^-$ final state is attributed both of these characteristics. On the one hand, as explained above, the distinction between ggF- and VBF-induced HH production could give rise to the $VVHH$ coupling strength that can only be probed by the latter production mode. On the other hand, the analysis is dominated by backgrounds originating from QCD multijet and Drell-Yan processes, as well as processes involving top quark pairs ($t\bar{t}$ and $t\bar{t}H$). While the two former impact the analysis greatly due to their vast rate, the two latter processes exhibit similarities in their final state with respect to the expected signal signatures and are therefore to be included in the list of background processes to distinguish. Moreover, contributions from $t\bar{t}$ are subdivided further according to the decays of the two W bosons, namely dilepton (DL), semi-lepton (SL), and full-hadron decays (FH), and contributions from $t\bar{t}H$ subdivided into events with either $H \to b\bar{b}$ or $H \to \tau\bar{\tau}$ decays, which both have partial similarities to the signal processes. This results in a total of six significant background processes.

The multi-class classification strategy can be described as follows with a supporting illustration shown in Fig 5.20. Per event, a predefined set of input variables is fed into a network whose architecture and weights are subject to a training and optimization process. The network evaluates the variables and outputs per event a vector of nine floating point numbers with a sum of one, where each number describes a probability estimate for the event to originate from a particular process under investigation. These processes are defined based on the six significant background processes listed above, the HH (ggF) signal process, the HH (VBF) signal process with $C2V = 1$, and the same HH (VBF) signal process with $C2V = \{0, 2\}$.

The decision to split the HH (VBF) process into a SM- and a BSM-like component is based on the expected variation of the event topology for different values of $C2V$. Therefore, only passing information about the SM configuration could lead to an artificial

bias of the neural network performance towards the SM-like realization of the HH (VBF) process, impeding its ability to exclude certain BSM scenarios. Contributions from QCD are determined by means of a data-driven approach and are therefore not attributed an output value by the DNN.

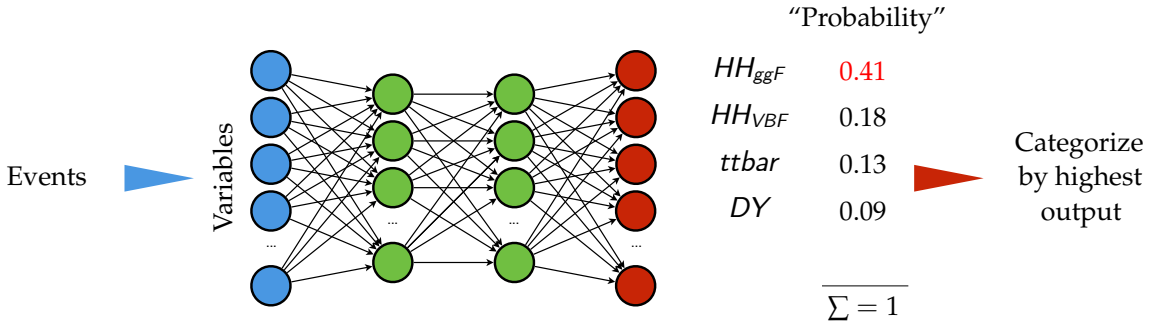


Figure 5.20: Concept of the event categorization approach using DNNs that perform a multi-class classification. Each event is attributed even values that express its probability to originate from either the HH (ggF) or the HH (VBF) signal, or one of the six relevant background processes to consider. An event is categorized by the process class that received the highest probability (HH (VBF) in the depicted example).

The subsequent categorization is performed in two steps. First, the output values of the three $t\bar{t}$, the two $t\bar{t}H$, and the two HH (VBF) subprocesses, respectively, are added such that a total of five distinct outputs remains: HH (VBF), HH (ggF), $t\bar{t}H$, $t\bar{t}$, and Drell-Yan. While this *merging* of output values is found to have only a negligible impact on the signal extraction results, it allows to pass more useful information to the neural network during the training process, possibly improving its separation performance. The eventual categorization into one of the remaining five *classes* is then derived according to the class that received the highest output value. This approach unambiguously assigns each event to a category that is intended to be enriched with events of similar topology corresponding the *most probable process*.

Following this categorization scheme, the choice of the discriminant to be used in the signal extraction procedure remains arbitrary (c.f. Figure 5.19). Here, we employ the DNN output distribution as described Sec. 5.4.6 due to the similarities between the ggF signal process and the VBF signal process with non-SM $C2V$ values (especially $C2V = \{0, 2\}$) (c.f. Fig. 5.22, top right plot).

Input Features

Table 5.8 lists 74 kinematic observables that are considered as input variables to the neural networks.

Training Procedure

In order to exploit the full amount of the simulated event samples, without the risk of introducing a bias in the DNN output due to potential overtraining, a two-fold cross validation approach is performed. First, the available amount of simulated events is divided into two statistically independent datasets, using the event number as a splitting criterion (*even* or *odd*). On each of these two sets, a dedicated training is performed using

Variable name(s)	Description
is_201{6,7,8}	Flag denoting the campaign / year of input events.
is_{etau,mutau,tautau}	Flag denoting reconstructed lepton channel.
lep{1,2}_{e,pt,eta,phi}	Four-vector components of the two leptons.
bjet{1,2}_{e,pt,eta,phi}	Four-vector components of the two b-tagged jets.
bjet{1,2}_{deepflavor,hhbtag}	b-tag and HH-b-tag of the two b-tagged jets.
vbfjet{1,2}_{e,pt,eta,phi}	Four-vector components of the two VBF jets.
vbfjet{1,2}_{deepflavor,hhbtag}	b-tag and HH-b-tag of the two VBF jets.
ctjet{1,2,3}_{e,pt,eta,phi}	Four-vector components of the three additional central jets.
ctjet{1,2,3}_{deepflavor,hhbtag}	b-tag and HH-b-tag of the three additional central jets.
fwjet{1,2}_{e,pt,eta,phi}	Four-vector components of the two additional forward jets.
met_{pt,phi}	Missing transverse energy and azimuthal direction.
bh_{e,pt,eta,phi}	Four-vector components of the reconstructed $H_{b\bar{b}}$ candidate.
tauh_{e,pt,eta,phi}	Four-vector components of the reconstructed $H_{\tau\bar{\tau}}$ candidate.

Table 5.8: Names and descriptions of variables considered as inputs to the neural networks. In case an event exhibits less VBF jets, additional central jets or forward jets than described, the network receives 0 values instead.

35% of the events as the actual training data and the remaining 15% for immediate cross validation. The latter is required since both the network architecture and the training process are subject to a variety of hyper-parameters whose optimal values are selected through a grid-based scanning procedure. In practical terms, this procedure constitutes another optimization process, which itself demands a dedicated, independent dataset. In general, a training proceeds for up to 15000 forward-pass and backpropagation steps. However, it is terminated in case the overall classification accuracy is found to diverge between training and validation events, or when no further improvement is detected within the last 1000 steps. The training with the highest overall classification accuracy is retained for further analysis.

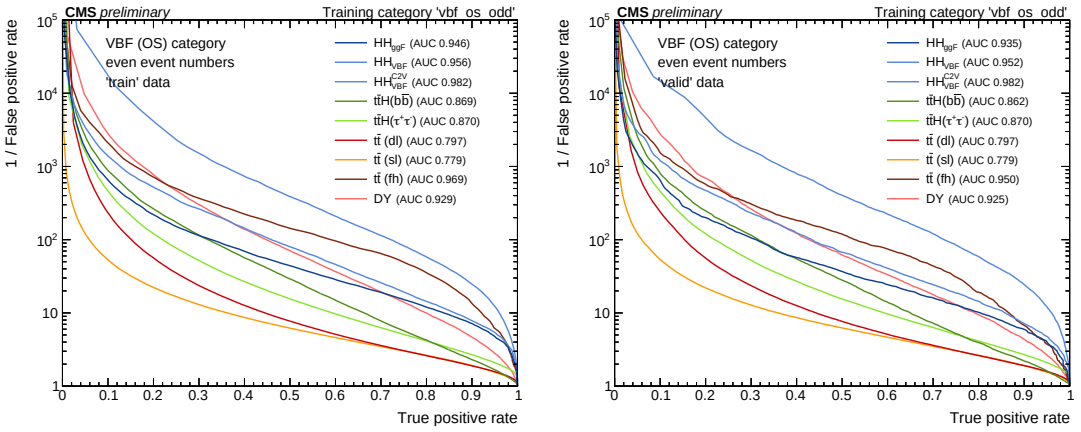
This procedure is performed independently for events with either *even* or *odd* event numbers, resulting in two potentially different network architectures. As described above, information on the data-taking period, i.e., 2016, 2017 and 2018, is incorporated through a categorical input feature to allow the network to infer differences related to detector conditions and tuning parameters in event simulation code. To increase the statistical robustness, ensembles of ten networks with different random seeds are trained and merged with equal weight after each particular training process concluded. For the inference of DNN output distributions, each event is evaluated by the network that corresponds to its opposite *even* or *odd* setting to ensure statistical independence. [2]

Training and Categorization Results

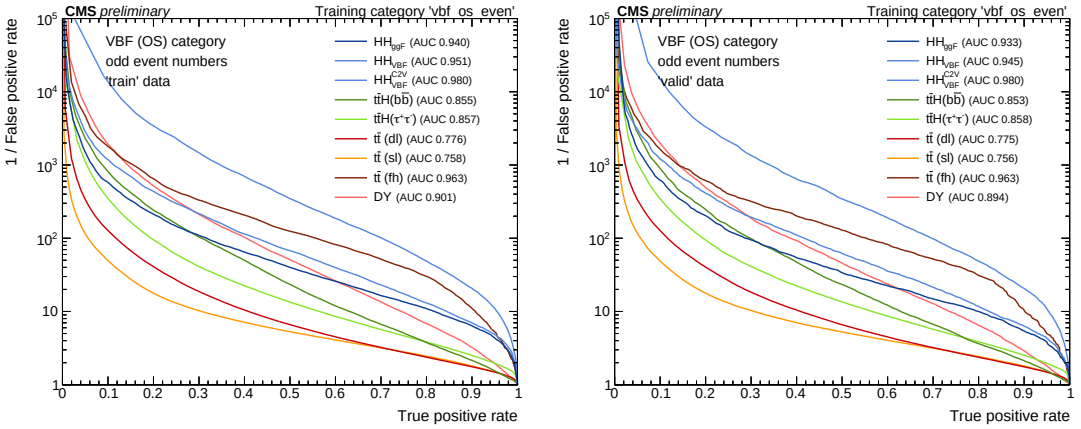
ROC curves of the trained networks with even and odd event numbers are shown in Figure 5.21 separately for training and validation events. No significant differences can be observed between these datasets which allows the conclusion that overtraining is sufficiently suppressed. The separation performance of the training networks is expressed by reasonably large area-under-curve (AUC) values in shown in the legend of Figure 5.21 with values consistently above 0.93 for HH signals and above 0.75-0.95 for backgrounds.

The distributions of the raw DNN outputs are shown in Fig. 5.22 for 2018. In

[2]The architecture of the network is described in CMS internal document AN-2020/095.



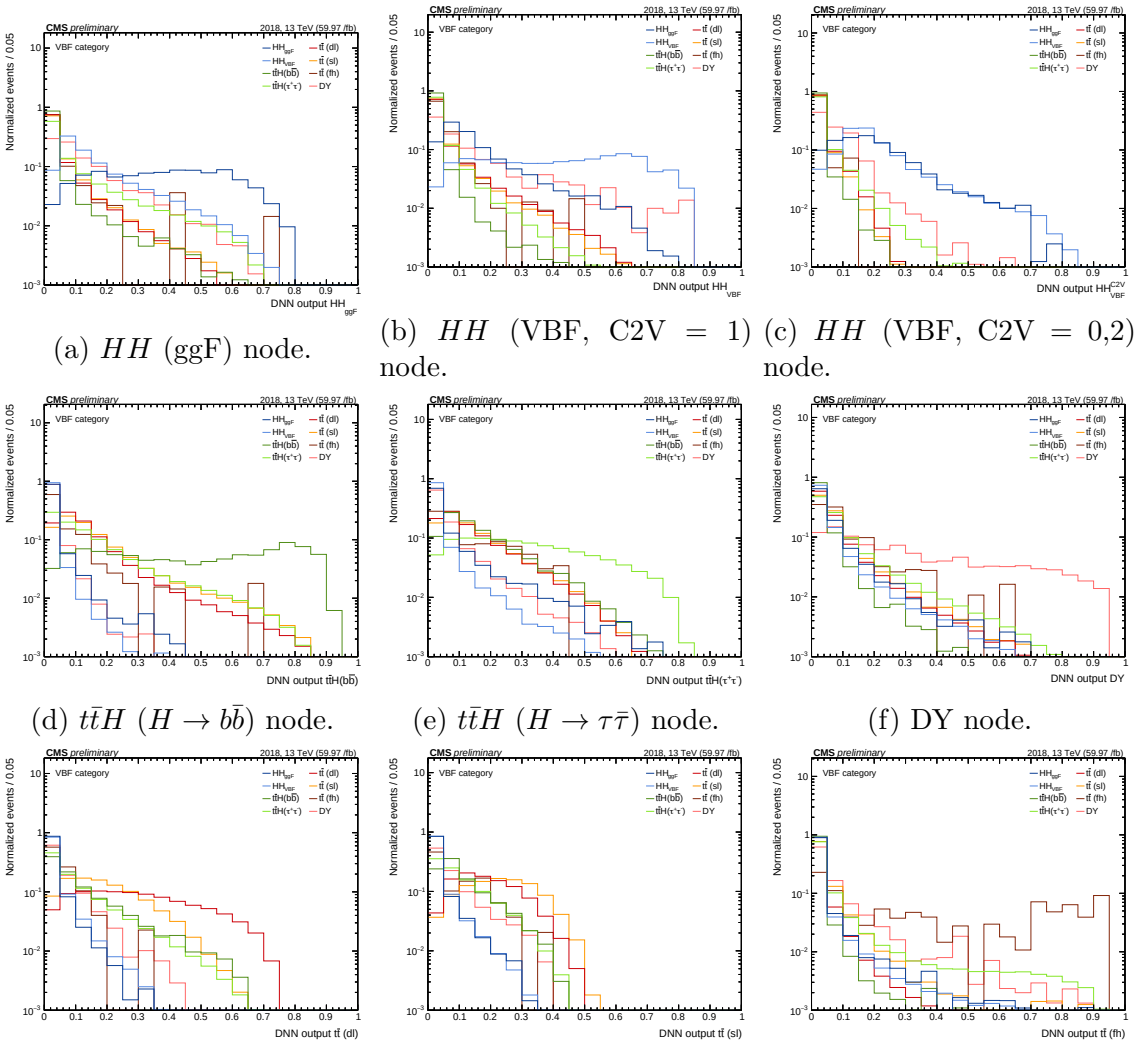
(a) Training data, even event numbers. (b) Validation data, even event numbers



(c) Training data, odd event numbers (d) Validation data, odd event numbers

Figure 5.21: ROC curves showing the relation between the false and true positive rates, evaluated for a particular physics process against all other processes separately on the training (left) and validation dataset (right), and for even (top) and odd (bottom) event numbers. Corresponding area-under-curve values are shown in the legend. The differences between training and validation events are acceptably small to conclude that overtraining is sufficiently suppressed.

each of the output nodes, the respective physics process is clearly emphasized at higher output values compared to other processes. Hence, the presented approach can indeed to employed to create distinct categories that are enriched by different physics processes.



(a) HH (ggF) node. (b) HH (VBF, C2V = 1) node. (c) HH (VBF, C2V = 0,2) node.

(d) $t\bar{t}H$ ($H \to b\bar{b}$) node. (e) $t\bar{t}H$ ($H \to \tau\bar{\tau}$) node. (f) DY node.

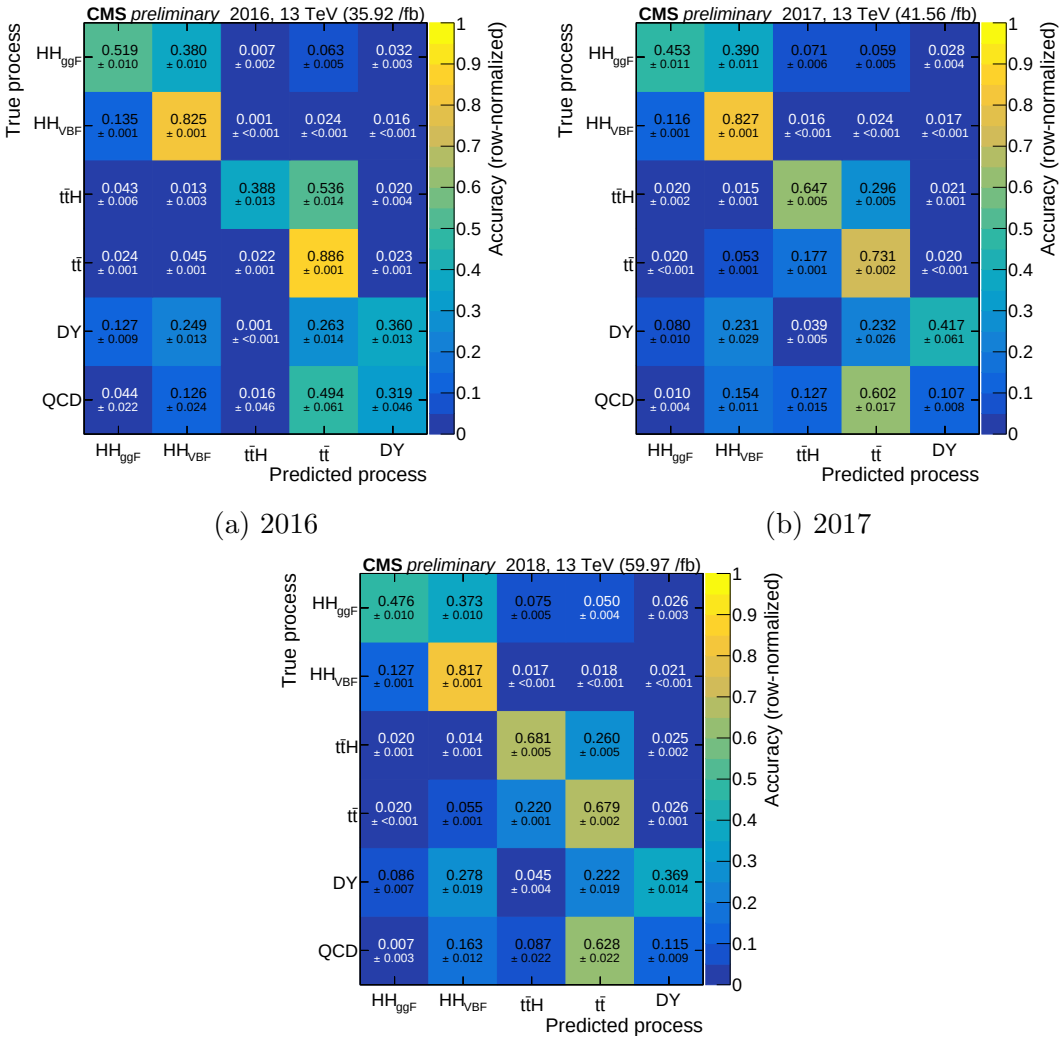
(g) $t\bar{t}$ (DL) node. (h) $t\bar{t}$ (SL) node. (i) $t\bar{t}$ (FH) node.

Figure 5.22: Normalized DNN output distributions for simulated events in 2018.

