SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Fisica "Giuseppe Occhialini"

Dottorato di Ricerca in / PhD program **Fisica e Astronomia** Ciclo / Cycle **XXXVIII**

Curriculum in (se presente / if it is) **Fisica subnucleare**

TITOLO TESI / THESIS TITLE

Investigation of the Higgs self-coupling in the bbττ final state with the CMS experiment at the LHC and preparation for high-luminosity with the addition of the MIP Timing Detector

Cognome / Surname **Palluotto** Nome / Name **Simona**

Matricola / Registration number **825723**

Tutore / Tutor: **prof. Alessio Ghezzi**

Cotutore / Co-tutor: **prof. Marco Toliman Lucchini**
(se presente / if there is one)

Supervisor: **dott.ssa Martina Malberti**
(se presente / if there is one)

Coordinatore / Coordinator: **prof. Stefano Ragazzi**

ANNO ACCADEMICO / ACADEMIC YEAR 2024/2025

Insanum quiddam esset,
et in se contrarium,
existimare ea, quae adhuc nunquam facta sunt, fieri posse,
nisi per modos adhuc nunquam tentatos.

— *Francis Bacon*

Summary

Using the data collected at the LHC, a great precision has been achieved in measuring many properties of the Higgs boson, including the production cross section and its couplings to particles. However, its role in electroweak symmetry breaking and the origin of the mass of fundamental particles has not yet been fully validated through a direct test of the Higgs potential. The measurements of Higgs bosons pair production provide direct access to the Higgs self-coupling λ, describing the shape of the Higgs potential, and therefore represent a crucial test of the Standard Model (SM). The SM HH production is a very rare process, currently under investigation at the LHC but not yet observed. The upcoming high-luminosity phase of the LHC (HL-LHC) is expected to deliver a dataset large enough to enable its observation, thereby allowing precise tests of the SM predictions for its value and potentially opening new avenues for BSM physics.

This thesis focuses on a dual approach to increase the sensitivity of searches for HH in the CMS experiment. On one side, it is dedicated to contributing to the searches of such processes using current LHC data collected at the CMS. On the other side, it addresses the development and characterization of a detector capable of precisely measuring the production time of particles. This precise timing information will, for the first time, be integrated into event reconstruction, enabling the mitigation of pile-up effects and improving the reconstruction of complex final states in the high-vertex-density environment expected at the HL-LHC.

In the search for HH production using CMS data, this thesis focuses on the final state consisting of two b-quarks and two τ-leptons. This channel is among the most sensitive ones, offering an optimal balance between a relatively high branching fraction and reduced background contamination. My work focuses on the study of $H \to bb$ events where the Higgs boson is produced with a significant Lorentz boost, referred to as the boosted topology. The study of the boosted topology and its impact on the overall sensitivity was first conducted on Run 2 data within the context of the search for resonant production of $HH \to bb\tau\tau$. Expected upper limits at 95% CL on the production cross section for resonant HH are determined for two spin hypotheses of resonances produced via gluon fusion, covering a mass range from 250 GeV to 3 TeV. The boosted category provides the largest sensitivity for resonance masses above 1.5 TeV, demonstrating the benefit of including the boosted phase space. During the development of the non-resonant HH search in the same final state, using the first Run 3 data (2022–2023) corresponding to an integrated luminosity of 62.4 fb^{-1}, the boosted category was fully optimized and integrated into the analysis. The expected upper limit at 95% CL on non-resonant HH production is found to be 5.7 times the SM prediction. This corresponds approximately to the same sensitivity achieved in the same analysis using Run 2 data (corresponding to 138 fb^{-1}), despite using only about half of the integrated luminosity. This improvement can be attributed to several significant updates introduced with Run 3, such as new trigger strategies and advanced tagging algorithms. In particular, for the boosted category, the improvement with respect to Run 2, when normalized to the same integrated luminosity, amounts to approximately 15% in sensitivity. The Higgs self-coupling modifier derived from this study is constrained to the range $[-3.75, 10.5]$ at 95% CL, while the $HHVV$ coupling modifier is constrained to $[-0.75, 2.95]$, both results being consistent with those obtained in Run 2.

Despite the advances achieved in increasing the sensitivity of di-Higgs searches across LHC Runs, the observation of this process remains elusive, requiring larger statistics to probe such a rare phenomenon. Current projections for the HL-LHC indicate that an observation could be achieved thanks to a tenfold increase in integrated luminosity. However, delivering this amount of data comes at the cost of a very challenging data-taking environment: detectors will need to withstand extreme radiation doses, and event

reconstruction will have to identify interesting hard-scattering events among a very high density of spurious interactions, i.e. pileup events. To mitigate the pileup-induced degradation in vertex reconstruction quality, which in turns affects the quality of reconstructed physics objects, a completely new concept will be introduced into the reconstruction process: the time dimension. Among the various benefits, the inclusion of precise timing information is expected to improve the significance of HH analyses by an amount equivalent to approximately two to three additional years of data-taking. Achieving this impact on analyses requires extremely demanding detector performance. To profit of timing information, the detector must provide time measurements with an accuracy of 30-60 ps during operation. The MIP Timing Detector (MTD) has been designed to meet this performance target in both the barrel and endcap regions of CMS, employing different technologies in the two subsystems.

This thesis presents the R&D studies that led to the definition of the final design of the barrel part of the MTD, the Barrel Timing Layer (BTL). It includes a detailed and systematic investigation of the factors affecting the overall time resolution, as well as the development of procedures for large-scale production, quality control and quality assurance. Covering a pseudorapidity range up to 1.48 and designed to withstand up to a 1 MeV equivalent neutron fluence of 2×10^{14} n/cm^2, the BTL is equipped with sensor modules composed of 16 LYSO:Ce crystal bars, each coupled at both ends to silicon photomultipliers (SiPMs). Radiation exposure is expected to increase the SiPM dark count rate up to 20 GHz by the end of HL-LHC operations, degrading the sensor response and introducing an additional time jitter of 35 ps. Several mitigation strategies were implemented, including optimized thermal management that enables SiPM operation at -45°C, significantly suppressing dark counts, while allowing controlled heating up to 60°C during technical stops for local in-situ annealing, partially recovering degraded performance. The focus of this thesis was on studies aimed at enhancing the signal amplitude, a crucial parameter affecting all contributions to the total time resolution. Specifically, this was achieved by investigating SiPMs with different cell size dimensions and sensor modules of varying thickness. These characterization studies were validated under near-final conditions using test beam data. Results from several test beam campaigns are presented, characterizing sensor modules with different thicknesses and SiPM cell sizes. This comparative phase of measurements enabled the development of a detailed model describing all factors affecting performance at both the beginning and end of operation. Such studies guided the final choice of SiPMs with a 25 μm cell size and sensor modules with a thickness of 3.75 mm. Furthermore, studies on modules irradiated to different levels and operated at different temperatures showed that the gradual performance degradation with fluence arises from the combined effect of increasing DCR and a radiation-induced drop in signal amplitude, with only the former being partially mitigated by cooling or annealing.
Sensor modules built to final specifications demonstrated a time resolution of 25 ps for non-irradiated modules and 55 ps for irradiated modules, representing beginning and end of BTL operation conditions at HL-LHC, respectively, with uniform response across the full active area. Dedicated studies of particle impact angles confirmed that the time resolution scales consistently with deposited energy, evaluating timing performance throughout the detector acceptance. The developed and validated model of each contribution allows verification of the main factors defining time performance over the detector lifetime, showing that, at the end of operation, performance is dominated by photo-statistics fluctuations, while electronic noise and DCR contributions have been significantly reduced through the implemented optimizations.

This extensive R&D effort confirmed the target performance and marked the start of the production phase. Procedures for assembly and QA/QC of BTL components were defined to ensure uniformity and reproducibility across the four assembly centers currently producing the detector components. These procedures, largely based on the experience gained during the characterization studies, have so far achieved a sensor module production yield of 95%. Approximately 25% of the BTL detector segments have been produced, with full integration expected by the end of 2026.

Contents

Chapter 1

The Higgs boson in the Standard Model and the investigation of its self coupling

The investigation of nature has ancient roots in humankind. Strongly driven by the philosophical and mathematical knowledge of each era, humanity has always tried to understand the reason behind interactions and the nature of matter itself. The current theoretical framework capable of describing three of the four known fundamental forces acting on matter is the Standard Model (SM), initially formulated about forty years ago. It predicts the existence of six quarks, three charged leptons and three neutrinos, as well as three interactions mediated by gauge bosons: electromagnetism, the strong force, and the weak force, while gravity is not yet included.

In 2012, the last missing piece of this theory was confirmed with the discovery of the Higgs boson, the particle responsible for the origin of mass of the fundamental constituents. Over the years, increasingly precise measurements have been performed at the LHC, thanks to its unprecedented energy and luminosity as a particle collider. All experimental tests of the Standard Model, including those involving the Higgs sector, have so far confirmed its predictions. At present, the properties of the Higgs boson, such as its interactions with other particles and the strength of these couplings, have been measured with remarkable precision. However, among all its properties, the self-interaction of the Higgs boson remains unobserved, as it corresponds to an extremely rare process. This measurement is of crucial importance, as it provides direct access to the Higgs boson's self-coupling and thus to the shape of the Higgs potential, offering fundamental insights into the mechanism of spontaneous symmetry breaking responsible for the generation of particle masses.

In this chapter, a historical overview of the developments leading to the formulation of the Standard Model is presented in Section 1.1.1, followed by a brief description of the Standard Model and the mechanism of spontaneous symmetry breaking in Section 1.1.2. A review of the main experimental results concerning the properties of the Higgs boson is reported in Section 1.2, with a particular focus on production and coupling measurements. The production of Higgs boson pairs at the LHC is then discussed, describing the main production mechanisms and summarizing the current results obtained by the ATLAS and CMS experiments, including their most recent combinations. Higgs boson pair production also provides opportunities to search for new physics, for instance through resonant production, where a new particle predicted by Beyond the Standard Model (BSM) theories may decay into two Higgs bosons, as shortly presented in Section 1.2.3. Finally, an outlook on future prospects and projections for Higgs measurements, including the self-coupling measurement, at ATLAS and CMS during the High-Luminosity LHC phase is reported in Section 1.3, highlighting the expected performance and improvements.

1.1 The theoretical framework of particle physics

1.1.1 Historical perspective

Studying the history of science allows us to see scientific progress as part of humanity's broader quest to understand itself and the world [1]. Advances in science have historically arisen both from the search for answers to existing questions and from the ability to formulate better ones, with discoveries often shaped by their social and historical contexts. In particular, in the field of physics, humankind has, since the very beginning, tried to explain and understand the world around us through the laws of nature. Simplifying somewhat, the description of nature in physics is based on the study of forces. Historically, electricity, magnetism, and gravitation were the first forces to be investigated, each to varying degrees of depth. Efforts to describe these different forces were made in various ways, constrained by the mathematical knowledge available at the time. Even today, a fully unified theory that encompasses all forces remains beyond our reach. At present, we describe interactions in terms of four fundamental forces, and the framework of particle physics successfully unifies three of them: the electromagnetic, strong, and weak interactions. Gravitation, however, is described separately by other theories, and a comprehensive unified framework is still lacking. The particle physics model on which we currently rely, the Standard Model, was formulated in the 1970s, following a long intellectual journey that began with the earliest studies of electricity and magnetism.

The ancients already knew of magnets and used them as compasses, while static electricity was observed and considered analogous to magnetism. It was in the sixteenth century that William Gilbert systematically studied magnetism, and by the late eighteenth century, Coulomb started investigating electricity and discovered that electrical attraction follows an inverse-square law, similar to Newton's law of gravitation. Around the same time, Benjamin Franklin proposed the existence of two types of electric charge to explain attraction and repulsion. Studies on the Leyden jar, reminiscent of lightning, eventually led to the invention of the lightning rod. In 1800, Volta invented the battery, producing electricity from copper and zinc in salt water, thereby opening the way to electrodynamics. Shortly afterwards, electrolysis was discovered, and its reverse process, electric currents driving chemical reactions, gave rise to electrochemistry and the discovery of new elements. In 1819, Ørsted observed that a current-carrying wire deflects a compass needle, revealing electromagnetism. The following year, Ampère showed that currents in wires attract or repel each other, establishing the first connection between electricity and magnetism. In 1831, Faraday discovered electromagnetic induction, demonstrating that a changing magnetic field produces electric currents. These discoveries were unified mathematically in the 1860s by Maxwell, who formulated Maxwell's equations and predicted that light is an electromagnetic wave. Meanwhile, studies of light and optics progressed in parallel. Newton described light as particles, while Hooke and Huygens proposed a wave theory. In 1801, Thomas Young demonstrated interference effects, confirming the wave nature of light. Faraday later discovered that magnetic fields can alter light polarization (the Faraday effect), inspiring Maxwell to link electricity, magnetism, and light. By the end of the century, Hertz (1887) had produced radio waves, and Röntgen (1895) had discovered X-rays, extending the electromagnetic spectrum. At the turn of the century, Planck (1900) introduced the quantization of energy to explain black-body radiation. Einstein (1905) applied this concept to light, showing that it consists of photons, capable of ejecting electrons from metals (the photoelectric effect). In the same year, his explanation of Brownian motion provided strong evidence for the reality of atoms. Studies of cathode rays in the late 19th century led J. J. Thomson (1897) to identify electrons as negatively charged particles. In 1923, Louis de Broglie proposed that electrons also have wave-like properties, an idea later confirmed by electron diffraction and double-slit experiments. These developments laid the groundwork for quantum mechanics and, eventually, quantum electrodynamics.

Building on his earlier work, Einstein developed the special theory of relativity (1905) and general relativity (1915). The fusion of relativity with quantum mechanics later gave rise to quantum field theory (QFT), the standard framework for describing matter and interactions. While Newton's law of gravitation was the first classical "field theory", the concept of fields was formally introduced by Faraday. By the late 19th century, however, classical theories could no longer explain black-body radiation. Planck's quantization (1900) and Bohr's atomic model (1913) marked the beginning of a new framework in which electrons occupy discrete energy levels and emit photons when transitioning between them. Between

1925 and 1927, Heisenberg, Schrödinger, Born, Dirac, Pauli, and Jordan developed a coherent quantum theory, showing that both particles and fields must be described quantum mechanically. Quantizing the electromagnetic field revealed that each mode carries discrete quanta of energy, the photons. By the late 1920s, it became clear that particles themselves could be viewed as quanta of their respective fields, establishing the foundation of QFT. This framework implied that matter could be created from fields given sufficient energy. In 1930, Pauli postulated the neutrino to account for missing energy in beta decay, and Fermi (1932) developed the first quantitative theory of this process. In the same year, Chadwick discovered the neutron, confirming that nuclei consist of protons and neutrons. Dirac's relativistic wave equation (1928) predicted the existence of antiparticles, experimentally confirmed by Anderson in 1932 with the discovery of the positron. QFT naturally incorporates the creation and annihilation of particle–antiparticle pairs. Interactions can be described as exchanges of virtual particles, with the range of a force inversely related to the mediator mass: massless photons mediate electromagnetism, while Yukawa (1935) proposed massive pions to account for the finite range of nuclear forces (now known to be mediated by gluons). The formulated theory seemed to provide a satisfactory explanation of what was observed so far, but challenges soon appeared. In 1930, Oppenheimer showed that an electron interacting with its own field leads to infinite energy corrections. Similar divergences arose in related problems. The solution came through renormalization: the observed mass and charge are understood as the sum of "bare" parameters and divergent self-interaction terms that cancel, leaving finite, measurable quantities. Weisskopf (1936) suggested this approach, later developed by Feynman, Schwinger, and Tomonaga, who formulated QED using Feynman diagrams to calculate interaction probabilities consistent with relativity and particle–antiparticle symmetry [2].

QED achieved remarkable precision by the 1940s. However, Fermi's original weak interaction theory was non-renormalizable, and the strong interaction's coupling constant prevented perturbative methods. A deeper reformulation was required. Yang and Mills (1954) [3] introduced gauge theories for the weak and strong interactions, mediated by vector bosons (weak) and gluons (strong). Gauge interactions are based on symmetries, remaining invariant under local spacetime transformations. A major obstacle, however, was that gauge symmetry does not allow gauge bosons to have mass and if they were massless, they should already have been detected. Introducing mass by hand destroyed the rationale of a gauge theory [4]. Physicists later turned to the idea of spontaneously broken symmetry to explain masses, where symmetries of the Lagrangian are not symmetries of the vacuum. In 1961, Goldstone proved that for every spontaneously broken symmetry there must exist a massless spinless particle. In 1964, Higgs showed how this reasoning could apply to gauge symmetry: the later identified as Goldstone boson is reinterpreted as the helicity zero component of a gauge boson, which thereby acquires mass. Independently, Englert and Brout discovered the same mechanism [5, 6, 7, 8, 9].

In the 1950s, Goldberger and Treiman demonstrated correlations between pion decay and strong and weak coupling constants. This led to the idea of a conserved axial-vector current. Developments in spontaneous symmetry breaking and symmetry groups then provided a more coherent theoretical framework. A broken $\mathrm{SU}(2)\times\mathrm{SU}(2)$ symmetry, consisting of isotopic spin and chiral isotopic spin transformations acting oppositely on left and right-handed nucleon fields, implied an exact symmetry that is spontaneously broken. This allowed precise predictions for low-energy interactions of the massless Goldstone bosons, which in this case are the pions. This reasoning culminated in 1967, when Weinberg, Salam, and Glashow independently formulated the electroweak theory. They all arrived at the same $\mathrm{SU}(2)\times\mathrm{U}(1)$ group structure, an exact but spontaneously broken gauge symmetry. Its spontaneous breaking gives mass not only to the intermediate vector bosons of the weak interaction but also to the leptons. The theory enabled precise calculations of the W and Z boson masses and their couplings, with a single unknown: the weak mixing angle θ_W. The predicted masses were large enough to explain why these bosons had not yet been observed. The electroweak theory also predicts three gauge bosons associated with spontaneously broken symmetries. These bosons are "eaten" by the Higgs mechanism, leaving only one observable massive neutral scalar particle. This particle had predicted couplings, although its mass remained unknown for fifty years. Renormalization of the electroweak theory was provided by Veltman and 't Hooft in 1971 via path integration. They defined gauges in which spontaneously broken non-Abelian gauge theories have a finite number of divergences to all orders of perturbation. Later work showed that the number of divergences matches the number of parameters in the original theory. This allows all infinities to be absorbed by parameter redefinitions. Renormalizability is important not only for removing divergences, but

also for describing weak and electromagnetic interactions at energies much higher than the characteristic mass used as the expansion parameter, i.e. the W boson mass. Eventually, the first experimental confirmation of the electroweak theory came in 1973 with the observation of neutral currents at CERN [10, 11].

After the success of Yang-Mills theory, Gross and Wilczek independently discovered the property of asymptotic freedom in non-Abelian gauge theories. At very high energies, or equivalently at short distances, the strength of the force decreases. At that time, it was already known [12, 13, 14] that quarks come in three colors, which led to the consideration of an SU(3) gauge group for strong interactions. Initially, gluons were thought to acquire mass via spontaneous symmetry breaking, similar to the electroweak theory. Later, it was found that they are massless. However, gluons are not observed freely due to the infrared properties of non-Abelian gauge theories, which imply that color is confined. Colored particles, the quarks, are therefore confined and can only exist in bound, color-neutral states. Another unresolved issue was strangeness non-conservation in weak interactions: charge exchange in semileptonic interactions violated strangeness conservation. This problem was solved in 1970 by Glashow, Iliopoulos, and Maiani, who showed that the violation disappears with the addition of two quark doublets. This implied a fourth quark, called the charm quark. The charm-anticharm bound state was observed in 1974 [15, 16], confirming both the existence of the fourth quark and the reality of quarks. The final piece to complete the Standard Model was the hypothesis of a third generation of quarks. This introduced a new mechanism for CP violation through the complex phase in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which manifests in semileptonic weak interactions. Further confirmation of the theory came in 1983 with the discovery of the W and Z vector bosons, along with precise measurements of their masses [17, 18]. After that, the search for the Higgs boson became one of the main goals of particle physics and an important motivation for the construction of the CERN Large Hadron Collider (LHC). Its observation required overcoming many challenges due to its large mass, which was beyond the reach of previous electron-positron colliders such as the Large Electron Positron (LEP) collider, as well as its low production cross section and relatively complex decay channels, which made detection difficult even at earlier hadron colliders such as the Tevatron. Finally, in July 2012, the ATLAS and CMS collaborations at the CERN LHC announced the observation of a new particle with a mass of about 125 GeV, consistent with the Higgs boson predicted by the Standard Model. This discovery further validated the theory and marked a milestone in our understanding of fundamental physics.

1.1.2 The building blocks

The finalization of the electroweak theory, formulated in 1974, is now known as the Standard Model and has proven remarkably accurate in predicting particle physics phenomena. Its complete formulation arises from a long series of observations and experiments, which have been interpreted and coherently expressed within a mathematical framework. This framework is currently capable of describing fundamental particles and the interactions between them, with the exception of gravity, which remains outside the scope of the theory. Many textbooks and review articles on the Standard Model are available in the literature (e.g. Refs. [19, 20, 21]). The goal of this section is to provide a brief overview of the theory without attempting a comprehensive description.

The Standard Model, as thus formulated, is a quantum field theory [22] describing fundamental particles and their interactions under three of the four fundamental forces: the electromagnetic, weak, and strong interactions. These forces exhibit specific characteristics that can be expressed as local gauge symmetries. More specifically, the SM is based on local gauge invariance under the symmetry group [23], expressed in Equation 1.1, where C denotes color, L the weak isospin, and Y the hypercharge:

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{1.1}$$

The requirement of local gauge invariance introduces one gauge field for each generator of the group: eight gluon fields $G_\mu^{1,\cdots 8}$ associated with $SU(3)_C$, three weak fields $W_\mu^{1,2,3}$ associated with $SU(2)_L$, and a single hypercharge field B_μ associated with $U(1)_Y$. These fields correspond to the mediators of the strong and electroweak interactions, which after spontaneous symmetry breaking give rise to the physical gauge bosons, the gluons, the photon, and the massive $W^\pm$ and Z^0 bosons.

The particles that experience these forces are fermions, which have spin 1/2. Matter is composed of quarks and leptons, which are experimentally observed to be organized into three generations, or families, differing in mass. Each generation contains a doublet of quarks and a doublet of leptons, with each particle accompanied by a corresponding antiparticle. In total, there are six leptons: the electron, muon, and tau, each with its associated neutrino, and six quarks: up, down, strange, charm, bottom, and top. Each fermion has left- and right-handed components. The weak force has been experimentally observed [24] to interact only with the left-handed components, while the electromagnetic and strong forces act on both chiralities. In the Standard Model, left-handed components form weak-isospin doublets under $SU(2)_L$, whereas right-handed components are weak-isospin singlets. Each particle belongs to a specific representation of the gauge group, which determines how it interacts with the various forces. Quarks are particles with a color charge and are subjected to all three forces. Quarks exist only as bound states with other quarks and antiquarks, collectively known as hadrons, due to QCD confinement, which prevents direct observation of color charges under the temperature and energy density conditions of ordinary matter. However, accelerating and colliding particles can create extreme conditions in which hadronization becomes accessible, i.e. enabling the transition of a system of quarks and gluons into color-neutral composite particles, known as hadrons. A schematic representation of the particles and their interactions is shown in Figure 1.1.

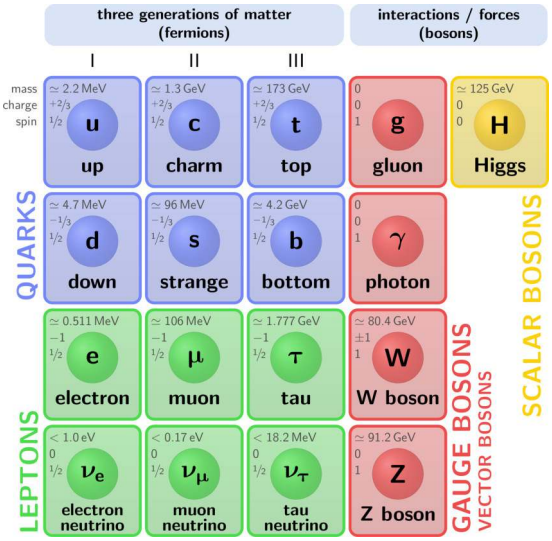

Figure 1.1: Schematic overview of the Standard Model particles, including quarks leptons, gauge bosons and the Higgs boson [25].

The interactions of these particles within the theory can be expressed via the Standard Model Lagrangian in Equation 1.3:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{YM}} + \mathcal{L}_{\text{ferm}} + \mathcal{L}_H + \mathcal{L}_{\text{Yuk}} \tag{1.2}$$

$$= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + |D_\mu\phi|^2 - V(\phi) + \psi_i y_{ij}\psi_j\phi + \text{h.c.} \tag{1.3}$$

where $\mathcal{L}_{\text{YM}}$ describes the free propagation and self-interactions of the gauge bosons, and $\mathcal{L}_{\text{ferm}}$ represents the fermion part, both terms are considered without mass terms to preserve gauge invariance. However, since these particles are experimentally observed to have mass, an extension of the theory is required. This is achieved through electroweak spontaneous symmetry breaking, which generates masses for the fermions and the weak gauge bosons while preserving the gauge invariance of the theory and it adds the $\mathcal{L}_H$ and $\mathcal{L}_{\text{Yuk}}$ terms to the SM lagrangian.

The introduction of electroweak symmetry breaking (EWSB) in the gauge sector represented a historical milestone in the development of the Standard Model and was experimentally confirmed in 2012 with the observation of the Higgs boson at the LHC. The idea explaining mass generation via the Higgs boson,

and involving EWSB, is known as the Brout–Englert–Higgs (BEH) mechanism. Without this mechanism, the theory would predict massless gauge bosons. To provide mass to the W^+, W^-, and Z^0 bosons while leaving the photon massless, a complex scalar doublet field is introduced, as in Equation 1.4:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.4}$$

Within the Higgs term $\mathcal{L}_H$ of the Standard Model Lagrangian, the self-interaction of the scalar field can be expressed through the potential in Equation 1.5:

$$V(\Phi) = -\mu^2(\Phi^\dagger\Phi) + \lambda(\Phi^\dagger\Phi)^2, \tag{1.5}$$

where the potential has a ground state corresponding to the vacuum state v, and it must be positive and real ($\lambda > 0$). The shape of the potential can either be a simple parabola (if $(-\mu^2) > 0$) or a "Mexican hat" (if $(-\mu^2) < 0$), illustrated in Figure 1.2. In the first case, the unique minimum at $\Phi = 0$ preserves the symmetry of the Lagrangian. In the second case, there is no unique ground state, but rather a set of degenerate vacuum states lying on a circle in the complex plane. Choosing one of these vacuum states spontaneously breaks the symmetry of the Lagrangian: applying a symmetry transformation to any of the vacuum states produces a physically distinct configuration.

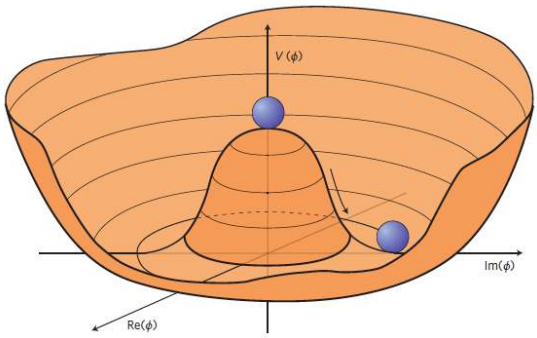

Figure 1.2: The illustration depicts the Higgs potential in the case where $(-\mu^2) < 0$, giving it the characteristic "Mexican hat" shape. The blue ball represents the Universe status. When it was located at the center of the potential, along the imaginary axis, all particles were massless. However, this configuration corresponded to a high-energy state. As the Universe expanded and cooled, it dropped into the trough of the potential, a process that endowed particles with inertia and, consequently, mass [26].

Requiring the vacuum expectation value to be electrically neutral forces the upper component of the field to vanish, leading to the vacuum configuration shown in Equation 1.6. This spontaneous symmetry breaking reduces the Standard Model gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ to $SU(3)_C \times U(1)_{\text{em}}$.

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \qquad v = \sqrt{\frac{\mu^2}{\lambda}}. \tag{1.6}$$

This breaking implies the existence of three massless Goldstone bosons, which correspond to three of the four degrees of freedom of the Higgs field. These massless modes become the longitudinal components of the physical $W^\pm$ and Z gauge bosons, providing them with masses given in Equation 1.7, where g and g' are the $SU(2)$ and $U(1)$ gauge couplings, and θ_W is the weak mixing angle (Weinberg angle).

$$m_W = \frac{1}{2} g v \qquad m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}\, v = \frac{m_W}{\cos\theta_W} \tag{1.7}$$

The gauge symmetries that remain unbroken correspond to massless gauge fields. In particular, the $U(1)_{\text{em}}$ and $SU(3)_C$ symmetries remain conserved, so the photon and the gluons remain massless. Therefore, among the four degrees of freedom introduced with the scalar field, three are absorbed by the W^+, W^-, and Z bosons, while the fourth corresponds to the physical Higgs boson. The Higgs is electrically neutral and transforms as a singlet under QCD, meaning that it does not couple directly to photons or

gluons at tree level.

Fermions also acquire mass in the Standard Model through gauge-invariant Yukawa interactions with the Higgs field. These terms, represented by the $\mathcal{L}_{\text{Yuk}}$ contribution to the Lagrangian, generate fermion masses after electroweak symmetry breaking:

$$m_{f_i} = \frac{y_{f_i} v}{\sqrt{2}}, \tag{1.8}$$

where y_{f_i} is the Yukawa coupling of the fermion f_i, and $i = 1, 2, 3$ labels the three generations. However, the observed hierarchy of Yukawa couplings is not explained by EWSB, leaving the wide range of fermion masses an open question.

After EWSB, the Higgs potential can be expressed as in Equation 1.9, in which the Higgs acquires trilinear and quadrilinear self-interactions:

$$V(H) = \frac{1}{2} m_H^2 H^2 + \lambda v H^3 + \frac{\lambda}{4} H^4 \tag{1.9}$$

both proportional to the Higgs self-coupling constant λ (or trilinear coupling, i.e. the interaction of three Higgs bosons), which is related to the vacuum expectation value and the Higgs mass via Equation 1.10:

$$m_H = \sqrt{2\lambda}\, v. \tag{1.10}$$

The Higgs mass has been measured at the LHC to be approximately 125 GeV, while the vacuum expectation value is determined from electroweak measurements to be $v \approx 246$ GeV. This corresponds to a value of $\lambda \approx 0.13$, with any deviation from this value indicating a modification of the Higgs potential. The trilinear coupling represents the lowest-order self-interaction and provides the most direct probe of the Higgs potential. This can be studied via Higgs boson pair production, however, this process is extremely rare, with a cross-section roughly 1000 times smaller than that of single Higgs production, making its observation at the current LHC challenging. Current LHC analyses are probing this process and setting upper limits on its rate, but no observation has yet been made, while the upcoming HL-LHC dataset, with a tenfold increase in luminosity, is expected to allow observation of di-Higgs production.

1.2 Measurements of the Higgs boson at the LHC

The first experimental investigations of the Higgs boson began with the operation of LEP and the Tevatron, which set limits on its mass through indirect measurements derived from precision electroweak measurements. A broader program became accessible at the LHC, where proton-proton collisions were performed at center-of-mass energies ranging from 8 to 14 TeV, to be compared with the approximately 200 GeV and 2 TeV at LEP and Tevatron, respectively. The discovery of the Higgs boson in 2012 marked a major milestone in our understanding of the Standard Model and the BEH mechanism. Since its discovery, many studies at the LHC have focused on a detailed investigation of the nature and properties of the Higgs boson. Early analyses focused on the dominant decay modes, while rarer and indirect decay channels, such as photon pairs produced via fermion and W boson loops, have also been explored. Couplings to fermions through the Yukawa interaction have been probed, and both the Higgs boson mass and total width have been measured with high precision. In the following, the main results related to the characterization of the Higgs boson are presented, with particular emphasis on latest CMS measurements based on an integrated luminosity of 138 fb^{-1} collected at a center-of-mass energy of 13 TeV at the LHC. Comparable results have also been reported by the other general-purpose LHC experiment, ATLAS.

1.2.1 The Higgs boson properties

In a hadron collider operating at energies from 8 to 14 TeV, the dominant mechanism for the production of a Higgs boson with a mass of 125 GeV is expected to be gluon-gluon fusion (ggF, $gg \to H$) via loops of heavy colored particles, followed by the vector boson fusion production (VBF, $qq \to qqH$), associated production with a vector boson (VH), and associated production with a pair of heavy quarks ($q\bar{q}H$). The

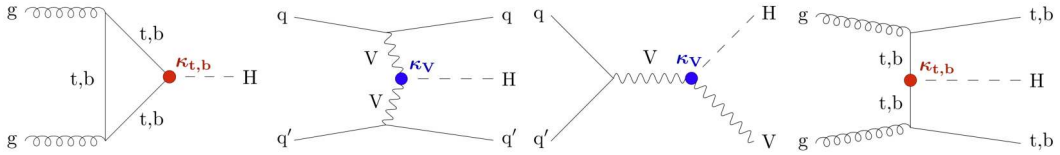

Figure 1.3: Feynman diagrams of the main Higgs boson production modes are shown. From left to right, the gluon-fusion, vector-boson-fusion, vector-boson-associated, and heavy-quark-associated production modes are reported. The parameters κ_i at the Higgs vertices are the coupling modifiers used to parametrize deviations from the SM expectations [27].

leading order Feynman diagrams for these production channels at hadron colliders are shown in Figure 1.3.

The cross sections of the various Higgs boson production processes depend on the center-of-mass energy of the collisions and on the Higgs boson mass. For a fixed Higgs boson mass and collision energy, the production cross section can be theoretically predicted and compared with experimental measurements. Considering the measured Higgs boson mass of 125.38 $\pm$ 0.14 GeV [28], the total production cross section at a center-of-mass energy of 13 TeV is 55.4 ± 2.6 pb, with the gluon-fusion process accounting for approximately 87% of the total, 48.3 $\pm$ 2.4 pb. It is followed by the vector boson fusion production where two quarks radiate virtual vector bosons that then combine to generate a Higgs boson, which accounts for a cross section of 3.77 $\pm$ 0.80 pb. Other production modes with smaller contributions to the total cross section are the associated production with vector bosons (WH with 1.359 $\pm$ 0.028 pb, and ZH with 0.877 $\pm$ 0.036 pb), the associated production with pairs of heavy-flavor quarks ($t\bar{t}H$ with 0.503 $\pm$ 0.028 pb and $b\bar{b}H$ with 0.482 $\pm$ 0.097 pb) [29]. After EWSB, the Higgs field provides mass to the vector bosons through the gauge couplings and to the fermions through the Yukawa couplings. The amplitude of the Higgs boson coupling to vector bosons is proportional to the square of the gauge boson mass, while the amplitude of the coupling to fermions is directly proportional to the fermion mass. For this reason, the Higgs boson can decay into a variety of final states, but preferentially into the heaviest kinematically allowed elementary particles. The Higgs boson can also decay into massless particles via quantum loops. Measuring the Higgs boson decays to various particles and determining their branching fractions provides an important test of the Standard Model, as any deviations from its predictions could indicate new physics. The branching fraction is calculated as the ratio between the partial width of a specific decay and the total width, which for the Higgs boson with a mass of 125 GeV is measured to be $3.2^{+2.4}_{-1.7}$ MeV [30]. A wide range of decay products of the Higgs boson have been measured, with different branching fractions. The $H \to bb$ decay features the highest branching fraction with a measured value of 57.63 $\pm$ 0.70%, followed by the decay into two W bosons (22.00 $\pm$ 0.33%), a pair of τ-leptons (6.21 $\pm$ 0.09%), two Z bosons (2.71 $\pm$ 0.04%), and two photons (0.2%) [29].

Observed data and expectations are usually compared via the signal strength parameter μ, defined as the parameter extracted from fitting signal yields in data with a model, without altering the shape (for example, for a production cross section σ, the signal strength is $\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$). The signal strengths measured at CMS, connecting all production and decay modes, are summarized in Figure 1.4, both individually for each production and decay channel and for their combination. In each production mode, the measured parameters are compared to the SM expectations ($\mu = 1$), showing an excellent agreement with the theory. Any deviations of the signal strength from a value of 1 would indicate physics beyond the Standard Model and deviations are commonly studied within the κ framework. A set of coupling modifiers is used to parameterize potential deviations from the Standard Model expectations of the Higgs boson couplings to particles. The set consists of six coupling modifiers corresponding to the tree level Higgs boson couplings: κ_W, κ_Z, κ_t, κ_b, κ_τ, and κ_μ. In addition, the effective coupling modifiers κ_g, κ_γ, and $\kappa_{Z\gamma}$ are introduced to describe the ggH production and the $H \to \gamma\gamma$ and $H \to Z\gamma$ decay processes. The coupling modifier values measured by CMS are reported in Figure 1.5, where the Higgs coupling modifiers to fermions and the square root of the coupling modifiers to gauge bosons are shown as a function of the particle mass. This corresponds to a p-value of 37.5% with respect to the Standard Model expectation, illustrating the wide range of mass values probed in terms of Higgs couplings and the predictive power of the theory across the entire range.

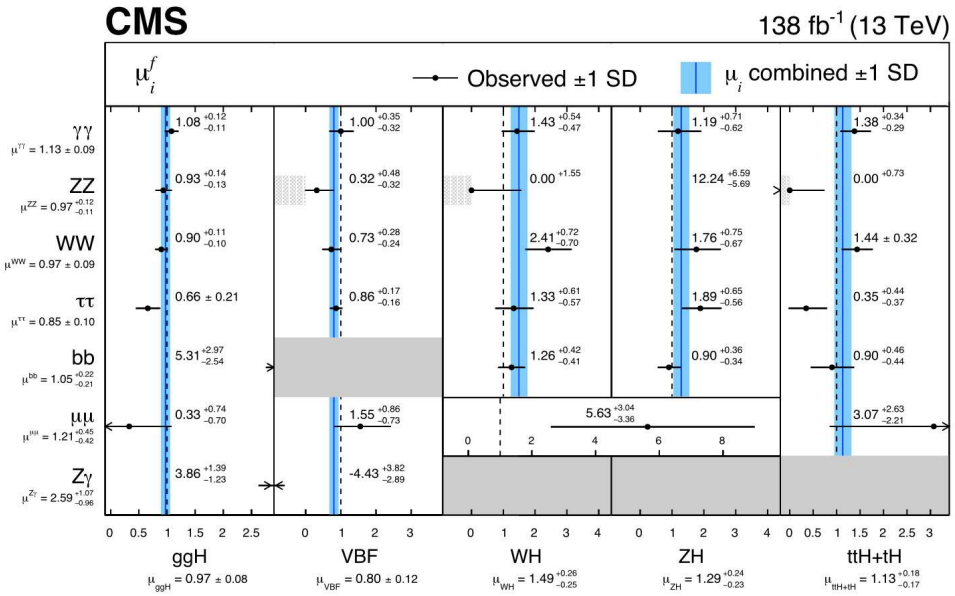

Figure 1.4: The signal strength is reported separately for each production mode and for each decay channel, as well as for their combination. The SM expectation is indicated by a dashed vertical line at $\mu = 1$, while the combined value of all decay channels for a given production mode is shown as a blue vertical line, with the corresponding 68% confidence level (CL) interval represented by the shaded band. Dark grey boxes indicate cases where no measurement is available, while light grey boxes correspond to cases where the signal strength is constrained to be positive [29].

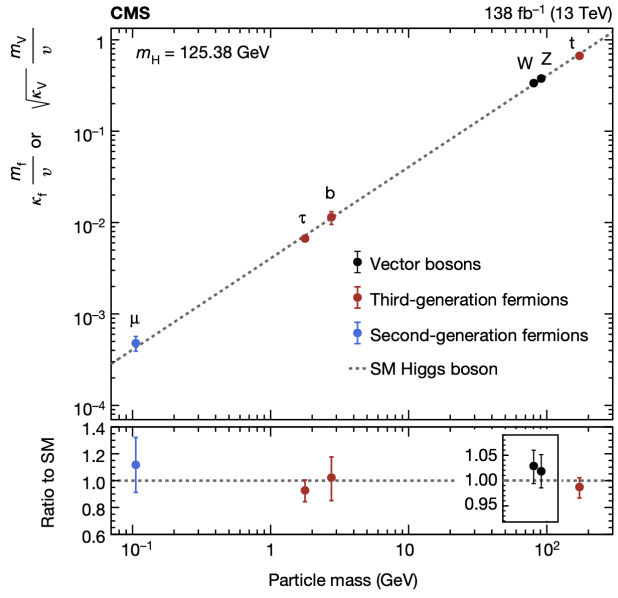

Figure 1.5: Measured coupling modifiers at the CMS experiment are reported for both fermions and gauge bosons. For the latter, the square root of the coupling is shown in order to preserve a linear dependence on the mass [27].

The Higgs boson mass is not predicted by the SM, but once measured, all its other properties become precisely predicted by the theory. Many of these properties have been measured with a precision better than 10%, all in agreement with SM expectations, representing an impressive success of the theory.

However, despite its success, the SM does not provide explanations for several open questions, suggesting it may be a low-energy approximation of a more complete theory. Outstanding issues include the naturalness problem, i.e. why the Higgs boson is a light scalar at the electroweak scale, the nature of the electroweak phase transition, the metastability of the SM vacuum, the inflationary phase of the early Universe, and the matter–antimatter asymmetry. Addressing these questions requires a deeper investigation of the Higgs sector. Higgs physics is therefore crucial for advancing our understanding of both particle physics and cosmology, providing a unique probe of the early Universe, the stability of the vacuum, and the mechanisms behind structure formation immediately after the Big Bang. Investigating further the Higgs sector is at center of the physics program of the new high luminosity phase envisaged for the LHC, which has been designed to deliver nearly ten times the luminosity of the LHC Run 3. This offers ATLAS and CMS the opportunity to significantly enhance sensitivity to many measurements, such as the Higgs self-coupling, probe weaker coupling modifiers, such as those involving charm quarks, and search for possible exotic Higgs decays.

1.2.2 The Higgs boson self interaction

A process that remains currently not observed is the production of a pair of Higgs bosons. The measurement of di-Higgs production is key to evaluating the Higgs self-coupling and, therefore, the Higgs potential. At the LHC, this process is expected to occur mainly via gluon fusion, represented by the Feynman diagrams in Figure 1.6, where the two Higgs bosons can appear either from a top or bottom quark loop, or from a virtual Higgs boson produced via a top or bottom quark loop, which then decays into two Higgs bosons. The latter production mode provides direct access to the Higgs self-interaction. These two modes are comparable in magnitude but interfere destructively, making the overall pair production cross-section small. The latest theoretical predictions for gluon fusion yield

$$\sigma_{\mathrm{ggHH}} = 30.8^{+2.0}_{-7.1} \text{ fb at } \sqrt{s} = 13 \text{ TeV}, \quad \sigma_{\mathrm{ggHH}} = 34.1^{+2.2}_{-7.9} \text{ fb at } \sqrt{s} = 13.6 \text{ TeV}$$

for a Higgs boson mass of 125 GeV [31]. The next most significant production mode is vector-boson fusion, proceeding through a virtual Higgs, a four-point interaction, or vector boson exchange. The corresponding cross sections are

$$\sigma_{\mathrm{VBFHH}} = 1.69 \pm 0.05 \text{ fb at } 13 \text{ TeV}, \quad \sigma_{\mathrm{VBFHH}} = 1.87 \pm 0.05 \text{ fb at } 13.6 \text{ TeV}$$

for a Higgs boson mass of 125 GeV [32, 33]. Given the small cross-section, observation of this process has so far been beyond the capabilities of the LHC dataset. Each Higgs boson produced in a pair can

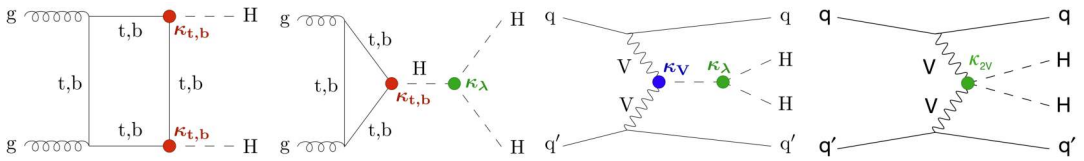

Figure 1.6: The main Feynman diagrams for the production of Higgs boson pairs are shown: from the left, the first two correspond to box and triangle production via gluon fusion, while the two diagrams on the right represents vector boson fusion production mode [27].

decay differently, resulting in a wide variety of possible final states, as shown in Figure 1.7, where the branching fractions of the most sensitive final states are reported. The current approach at the LHC is to study different production modes and decay channels independently and then combine them. Analyses typically employ the κ framework, defining the coupling modifier κ_λ as the scale of the Higgs self-coupling λ relative to its SM value, with $\kappa_\lambda = 1$ corresponding to the SM prediction. The investigation of the SM process proceeds via the non-resonant production modes shown in Figure 1.6. However, Higgs boson pair production can also serve as a tool for searching for new physics through resonant processes, looking for a hypothetical resonance particle decaying into two Higgs bosons, as discussed later in this section.

Within non-resonant di-Higgs production searches, the most sensitive channels have been analyzed first by both the CMS and ATLAS experiments, owing to their relatively large branching fractions and

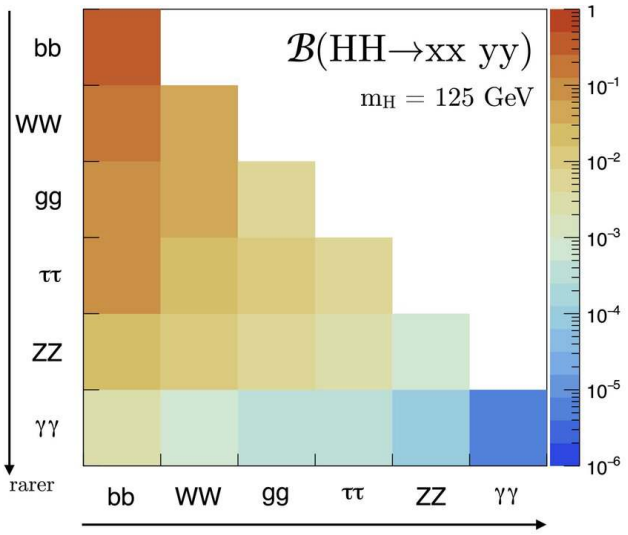

Figure 1.7: The branching fractions of the most sensitive channels are reported for all their combinations, assuming a mass of the Higgs boson of 125 GeV [34].

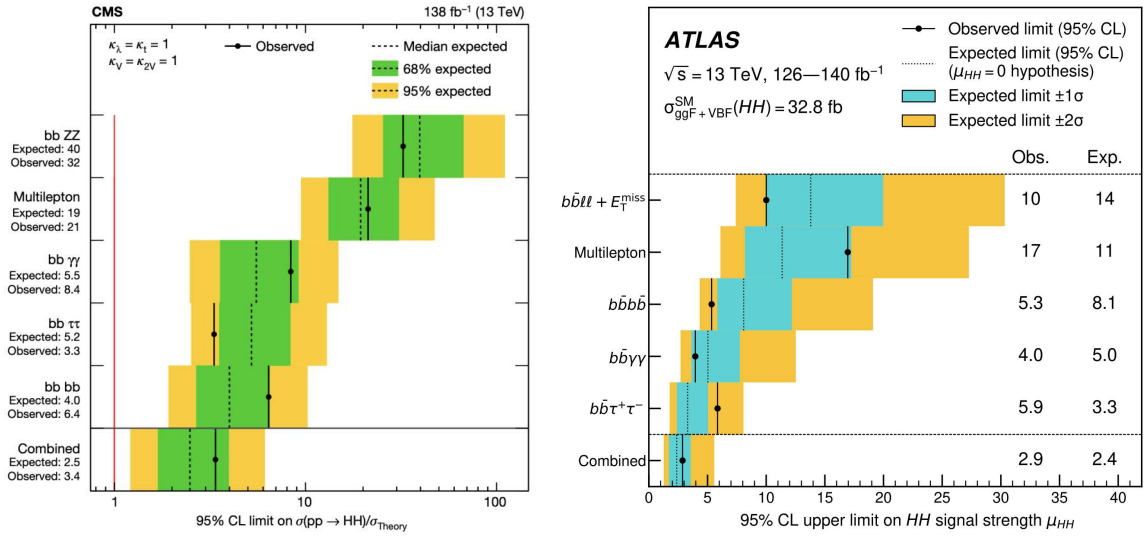

Figure 1.8: Limits on Higgs boson pair production. The observed and expected upper limits at 95% CL on the ratio of the measured cross section to the Standard Model prediction are shown for individual final states and their combination for CMS [27] and ATLAS [35]. Each expected value is reported with the corresponding colored bands representing the one standard deviation and 2 standard deviation beyond the expected limit.

clean experimental signatures. These include the decays of the two Higgs bosons into four b quarks, two b quarks and two τ leptons, two b quarks and two photons, multilepton final states, and two b quarks and two Z bosons. While no observation of the process has yet been achieved, upper limits at 95% confidence level [36] have been set by combining all channels, corresponding to less than 3.4 times the SM expectation in CMS and 2.9 times in ATLAS. The results from both experiments are shown in Figure 1.8. The expected and observed limits have also been evaluated as functions of the coupling modifiers κ_λ (related to the Higgs self-coupling) and κ_{2V} (the quartic vector-boson–Higgs coupling). The results of the CMS combination are reported in Figure 1.9, where the red curves represent the theoretical dependence of the cross section on the coupling modifiers: a dip near $\kappa_\lambda = 1$ arises due to the destructive interference between the contributing production amplitudes. Using the data included in these CMS analyses, the self-coupling modifier κ_λ is constrained within the range $-1.24 < \kappa_\lambda < 6.49$ at 95% CL, while κ_{2V} is constrained to be positive in the interval $0.64 < \kappa_{2V} < 1.38$. These results define, with a significance of 6.6 standard deviations, the existence of the quartic $VVHH$ coupling.

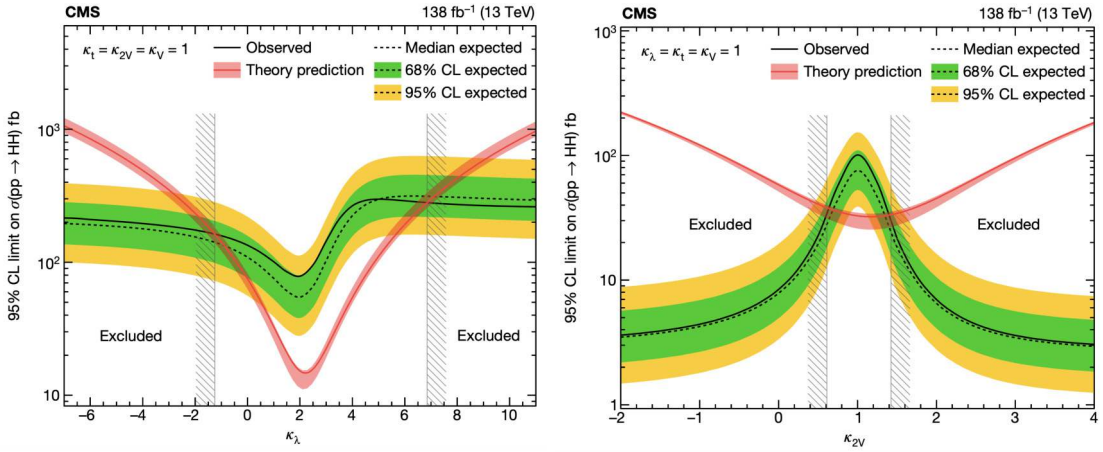

Figure 1.9: Expected and observed 95% CL upper limits on the di-Higgs production cross section as functions of κ_λ (left) and κ_{2V} (right), assuming the other coupling modifiers to be 1. The red line represents the theoretical prediction for the process and the hatched regions represent the excluded values at 95% CL [27].

The results obtained during Run 2 provide valuable constraints on the Higgs self-coupling parameter κ_λ. The combination of different search channels, despite their varying challenges, has led to a consistent picture compatible with the SM, with no significant deviations [27]. However, the results also highlight the current limitations of experimental sensitivity. Background processes, particularly from QCD multi-jet events, and the small branching ratios of some decay channels remain significant challenges. Nevertheless, the application of advanced machine learning techniques has substantially improved the ability to distinguish signal from background, allowing us to achieve limits close to a few times the SM expectation. The analysis sensitivities achieved with the data collected during the ongoing LHC Run 3 are already showing significant improvements compared to LHC Run 2, driven by the use of more sophisticated analysis techniques. For instance, the CMS experiment is operating under conditions where Run 3 is expected to deliver an integrated luminosity of 300 fb^{-1} and an average of about 60 pileup interactions, i.e. the mean number of interactions per bunch crossing, as shown in Figure 1.10. The forthcoming High-Luminosity LHC (HL-LHC) phase will provide roughly an order of magnitude more integrated luminosity, thereby offering new opportunities to further enhance the sensitivity of analyses targeting the measurement of the Higgs self-coupling. In particular, the introduction of new triggers specifically optimized for di-Higgs production, as discussed in Section 2.2.2, together with the adoption of more advanced algorithms for object identification, is expected to lead to a substantial improvement in sensitivity, providing a unique opportunity to probe the Higgs sector with unprecedented precision and depth.

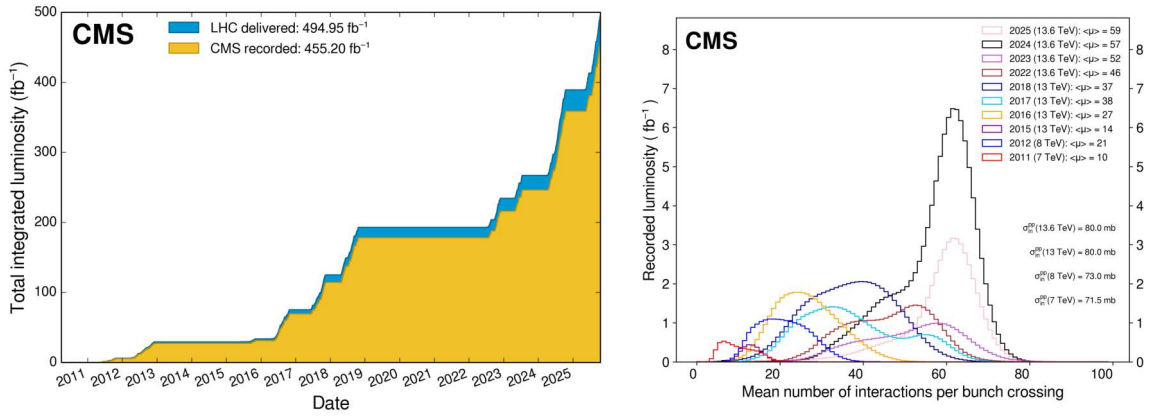

Figure 1.10: Left: cumulative LHC delivered and CMS recorded luminosity (blue and yellow) for proton-proton collisions during the three LHC Runs operated so far. Right: distribution of the mean number of interactions per bunch crossing (pileup) for each CMS data-taking period [37].

1.2.3 Di-Higgs as a tool for probing new physics

In addition to providing direct access to the study of the Higgs self-coupling, the production of Higgs boson pairs also offers a powerful probe for processes beyond the Standard Model. Within the SM, di-Higgs production occurs via non-resonant production, without any intermediate heavy particle decaying into two Higgs bosons. Therefore, the observation of resonant di-Higgs production would represent clear evidence of new physics, as predicted by several BSM scenarios featuring a heavy particle decaying into two Higgs bosons. Almost all extensions of the Higgs sector include such heavy states, as in models with weak-isospin singlets, doublets, triplets, or theories with extra spatial dimensions. The resonant production expected in these models typically has a cross section larger than the SM non-resonant production, depending on the model. A few examples are briefly discussed below, while a more detailed discussion can be found in Refs. [38, 39, 40, 41].

The Two-Higgs-Doublet Model (2HDM) provides a perturbative extension of the SM by introducing a second Higgs doublet. This leads to five physical Higgs bosons: one light neutral CP-even state, one heavy neutral CP-even state, one neutral CP-odd state, and two charged Higgs bosons $H^{\pm}$. The Minimal Supersymmetric Standard Model (MSSM) extends the SM through supersymmetry, predicting a superpartner for each SM particle and an extended Higgs sector equivalent to a 2HDM, resulting in the same five Higgs states with similar properties. Warped Extra Dimension (WED) models [41], as proposed by Randall and Sundrum, postulate the existence of an extra spatial dimension compactified between two fixed branes: one associated with the electroweak scale (TeV brane) and the other with the Planck scale. This setup introduces a warp factor that naturally explains the hierarchy between the Planck and electroweak scales. Assuming an exponential metric along the fifth dimension, bounded by the branes, this model predicts the existence of two resonances that can decay into Higgs boson pairs: the spin-0 radion, associated with fluctuations of the inter-brane distance, and the spin-2 Kaluza–Klein excitation of the graviton.

1.3 Prospects for Higgs measurements at the HL-LHC

To push further the physics reach of current LHC machines, a new high luminosity phase of LHC is expected to start in 2030 and to provide an integrated proton-proton luminosity of 3,000 fb^{-1}, in ten years of operation. HL-LHC will enable more precise studies of the Standard Model, probing its validity at unprecedented energy scales and with higher precision, while extending the sensitivity to potential deviations. The increased dataset will improve measurements of Higgs boson interactions and self-coupling, providing additional tests of the SM. New sensitivities to BSM physics, both through direct and indirect searches, will become accessible, enabling a deeper understanding of Higgs interactions. The Higgs sector

remains a key focus of the HL-LHC physics program. The HL-LHC is expected to operate as a "Higgs boson factory", producing about 170 million Higgs bosons and roughly 120,000 Higgs boson pairs. With this large dataset, precise measurements of Higgs production cross sections and couplings will be possible. Higgs decay channels that are currently statistically limited at the LHC will benefit from the increased amount of data available at the HL-LHC, enabling the determination of more precise uncertainties on many Higgs decay channels. Rare channels such as $H \to \mu\mu$ and $H \to Z\gamma$ will be measured with greater precision. Moreover, an enhanced precision in measuring the Higgs boson couplings is expected, as shown in Figure 1.11.

Projections of the achievable precision with HL-LHC data have been evaluated by both the ATLAS and CMS Collaborations, including estimates of cross sections for the main production modes, branching fractions for various decay channels, and coupling modifiers. Detailed projections of the HL-LHC physics program are reported in Refs. [42, 43]. A selection of relevant results is summarized below, focusing on the Higgs boson couplings to particles and to itself. These projections account for theoretical, statistical, and experimental uncertainties. The expected measurement precision is evaluated under the anticipated operating conditions of the upgraded detectors at the HL-LHC. Simulations used for these projections are based on current event generators adapted to HL-LHC conditions, and the extrapolations rely on established statistical frameworks used to assess expected sensitivities. Scaling factors account for the increased center-of-mass energy and improved detector performance (further details on the expected performance of ATLAS and CMS can be found in Refs. [44, 45]). Theoretical uncertainties are assumed to be reduced by a factor of two relative to current values, owing to higher-order calculations and improved knowledge of parton distribution functions. Statistical uncertainties are scaled from Run 2 estimates according to the expected integrated luminosity. Systematic uncertainties related to detector performance are assumed to remain similar to Run 2 levels, under the expectation that the detector upgrades will compensate for the increased radiation damage and pileup at the HL-LHC.

Projection results from the combination of CMS and ATLAS measurements are reported in Figure 1.11. Each result is represented as the total expected relative uncertainty and is further broken down into its different components: statistical, experimental, and theoretical uncertainties, under the assumption that these components are independent within each experiment [42]. Results are shown for the expected total uncertainties on Higgs boson coupling modifiers in the left panel of Figure 1.11: κ_γ (Higgs-photon), κ_W (Higgs-W boson), κ_Z (Higgs-Z boson), κ_g (Higgs-gluon), κ_t (Higgs-top), κ_b (Higgs-bottom), κ_τ (Higgs-tau), κ_μ (Higgs-muon), and $\kappa_{Z\gamma}$ (Higgs-Z-photon). It can be noted that theoretical uncertainties are expected to be the main limiting factor for most couplings, highlighting the importance of improving the precision of theoretical predictions. To reduce the effect of common systematic uncertainties, projected measurements of ratios of Higgs boson coupling modifiers are also reported. A reference parameter, $\kappa_{gZ} = \kappa_g\kappa_Z/\kappa_H$, is defined to include the high-precision determination of the $gg \to H \to ZZ$ production rate. Cross-sections in the other channels are subsequently re-expressed in terms of κ_{gZ} and a set of coupling strength ratios, defined as $\lambda_{XY} = \kappa_X/\kappa_Y$. The corresponding results are shown in the right panel of Figure 1.11. The HL-LHC will also provide increased sensitivity to studies of electroweak symmetry breaking. In particular, di-Higgs production is expected to be observed individually by both ATLAS and CMS, with a combined significance greater than 7σ. The trilinear coupling, predicted in the SM to be $\lambda = 0.13$, is expected to be measured with a precision better than 30%. Any deviation larger than about 75% from the predicted value would allow the collaborations to exclude the SM prediction, offering an important test of the Higgs sector. Given the increased sensitivity achieved by analyses recently thanks to advanced machine learning techniques for object identification, in addition to the previous assumptions, improved efficiencies in the identification of heavy flavor jets and hadronically decaying τ-leptons are assumed (more details on the different scenarios can be found in Refs. [46, 47]). The most sensitive final states have been considered for the combinations, including $bbbb$, $bb\tau\tau$, $bb\gamma\gamma$, multilepton, and $bbll$. Results are shown in Figure 1.12. The opportunities offered by the HL-LHC will significantly advance our understanding of particle physics over the coming decades, addressing crucial questions on electroweak symmetry breaking and providing potential for new BSM discoveries. Achieving these ambitious goals requires extensive technical upgrades to both the accelerator and the experiments. A discussion of the HL-LHC, and specifically the upgrades planned for CMS, is presented in Chapter 4, with particular emphasis on the concept of precision timing measurements as a novel tool for future colliders to cope with high pileup conditions.

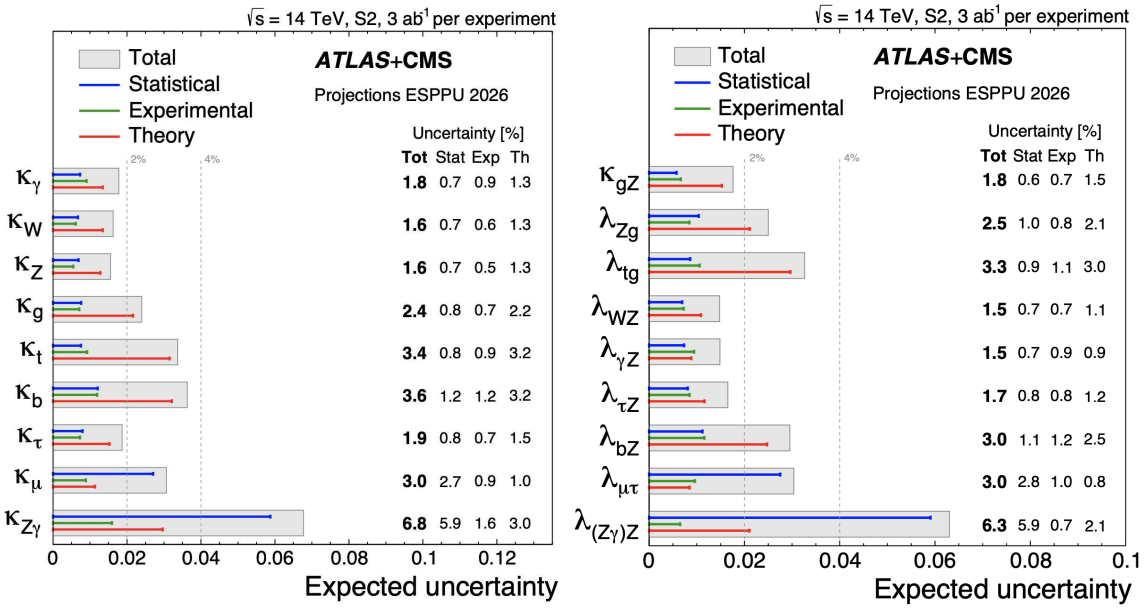

Figure 1.11: Combined expected ±1 uncertainties from ATLAS and CMS with 3,000 fb^{-1}. Each measurement represents the total uncertainty by a gray box, within which the individual contributions from statistical (blue), theoretical (red), and experimental (green) uncertainties are indicated. The expected uncertainties on coupling modifier parameters are shown on the left, and the expected uncertainties on their ratios are shown on the right [43].

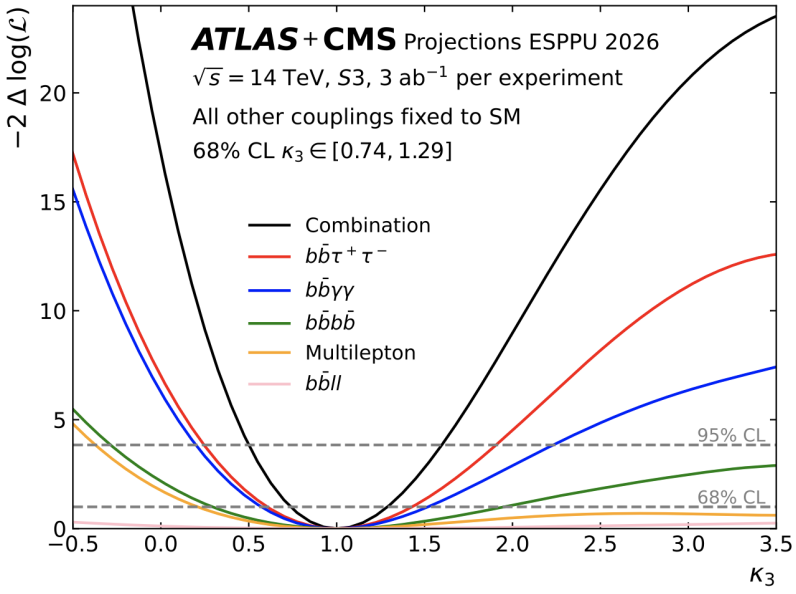

Figure 1.12: Expected κ_λ (or κ_3) likelihood scans for the combination of ATLAS and CMS projections at the HL-LHC in the most sensitive channels: $bbbb$, $bb\tau\tau$, $bb\gamma\gamma$, multilepton, and $bbll$ [43].

Chapter 2

The CMS experiment at the LHC

Fundamental physical laws have historically been investigated by probing ever smaller scales in space and time. Exploring such scales requires the development of technologies capable of reaching high energies and therefore small distance scales. In modern times, the discovery of the Higgs boson by the ATLAS and CMS collaborations, using approximately 25 fb^{-1} of data collected at the Large Hadron Collider, represented a turning point for particle physics and a milestone in humankind's understanding of nature. The LHC is a powerful machine enabling the exploration of extreme energy scales and the investigation of currently unknown phenomena. With its unprecedented collision energy and luminosity, the LHC, overviewed in Section 2.1, has enabled the collection of a large dataset across its four main experiments: ALICE, ATLAS, CMS, and LHCb. Substantial technical and computing challenges have been successfully addressed and continue to be improved upon. The LHC has established itself as a discovery machine, performing exceptionally well across a wide range of physics results and investigating fundamental questions such as the origin of elementary particle masses, the nature of dark matter [48], the unification of forces, the behavior of QCD under extreme conditions [49], and the matter-antimatter asymmetry [50].

Within this framework, the Compact Muon Solenoid experiment, together with ATLAS, is one of the two general-purpose detectors specifically designed to fully exploit the physics potential of the LHC. Following the discovery of the Higgs boson, the CMS experiment has delivered increasingly precise measurements of its properties and continues to push the limits of accuracy in di-Higgs searches using the data provided by the LHC. The CMS detector is designed to achieve excellent muon identification and momentum resolution over a wide kinematic range, precise charged-particle momentum resolution and reconstruction efficiency in the tracker, and outstanding electromagnetic energy resolution, resulting in excellent diphoton and dielectron mass resolutions, as detailed in Section 2.2. At CMS, proton bunches collide at a rate of 40 MHz, with a beam spacing of 25 ns. To extract physics results from billions of collisions per second, fast electronics, high-performance reconstruction algorithms are essential, and a highly selective triggering system is required, as discussed in Section 2.2.2. The Particle-Flow (PF) algorithm is employed to identify and reconstruct all final-state particles, as outlined in Section 2.3.1. The resulting events consist of a variety of particles traversing the detector and interacting with its sub-sections depending on their nature. Alongside the collected data, simulated events are produced using Monte Carlo (MC) techniques to model the collision process from the underlying theoretical framework up to the generation of digital signals in the CMS detectors. MC simulation plays a crucial role in most analyses, enabling data-to-simulation comparisons. It is therefore essential to ensure a good agreement between data and simulation by validating the latter using well-understood reference processes. Corrections are applied at every stage of the simulation chain, as described in Section 2.3.2, and dedicated CMS working groups provide validated corrections and studies used across all analyses within the collaboration. Simulation is employed both to model the signal process under study and to reproduce background processes arising in collisions that can obscure the signal reconstruction. Achieving the final physics result requires the development of accurate methods to discriminate signal events from background, enabling fits to be performed on discriminating variables. Finally, statistical analysis is performed on simulated and data events to evaluate the assumed model, with a brief overview given in Section 2.3.3.

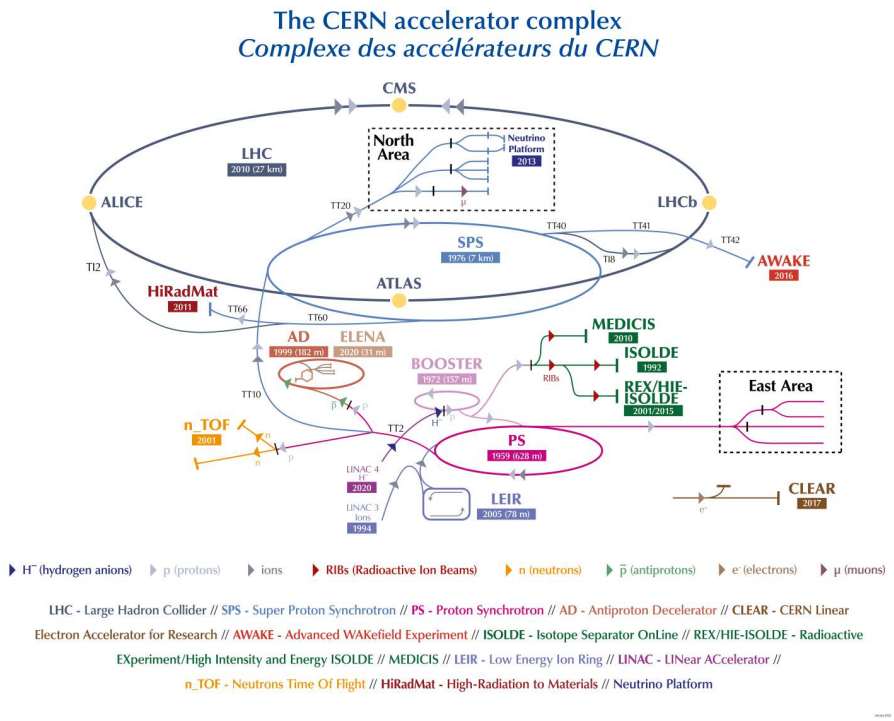

Figure 2.1: The LHC accelerator complex consists of several acceleration stages that serve both the LHC itself and a variety of other experimental programs. Each element of the chain contributes to increasing the beam energy before it is injected into the subsequent machine. The final stage takes place within the LHC ring, where the particles are accelerated to their nominal energy and brought into collision at specific interaction points, where dedicated experiments detect and record the resulting collision products [52].

2.1 The Large Hadron Collider

Following the closure of LEP, the LHC was conceived to achieve even higher center-of-mass energies in proton–proton collisions, with the aim of further exploring the high-energy frontier of particle physics. It was designed to continue direct tests of the Standard Model, most notably the search for the Higgs boson, and to investigate physics beyond the Standard Model, such as supersymmetry and dark matter candidates. The LHC is a 26.7 km hadron collider with two counter-rotating rings, constructed between 1984 and 1989 to host the former CERN Large Electron–Positron Collider [51]. The tunnel lies between 45 m and 107 m below the surface, on a plane inclined by 1.4% towards Lake Geneva. The underground and surface infrastructures at Points 2 and 8, inherited from LEP, are still in use, while Points 1 and 5, hosting ATLAS and CMS, were newly constructed for the LHC. The European Organization for Nuclear Research (CERN) approved the LHC project in December 1994, and the machine has been operating since 2009. It has reached center-of-mass energies of up to 14 TeV and remains the largest and most powerful particle accelerator ever built. Hadrons are accelerated to nearly the speed of light in two counter-rotating beams traveling in separate ultra-high-vacuum beam pipes, housed within the same tunnel. The LHC enables the study of proton–proton collisions at a center-of-mass energy of up to 14 TeV with a design luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, as well as heavy-ion collisions at a center-of-mass energy per nucleon pair of 2.76 TeV with a luminosity of about $10^{27}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Superconducting magnets guide and focus the beams, which collide at four main interaction points where the experiments ATLAS, ALICE, CMS, and LHCb are located.