This finding is further underlined by so-called confusion matrices in Figure 5.23 that arise when performing the above mentioned categorization procedure. The relative numbers are normalized per row and quote the probability for an event emerging from a true physics process on the vertical axis to be assigned to a predicted process on the horizontal axis. Uncertainties are obtained by propagating the effect of the limited amount of available simulated events. Values on the pronounced diagonal signalize reasonably high classification efficiencies. The two left-most columns describe the relative amount of background contamination in the two HH signal categories.

5.4.6 Deep neural network for signal extraction

The general approach used to train the deep neural networks (DNNs) [64], follows that which was used in the CMS di-Higgs HL-LHC projection analysis for the $b\bar{b}\tau^+\tau^-$ channel



(a) 2016

(b) 2017

(c) 2018

Figure 5.23: Row-normalized process classification accuracies and confusion rates for simulated events in different years. Quoted uncertainties are propagated from the limited amount of simulated events. It should be noted that contributions from QCD are determined by means of a data-driven approach and are therefore not attributed an output value by the DNN. However, they appear to be assigned mostly to the $t\bar{t}$ category which seems desirable given their similar expected event topology.

in Ref.[65]. In this approach the Monte Carlo data are split evenly into two sets. A pair of neural-network discriminators are then trained, each on different halves of the data. At inference time the discriminators are used to predict the classes of events in the halves of the data on which they were not trained. In doing so, all of the Monte Carlo data may then be used for inference, thereby reducing the associated statistical uncertainty on the sample density-distributions, whilst not being subject to a systematic bias on the predictions which would otherwise arise if the discriminators were applied to data on which they were trained.

It should be noted that this approach does not provide an explicit validation sample, and does therefore not permit any fine-tuning of hyper-parameters without the risk of unknowingly biasing the model to the testing data. The study performed in Ref. [66] found that the architecture and training scheme that is used has only a weak dependence on architecture parameters, so the lack of tuning is not expected to damage the potential performance too drastically.

Each neural network is trained to identify the class of events, either signal (SM non-resonant $b\bar{b}\tau^+\tau^-$ via gluon fusion) or background (all MC backgrounds considered), by assigning a single prediction per event: closer to zero indicates "background-like" and closer to one indicates "signal-like" events.

Training data

The training data consist of: all MC-driven backgrounds (post-corrections), and SM ($\kappa_\lambda = 1$) non-resonant $b\bar{b}\tau^+\tau^-$ via gluon fusion for all three years. All considered search channels are included, and the jet categories included are res1b, res2b, boosted, and VBF (as defined in Sec. 5.4.2).

Training is performed considering the weight associated to each event. In order to train an unbiased classifier, the total weights of signal and background are both normalised to one, such that the classifier does not focus mainly on accurately classifying the background. The reweighting is performed separately in each channel and category: it was already demonstrated that this mode of reweighting further improved signal sensitivity, by magnifying the importance of the less populated, but more discriminant categories, such as boosted $\tau_h\tau_h$.

Feature selection

A large range of over 100 features of the data are computed. These include final-state 4-momenta components, relative angles of the final state objects in a variety of rest-frames, reconstructed masses, transverse masses, and features already produced by the reconstruction and tagging algorithms. It is assumed that the full menu of features is larger than what is required to train well-performing models, and that it will contain redundant information. Therefore the menu is run through a process of feature selection, in order to narrow it down to only the most useful and necessary features for our particular classification task.

Initial removal It was already found that of the final-state 4-momenta, only the transverse momenta were potentially useful and that the networks did not perform better when provided the full range of 4-momenta components. These were therefore removed. Next, any features with values that were constant throughout the datasets were removed. Only

one such feature was found: a flag indicating whether SVfit converged (it always converged in the selected data).

Monotonically related features Feature selection is primarily driven by evaluating the permutation importance of each feature, as discussed in Sec. 5.4.6. This method fails, however, when one or more features are correlated with one another, since the importance of the set of correlated features is instead shared across all features within the set, reducing the apparent importance of any one of the features. It is therefore necessary to first filter down the presence of correlated features.

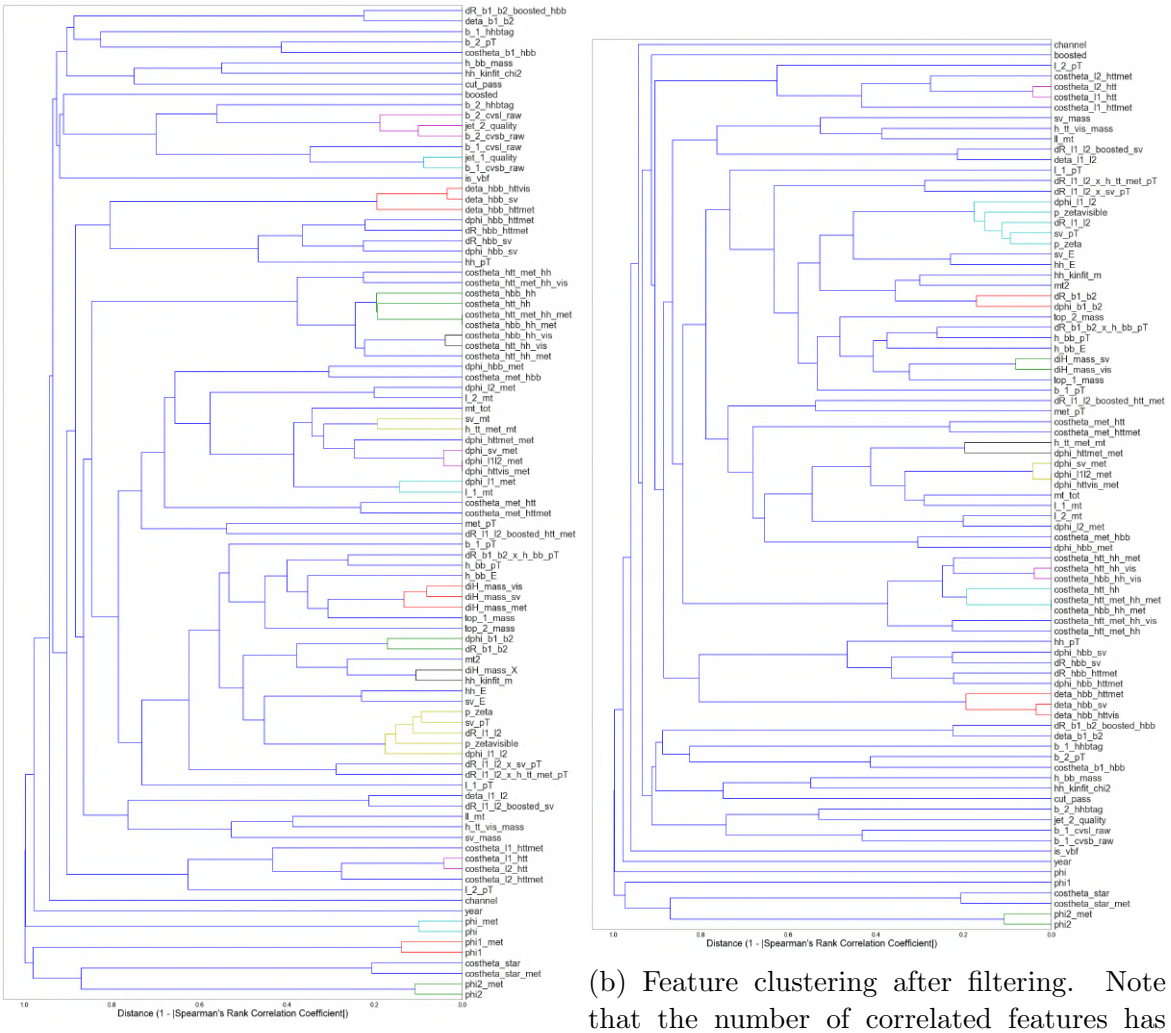
Groups of monotonically related features are identified via hierarchical clustering of features according to the Spearman's rank-order correlation coefficient [67] (SROCC) of each pair of features. Groups of features clustered with a SROCC lower than 0.2 are considered *correlated* with one another. Random Forest classifiers [68] are then used to check whether features from each group can be removed without damaging the performance of the classifier. First, models are trained on the full set of remaining features, next a correlated feature is temporarily removed and new models are trained. If the average performance does not decrease, then the feature is considered safe for removal. if the removal of two or more features from a correlated cluster results in no performance loss, then the single feature whose removal gives the largest increase in performance is removed. The test repeats on the remaining features in the cluster until only one feature remains, or no features can be safely removed. An example of clustering before and after filtering is shown in Fig. 5.24.

Permutation importance The main selection process for features is based on their permutation importance (PI) - a measure of how much the performance of a trained model degrades if a given feature of the input data was randomly shuffled and all others kept the same. If the feature was important, then shuffling it, and thus destroying the information it carries, should reduce the accuracy of the predictions.

PI, here, is computed using Random Forests. It is possible, though, that the importance of features can change depending on model training and the particular training dataset. To account for this multiple models are trained on a subsample of the original dataset, and features with an average PI greater than, or equal to, a prescribed threshold are said to be *important*. This process is then repeated on new subsamples and features are then selected based on the fraction of times they are *important*. Figure 5.25 illustrates and example of feature importance, and why repeated evaluation is beneficial.

Mutual dependence The checks performed in Sec. 5.4.6 were only sensitive to monotonically related features. It is possible, though, that features may display more complicated relationships. One can check for this *mutual dependence* by attempting to regress (here, using Random Forests) to the values of one feature using the other features as inputs. The feature importance of the inputs to the regressor can then be used to identify which features are used to predict the values of the target feature. Accurately regress-able features are targets for removal as per the removal technique in Sec. 5.4.6. Figure 5.26 illustrates an example of removal by mutual dependence.

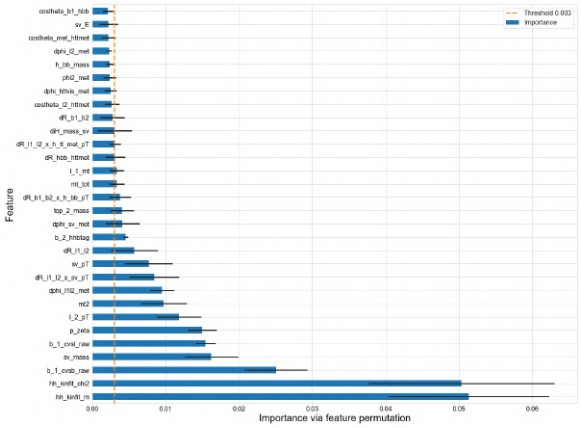
Ideally, this process should be run before the selection by permutation importance, however it is comparatively slow to run on a large number of features. Instead, the feature selection by permutation importance is kept moderately loose, to avoid rejecting



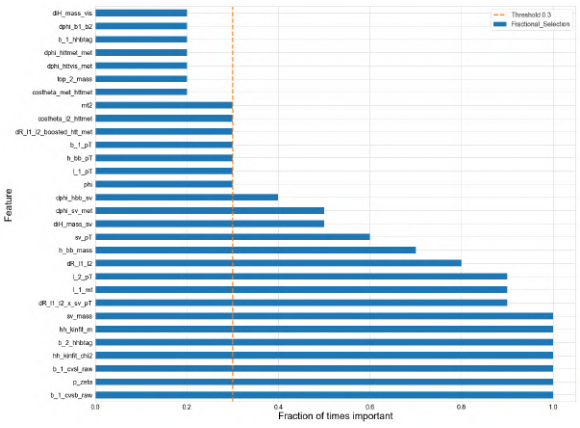
(a) Feature clustering prior to filtering. Correlated clusters are indicated with non-blue connection line

(b) Feature clustering after filtering. Note that the number of correlated features has been reduced, but some correlated features were unable to be removed safely.

Figure 5.24: An example of feature removal based on hierarchical clustering via Spearman's rank-order correlation coefficient.



(a) Example of average permutation importance for a certain subsample of data.



(b) Example of repeated evaluation of permutation importance on multiple subsamples of data and checking the fraction of times features are deemed *important*.

Figure 5.25: Feature selection based on permutation importance. In the left plot, some features are close to the selection threshold and could easily pass or fail based on slight changes in the data or model. In the right, after repeated evaluations, one can see that some features are consistently important, whereas some are only occasionally important.

important features with mutual dependence and instead the filtering of these is left to the mutual dependence selection.

Feature pre-processing

Continuous features are normalized to have a mean of zero and a standard deviation of one. The necessary transformations are computed twice, once on each of the two halves of the dataset. When training each half is pre-processed using the transformation that was fitted to the half of the dataset. During application, when each discriminator is applied to the opposite halves of the data, the features are pre-processed using the opposite transformation as well, in order to ensure optimal compatibility.

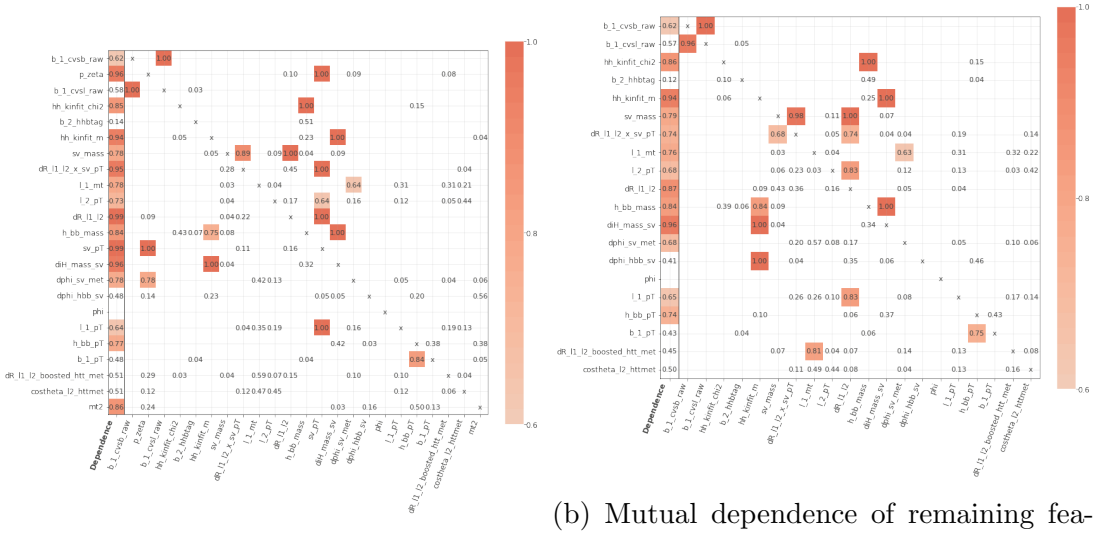
The transformations are computed whilst invalid or undefined feature values are still `NaN`, in a way that ignores `NaN` values. Following the application of the transformations, any `NaN` values are replaced with zeros [3]

Input features

The final continuous input features are listed here in descending order of importance. The ranking is based on Random Forest models: features at the top are consistently important, features in the middle are usually found to be important, and features near the bottom are only important at least 30% of the time.

- DeepFlavour charm tagger $CvsB$, binned in working points, of the first b-jet.
- DeepFlavour charm tagger $CvsL$, binned in working points, of the first b-jet.
- The χ^2 of the kinematic fit to the invariant mass of the $HH \to b\bar{b}\tau^+\tau^-$ system.

[3] The architecture of the network is described in CMS internal document AN-2019/188.



(a) Mutual dependence of features prior to filtering.

(b) Mutual dependence of remaining features after filtering. Some remaining features display mutual dependence, however they cannont be safely be removed.

Figure 5.26: Mutual dependence of features as estimated via Random Forest Regression. The "Dependence" column indicates ease of prediction of that row's feature values, and the other columns indicate the relative importance of each feature in predicting the dependent feature.
As an example, in 5.26a `sv_pT` is easily predictable (dependence of 0.99), and the most important feature in predicting its values is `p_zeta` (importance of 1.00). When testing the removal of `sv_pT`, it is found that its presence as a training feature has no impact on the average performance of a classifier and so it is removed.

- The invariant mass of the $HH \to b\bar{b}\tau^+\tau^-$ system from the kinematic fit.
- The invariant mass of the $\tau\tau$ system from the SVfit algorithm.
- The product $\Delta R(\tau_1, \tau_2) \cdot p_T(H_{\tau\tau}^{SVfit})$.
- The transverse mass of the first tau lepton: the e or μ in case of semi-leptonic decays, or the most isolated τ_h in the fully hadronic decay case.
- The transverse momentum of the second tau lepton: the τ_h case of semi-leptonic decays, or the second most isolated τ_h in the fully hadronic decay case.
- The ΔR separation between the two tau lepton candidates.
- The $\Delta\phi$ separation between the $H \to \tau\tau$ candidate, reconstructed with the SVfit algorithm, and the MET.
- The mass of the $H \to bb$ candidate.
- The HH-Btag score of the second b-jet.
- The invariant mass of the system $H_{bb} + H_{\tau\tau}^{SVfit}$.
- The $\Delta\phi$ separation between the H_{bb} and the $H_{\tau\tau}^{SVfit}$ systems.
- The transverse momentum of the H_{bb} system.
- The ΔR separation between the two tau candidates in the rest frame of the $H_{\tau\tau} + MET$ system.
- The transverse momentum of the first tau lepton: the e or μ in case of semi-leptonic decays, or the most isolated τ_h in the fully hadronic decay case.
- The transverse momentum of the first b-jet.
- Angle between the decay planes of the four final state particles expressed in the rest frame of the HH system.
- Cosin of theta angle between the second tau lepton and the direction of flight of $H_{\tau\tau} + MET$ system in the $H_{\tau\tau} + MET$ rest frame.

In addition, six categorical features are used as input to the neural network:

- If the event is boosted or not.
- The $\tau\tau$ decay mode.
- If the event has two VBF jet candidates, as defined in Section 5.4.2.
- Highest working point of the DeepFlavour algorithm passed by the first b-jet.
- Highest working point of the DeepFlavour algorithm passed by the second b-jet.
- The year of the data taking.

5.5 Background Estimation

Since multiple sources of background affect the $b\bar{b}\tau^+\tau^-$ analysis and different signal topologies are explored in the search, an accurate modeling of all processes involved is crucial in order to optimize the analysis strategy and techniques, and to obtain a valid comparison between the observed data and the theoretical predictions.

The backgrounds can be classified in two main categories, either as "irreducible" or as "reducible" contributions. The former are composed of processes that lead to the exact same final state as in the $HH \to b\bar{b}\tau^+\tau^-$ decay, that is the object of the search described in this thesis. The two most important contributions in this category originate from the $t\bar{t} \to b\bar{b}WW \to b\bar{b}l\nu_l\tau\nu_\tau$ decay and from the Drell-Yan production of a tau pair in association to a b quark pair. On the other hand, reducible backgrounds arise from the misidentification of objects due to experimental detector effects, the most striking case being the erroneous identification of gluon or light quark initiated jets with a τ_h candidate or as a b jet. The perfect example of reducible background is the QCD multi-jet contributions, especially relevant in the $\tau_h\tau_h$ channel.

In order to handle these two categories of background sources, different strategies are exploited. The reducible contributions are suppressed through the application of tight quality selection requirements that aim at guaranteeing an high background rejection efficiency. Since these selections can result in a loss in signal acceptance the optimal working point is the one that balance the two effects. Irreducible background sources can be tackled only by exploiting the kinematic differences with respect to HH signal events.

The strategies put in place in order to reject background events are thoroughly described in Section 5.3 and Section 5.4.3 for the reducible and irreducible contributions, respectively.