The accelerator chain, illustrated in Figure 2.1, consists of a succession of accelerators that increase the energy of particles and inject them into the next machine in the chain. Until 2018, the sequence of acceleration started with LINAC 2, where protons extracted from hydrogen gas were accelerated via a radio-frequency system up to an energy of 50 MeV. During the Long Shutdown 2, LINAC 2 was replaced by LINAC 4, which became the source of proton beams and accelerated negative hydrogen ions to 160

MeV before injection into the Proton Synchrotron Booster (PSB). During injection, the ions are stripped of their electrons, leaving only protons. These are then accelerated up to 2 GeV for injection into the Proton Synchrotron (PS), where particles reach 25 GeV. Finally, particles are accelerated in the Super Proton Synchrotron (SPS), where they are brought up to 450 GeV. After the SPS, protons are sent to the LHC beam pipes, where one beam travels clockwise while the other circulates anticlockwise. Filling the LHC ring requires 4 minutes and 20 seconds, with a total of about 20 minutes needed to reach the maximum energy of 6.8 TeV via high-frequency accelerating cavities. The beams can then circulate for many hours inside the beam pipes, kept in orbit by 1232 superconducting dipole magnets distributed across eight arcs. The superconducting magnets are made of a niobium–titanium alloy and exhibit superconductivity when cooled to 1.0 K, enabling the production of a magnetic field of about 8.3 T and a current of 11 kA. The transverse dimensions of the beams are focused and reduced by quadrupole magnets, while octupole magnets are installed to control the energy spread of the beam. Approximately 2808 bunches of particles per beam are accelerated, each containing 1.2×10^{11} protons at the start of the fill (decreasing as collisions proceed). With the stabilization of beam properties and the achievement of nominal energy, the accelerated bunches are further focused by quadrupoles located before the four interaction points to squeeze the beams and increase the likelihood of collision. The collisions occurring at the four interaction points are recorded by the four experiments, where the total energy at the collision point reaches 13.6 TeV. The LHC was designed not only to accelerate and collide protons but also to perform lead and heavy-ion collisions, which will not be discussed in the present work, where the focus is on analyses with proton–proton collisions.

The performance goal of the LHC is defined both in terms of center-of-mass energy and in terms of luminosity, i.e. a quantity depending on the beam conditions that linearly affects the number of events per second generated in LHC collisions. Luminosity therefore represents a crucial parameter that determines the discovery potential of the LHC. The luminosity delivered by the LHC, L, can be expressed as in Equation 2.1, where γ is the relativistic factor, n_b the number of bunches per beam colliding at the interaction point, N the bunch population, f_{rev} the revolution frequency, β^* the beta function of the beam at the collision point (defining the focal length), ϵ_n the transverse normalized emittance, R the luminosity geometrical reduction factor due to the crossing angle, θ_c the full crossing angle between colliding beams, and σ and σ_Z the transverse and longitudinal RMS beam sizes, respectively.

$$L = \gamma \frac{n_b N^2 f_{\text{rev}}}{4\pi\beta^*\epsilon_n} R \qquad \text{with} \qquad R = \frac{1}{\sqrt{1 + \frac{\theta_c \sigma_Z}{2\sigma}}} \tag{2.1}$$

The LHC has been operating and providing collisions over three runs, alternated with long shutdowns required for collider and detector maintenance and upgrades. For instance, the delivered energy at CMS during Run 1, in the 2010-2011 and 2012 data-taking periods, was of approximately 7 TeV and 8 TeV, with 6.1 fb^{-1} and 26.3 fb^{-1}integrated luminosity, respectively. During Run 2 (2015–2018), the energy was increased to 13 TeV, and it provided 138 fb^{-1} of pp collision data to CMS. The currently ongoing Run 3 has collected 62.4 fb^{-1} in 2022 and 2023, and it has recently reached 100 fb^{-1} recorded, at a centre-of-mass energy of 13.6 TeV. Run 3 started in 2022 and is expected to deliver about 250 fb^{-1} of integrated luminosity, followed by Long Shutdown 3 (LS3), during which the experiments will undergo major upgrades in preparation for the High-Luminosity LHC, discussed in Section 4.1. The LHC plan, along with the envisaged upgrade plans of the LHC, is shown in Figure 4.1.

2.2 The Compact Muon Solenoid

Within the LHC physics program, the CMS experiment, together with ATLAS, has played a major role, providing in 2012 the observation of a new particle with a mass of 125 GeV, compatible with the Higgs boson. The data collected at the LHC have since then enabled precision measurements of many Standard Model properties, along with the exploration of hints of new physics, constraining the theory. The CMS detector has been designed as a general-purpose experiment, investigating a wide variety of processes produced in hard scattering interactions and exploring phenomena from electroweak to Higgs, top-quark, and beyond-the-Standard-Model physics, as well as flavor physics. During LHC operation, CMS has demonstrated the ability to operate both as a discovery machine, observing the Higgs boson and new

processes such as vector boson scattering, and as a precision machine, enabling accurate measurements of cross sections for many processes.

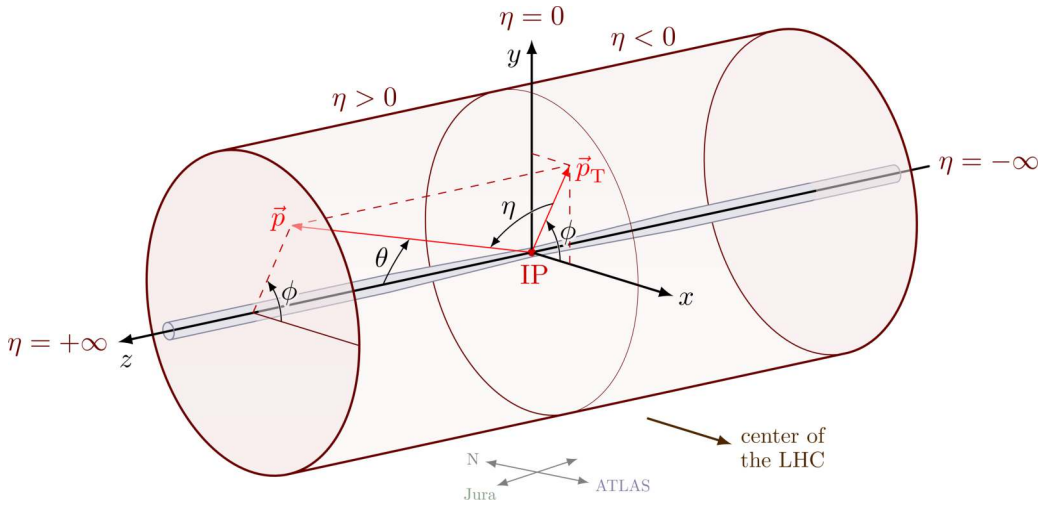

Figure 2.2: A graphical representation of the CMS coordinate system generated using a public image generator [25], where the dimensions associated with the CMS detector are shown. In particular, the figure illustrates the pseudorapidity and its directions, as well as the definitions of the z and ϕ axes with respect to the CMS orientation and the interaction point.

CMS is designed to cover the entire solid angle around the collision point, using a Cartesian coordinate system as a reference, as shown in Figure 2.2. The collision point is defined as the origin of the coordinate system. The y-axis points vertically upward, the x-axis points radially inward toward the centre of the LHC ring, and the z-axis points along the direction of the counterclockwise proton beam when viewed from above the LHC. The azimuthal angle ϕ is measured in the (x, y) plane from the x-axis, with the radial coordinate denoted as r. The polar angle θ is defined in the (r, z) plane from the z-axis. The pseudorapidity (η) is commonly used as a reference and is defined as in Equation 2.2.

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \tag{2.2}$$

The angular variable R is also commonly used to define the distance between objects, as shown in Equation 2.3, where $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and azimuthal angle between the objects.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.3}$$

Based on this reference definition, some useful variables are defined as follows: the component of the momentum transverse to the z-axis is denoted as p_T and the transverse energy is defined as $E_T = E \sin\theta$. The missing transverse momentum, $\vec{p}_{\text{miss}}$, is the vector sum of the transverse momenta of undetectable particles.

The main features of the detector have inspired its name: with its 15 m height, 21 m length and 14,000 tons, CMS is a remarkably compact experiment compared, for instance, to ATLAS, which has a length of 46 m, a diameter of 25 m and a weight of 7,000 tons. At the heart of the detector, a solenoidal magnet with a 6 m internal diameter generates a strong magnetic field of about 3.8 T. Due to this powerful magnetic field and the dedicated muon detection system, CMS is capable of high-precision muon identification and momentum measurement. The bore of the magnet is large enough to contain both the tracking and calorimetry systems, each consisting of a barrel section covering the central pseudorapidity region and two endcap sections. This configuration minimizes the amount of material in front of the calorimeters, therefore reducing the energy losses due to particles showering and facilitating the PF reconstruction. The electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL) cover up to a pseudorapidity of approximately 3. A hadron forward (HF) calorimeter extends the pseudorapidity coverage. The overall design, shown in Figure 2.3, consists of several layers of subdetectors. Events are selected using a two-

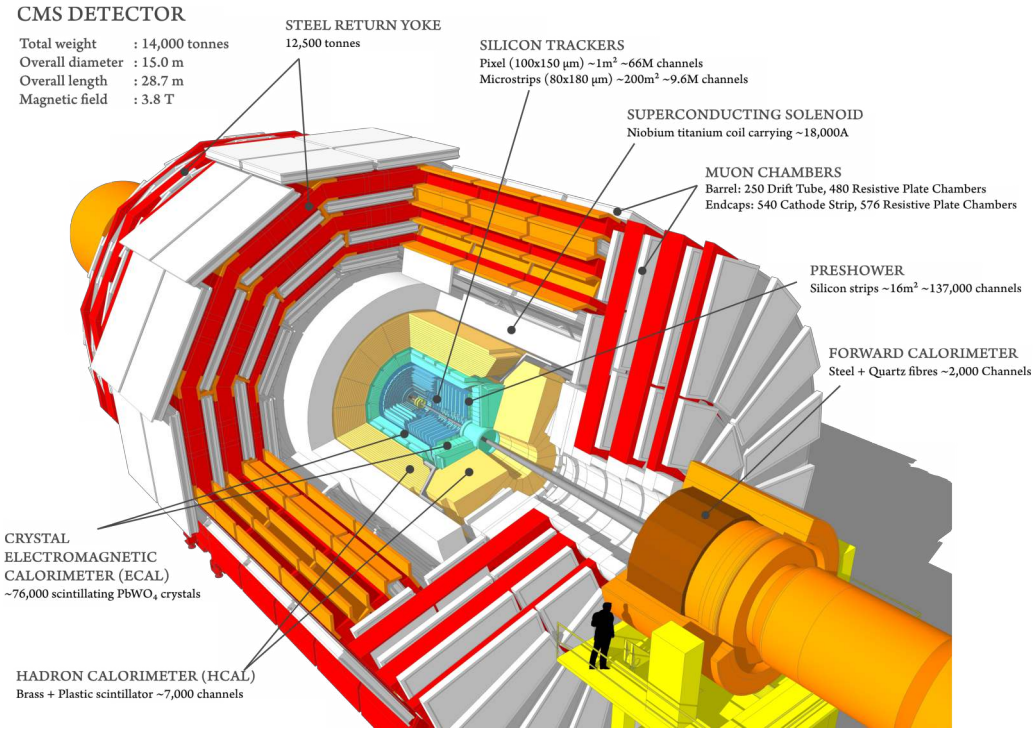

Figure 2.3: Representative visualization of a cross section of the CMS experiment. The detector provides full hermetic coverage of the solid angle around the collision point, where proton-proton interactions take place. General properties of the detector are shown, and arrows indicate all the subdetectors of the experiment [53].

tier trigger system, as described in Section 2.2.2. The first-level trigger consists of custom hardware processors selecting events at a rate of about 100 kHz, while the second-level trigger runs optimized object reconstruction algorithms for fast processing, delivering an event rate of about 1 kHz.

2.2.1 The CMS design

The CMS detector was designed with the main goals of providing excellent muon identification and momentum resolution, achieving a di-muon mass resolution of about 1% at 100 GeV and an unambiguous determination of the muon charge for momenta up to 1 TeV. The design also required a set of other performance targets: good momentum resolution and reconstruction efficiency for charged particles in the inner tracker, electromagnetic energy resolution at the percent level around 100 GeV, wide geometric coverage, strong π^0 rejection, and effective isolation of photons and leptons at high luminosities. In addition, precise measurements of the missing transverse energy and di-jet mass resolution demanded large hermetic coverage and fine granularity in the hadron calorimeters. A comprehensive description of the CMS design can be found in Ref. [54], with details on recent developments for LHC Run 3 provided in Ref. [55]. An overview of the main detector components is presented below.

At the center of the CMS design lies the superconducting solenoid, made of NbTi, which provides a magnetic field of 3.8 T, ensuring strong bending power for charged particles. The first detector layers traversed by particles produced at the interaction point form the tracking system. The tracker is designed to measure the transverse momentum p_T of particles down to 50 MeV within a pseudorapidity region of $|\eta| < 2.5$, offering the granularity and precision needed to reconstruct tracks in high-multiplicity environments. The tracker is a cylindrical detector with an outer radius of 1.20 m and a length of 5.6 m, composed of two subsystems: the inner pixel detector and the outer silicon strip tracker. Its layout has evolved over the various LHC runs, with different numbers of active layers in the barrel and endcap regions, as shown in Figure 2.4. The pixel detector currently consists of four barrel layers (BPIX) and three disks on each end (FPIX). Each pixel sensor contains 160 × 416 pixels with dimensions of 100 × 150 μm^2, resulting

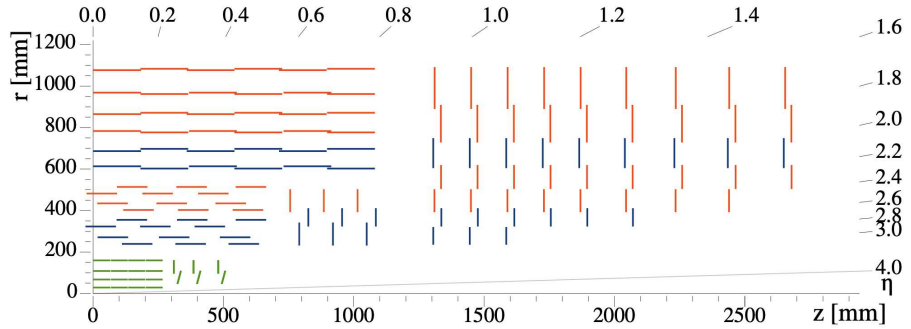

Figure 2.4: Representation in the $r - z$ plane of the components of the pixel layers, represented in green, and the strip detector represented in red and blue [55].

in a total of about 124 million readout channels for both regions. The achieved position resolution is approximately 11 μm in the r–ϕ direction and 20 μm in z (with slightly better performance in the outer layers due to lower radiation damage). Together with the pixel system, the silicon strip tracker provides track measurements for $|\eta| < 2.5$, with ten layers in the barrel region (the Tracker Inner Barrel, TIB, and Tracker Outer Barrel, TOB), three inner disks at each end (TID), and the Tracker Endcaps (TEC). Sensors with a thickness of 320 μm are used in the TIB, TID, and part of the TEC, while 500 μm thick sensors are used in the TOB and the remaining TEC modules. Silicon pixel sensors in the barrel region, together with two forward disks, seed track reconstruction and enable the measurement of impact parameters such as secondary vertices, which are crucial for heavy-flavor object identification (see Section 2.3.1).

The electromagnetic calorimeter (ECAL) covers pseudorapidities up to $|\eta| < 3$ and uses lead tungstate ($PbWO_4$) crystals as the active sensor in both the barrel and endcap regions. The barrel covers $|\eta| < 1.479$, while the endcaps cover $1.479 < |\eta| < 3.0$. The light produced by the scintillators is read out with avalanche photodiodes (APDs) in the barrel and with vacuum phototriodes (VPTs) in the endcaps. The crystals provide a radiation length of approximately 25 X_0, sufficient to absorb more than 98% of the energy of 1 TeV electrons and photons. Preshower detectors are installed in front of the endcaps and consist of two planes of silicon sensors interleaved with lead layers. The ECAL provides excellent electron and photon identification, measuring their positions and energies through clustering algorithms, and also contributes to jet and missing transverse momentum reconstruction. During LHC operations, the ECAL performance has been closely monitored due to the reduction in light transmission of the crystals caused by radiation exposure. To monitor transparency variations and ensure the correct calibration of ECAL signals, a laser monitoring system has been operated continuously. Performance studies during Run 2 showed a stable energy scale and an electron energy resolution ranging from 1.6% to 5% for electrons with $p_T \approx 45$ GeV from $Z \to e^+e^-$ decays. The ECAL also provides timing information for electrons and photons, with a time resolution of $\mathcal{O}(150)$ ps for energies of $\mathcal{O}(50)$ GeV.

The hadron calorimeter (HCAL) measures the energy of charged and neutral hadrons and plays a key role in reconstructing jets and missing transverse momentum. It is composed of four main subdetectors: the Hadron Barrel (HB), Hadron Endcap (HE), Hadron Outer (HO), and Hadron Forward (HF) calorimeters. The HB and HE are housed inside the solenoid, covering pseudorapidities up to $|\eta| < 3$, and are made of brass plates interleaved with layers of scintillating tiles. The HF is located outside the solenoid, covering the pseudorapidity range $3 < |\eta| < 5.2$, and is made of steel with embedded quartz fibers. The HO consists of plastic scintillators and covers the region $|\eta| < 1.26$. A schematic view of the HCAL sections is shown in Figure 2.5. Light collected from layers grouped into "depth" segments is read out by a single photodetector. The hybrid photodiodes that originally received light from clear plastic fibers were later replaced with silicon photomultipliers (SiPMs), improving radiation tolerance and enabling finer segmentation. The segmentation follows a tower structure in η–ϕ space. The performance of the HCAL, when combined with information from other detectors, yields a jet energy resolution ranging from 15–20% at 30 GeV down to approximately 5% at 1 TeV. For pions with momenta between 40 and 60 GeV showering in the HCAL, the energy resolution was measured to be 19.4%, 18.8%, and 23.6% in the HB, HE, and transition regions, respectively. Continuous monitoring is performed using a laser

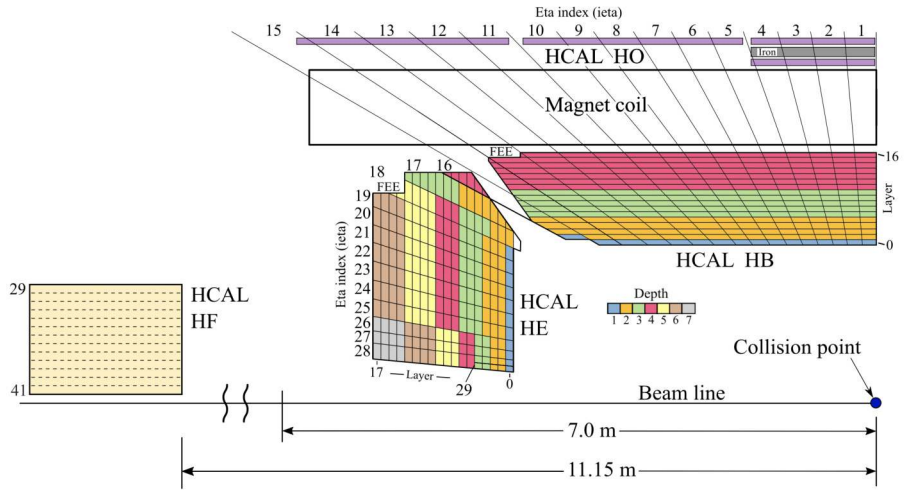

Figure 2.5: Longitudinal and transverse sections of the HCAL design for operation during LHC Run 3 [55].

calibration system and by studying the response to a ^{60}Co source to evaluate signal loss due to radiation damage in the HE, where the effect is most pronounced.

A key component of the CMS experiment is the muon system, which provides powerful triggering and detection capabilities for muons. The CMS muon system consists of four detector technologies located within the steel flux-return yoke of the solenoid. These four subsystems are drift tubes (DTs), cathode strip chambers (CSCs), resistive plate chambers (RPCs), and the recently added gas electron multipliers (GEMs). A schematic representation of the overall design is shown in Figure 2.6.

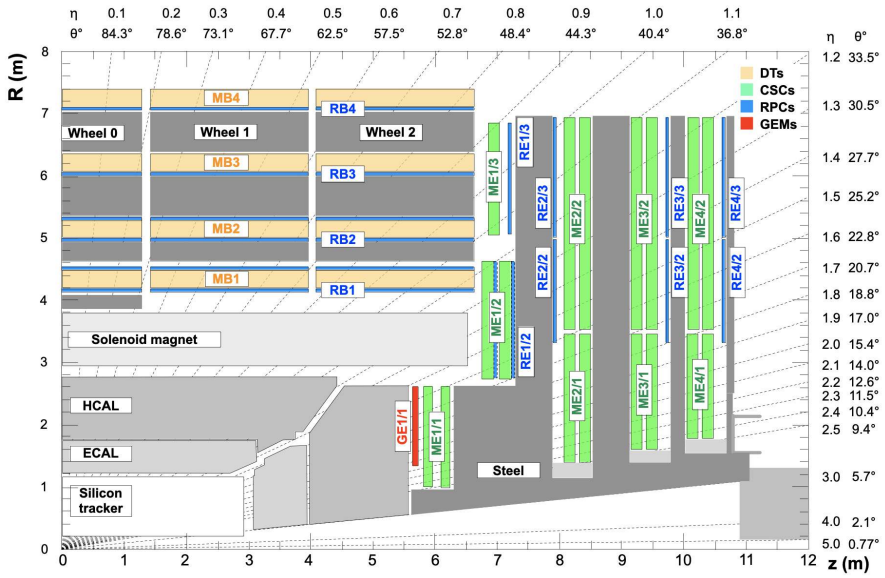

Figure 2.6: Schematic view of the CMS muon system in the r–z plane. DTs are shown in orange and labeled MB, CSCs in green and labeled ME, RPCs in blue and labeled RB and RE, and GEMs in red and labeled GE. The grey areas represent the magnet yoke [55].

The DTs are chambers filled with an Ar/CO_2 gas mixture, with an anode wire at the center of each tube and cathodes at the tube boundaries. The electric field generated inside the tube produces a uniform drift of ionization electrons created by the passage of charged particles through the gas. The particle trajectory is reconstructed from the signal collection times on the wires. The CSCs are filled

with a gas mixture of Ar, CO_2, and CF_4, with signals induced on anode wires and cathode strips. The high segmentation provided by the strips, combined with fast readout electronics, ensures precise timing and excellent spatial resolution for muon identification. The RPCs consist of two high-pressure laminate plates separated by a gas mixture of $C_2H_2F_2$–i-C_4H_{10} and SF_6. The readout strips are located between the two plates, and a high voltage is applied to electrodes coating each plate. This detector provides excellent time resolution, ensuring accurate muon timing and bunch-crossing identification. Finally, the GEMs consist of perforated insulating polymer foils located between conductors. The gas mixture used in these chambers is Ar/CO_2, and a voltage applied across the foils generates a strong electric field within the holes. Electrons produced by ionization are accelerated through the holes and collected on segmented readout strips, providing precise spatial resolution and fast timing.

2.2.2 The CMS trigger system

With the unprecedented amount of data produced by the LHC experiments, exceeding 10^7 readout channels operating at a 40 MHz frequency, the design of the trigger and data acquisition (DAQ) systems must account for physics requirements as well as technological and financial constraints. Flexibility, modularity, and redundancy are critical criteria in designing an effective trigger system that can maintain the same speed of the rapid evolution of technology, adapt to changing operational conditions such as beam background, instantaneous luminosity, or detector noise, enable efficiency measurements directly from data, and accommodate evolving physics priorities. The CMS trigger system [56] employs two levels to select events of potential physics interest. The first level (L1), implemented in custom hardware, selects events in which a candidate object is found (e.g. ionization deposits consistent with a muon or an energy cluster consistent with an electron or photon). On top of this, a second-level trigger, the High-Level Trigger (HLT), selects events according to a programmable set of algorithms that build on the objects identified at L1. The L1 thresholds are configured to reduce the event rate to about 100 kHz, while the HLT further refines the selection, reducing the output rate to approximately 400 Hz for offline storage. Prescale factors can be applied to events passing specific selection criteria to regulate the total output rate of both the L1 trigger and the HLT. The set of object definitions and algorithms used to make trigger decisions defines the *trigger menu*. In addition to selecting and collecting collision data, the trigger and DAQ systems are also responsible for recording information used for detector monitoring and calibration, although these data volumes are typically small compared to the physics event data.

The L1 trigger is a hardware-based system with fixed latency. Within about 4 μs of a collision, it decides whether an event should be accepted or rejected, based on information from the calorimeter and muon detectors. The elementary inputs, called *trigger primitives*, are generated by the ECAL, HCAL, DT, CSC, and RPC systems. Each trigger primitive is evaluated via custom-built electronics and field-programmable gate arrays (FPGAs). Trigger information from the calorimeters and muon detectors is first processed in parallel and then combined in the Global Trigger, which makes the final decision. This decision is then sent to the tracker, calorimeters, and muon systems via the trigger, timing, and control system, at which point the data acquisition system reads out data from the subsystems for offline storage. The Level-1 calorimeter trigger evaluates the transverse energies and quality flags from more than 800 ECAL and HCAL trigger towers, i.e. arrays of detector sensors, covering the pseudorapidity range up to $|\eta| < 5$. It processes this information and produces as output e/γ candidates, which are indistinguishable at this stage, as well as regional transverse energy sums based on groups of 4×4 towers. These e/γ candidates are further sorted, jets are identified using the E_T sums, and global quantities such as missing transverse energy are computed. The output consists of e/γ candidates, central jets, τ-jets and forward jets, along with several global quantities. The L1 muon trigger uses information from all three muon detector systems. Hit information is collected in multiple detector planes, and dedicated pattern recognition algorithms are used to identify muon candidates and measure their momenta from the curvature in the magnetic field. Muon candidates are identified and sorted by the three regional track finders of the DT, CSC, and RPC systems, which also assign transverse momentum, quality codes, and (η, ϕ) coordinates. These candidates are then merged and further filtered with quality requirements to reject low-quality signals.

The list of objects produced by these steps, electrons/photons, muons, jets, and τ leptons, is then passed to the Global Trigger, which constitutes the final stage of the CMS L1 trigger system. The

Global Trigger applies a set of selection criteria to the final list of objects, ranging from simple transverse energy requirements to coincidences of multiple objects with specific topological conditions. These algorithms combine kinematic requirements with basic logical operations (`AND`, `OR`, `NOT`). Ultimately, the Global Trigger uses the information provided by the calorimeter and muon systems, such as transverse energy or momentum, position, and object quality, to make its decisions on whether to accept or reject each physics event at every LHC bunch crossing. Prescale factors are adjusted dynamically during data-taking to maintain a roughly constant trigger rate of about 100 kHz. The L1 trigger menu and prescales are continuously adjusted to maximize signal acceptance under evolving operating conditions. For example, as the transparency of the ECAL crystals decreases over time due to radiation damage, energy corrections are periodically computed and applied to maintain high trigger efficiency and low trigger thresholds. During Run 1, the L1 trigger operated with 128 different selections (seeds), which later increased to 400–500 seeds in Runs 2 and 3, accompanied by an expansion from a single FPGA to three and then six, thereby enabling a more demanding physics program. In Run 3, the L1 trigger rate was increased to up to 110 kHz, and several algorithmic improvements were implemented, including the identification of long-lived particle jets, the measurement of muon displacements from the primary interaction point, and the handling of high hit multiplicities. In addition, new data-scouting dedicated hardware was introduced, representing the first large-scale demonstration of a broader system planned for the Phase-2 upgrades. Data scouting is a novel technique that uses trigger-level reconstruction for physics analyses, reducing the event data format to the minimum necessary to perform analyses without saving the full raw event.

The physics objects produced as output of the L1 trigger, together with tracker information, are subsequently processed by the High-Level Trigger. For each event, the HLT applies reconstruction and identification criteria to select only those events that might be of interest for analyses. The software computing farm, located in the surface building at LHC Point 5, executes the filtering process. This consists of a reconstruction sequence similar to the offline one (discussed in Section 2.3.1), based on HLT paths that form the HLT menu: sets of algorithms executed in a fixed order that reconstruct physics objects and apply selections of increasing complexity and refinement. Selections using calorimeter and muon information are applied first, followed by more computationally expensive reconstructions using tracking information. The selected events are then sent to the storage manager for archival and transfer to the CMS Tier-0 computing center for offline processing and permanent storage. Most data are processed as soon as possible, except for parked data, which consist of lower-priority samples that are not analyzed until the end of the run. More specifically, data parking and data scouting are strategies developed to overcome the limited size of the so-called raw data storage at Point 5 and the constraints in DAQ bandwidth, which impose restrictions on the data volume and computation required for storing interesting events. In the case of data parking, computing resources are used to reconstruct those events at a later time, when resources are more readily available. These approaches address the limits of real-time processing and open new possibilities in many analyses where object reconstruction is constrained by trigger requirements[1].

For Run 3, the HLT has started using graphical processing units (GPUs) in the trigger filter farm, which, together with developments in computing and parallelization of reconstruction algorithms, has reduced event processing time by about 40% [58]. HLT tracking efficiency has also been significantly improved via new algorithms. The identification of jets originating from b-quark decays has undergone substantial improvement thanks to the use of new neural network taggers. For instance, the ParticleNet tagger, which provides a multinomial jet-flavor classification for categories of b, c, light quarks, gluons, and hadronically decaying τ-leptons, significantly reduces the light-flavor jet misidentification rate, as shown in Figure 2.7, compared to the DeepCSV tagger used in Run 2. A relevant example of the use of these triggers is in di-Higgs searches. Building on more performant heavy-flavor tagging, new trigger strategies have been developed at the HLT level to target HH production. The trigger applies a selection on the jet transverse momentum of 280 GeV, requires at least four jets with $p_T > 30$ GeV, and imposes an average ParticleNet probability of at least 0.55 for the two b-jets to be identified as b-jets. Data parking is exploited to increase both the rate and the acceptance, providing valuable opportunities for HH analyses. In 2023, the parking strategy was extended by improving the purity of the dielectron

[1] A comprehensive review of data parking and data scouting can be found in Ref. [57]

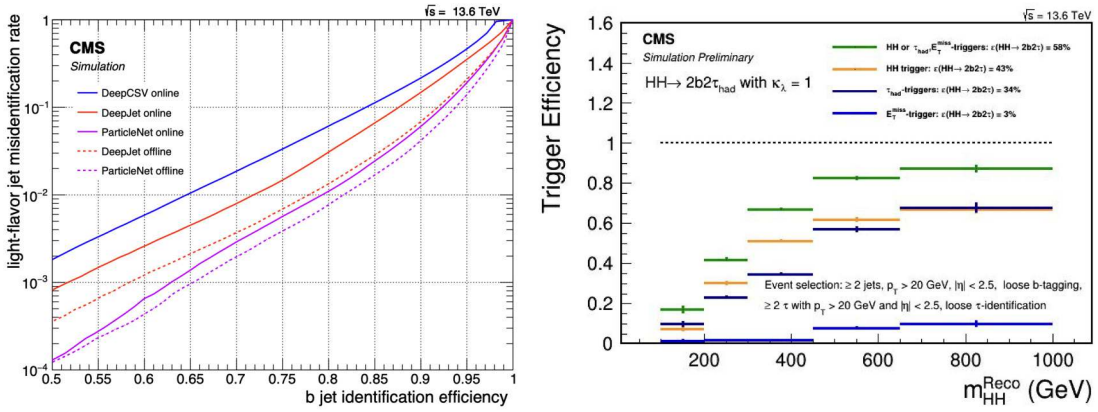

Figure 2.7: Left: The light-flavor jet misidentification rate is reported as a function of the b-jet identification efficiency when using different jet taggers. The solid curves demonstrate the performance of the algorithms in the HLT, while the dashed curves show the corresponding performance when using offline reconstruction and training. The DeepJet algorithm brings a reduction in jet misidentification rate by a factor of 3 compared to DeepCSV, i.e. an algorithm used for Run 2, while ParticleNet brings an additional factor of 2.5 reduction [55]. Right: The trigger efficiency as a function of the reconstructed mass of the di-Higgs system is shown for the simulated Standard Model process of $HH \rightarrow bb\tau\tau$. The Run 3 hadronic triggers are shown in blue, the dedicated HH trigger for a sample of 27.8 fb^{-1} in orange, and the combination of both triggers in green. The efficiency achieved with the Run 3 triggers, compared to triggering on the missing transverse energy only, is 43%, while the use of only τ-lepton triggers yields 34%. The use of both triggers together with the missing transverse energy reaches a trigger efficiency of 58% [59].

triggers, allowing CMS to trigger also on events with two b-tagged jets, Higgs boson production via vector boson fusion, and long-lived particle (LLP) signatures. For instance, in the HH→bb$\tau\tau$ channel, the trigger efficiency as a function of the reconstructed invariant mass of the HH system for the simulated SM process has been studied using different Run 3 triggers, addressing either hadronic decays of τ-leptons or HH processes or a combination of both [59], as shown in the right panel of Figure 2.7. The new VBF parking triggers have, for example, been employed in the non-resonant search for HH production in the $bb\tau\tau$ final state, as discussed in Section 3.2.

2.3 CMS physics analysis workflow

The raw data selected by the trigger system are subsequently processed by dedicated reconstruction algorithms, designed to target each physics object differently. Within the CMS experiment, the particle-flow approach is employed, reconstructing particles from the primary vertex, where the hard collision occurs, up to the signals detected by the various sub-detectors. By exploiting the expected interactions of each object with the different detector layers, customized techniques are used to precisely identify every particle. In recent years, with the development of increasingly advanced machine learning techniques, reconstruction methods have been enhanced through the use of new algorithms. In addition to selecting events from collisions, the data analysis is integrated with the production of extensive Monte Carlo simulations, which model the interactions of particles produced in collisions within the CMS detector. Precise calibration of these simulated events is essential to accurately describe many processes and to perform high-precision measurements of particle and process properties. Ultimately, data are compared with simulation, and statistical analyses are performed on the observed and predicted distributions to extract the desired measurements.

2.3.1 Physics objects reconstruction

Every collision event can be decomposed into a subsequent set of objects. Collisions occur at the beam interaction region, and the produced particles first traverse the tracker, where charged particles leave a

signature (*tracks*). Their origin points (*vertices*) are reconstructed from signals in the sensitive layers (*hits*). The particle identification through the detector layers is illustrated in Figure 2.8, where each particle exhibits a characteristic behaviour inside CMS under the influence of the magnetic field and depending on the nature of its interaction.

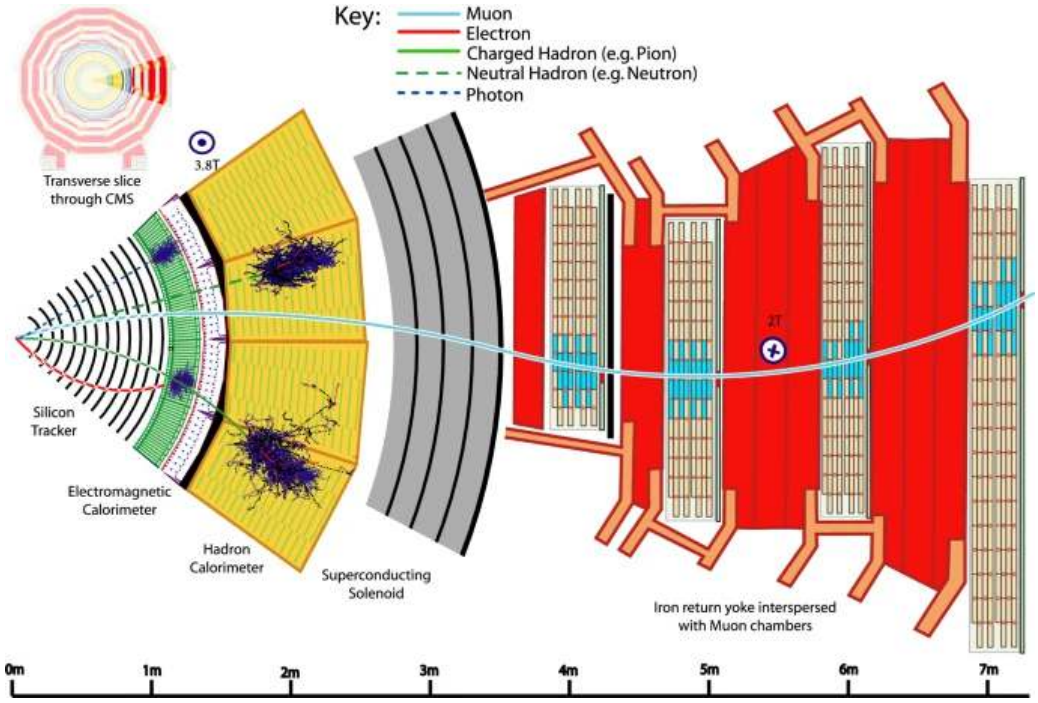

Figure 2.8: Illustration of particle trajectories in a slice of the CMS detector, from the interaction point to the muon system. For each particle, the path traversed is shown depending on its nature [60]. Electrons leave a track in the tracker and are absorbed in the electromagnetic calorimeter, while photons do not interact with the tracker and are absorbed in the ECAL. Muons do not interact significantly with the tracker or the calorimeters, but are eventually detected in the muon chambers. Hadrons are detected via their absorption in the hadron calorimeter, with charged hadrons also leaving a track in the tracker.