5.5.1 QCD multi-jet background

One of the main difficulties in the $b\bar{b}\tau^+\tau^-$ analysis is the correct identification of hadronically decaying taus. Thus, together with the $t\bar{t}$ processes, the multi-jet QCD events represent the main background source, especially in the fully hadronic $\tau_h\tau_h$ final state.

Most of the CMS analyses involving $\tau\tau$ pairs in the final state evaluate this contribution either estimating the $jet \to \tau_h$ rate in data sidebands, or computing the yield and shape of multi-jet distributions in jet enriched regions in data, as in the case of the $HH \to b\bar{b}\tau^+\tau^-$ search here described. The use of MC simulation to evaluate the QCD contribution is disfavored due to two main factors: firstly, the probability for a quark or a gluon jet to be identified as a τ_h object is very low and it has to be combined with the equally poor probability to have in the event two additional jets that pass the medium working point of the b tagging discriminator. To cope with these rates, the QCD sample generated with a MC simulation would require a too large number of events to ensure a sufficient number of them is present in the phase space considered in the analysis. The second reason why the data-driven method is preferred is the fact that the misidentification rate for τ_h objects is mainly lead by detector effects that are very complex to simulate properly and can change over time in account of many external factors impossible to predict in advance.

In the analysis described in this thesis, the so-called ABCD method is adopted in order to model and estimate the QCD multi-jet background from jet enriched regions in data. The phase space of the events is divided in four regions, using the values of

two uncorrelated variables, tau pair sign and tau isolation (the sub-leading tau in the $\tau_h\tau_h$ channel). A schematic representation can be seen in Fig. 5.27:

- Region A: Represents the signal region as defined in Section 5.4.1 and contains a pair of opposite sign electric charge (OS) tau leptons (either $\tau_{e/\mu}/\tau_h$ or $\tau_h\tau_h$) and where all τ_h objects pass the medium working point of the tau isolation discriminant.
- Region B: It is defined with the same isolation selections, but the pair charge requirement is inverted (same sign or SS).
- Region C: It is composed by events with an opposite sign tau pair where τ_h objects pass the VVLoose, VLoose or Loose working point of the tau isolation discriminant, but are required to fail the medium WP that defines the signal region. In the $\mu\tau_h$and $e\tau_h$channels this tau isolation selection is applied to the only τ_h candidate present in the event, while in the $\tau_h\tau_h$final state it is applied only to the lowest p_T τ_h candidate selected.
- Region D: It is the region most different from the signal phase space as it has the same tau isolation criteria of region C, but it also requires that the leptons in the tau pair have the same electric charge.

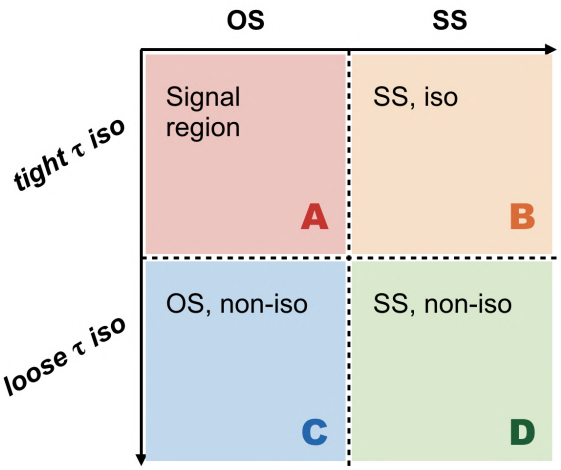


Figure 5.27: Schematic representation of the four regions used to estimate the QCD multijet background

In order to properly estimate the multi-jet yield contributions coming from the other backgrounds estimated with MC simulation are subtracted from the data yield in the B, C and D regions. The QCD background yield in the signal region A is then estimated from region B through the following extrapolation factor:

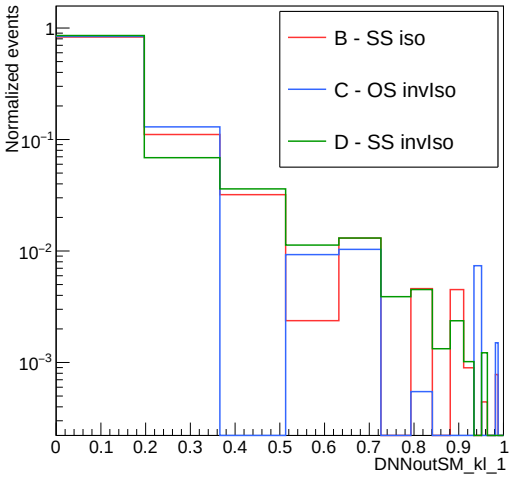
$$N_A = N_C \times \frac{N_B}{N_D} \tag{5.34}$$

In the boosted category, due to the limited statistics, the correction factor N_B/N_D is estimated from events that pass all the selections that define the boosted category (as detailed in Section 5.4.4) excluding the b-tagging requirements. For the same reason in each of the VBF sub-categories (VBF-HH, GGF-HH, ttH, $t\bar{t}$, DY as defined in Section 5.4.4)

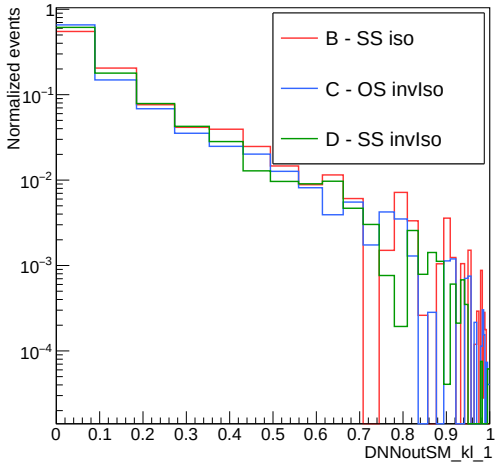
the correction factor N_B/N_D is estimated from an inclusive category that includes all the five sub-categories themselves.

The differential distribution (or "shape") of the QCD background is evaluated by subtracting from the data all the MC simulation contributions in each bin of the distributions in the C region.

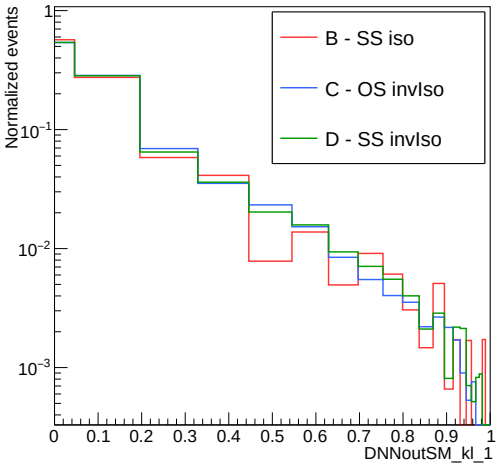
It is assumed that the shapes of the QCD background in the B, C and D regions are compatible. This assumptions turns out to be a good approximation as is showed in Fig. 5.28 where the Data - MC shapes are reported for the three analysis channels in the resolved 1b1j category.



(a) $e\tau_h$ channel



(b) $\mu\tau_h$ channel



(c) $\tau_h\tau_h$ channel

Figure 5.28: Data - MC shapes of the final discriminating variable in the B, C and D control regions for the $e\tau_h$ (a), $\mu\tau_h$ (b) and $\tau_h\tau_h$ (c) channels in resolved 1b1j category. 2018 data and MC samples are used.

5.5.2 Drell-Yan $Z/\gamma^* \to \tau\tau$ background

As already mentioned in the introduction to this Chapter, Drell-Yan decays in association with the production of two jets represent one of the main background contributions for

the $b\bar{b}\tau^+\tau^-$ analysis.

Given the narrow phase space studied in the analysis, the DY statistics is increased by combining the inclusive sample with complementary ones, where the emission of 1, 2, 3 or 4 additional jets, or the emission of 1 or 2 b jets, is required. Since the full MC simulation process is quite consuming, in terms of computing time and power, in all the events of these samples the invariant mass m_{ll} is forced to be larger than 50 GeV without loosing any information from low mass events that would be in any case be excluded from the analysis by the selections used to define the signal region.

If, on one hand, the differential distributions of the DY events show a good agreement with the observed data, the modelization of the yield, especially when the production is in association with multiple jets, is known to be imperfect and thus requires a correction, that in the case of the $b\bar{b}\tau^+\tau^-$ analysis is computed from data in a di-muon control region.

The trigger requirements to select $Z \to \mu\mu$ events are shown in Table 5.9 for 2016, 2017 and 2018 data-taking period. In the data takings where more than one HLT path is available a logical OR between the HLT paths has been used.

Year	HLT Path	Recorded Lumi (fb$^-$1)
2016	`HLT_IsoMu22_v*`	28.6
2017	`HLT_IsoMu24_v*`	38.05
	`HLT_IsoMu27_v*`	41.53
2018	`HLT_IsoMu24_v*`	59.74
	`HLT_IsoMu27_v*`	

Table 5.9: Trigger requirements to select the Z $\to \mu\mu$ + jets sidebands region for 2016, 2017 and 2018 data-taking period.

Offline selections require the presence of two muons with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$. The vertex of both the muons must be close to the main primary vertex within $\Delta_{xy} < 0.045$ mm and $\Delta_z < 0.2$ mm. The muons should also pass the tight muon selection criteria provided by the Muon POG and have a relative isolation $I^{\mu}_{rel} < 0.15$ calculated in a cone of size $\Delta R < 0.4$ around the muon. The two selected muons should be separated by $\Delta R > 0.1$ and have an opposite charge. The muon pair should have an invariant mass of $m_{\mu\mu} > 50\,\text{GeV}$. The jet selection criteria and the requirement of third lepton veto are the same as for the other final states.

To reduce the QCD and the $t\bar{t}$ contributions in the sideband region, a cut on the missing transverse momentum is applied requiring $MET < 45$ GeV.

Both data and MC simulated events thus selected are split in three regions according to the number of jets that pass the medium working point of the b tag discriminant.

- **Z + light jets control region.** The contribution of Z+light jets to the signal region is due to the erroneous identification of light jets as b-jets. In order to estimate this contribution the two jets are required to fail the medium working point of the b-tag discriminant.

- **Z + one b-tagged jet control region.** The second control region is defined in order to enrich the contributions of Z + one b-jet events. The selections are the same as for the previous region, but one jet is required to pass the medium working point of b-tagging discriminant, while the second one is still required to fail it.

- **Z + two b jets control region.** The most important source of background is the production of Z + two b jets. The final state in this case is the same of the signal, the only difference is that the b jets are not coming from the Higgs decay. In this region both jets are required to be b-tagged.

Each of these control regions is further divided in 6 sub-regions based on the p_T of the reconstructed Z candidate. The p_T intervals are defined according to the level of data-MC agreement and are slightly different for 2016 with respect to 2017 and 2018.

The data distribution of the invariant mass of the muon pair is fitted simultaneously in all the categories with the 19 MC templates described above (the additional MC template combines all other backgrounds). The normalization of the MC templates are kept floating during the simultaneous fit. The scale factors resulting from the simultaneous fit will be applied to the corresponding DY contributions henceforth in the analysis. The errors on the scale factors and their covariance matrix are obtained from the fit and used as source of systematic uncertainty.

The effect of these scale factors can be seen in Fig.5.29 where the distributions of the invariant mass and angular separation of the two muons are reported before and after the application of the SF.

5.5.3 $t\bar{t}$ background

The main contribution of background in the semileptonic channels originates from $t\bar{t}$ events. Three different samples that cover the total phase space (fully hadronic, fully leptonic and semileptonic) are used in this analysis, they are centrally produced by CMS assuming a top quark mass of 172.5 GeV. Measurements of the $t\bar{t}$ differential production cross section in CMS show that this process is well described by the MC simulation but the agreement with the observed data is not satisfactory when two or more jets are required to be b-tagged, as shown in Fig. 5.30 where the $H \to bb$ and $H \to \tau\tau$ visible masses are reported in the resolved 2b0j category of the $e\tau_h$ channel using the 2016 data and MC samples. This disagreement has also been observed in other analysis within the di-Higgs CMS group. In order to take this disagreement into account a $t\bar{t}$ control region has been used to derive the correct normalization for this sample. In the following, this scale factor will be referred to as ttSF.

The definition of the CR follows four main considerations: the CR must be enriched in $t\bar{t}$ events, it must be as much depleted as possible of the other backgrounds, it must be as similar as possible to the SR while remaining orthogonal to it, and it must not depend on the τ pair decay mode (i.e. must not depend on the $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ channels).

To make the CR independent of the τ pair categorization, one single CR is defined for each year. To do so, the following procedure is applied: events are first categorized following the flow outlined in Section 5.4.1, then the $\mu\tau_h$, $e\tau_h$, and $\tau_h\tau_h$ events are merged in a single CR for each year. The decision of using three differnt CRs for the three years is based on two fact. Firstly, the CMS experimental apparatus has undergone several changes throughout the years considered; secondly, the disagreement observed is different in three years both in our analysis and in the other analyses that see the discrepancy. Nevertheless, it should be noted that the disagreement for the 2017 and 2018 data taking periods is comparable between the two and smaller with respect to 2016.

To make the CR as similar as possible to the SR, and to make it enriched in $t\bar{t}$ events, the same selections outlined in Section 5.4.2 for the *resolved 2jet–2tag* are applied. This ensures that a genuine $b\bar{b}$ pair, conceivably coming from the $t\bar{t}$ decay, is found.

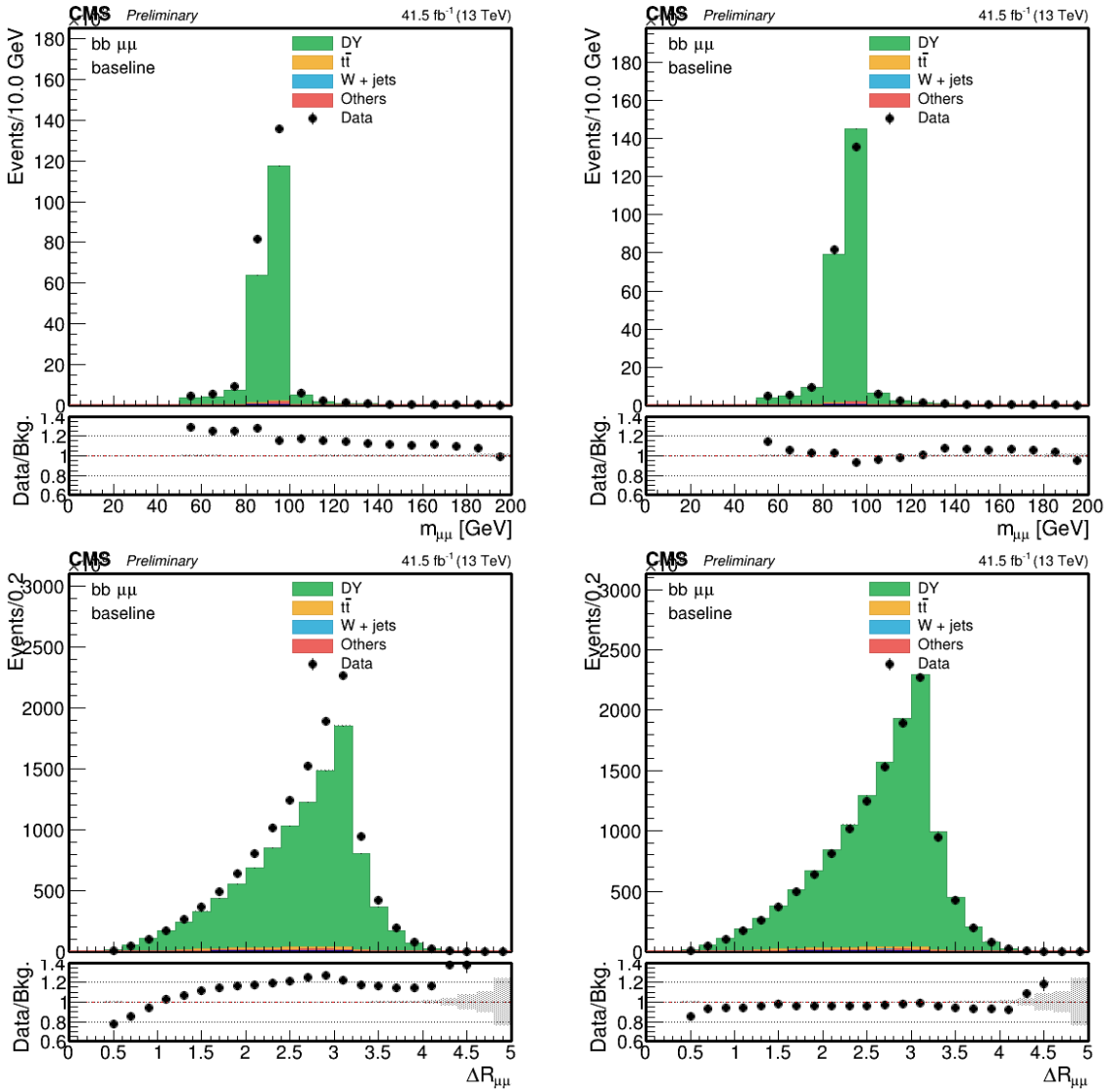


Figure 5.29: Distributions of the di-muon invariant mass and angular separation of 2017 data and MC simulated samples with (right) and without(left) scale factors.

To make the CR orthogonal to the SR, the elliptical mass cut defined in Section 5.4.3 is inverted. Therefore, the mass cut used for the $t\bar{t}$ CR is:

$$\frac{(\mathrm{m}_{\tau\tau} - 129\,GeV)^2}{(53\,GeV)^2} + \frac{(\mathrm{m}_{bb} - 169\,GeV)^2}{(145\,GeV)^2} > 1 \tag{5.35}$$

This mass cut not only guarantees orthogonality with the SR but also ensures that the CR is enriched in $t\bar{t}$ events and at the same time depleted of signal events.

Since no shape trend is observed in the disagreement, the variable to be used for the fits can be arbitrarily chosen. For simplicity, the score of the DNN described in Section 5.4.6 applied to the $t\bar{t}$ CR is used. This choice has been taken only to keep a certain uniformity with the signal extraction procedure. The CR only fits are performed under three different binning paradigms of the DNN score: one single bin, three variable widths bins (i.e. [0, 0.05, 0.3, 1]) and twenty constant width bins.

The results of the CR only fit are reported in Table 5.10 for the three years and the three binning paradigms described above. As can be seen, no shape dependence of the fit can be spotted. In fact, the central value of the fit remains almost constant at the variation of the binning paradigm, and when it changes, the variations are well compatible within the errors.

	CR (1bin)	CR (3bin)	CR (20bin)
2016	0.903 ± 0.007	0.903 ± 0.007	0.903 ± 0.007
2017	0.963 ± 0.006	0.963 ± 0.006	0.963 ± 0.006
2018	0.957 ± 0.005	0.956 ± 0.005	0.956 ± 0.005

Table 5.10: Results of the CR fits under the three binning paradigms for the three years. The ttSF fitted values and their errors are reported.

The same distributions showed in Fig. 5.30 are reported in Fig. 5.31 with the correct normalization for the $t\bar{t}$ sample.

5.5.4 Other backgrounds

Other background contributions show a very limited presence in the phase space considered in the $b\bar{b}\tau^+\tau^-$ analysis and their contribution and modeling, both in shape and event yield, are assessed relying solely on Monte Carlo simulation.

Standard model single Higgs The single Higgs production cross section is small compared to the other backgrounds. However the production of single Higgs (through gluon fusion or vector boson fusion) and the production in association with a vector boson (VH) or with a pair of top quarks (ttH) have similar final states with the $HH \to bb\tau\tau$ signal and are therefore considered as backgrounds in this analysis. The MC samples are generated considering a SM Higgs boson with $m_H = 125\,GeV$.

Di- and Tri-boson Processes involving the presence of a pair of vector bosons (ZZ, WW or WZ) are modeled using exclusive decay samples, while events involving three vector bosons (WWW, WWZ, WZZ or ZZZ) are modeled with inclusive samples.

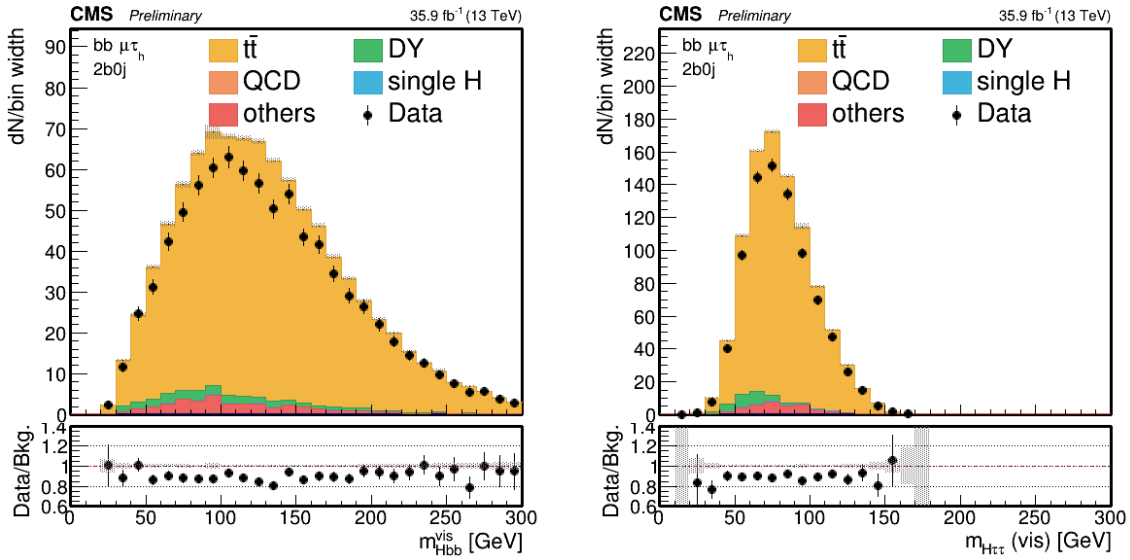


Figure 5.30: Distributions of the $H \to bb$ and $H \to \tau\tau$ visible masses in the resolved 2b0j category of the $\mu\tau_h$ channel using 2016 data and MC samples. A disagreement between data and MC simulation is clearly visible.

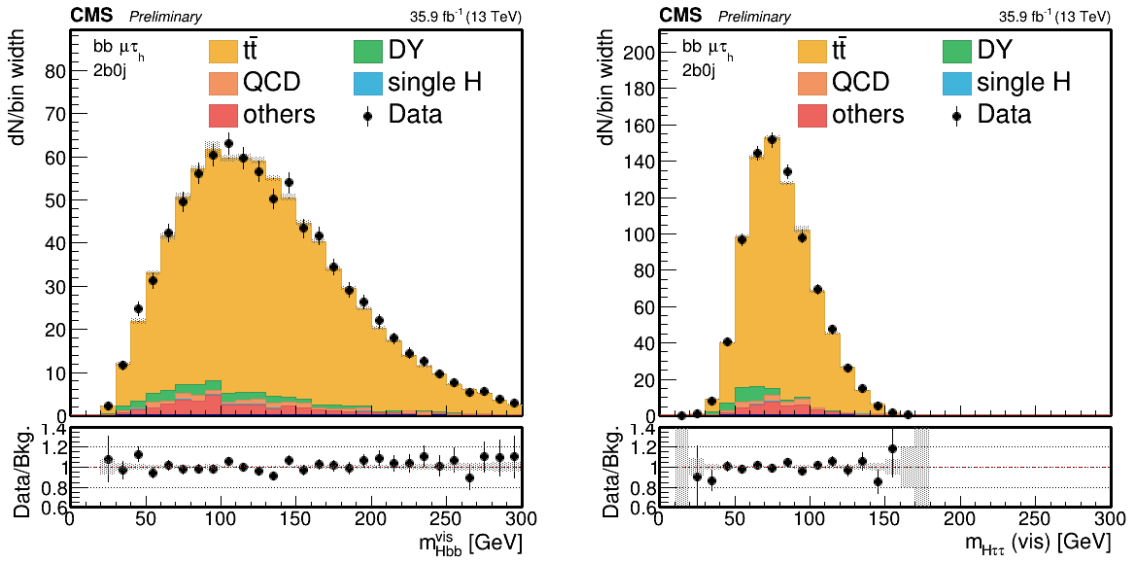


Figure 5.31: Distributions of the $H \to bb$ and $H \to \tau\tau$ visible masses in the resolved 2b0j category of the $\mu\tau_h$ channel using 2016 data and MC samples with the correct $t\bar{t}$ sample normalization.

Other processes The contribution of W+jets background is highly suppressed after the requirement of 2 b-tagged jets, while the contribution arising from electroweak processes (V+2jets) and from single top quark production, both in the `s-` and `t-channels`, are extremely small.

Processes with a $t\bar{t}$ pair produced in association with one or two vector bosons have very small cross sections, nevertheless they are taken into account in this analysis as they represent one of the main sources of background in the vector boson fusion categories.

5.6 Signal Modeling

The exploration of various BSM scenarios requires the modeling of the signal for several different sets of couplings. Since only a limited set of simulated samples can be produced, weighting techniques are implemented both for the gluon fusion and the VBF signals, in order to model additional BSM scenario starting from a small set of fully simulated BSM signals.

For gluon fusion HH samples, two different reweighting methods are used: one applies to the samples simulated at leading order (LO), it is described in paragraph 5.6.1 and it was also used in the 2016 $bb\tau\tau$ analysis [69] and in the 2016 HH CMS combination [70]. On one hand, this technique provides a large number of signal events: all the LO ggF data sets are merged together and reweighted to the target BSM scenario. On the other hand, the LO modelling does not consider parton emission at the matrix element level. Thus, a new reweighting procedure to be applied to NLO samples has been adopted by all HH analyses in view of the CMS Run2 Legacy results, and it is described in paragraph 5.6.2. The ggF LO modelling is used to produce events for the training of the machine learning techniques implemented in this analysis, while the ggF NLO modelling is used for the final signal extraction.

The VBF modelling strategy is similar to that used for ggF NLO samples, as described in paragraph 5.6.3.

5.6.1 LO Gluon Fusion modelling

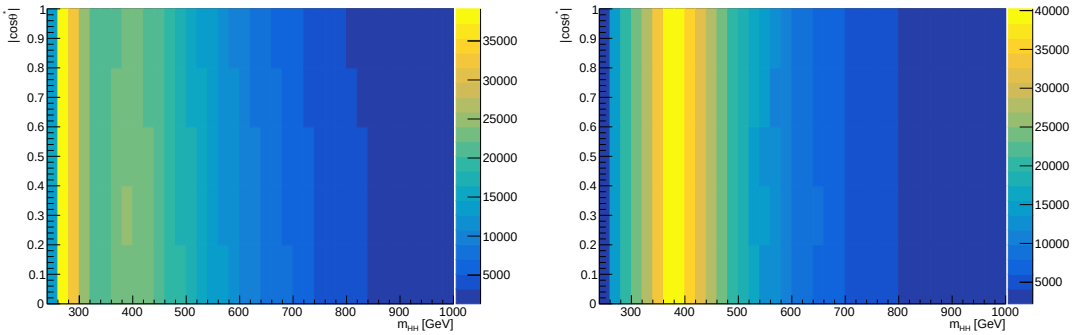
The weighting technique for the LO gluon fusion modelling allows various scenarios to be explored in terms of the effective Lagrangian parametrization described in [71], where the Higgs pair production is regulated by the five couplings y_t, λ_{HHH}, c_2, c_{2g} and c_g; the variations from the SM values of the couplings are expressed as $k_\lambda = \lambda_{HHH}/\lambda_{HHH}^{SM}$ and $k_t = y_t/y_t^{SM}$.

The two Higgs bosons are produced back-to-back in the reference frame of the center of mass; before any hadronization effect, they have identical transverse momenta and opposite azimuthal angle. At this level, the kinematics of the event is totally determined by two parameters: the invariant mass of the HH system and the angle $\cos\theta^*$ between one Higgs boson and the beam axis [71]. These variables are exploited by the weighting procedure.

For each of the identified BSM benchmarks, signal samples are produced centrally by CMS at leading order (LO) precision with MadGraph5_aMC@NLO; the samples used in this search are marked as "GGF LO" and reported in tables 5.11, 5.12 and 5.13.

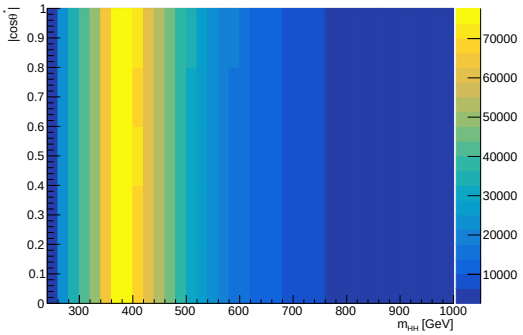
The events of all the gluon fusion HH signal samples are combined to build a 2D distribution as a function of m_{HH} and $|\cos\theta^*|$, computed using simulated Higgs boson properties after the hard scatter and before hadronization effects. The 2D distributions

of events $f(\mathrm{m}_{HH}, |\cos\theta^*|)$ are shown in Figure 5.32. An identical histogram is filled using the SM signal sample only. The content of a bin j in the two bidimensional distributions, normalised to unity, is denoted as f^j_{comb} and f^j_{SM}.



(a) 2016

(b) 2017



(c) 2018

Figure 5.32: Distribution of the simulated events of non-resonant HH production in the $[\mathrm{m}_{HH}, |\cos\theta^*|]$ phase space for the combination of all the available samples in each year: 2016 in Figure 5.32a, 2017 in Figure 5.32b and 2018 in Figure 5.32c.

The ratio of the total HH cross section over the SM prediction, can be expressed as a function of the j bin number as

$$R^j_{HH} = \frac{\sigma^j_{HH}}{\sigma^{j,\mathrm{SM}}_{HH}} = A^j_1 k_\mathrm{t}^4 + A^j_2 c_2^2 + (A^j_3 k_\mathrm{t}^2 + A^j_4 c_g^2)k_\lambda + A^j_5 c_{2g}^2 + (A^j_6 c_2 + A^j_7 k_\lambda k_\mathrm{t})k_\mathrm{t}^2 + (A^j_8 k_\mathrm{t} k_\lambda + A^j_9 c_g k_\lambda)c_2 + (A^j_{10} c_2 c_{2g} + (A^j_{11} c_g k_\lambda + A^j_{12} c_{2g})k_\mathrm{t}^2 + (A^j_{13} k_\lambda c_g + A^j_{14} c_{2g})k_\mathrm{t} k_\lambda + A^j_{15} c_g c_{2g} k_\lambda. \tag{5.36}$$

The ratio R^j_{HH} is computed using simulated events with different sets of couplings; thus, the A^j_i coefficients are extracted from its interpolation as a function of the couplings.

Finally, event-by-event weights are computed as

$$\omega = \frac{\Omega}{\sum_n \Omega} \tag{5.37}$$

where

$$\Omega(k_\lambda, k_\mathrm{t}, c_2, c_g, c_{2g}; j) = \frac{f^j_{\mathrm{SM}}}{f^j_{comb}} \cdot \frac{R^j_{HH}(k_\lambda, k_\mathrm{t}, c_2, c_g, c_{2g})}{R_{HH}(k_\lambda, k_\mathrm{t}, c_2, c_g, c_{2g})} \tag{5.38}$$

and the sum goes over the number of simulated events; thus, only the differential event distribution is modified and not the global normalization.

5.6.2 NLO Gluon Fusion modelling

A different procedure, based on the sum of different samples, is used to reweight NLO samples due to the presence of an additional parton at matrix level that prevents the usage of the same technique used for LO samples.

At leading order the the amplitude of the gluon fusion HH production process can be written as:

$$\mathcal{A} = k_\lambda k_{top} T + k_{top}^2 B \tag{5.39}$$

where, T and B can be directly associated to the triangle and box diagrams of Fig. 5.1. The HH production cross section via gluon fusion can thus be written as the square of the amplitude:

$$\sigma(k_\lambda, k_{top}) \sim |\mathcal{A}|^2 = k_{lambda}^2 k_{top}^2 t + k_{top}^4 b + k_{lambda} k_{top}^3 i \tag{5.40}$$

where $t = |T|^2$, $b = |B|^2$, and $i = |TB^* + B^*T|$.

Equation (5.40) can also be rewritten in a matricial form as

$$\sigma(k_\lambda, k_{top}) = \mathbf{c}(k_\lambda, k_{top}) \cdot \mathbf{v} \tag{5.41}$$

where $\mathbf{c}(k_\lambda, k_{top} = (k_\lambda^2 k_{top}^2,\ k_{top}^4,\ k_\lambda k_{top}^3)$ is the vector of the coupling functions and $\mathbf{v} = (t,\ b,\ i)$ is the vector of the values of the three components.

For a fixed set of (k_λ, k_{top}) values, the HH signal can thus be described as the linear sum of three components.

Given three samples $(s_1,\ s_2,\ s_3)$ produced at NLO using POWHEG with a specific choice of k_λ and, denoting with $(c_i^1,\ c_i^2,\ c_i^3)$ the 3 elements of the coupling vector for each s_i samples, we can express equation (5.40) in terms of a linear combination of these three samples:

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix} = \begin{pmatrix} c_1^1 & c_1^2 & c_1^3 \\ c_2^1 & c_2^2 & c_2^3 \\ c_3^1 & c_3^2 & c_3^3 \end{pmatrix} \begin{pmatrix} t \\ b \\ i \end{pmatrix} \tag{5.42}$$

or again in matricial form:

$$\boldsymbol{\sigma} = \mathbf{C}\mathbf{v} \tag{5.43}$$

Given that the three values of σ_i are provided from the POWHEG generator, the matrix $\mathbf{C}$ can be inverted to express:

$$\mathbf{v} = \mathbf{C}^{-1}\boldsymbol{\sigma} \tag{5.44}$$

so that by simple substitution equation (5.41) becomes:

$$\sigma(k_\lambda) = \mathbf{c}^T(k_\lambda)\mathbf{C}^{-1}\boldsymbol{\sigma} \tag{5.45}$$

This can be applied to build the differential $d\sigma/dx$ distributions as well.

A generic value of k_λ can therefore be easily modelled by manipulating a few input histograms rather than going through an event-by-event reweighting. The gluon fusion NLO samples used are provided centrally by CMS and they are reported in tables 5.11, 5.12 and 5.13, marked as "GGF NLO".

5.6.3 LO Vector Boson Fusion modelling

A similar procedure as described in 5.6.2 can be applied to the Vector Boson Fusion modelling. The HH production cross section via VBF can be written as the square of the amplitude of the LO diagrams represented in 5.2, i.e.

$$\begin{aligned}\sigma(c_{\mathrm{V}}, c_{2\mathrm{V}}, k_\lambda) \sim |\mathrm{A}c_{\mathrm{V}}k_\lambda + \mathrm{B}c_{\mathrm{V}}^2 + \mathrm{C}c_{2\mathrm{V}}|^2 = \\ = \mathrm{a}c_{\mathrm{V}}^2k_\lambda^2 + \mathrm{b}c_{\mathrm{V}}^4 + \mathrm{c}c_{2\mathrm{V}}^2 + \mathrm{i_{ab}}c_{\mathrm{V}}^3k_\lambda + \mathrm{i_{ac}}c_{\mathrm{V}}c_{2\mathrm{V}}k_\lambda + \mathrm{i_{bc}}c_{\mathrm{V}}^2c_{2\mathrm{V}}\end{aligned} \tag{5.46}$$

where $\mathrm{a} = |\mathrm{A}|^2$, $\mathrm{b} = |\mathrm{B}|^2$, $\mathrm{c} = |\mathrm{C}|^2$ and i_{ij} are the interference terms. Therefore, the cross section, as well as any differential distribution $\mathrm{d}\sigma/\mathrm{d}x$, depends on six components.

Thus, the VBF signal can be modelled through the sum of six components $\mathbf{V} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{i_{ab}}, \mathrm{i_{ac}}, \mathrm{i_{bc}}\}$, each scaled by a function of c_{V}, $c_{2\mathrm{V}}$ and k_λ; denoting as

$$\mathbf{K} = \{c_{\mathrm{V}}^2k_\lambda^2, c_{2\mathrm{V}}^4, c_{2\mathrm{V}}^2, c_{\mathrm{V}}^3k_\lambda, c_{\mathrm{V}}c_{2\mathrm{V}}k_\lambda, c_{\mathrm{V}}^2c_{2\mathrm{V}}\} \tag{5.47}$$

the vector of the functions of the couplings, the (5.46) can be expressed as

$$\sigma = \mathbf{K}^{\mathrm{T}}\mathbf{V}. \tag{5.48}$$

However, the generator used does not allow for generating the individual $\mathbf{V}_i$ components. Instead, they can be determined by solving a system of equations using six samples corresponding to different combinations of $(c_{\mathrm{V}}, c_{2\mathrm{V}}, k_\lambda)$. Denoting these samples as $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6\}$, where $\sigma_i = \sigma(c_{\mathrm{V},i}, c_{2\mathrm{V},i}, k_{\lambda,i})$, they can be represented as

$$\boldsymbol{\sigma} = \mathbf{M}\ \mathbf{V} \tag{5.49}$$

where $\mathbf{M}$ is the 6×6 coefficients matrix; its solution is

$$\mathbf{V} = \mathbf{M}^{-1}\boldsymbol{\sigma}. \tag{5.50}$$

Thus, the cross section σ_{target} of a given $(c_{\mathrm{V}}, c_{2\mathrm{V}}, k_\lambda)$ combination can be computed as

$$\sigma_{target} = [\mathbf{K}^{\mathrm{T}}\mathbf{M}^{-1}]\boldsymbol{\sigma}. \tag{5.51}$$

The (5.51) can be equally applied to build the differential distribution as a function of a given observable and for a given $(c_{\mathrm{V}}, c_{2\mathrm{V}}, k_\lambda)$ combination; in that case, the unknowns $\mathbf{V}(x)$ are a function of the observable x and

$$h(x)_{target} = [\mathbf{K}^{\mathrm{T}}\mathbf{M}^{-1}]\mathbf{h}(x), \tag{5.52}$$

where $\mathbf{h}(x)$ contains the differential distributions. Thus, the shape of a signal can be easily obtained by manipulating a few input histograms, rather than going through an event-by-event reweighting; to do so, only six fully simulated combinations of $(c_{\mathrm{V}}, c_{2\mathrm{V}}, k_\lambda)$ are needed. The six samples used are provided centrally by CMS and they are reported in tables 5.11, 5.12 and 5.13, marked as "VBF LO".

5.7 Systematic Uncertainties

Residual differences between data and MC simulation can be attributed to uncertainties on theoretical predictions, unforeseen detector responses as well as on statistical uncertainties affecting the data-driven methods. In order to properly account for all these effects they are included in the final fit as systematic uncertainties.