The trajectories of charged particles are bent by the magnetic field produced by the solenoid, enabling the measurement of charge and momentum. Electrons and photons are absorbed in the ECAL, where their energy deposit is reconstructed as a *cluster* of energy across neighboring cells. Charged and neutral hadrons may initiate a shower in the electromagnetic calorimeter and are fully absorbed in the HCAL, where clusters are then reconstructed. Muons and neutrinos traverse the calorimeters almost unaffected, as they interact only weakly with the detectors. Muons eventually leave hits in the tracking layers of the muon system, while neutrinos do not leave any detectable signature. Muons are the only particles able to reach the outermost muon detector system and being detected, and their identification relies on information from this sub-system. Isolated[2] photons and electrons are primarily reconstructed from the information collected in the ECAL, while jets, i.e. sprays of particles (mainly hadrons and photons) originating from quark or gluon fragmentation, are reconstructed from energy deposits in the calorimeters. Jets originating from heavy-flavor quarks (e.g. b-quarks or c-quarks) or from hadronic τ decays are tagged based on the properties of particle tracks, and thus also require information from the tracker. An event can therefore be described by coherently collecting all the information provided by basic interactions in each detector layer (such as tracks and clusters), and by combining them to reconstruct the properties of the particles. This approach is known as particle-flow reconstruction [60]. It starts from a reconstruction of tracks and clusters and later combines and links this information to reconstruct physics objects, such as electrons, muons, jets, and missing transverse energy.

[2]Photon isolation is defined by considering the energetic activity within a cone centered on the photon, as defined in Equation 2.4.

A track is a signature left by a charged particle in the tracking system as it traverses the detector. However, thousands of charged particles are expected to cross the tracker at each bunch crossing, posing a significant challenge for efficiently reconstructing tracks while keeping the fraction of fake tracks low. Fake tracks can arise either from a random combination of unrelated hits or from a genuine trajectory that is incorrectly reconstructed due to spurious hits. The tracking logic used in CMS is called the Combinatorial Track Finder (CTF) [61], which is an adaptation of the combinatorial Kalman filter, itself an extension of the Kalman filter pattern recognition approach [62]. Track reconstruction is performed through multiple iterations of the reconstruction sequence (*iterative tracking*). The initial phase focus on tracks that are easier to identify, such as high-p_T tracks or tracks close to the interaction region. After each iteration, hits associated with tracks reconstructed in the previous iteration are removed from the overall collection, reducing the combinatorial complexity for subsequent iterations. This method, however, produces a large number of track candidates, among which several fake tracks. Therefore, quality criteria are also applied to keep good tracks while rejecting fake ones. A more detailed description of the various steps can be found in Ref. [61].

Clusters consist of the energy deposit in several neighboring cells of the calorimeter, and a specific algorithm was developed for their reconstruction. Every sub-detector has its own clustering: ECAL, HCAL, and preshower clustering. It starts with the identification of cluster seeds, where cells with an energy larger than a specific threshold and larger than their neighbors are considered. Topological clusters are then formed around these seeds, exploiting the expected spatial patterns of energy deposits in the crystals. Various algorithms have been used to reconstruct the final clusters, with more recent approaches employing machine learning techniques, while still leveraging the topological characteristics of the energy deposits The link between tracks and clusters is then performed via an algorithm that connects the elements based on their nature, and it defines PF blocks that are then used to identify objects. A brief description of the objects reconstruction is provided below.

2.3.1.1 Primary vertex reconstruction

In every collision event, it is crucial to identify the origin of the tracks, i.e. the primary or secondary vertex. Vertex reconstruction proceeds by first selecting the tracks, clustering them into groups that appear to originate from the same interaction vertex, and then fitting the position of each vertex using the track parameters. Track clustering is performed using a deterministic annealing algorithm [63], a method for finding the global minimum of a problem with many degrees of freedom. The idea is analogous to reaching a state of minimal energy through successive reductions in temperature (hence the name). Using a free-energy function, the algorithm searches for vertex positions that maximize the compatibility between tracks and potential vertices. This procedure is not robust against outliers, such as secondary vertices or mis-measured tracks. Therefore, additional conditions, such as cutoffs, are applied to prevent the generation of fake vertices, and an adaptive fit is performed on the remaining vertices to assign weights reflecting their quality. Finally, a quantity strongly correlated with the number of tracks consistent with originating from the interaction region is defined, which can be used to select true proton-proton interactions. The primary vertex associated with the proton-proton collision is identified as the one with the largest number of associated tracks and the highest sum of transverse momentum, reducing the probability of misidentifying a pileup vertex. Secondary vertices are typically due to the decays of heavy-flavor hadrons, τ-leptons, etc. and they are found among the tracks not belonging to the primary vertex and pointing to a location significantly displaced from it. Pileup vertices are identified as the remaining vertices distributed along the z-axis and can be used to remove pileup tracks from computations and isolation, thus reducing contamination in the reconstruction of physics objects.

2.3.1.2 Muon reconstruction

Muons are reconstructed by combining information from the inner tracker and the muon chambers. Three types of muon objects are defined: *standalone muons*, *global muons*, and *tracker muons*. Hits in the CSC, DT, and RPC detectors are first clustered into track segments. These segments are then combined and fitted to form a complete trajectory in the muon system, referred to as a standalone muon. Standalone muons suffer from lower efficiency due to contamination from cosmic-ray muons. A global muon is obtained when a standalone muon is found to be compatible with a track reconstructed in the inner tracker.

Tracker muons are instead defined starting from tracks reconstructed in the inner tracker, which are extrapolated to the muon system. If the extrapolated track is matched to at least one compatible segment in the muon detectors, the track is identified as a tracker muon. Thanks to the high efficiency of both the tracker and muon systems, about 99% of muons traversing CMS are reconstructed either as global or tracker muons. At low momenta, tracker muon reconstruction is more efficient, since it requires only a single muon segment, while global muon reconstruction achieves higher efficiency at higher momenta. If the same tracker track gives rise to both a tracker muon and a global muon, the two are merged into a single candidate.

Muon identification is applied through a set of selection criteria designed to ensure good isolation and reject contamination from hadrons. These criteria include requirements on the track fit χ^2, isolation, and spatial matching between tracks. For isolated global muons, isolation is computed within a cone of $\Delta R = 0.3$, requiring the scalar sum of transverse momentum and transverse energy within the cone to be less than 10% of the muon transverse momentum. Non-isolated muons are subject to additional quality requirements, aimed at reducing contamination from high-p_T hadrons misidentified as muons and from incorrect associations between tracker and standalone tracks. This muon identification strategy is designed to reconstruct prompt muons, such as those originating from Z or W boson decays, as well as muons from heavy-flavor hadron decays and from light hadrons, while minimizing the probability of misidentifying charged hadrons as muons. Identification working points (loose, medium, and tight) are defined by applying increasingly stringent requirements on variables such as the track fit χ^2 and the alignment between tracker and standalone tracks. These working points correspond to higher purity at the expense of reduced efficiency, and the choice of working point depends on the needs of the analysis. In analyses, muons are usually also required to be compatible with the primary vertex, with $\Delta_{xy} < 0.045$ cm in the transverse plane and $\Delta_z < 0.2$ cm along the longitudinal axis (see Figure 2.2 for CMS coordinates). Isolation requirements may also be applied to reduce contamination from muons coming from semileptonic decays of b or c quarks inside jets. The isolation parameter is defined as in Equation 2.4, where $\sum p_T^{\text{charged had.}}$, $\sum p_T^{\text{neutral had.}}$, and $\sum p_T^{\gamma}$ are the scalar sums of the transverse momenta of charged hadrons, neutral hadrons, and photons from the primary vertex, respectively, while $\sum p_T^{\text{PU}}$ is the scalar sum of the transverse momenta of charged hadrons from pileup interactions.

$$I_{rel}^{\ell} = \frac{\sum p_T^{\text{charged had.}} + \max\Big(0,\ \sum p_T^{\text{neutral had.}} + \sum p_T^{\gamma} - \frac{1}{2}\sum p_T^{\text{PU}}\Big)}{p_T^{\ell}}, \qquad \text{with} \quad \ell = \mu, e \tag{2.4}$$

2.3.1.3 Reconstruction of electrons and photons

Electrons produced in collisions first traverse the tracker, where part of their energy may be emitted as bremsstrahlung photons, and then reach the ECAL, where they are absorbed and reconstructed as clusters. To account for the full trajectory through the detector layers, the energy deposited in the ECAL alone is not sufficient and information from bremsstrahlung photons outside the main calorimeter cluster must also be included. Clusters in the ECAL, together with those from bremsstrahlung photons, are grouped into a *supercluster*, which has a limited extension in η and a wider window in ϕ around the electron direction. However, when an electron is inside a jet, its energy and position may be biased by contamination from nearby particles, leading to inefficiencies. To mitigate this, additional isolation requirements are applied, using a threshold on the isolation parameter defined in Equation 2.4.

Electron reconstruction starts from track-based seeding: electrons traversing the tracker leave different types of tracks depending on whether they radiated photons or not.[3] A projection of the track to the ECAL helps identify the full trajectory, but this is not always straightforward. Therefore, a Gaussian-Sum Filter (GSF) is used to reconstruct the electron track, as it models the energy loss as a sum of Gaussian distributions, making the reconstruction more accurate than with the Kalman Filter. To refine the selection, a multivariate analysis (MVA) based on a boosted decision tree (BDT) is employed. This combines information such as the number of hits, fit quality, energy loss, and the distance between the extrapolated track and the closest cluster. The resulting reconstruction provides a significant gain in efficiency, particularly for electrons within jets, while still offering a modest improvement for isolated

[3]The fact that they radiate also helps to distinguish electrons from charged hadrons.

electrons. Electron and photon candidates are then reconstructed differently: electrons are identified from a GSF track associated with an ECAL cluster, while photons are reconstructed from a supercluster with no associated track. In both cases, an isolation requirement is applied, namely that the HCAL energy within a region close to the supercluster does not exceed 10% of the ECAL supercluster energy. In analyses, electrons are also required to be compatible with the primary vertex, with $\Delta_{xy} < 0.045$ cm in the transverse plane and $\Delta_z < 0.2$ cm along the longitudinal axis.

After the reconstruction of muons, electrons, and isolated photons, what remains in the PF block is identified as non-isolated photons or hadrons. These may be reconstructed as charged hadrons, neutral hadrons, or non-isolated photons (e.g. from π^0 decays). Non-isolated photons and neutral hadrons are identified as clusters without any associated tracks in ECAL and HCAL, respectively. These objects are then grouped and reconstructed as jets.

2.3.1.4 Jet reconstruction

In the hard scattering process, quarks and gluons are produced and hadronize, as they do not exist as free states. This hadronization occurs within the CMS detector, resulting in particles clustered into cone-shaped objects called jets. Jets are reconstructed using the ParticleFlow algorithm in combination with the anti-k_t algorithm [64], which is designed to separate isolated particles from jet constituents. This algorithm is both infrared-safe, meaning that adding an infinitesimally small amount of soft radiation does not change the reconstruction, and collinear-safe, meaning that splitting a parton into two parallel components does not affect the reconstruction. It is also resilient to soft radiation, ensuring that the overall jet shape is not significantly influenced by low-energy emissions. The distances between entities i and j, such as particles and pseudo-jets, and between an entity i and the beam B, are defined as in Equation 2.5, with the parameter $p = -1$ ensuring that the reconstruction proceeds starting from the particle with the highest transverse momentum in the event:

$$d_{ij} = \min(k_{ti}^{2p},\ k_{tj}^{2p}), \quad d_{iB} = k_{ti}^{2p}, \quad \text{with} \quad p = -1. \tag{2.5}$$

The algorithm iterates over all particles, identifying the smallest distance between objects. If the smallest distance is found between two entities, i and j, they are combined to form a new pseudo-jet. If the smallest distance is instead between an entity i and the beam, the clustering stops, and jet i is defined as a final jet. This procedure continues until no unclustered entities remain. The size of each jet is characterized by the parameter R, defined in Equation 2.3. Depending on the chosen value of R, different types of jets are reconstructed in CMS. In particular, jets with $R = 0.4$ (AK4 jets) typically correspond to hadronization from objects with moderate transverse momentum, while jets with $R = 0.8$ (AK8 jets) are generally associated with boosted objects originating from high-momentum kinematic processes. Distinguishing highly boosted objects that undergo hadronization and result in very collimated jets from QCD multi-jet events is crucial for modern tagging algorithms. A schematic representation of a jet is shown in Figure 2.9. It is equally important that these algorithms are decorrelated from the invariant mass of the particle, ensuring that the tagger does not introduce spurious mass peaks. Significant recent developments have focused on these aspects, improving the precision of boosted-event tagging and optimizing the relevant kinematic phase-space regions.

The reconstruction of jets may include soft, wide-angle radiation originating from initial-state radiation, pileup interactions, or underlying events, which are events associated with the primary vertex that cannot be removed and are connected to the hard collision. To mitigate this contamination, a soft-drop grooming algorithm can be used to reconstruct the invariant mass, which recursively remove soft wide-angle radiation and it results in an overall improvement of the jet mass resolution. Pileup particles are removed from the reconstruction using dedicated algorithms, such as charged-hadron subtraction (CHS) or pileup per particle identification (PUPPI) [65]. CHS exploits the fact that all charged hadron candidates are associated with a track and checks whether the track originates from the primary vertex. If this is not the case, the charged hadron is removed. However, CHS is limited by its inability to remove pileup contributions from neutral particles. To overcome this limitation, PUPPI was developed to build upon CHS by calculating, for each neutral particle, the probability of originating from a pileup interaction, and scaling its momentum by this weight. The resulting list of particles, with each neutral particle's momentum scaled according to its weight, is then used as input for jet clustering. Moreover, jets can be

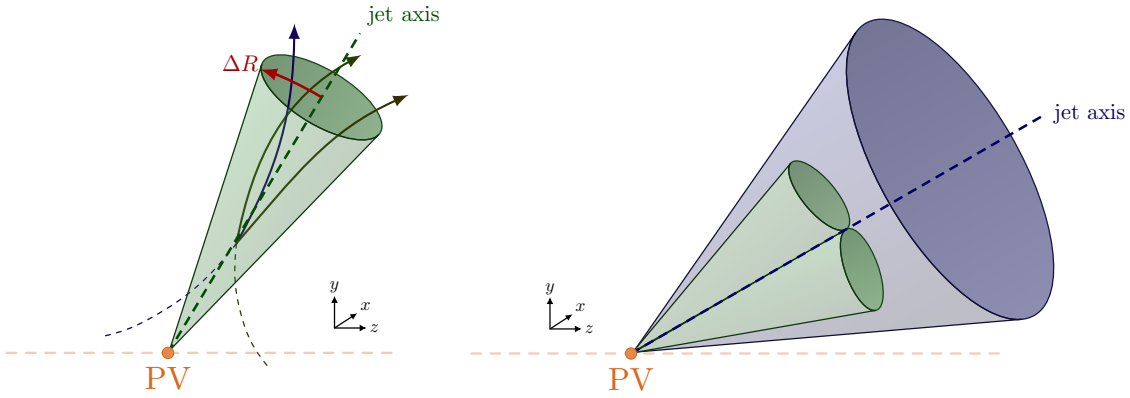

Figure 2.9: Left: A pictorial representation of a jet cone is shown. The jet axis is depicted as a green dashed line, with colored solid lines representing the particle tracks, and the angular extent of the cone is schematically indicated in red. Right: A representation of a boosted jet, in which the constituents have a high transverse momentum, causing them to be highly collimated and reconstructed as a large-radius jet [25].

affected by detector noise or other experimental artifacts. Therefore, a dedicated jet cleaning procedure is applied by imposing thresholds on several jet observables for identification purposes, resulting in an efficiency above 99% and a comparable level of background rejection.

The origin of a jet strongly influences its properties. For example, gluon jets usually exhibit a broader energy spread than quark jets, while high-momentum heavy particles (e.g. from top quarks, W, Z, or Higgs boson hadronic decays) produce jets with distinct multi-prong substructures. Therefore, identifying the origin of a jet from its reconstructed features is essential. This motivates the development of *flavor tagging* algorithms, which classify jets as b-jets, c-jets, or light-flavor jets. Several jet-tagging algorithms have been developed within CMS, and with the advent of advanced machine learning techniques, new strategies have been introduced. For instance, during LHC Run 2, the DeepJet [66] algorithm, based on a recurrent neural network (RNN) architecture, was widely employed. Towards the end of Run 2, a new approach based on Graph Neural Networks (GNNs), named ParticleNet [67], was introduced and gradually preferred due to its superior performance.

For instance, jets originating from b-quark and c-quarks decays are particularly distinguishable because they contain open-bottom or open-charm hadrons with relatively long lifetimes ($c\tau \approx 0.5$ mm for bottom hadrons and 0.3 mm for charm hadrons). Consequently, these hadrons often decay at a measurable distance from the primary vertex, making the presence of a secondary vertex a distinctive feature of such jets. The identification of b and c-jets is referred to as heavy-flavor tagging and is crucial for many analyses. Minimizing contamination from jets originating from light quarks or gluons requires specific tagging algorithms to reliably separate these objects. Jet flavor classification typically exploits hadron fragmentation properties, lifetime information, and tracking displacement variables. Due to the large number of observables to account for, machine learning methods are particularly well suited for this task. The DeepJet algorithm, for instance, simultaneously uses information from all jet constituents, both charged and neutral particles, together with displacement and global event variables. DeepJet uses information from the secondary vertex properties and the variables of the PF-jet components as inputs of the DNN, the output is the classification of jets in light, b or c-jets. ParticleNet, on the other hand, adopts a completely new paradigm for jet tagging. Instead of treating jets as ordered collections of constituents, it represents them as unordered point clouds of PF candidates. Based on a Dynamic Graph Convolutional Neural Network (DGCNN) architecture, ParticleNet effectively learns the relationships among particles within the jet. Each jet is thus reconstructed from the spatial and kinematical information of the particles. This approach achieves a substantial improvement in performance compared to existing methods. The ParticleNet tagger targets both resolved and boosted topologies. It is indeed capable of identifying hadronic decays of boosted objects (e.g. W, Z, Higgs bosons) and determine the likelihood of

each jet to be a $X \to b\bar{b}$, $X \to c\bar{c}$ or $X \to q\bar{q}$ resonance decay or a QCD multi-jet event. The probability is usually evaluated as in Equation 2.6.

$$score_{\ ParticleNet} = \frac{P(X \to b\bar{b})}{P(X \to b\bar{b}) + P(QCD)} \tag{2.6}$$

Given its superior performance compared to other heavy-flavor tagging algorithm, ParticleNet has been used in the searches addressed in this work, as it is discussed in Section 3.

2.3.1.5 τ-leptons reconstruction

τ-leptons are unstable, with a mean lifetime of $c\tau = 87.03\,\mu$m and decay far from the primary vertex creating a secondary vertex. They decay in a τ-neutrino and a W boson, which in turns can decay either leptonically or hadronically, as shown in Figure 2.10. In leptonic decays, the τ-lepton produces an elec-

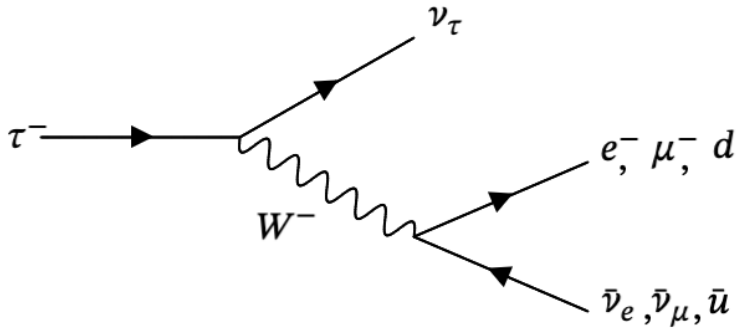

Figure 2.10: The τ-lepton decay is illustrated in its Feynman diagram. Each τ-lepton decays into a tau neutrino and a W boson. The W boson can then decay either into a lepton and its corresponding neutrino, or into a quark-antiquark pair.

tron or a muon of the same charge, which are reconstructed with the methods described in the previous Sections, along with two neutrinos ($\tau \to \ell\nu_\ell\nu_\tau$), emitted along the τ-lepton flight direction, which escape from the detector without leaving a direct trace. However, the two neutrinos leave a momentum unbalance in the transverse plane, thus their presence can be inferred through the measurement of a significant missing transverse momentum in the event. This decay mode accounts for approximately 35% of the τ-lepton total width. The main challenge in identifying such decays lies in reconstructing the missing transverse energy (MET), which serves as an indirect measurement of the neutrinos. Its reconstruction requires the utilization of all physics objects in the event, as described in Section 2.3.1.6.

On the other hand, hadronic τ decays involve one or more mesons, which are reconstructed as hadron jets and a neutrino, and occur in approximately 64.8% of cases. Intermediate resonances of $\rho(770)$ or $a_1(1260)$ might occur, resulting in a variety of charged and neutral hadrons, which are usually called *prongs*. The hadronic decay of τ leptons involves several charged hadrons, together with photons originating from the π^0 decay. The algorithm used for this reconstruction must identify such objects, determine the decay mode, and evaluate the kinematic properties. The Hadron Plus Strips (HPS) algorithm is used to reconstruct and identify τ_h candidates and to check the compatibility of all PF jet candidates with a τ lepton. The algorithm first identifies PF jets as seeds and considers all particles within a cone of $\Delta R < 0.5$ around the hadronic jet. It reconstructs π^0 candidates using either photons or PF electrons within the jet (photons from π^0 decays may convert into electron–positron pairs). These reconstructed objects are referred to as *strips*. Candidates within the clustering region around each strip are gradually added, and the strip position is recomputed using a p_T-weighted average. The $\eta \times \phi$ window is dynamically adjusted at each iteration to match the strip p_T and ensure optimal energy collection. This procedure is repeated until no further candidates are found. At this stage, all possible combinations of PF charged hadrons and strips are considered to reconstruct the potential decay mode hypotheses. There are seven different τ_h decay modes, as listed in Table 2.1. Among all candidates within a seed region, a single object is selected,

corresponding to the one with the highest p_T. τ_h candidates are then classified as originating from genuine τ-leptons or as misidentified objects, such as jets, electrons, or muons. QCD multi-jet events constitute a significant background due to their abundant production in proton-proton collisions at the LHC. Moreover, as done for other leptons, it is common in many analyses to enforce the compatibility of the τ_h candidate with the primary vertex by requiring constraints on Δ_z. In recent years, a new deep neural

Decay mode	**Meson resonance**	**B [%]**
$e\nu_e\nu_\tau$	—	17.8
$\mu\nu_\mu\nu_\tau$	—	17.4
All leptonic decays	—	35.2
$h^\pm\nu_\tau$	—	11.5
$h^\pm\pi^0\nu_\tau$	$\rho(770)$	26.0
$h^\pm\pi^0\pi^0\nu_\tau$	$a_1(1260)$	9.5
$h^\pm h^\mp h^\pm\nu_\tau$	$a_1(1260)$	9.8
$h^\pm h^\mp h^\pm\pi^0\nu_\tau$	—	4.8
Other hadronic decays	—	3.2
All hadronic decays	—	64.8

Table 2.1: All possible τ decay modes are listed, divided into leptonic and hadronic decays, including the meson resonances for some decay modes. For each decay mode, the corresponding branching fraction is given in percent.

network (DNN) specifically developed to reconstruct τ_h candidates has been adopted in CMS analyses: the DeepTau algorithm. It is a multiclass τ identification algorithm based on a deep convolutional neural network (CNN), which discriminates genuine τ leptons from electrons, muons and jets. The algorithm uses information from all reconstructed particles near the τ_h candidate axis, together with high-level information from the reconstructed τ leptons. Two different versions of the DeepTau algorithm were employed for Run 2 and Run 3. The main improvements in Run 3 include reducing discrepancies between data and simulation in regions of high hadronic τ purity, integrating domain adaptation techniques, and improving data handling, resulting in enhanced overall performance. This new algorithm has been shown to significantly improve τ_h identification and reconstruction compared to previous algorithms [68]. For example, for a given jet rejection level, the efficiency for genuine τ_h candidates is increased by 10–30%.

When τ leptons are produced from highly boosted resonances, their decay products become collimated within a narrow cone, often forming a single jet, similar to the behavior observed for AK8 jets. While for LHC Run 2 a dedicated reconstruction technique was not available, with the beginning of Run 3, an extension of the HPS algorithm has been specifically developed to reconstruct such high-momentum candidates by exploiting the jet substructure of the τ_h candidates. Jets are reconstructed using the Cambridge-Aachen (CA) clustering algorithm, with a distance parameter of 0.8. The algorithm reverses the last step of the CA clustering sequence, identifying the two sub-jets that are expected to correspond to the decay products of the boosted τ lepton. Requirements on the sub-jets are imposed to reduce the misidentification of QCD jets. If no valid sub-jets are found, τ_h reconstruction is not attempted for that jet, and the algorithm proceeds to the next one. Candidates containing electrons or muons are also rejected, as they are interpreted as semileptonic decays. This algorithm achieves superior performance compared to the standard HPS in reconstructing highly boosted τ leptons and it maintains a misidentification rate on the order of 10^{-4}. Eventually, the selected boosted τ candidates are passed to the HPS algorithm that performs the standard τ reconstruction.

2.3.1.6 MET reconstruction

Neutrino particles do not interact with the CMS sub-detectors and travel all the way from the collision point out of the detector without being detected. Therefore, their presence must be inferred indirectly. The imbalance in the plane transverse to the proton beams is quantified by the missing transverse momentum or missing transverse energy[4], and it is defined as in Equation 2.7, where the momentum

[4]The longitudinal component cannot be reconstructed because the parton momentum fractions within the protons are unknown on an event-by-event basis.

$\overrightarrow{p}_{T,j}$ of each PF jet with $p_T > 10$ GeV is corrected via jet energy corrections to account for detector and software limitations (e.g. tracking inefficiencies, calorimeter energy thresholds, and non-linear detector responses).

$$\overrightarrow{p}_T^{\text{miss}} = -\sum_{i=1}^{N_{\text{particles}}} \overrightarrow{p}_{T,i} - \sum_{j=1}^{N_{\text{jets}}} (\overrightarrow{p}_{T,j}^{\text{corr}} - \overrightarrow{p}_{T,j}) \tag{2.7}$$

Given the elusiveness of neutrinos, MET alone may not be sufficient to reconstruct their energy and momentum efficiently. Therefore, extensive effort has been dedicated to develop techniques to reconstruct the mass of initial states involving neutrinos in their decay products. Of particular interest in this work are the algorithms developed to reconstruct the $H \to \tau\tau$ system, which involves at least two neutrinos in its decay. The analyses described in Chapter 3 make use of FastMTT [69]. The methods use a likelihood approach to quantify the compatibility between the Higgs boson mass hypothesis and the measured momenta of objects reconstructed as decaying from the τ candidates. This results in an improved mass resolution of the $H \to \tau\tau$ system by about 30% compared to the visible invariant mass of the system. A slight shift from the expected Higgs boson mass is still observed due to the assumption that the source of missing transverse energy originates solely from neutrinos from τ decays, while additional contributions may arise from B-hadron and meson decays from b-quark hadronization. It demonstrated to achieve a mass resolution of 22 GeV for Z bosons and 34.5 GeV for Higgs bosons.

2.3.2 Simulated processes

In addition to data collection, each experiment aims to describe collisions between proton-proton beams and their outcomes using dedicated Monte Carlo simulations. The results from both data-triggered and simulation-triggered events are then reconstructed into objects in an analogous way, as described in the previous section, and are ultimately used in all analyses.

The event simulation begins with the generation step, in which simulated collision events are produced using theoretical models of particle interactions. These models are implemented in specific event generators, such as MadGraph5_aMC@NLO [70] or Powheg [71], which are Matrix Element generators including calculations at Leading Order (LO) or Next-to-Leading Order (NLO) and produce the hard scattering and the resulting particles before fragmentation. They are then combined with other generators such as Pythia [72] for the parton shower, hadronization, and the production of the underlying events to generate particle-level information prior to simulation and reconstruction. Detector simulation models are then used to emulate how particles interact within the detectors, accurately reproducing real data conditions. Within the CMS context, the Geant4 [73] framework is employed. This software simulates particles and their interactions with matter, including functionalities such as tracking, geometry, physics models, and hits, covering a wide range of physics processes. CMS uses Monte Carlo programs to perform realistic, data-driven simulations, which are built upon the CMS-specific software package CMSSW and continuously evolve in line with the experiment's performance. Monte Carlo simulations track particles through different detector materials, emulating interactions in every layer, including energy loss, multiple scattering, and secondary particle production. The interaction of particles within the detector produces signals that are digitized via dedicated electronics. This step is emulated in simulations to reproduce the digital output of the sub-systems, including real effects such as noise and pulse shaping, and simulation also accounts for pileup events. Since the LHC collides many protons simultaneously, several interactions can overlap within a single event. CMS uses minimum-bias MC samples to simulate these additional interactions, reproducing realistic detector conditions. Finally, the digitized signals are converted into a format that emulates the requirements imposed by the HLT, bringing simulated events to the same level as collected data. Events are then reconstructed as described in Section 2.3.1. The primary datasets in CMS are collected as physics objects (e.g. muon, jet, τ-lepton, etc.).

2.3.2.1 Corrections to simulation events

Many analyses rely on MC samples to model a wide range of phenomena, therefore, ensuring good agreement between observed data and simulated events is essential. Any discrepancies must be carefully evaluated and, when necessary, corrected. Corrections can be applied through dedicated weights assigned to MC events. When reweighting alone is insufficient to account for observed deviations between data

and simulation, data-driven techniques are employed to estimate and correct the remaining discrepancies. Different data-taking periods can have distinct sets of weights w depending on their specific conditions. Several CMS analysis groups focus specifically on evaluating these corrections through Scale Factors (SFs), which account for differences between data and simulation. These corrections are often derived per object (e.g. electrons, muons, jets) or triggers etc. They are usually developed at a specific operating point, i.e. a working point (WP), defined at a given efficiency level. SFs are defined as the ratio between the efficiency measured in data and that measured in simulated events, as shown in Equation 2.8.

$$SF = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}} \tag{2.8}$$

A common strategy to measure these efficiencies is the *tag-and-probe* technique, in which a well-understood and well-reconstructed process is used as a clean source of target objects. One subset of events, called the *tag*, satisfies tight selection criteria to ensure that genuine signal events are retained while suppressing background contamination. The other subset, called the *probe*, is associated with the tag event and selected with looser criteria and it is used to study the efficiency of the selection under investigation. The efficiency is then computed as the ratio of probe events passing the selection under study to the total number of probe events. SFs are usually computed "centrally" by dedicated CMS working groups. Some relevant examples of simulation correction evaluations, particularly relevant to the analyses presented in this thesis, are discussed below.

Pileup reweighting

The beam conditions in terms of energy and luminosity are considered when producing the theoretical models of hard scattering. However, simulated events may exhibit slightly different distributions of the number of interactions per vertex compared to data. Pileup interactions affect both the detector response and object reconstruction, therefore, this discrepancy between data and simulation must be corrected to achieve a better match. A reweighting procedure is applied to simulated events using the ratio between the pileup distribution observed in data and that predicted by simulation. This procedure ensures a precise morphing of the simulated PU distribution to match that observed in data.

Trigger corrections

Each analysis relies on a specific set of trigger requirements based on the targeted objects, which are applied consistently to both data and simulated events. Although MC simulations aim to reproduce the trigger behavior observed in data, residual differences may remain and must be corrected using dedicated scale factors derived from trigger efficiencies measured in data and simulation. When multiple triggers cover the same phase space, the overall trigger efficiency must be computed accounting for the logical `OR` of the individual triggers, to avoid double-counting events. Depending on the trigger path, different approaches can be employed to derive these efficiencies. For example, for single-muon triggers, the efficiency is measured using the tag-and-probe technique applied to a sample of $Z \to \mu\mu$ events, exploiting the clean muon signature. One "tag" muon is required to pass the single-muon trigger and tight identification criteria, while the other muon, the "probe", is selected without trigger requirements but constrained by the invariant mass of the Z boson to enhance sample purity. The resulting efficiency is computed as the ratio between the number of events where the probe muon passes the single-muon trigger requirement and the total number of selected events. The ratio of this efficiency measured in data to that measured in simulation defines the SF, which is studied as a function of the reconstructed muon momentum and pseudorapidity. A similar strategy is adopted for single-electron triggers with $Z \to ee$ events.

Muon corrections

SFs for muon identification are evaluated through efficiency measurements performed via tag-and-probe technique on different samples, depending on the type of reconstructed muon. The efficiency for *prompt* muons (i.e. muons originating directly from primary interactions such as $Z \to \mu\mu$ decays) is measured using $Z \to \mu\mu$ events, while QCD multi-jet samples are used to evaluate the efficiency for identifying non-prompt muons.

Electrons corrections

Corrections for electrons are obtained in a similar way to those used for muons, by selecting a clean $Z \to e^+e^-$ enriched region and applying a tag-and-probe technique to derive the efficiencies. In addition, energy scale and resolution corrections are computed for the calibration of electrons and photons in both data and MC, evaluated using the $Z \to e^+e^-$ peak. Reconstructed energies in data tend to be slightly underestimated with respect to expectations from theory; therefore, the position of the Z peak is compared between data and MC, and the energies are corrected accordingly. Resolution in MC is typically narrower than in data, so an additional smearing is applied to MC to match the observed resolution.

Jets corrections

Jets reconstructed in data require corrections to their energy scale and energy resolution in order to calibrate the detector response, accounting for pileup contributions, non-linearities in the calorimeter response, and potential variations between data and simulation. These corrections are referred to as *Jet Energy Scale* (JES) and *Jet Energy Resolution* (JER). The JES evaluation starts with pileup subtraction, then accounts for the detector response, and finally derives residual corrections for differences between data and simulation. Calibration studies are performed using dijet and multi-jet samples, such as the Z/γ + jet p_{T} balance, where the photon or the Z boson is used as a reference and momentum conservation is exploited to derive corrections [74]. To account for the worse energy resolution observed in data, a process of *smearing* (i.e. worsening the jet energy resolution) is applied on simulated events to recover agreement between data and simulated events.

Specific corrections are also needed to account for deviations in efficiency between data and simulation related to the identification of heavy-flavor jets, such as b-jets. For resolved jets, depending on how the tagger is used, two different approaches can be applied. The first approach involves deriving a correction to account for efficiency differences when the tagger is only used for event selection. The second approach is required when the full shape of the tagger distribution is used as input to a multivariate analysis or a final fit; in this case, proper reweighting must be applied. For the reweighting approach, scale factors must be computed on a per-event basis to modify the shape of the b-tagging discriminant and to correct the simulation so that it matches the behaviour observed in data. These weights are functions of the jet transverse momentum and pseudorapidity. They are expected to modify only the shape of the discriminant; therefore, additional reweighting is performed in each analysis to ensure that the overall yield before and after applying these weights is preserved. In both approaches, working points are defined for different tagger efficiencies and the corresponding calibrations are estimated. For instance, in the case of the DeepJet tagger used for AK4 jets in the LHC Run 2 data-taking, the working points defining loose and medium selections correspond to cut values in the discriminator distribution for which the misidentification rate of light jet as a b-jet is 10% and 1%, respectively [66, 75]. Recently, new calibration methods have been developed for taggers addressing boosted jets, such as AK8 jets, where *proxy jet* methods are employed. For instance, for the ParticleNet tagger used to identify AK8 jets originating from resonances decaying to $b\bar{b}$ or $c\bar{c}$, different working points are defined for different efficiency levels and calibrations are performed using proxy jets from gluon splitting into a pair of b or c quarks, along with other techniques designed to improve the similarity between the selected proxy and the signal jet. In particular, for the Run 2 data-taking period, three working points are defined for $H \to b\bar{b}$ signal jets with efficiencies of 80%, 60%, and 40%, referred to as Low Purity (LP), Medium Purity (MP), and High Purity (HP), respectively.

τ corrections

The efficiencies in identifying and isolating τ-leptons are computed analogously to what is done for other leptons, using a tag-and-probe technique on a $Z \to \tau\tau$ sample. Additionally, different sets of scale factors are applied to the energy scale of the τ-leptons depending on the scenario: specific SFs are applied for genuine τ-leptons, genuine electrons misidentified as τ-leptons, and genuine muons misidentified as τ-leptons. The efficiencies may vary from about 40% for jets up to 99.95% for muons, depending on the object type and transverse momentum. Significant updates are being implemented, mainly for Run 3 analyses, including the use of newer and extended datasets for training, improved training techniques,

and optimized hyperparameter tuning. The efficiency in identifying τ-leptons against other particles must also be considered and usually depends on the decay mode of the τ-lepton.

2.3.3 Statistical analysis and interpretation

The datasets obtained from data and MC simulations, corrected for the effects described in the previous section, are eventually selected according to the goals of the specific analysis. Events are then analyzed differently depending on the target of the measurement. In this section, however, the focus is specifically on the statistical methods employed in Higgs searches. It is common to aim at measuring the cross section of a process, and it is necessary to test the compatibility of the observed data with the expectation given by the signal-plus-background hypothesis, in comparison to the hypothesis that the data are well described by a background-only model. To this end, statistical tools based on likelihood functions and hypothesis testing are employed, as extensively described in Ref. [76] and effectively reviewed in Ref. [77]. In the following, a brief overview is provided, serving as a foundation for the discussion in the next section.