	Data set Name
GGF LO	/GluGluToHHTo2B2Tau_node_box_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_SM_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_2_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_9_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_10_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_11_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_12_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_13_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
GGF NLO	/GluGluToHHTo2B2Tau_node_cHHH5_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH2p45_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH1_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH0_TuneCUETP8M1_PSWeights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
VBF LO	/VBFHHTo2B2Tau_CV_0_5_C2V_1_C3_1_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_5_C2V_1_C3_1_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_0_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_1_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_2_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_2_C3_1_13TeV-madgraph/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM

Table 5.11: List of non-resonant signal samples MC datasets for 2016.

	Data set Name
GGF LO	/GluGluToHHTo2B2Tau_node_SM_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_2_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_3_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_4_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_7_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_9_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_12_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
GGF NLO	/GluGluToHHTo2B2Tau_node_cHHH5_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH2p45_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH1_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH0_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
VBF LO	/VBFHHTo2B2Tau_CV_1_5_C2V_1_C3_1_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_0_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_1_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_2_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_2_C3_1_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_0_5_C2V_1_C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v3/MINIAODSIM

Table 5.12: List of non-resonant signal samples MC datasets for 2017.

	Data set Name
GGF LO	/GluGluToHHTo2B2Tau_node_SM_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_2_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_3_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_4_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIMM
	/GluGluToHHTo2B2Tau_node_5_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_6_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_7_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_8_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_9_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_10_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_11_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_12_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
GGF NLO	/GluGluToHHTo2B2Tau_node_cHHH5_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH2p45_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH1_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/GluGluToHHTo2B2Tau_node_cHHH0_TuneCP5_PSWeights_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
VBF LO	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_0_5_C2V_1_C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_5_C2V_1_C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_0_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_1_C3_2_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	/VBFHHTo2B2Tau_CV_1_C2V_2_C3_1_TuneCP5_PSWeights_13TeV-madgraph-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM

Table 5.13: List of non-resonant signal samples MC datasets for 2018.

The systematic uncertainties affecting only the yield of a given process (either signal or background) are detailed in Section 5.7.1.

The systematic uncertainties affecting the shape of the final discriminating variable are detailed in Section 5.7.2. Alternative shapes of the discriminating variable for the simulated processes are computed by varying the scale of the objects affected by each uncertainty.

5.7.1 Normalisation Uncertainties

Luminosity

The uncertainty on the luminosity measurement is estimated by the CMS LumiPOG. Different sources of uncertainties on the luminosity measurement are considered and their correlations are reported in Table 5.14.

Uncertainty	2016	2017	2018
2016 Specific	2.2	-	-
2017 Specific	-	2.0	-
2018 Specific	-	-	1.5
X-Y factorization	0.9	0.8	2.0
Length scale	-	0.3	0.2
Beam-beam deflection	0.4	0.4	-
Dynamic beta	0.5	0.5	-
Beam current calibration	-	0.3	0.2
Ghosts and satellites	0.4	0.1	-

Table 5.14: Year-to-year correlations between uncertainties on the luminosity measurement.

These values are obtained from dedicated Van-der-Meer scans during the data taking. In these scans, the beam separation is varied over time and the resulting rate of some physical observables (e.g., number of charged hadrons passing through a silicon detector or energy deposited in a calorimeter) as a function of separation can be used to extract the effective beam size. The absolute luminosity at one point in time can then be calculated from the measurable beam parameters.

These uncertainties are applied to the signals and to all the background processes estimated only from Monte Carlo simulation. The normalizations of the QCD and Z+jets backgrounds are obtained from data and hence not subject to the luminosity uncertainties.

Electrons and muons isolation and identification efficiencies

The uncertainties on electron and muon identification and isolation efficiencies are determined from the uncertainties on the Monte Carlo-to-data scale-factors. A value of 1% for both electrons and muons is obtained.

L1 Prefiring

An uncertainty of 2% that modifies the overall yield of each MC process is associated to the L1 prefiring corrections applied to 2016 and 2017 Monte Carlo simulations, as

discussed in Section 5.3.8.

PU reweighting uncertainty

An uncertainty on the PU reweighting technique is estimated applying to each MC event the up and down variations of the PU weights computed as follows:

- the distribution of the number of PU interactions in the data is recomputed varying the total cross section of proton-proton interactions by ±4.7%
- new set of weights are computed performing a bin-by-bin division between the PU distribution in the MonteCarlo samples (in this analysis $t\bar{t}$ samples are used) and the aforementioned data distributions

The uncertainty is estimated to be negligible in the $\tau_h\tau_h$ final state, while it has a value of 1% in the $\mu\tau_h$ and $e\tau_h$ channels.

Drell-Yan Scale Factors uncertainties

The normalization of the Drell-Yan Z+jets background is corrected using a fit to 18 different control regions in data, as described in Section 5.5.2. The uncertainties on the SF obtained in these control regions are propagated to the signal regions taking into account their correlation.

QCD estimation uncertainty

The QCD background, which is estimated from data using the ABCD method as described in Section 5.5.1, is potentially affected by different sources of systematic uncertainty. Assuming that the contribution of the QCD background is constant in all the 4 regions (A, B, C and D) one can define three uncertainties:

- **Shape uncertainty:** since the shape of the QCD contribution is estimated from region C and normalized to the yield of $C \times B/D$ in the main analysis, a second histogram can be built using the shape obtained from B. This second histogram is then used to define the shape uncertainty on the QCD process.
- **Uncertainty on yield correction factor:** the statistical uncertainty on the correction factor B/D (as described in Section 5.5.1) is a normalization uncertainty defined as the sum in quadrature of the statistical uncertainties on the event yields in regions B and D. The value of this uncertainty ranges from 5% to 100% depending on the category, channel and year considered.
- **Additional uncertainty:** this normalization uncertainty is added only in the few cases where the correction factor is not constant as function of the DeepTau WP used to define the anti-isolated QCD regions.

Theoretical HH cross section

The uncertainties on the theoretical cross sections of di-Higgs pairs production have the following values:

- **GGF:** $+2.2\%/-5.5\%$ (scale), $\pm 3\%$ (PDF + α_S) and $\pm 2.6\%$ (m_{top})

- **VBF:** $+0.03\%/-0.04\%$ (scale), $\pm 2.1\%$ (PDF + α_S)

These uncertainties are included only when computing cross section upper limit in the SM point and for the likelihood scans, where a k_λ-dependent uncertainty is also added.

Final state branching fraction

Two different normalization systematics are obtained propagating the theoretical uncertainties on the Higgs decays branching fractions. Assuming $m_H = 125$ GeV:

- $\mathcal{B}(H \to bb) = 0.5824 \ \pm 0.65\%(\text{theory}) \ ^{+0.72\%}_{-0.74\%}(m_q) \ ^{+0.78\%}_{-0.80\%}(\alpha_S)$
- $\mathcal{B}(H \to \tau\tau) = 0.06272 \ ^{+1.16\%}_{-1.17\%}(\text{theory}) \ ^{+0.98\%}_{-0.99\%}(m_q) \ \pm 0.62\%(\alpha_S)$

The total uncertainties are $+1.25/-1.27\%$ for the $H \to bb$ decay, and $\pm 1.65\%$ for the $H \to \tau\tau$ decay.

These uncertainties are included only when computing cross section upper limit in the SM point.

Cross sections

The uncertainties of the cross sections of the $t\bar{t}$, W+jets, single top, single Higgs, and multi-boson background processes are reported in Table 5.15.

Uncertainty	Value [%]
QCDscale_ZH	$+3.8/-3.1$
pdf_ZH	± 1.6
QCDscale_WH	$+0.5/-0.7$
pdf_WH	± 1.9
QCDscale_ttH	$+5.8/-9.2$
pdf_ttH	± 3.6
QCDscale_ggH	$+7.6/-8.1$
pdf_ggH	± 3.1
QCDscale_vbfH	$+0.4/-0.3$
pdf_vbfH	± 2.1
QCDscale_ttbar	$+4.8/-5.5$
QCDscale_singleT	$+4.2/-3.5$
QCDscale_tW	± 5.4
QCDscale_W	$+0.8/-0.4$
QCDscale_EWK	± 2
QCDscale_VV	± 10
QCDscale_VVV	± 10

Table 5.15: Theorical uncertainties on the background production cross sections.

5.7.2 Shape Uncertainties

In order to evaluate the effect of a systematic uncertainty on the shape of the discriminating variable it is necessary to recompute the shape after shifting each object affected by that uncertainty (*e.g.* when shifting taus also the MET in the event is shifted as well and the SVfit algorithm is re-runned to compute the $H \to \tau\tau$ mass).

A shape uncertainty is assigned to each bin of the distribution to account for the finite MonteCarlo samples statistics following the Barlow-Beeston approach described in [72].

τ_h Energy Scale

The uncertainty on the measurement of the energy of τ leptons decaying into hadrons can lead to a change in the distribution of the variables used to search for a signal contribution. An uncertainty on the energy scale of each τ_h candidate is considered as recommended by the Tau POG. The uncertainties are derived combining two different measurements: the low and high-p_T measurements in $Z \to \tau\tau$ and in $W^* \to \tau\nu$ events, respectively. The following scheme is applied:

- $p_T(\tau) \leq 34$ GeV : from low-p_T measurement
- $34 < p_T(\tau) \leq 170$ GeV : linearly interpolated between the uncertainties in the low and high p_T measurement
- $p_T(\tau) \geq 170$ GeV : from high-p_T measurement

Four different uncertainties are provided by the Tau POG in order to take into account the different tau decay modes considered in this analysis: $DM = 0,\ 1,\ 10$ and 11. When considering the uncertainty for a particular decay mode, the shift is applied only to the τ_h candidates that are reconstructed with that particular decay mode, all other τ_h candidates are left unchanged.

Electrons and muons faking taus Energy Scales

The uncertainty on the energy scale of electrons reconstructed as hadronic taus is provided by the Tau POG. As for the τ_h uncertainties, also for the $e \to \tau_h$ energy scale, two different uncertainties are implemented in order to take into account the decay modes 0 and 1.

The recommended uncertainty on the energy scale for muons reconstructed as hadronic taus is 1%, uncorrelated in decay mode.

Jet Energy Scale

Uncertainties on the jet energy scale (JES) are provided by the JetMET POG. A reduced set of 11 sources is used in this analysis, as defined by the JetMET POG itself. The sources used for the three years are:

- **2016:** 'Absolute', 'Absolute_2016', 'BBEC1', 'BBEC1_2016', 'EC2', 'EC2_2016', 'FlavorQCD', 'HF', 'HF_2016', 'RelativeBal', 'RelativeSample_2016'
- **2017:** 'Absolute', 'Absolute_2017', 'BBEC1', 'BBEC1_2017', 'EC2', 'EC2_2017', 'FlavorQCD', 'HF', 'HF_2017', 'RelativeBal', 'RelativeSample_2017'

- **2018:** 'Absolute', 'Absolute_2018', 'BBEC1', 'BBEC1_2018', 'EC2', 'EC2_2018', 'FlavorQCD', 'HF', 'HF_2018', 'RelativeBal', 'RelativeSample_2018'

Sources that have the same name across the years are treated as 100% correlated, while those labelled with the year are treated as uncorrelated. A 12^{th} uncertainty, labelled as *Total*, is also provided for reference and represents the convoultion of all other uncertainties in each year.

DeepTau scale factors

The uncertainties arising by the application of the tau ID scale factors do not modify the single objects (*e.g.* the tau_h candidates), but modify the overall weight of each event.

***vsJet* discriminator**

The uncertainties on the *vsJet* DeepTau scale factors are provided by the Tau POG and are treated as uncorrelated across tau p_T bins. Five different uncertainties are identified for the five p_T bins defined by the boundaries [20, 25, 30, 35, 40, infinity]. All five uncertainties are used in the $\mu\tau_h$ and $e\tau_h$ channels, while, since in the $\tau_h\tau_h$ channel both leptons are required to have a transverse momentum larger than 40 GeV, only the last uncertainty ($40\ GeV < p_T < \ Inf$) is used.

***vsEle* discriminator**

The uncertainties on the *vsEle* DeepTau scale factors are provided by the Tau POG and are treated as uncorrelated across tau η bins: two different uncertainties are identified, one for the barrel and one for the endcap. Both uncertainties are used in all the channels.

Trigger scale factors

Four different uncertainties are provided by the Tau POG, one for each decay mode of the hadronic taus: $DM\ = 0,\ 1,\ 10$ and 11 and they are applied to the hadronically decaying leg of the $\tau\tau$ pair. Two more trigger uncertainties are used to cover the cases where the tau lepton decays in an electron or in a muon.

PU jet ID scale factors

Uncertainties on the PU jet identification scale factors are provided by the JetMET POG as function of jet p_T and η and are applied on an event-by-event base.

Custom tauID scale factors

For the $\tau_h\tau_h$ channel in 2017, custom SF are applied to the events with real taus, as explained in Section 5.3.8. Four additional uncertainties are defined to take into account the errors provided by the fit used to derive these scale factors: since the fits are done independently for each decay mode of the tau, the four uncertainties are treated as uncorrelated.

5.8 Results

In order to compute expected results, i.e. not including the observed data, the asymptotic limit approach described in [73], has been adopted.

5.8.1 Statistical treatment

In order to asses the presence or absence of signal events in the selected finaldistributions, a statistical procedure is needed. In the context of Higgs searches, the ATLAS and CMS collaborations use a modified frequentist approach referred to as CL_s and defined in [74].

A binned maximum likelihood fit is performed on the discriminating variable described in Section 5.4.6. The expected event yield of the signal and the total background are denoted as s and b, respectively, and the signal normalization is scaled by a signal strength modifier μ (the signal treatment is extensively described in 5.6). Given that the variables used are binned, s and b are vectors containing the yield expectations in each bin of the distributions.

The systematic uncertainties, described in Section 5.7, are included in this model as nuisance parameters θ_i, collectively denoted as θ, that affect the expected event yield for both signal and background processes which can thus be written as $s(\theta)$ and $b(\theta)$.

The likelihood function can be written as

$$\mathcal{L}(n, \tilde{\theta}|\mu, \theta) = P(n|\mu s + b) \cdot p(\tilde{\theta}|\theta) \tag{5.53}$$

where P denotes the probability density function of the observation of n events in a particular bin given by the sum of signal μs and background b expected events. For binned distributions, P is the product of the Poisson distributions for every bin:

$$P(n|\mu s + b) = \prod_j \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \tag{5.54}$$

The second term on the right side of Equation 5.53 represents the knowledge about the values of the nuisance parameters: each term reflects the probability for the true value to be equal to θ_i, given the best estimate $\tilde{\theta}_i$ obtained from auxiliary measurements on control region events or directly from the MC simulation. Since all systematic uncertainties are assumed to be uncorrelated, the combined term is the product of the single uncertainties $p(\tilde{\theta}|\theta) = \prod_i p_i(\tilde{\theta}_i|\theta_i)$.

The functional form of $p_i(\tilde{\theta}_i|\theta_i)$ depends on the type of uncertainty described. Uncertainties which arise from independent measurements, such as luminosity or trigger efficiencies, are modeled with log-normal functions:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi} ln(k)} exp\left(-\frac{(ln(\theta/\tilde{\theta}))^2}{2(ln(k))^2}\right)\frac{1}{\theta} \tag{5.55}$$

where k is the parameter that defines the width of the log-normal distribution and thus represents the interval of possible variations of the observable.

Uncertainties on the template shapes are taken into account during the fit procedure, using the *Vertical Template Morphing* technique: for each quantity that affects the shape, multiple instances of the templates are produced from the simulated events by varying that quantity by $\pm 1\sigma$ and bin-by-bin interpolation is performed between them. A nuisance that represents the variation of such quantities from the nominal value, is added to the likelihood model.

In order to evaluate the expected results the observed data are replaced by a so-called "Asimov dataset" constructed in such a way that the event yield n is equal to $\mu s + b$, without statistical fluctuations, as described in [73].

Two hypotheses are tested for the signal plus background or background only cases, $H_{\mu s+b}$ and H_b respectively. To set an exclusion limit on the presence of a signal, one has to find the value of μ that allows to reject the $H_{\mu s+b}$ in favor of H_b.

The test statistic chosen to set the exclusion limit is the likelihood ratio:

$$q_\mu = \begin{cases} -2ln\frac{\mathcal{L}(data|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data|0,\hat{\theta}_0)} & \text{with } \hat{\mu} < 0 \\ -2ln\frac{\mathcal{L}(data|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu},\hat{\theta})} & \text{with } 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{with } \hat{\mu} > \mu \end{cases} \tag{5.56}$$

where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ given the signal strength modifier μ, while "data" refers to the set of event yields n_i observed in all bins of the observed variables. The pair of parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood defined in Equation 5.53. The lower constraint on $\hat{\mu}$ is dictated by physics (positive signal rate only), while the upper constraint is imposed by hand in order to guarantee a one-sided confidence interval. This definition of the test statistic implies that larger values of q_μ represent an increasing incompatibility between the data and the hypothsized value of μ.

Given a signal strength modifier μ, the observed value q_μ^{obs} is obtained through the test statistic using the observed data n. In order to quantify the degree of compatibility of the observed data to the signal plus background or background only hypotheses, the probabilities for q_μ to be equal or larger than q_μ^{obs} are computed:

$$\begin{aligned} CL_{s+b}(\mu) &= P(q_\mu \geq q_\mu^{obs}|H_{\mu s+b}) \\ CL_b(\mu) &= P(q_\mu \geq q_\mu^{obs}|H_b) \end{aligned} \tag{5.57}$$

Finally, CL_s is computed, for the given value of μ under test, by the ratio of these probabilities:

$$CL_s(\mu) = \frac{CL_{s+b}(\mu)}{CL_b(\mu)} \tag{5.58}$$

A signal of strength μ is said to be excluded at a Confidence Level $1-\alpha$ if $CL_s(\mu) \leq \alpha$. In the $bb\tau\tau$ search we adopted $\alpha = 0.05$ and varied the parameter μ until the condition $CL_s(\mu) \leq \alpha$ is met so that the exclusion limits are always quoted with a 95% Confidence Level. The value of μ thus obtained can be converted into a limit on $\sigma_{HH} \times \mathcal{B}(HH \to bb\tau\tau)$.

5.8.2 Examples of final discriminant distribution

Figures 5.33, 5.34 and 5.35 show the blinded distributions of the final discriminant for the resolved 1b category, in the different years and channels. A very good agreement between data and simulation is found in the unblinded, background dominated, regions.

5.8.3 Inclusive GGF+VBF HH production

95% CL upper limits on production cross section vs k_λ

The expected sensitivity for the inclusive (GGF + VBF) HH production cross section are reported here as function of k_λ at 95% Confidence Level. As previously described in Section 5.4.4 eight categories in three channels are fitted simultaneously in the three

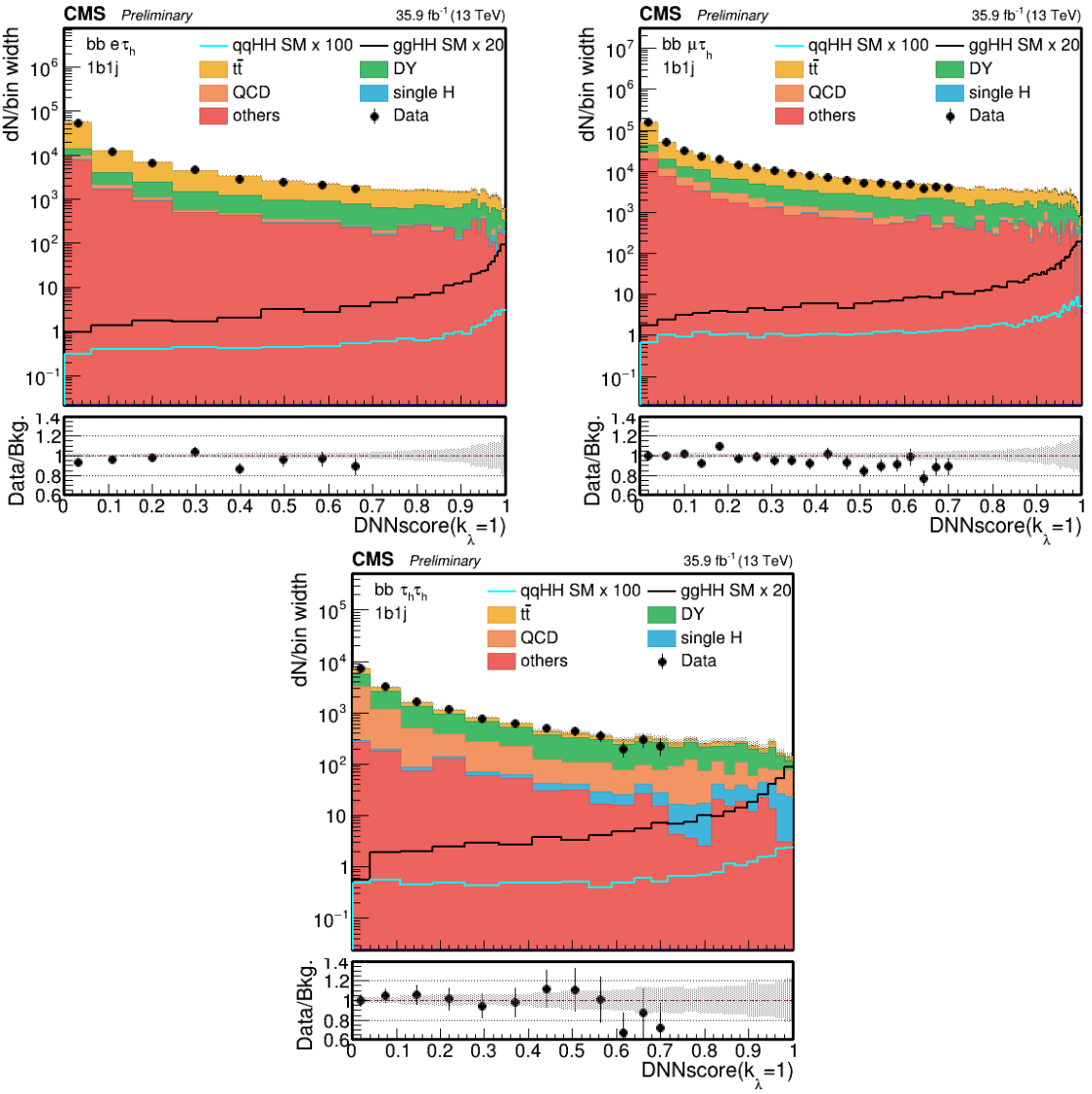


Figure 5.33: DNN score distributions for the resolved 1b category in 2016. The shaded band in the plots represents the statistical uncertainty only.

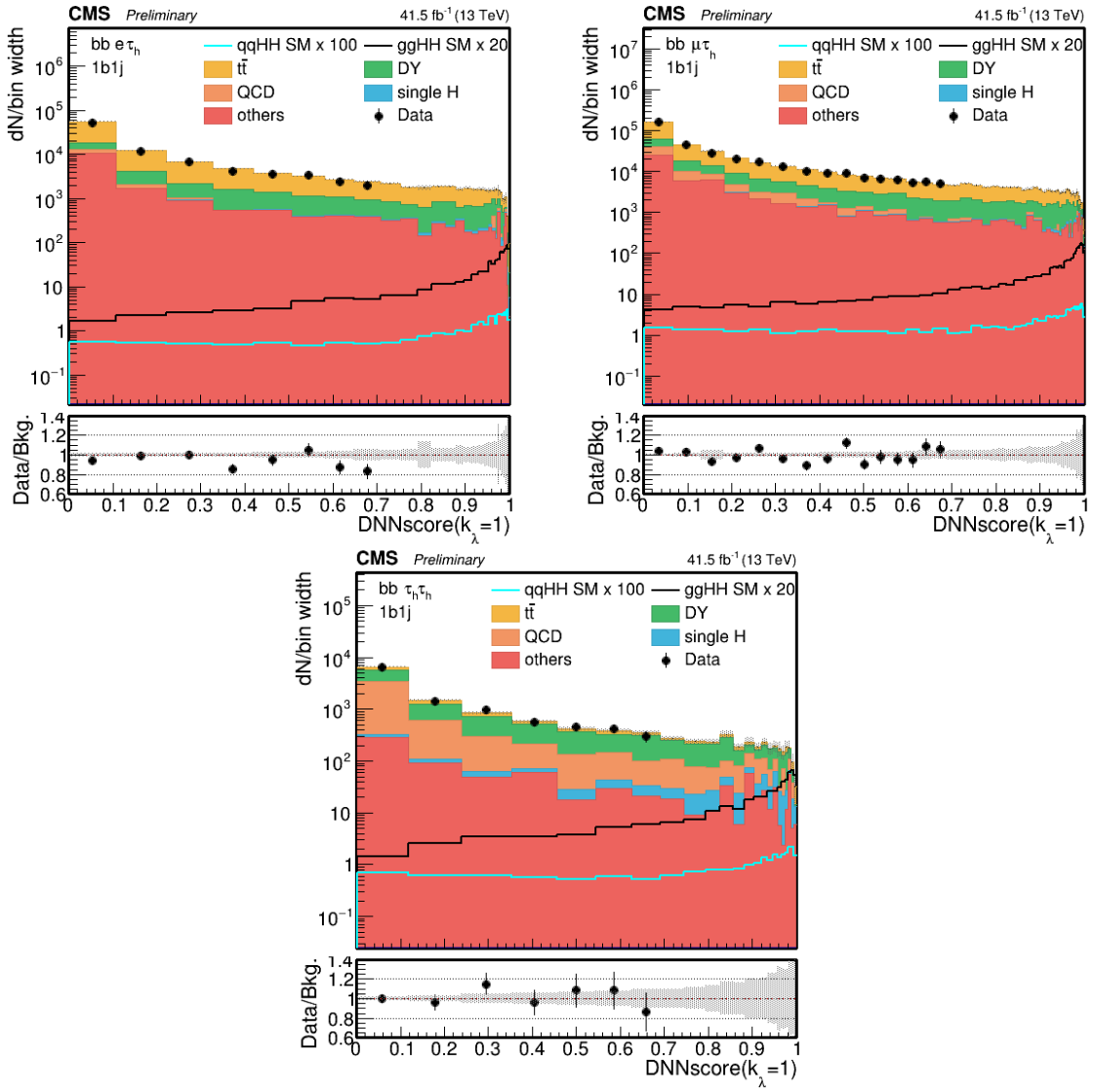


Figure 5.34: DNN score distributions for the resolved 1b category in 2017. The shaded band in the plots represents the statistical uncertainty only.

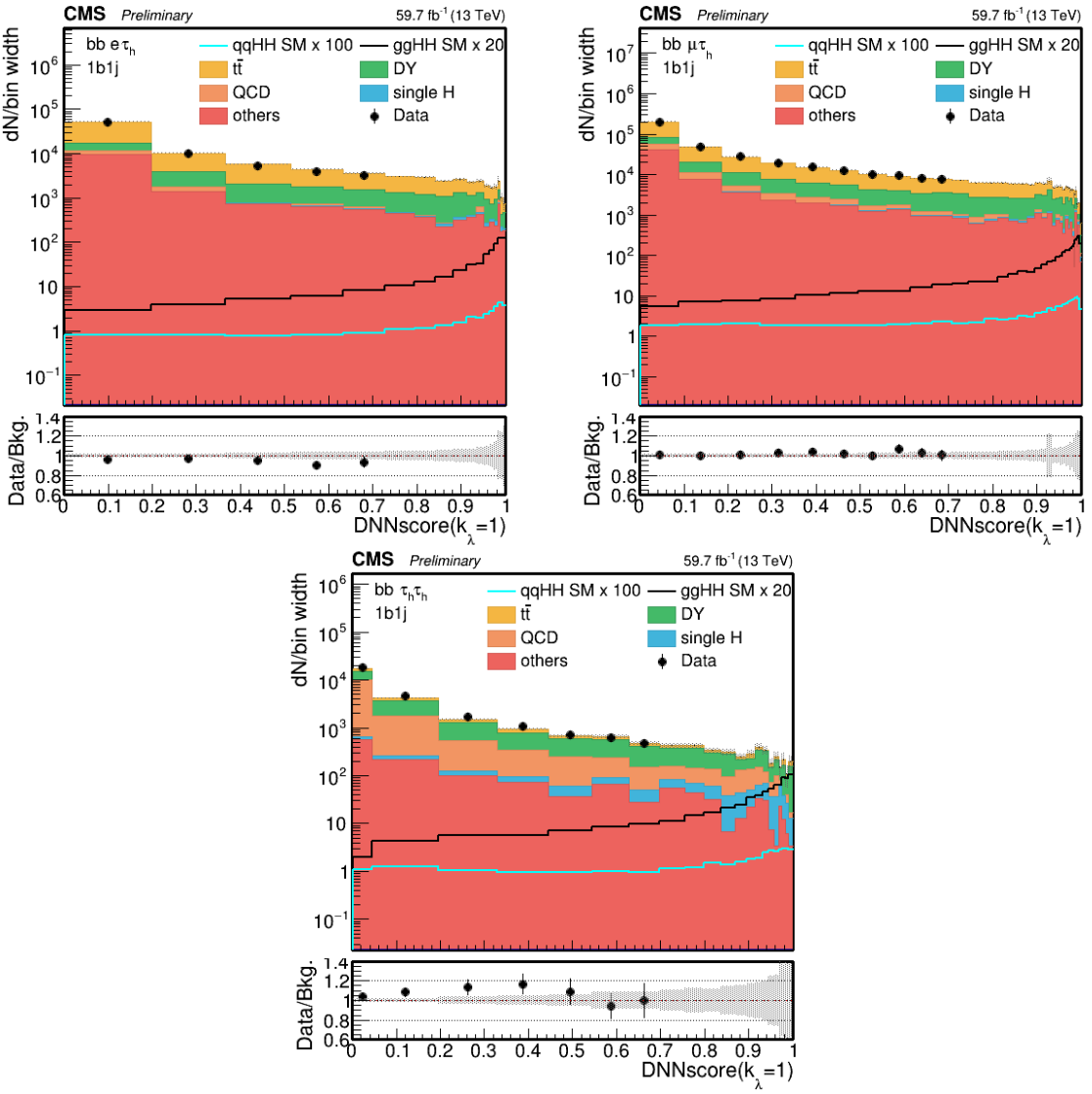


Figure 5.35: DNN score distributions for the resolved 1b category in 2018. The shaded band in the plots represents the statistical uncertainty only.

years to obtain the final sensitivity. The expected constraint of k_λ is $-2.5 < k_\lambda < 9.5$, as shown in Figure 5.37 for the three years, and in Figure 5.38 for the full Run 2 combination. The expected limits for the SM point ($k_\lambda = 1$) are listed in Table 5.16 (statistical uncertainty only) and Table 5.17 (statistical and systematic uncertainty). The results including systematic uncertainties are also summarized in Figure 5.36.

Expected limit	2016	2017	2018	Run 2
$\sigma_{ggF+VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)/\sigma^{SM}_{ggF+VBF}$	9.5	8.6	6.4	4.2
$\sigma_{ggF+VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)$ [fb]	22.7	20.5	15.3	10.0
$\sigma_{ggF+VBF}(\text{pp} \to \text{HH})$ [fb]	310.4	280.6	208.9	136.7

Table 5.16: 95% CL expected limits (without systematics) for the SM point (k_λ=1), where $\sigma^{SM}_{ggF+VBF} = (31.05 + 1.726) \cdot \mathcal{B}(HH \to bb\tau\tau) = 32.776$ fb $\cdot\, 0.073 = 2.39$ fb.

Expected limit	2016	2017	2018	Run 2
$\sigma_{ggF+VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)/\sigma^{SM}_{ggF+VBF}$	10.3	9.1	7.0	4.5
$\sigma_{ggF+VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)$ [fb]	24.5	21.8	16.7	10.7
$\sigma_{ggF+VBF}(\text{pp} \to \text{HH})$ [fb]	336.0	299.0	228.4	147.0

Table 5.17: 95% CL expected limits (including systematics) for the SM point (k_λ=1), where $\sigma^{SM}_{ggF+VBF} = (31.05 + 1.726) \cdot \mathcal{B}(HH \to bb\tau\tau) = 32.776$ fb $\cdot\, 0.073 = 2.39$ fb.

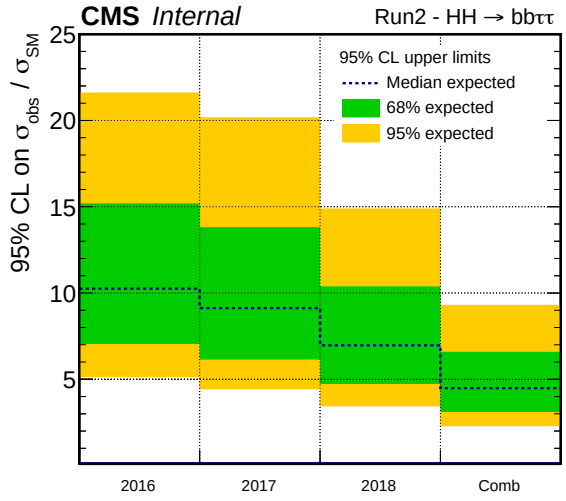


Figure 5.36: Visualization of the 95% CL expected limits, including the systematic uncertainties, for the SM point (k_λ=1), as reported in Table 5.17.

95% CL upper limits for different categories and channels

The expected sensitivity for the inclusive (GGF + VBF) HH production cross section are reported here as function of k_λ at 95% Confidence Level highlighting the contribution of different channels and categories. Results are reported for 2016, 2017 and 2018 separately in Figures 5.39, 5.40 and 5.41, respectively.

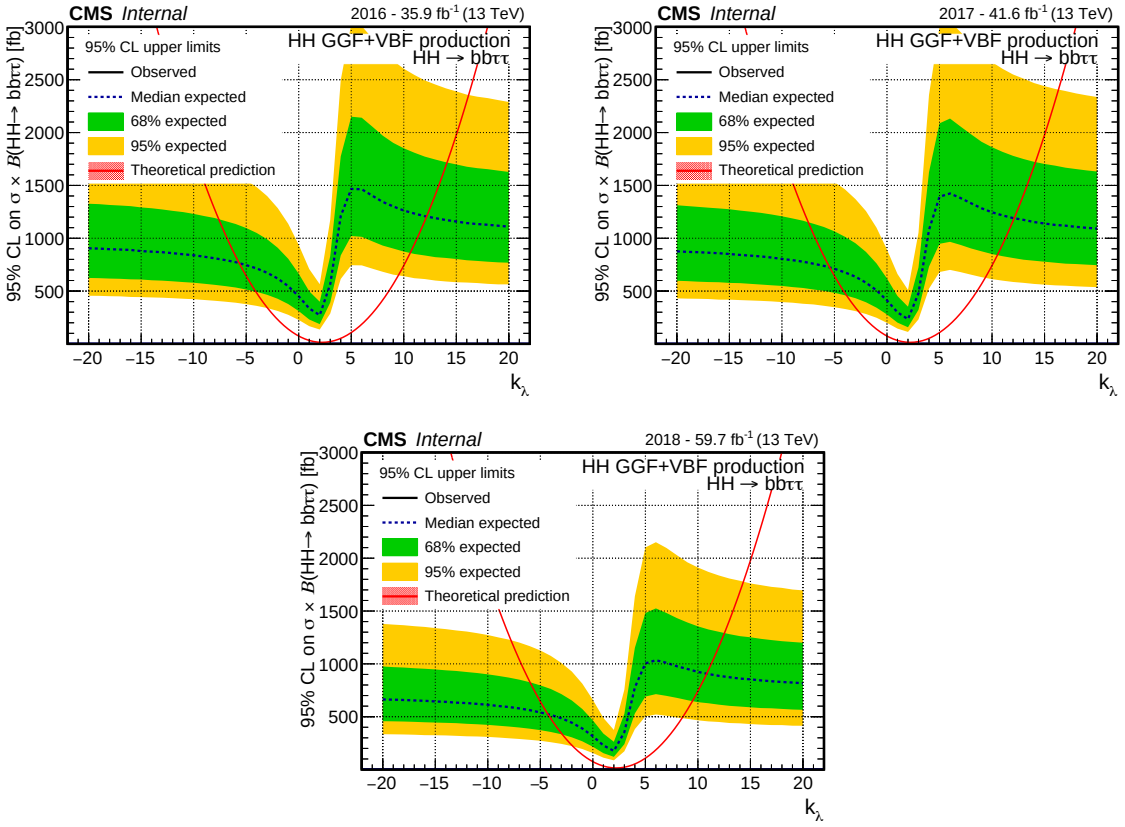


Figure 5.37: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the k_λ coupling. The plots correspond to the combination of the channels and categories within each year: 2016 (top left), 2017 (top right) and 2018 (bottom).

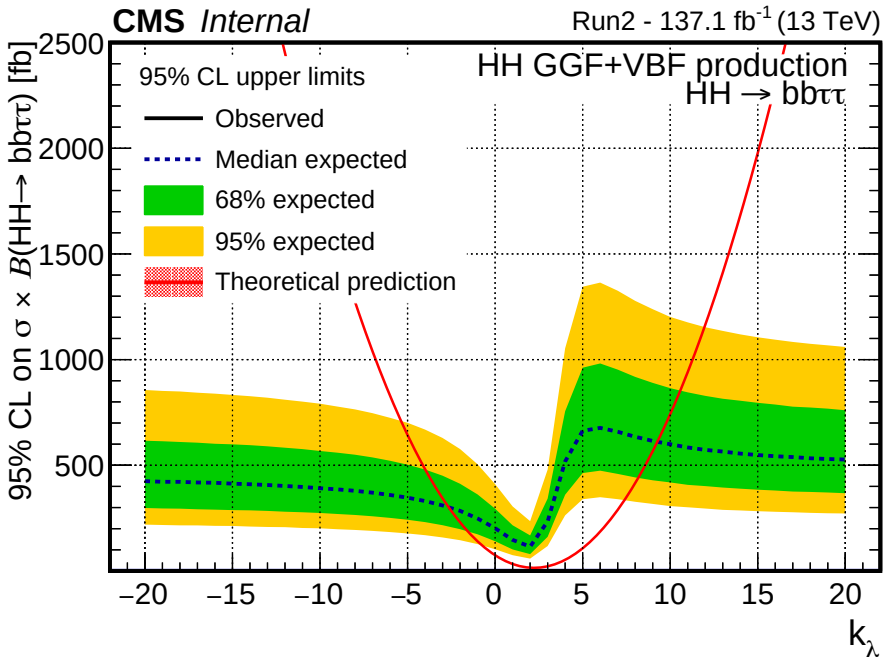


Figure 5.38: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the k_λ coupling. The plot corresponds to the combination of all the channels and categories for the full Run 2.