It is common practice in HEP analyses to develop the entire analysis workflow without revealing the data before the analysis procedure is completely established and endorsed by the collaboration, thus following a "blinding" procedure to avoid unintentional biases during the development phase. For this reason, data in the most sensitive kinematic regions are not shown until the complete analysis workflow is approved. In a blinded analysis, pseudo-data are used in the statistical analysis instead of the observed data, which in the following is referred to just as data. To perform the statistical analysis, a discriminant variable sensitive to differences between signal and background is chosen. The distributions of this discriminant variable are then used to perform a binned fit of the model to data, in which each bin contains a count that can be described by a Poisson distribution. The overall probability density function is the joint product of the probabilities of observing the data in each bin, given the model, as shown in Equation 2.9, where N_i is the observed number of events in data and Λ_i is the expected number from the model in the i-th bin.

$$P(\text{data} \mid \text{model}) = \prod_i P(N_i \mid \Lambda_i) \qquad \text{with} \quad P(N_i \mid \Lambda_i) = \frac{\Lambda_i^{N_i} e^{-\Lambda_i}}{N_i!} \tag{2.9}$$

The expected value Λ_i can be expressed as the sum of the signal yield s and the background yield b. However, one must also take into account the systematic uncertainties affecting the yield in each bin. These are included in the model as nuisance parameters (θ vector). The expected value can then be written as in Equation 2.10, and the Poisson expression can be re-expressed as in Equation 2.11.

$$\Lambda_i(\mu, \theta) = \mu\, s_i(\theta) + b_i(\theta) \tag{2.10}$$

$$P(N_i \mid \mu s_i + b_i) = \frac{(\mu s_i + b_i)^{N_i} e^{-(\mu s_i + b_i)}}{N_i!} \tag{2.11}$$

Each nuisance parameter is included in the statistical model with an associated probability density function that encodes the prior knowledge about its expected value and uncertainty. The analytical form of this probability depends on effect addressed by the systematics. For example, luminosity uncertainties, which arise from independent measurements, are typically modeled using a log-normal distribution.

To infer the parameter of interest, typically the signal strength μ, defined in Equation 2.12, the maximum likelihood estimation method is used to determine the value that best describes the observed data:

$$\mu = \frac{\sigma_{\text{obs}}}{\sigma_{\text{SM}}} \tag{2.12}$$

The likelihood function accounts for both the probability of observing the data given the parameters and the prior knowledge associated with the nuisance parameters. It can be written as in Equation 2.13, where "data" represents either the observed data or pseudo-data, and $p(\tilde{\theta} \mid \theta)$ is the probability density function of the systematic uncertainties, which, as discussed in Ref. [77], can be interpreted in a frequentist framework.

$$\mathcal{L}(\text{data} \mid \mu, \theta) = \text{Poisson}\big(\text{data} \mid \mu\, s(\theta) + b(\theta)\big)\; p(\tilde{\theta} \mid \theta) \tag{2.13}$$

To determine whether the data are compatible with the background-only hypothesis or the signal+background hypothesis, with the signal yield scaled by a signal strength parameter, the test statistic $\tilde{q}_\mu$ [78] is defined

based on the profile likelihood ratio, as given in Equation 2.14, where $\hat{\theta}_\mu$ are the conditional maximum likelihood estimators of the nuisance parameters for a given μ, and $(\hat{\mu}, \hat{\theta})$ is the global maximum of the likelihood.

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})} \tag{2.14}$$

The test statistic is computed for the observed data $\tilde{q}_\mu^{\text{obs}}$, along with the nuisance parameters that best describe the data by maximizing the likelihood, both under the background-only and signal-plus-background hypotheses, $\hat{\theta}_\mu^{\text{obs}}$ and $\hat{\theta}_0^{\text{obs}}$, respectively. A MC toy is generated to construct the probability density functions of observing a test statistic $\tilde{q}_\mu$ given the signal strength or the null (background-only) hypothesis, using the observed nuisance parameters that best describe the data $\hat{\theta}_\mu^{\text{obs}}$ and $\hat{\theta}_0^{\text{obs}}$. These functions are then used to define the p-values for the observation under the signal-plus-background and background-only hypotheses, denoted as p_μ and p_b, respectively, and defined as in Equation 2.16

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} \mid signal + background) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f_\mu(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})\, d\tilde{q}_\mu, \tag{2.15}$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} \mid background - only) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f_0(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})\, d\tilde{q}_\mu. \tag{2.16}$$

The confidence level CL_s is then computed as the ratio between the probability of observing a test statistic under the signal-plus-background hypothesis and the probability of observing a test statistic greater than the observed one under the background-only hypothesis, as in Equation 2.17:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}. \tag{2.17}$$

Results are usually reported in terms of a 95% confidence level (CL), meaning that $CL_s(\mu) \leq 0.05$ for those results, thereby excluding the corresponding signal strength at the 95% CL. Expected upper limits are also commonly reported. These are evaluated along with their $\pm 1\sigma$ and $\pm 2\sigma$ bands using large samples of pseudo-data generated under the background-only hypothesis. The median value is computed as the 50th percentile of the distribution, while the $\pm 1\sigma$ ($\pm 2\sigma$) bands correspond to the 16th and 84th (2.5th and 97.5th) percentiles.

Chapter 3

Search for di-Higgs production into the bb$\tau\tau$ final state with LHC data at CMS

The advanced CMS design, combined with precise object reconstruction, has enabled a wide variety of measurements, allowing for stringent tests of the Standard Model, detailed studies of top quark and Higgs physics, and searches for beyond the Standard Model phenomena. Currently, Higgs physics can only be investigated at the LHC, offering a rich opportunity to test the validity of the SM while probing for hints of new physics. As discussed in Chapter 1, among all Higgs measurements, searches for di-Higgs production with the CMS experiment at the LHC are of particular importance. Di-Higgs production provides a direct probe of the Higgs boson self-coupling predicted by the SM, giving access to the Higgs potential and the BEH mechanism. The study of the Higgs self-coupling therefore constitutes a central objective of the current LHC program, the future high-luminosity LHC, and the strategic plans for future collider experiments. In this thesis, the study of the HH production is conducted on the final state in which the Higgs bosons decay into two b-quarks and two τ-leptons, which constitutes a channel among the most sensitive [27], combining a relatively large branching ratio with reduced background contamination compared to other final states.

At the time of this thesis, data from the LHC Run 2 data-taking period are available, while data from Run 3 are gradually becoming accessible. Aiming to study the Higgs boson self-coupling, this thesis explores the possibility of enhancing the sensitivity of the search for $HH \to bb\tau\tau$ by including a dedicated phase space optimized for $H \to bb$ events with a high Lorentz boost, where the resulting b-quarks are reconstructed as large-radius (AK8) jets. Although this boosted phase space represents a statistically limited subset of events, it typically achieves a higher signal-to-background ratio than the majority of events selected by the so-called resolved analysis, thanks to its characteristic topology and lower QCD multi-jet backgrounds. Recent advancements on boosted jet topologies analysis methods like the ParticleNet tagger (see Section 2.3.1.4) offered also interesting prospects to further improve the sensitivity in this boosted regime. While the non-resonant search for $HH \to bb\tau\tau$ using the LHC Run 2 dataset has already been published in Ref. [79], a first test of the ParticleNet tagger performance and development of the strategies for handling this phase-space was conducted initially in the context of the search for resonant di-Higgs production in the same $bb\tau\tau$ final state using the full Run 2 available dataset. Once the 2022 and 2023 Run 3 data became available, I contributed to the development of the search for non-resonant $HH \to bb\tau\tau$. Building on the experience gained from the studies of boosted topologies in the resonant search, this thesis presents a detailed optimization of the treatment of such events. In particular, my work focused on optimizing the selection of boosted $H \to bb$ events, defining custom selection strategies, deriving the corresponding correction factors for background processes, and developing a dedicated method to discriminate signal from background in this kinematic regime.

The search for resonant HH production in the $bb\tau\tau$ final state using Run 2 proton–proton collision data at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}, is briefly described in

Section 3.1. The non-resonant search using Run 3 data at $\sqrt{s} = 13.6$ TeV, collected during the 2022 and 2023 data-taking periods and corresponding to an integrated luminosity of 62.4 fb^{-1}, is discussed in Section 3.2. Both analyses target three $\tau\tau$ final states, where one τ lepton decays hadronically (τ_h) and the other decays either hadronically or leptonically into an electron (τ_e) or a muon (τ_μ), accompanied by neutrinos. These three modes (referred to as channels) cover 88% of all $\tau\tau$ decays, corresponding to six out of nine possible combinations. The remaining fully leptonic modes (both τ leptons decaying into electrons or muons) are not considered in this work, as they involve four neutrinos, which complicates event reconstruction and worsens the $\tau\tau$ mass resolution, an important variable for background suppression. Moreover, fully leptonic channels with two same-flavor leptons suffer from significant contamination from Drell–Yan (DY) production events [80]. In the three selected channels, the main reconstruction challenges arise from the presence of neutrinos and from distinguishing hadronic τ decays from background processes. The signal processes are extremely rare compared to other SM processes with similar experimental signatures, and several SM backgrounds mimic the signal. The most abundant ones are top–antitop ($t\bar{t}$), DY, and QCD multi-jet events. In many cases, these backgrounds contain genuine b-jets or genuine τ leptons; in others, objects such as light jets or leptons may be misidentified as b-jets or τ candidates (for example, in DY events, the τ leptons may be genuine, while other leptons may be misreconstructed as τ_h candidates). The first step in selecting signal events is to ensure that the particles originating from Higgs boson decays are correctly identified, reducing instrumental and non-prompt backgrounds. This provides an initial level of background suppression. Sophisticated algorithms such as DeepTau and ParticleNet are used to further suppress jets originating from non-signal sources (see Section 2.3.1). Finally, kinematic properties and particle correlations are exploited to enhance signal sensitivity and further reduce background contributions.

Given the complexity and breadth of expertise required for such analyses, it is common within large HEP collaborations such as CMS for these studies to be carried out collaboratively by several international teams. Both searches build extensively upon the previous CMS analysis of non-resonant $HH \to bb\tau\tau$ production using LHC Run 2 data, published in Physics Letters B in 2021 [79]. The overall logic followed in both analyses can be divided into several steps, each described in its corresponding section, while sharing a common workflow. The starting datasets consist of the data collected by CMS during LHC Run 2 or Run 3, which, as described in the previous section, are organized according to reconstructed physics objects. Simulated events produced with Monte Carlo techniques are considered on a per-sample basis. A dedicated set of triggers, aimed at isolating the objects relevant to this analysis, is applied to both data and simulated samples. τ leptons and b quarks are selected through specific identification criteria and kinematic requirements designed to maximize signal acceptance while minimizing background contamination. Subsequently, the Higgs boson candidate decaying into $\tau\tau$ is reconstructed through the three τ decay channels mentioned above, whereas for the Higgs boson decaying into $b\bar{b}$, two jet topologies are identified: boosted or resolved, depending on the event kinematics. Additional selections are applied to the invariant mass of the reconstructed systems to further suppress background contributions, and events are categorized into distinct classes. The discrimination between signal and background is achieved using a custom deep neural network (DNN), whose output is used as input to the statistical analysis to extract constraints on Higgs boson production processes. The searches are presented following a common structure, with particular emphasis on the developments related to the boosted category. The document structure is organized as follows: datasets and triggers (Sections 3.1.1 and 3.2.1); object reconstruction (Sections 3.1.2 and 3.2.2); event selection and categorization (Sections 3.1.3 and 3.2.3); background modeling (Sections 3.1.4 and 3.2.4); systematic uncertainties (Sections 3.1.6 and 3.2.6); and finally, results (Sections 3.1.7 and 3.2.7), for the resonant Run 2 and non-resonant Run 3 searches, respectively. The searches presented here are currently under scrutiny by the CMS collaboration prior to final approval. Therefore, the results are shown without data in the sensitive phase space, as this region remains blinded (see Section 2.3.3).

3.1 The boosted $H \to bb$ category in the resonant search for $HH \to bb\tau\tau$

The search for resonant production of a pair of Higgs bosons decaying into $bb\tau\tau$ uses proton-proton collision data at $\sqrt{s} = 13$ TeV collected by the CMS experiment at the LHC between 2016 and 2018,

corresponding to an integrated luminosity of 138 fb^{-1}. The search targets heavy resonances X that are assumed to be produced via gluon fusion and to decay into two on-shell Higgs bosons. Models involving WED (see Section 1.2.3) predict the existence of new resonances such as the spin-0 radion and the spin-2 Kaluza-Klein graviton. This search targets both spin hypotheses, assuming narrow resonances ($\Gamma_X = 1$ MeV) with masses in the range between 250 GeV and 3 TeV. Due to this wide mass range, different kinematic configurations must be addressed. At low resonance masses, the Higgs bosons typically decay into particles which are produced back-to-back, as most of the resonance mass is used by the Higgs boson masses themselves, leaving little available transverse momentum. At higher resonance masses, however, the Higgs bosons are produced with a significant Lorentz boost, resulting in more collimated decays. My contribution to this search focused on defining a dedicated treatment of boosted $H \to bb$ events through the development of a new boosted category. Events in this category are selected using the ParticleNet tagger and serve as an extension of the accessible phase space, enhancing the sensitivity of the analysis in the high-mass resonance regime. The importance of an efficient boosted category is primarily kinematic: at high resonance masses (up to 3 TeV), Higgs bosons are produced with large boosts, causing their decay products to become collimated and often reconstructed within a single merged jet cone. The boosted category therefore becomes increasingly significant as the resonance mass increases. While the inclusion of the boosted $H \to bb$ category provides a gain in sensitivity, the analysis is not fully optimized for the boosted regime overall, since the reconstruction of boosted τ leptons only became available with the beginning of Run 3 and they are therefore not included here.

A brief overview of the search workflow is presented in this section to contextualize the boosted $H \to bb$ category and its role within the overall analysis strategy, with particular emphasis on the aspects developed in this work. A schematic representation of the main analysis steps is shown in Figure 3.1.

3.1.1 Datasets and triggers

This search uses data recorded at CMS during the 2016, 2017, and 2018 LHC runs, corresponding to integrated luminosities of 36.3 fb^{-1}, 41.5 fb^{-1}, and 59.8 fb^{-1}, respectively. During CMS data taking, the trigger system, composed of L1 and HLT stages, selects events as described in Section 2.2.2. Events are grouped into primary datasets according to the HLT triggers used for their recording (e.g. electron/photon, muon, τ, and MET primary datasets).

As outlined in Section 2.3.2, both signal and background MC samples in CMS are generated using state-of-the-art theoretical calculations (LO, NLO, or NNLO), with event generation performed by MadGraph5_aMC@NLO (used for every sample except for $t\bar{t}$, single top and single Higgs samples) or POWHEG (used for generating $t\bar{t}$, single top and single Higgs samples). Multi-parton interactions, parton showering, and hadronization are modeled with PYTHIA, and the CMS detector response is simulated using Geant4. Dedicated reweighting procedures are applied to reproduce the pileup conditions observed in data. Moreover, as discussed in Section 2.3.2.1, correction factors are derived and applied to simulation events to account for the measured trigger efficiencies, as well as for any differences between data and simulation arising from objects reconstruction and identification. The simulated signal samples correspond to narrow-width resonances produced via gluon fusion, including spin-0 radions and spin-2 gravitons, generated at leading-order precision for resonance masses between 250 GeV and 3 TeV. The dominant backgrounds depend on the kinematic region but are generally driven by $Z\gamma^* \to \ell\ell$ and $t\bar{t}$ contributions, simulated at NLO precision. W+jets processes, which can fake signal objects, are simulated at LO precision. Single-top (t+X) and single-Higgs (H+X) production are simulated at NLO, diboson (WW, ZZ, WZ) production at LO, and triboson (ZZZ, ZZW, ZWW, WWW) at NLO. $t\bar{t}$+X production is simulated either at LO or NLO, depending on the specific process.

For each of the three channels considered ($\tau_e\tau_h$, $\tau_\mu\tau_h$, $\tau_h\tau_h$), specific trigger requirements are applied. Both general-purpose and dedicated triggers are employed, requiring single electrons, muons, or taus, as well as lepton–tau combinations. For the semi-leptonic $\tau_e\tau_h$ and $\tau_\mu\tau_h$ channels, single-lepton triggers requiring an isolated lepton are used in combination with cross-lepton triggers that select events containing an isolated lepton and a hadronically decaying τ-lepton. The cross-lepton triggers help lower the lepton p_T thresholds and enlarge the accessible phase space. In the $\tau_h\tau_h$ channel, triggers require a pair of hadronically decaying τ-leptons. In addition, single-τ triggers targeting high-momentum τ-leptons and MET triggers targeting events with large missing transverse energy are used to enhance the sensitivity

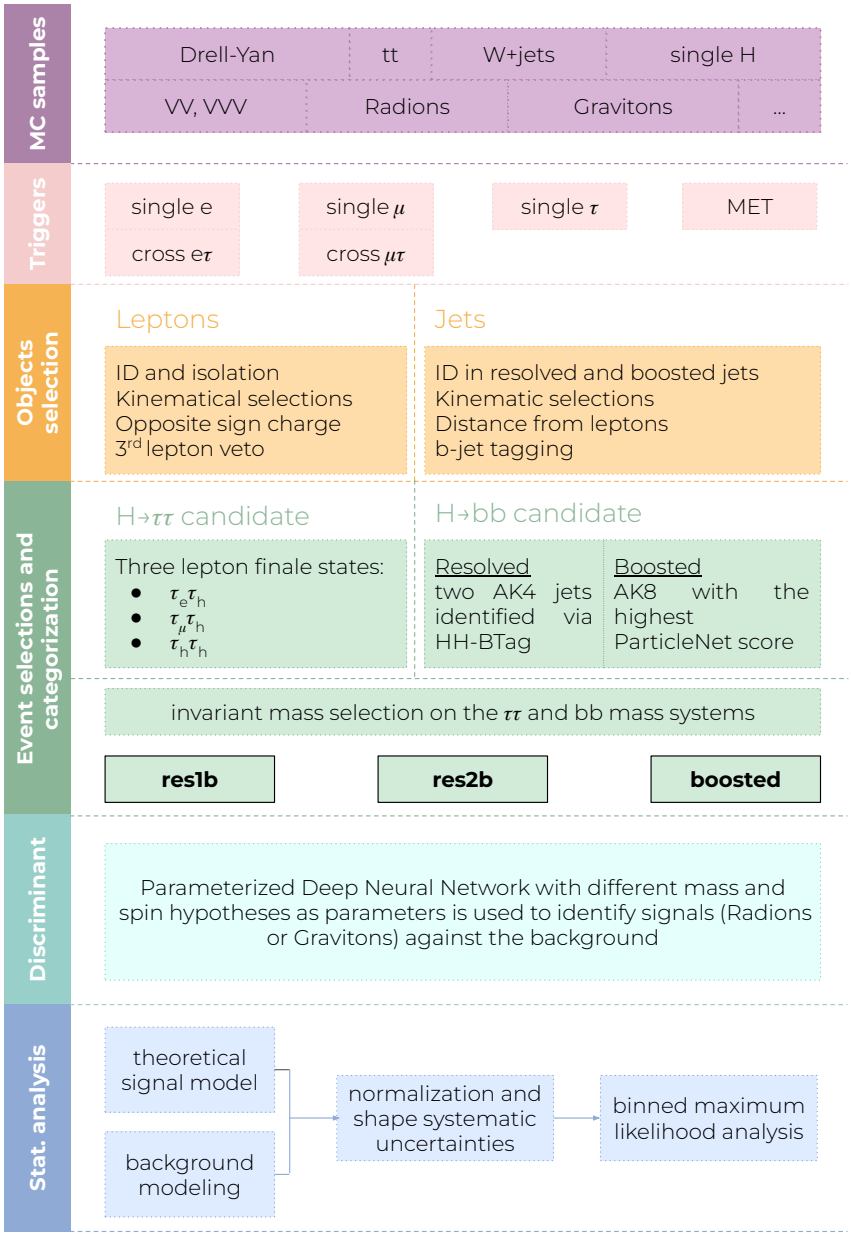

Figure 3.1: A representative schematic of the analysis workflow is here reported. The datasets are the data collected by CMS during the Run 2 data taking. Simulated events produced with Monte Carlo techniques are considered on a per-sample basis. A dedicated set of triggers, designed to isolate the objects relevant for this analysis, is applied consistently to both data and simulated samples. τ-leptons and b-quarks are selected according to efficiency and kinematic criteria optimized to maximize signal acceptance while minimizing background contamination. The Higgs boson candidate decaying into $\tau\tau$ is reconstructed through its three main decay channels. For the Higgs candidate decaying into $b\bar{b}$, jets are classified as either boosted or resolved depending on their topology. Further selections are applied on the invariant mass of the reconstructed systems to suppress background contributions. Finally, events are grouped into three distinct categories. Signal and background are discriminated using a dedicated deep neural network, whose output distributions are then employed in the statistical analysis to extract limits on the Higgs boson production processes.

of the analysis to heavy resonances. To ensure high trigger efficiency, specific transverse momentum thresholds are required for each trigger, depending on the data taking period: electrons have thresholds in the range 24–35 GeV, muons in the range 20–27 GeV, τ_h pairs within range values of 35-40 GeV, τ_h leptons firing the cross triggers in the range 27–30 GeV, single-τ_h triggers between 120–180 GeV, and MET triggers around 120 GeV. Conditions on pseudorapidity ranges are also applied to ensure high trigger efficiency.

3.1.2 Physics objects reconstruction

Events are reconstructed using the particle-flow algorithm and dedicated techniques developed within the CMS framework, allowing for the reconstruction of all physics objects (see Section 2.3.1). The objects relevant for this study are reconstructed as follows. Muons used in the signal regions are required to satisfy an isolation criterion (see Equation 2.4) $I^{\ell}_{rel} < 0.15$ and a tight identification criterion, resulting in an efficiency ranging from 90% to 99%, depending on p_T, while muons used to veto events are required to pass a looser isolation requirement of 0.3 and a medium identification requirement, corresponding to 99% efficiency. Signal electrons are identified using a tight working point of the BDT discriminator, with an associated efficiency of approximately 80%, while veto electrons pass the medium criterion. τ-leptons are reconstructed with the hadrons-plus-strips algorithm (Section 2.3.1.5) and the DeepTau algorithm is used for identification, discriminating τ-leptons against muon, electrons and jets, with tight, very-very loose and medium requirements, respectively. Jets are reconstructed with the anti-k_T algorithm, either as resolved or boosted jets (Section 2.3.1.4). Resolved jets are required to pass a medium working point on b-tagging, corresponding to an efficiency of 84%, while boosted jets must pass a low-purity working point with 80% efficiency. Object reconstruction is applied consistently to both data and MC samples, with corrections being applied to simulated events to match data, as discussed in Section 2.3.2.

3.1.3 Event selection and categorization

With all objects satisfying the selection criteria described above, and aiming to reconstruct the $bb\tau\tau$ final state, the decay products of the two Higgs boson candidates must be then identified. The reconstruction of the Higgs candidates begins with identifying the $H \to \tau\tau$ candidate. Events are classified into three channels depending on the decay mode of the τ-lepton. The reconstruction of the $H \to \tau\tau$ candidate requires at least one τ to decay hadronically, while the other may decay into a muon or an electron (if it passes the selection criteria discussed above) or hadronically. If a selected muon is found, the event is categorized as $\tau_\mu\tau_h$. If no muons are present but an electron candidate is identified, the event is classified as $\tau_e\tau_h$. If neither muons nor electrons are selected but a second hadronic τ_h is reconstructed, the event falls into the $\tau_h\tau_h$ channel. When multiple candidates are present, the most isolated muon, electron, or τ_h is chosen. If two τ_h candidates have equal isolation, the one with the highest transverse momentum is selected. The selected leptons or hadronic τ candidates must be consistent with originating from the primary vertex of the event. Specifically, a condition of the distance between the track and the vertex in the z direction of $|\Delta_z| < 0.2$ cm is imposed, and electrons and muons must satisfy a requirement of $|\Delta_{xy}| < 0.045$ cm in the transverse (x-y) plane. To suppress overlaps and potential misidentification, the two $\tau\tau$ candidates must be separated by an angular separation of $\Delta R > 0.5$ and have opposite electric charge, as expected from charge conservation in Higgs decays. To reduce contamination from $Z\gamma^* \to \ell\ell$ processes, a third-lepton veto is applied: events containing additional electrons or muons are rejected.

The $H \to bb$ candidate is identified from the jets passing the mentioned criteria. Its reconstruction depends on the topology of the $H \to bb$ decay. In the resolved regime, at least two jets with p_T>20 GeV and $|\eta| < 2.5$ are required, separated from the τ candidate by $\Delta R > 0.5$. In boosted topologies, minimal cuts are applied on the mass and p_T of the $H \to bb$ system. The typical angular separation of the two b-jets can be estimated as in Equation 3.1, where m_H and p_T^H are the mass and transverse momentum of the Higgs boson. A selection of p_T> 300 GeV ensures boosted collimation, with decay products merging within $\Delta R < 0.8$. Additionally, a soft-drop mass requirement of $m_{SD} > 30$ GeV is applied to reduce contamination from the tail of the Higgs mass spectrum, and a minimum separation from the τ candidate of $\Delta R > 0.8$ is imposed.

$$\Delta R(b,b) = \frac{m_H}{p_T^H} \tag{3.1}$$

Specific taggers are used to identify b-jets. In the resolved regime, the HH-BTag neural network assigns to each jet a likelihood of originating from a Higgs boson decay. Developed and validated for non-resonant $HH \to bb\tau\tau$ searches [79], it is based on a Recurrent Neural Network (RNN) architecture, which provides each jet with a probability score of being Higgs candidate between 0 and 1. The $H \to bb$ is then identified as the two resolved jets with associated the two highest scores. In the boosted regime, b-jet identification ranks AK8 jets according to their ParticleNet score (Equation 2.6), with the highest-scoring jet chosen as the $H \to bb$ candidate.

The described procedure is designed to identify the most likely $H \to \tau\tau$ and $H \to bb$ candidates in each event. Categorization in three orthogonal categories based on the topology of the $H \to bb$ system is then defined to improve the analysis sensitivity and optimize discrimination between signal and background across different kinematic regions. The categories are listed below in priority order, which is used to resolve any overlap between categories in case an event satisfies multiple definitions:

- resolved-2b (res2b): the two b-jets identified as $H \to bb$ candidate pass a medium working point of the DeepJet algorithm.
- boosted: one boosted b-jet is found in the event, passing the mentioned selections and satisfying a low-purity working point of the ParticleNet algorithm.
- resolved-1b (res1b): only one b-jet, among the two b-jets identified as originating from the Higgs decay, passes a medium working point of the DeepJet algorithm.

Additional selections on the invariant mass of the two Higgs candidates are applied to minimize background contamination. These cuts remove background tails from the mass spectrum and enhance the analysis sensitivity. They also enable the definition of control regions (CRs) with low signal contamination, allowing dedicated evaluations of background modeling. In the resolved regime, the maximum and minimum values of the $H \to bb$ and $H \to \tau\tau$ candidate masses are set according to the 99% and 0.5% quantiles. The resulting mass windows are 20–130 GeV for the visible mass of the $H \to \tau\tau$ system and 40–270 GeV for the $H \to bb$ system. In the boosted category, the visible mass of the $H \to \tau\tau$ system is required to satisfy $m_{\tau\tau}^{\text{vis}} < 130$ GeV.

3.1.3.1 Boosted category definition

To properly target events in which the two b-jets merge into a single jet at high transverse momentum, a dedicated strategy has been developed, leveraging the performance of ParticleNet, which models jets as unordered particle clouds and is particularly effective in this regime compared to other taggers available at the time of this work, as detailed in Ref. [81]. The ParticleNet tagger score has been included into the analysis framework, and in the boosted regime the $H \to bb$ candidate is defined as the jet with the highest ParticleNet score, which is defined in Equation 2.6. A preliminary assessment of the tagger's performance was carried out by studying the distributions of the ParticleNet score for signal and background samples. To quantitatively evaluate the separation between signal and background processes, the efficiencies were determined and the corresponding ROC curves were constructed. The ROC curves are obtained using Radion signal samples spanning the full range of masses considered in this analysis, while the background is defined as the sum of all background processes. The resulting ROC curves for different resonance masses are shown in Figure 3.2. These results demonstrate that the tagger effectively distinguishes signal from background events, particularly at higher resonance masses where Higgs bosons are more likely to decay with a large Lorentz boost. In particular, at higher resonance masses, for a signal efficiency of 80%, a background efficiency smaller than 10% is achieved, highlighting the tagger's capability to identify boosted $H \to bb$ candidates while suppressing background contamination. CMS central recommendations define three working points, each associated with corresponding corrections for simulated events, as described in Section 2.3.1.4. For this analysis, the loosest working point was chosen to select events in the boosted category. This strategy maintains low background contamination while maximizing signal acceptance, which is particularly important given the limited statistics available in this kinematic region. Although corrections are provided for jets identified as $H \to bb$ in signal events, the topology of events reconstructed as $H \to bb$ from background can vary depending on the analysis. Therefore, a dedicated evaluation of the corrections is required for each analysis; for the current study, this evaluation is discussed in Section 3.1.4.1.

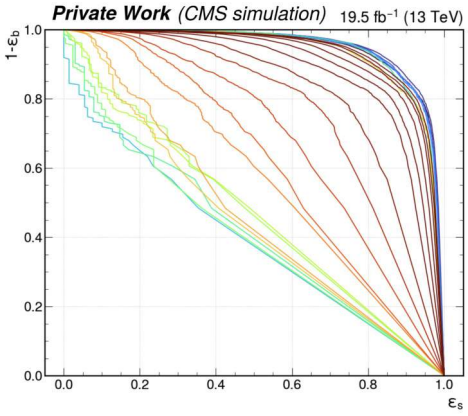

Figure 3.2: The ROC curves corresponding to different ParticleNet score thresholds, evaluated for various signal resonance masses, are shown in the figure. The color map ranges from shades of green to blue, representing masses from 250 GeV to 3000 GeV.

3.1.4 Background modeling

A precise understanding of the signal processes and of the background processes mimicking the signal constitutes a fundamental step in every analysis. In this study, every background contribution except QCD, which is estimated with a data-driven method, is modeled using MC simulations. Residual corrections in the modeling, arising from imperfect emulation of the processes and their interaction with the detector, or from uncertainties on theoretical calculations, are accounted for as described in Section 2.3.2 and their uncertainties are treated as systematic uncertainties in the statistical analysis.

The contribution from $Z\gamma^* \to \ell\ell$ is modeled with NLO samples generated either inclusively or in different bins of the Z boson transverse momentum and different jet multiplicities. Control regions enriched in DY events are defined to validate the modeling of this process: events with two oppositely charged muons and two b-jets are selected. An example of the agreement between data and MC in this region is shown in Figure 3.3. In the $t\bar{t}$ processes, top quarks decay to a b quark and a W boson, with the latter

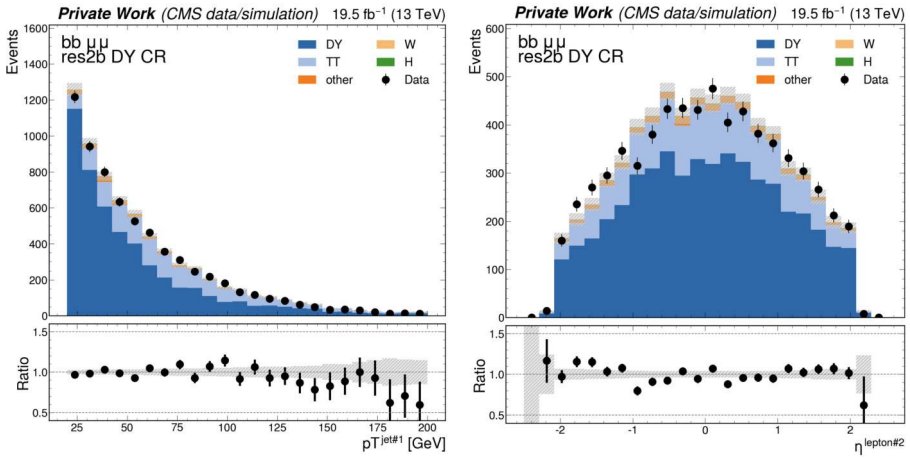

Figure 3.3: Distributions of transverse momentum of the leading b-jet (left) and pseudorapidity the sub-leading lepton (right) identified in a DY-enriched control region, using a dataset corresponding to 19.5 fb^{-1}. The data and simulated event distributions are overlaid, showing, as evidenced by the ratio plot, a generally good agreement between data and simulation events, within the uncertainties.

decaying leptonically or hadronically. As a result, the analysis may reconstruct genuine b-jets and may misidentify leptons originating from W boson decays or other nearby particles. To validate the modeling of backgrounds with MC simulation, a control region enriched in $t\bar{t}$ events is defined in the $\tau_\mu\tau_h$ channel,

applying the same requirements as in the resolved-2b category but with an additional cut on the $H \to \tau\tau$ mass ($m_{\tau\tau} > 130$ GeV), reducing DY and signal contamination, as shown in Figure 3.4. Other minor

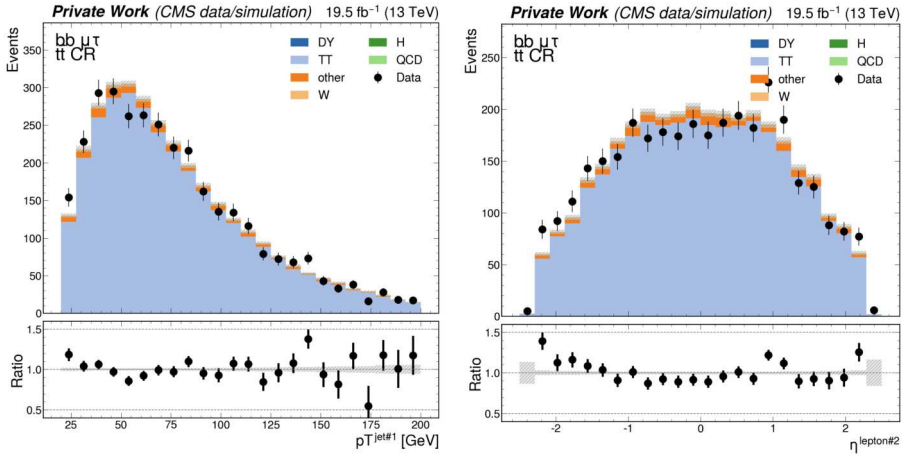

Figure 3.4: Similarly to what is shown in Figure 3.3, the distributions of transverse momentum of the leading b-jet (left) and pseudorapidity the sub-leading lepton (right) are also evaluated in a $t\bar{t}$ control region, showing, also in this case, an overall good agreement between data and simulation.

contributions are also modeled through MC simulations. These include the production of two or three vector bosons (WW, WZ, ZZ, WWW, WWZ, WZZ, ZZZ), single Higgs production via gluon fusion, vector boson fusion, or associated with a vector boson or a top-quark pair, which represent irreducible backgrounds due to their similar final states. Non-resonant $HH \to bb\tau\tau$ production is also included as a background, assuming the SM trilinear Higgs coupling. Additional processes considered are W+jets events, electroweak processes such as V+2 jets, and single top production in both the s- and t-channels, although their contribution is nearly negligible. Processes involving $t\bar{t}$ pairs in association with one or more vector bosons are also accounted for.

During collisions, numerous jets are produced via QCD processes, and due to the limited understanding of these processes, simulations may be insufficient or inaccurate. Therefore, an alternative method is necessary to estimate these backgrounds, which are abundant in a pp collider like the LHC. A data-driven technique known as the ABCD method is used in this study to estimate the multi-jet component. This method relies on selecting two mutually independent variables that effectively characterize the background. In this study, the chosen variables are the isolation quality of the hadronic τ-lepton and the lepton charge in the $\tau\tau$ channel. In the $\tau_h\tau_h$channel, the isolation condition is applied to the τ_h candidate with the lowest-quality isolation. Four orthogonal regions (A, B, C, D) are then defined using selections on these two variables. Region A is signal-enriched, while the other three regions (B, C, D) are background-enriched. In particular region A selects opposite-sign taus with isolation criteria, region B selects same-sign taus with isolation criteria, region C selects opposite-sign non isolated taus, while region D selects same-sign non isolated taus. The number of QCD events in the background regions (B, C, D) is estimated by subtracting the yield of all other background contributions from the observed data. The QCD background shape can be extracted from either region B or region C, with a correction to its normalization derived using the C/D or B/D ratios, respectively. Specifically, one shape is obtained by taking the distribution in region B and normalizing it using the C/D ratio, while the other is derived from region C and normalized using the B/D ratio. The final QCD background shape is taken as the average of these two results, with the difference between them assigned as a systematic uncertainty.

3.1.4.1 Monte Carlo corrections for background processes in the boosted category

The defined selections for identifying boosted objects require corrections to guarantee good data/MC agreement. As described in Section 2.3.1.4, corrections for signal-like jets (resonances decaying to b-quark pairs) are provided centrally by CMS working groups. However, these cannot be applied to background processes, where the boosted jet composition may not correspond to the signal structure and requires

phase-space-dependent corrections. The two main background contributions in this category are DY and $t\bar{t}$ processes. In DY events, the boosted jet may result from gluon splitting or other jets (e.g. ISR/FSR). In $t\bar{t}$ events, one b-quark comes from top decay while the second nearby b-quark may originate from another jet. Given the multiplicity of potential processes reconstructed as a signal-like boosted jet, an effective approach is proposed. Three classes of corrections to simulated events are considered:

- signal-like jets: boosted jets matching a Z or Higgs boson decaying to b-quarks are corrected using central CMS factors.
- background jets: boosted jets not matching Z or Higgs decaying to b-quarks are subdivided into:
  - DY-like events: corrected with values derived in dedicated DY control regions (e.g. W+jets, multi-boson events);
  - $t\bar{t}$-like events: corrected with values derived from $t\bar{t}$ control regions (e.g. single top, $t\bar{t}$V, single Higgs+top).

Effective corrections on background processes are derived by computing the data/MC ratio before and after applying the ParticleNet score selection, isolating the effect of the score selection by ensuring good agreement between data and simulated events prior to its application. The boosted jet p_T spectrum is analyzed for each control region, either DY or $t\bar{t}$-enriched. To preserve sensitivity to the distribution shape while maintaining sufficient statistics per bin, the p_T spectrum is divided into three bins. Distributions are considered together with their associated statistical uncertainties. Efficiencies are calculated as in Equation 3.2 for the specific MC sample under study (i.e. DY or $t\bar{t}$) after subtracting from data the MC yields of all other background contributions. The corresponding uncertainty is evaluated using standard uncertainty propagation, taking into account statistical uncertainties only.

$$\text{efficiency(sample)} = \frac{\text{selected boosted events passing low\_purity working point}}{\text{selected boosted events}} \tag{3.2}$$

The correction is then calculated as in Equation 3.3, with an associated uncertainty evaluated via standard uncertainty propagation of the statistical uncertainty.

$$\text{correction} = \frac{\text{efficiency(data} - \text{otherBackgrounds)}}{\text{efficiency(MC)}} \tag{3.3}$$

The same method is applied to both DY and $t\bar{t}$ enriched region, defined with specific requirements to reduce the contribution from other background that the one under study.

DY-enriched regions are obtained by selecting events containing two muons passing the selections described in Section 3.1.3, defining the $\tau_\mu\tau_\mu$ channel. In addition to the selections applied for signal events, a mass window cut around the Z boson mass ($80 < \text{m}_\text{Z} < 100$ GeV) is applied to further enrich the DY content. Distributions before and after the selection on the ParticleNet score are reported in Figure 3.5. Corrections are then derived in this region and are reported in p_T bins in Figure 3.6.

The same approach is used to extract corrections in the $t\bar{t}$-enriched region, which is defined by applying the signal selections in the $\tau_e\tau_h$ and $\tau_\mu\tau_h$ channels, with an additional cut on the invariant mass of the $\tau\tau$ system $\text{m}_{\tau\tau} > 130$ GeV to ensure orthogonality with the signal region. Distributions before and after the selection on the ParticleNet score are reported in Figure 3.7. The derived corrections are shown in Figure 3.8 and resulted compatible with unity within 1.5σ. These values are applied as weights to boosted events to correct for the ParticleNet score selection, and their uncertainties are propagated to the final statistical analysis as systematic uncertainties.

Eventually, all corrections to the simulation, including pileup reweighting, object identification scale factors, trigger efficiencies, jet energy corrections etc., are applied as described in Section 2.3.2. As an example, the resulting distribution of the transverse momentum of the boosted $H \to bb$ candidate reconstructed in the three channels in the boosted category are reported in Figure 3.9 for the data collected in one data-taking period, demonstrating a sufficient data/MC agreement within the statistical uncertainties after the application of all the corrections.

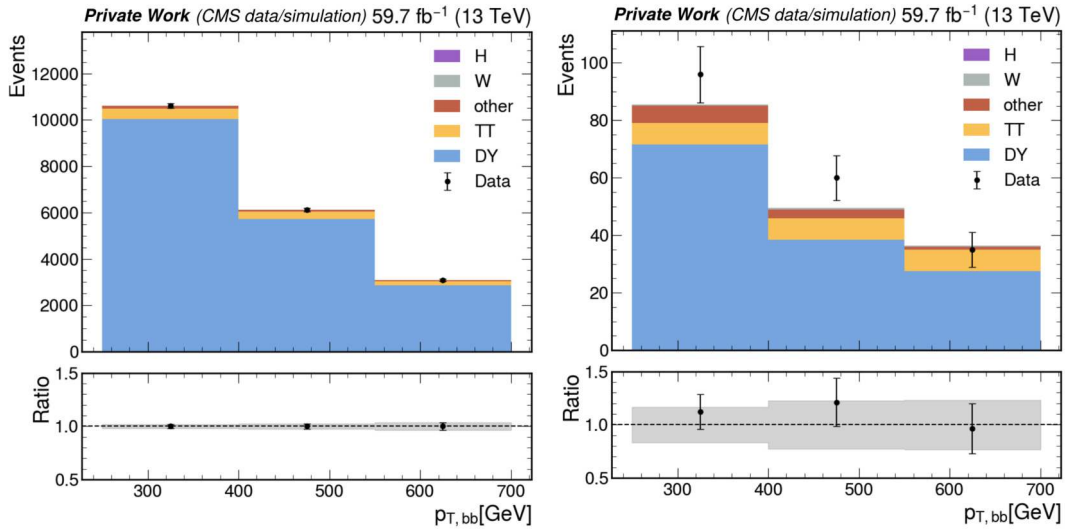

Figure 3.5: Distributions of p_T are reported for the DY-enriched region, before (left) and after (right) applying the ParticleNet score selection. The individual contributions of each background process are shown, along with the data. The selections applied to enrich the DY-enriched region are particularly effective, significantly reducing the contributions from other processes.

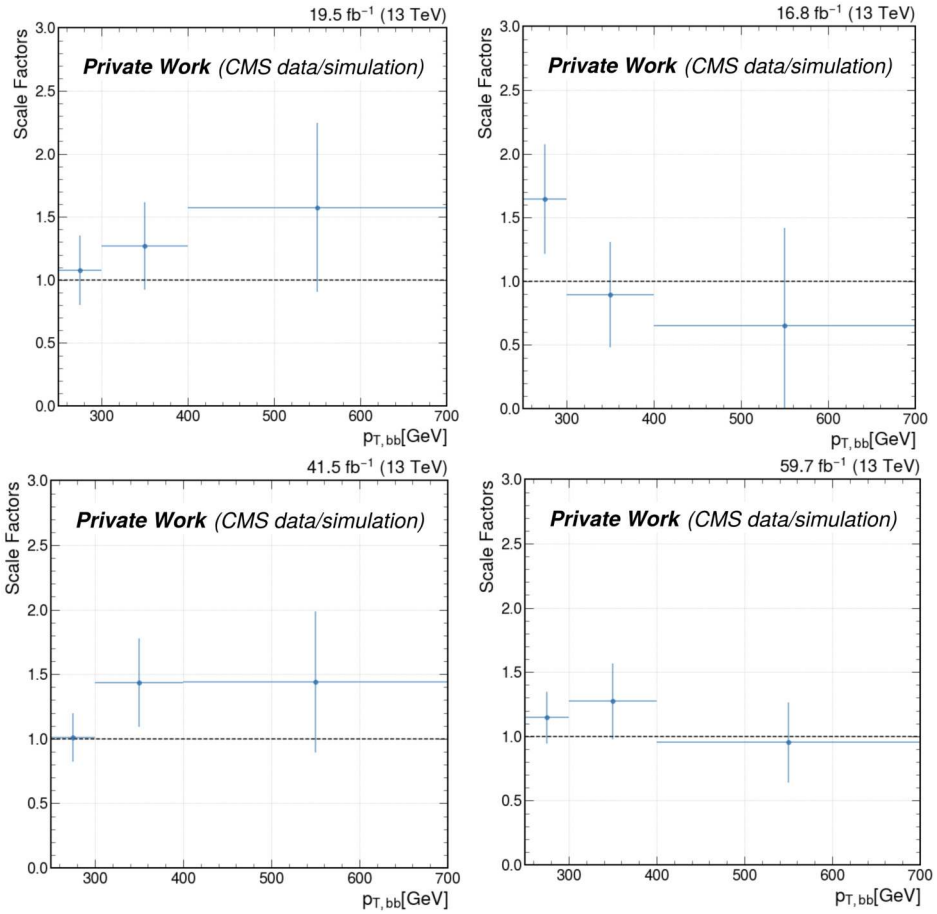

Figure 3.6: The correction factors, evaluated for boosted jets identified as $H \to bb$ candidates and selected with a ParticleNet score cut in DY-enriched regions, are reported with their corresponding uncertainties for each period, divided into three exclusive p_T bins. The corrections are provided for four datasets corresponding to the different data-taking periods.

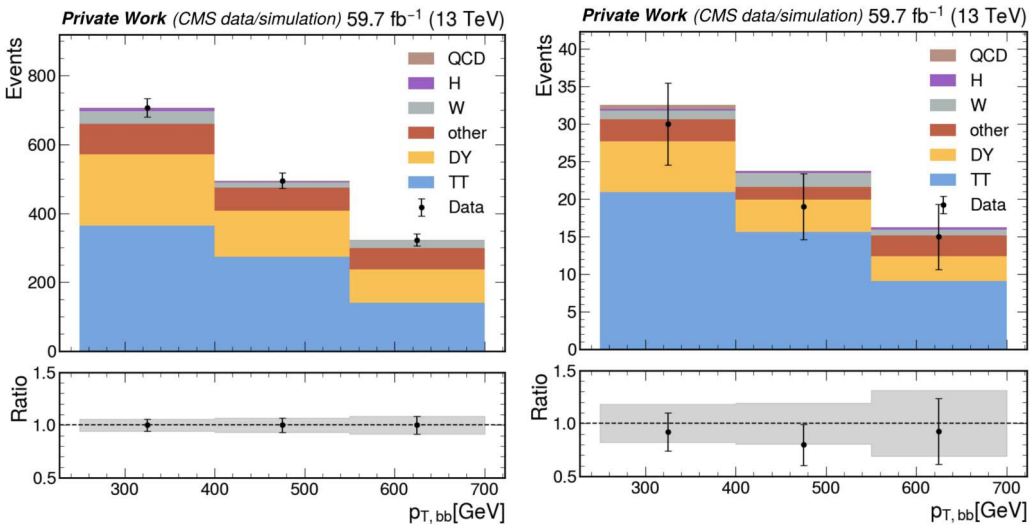

Figure 3.7: Distributions of p_T are reported for the $t\bar{t}$ region, before (left) and after (right) applying a ParticleNet score selection. The individual contributions of each background process are shown, along with the data. The selections applied to enrich the $t\bar{t}$ region overall achieve a good purity of the process.

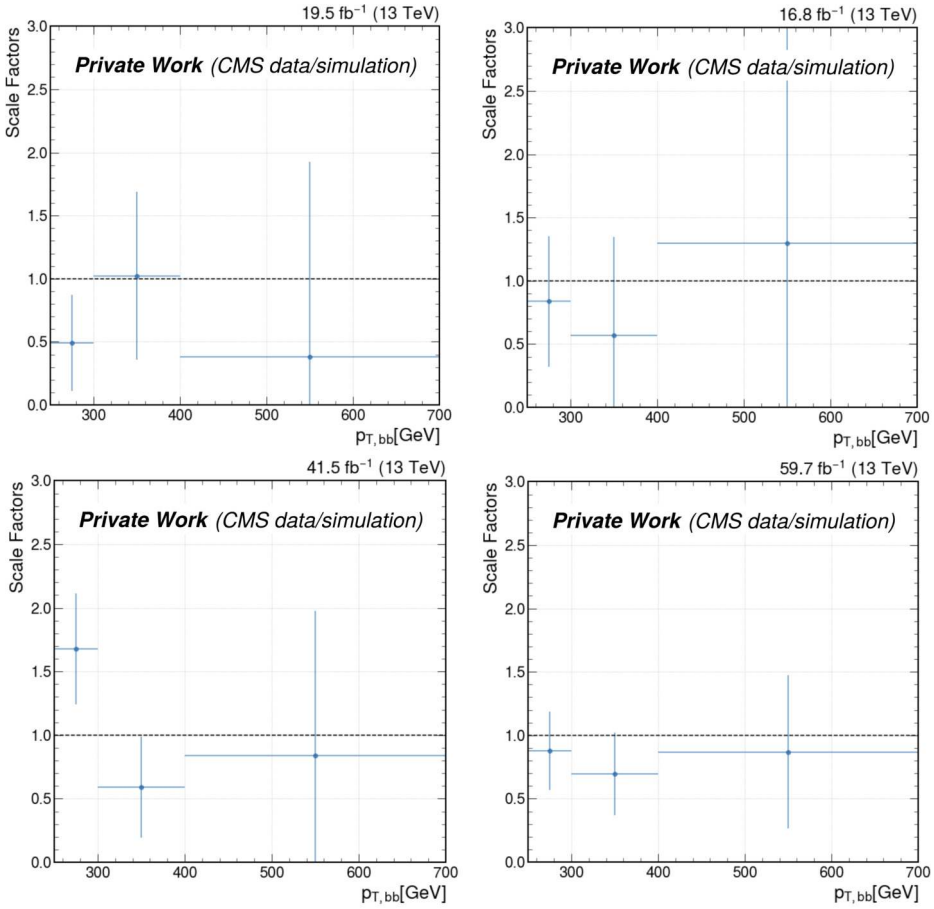

Figure 3.8: The correction factors, evaluated for boosted jets identified as $H \to bb$ candidates and selected with a ParticleNet score cut in $t\bar{t}$-enriched regions, are reported with their corresponding uncertainties for each period, divided into three exclusive p_T bins. The corrections are provided for four datasets corresponding to the different data-taking periods.

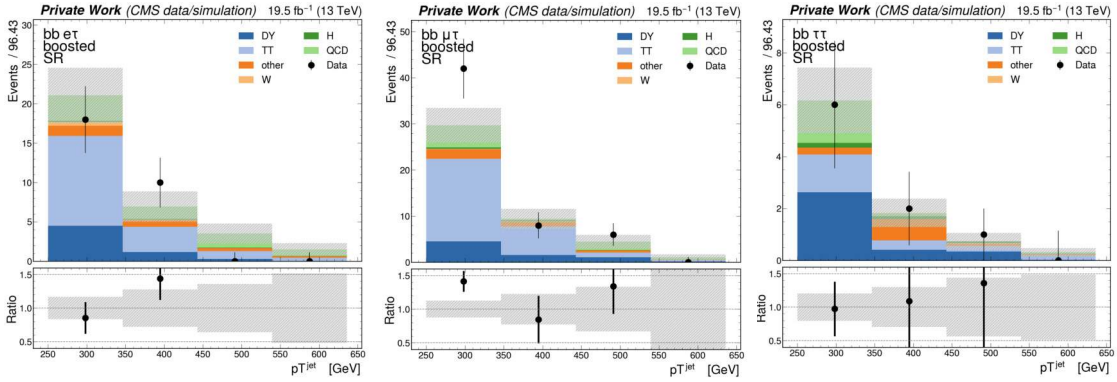

Figure 3.9: Distributions of the transverse momentum of the boosted b-jet are shown for each channel ($\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$) in the signal region of the boosted category. The dataset corresponds to 19.5 fb^{-1}. Both data points and simulated events for each process are displayed, with the simulated backgrounds stacked and associated with a statistical uncertainty represented as a shaded gray region.

3.1.5 Discriminating signal processes from background

The events selected using the methods described in the previous sections are subjected to a statistical analysis to determine whether the observed data are consistent with the background-only hypothesis or if an excess is present. The sensitivity of any analysis relies on the discriminating variable chosen to separate signal from background distributions. The signal targeted in this study would manifest as a bump in the invariant mass spectrum of the final-state particles, making this kinematic variable a natural candidate. However also other variables are expected to be useful to discriminate signal and background events. Recent advances in machine learning have demonstrated remarkable effectiveness in classifying events in high-energy physics analyses, significantly improving both performance and sensitivity. For instance, a DNN was implemented in the non-resonant $HH \to bb\tau\tau$ search, where it was shown to outperform traditional kinematic discriminants [79].

A parameterized DNN was developed, which is parameterized on different mass and spin hypotheses (two spin hypotheses and 25 mass points). A comprehensive set of input features is used for training, including kinematic variables of the τ, the $H \to \tau\tau$, the b jets and the $H \to bb$ candidates, as well as the decay modes and charges of the τ leptons, the data-taking period, the resonance spin and mass, the missing transverse momentum, and the tagging scores of the resolved b jets. The discriminating network is trained using both signal and the main background samples from all data-taking periods. These include signal samples of both spin hypotheses across the full mass range considered, DY samples, $t\bar{t}$ samples, and $t\bar{t}$H events. The DNN output eventually is a score that is used as final discriminant. Events are evaluated for every analysis category, data-taking period, mass point, and spin hypothesis, resulting in 1800 distributions. The binning of these distributions is chosen so that the signal yield is constant across the spectrum, while ensuring a minimum yield for the main backgrounds (DY and $t\bar{t}$) to allow meaningful statistical analyses when performing the binned maximum-likelihood fit. DNN score distributions for the analysis categories are shown in Figure 3.10 as an example for a spin-0 resonance with a mass of 1000 GeV. While the simulation distributions are visible across the entire range, data are not shown in the most significant bins since the analysis is currently still "blinded" (see Section 2.3.3).

3.1.6 Systematic uncertainties

The description of signal and background events in the analysis phase space is affected by several sources of uncertainty. Limited knowledge of the underlying processes, discrepancies between MC samples and data, and the finite accuracy of the detector simulation influence the overall modeling, as extensively discussed in Section 2.3.2.1. Uncertainties are typically categorized as normalization and shape uncertainties: the former affect the total yield of a process, while the latter also impact its differential distributions. The normalization uncertainties considered in this search are all implemented as log-normal nuisance parameters and are listed below:

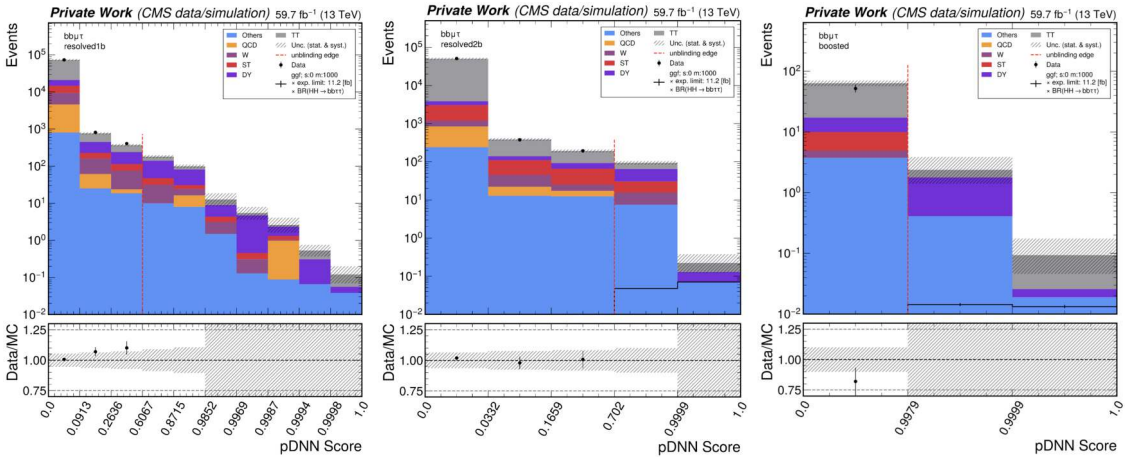

Figure 3.10: Example of parameterized DNN score distributions in the analysis category for a resonance with mass 1000 GeV and spin 0. From left to right, distributions in the res1b, res2b, and boosted are reported. The signal has been scaled for a better visualization. Background bins are represented with their statistical uncertainty shown as a shaded gray region, while uncertainties on data are shown as error bars. The bins most sensitive to the signal have been blinded.

- Luminosity is used to normalize every MC event, therefore the corresponding uncertainties must be accounted for. The uncertainties related to the measurement of the integrated luminosity affect both the signal and background normalizations. Correlated and uncorrelated components among different data-taking periods are considered, with values ranging approximately between 0.5% and 2%.
- The QCD background estimation method has a normalization uncertainty ranging from 5% to 60%, depending on the category, channel, and data-taking period considered.
- The total uncertainties related to the Higgs boson branching fractions are $^{+1.25\%}_{-1.27\%}$ for the $H \to bb$ decay and $\pm 1.65\%$ for the $H \to \tau\tau$ decay [82]. These uncertainties are included because the final cross section is evaluated taking into account the branching fractions for this specific final state.
- Theoretical predictions used to generate several processes are affected by uncertainties related to the parton distribution functions (PDFs), QCD scale variations, and the strong coupling constant α_s. Uncertainties range from 0.4% to 10% for the QCD scale, between 1.3% and 4.2% for the PDFs, and between 0.5% and 2.6% for α_s, depending on the process considered (e.g., $t\bar{t}$, W+jets, single top, single Higgs, diboson, and triboson samples production) [82].

Shape-related systematic uncertainties are evaluated using templates obtained by varying each source by ± 1 standard deviation. The shape uncertainties considered in this search are the following: QCD scale and PDF uncertainties; inefficiencies in the L1 trigger affecting data collected during the early Run 2 period due to timing misalignments in the ECAL and muon chambers; pileup reweighting and pileup identification efficiency; electron reconstruction and identification efficiency; τ identification; τ energy scale; electron and muon energy scale and resolution; jet energy scale and resolution; trigger efficiencies; b-tagging efficiencies. More details on the origin of these corrections can be found in Section 2.3.2.1. In addition, uncertainties related to the limited number of simulated events are considered on a bin-by-bin basis.

3.1.7 Results

The DNN scores serve as input to a binned maximum likelihood analysis (see Section 2.3.3 for details on the statistical analysis). The uncertainties detailed in Section 3.1.6 are used as nuisance parameters in the likelihood function. Expected upper limits at 95% confidence level [36], corresponding to the value below which the signal is expected to lie in 95% of repeated experiments, are extracted for the X $\to$ HH cross section both in the case of a spin-0 and a spin-2 resonance, for a wide range of mass points spanning from

250 GeV to 3 TeV. Expected upper limits at 95% CL are shown in Figure 3.11 for the final combination of all categories, exhibiting similar results for the spin-0 and spin-2 hypothesis. Maximum sensitivity is expected near 1 TeV, after which it gradually decreases. This is because, at higher masses, the likelihood of producing boosted objects increases, and the current search is not optimized for boosted $\tau\tau$ events. The results are also shown in Figure 3.12, separated by category as a function of the resonance mass.

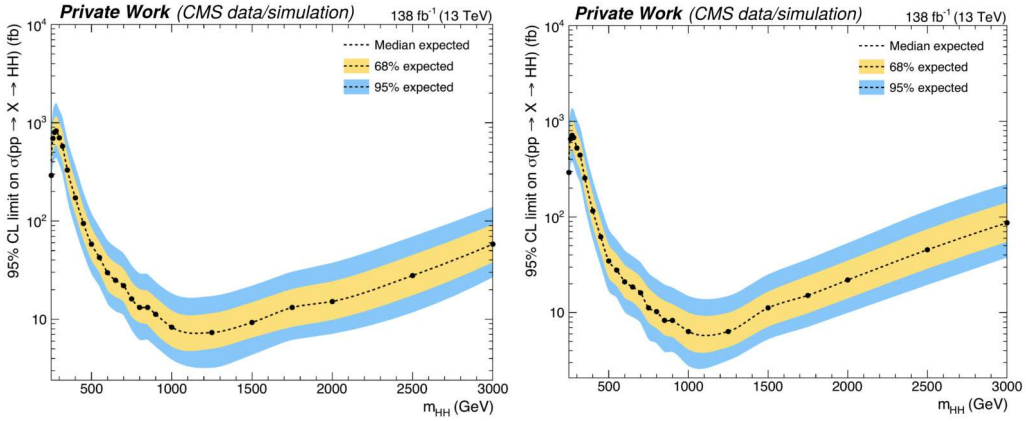

Figure 3.11: Expected upper limits at 95% CL on the cross section for the production of a resonance particle X with spin 0 (left) and spin 2 (right), decaying to a pair of Higgs bosons, are reported as a function of the resonance mass X, assumed to be equal to the HH invariant mass. The expected median values are shown as dashed lines, while the yellow and blue bands represent the 68% and 95% confidence intervals, respectively.

The boosted category becomes the dominant contributor beyond approximately 1.5 TeV, where heavy resonances provide Higgs bosons and their decay products with a high Lorentz boost. This search provided a valuable opportunity to exercise techniques that will be needed for Run 3, specifically the definition of a dedicated boosted $H \to bb$ category and the derivation of custom corrections for background processes in this phase space.

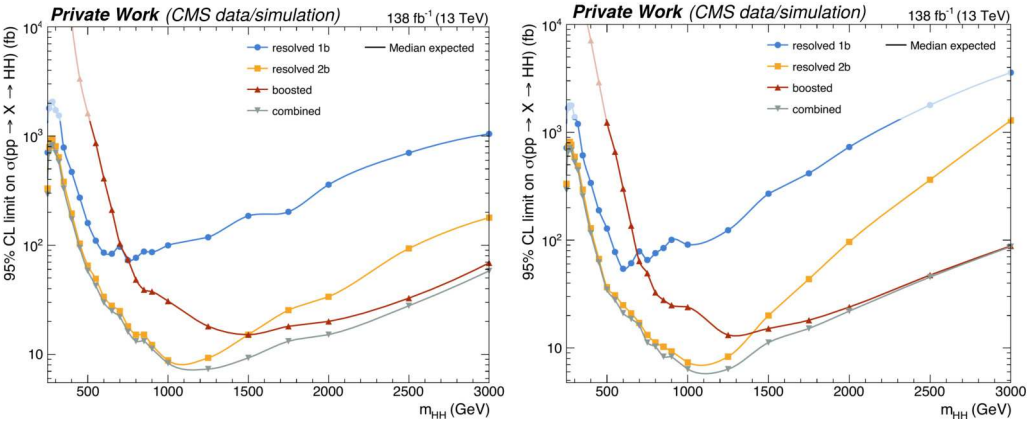

Figure 3.12: Expected upper limits at 95% CL on the cross section for the production of a resonance particle X with spin 0 (left) and spin 2 (right), decaying to a pair of Higgs bosons, are shown as a function of the resonance mass, for the individual contributions from the different categories and for their combination.

3.2 Search for pair production of Higgs bosons in the $bb\tau\tau$ final state

Searches for non-resonant production of di-Higgs investigate the existence of this process by measuring its production cross section and they also target the measurement the Higgs self-coupling. However, given the rarity of this process and the available statistics, current searches at the LHC have not observed this process yet. The search presented here follows the work carried out on Run 2 data, documented in Ref. [79]. It investigates the production of Higgs boson pairs via gluon fusion and vector boson fusion, using 62.4 fb^{-1} of data recorded at the CMS experiment during LHC Run 3 in 2022 and 2023. Significant advancements at both the trigger level and in offline object reconstruction, compared to Run 2, allow for improved performance in this analysis. In particular, this non-resonant search introduces new triggers, such as one targeting a τ-lepton pair and a jet, a VBF parking strategy (see Section 2.2.2), improved τ identification via DeepTau at the HLT level, and a boosted $\tau\tau$ channel. Additional improvements include a VBF tagging algorithm, a fake-factor method for estimating fake τ-leptons from multi-jet events, and a newly developed DNN for signal extraction in different regions. Building on my experience defining a boosted category in the resonant search (Section 3.1), my contribution focused on developing a new boosted category targeting the Higgs decay into two b-quarks. Compared to the previous non-resonant iteration, selections have been re-defined and optimized, with updated identification criteria for the Higgs candidate in this kinematic region. A DNN is specifically optimized to discriminate signal from background processes, and the working point is tuned for this analysis, including in situ evaluation of background corrections. Apart from these improvements, the analysis workflow and methods remain largely similar to those described in Section 3.1, and many aspects will therefore be summarized briefly here. Figure 3.13 provides a schematic overview of the main steps of the analysis.

3.2.1 Datasets and triggers

This search uses the data collected by the CMS experiment at $\sqrt{s} = 13.6$ TeV, corresponding to an integrated luminosity of 62.4 fb^{-1} during the 2022 and 2023 data-taking periods, yielding 34.6 fb^{-1} and 27.8 fb^{-1}, respectively, at the LHC. Simulated samples of signal and background processes have been produced using state-of-the-art theoretical calculations at leading order, next-to-leading order, or next-to-next-to-leading order. The simulated background processes are the same as those described in Section 3.1.1. DY, W+jets, single-top with b-quark +lepton and neutrino, $t\bar{t}$ + di-boson, and tri-boson events are generated with MadGraph5_aMC@NLO, while $t\bar{t}$, single-top, single-Higgs, di-boson, and $t\bar{t}$H events are generated with POWHEG.

Triggers for $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$ channels follow the same strategy described in Section 3.1.1, selecting events with isolated leptons combined with cross-lepton triggers. Thresholds on transverse momentum are applied to ensure high trigger efficiency, specifically in the range between 12 and 30 GeV for electrons, 3 and 24 GeV for muons, 20-45 GeV for τ pairs, 20-30 GeV for the τ_h lepton firing the cross triggers. One of the main improvements compared to previous analyses comes from the addition of a trigger requiring two τ_h and a jet. This additional trigger was demonstrated to bring an integrated gain of approximately 13% in signal acceptance. Specific triggers with requirements on MET (with p_T greater than 120 GeV), as well as a new trigger introduced in Run 3 for boosted τ-leptons (with p_T greater than 230 GeV), are also used to target the boosted $\tau\tau$ topology, which, in this study, is treated as an additional, orthogonal channel to the three decay modes already described. The VBF processes are characterized by specific kinematic properties, which enable specific identification requirements such as two forward jets with a large pseudorapidity separation and a high invariant mass. In the non-resonant search with Run 2 data, a trigger targeting this topology was developed, however with Run 3, an increased number of triggers addressing this process became available and were therefore employed in this search, showing for instance that the inclusion of specific VBF triggers in the $\tau_h\tau_h$ channel leads to a signal gain of up to 30% compared to the previous strategy. In the latest dataset available, a new parking strategy was also introduced, extending to all channels.

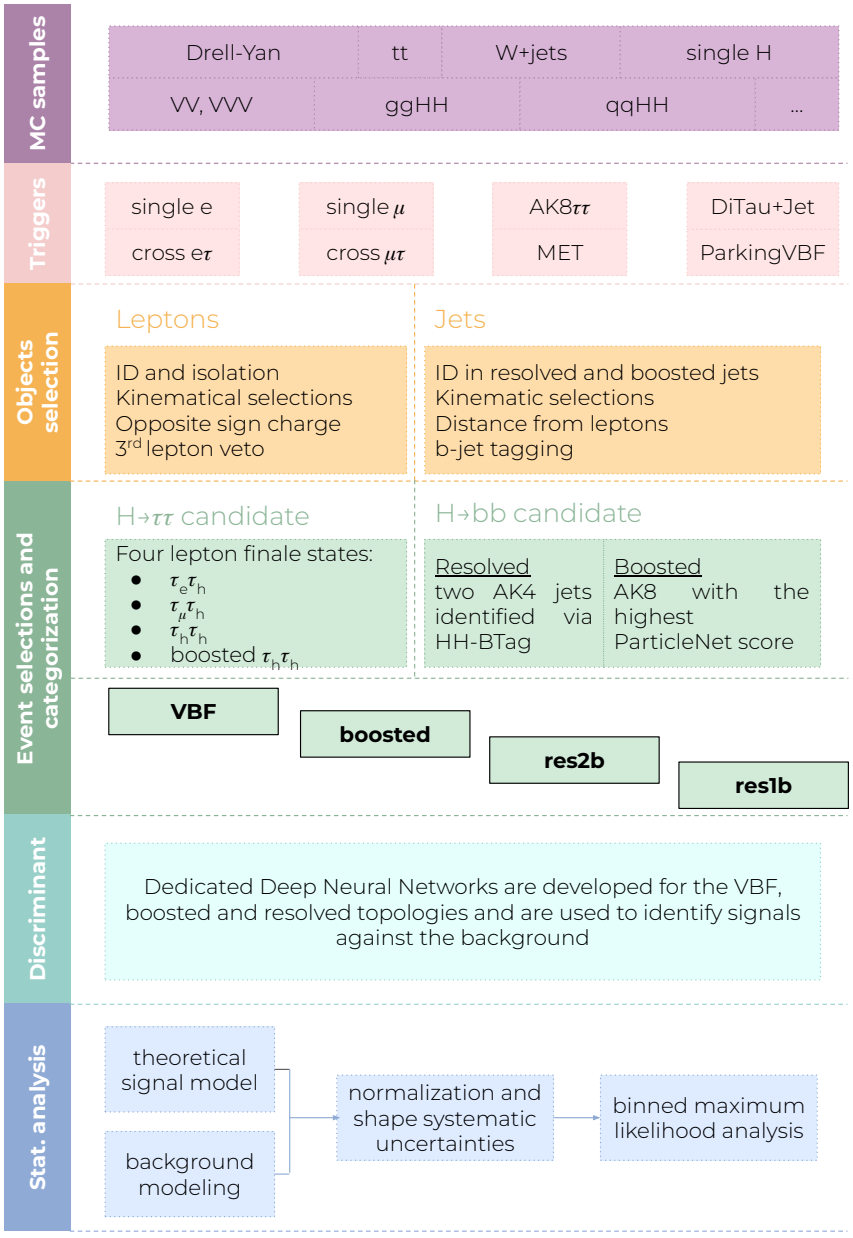

Figure 3.13: A representative schematic of the analysis workflow is shown here. The starting point consists of the datasets collected by CMS during Run 3. The workflow largely follows the procedure illustrated in Figure 3.1, with the main differences being the inclusion of new triggers available in Run 3, the definition of an additional channel targeting the boosted $\tau\tau$ final state, and the introduction of a VBF category. The rest of the analysis strategy remains similar to what was previously described, with dedicated DNNs adapted to the different topologies considered in this study.

3.2.2 Physics objects reconstruction

Object reconstruction follows the procedures already described in Section 2.3.1. In this search, identification is based on selections similar to those described in the resonant search discussed in Section 3.1. More specifically, signal muons are required to pass a tight WP for identification and have a relative isolation of I^{ℓ}_{rel} <0.15, while veto muons are defined with a medium WP and looser isolation I^{ℓ}_{rel} <0.3. Signal electrons are selected with a tight WP corresponding to 80% efficiency and I^{ℓ}_{rel} <0.1, while veto electrons pass a loose WP and the requirement on isolation is of I^{ℓ}_{rel} <0.3. τ-leptons are reconstructed with HPS and identified with the DeepTau algorithm discriminating τ-leptons from jets with a medium WP, muons with a tight WP and electrons with a very-very-loose WP. Additionally, boosted τ-leptons are reconstructed as described in Section 2.3.1.5, with an identification WP defined to achieve maximum significance. Ultimately, jets are reconstructed as resolved or boosted depending on their radius parameter. To reject noise, a tight WP of the jet identification discriminant is applied. Data-to-simulation corrections are applied to all objects as described in Section 2.3.2.1.

3.2.3 Event selection and categorization

Compatibility of leptons with the primary vertex is ensured with $|\Delta_z| < 0.2$ cm, with electrons and muons also satisfying $|\Delta_{xy}| < 0.045$ cm. Potential overlap or misidentification is reduced requiring the two $\tau\tau$ candidates to be separated by $\Delta R > 0.5$ and have opposite electric charge. Also, to reduce contamination from $Z\gamma^* \to \ell\ell$ processes, a third-lepton veto is applied. The reconstruction of Higgs candidates follows the same procedure described in Section 3.1.3, addressing three final states: $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$, with the addition of a new boosted $\tau_h\tau_h$ channel. The $H \to bb$ reconstruction is also based on the methods of the previous search, but now uses ParticleNet instead of DeepJet in the resolved regime, where a dedicated retraining of the HH-BTag algorithm was performed.

For the resolved regime, a dedicated retraining of HH-BTag was carried out for this search by the developing team of the tagger. For non-resonant signal samples across all production modes and data-taking periods, the retrained HH-BTag achieves purities between 94–98%, outperforming ParticleNet (91–95%). These results favored the use of the retrained HH-BTag for identifying the two resolved b-jets. In the boosted regime, the $H \to bb$ candidate is required to have p_T> 300 GeV and is identified as the boosted jet with the highest ParticleNet score, defined as in Equation 2.6, following the methods described in Section 3.1.3. Within the context of this analysis a neural network dedicated to the VBF topology identification, similar to the HH-BTag architecture, was used for tagging specifically VBF jets. This approach provides a purity above 90%, to be compared to the level of 70% that was achieved by simply selecting the jets with the highest invariant mass.