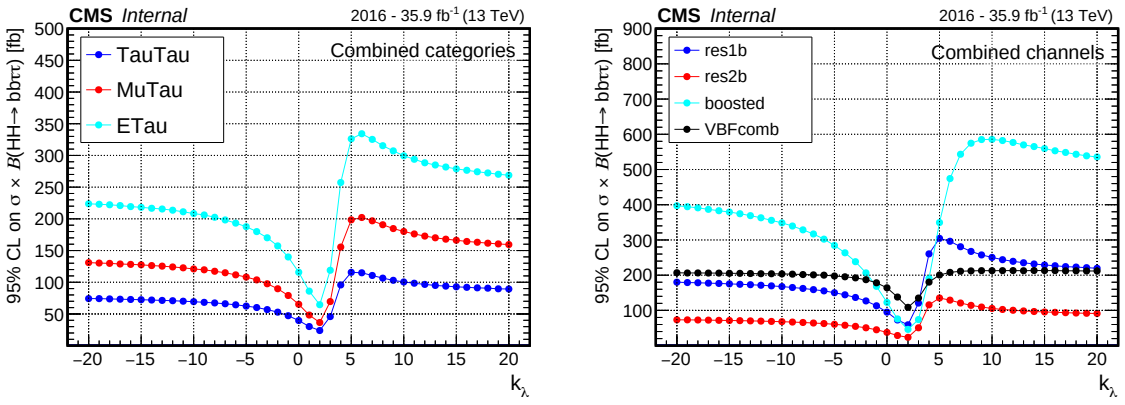


Figure 5.39: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the k_λ coupling. The plot on the left shows the result obtained in 2016 combining all the categories in different $\tau\tau$ decay channels. The plot on the right shows the result obtained in 2016 combining all three $\tau\tau$ decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

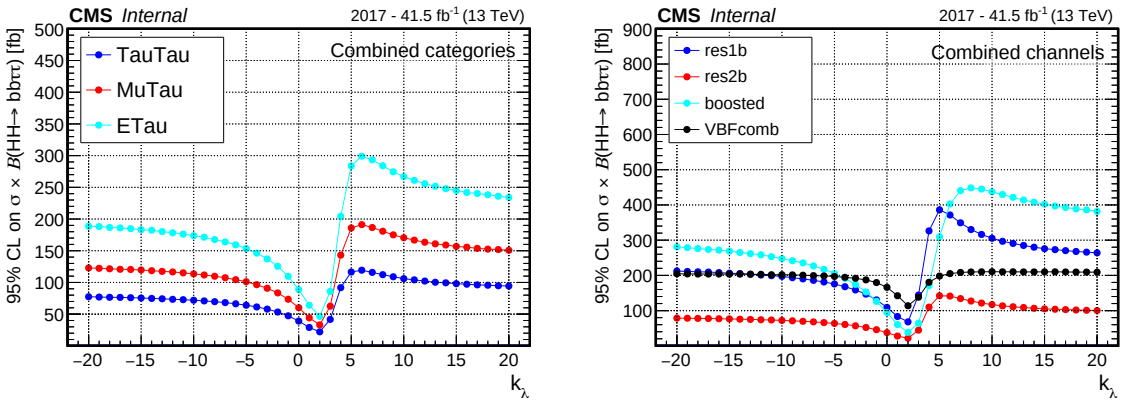


Figure 5.40: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the k_λ coupling. The plot on the left shows the result obtained in 2017 combining all the categories in different $\tau\tau$ decay channels. The plot on the right shows the result obtained in 2017 combining all three $\tau\tau$ decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

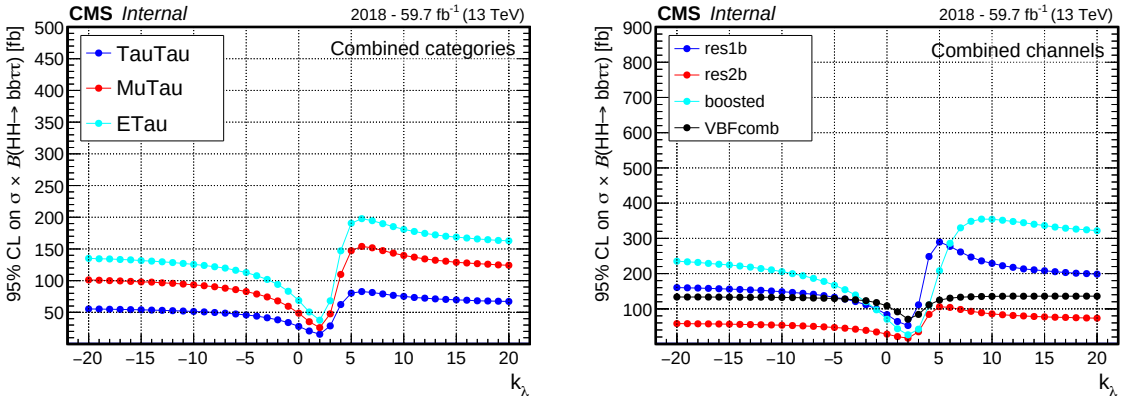


Figure 5.41: 95% CL expected exclusion of the GGF + VBF HH production cross section as a function of the k_λ coupling. The plot on the left shows the result obtained in 2018 combining all the categories in different $\tau\tau$ decay channels. The plot on the right shows the result obtained in 2018 combining all three $\tau\tau$ decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

Likelihood scans vs k_λ coupling

Assuming that VBF and GGF HH events are produced with the properties predicted by the SM, it is possible to measure the value of the k_λ coupling. The scan of the expected likelihood as a function of k_λ is shown in Figure 5.42 for each of the three years and for the full Run 2. The expected confidence interval on k_λ corresponds to [-1.6,8] at 68% CL and to [-3.5,10] at 95% CL.

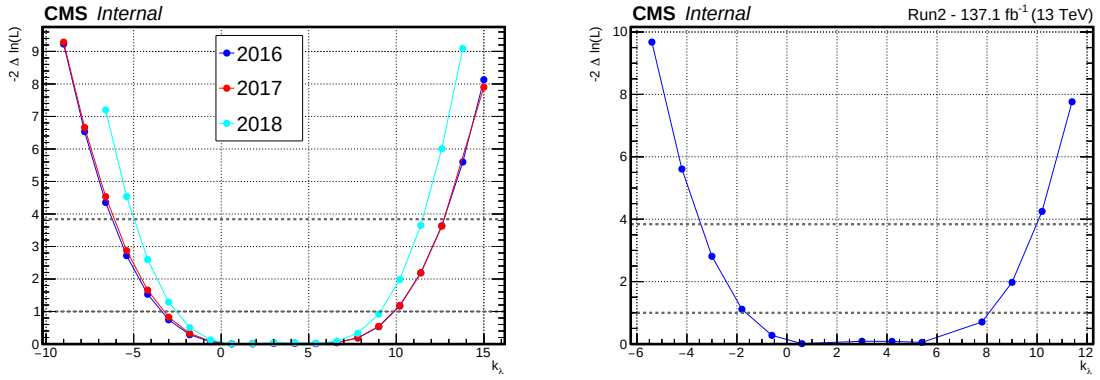


Figure 5.42: Expected likelihood scan as a function of k_λ. The plot on the left shows the result split for the three years, while the plot on the right represents the combination of the full Run 2.

5.8.4 VBF HH production

95% CL upper limits on VBF production cross section vs c_{2V}

The expected sensitivity for the inclusive VBF HH production cross section are reported here as function of c_{2V} at 95% Confidence Level. Also for VBF studies eight categories in three channels are fitted simultaneously in the three years to obtain the final sensitivity. The expected constraint of c_{2V} is $-0.5 < {}_{2V} < 2.6$, as shown in Figure 5.44 for the three years, and in Figure 5.45 for the full Run 2 combination. The expected limits for the SM point ($c_{2V} = 1$) are listed in Table 5.18 (statistical uncertainty only) and Table 5.19 (statistical and systematic uncertainty). The results including systematic uncertainties are also summarized in Figure 5.43.

Expected limit	2016	2017	2018	Run 2
$\sigma_{VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)/\sigma_{VBF}^{SM}$	314.0	327.5	186.0	130.5
$\sigma_{VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)$ [fb]	542.0	565.3	321.0	225.2
$\sigma_{VBF}(\text{pp} \to \text{HH})$ [fb]	39.6	41.3	23.4	16.4

Table 5.18: 95% CL expected limits (without systematics) for the SM point (c_{2V}=1), where $\sigma_{VBF}^{SM} = 1.726 \cdot \mathcal{B}(HH \to bb\tau\tau) = 1.726$ fb $\cdot\, 0.073 = 0.126$ fb.

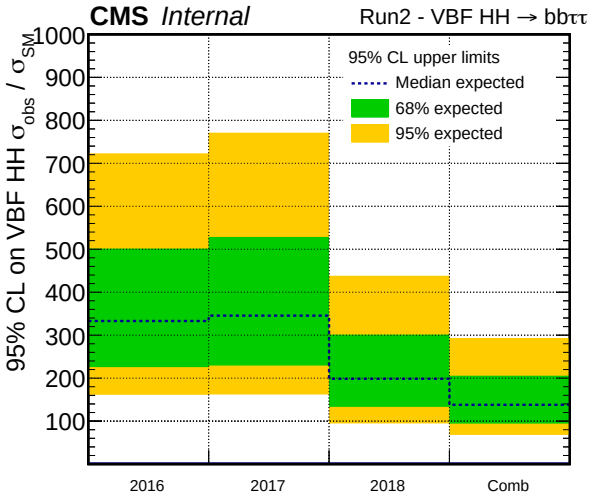


Figure 5.43: Visualization of the 95% CL expected limits, including the systematic uncertainties, for the SM point (c_{2V}=1), as reported in Table 5.19.

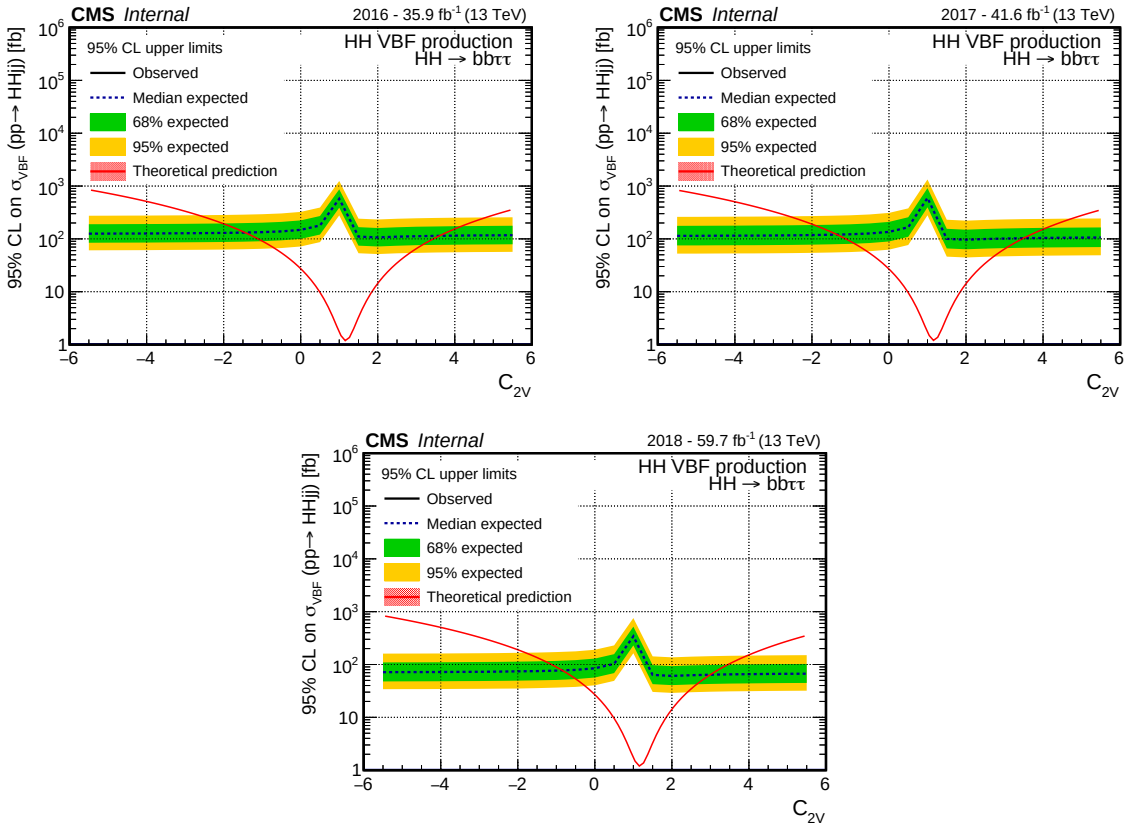


Figure 5.44: 95% CL expected exclusion of the BF HH production cross section as a function of the c_{2V} coupling. The plots correspond to the combination of the channels and categories within each year: 2016 (top left), 2017 (top right) and 2018 (bottom).

Expected limit	2016	2017	2018	Run 2
$\sigma_{VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)/\sigma_{VBF}^{SM}$	333.0	345.5	198.5	138.0
$\sigma_{VBF}(\text{pp} \to \text{HH} \to bb\tau\tau)$ [fb]	574.8	596.3	342.6	238.2
$\sigma_{VBF}(\text{pp} \to \text{HH})$ [fb]	42.0	43.5	25.0	17.4

Table 5.19: 95% CL expected limits (including systematics) for the SM point (c_{2V}=1), where $\sigma_{VBF}^{SM} = 1.726 \cdot \mathcal{B}(HH \to bb\tau\tau) = 1.726 \text{ fb} \cdot 0.073 = 0.126 \text{ fb}$.

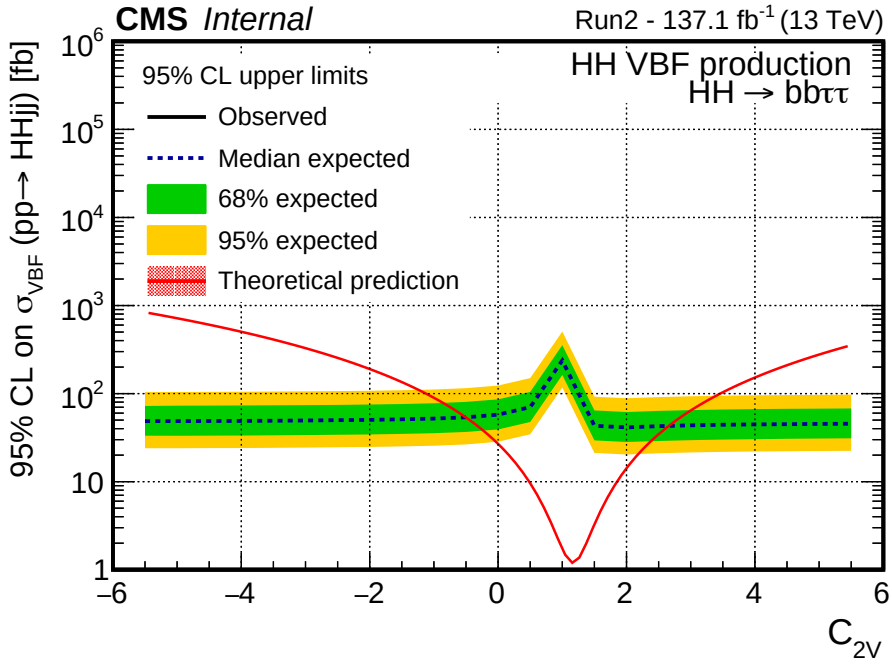


Figure 5.45: 95% CL expected exclusion of the VBF HH production cross section as a function of the c_{2V} coupling. The plot corresponds to the combination of all the channels and categories for the full Run 2.

95% CL upper limits on VBF production cross section for different categories and channels

the inclusive VBF HH production cross section are reported here as function of c_{2V} at 95% Confidence Level highlighting the contribution of different channels and categories. Results are reported for 2016, 2017 and 2018 separately in Figures 5.46, 5.47 and 5.48, respectively.

Likelihood scans vs c_{2V} coupling

Assuming that VBF HH events are produced with the properties predicted by the SM, it is possible to measure the value of the c_{2V} coupling. The scan of the expected likelihood as a function of c_{2V} is shown in Figure 5.49 for each of the three years and for the full Run 2. The expected confidence interval on c_{2V} corresponds to [0,2.15] at 68% CL and to [-0.5,2.55] at 95% CL.

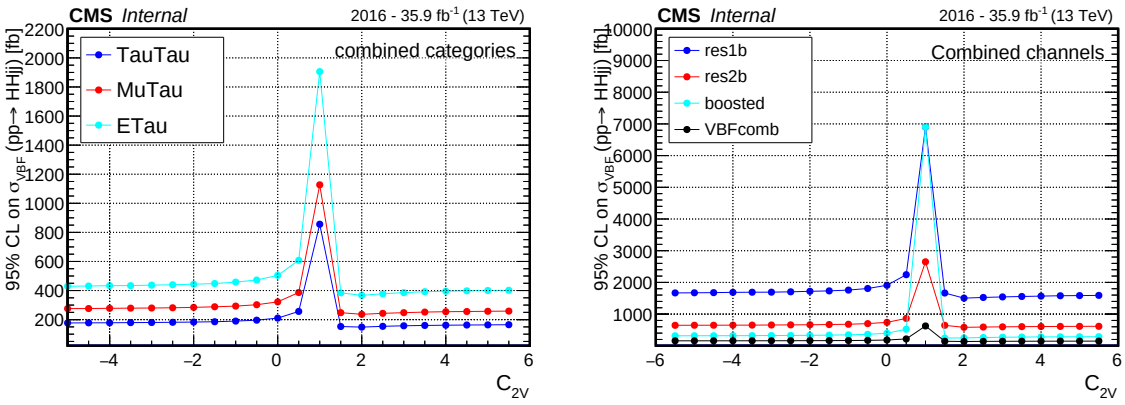


Figure 5.46: 95% CL expected exclusion of the VBF HH production cross section as a function of the c_{2V} coupling. The plot on the left shows the result obtained in 2016 combining all the categories in different $\tau\tau$ decay channels. The plot on the right shows the result obtained in 2016 combining all three $\tau\tau$ decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

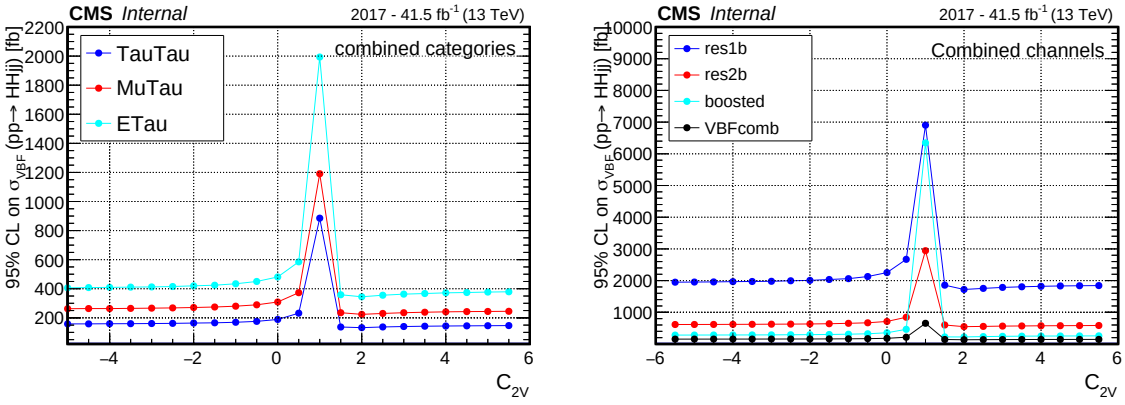


Figure 5.47: 95% CL expected exclusion of the VBF HH production cross section as a function of the c_{2V} coupling. The plot on the left shows the result obtained in 2017 combining all the categories in different $\tau\tau$ decay channels. The plot on the right shows the result obtained in 2017 combining all three $\tau\tau$ decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

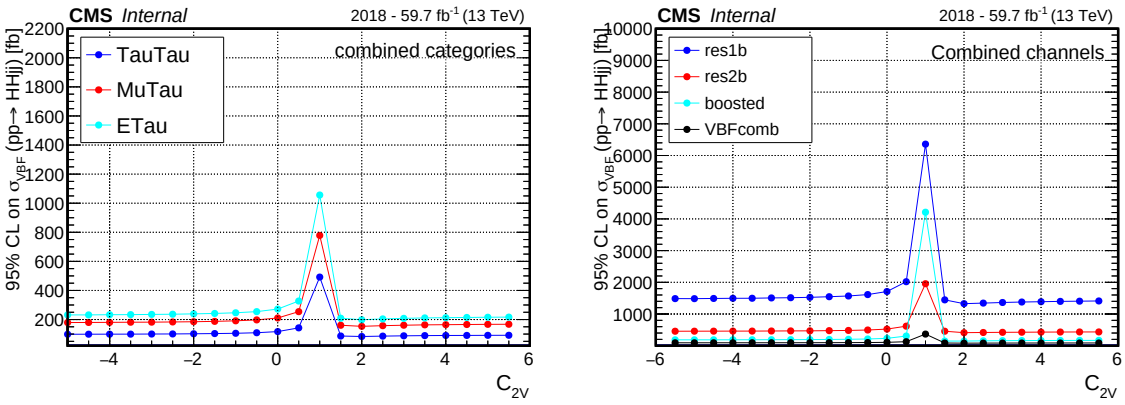


Figure 5.48: 95% CL expected exclusion of the VBF HH production cross section as a function of the c_{2V} coupling. The plot on the left shows the result obtained in 2018 combining all the categories in different $\tau\tau$ decay channels. The plot on the right shows the result obtained in 2018 combining all three $\tau\tau$ decay channels in different categories: res1b, res2b, boosted and the combination of the 5 multicategories in the VBF phase-space (in the plot denoted as *VBFcomb*).