Four mutually exclusive categories are defined and are listed below, in order of priority:

- VBF: events with two VBF jet candidates with $m_{jj} > 500$ GeV and $\Delta\eta_{jj} > 3.0$, specifically selected by the VBF tagger.
- boosted: events with a boosted $H \to bb$ candidate passing the custom WP.
- resolved-2b (res2b): events with two resolved jets identified with the highest HH-BTag score and passing the medium ParticleNet WP.
- resolved-1b (res1b): events with exactly one jet passing the medium WP.

Following the strategy used in the Run 2 search [79], additional mass-based selections are imposed on Higgs candidates in the resolved categories to enhance sensitivity and retaining 99% of the signal. Kinematic selections for the boosted category were studied, and the optimization of the expected significance in this phase space was a main focus of my contribution. These aspects will be discussed in the following sections.

3.2.3.1 Definition of the boosted category

The definition of a boosted category for $H \to bb$ events was first achieved by evaluating the optimal kinematic selections to maximize signal acceptance while reducing background contributions. Subsequently, a custom working point for the ParticleNet tagger was optimized to identify the selection that provides the best sensitivity in this phase space.

Kinematic selections

Aiming to enhance signal purity by retaining events around the Higgs boson mass while suppressing background-enriched tails, specific kinematic selections have been studied. In particular, the impact of applying a soft-drop mass window of $80 < m_{SD} < 170$ GeV and a transverse momentum requirement of $p_T > 300$ GeV (motivated by Equation 3.1) has been evaluated for both signal and background processes. To validate the impact of these selections, both individually and in combination, distributions of the ParticleNet score for signal and background samples have been studied. The results, shown in Figure 3.14, indicate that the combination of the mass and transverse momentum requirements achieves the best background suppression while maintaining a comparable signal acceptance. To ensure that these selections do not degrade the event purity, a generator-level matching to signal events has been performed, showing that both the individual and combined selections improve the overall signal purity (defined as the fraction of matched signal events over the total), compared to the scenario in which no kinematic selections are applied. Therefore, both the p_T and m_{SD} selections are retained in the category definition. Extending the methodology used in the search described in Section 3.1, the boosted jet is identified as the

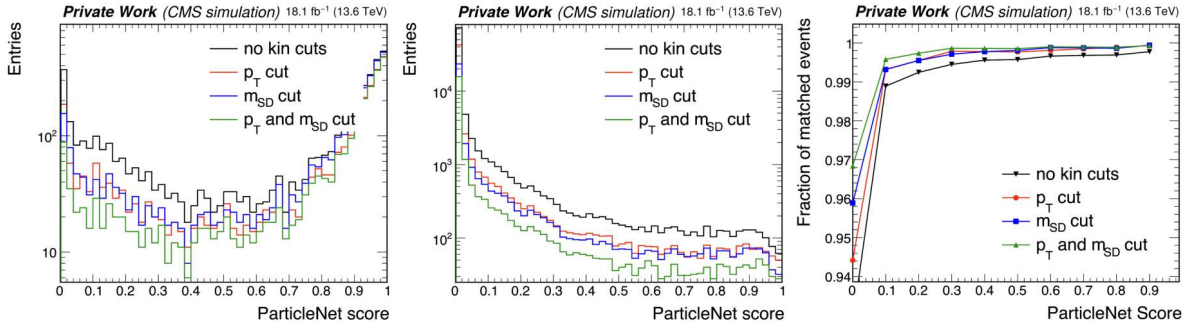

Figure 3.14: Impact of kinematic selections on the ParticleNet score and generator matching. Left and middle: Distributions of the ParticleNet score for signal and background events under different kinematic selections. Right: Fraction of events being matched at generator level with signal events, in a signal sample, as a function of ParticleNet score.

$H \to bb$ candidate by selecting the AK8 jet with the highest ParticleNet score (Equation 2.6). Given the limited statistics, the boosted category combines the three channels: $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$, which will be collectively referred to as "$X\tau_h$" in the following. At the time of this analysis, dedicated CMS-wide corrections for simulated boosted jet tagged with the ParticleNet were not yet available for Run 3. Therefore, a custom working point was defined to balance signal efficiency and background rejection and its optimization is described in Section 3.2.3.1.2.

Working point optimization

To improve the overall sensitivity of the analysis in this phase space, an optimization procedure was performed on the selection threshold of the ParticleNet score, which is used to identify the boosted $H \to bb$ candidates. The goal of this procedure is to determine the working point that provides the best expected performance in terms of signal sensitivity. The choice of ParticleNet threshold affects both the signal efficiency and the background rejection. In a statistically limited category such as the boosted one, applying a very tight selection can significantly reduce the number of signal events. To evaluate the optimal balance, a scan is performed over a range of ParticleNet score thresholds spanning from 0.1 to 0.9. For each threshold value, the full statistical analysis chain is repeated using the score distributions obtained from the dedicated DNN developed for signal extraction in this kinematic region, as detailed in

Section 3.2.5.1. Specifically, for each ParticleNet threshold considered in the scan, the event selection is re-applied accordingly and the expected upper limits at the 95% CL on the non-resonant Higgs boson pair production cross section are computed.

Evaluating the expected upper limit over the entire range of the ParticleNet score enables the identification of a minimum, corresponding to a maximum in sensitivity. The ParticleNet score cut applied at that point is therefore identified as the optimal working point for the boosted $H \to bb$ phase space. This approach allows for a performance-based selection of the ParticleNet cut. The results of the scan are summarized in Figure 3.15, where the expected limits as a function of the ParticleNet threshold are shown. According to these results, the ParticleNet score selection at 0.75 is selected as optimal because it achieves the best balance between retaining signal events and rejecting background, leading to the best expected limit. A similar behavior as a function of the ParticleNet score is observed

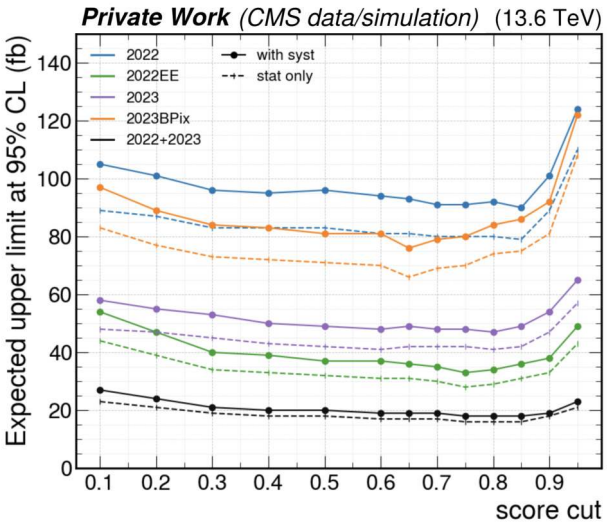

Figure 3.15: Expected upper limits at 95% CL on the non-resonant $HH \to bb\tau\tau$ production cross section as a function of the ParticleNet score cut applied in the boosted category. Different line colors correspond to different data-taking periods (listed separately to account for variations in experimental conditions), although the results among periods only differ in the amount of statistics. Each curve represents the result obtained using the score of the DNN optimized for the boosted category as the final discriminant. Results without systematic uncertainties are shown with dashed lines, while those including systematics are displayed with solid lines.

for the different data-taking periods and for their combination, thereby motivating the choice of a common working point of 0.75 for selecting $H \to bb$ candidates.

3.2.4 Background modeling

Signal and background modeling follows the procedure described in Section 3.1.4, with the exception of QCD multi-jet estimation, for which a new methodology based on the Fake Factor (FF) estimation was employed. Corrections, including pileup reweighting, scale factors object identification, trigger, etc., are applied as described in Section 2.3.2.1.

Part of the background contamination, including QCD multi-jet, $t\bar{t}$, and W+jets, originates from jets misidentified as τ_h candidates. In $t\bar{t}$ events, these fakes typically come from jets from top decays, while in QCD multi-jet events both jets from quarks and jets from gluons can be misidentified. In W+jets events, fakes mostly come from initial or final state radiation or from the W decay. Since the modeling of such jets in MC is not always reliable, their contributions are estimated from data. In this analysis, unlike previous studies targeting this final state that relied on the ABCD method, the FF method is employed to estimate the contribution of jets misidentified as τ_h. The FF method was developed within the context of other analyses, with a detailed description reported for instance in Refs. [83, 84]. FFs

are measured in three control regions enriched in specific backgrounds: $t\bar{t}$ (95% purity), QCD multijet (99%), and W+jets (80%). They quantify the probability for a jet to pass the nominal τ_h identification criteria and are used to estimate the fake-τ_h background in the signal region.

Studies were performed on the DY modeling in dedicated control region, focusing on final states with two electrons or two muons originating from τ decays and selecting events within a mass window around the $H \to \tau\tau$ candidate. A reweighting of the DY process is applied to correct the MC simulation and achieve good agreement with data. Several variables are evaluated, and the number of central jets, defined as jets with $p_T > 20$ GeV and $|\eta| < 2.5$, is found to be the most sensitive observable. By fitting the discrepancy between data and MC in this variable for each category and data-taking period, an effective correction is derived for the DY samples. Different functional forms were tested, and the quadratic fit is found to provide the best performance. The closure of this method is verified by checking the agreement in other kinematic variables before and after applying the DY weights. The $t\bar{t}$ and DY processes are evaluated in the $\tau_e\tau_\mu$ and $\tau_\mu\tau_\mu$ final states, which are not included in this analysis as final states. An example is shown in Figure 3.16, where good agreement between data and MC is observed for both the DY- and $t\bar{t}$-enriched regions.

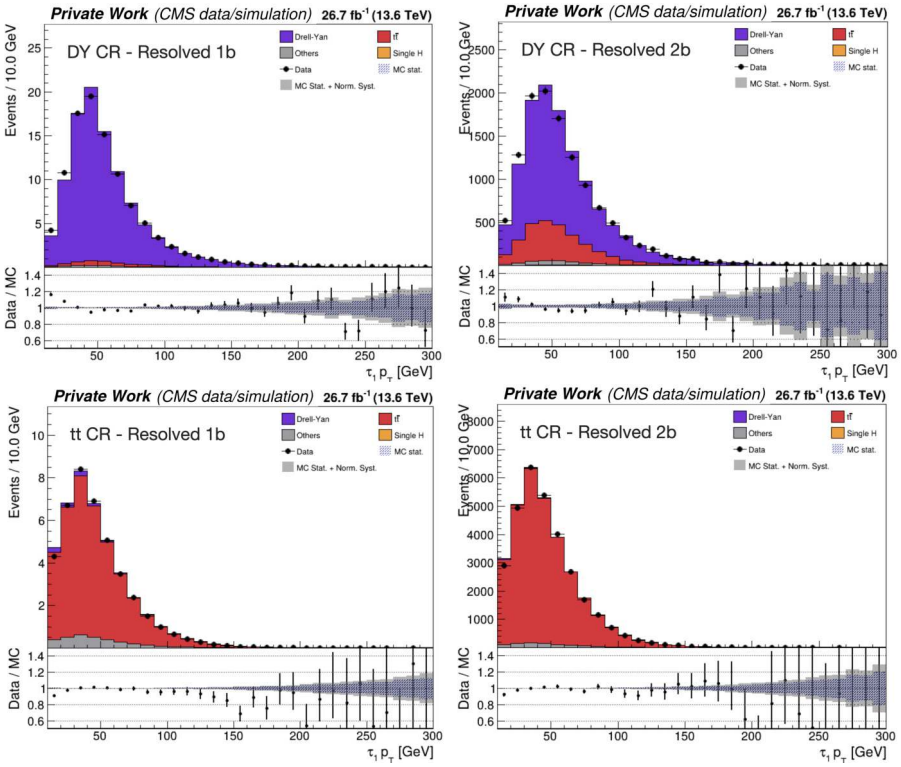

Figure 3.16: The transverse momentum distributions of the leading τ lepton are shown in the resolved categories for the DY- and $t\bar{t}$-enriched control regions in the $\tau_\mu\tau_\mu$ and $\tau_e\tau_\mu$ final states, respectively. The dataset corresponds to 26.7 fb^{-1}. Both data points and simulated events for each process are shown, with simulation backgrounds stacked. Good agreement between data and Monte Carlo is observed within uncertainties.

3.2.4.1 Monte Carlo corrections for the boosted $H \to bb$ candidates

The working point selection of the boosted category allows efficient identification of boosted objects, but corrections are needed to ensure good agreement between data and simulation. As discussed in Section 2.3.2.1, a dedicated CMS working group usually provides collaboration-wide prescriptions for

[0]The data-to-simulation agreement in the boosted category has been further studied, and additional corrections have been evaluated, as described in Section 3.2.4.1.

signal-like jet corrections, i.e. for resonances decaying into pairs of b quarks, such as $H \to bb$ or $Z \to bb$ events. However, at the time of writing, these results are not available yet. Therefore, considering that the values obtained for Run-2 were compatible with unity, and considering that no significant variations are expected a part from statistics, a correction factor of 1 is assumed (corresponding to a good emulation in simulation) with an associated conservative uncertainty of 20% to account for possible deviations from the observed behavior. For background processes, phase-space-dependent corrections are required. The same method employed in Section 3.1.4.1 is used, with a few exceptions. Due to much more limited statistics, corrections are computed by combining all data-taking periods together and two p_T bins are used. The resulting distributions of the boosted jet momentum are shown in Figure 3.17, before and after applying the ParticleNet score cut, for both DY $t\bar{t}$ control regions.

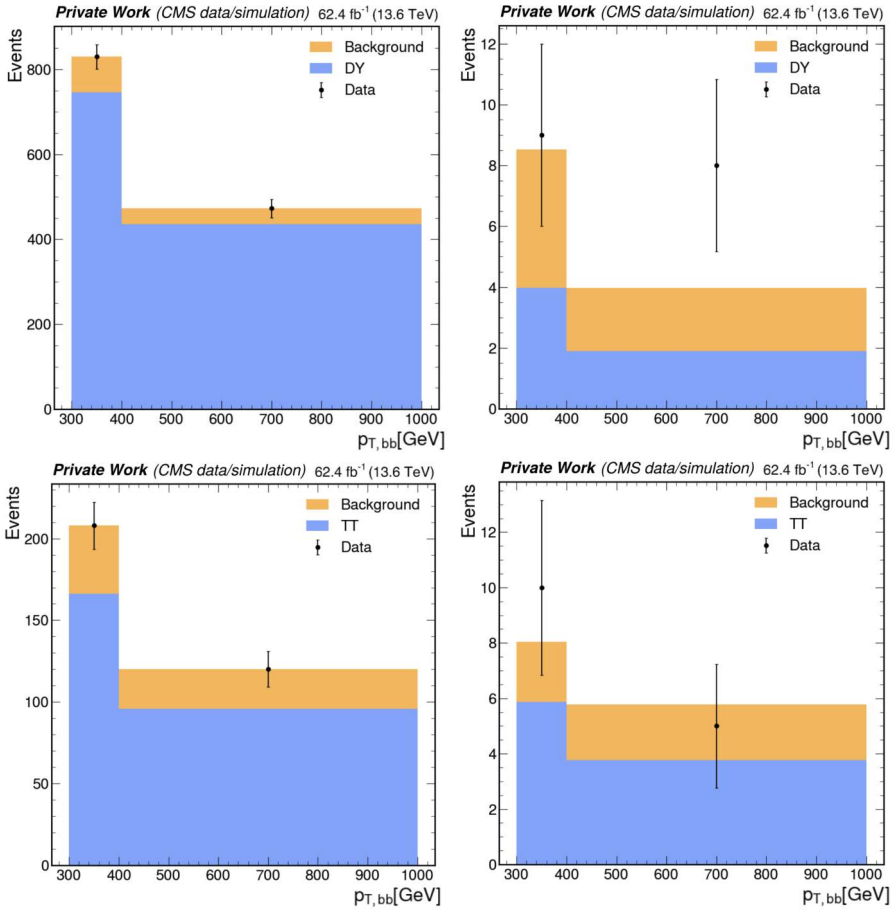

Figure 3.17: Distributions of p_T are reported for the DY-enriched region (top row) and the $t\bar{t}$ region (bottom row), before (left) and after (right) applying a ParticleNet score selection. The selections applied to enrich the control region in the process under study (either DY or $t\bar{t}$) achieve overall good purity of the process.

Corrections are then evaluated following the same procedure described in Section 3.1.4.1. The results are consistent with unity within 2σ and are summarized in Figure 3.18, both for DY and $t\bar{t}$ processes. In the analysis, these corrections are applied as event weights to events with boosted jets that do not match a signal-like jet at generator level and pass the ParticleNet score selection, and their uncertainties are propagated to the final statistical analysis as systematic uncertainties.

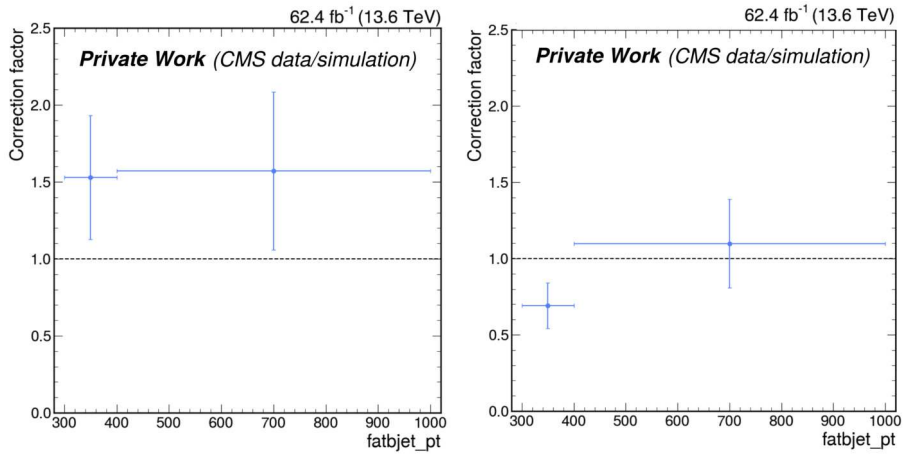

Figure 3.18: Boosted corrections computed using the full Run-3 dataset in two exclusive p_T bins. Results for DY and $t\bar{t}$ processes are shown in the left and right panels, respectively.

3.2.5 Discriminating signal processes from background

In the resolved categories (res1b and res2b), signal extraction uses a DNN similar to the one mentioned in Section 3.1.5. In the VBF category, the same DNN architecture is used, but trained on VBF samples with loose selections. The output provided by the DNN is used as the discriminant variable between signal and background processes to perform the final statistical analysis. In all categories, to improve statistical stability and sensitivity in the limit-setting procedure, the DNN output histograms undergo a bin-merging procedure. This algorithm adjusts the bin boundaries such that the signal distribution is approximately flat in terms of event yield per bin. This technique reduces statistical fluctuations in low-statistics regions while preserving the discriminating power between signal and background shapes. This approach enhances the robustness and stability of the fit.

3.2.5.1 Optimization of a DNN discriminant for the boosted category

A different DNN was specifically developed for the signal extraction in the boosted category. A wide range of input variables was assessed along with their correlation matrix, to determine the most sensitive variables to the kinematic differences between signal and background. Studying the correlation and sensitivity of several variables resulted in defining a set of twelve variables used in the DNN, as shown in Figure 3.19. This includes the p_T of the di-Higgs system; the mass and p_T of the $H \to \tau\tau$ and $H \to bb$ candidates; the p_T of the leading and sub-leading τ candidates; the pseudorapidity separation between the τ candidates; the angular separation between each τ candidate and the $H \to bb$ candidate; the pair type (i.e. the final state $\tau_\mu\tau_h$, $\tau_e\tau_h$, $\tau_h\tau_h$); and the scalar sum of hadronic central ($|\eta| < 2.4$) jets.

The DNN architecture consists of an input layer with 12 features, followed by three hidden layers with 128, 64, and 32 neurons, respectively. Each hidden layer employs batch normalization, ReLU activation, and a 20% dropout rate. L2 regularization (10^{-4}) is applied to the first layer. The output layer uses a sigmoid activation function. The network is optimized using the `Adam` optimizer [85] with an initial learning rate of 0.001 and a learning rate scheduler that reduces the rate after 5 epochs. Early stopping with a patience of 5 epochs is applied, and the model is trained using the binary cross-entropy loss function. A 2-fold cross-validation procedure is employed to evaluate the model performance. The training includes all background samples and HH signal samples corresponding to $\kappa_\lambda = 0.00$, 1.00, 2.45, and 5.00 across all data-taking periods. Due to limited statistics, all final state channels are merged into a single inclusive dataset. During training, the ParticleNet selection was relaxed to 0.1 to increase the available statistics, with an evaluation of the shapes of the input variables confirming that no bias was introduced in the input distributions by choosing a looser selection. Distributions of the input kinematic variables are shown in Fig. 3.20.

A weight, computed as the product of the cross section and the integrated luminosity, was assigned to

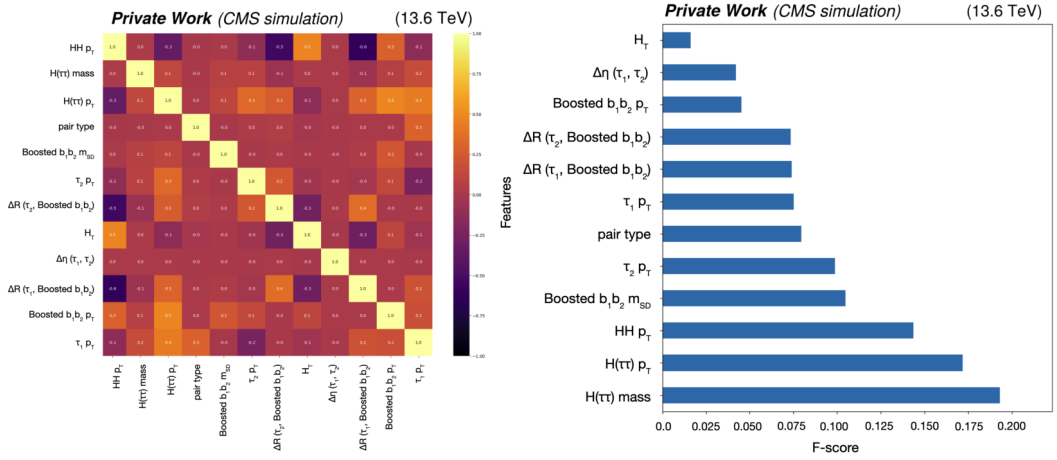

Figure 3.19: Left: The correlation matrix, showing the Pearson correlation coefficients between each pair of input variables, is displayed. The selected set of variables exhibits coefficients ranging approximately from -0.6 to 0.5, thus reducing redundancy among features. The coefficients with higher absolute values are physically reasonable, being some variables expected to be correlated for kinematic reasons. Right: The sensitivity of each input variable to the separation between signal and background is quantified using the F-score, represented on the x-axis for the variables listed on the y-axis. The three most relevant input variables are found to be the mass and p_T of the $H \to \tau\tau$ candidate, as well as the p_T of the di-Higgs system.

each event, and a re-sampling was performed to balance the signal and background statistics. The model outputs a score between 0 and 1, indicating the event's likelihood of being signal-like. The performance of the model is summarized in Figure 3.21. The described DNN is used to provide a discriminant distribution for boosted $H \to bb$ events, which is subsequently used as input to the statistical analysis.

3.2.6 Systematic uncertainties

Normalization and shape uncertainties follow the definitions in Section 3.1.6, with values updated for Run 3 experimental conditions. Since all these sources have already been described in Section 2.3.2.1 and discussed in the previous search, only a summary of the uncertainties included in the statistical analysis is reported here. Normalization uncertainties include the integrated luminosity uncertainty, with a value between 1.3% and 1.4%. Theoretical uncertainties affecting the inclusive HH production cross section include variations in the QCD renormalization and factorization scales, the choice of PDF set, and the value of the strong coupling α_s. For SM gluon-gluon fusion HH production, the combined scale and top-quark mass scheme uncertainty amounts to roughly $^{+6\%}_{-23\%}$, PDF uncertainties are $\pm 1.5\%$, and α_s variations are $\pm 1.7\%$ [31, 86]. For vector-boson fusion HH production, scale variations are $^{+0.05\%}_{-0.03\%}$ and PDF+α_s uncertainties are $\pm 2.7\%$ [32, 33]. The dependence of the inclusive HH production cross section and the event kinematic distributions on κ_λ is taken into account when interpreting the results for non-Standard-Model hypotheses. Branching fraction uncertainties are $^{+1.24\%}_{-1.26\%}$ for $H \to b\bar{b}$ and $^{+1.65\%}_{-1.63\%}$ for $H \to \tau^+\tau^-$. Theoretical uncertainties on the background normalizations, modeled solely from simulation, include the process-dependent PDF and α_s nuisances as reported in Ref. [87], with magnitudes ranging from 2% to 10%. All these uncertainties are implemented as nuisance parameters described by a log-normal distribution in the statistical analysis. Shape uncertainties account for corrections affecting event kinematics and object reconstruction and are included via templates obtained by varying quantities with ± 1 standard deviations. These include: energy scale and resolution of electrons, muons, and τs; reconstruction, identification, and isolation efficiencies of electrons, muons, and τs; efficiencies for τ_h discrimination against jets, electrons, and muons; pileup reweighting; jet energy scale and resolution; b-tagging efficiency; trigger efficiency; and fake factor uncertainties. More details on the origin of these corrections can be found in Section 2.3.2.1. In addition, uncertainties related to the limited number of simulated events are considered on a bin-by-bin basis.

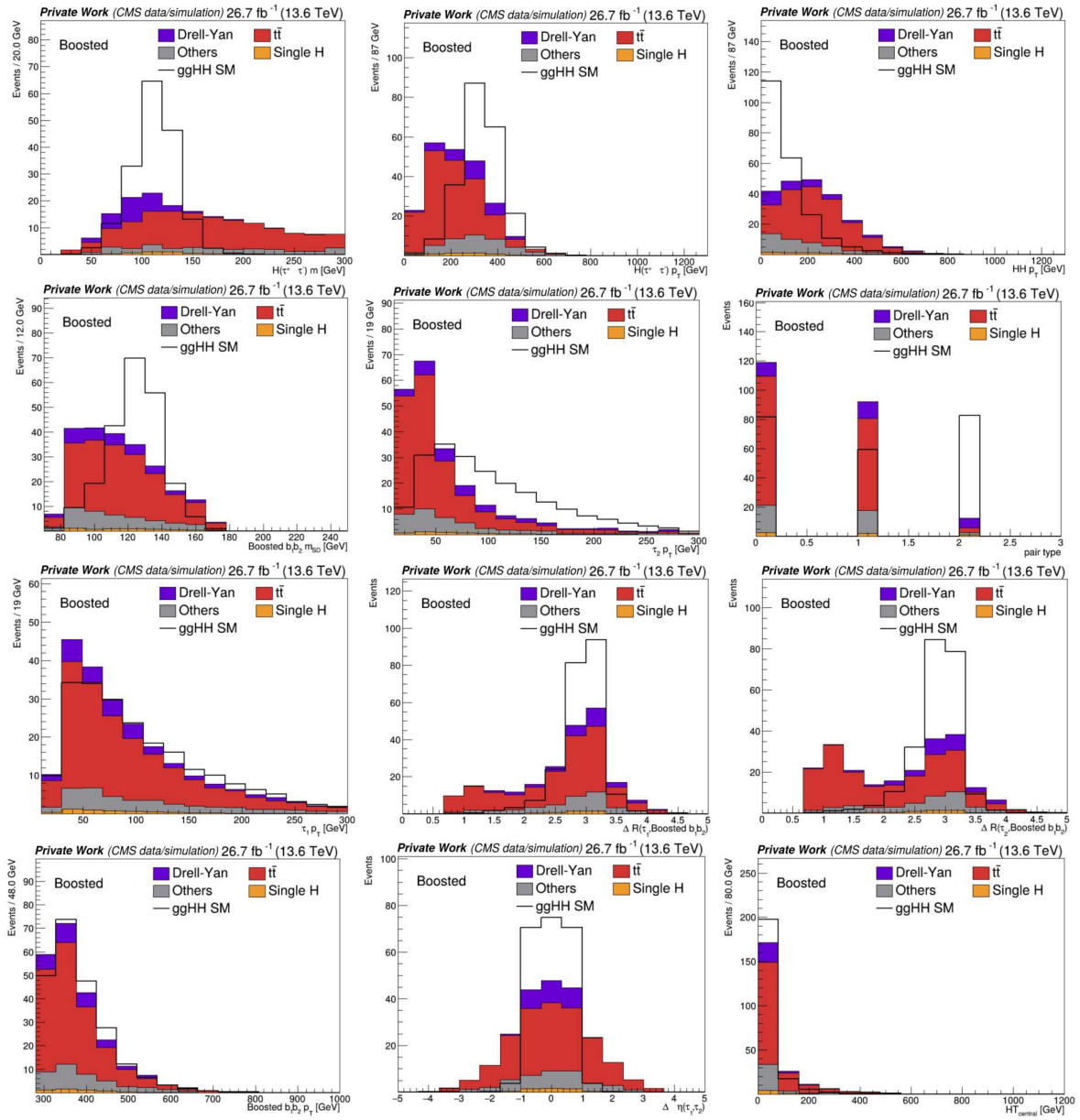

Figure 3.20: Background and signal distributions normalized to the background are shown for all input variables used by the DNN to discriminate signal from background. The distinction between the events is more prominent for the input features with high F-score, as expected. The represented dataset corresponds to an integrated luminosity of 26.7 fb^{-1}. The pair type refers to the final state, with 0 representing the $\tau_\mu\tau_h$, 1 the $\tau_e\tau_h$, and 2 the $\tau_h\tau_h$.

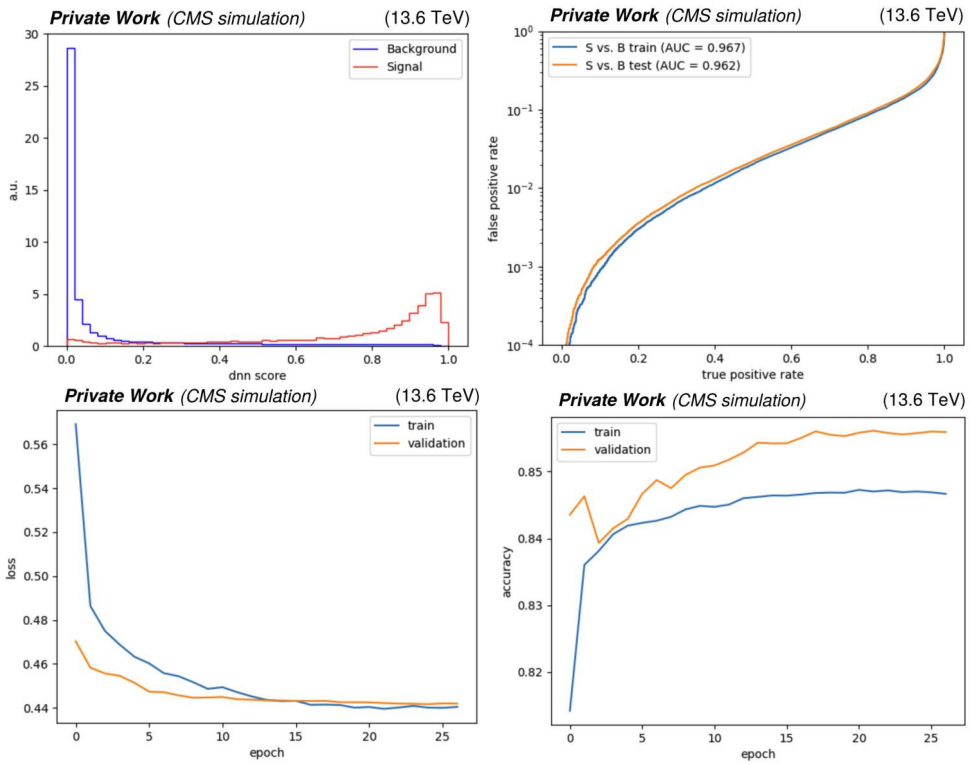

Figure 3.21: The DNN performance is summarized in terms of several key metrics in this figure. Top left: The DNN output score for signal (red) and background (blue) processes is shown, demonstrating good separation between the two cases. The signal distribution is obtained by concatenating all simulated signal samples used in the training (corresponding to $\kappa_\lambda = 0$, 1 and 2.45). The signal distribution peaks near 1, while the background peaks near 0, with small overlapping tails. Top right: The corresponding receiver operating characteristic (ROC) curve is displayed, demonstrating a good discriminating power with an area under the curve of about 0.96. Bottom left: The model loss, which quantifies how far the model predictions are from the true labels, decreases rapidly and stabilizes after approximately 15 epochs for both the training and validation samples. Bottom right: The model accuracy reaches its maximum value after roughly 15 epochs for both the training and validation samples.

3.2.7 Results

The DNN score distributions used as input to the statistical analysis for the resolved, boosted, and VBF categories are shown in Figure 3.22. The binning of the distributions is optimized following the procedure described in Section 3.2.5. Data in the bins with the highest DNN scores, corresponding to the regions with the largest signal sensitivity, are not shown since the analysis is performed in a blinded stage. For completeness, the expected event yields for the different background and signal processes are summarized in Table 3.1.

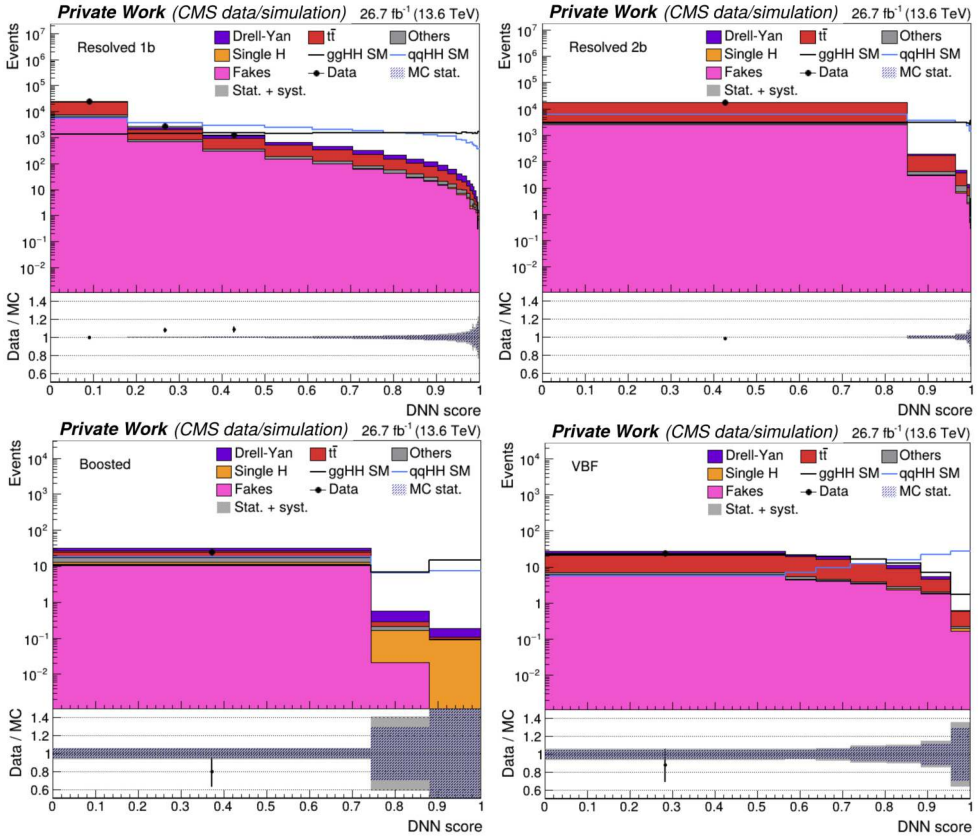

Figure 3.22: Distributions of the DNN score used in the statistical analysis for the different analysis categories. The distributions for the res1b and res2b categories are shown in the top row for the $\tau_\mu\tau_h$ channel, while the boosted and VBF categories are shown in the bottom row for the $X\tau_h$ combined channels, as used in the analysis. The distributions correspond to a data-taking period with an integrated luminosity of 26.7 fb^{-1}. The signal samples have been scaled for a better visualization (see Table 3.1 for the expected yields), background samples are stacked, while data are shown as points. The lower panels show the ratio of data to the total expected background. The inner band represents the Monte Carlo statistical uncertainty, while the outer band includes both statistical and normalization systematic uncertainties. The bins most sensitive to the signal have been blinded.

A binned maximum likelihood fit is performed on the DNN score distributions, with nuisance parameters corresponding to the systematic uncertainties (see Section 2.3.3 for more details on the statistical analysis). The analysis uses simulated events corresponding to a total of 62.4 fb^{-1}, combining four categories and four channels. The events are evaluated both inclusively, considering the gluon-gluon fusion and vector boson fusion production modes together, and separately for the individual processes.