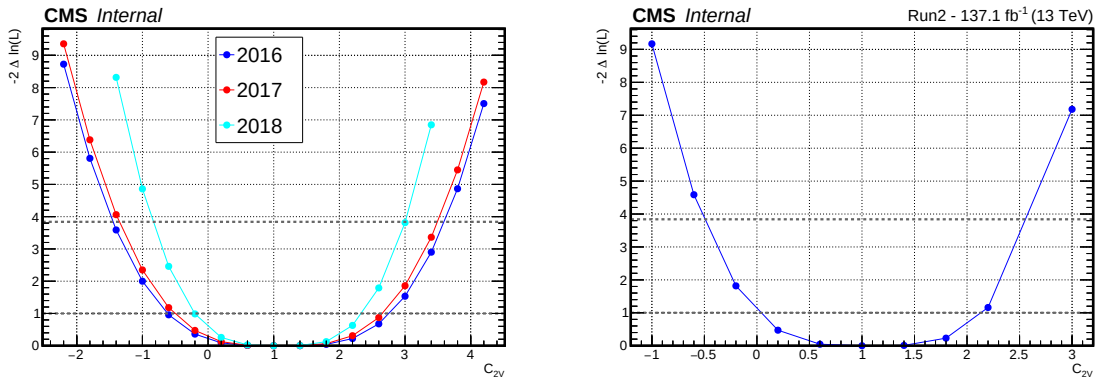


Figure 5.49: Expected likelihood scan as a function of c_{2V}. The plot on the left shows the result split for the three years, while the plot on the right represents the combination of the full Run 2.

Bibliography

[1] Yves Baconnier, Giorgio Brianti, P Lebrun, A G Mathewson, R Perin, and Yves Baconnier. *LHC: the Large Hadron Collider accelerator project*. CERN, Geneva, 1993.

[2] Thomas Sven Pettersson and P Lefèvre. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05-LHC, Oct 1995.

[3] Esma Mobs. The CERN accelerator complex. Complexe des accélérateurs du CERN. Jul 2016. General Photo.

[4] G Apollinari, I Béjar Alonso, O Brüning, M Lamont, and L Rossi. *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2015.

[5] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008.

[6] 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. *Nucl. Instrum. Meth. A*, 565:188–194, 2006.

[7] Q Ingram. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *Journal of Instrumentation*, 2(04):P04004–P04004, apr 2007.

[8] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:9009, 2013.

[9] S. Abdullin et al. The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c. *Eur. Phys. J. C*, 60:359–373, 2009. [Erratum: Eur.Phys.J.C 61, 353–356 (2009)].

[10] A.M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018.

[11] Andrey Starodumov, P. Berger, and M. Meinhard. High rate capability and radiation tolerance of the PROC600 readout chip for the CMS pixel detector. *JINST*, 12(01):C01078, 2017.

[12] A.M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[13] CMS Collaboration. The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical Report CERN-LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov 2017. Technical Design Report of the endcap calorimeter for the Phase-2 upgrade of the CMS experiment, in view of the HL-LHC run.

[14] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical Report CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC.

[15] T. T. Böhlen, F. Cerutti, M. P. W. Chin, A. Fassò, A. Ferrari, P. G. Ortega, A. Mairani, P. R. Sala, G. Smirnov, and V. Vlachoudis. The FLUKA Code: Developments and Challenges for High Energy and Medical Applications. *Nucl. Data Sheets*, 120:211–214, 2014.

[16] Alfredo Ferrari, Paola R. Sala, Alberto Fasso, and Johannes Ranft. FLUKA: A multi-particle transport code (Program version 2005). 10 2005.

[17] Maurice Garcia-Sciveres. The RD53A Integrated Circuit. Technical Report CERN-RD53-PUB-17-001, CERN, Geneva, Oct 2017.

[18] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[19] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[20] Vardan Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{\text{s}} = 8$ TeV. *JINST*, 10(06):P06005, 2015.

[21] A.M. Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):P10005, 2018.

[22] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[23] Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[24] Albert M Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *JINST*, 14(07):P07004, 2019.

[25] S. Ramo. Currents induced by electron motion. *Proceedings of the IRE*, 27(9):584–585, 1939.

[26] W. Shockley. Currents to conductors induced by a moving point charge. *Journal of Applied Physics*, 9(10):635–636, 1938.

[27] Julian Becker. Signal Development in silicon sensors used for radiation detection. Other thesis, 8 2010.

[28] Michael Moll. Radiation damage in silicon particle detectors: microscopic defects and macroscopic properties, 1999. Presented on 30 Nov 1999.

[29] Angela Vasilescu. The niel scaling hypothesis applied to neutron spectra of irradiation facilities and in the atlas and cms sct. Technical report, CERN, 1999.

[30] G Kramberger, V Cindro, I Mandić, M Mikuž, and M Zavrtanik. Effective trapping time of electrons and holes in different silicon materials irradiated with neutrons, protons and pions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 481(1):297 – 305, 2002.

[31] S.I. Parker, C.J. Kenney, and J. Segal. 3d — a proposed new architecture for solid-state radiation detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 395(3):328 – 343, 1997. Proceedings of the Third International Workshop on Semiconductor Pixel Detectors for Particles and X-rays.

[32] Fabio Ravera. The CT-PPS tracking system with 3D pixel detectors. *JINST*, 11(11):C11027, 2016.

[33] Cinzia Da Via et al. 3D silicon sensors: Design, large area production and quality assurance for the ATLAS IBL pixel detector upgrade. *Nucl. Instrum. Meth. A*, 694:321–330, 2012.

[34] Marcos Turqueti, Ryan A. Rivera, Alan Prosser, Jeffry Andresen, John Chramowicz, and /Fermilab. Captan: A hardware architecture for integrated data acquisition, control, and analysis for detector development. In *2008 IEEE Nuclear Science Symposium Conference Record*, Nov 2008.

[35] Ryan A. Rivera, Marcos Turqueti, and Alan Prosser. A software solution for the control, acquisition, and storage of captan network topologies. In *2008 IEEE Nuclear Science Symposium Conference Record*, 2008.

[36] V. Re, M. Manghisoni, L. Ratti, J. Hoff, A. Mekkaoui, and R. Yarema. Fssr2, a self-triggered low noise readout chip for silicon strip detectors. *IEEE Transactions on Nuclear Science*, 53(4):2470–2476, 2006.

[37] Hendrik Jansen, Simon Spannagel, Jörg Behr, Antonio Bulgheroni, Gilles Claus, Emlyn Corrin, David Cussans, Jan Dreyling-Eschweiler, Doris Eckstein, Thomas Eichhorn, Mathieu Goffe, Ingrid Maria Gregor, Daniel Haas, Carsten Muhl, Hanno Perrey, Richard Peschke, Philipp Roloff, Igor Rubinskiy, and Marc Winter. Performance of the eudet-type beam telescopes. *EPJ Techniques and Instrumentation*, 3(1):7, 2016.

[38] C. Hu-Guo, J. Baudot, G. Bertolone, A. Besson, A.S. Brogna, C. Colledani, G. Claus, R. [De Masi], Y. Degerli, A. Dorokhov, G. Doziere, W. Dulinski, X. Fang, M. Gelin, M. Goffe, F. Guilloux, A. Himmi, K. Jaaskelainen, M. Koziel, F. Morel, F. Orsini, M. Specht, Q. Sun, O. Torheim, I. Valin, and M. Winter. First reticule size maps

with digital output and integrated zero suppression for the eudet-jra1 beam telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 623(1):480 – 482, 2010. 1st International Conference on Technology and Instrumentation in Particle Physics.

[39] R. Diener, J. Dreyling-Eschweiler, H. Ehrlichmann, I.M. Gregor, U. Kötz, U. Krämer, N. Meyners, N. Potylitsina-Kube, A. Schütz, P. Schütze, and M. Stanitzki. The desy ii test beam facility. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 922:265 – 286, 2019.

[40] C. DaVia and S.J. Watts. The geometrical dependence of radiation hardness in planar and 3d silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 603(3):319 – 324, 2009.

[41] H.C. Kästli. Frontend electronics development for the cms pixel detector upgrade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 731:88 – 91, 2013. PIXEL 2012.

[42] G.-F. [Dalla Betta], M. Boscardin, G. Darbo, R. Mendicino, M. Meschini, A. Messineo, S. Ronchin, D.M.S. Sultan, and N. Zorzi. Development of a new generation of 3d pixel sensors for hl-lhc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 824:386 – 387, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[43] Gian-Franco Dalla Betta, Roberto Mendicino, Dms Sultan, Maurizio Boscardin, Gabriele Giacomini, Sabina Ronchin, Nicola Zorzi, Giovanni Darbo, Marco Meschini, and Alberto Messineo. Small pitch 3D devices. *PoS*, Vertex2016:028, 2017.

[44] Hans Bichsel. Straggling in thin silicon detectors. *Rev. Mod. Phys.*, 60:663–699, Jul 1988.

[45] Jan Hendrik Hoß. *Search for Supersymmetry with Multiple Charged Leptons at sqrts = 13 TeV with CMS and Radiation Tolerance of the Readout Chip for the Phase I Upgrade of the Pixel Detector*. PhD thesis, ETH Zurich, Zurich, 2017.

[46] M. Boscardin et al. Performance of new radiation-tolerant thin planar and 3D columnar n^+ on p silicon pixel sensors up to a maximum fluence of $\sim 5 \times 10^{15}$ n_{eq} /cm^2. *Nucl. Instrum. Meth. A*, 953:163222, 2020.

[47] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[48] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[49] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[50] J. Goldstone. Field theories with superconductor solutions. *Il Nuovo Cimento (1955-1965)*, 19:154–164, 1961.

[51] G.'t Hooft. Renormalizable lagrangians for massive yang-mills fields. *Nuclear Physics B*, 35(1):167 – 188, 1971.

[52] S. Amoroso, P. Azzurri, J. Bendavid, E. Bothmann, D. Britzger, H. Brooks, A. Buckley, M. Calvetti, X. Chen, M. Chiesa, L. Cieri, V. Ciulli, J. Cruz-Martinez, A. Cueto, A. Denner, S. Dittmaier, M. Donegà, M. Dührssen-Debling, I. Fabre, S. Ferrario-Ravasio, D. de Florian, S. Forte, P. Francavilla, T. Gehrmann, A. Gehrmann-De Ridder, L. Gellersen, E. W. N. Glover, P. Gras, C. Gwenlan, Y. Haddad, G. Heinrich, J. Hessler, T. J. Hobbs, M. Höfer, A. Huss, J. Huston, T. Ježo, S. P. Jones, S. Kallweit, M. Klasen, G. Knippen, A. Larkoski, M. LeBlanc, P. Loch, K. Long, D. Maître, S. Marzani, J. Mazzitelli, J. A. Mcfayden, E. Metodiev, J. K. L. Michel, M. Moreno Llácer, B. Nachman, P. Nadolsky, D. Napoletano, E. R. Nocera, C. Oleari, C. Pandini, M. Pellen, S. Pigazzini, J. Pires, S. Plätzer, S. Prestel, K. Rabbertz, E. Re, P. Richardson, F. Ringer, J. Rojo, J. Roloff, R. Röntsch, M. Schönherr, C. Schwan, F. Siegert, D. Soper, G. Soyez, M. Spira, M. R. Sutton, F. J. Tackmann, V. Theeuwes, S. L. Villani, J. Whitehead, H. T. Yang, and J. Zhou. Les houches 2019: Physics at tev colliders: Standard model working group report, 2020.

[53] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, Nov 1998.

[54] W. Erdmann. Offline primary vertex reconstruction with deterministic annealing clustering. *CMS Internal Note*, IN-2011/014, 2011.

[55] Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons. Oct 2019.

[56] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[57] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[58] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10:059, 2014.

[59] Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector. Nov 2018.

[60] Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016.

[61] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Reconstruction of the higgs mass in h to tautau events by dynamical likelihood techniques. *Journal of Physics: Conference Series*, 513(2):022035, 2014.

[62] CMS Collaboration. Search for pair production of Higgs bosons in the two tau leptons and two bottom quarks final state using proton-proton collisions at $\sqrt{s} = 13TeV$. CMS Physics Analysis Summary CMS-PAS-HIG-17-002, CERN, 2016.

[63] CMS Collaboration. Search for non-resonant higgs boson pair production in the $bb\tau\tau$ final state using 2016 data. CMS Physics Analysis Summary CMS-PAS-HIG-16-028, CERN, 2016.

[64] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1957.

[65] Prospects for HH measurements at the HL-LHC. Technical Report CMS-PAS-FTR-18-019, CERN, Geneva, 2018.

[66] Giles Chatham Strong. On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case. *Machine Learning: Science and Technology*, Jun 2020.

[67] Yadolah Dodge. *The Concise Encyclopedia of Statistics*, pages 502–505. Springer New York, New York, NY, 2008.

[68] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.

[69] Albert M Sirunyan et al. Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13TeV$. *Phys. Lett.*, B778:101–127, 2018.

[70] Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-17-030, CERN, Geneva, 2018.

[71] Alexandra Carvalho, Martino Dall'Osso, Tommaso Dorigo, Florian Goertz, Carlo A. Gottardo, and Mia Tosi. Higgs pair production: choosing benchmarks with cluster analysis. *Journal of High Energy Physics*, 2016(4):126, 2016.

[72] Roger J. Barlow and Christine Beeston. Fitting using finite Monte Carlo samples. *Comput. Phys. Commun.*, 77:219–228, 1993.

[73] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), 2011.

[74] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

Ringraziamenti

Prima di tutto voglio ringraziare i miei genitori. Mi ritengo un ragazzo fortunato: ho potuto studiare nella mia città, vivendo a casa mia, e questo offre notevoli vantaggi, specialmente quando si è giù di morale. Quando sono partito per il CERN, a Giugno del 2019, non potevo immaginare che mi sarei ritrovato lontano da casa per 4 mesi, battendo il record di 3 mesi ottenuto a Fermilab. Non è stato facile ritrovarsi dalla parte sbagliata delle Alpi durante una pandemia ma siamo sopravvissuti anche a questo. E siamo grati per questo.

Poi voglio ringraziare i miei colleghi più anziani: Dario, Luigi, Mauro e Sandra. Siete stati maestri insuperabili e compagni di avventura straordinari. Abbiamo condiviso tante giornate in U2 ma solo i più fortunati hanno potuto trascorrere con me intere giornate al test beam! Io riassumerei con "It was the best of the times, it was the worst of the times".

I rignraziamenti che seguono sono diretta conseguenza del primo assioma di Buffari: "Zuolo non conosce tutti ma tutti conoscono Zuolo".

Credo che ognuno di noi abbia un gruppo di amici che sente e vede più degli altri. Il mio si chiama "OttoSottoUnTetto". Il tetto è quello di Pagnano e gli altri sette sono Sara, Maurizio, Marcello, Claudio, Matteo, Alice e Jacopo. Abbiamo festeggiato insieme diciottesimi, lauree, Natali e Capodanni, i ricordi che ho con voi potrebbero riempire un'enciclopedia. Siete il mio punto di riferimento in una vita piena di trasferte e vi sono grato per questo.

Tutti abbiamo anche un gruppo con un nome banale ma pieno di significato. Il mio è "Serate" ed è un gruppo molto movimentato. Da una parte ci sono amici che conosco dal primo anno di superiori, dall'altro ci sono fidanzate che vengono e fidanzate che vanno. Passano gli anni ma il compleanno di Friso a Colico rimane una tradizione e sono felice di passare tante serate con voi.

Infine c'è una cosa che credo sia solo mia: la vita sociale al CERN. Ho inziato a frequentarlo nel 2018, facendo dei turni con reperibilità h24 7/7, per 6 settimane. Da Giugno 2019 a Maggio 2020 è stato la mia casa e i ragazzi che ho conosciuto sono stati la mia famiglia. Il problema è che io sono un animale un po' troppo sociale per cui nella strada dall'ingresso al 40 mi fermo ogni 10 metri a salutare qualcuno, stranendo un po' quelli che sono con me. Credo che per nominarli tutti impiegherei una mezza pagina per cui non la farò, loro sanno che ho un'ottima memoria. Vi sarò eternamente grato per tutte le esperienze che abbiamo condiviso, le gite, le fondute, le raclette, i crostoli, il cinefellow e chi più ne ha più ne metta: viva "Abbasso Geraldine".

In questi anni mi sono sempre stupito della reazione delle persone quando finivo di spiegare cosa faccio nella vita. Non ho mai avuto la senzione di fare qualcosa che richiedesse delle doti particolari, semplicemente di fare qualcosa che mi piace tanto. Non nascondo la speranza che la mia esperienza possa servire di esempio a qualcuno che vuole realizzare un sogno. Concludo ringraziando un amico che secondo me ha fatto sì qualcosa di davvero straordinario. Si chiama Salvatore Rossini ed è riuscito a diventare marito, padre, pallavolista di serie A e ingegnere. Ogni tanto mi chiedo come abbia fatto, non riesco a darmi una risposta ma sicuramente per me è un esempio.