In particular, to evaluate the signal strength of the individual production modes, the expected 95% CL upper limits are computed by treating either the ggF or the VBF signal strength as the parameter of interest. Results are shown both individually per category and combined for each production mode in Figure 3.23. An expected upper limit at 95% CL of 5.7 times the Standard Model expectation is obtained for ggF production, and 194 times the Standard Model prediction for VBF production. As expected, the

Category	Channel	DY	$t\bar{t}$	Fakes	Other bkgs	ggF SM	ggF k_λ=2.45	VBF SM	VBF c_{2V}=0
boosted	$X\tau_h$	4.9	9.0	11.6	6.4	0.15	0.10	0.0027	0.74
	$b\tau_h\tau_h$	1.69	2.6	1.98	1.57	0.016	0.014	0.00024	0.48
res1b	$\tau_e\tau_h$	3400	13000	5100	1300	0.39	0.16	0.012	0.27
	$\tau_\mu\tau_h$	2400	18000	7600	1900	0.61	0.25	0.019	0.41
	$\tau_h\tau_h$	850	600	1100	120	0.50	0.21	0.012	0.37
	$b\tau_h\tau_h$	67	68	760	22	0.0093	0.0072	0.00026	0.10
res2b	$\tau_e\tau_h$	340	10000	1800	420	0.37	0.15	0.0071	0.21
	$\tau_\mu\tau_h$	250	15000	2600	590	0.57	0.23	0.012	0.31
	$\tau_h\tau_h$	100	460	150	31	0.49	0.20	0.0074	0.30
	$b\tau_h\tau_h$	6.4	12	75	2.0	0.0051	0.0049	0.00012	0.055
VBF	$X\tau_h$	14	59	22	4.0	0.028	0.013	0.026	0.72
	$b\tau_h\tau_h$	0.62	0.88	2.8	0.095	0.00076	0.00056	0.00013	0.081

Table 3.1: Expected event yields per category, channel, and process are obtained by integrating over the full DNN score spectrum. The yields correspond to a data-taking period with an integrated luminosity of 26.7 fb^{-1}and are rounded to a few significant digits. Signal yields are shown for the SM hypothesis and for selected benchmark scenarios.

VBF category is the most sensitive phase-space for VBF production compared to the other categories, providing the most stringent upper limit in that production mode.

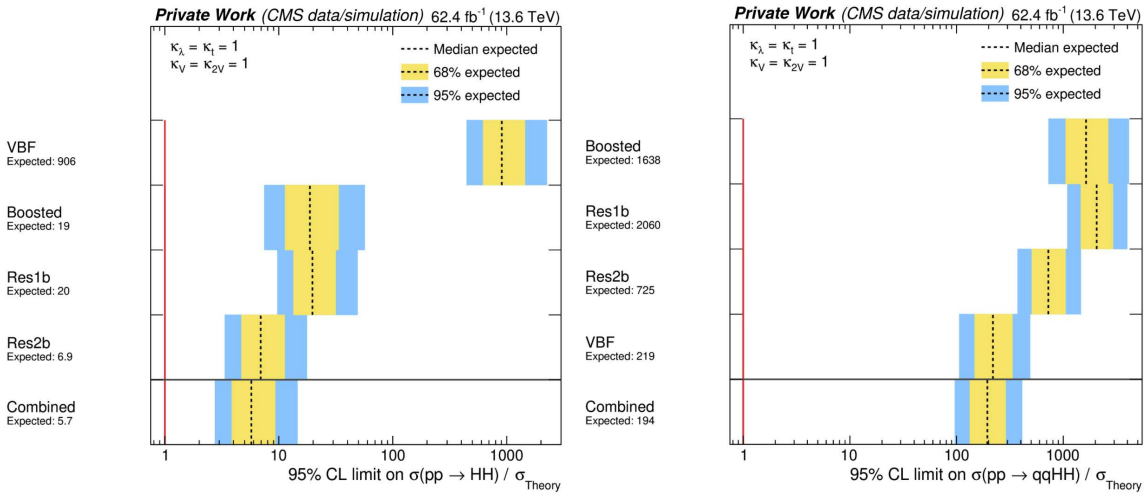

Figure 3.23: Expected upper limits at 95% confidence level on the HH cross section via gluon fusion (left) and vector boson fusion (right) production. The colored bands represent in yellow and blue the 1 and 2 standard deviations, respectively, on the expected limit.

The inclusive production can also be considered, in which the parameter of interest is the total signal strength, comprising both ggF and VBF. The expected upper limits at 95% CL on inclusive Higgs boson pair production are shown in Figure 3.24, both individually per category and per channel, as well as combined. The resulting expected upper limit at 95% CL for the combination of categories and channels on the inclusive cross section is 5.7 times the Standard Model expectation, dominated by the ggF sensitivity. The most sensitive category is res2b, followed by the boosted, res1b, and VBF categories. Despite the overall limited sensitivity in VBF phase space, including a VBF category is particularly relevant for probing the Higgs–vector boson coupling modifier κ_{2V}. The coupling modifiers are studied by evaluating the expected limits at 95% CL on the cross section as a function of different values of the Higgs boson self-coupling modifier (κ_λ), and of the coupling modifier for the HHVV vertex (κ_{2V}). The results are shown in Figure 3.25. This analysis sets expected constraints on κ_λ in the range $-3.75 < \kappa_\lambda < 10.5$, while κ_{2V} is expected to be constrained within $-0.75 < \kappa_{2V} < 2.95$. These results can be compared to those obtained by the same analysis performed on the Run 2 dataset [79], which used 138 fb^{-1} of data. In

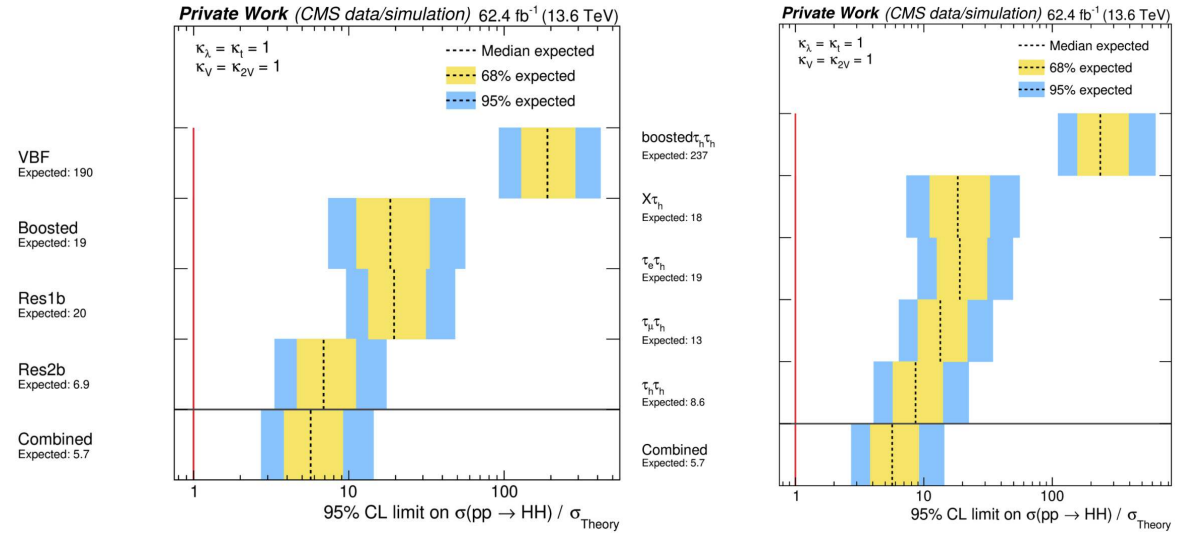

Figure 3.24: Expected upper limits at 95% confidence level on the inclusive HH production cross section are shown both for individual categories (left) and channels (right). The colored bands represent in yellow and blue the 1 and 2 standard deviations, respectively, on the expected limit.

that analysis, κ_λ was constrained to the range $[-2.9, 9.8]$ and κ_{2V} to $[-0.6, 2.8]$, despite having roughly twice the integrated luminosity.

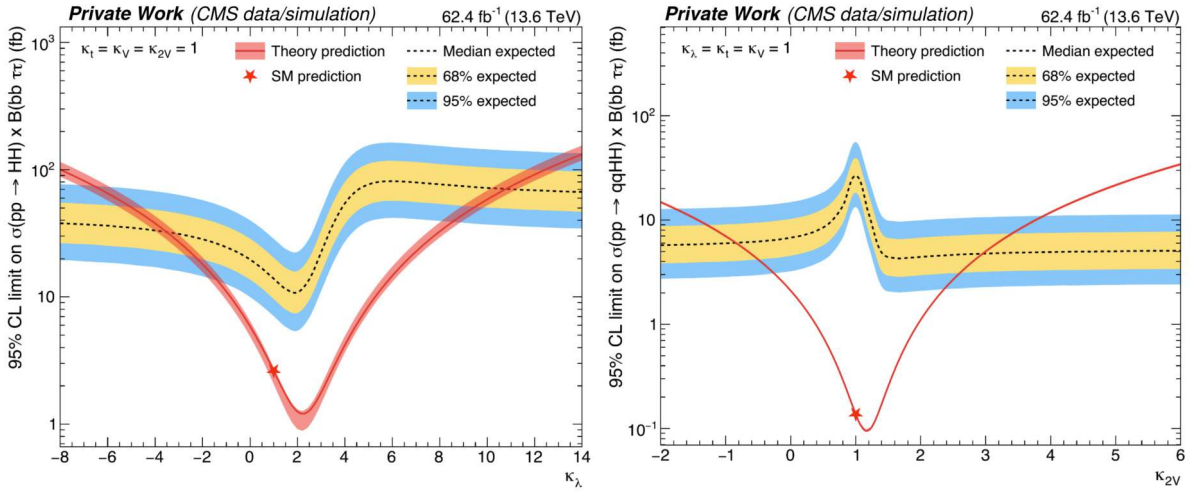

Figure 3.25: Expected upper limits at 95% confidence level on the HH production cross section times the $HH \to bb\tau\tau$ branching ratio as a function of κ_λ (left) and κ_{2V} (right). The red solid line shows the theoretical prediction for the HH production cross section, with its uncertainty represented by a band.

For the cross-section comparison, results from Run 2 and Run 3 searches are considered excluding the contribution from the VBF category, as its definition differs significantly in the two analyses. When comparing the combined results of Run 2 and Run 3, despite the Run 2 dataset having roughly twice the statistics of Run 3, the expected limits results to be similar. After rescaling the expected limits obtained with Run 3 data to the Run 2 integrated luminosity, an improvement of approximately 27% is observed. This indicates a significant gain beyond the pure luminosity scaling, and this is observed across all categories, as shown in Figure 3.26. This enhancement can be partially attributed to better trigger performance (the trigger requiring a pair of τ_h leptons and a jet alone provides an 8% improvement in the expected limits over the full phase space), the use of the fake-factor method for estimating misidentified τ_h leptons, and a dedicated strategy for handling boosted topologies. In particular, for the boosted

category, this improvement is estimated to be approximately 15% compared to previous results.

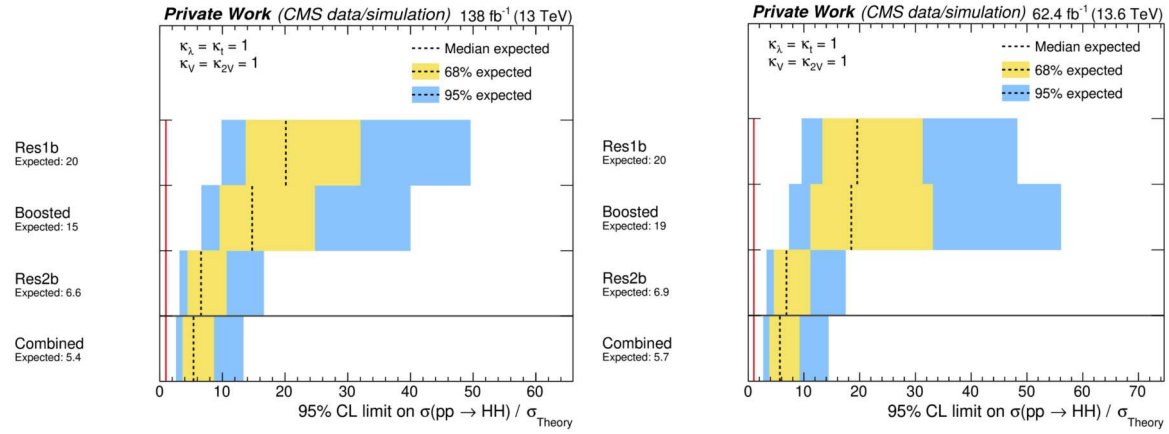

Figure 3.26: Expected upper limits at 95% CL are shown individually for the res1b, res2b, and boosted categories, comparing the results from Run 2 [79] (left) and this analysis with Run 3 data (right).

The expected results already highlight the significant sensitivity enhancement achieved through new triggers, reconstruction algorithms, and signal extraction methods. These results, based on data collected during the first two years of LHC Run 3 (62.4 fb^{-1}), demonstrate a strong capability to further constrain the Higgs boson pair production cross section in the $bb\tau\tau$ final state and the value of the Higgs self-coupling modifier. Once the full Run 3 dataset corresponding to approximately 300 fb^{-1} becomes available, the increased statistics is expected to further improve the constraints reaching about 2.6 time the SM cross section. All these developments and the integration of new techniques, largely driven by more sophisticated algorithms, provide a valuable foundation for refining analysis methods in preparation for the HL-LHC. The larger datasets expected at the HL-LHC will significantly enhance the sensitivity to the Higgs boson self-coupling, bringing us closer to a deeper understanding of the Higgs potential and the mechanism of electroweak symmetry breaking.

Chapter 4

Precision timing in CMS for the high luminosity phase of LHC

Designed to be a discovery machine, the LHC has also proven to be a powerful precision instrument. The large amount of data delivered, together with increasingly advanced techniques in event processing and reconstruction algorithms, has accelerated the progress of analyses, leading to ever more precise measurements. The Higgs boson properties have been measured with high precision during Run 2, with current uncertainties below 10% for most couplings. Constraints on the Higgs self-coupling have also been significantly improved, reaching observed upper limits of a few times the Standard Model prediction. New analyses, currently being finalized with Run 3 data, are expected to deliver even better performance than the naive statistical scaling, thanks to dedicated triggers and advanced tagging techniques. However, achieving the precision needed to probe very rare processes and further improve measurement accuracy requires a new phase of operation at higher luminosity.

The HL-LHC, scheduled to begin operation in 2030, will deliver a substantial increase in luminosity compared to previous runs, with proton-proton collisions at 14 TeV corresponding to integrated luminosities of 3000 fb^{-1} for CMS and ATLAS, 50 fb^{-1} for LHCb, and PbPb collisions with integrated luminosities of 13 nb^{-1} and 50 nb^{-1} for ALICE. Building on the achievements of the LHC physics program, this new phase is expected to deepen our understanding across a broad range of topics. The higher statistics will reduce statistical uncertainties, while new strategies for constraining systematic effects will enable unprecedented levels of precision in many analyses. Despite these benefits, operating under such extreme experimental conditions poses major challenges, requiring accelerators and detectors capable of withstanding high radiation levels and extreme particle rates.

The CMS experiment is currently undergoing a major upgrade program to maintain its excellent performance under high radiation levels and extreme pileup conditions, i.e. a large number of simultaneous proton-proton interactions, which are expected to reach up to 200 in the ultimate HL-LHC scenario. To mitigate the impact of high pileup on event reconstruction, CMS will integrate a new timing detector capable of measuring the arrival time of charged particles with a precision of about 30 ps at the start of HL-LHC operation, degrading to approximately 60 ps in the barrel region by the end of the run due to radiation damage effects. Incorporating timing measurements into the CMS event reconstruction workflow will significantly improve object reconstruction and enable new analysis capabilities. The CMS MIP Timing Detector (MTD) will provide coverage for both the barrel (Barrel Timing Layer, BTL) and the endcap (Endcap Timing Layer, ETL) regions, using different technologies optimized for the respective environments and performance requirements. Both systems have undergone extensive R&D studies and optimization campaigns to achieve the target performance, as well as the development of protocols for large-scale production. Mass production is currently ongoing for the BTL and is about to begin for the ETL. The BTL installation is expected to be completed by the end of 2026, while the ETL is planned to be installed by mid-2029. The addition of precise timing information will provide CMS with the capability to measure the time of flight of charged particles and will enhance numerous physics analyses, such as di-Higgs searches, through improved pileup rejection in vertexing and track association, as well as better object reconstruction.

This chapter provides a brief overview of the high-luminosity phase of the LHC in Section 4.1, including the main accelerator upgrades required to achieve the target performance. The CMS detector upgrades designed to preserve the current excellent experimental performance are summarized in Section 4.2, with particular focus on the MTD. Section 4.3 is dedicated to a description of the MTD, a key component of this work, including both the BTL and ETL subsystems and their shared infrastructures. Finally, the impact of precise timing information on event reconstruction and a few representative physics analyses is discussed in Section 4.4.

4.1 The High-Luminosity phase of the LHC

With the completion of Run 3, gaining statistical precision by simply continuing accelerator operation without major upgrades would be inefficient, as it would take more than ten years to halve the statistical error. In this context, CERN proposed the high-luminosity LHC project [88], aiming for a peak luminosity of 5×10^{34} cm^{-2}s^{-1} with levelling[1], and an integrated luminosity of approximately 250 fb^{-1} per year over 12 years. The "ultimate performance" is expected to produce up to 200 pileup events per bunch crossing, reaching a peak luminosity of 7.5×10^{34} cm^{-2}s^{-1} and integrating approximately 350 fb^{-1} per year. Luminosity, expressed in Equation 2.1, is a key parameter for assessing accelerator performance, as it provides a quantitative estimate of the number of collisions in a given time. Higher luminosity directly corresponds to more data collected and, therefore, more statistics for exploring rare processes. For instance, at the HL-LHC at least 15 million Higgs bosons will be produced per year, compared to around 3 million collected during LHC operation in 2017. Intense upgrades are required to achieve such an ambitious goal and the current plan is shown in Figure 4.1, anticipating HL-LHC operation by mid-2030.

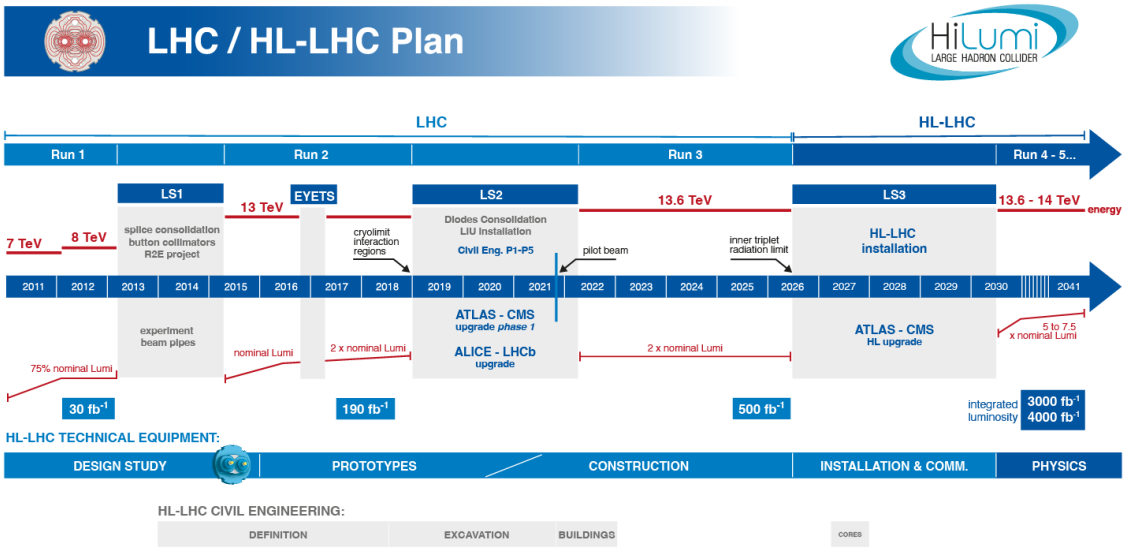

Figure 4.1: The current schedule of LHC and HL-LHC operations at the time of writing is shown. The top line indicates the collision energy, while the bottom line shows the luminosity. At present, we are approaching the completion of Run 3 data taking, with the Long Shutdown 3 expected to begin next year. Afterward, the accelerators will operate in the high-luminosity configuration, with the start of HL-LHC operations foreseen for 2030 [89].

To achieve a tenfold increase in luminosity, critical innovative technologies have been proposed for the new machine, upgrading over 1.2 km of the LHC accelerator. These upgrades aim to increase the

[1]The accelerator configuration is varied during operation to maintain constant luminosity.

number of protons per bunch while reducing the transverse beam size to enhance luminosity. This will be achieved through new optics and technologies, which are not only essential for delivering high collision rates in high-energy physics experiments, but also serve as technological drivers for future accelerators and colliders. To achieve the target luminosity, several parameters affecting luminosity have been carefully optimized, including:

- *Beam brightness:* Defined as the ratio of bunch intensity to transverse emittance. It must be maximized during beam generation and preserved along the injector chain to ensure stable operation. The goal is to double the number of protons per bunch without increasing emittance.
- β^*: β^* defines the beam size at the collision point. Smaller β^* yields higher peak luminosity, requiring more powerful quadrupole magnets and advanced dipoles. The target is 10–15 cm, to be compared with the current value of 55 cm of LHC. The HL-LHC will employ quadrupole magnets with a 150 mm aperture, compared to the current 70 mm in the LHC, in order to accommodate the upgraded beam and to shield the magnet coils from radiation. Such magnets will generate magnetic fields up to almost 12 T, compared with 8 T at the LHC. These high fields will be achieved using superconducting Nb_3Sn technology.
- *Crossing angle:* A small β^* implies a large crossing angle, which reduces R (Equation 2.1). Advanced superconducting RF crab cavities are designed to rotate the bunches, enabling nearly head-on collisions and minimizing luminosity loss. Crab cavities will be produced from bulk niobium, operate at a frequency of 400 MHz, and have dimensions about five times smaller than traditional elliptical cavities, featuring low noise and precise phase control.
- *Beam current:* Many LHC systems (RF, cryogenics, vacuum, monitoring, etc.) depend on the beam current. The current limit is 0.86 A, but HL-LHC aims to increase this by 30%, with bunches every 25 ns. High-current power lines are required for the magnet-related equipment and must carry 30–100 kA of direct current with minimal voltage drop, via new superconducting links made of MgB_2, operational up to 25 K.

The new technologies expected for this phase include compact superconducting cavities for beam rotation with precise phase control, novel beam collimation techniques, superconducting magnets expected to reach 11–12 T, and high-power superconducting links with minimal energy dissipation. A schematics of the upgrades anticipated for HL-LHC accelerator and colliding systems is reported in Figure 4.2.

Thanks to the potential offered by these innovations, direct and indirect searches will be pursued at the HL-LHC through studies of higher-order effective operators. Rare processes such as di-Higgs production, rare τ-lepton decays, and Higgs boson couplings to second-generation fermions will be central elements of the physics program. LHC results have demonstrated that the SM provides an excellent description of most observed particle physics phenomena. Nevertheless, there remain several processes that the SM cannot explain: it does not account for dark matter, it does not provide a unification with gravitation, motivating the search for new physics at the TeV scale. Deviations from SM expectations could point to new physics, motivating a continuous effort toward increasingly precise measurements. Currently, no clear hint of BSM physics has been observed, but the HL-LHC will offer new opportunities enabled by a significantly larger dataset. The HL-LHC physics program includes precision measurements within the SM, detailed studies of Higgs boson properties, searches for BSM processes, investigations of heavy-quark and lepton flavor physics, and studies of QCD matter at high density and temperature. A comprehensive description of this program can be found in Ref. [42], while a brief summary is provided below to illustrate the potential offered by this upgrade.

Unresolved fundamental questions, such as naturalness, the nature of dark matter, neutrino masses, and baryogenesis, require new BSM scenarios, which are expected to be more effectively probed with the increased statistics provided by the HL-LHC. New physics can also be investigated through flavor observables in several sectors, such as Higgs couplings to SM fermions, deviations in Yukawa couplings, searches for flavor-violating interactions, or flavor-changing transitions involving top, bottom, charm, and strange quarks or leptons. Moreover, QCD studies with ion and proton beams at the HL-LHC are expected to provide higher precision in the characterization of the macroscopic, long-wavelength properties of the Quark-Gluon Plasma (QGP). These studies aim to build a coherent picture of particle production and QCD dynamics, spanning from proton-proton to nucleus-nucleus systems, and to extend the study

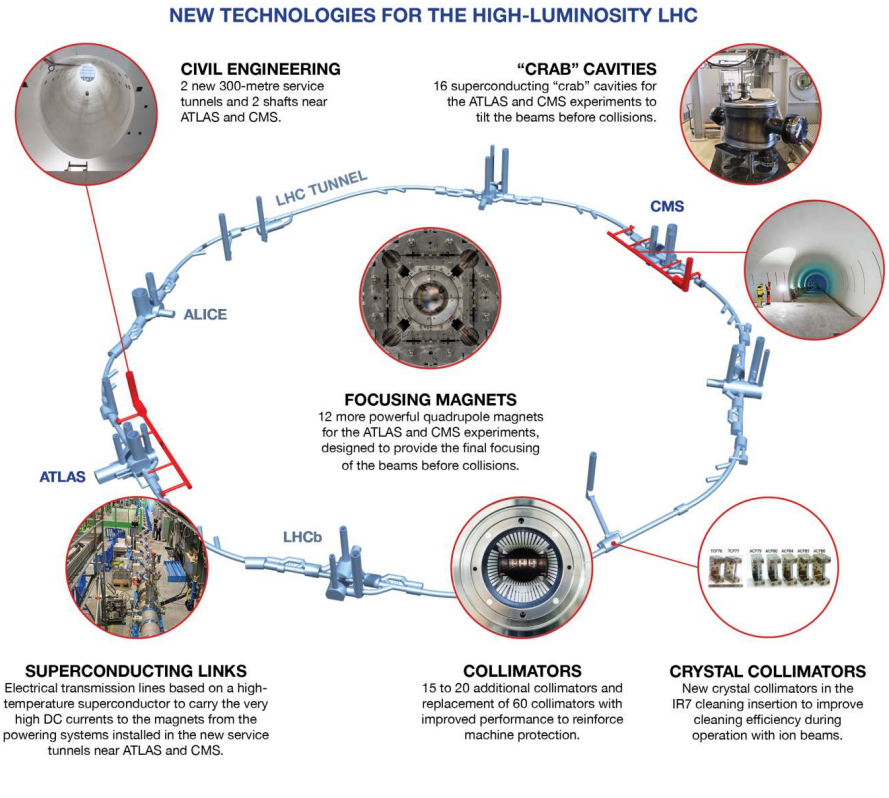

Figure 4.2: Sketch representing the main new technologies expected at the HL-LHC to optimize the collider's parameters. Some of these technologies are entirely new and represent innovative additions being developed specifically for the HL-LHC. A brief description of each element is shown in the figure and in the text [89].

of parton densities in nucleons over a wider kinematic range.

The tenfold increase in the dataset will allow for significant improvements in the precision of SM measurements. Through global fits of these measurements, it will be possible to further constrain PDF uncertainties, thereby improving the precision of many SM parameters, such as the weak mixing angle and the W boson mass. A coherent understanding of the origin of electroweak symmetry breaking remains one of the primary objectives of the HL-LHC physics program, as discussed in Chapter 1. The HL-LHC opens the possibility of addressing many unresolved questions. A precision at the percent level is expected for measurements of Higgs boson couplings, with the high statistics particularly beneficial for studying complex final states, such as those arising from $t\bar{t}H$ production. BSM physics can be probed indirectly through precision measurements, including searches for exotic Higgs boson decays (e.g. to light scalars, dark photons, axion-like particles, or long-lived particles), or through the production of additional charged or neutral Higgs bosons with masses above or below 125 GeV. The HL-LHC will enable the exploration of very rare exotic decay modes of the 125 GeV Higgs boson thanks to the large dataset, with branching ratios as small as $\mathcal{O}(10^{-6})$ being potentially accessible for sufficiently clean final states. Furthermore, the mass reach for new heavy Higgs bosons will extend up to a few TeV. The HL-LHC will be capable of probing new Higgs states as heavy as about 2.5 TeV, while precision coupling measurements will allow the exclusion of additional Higgs bosons with masses below approximately 1 TeV over a wide range of $\tan\beta$ values.

4.2 The CMS upgrades

The ambitious HL-LHC physics program requires each experiment to maintain the reconstruction capabilities demonstrated during LHC Run 1 and Run 2. Therefore, an extensive upgrade program in detector technologies, triggering, and software has been envisaged for all experiments. In particular, the CMS Phase II upgrade is described in detail in Ref. [90], and an overview of the main upgrades is provided here.

Simulation studies have guided the identification of the necessary changes to cope with beam conditions and radiation damage. The pileup levels expected at the HL-LHC demand improved tracking efficiency to correctly identify all primary pp collision vertices and maintain high-quality event reconstruction within the particle-flow framework. Radiation damage has been evaluated through test-beam studies to benchmark simulations of the doses and degradation expected at the HL-LHC. These studies have led to the decision to completely replace the tracker system (which is expected to experience up to three billion hits per second per square centimeter) and the endcap calorimeters for Phase II, to address the high pileup expected in the inner and forward detector regions. In addition, to mitigate the extreme pileup conditions, a new timing layer will be integrated between the tracker and the calorimeter, providing a precise measurement of the time of arrival of charged particles, i.e. the MIP Timing Detector. A luminosity detector will also be installed, and all existing electronics will be replaced.

The high pileup will cause additional hits within the tracking system, increasing the complexity of track reconstruction and the fake track rate, as well as contaminating energy clusters from particles originating from the hard scatter interaction point. Since both hits and clusters are affected, the overall result is a significant degradation of object reconstruction performance and efficiency. It also impacts trigger efficiency and increases the amount of data to be read out in each bunch crossing, consequently slowing down event reconstruction at the HLT and offline analysis stages. All these factors call for an increased detector granularity, which reduces the occupancy per readout cell at the cost of a larger bandwidth that needs enhanced trigger and DAQ capabilities. A new approach will be deployed in the trigger system to achieve low-energy trigger thresholds without loss of efficiency, by introducing tracking information at the hardware trigger level. Full particle-flow reconstruction will already be performed at the hardware level from the Level-1 trigger at a rate of 40 MHz, while timing information will be integrated in the HLT, running optimized algorithms on GPUs. Furthermore, to maximize acceptance in the forward regions, the muon system will be improved by adding new chambers, extending the muon coverage up to a pseudorapidity of approximately 3 [91].

The entire tracker system will be replaced, with an expected increase in granularity by factors of four

and six in the outer tracker and pixel detector, respectively, compared to the current design. The tracker design for HL-LHC is represented in the left panel of Figure 4.3. The silicon strips in the outer tracker will be shortened while maintaining the same pitch. Six barrel layers and five endcap disks are foreseen, with an increased forward acceptance up to a pseudorapidity of 4 and improved radiation hardness, sustaining fluences up to 2.3×10^{16} n_{eq}/cm^2 in the first pixel layer, as shown in the panel on the right in Figure 4.3. The inner tracker will consist of silicon pixels with a new 65 nm ASIC technology enabling 25×100 μm^2 pixels, for a total of 3900 modules and 4.9 m^2 pixel area. The sensors will be 3D pixels in the barrel and planar silicon sensors in the forward and endcap regions, with a 150 μm active thickness. The mechanical structure has been designed to facilitate installation and removal, allowing for replacement of damaged parts. To implement track information at the hardware trigger level, specific on-detector electronics measure the bending of tracks and select only hits from tracks with a transverse momentum greater than 2 GeV to be sent to the trigger. In the backend, tracks are fully reconstructed and fitted, and then sent to the central trigger, allowing for efficient background rejection.

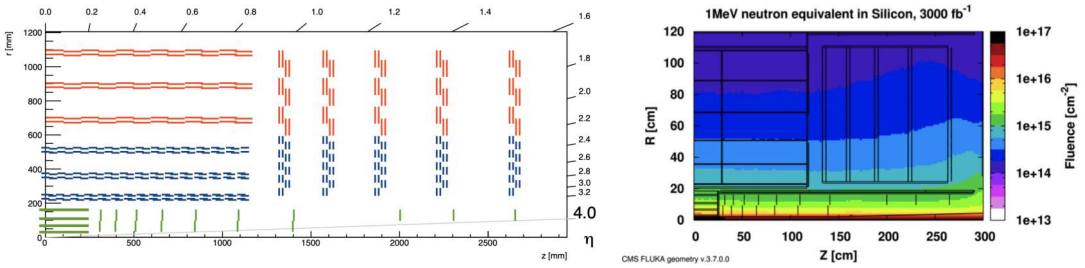

Figure 4.3: Left: A schematic view of the tracker design is shown, where the outer tracker components are illustrated in orange and blue, while the inner tracker is shown in green. Right: The expected particle fluence, expressed in terms of 1 MeV neutron equivalent in silicon, is shown as a map in the tracker volume for an integrated luminosity of 3000 fb^{-1} [91].

The lead tungstate crystals composing the barrel electromagnetic calorimeter are expected to remain efficient for the entire HL-LHC operation and will not be replaced. The only upgrade foreseen is to the front-end electronics, required to satisfy the new trigger requirements. The newly designed front end will perform 160 MHz sampling and signal readout, exploiting the full ECAL granularity and enabling reconstruction with more advanced algorithms. The APDs will operate at lower temperatures, from 18°C to 8°C, to mitigate aging-induced noise that would otherwise dominate the energy resolution. The front-end chip will also provide shorter signal shaping to further reduce APD noise.

In the endcap region, a completely new calorimeter concept will be installed: the High Granularity Calorimeter (HGCAL). It combines both electromagnetic and hadronic sections, ensuring excellent transverse and longitudinal segmentation for shower measurements, maximizing granularity, and fully exploiting particle-flow reconstruction. It is a radiation-tolerant calorimeter, with expected fluences above 10^{16} n_{eq}/cm^2 in the electromagnetic section and above 10^{14} n_{eq}/cm^2 in the hadronic section. It is capable of providing a readout with spatial, energy, and timing information. HGCAL is expected to achieve 30 ps in time resolution for clusters of particles with $p_T > 5$ GeV, to mitigate the pileup effect. The electromagnetic section consists of 26 active layers of silicon sensors interleaved with copper, copper–tungsten, and lead absorbers. The hadronic section is made of 21 active layers of silicon and scintillator tiles, interleaved with steel absorbers. A schematic view of the section of HGCAL is shown in Figure 4.4. The total silicon sensor area amounts to about 600 m^2, complemented by 400 m^2 of SiPM-on-tile modules, with a total of 6 million channels. The silicon sensors are hexagonal, forming a honeycomb-like geometry, with a cell size of 0.5 cm^2 in the innermost region and 1.2 cm^2 in the outer region. This design choice optimizes the use of circular silicon wafers, thereby minimizing costs. The scintillator tiles are used in the hadronic section at lower radiation compared to the regions of the silicon sensors. Both silicon sensors and SiPMs reading the light produced by scintillators will be operated at -35°C to minimize the leakage current due to radiation damage.

While many upgrades are currently under production, some have already been integrated into CMS, such as the new Gas Electron Multiplier (GEM) detectors, which have been used for muon detection

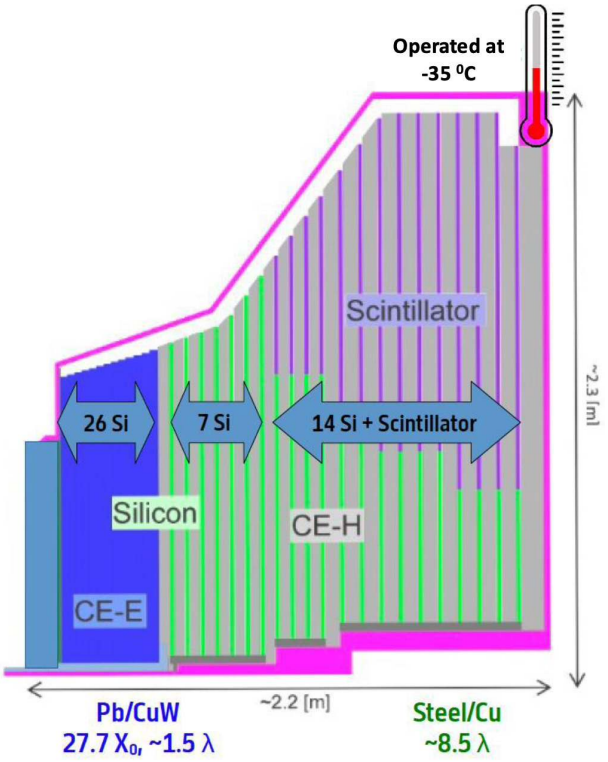

Figure 4.4: Pictorial view of a section of the upper half of one endcap calorimeter, showing all key components and parameters [92].

since LS2, with 144 chambers already fully installed and a second station expected before LS3. These additional chambers provide higher rate capability and good spatial resolution. Moreover, in the forward region, two additional stations will be added, implementing low-resistivity RPCs with lower granularity but good time resolution. All front-end electronics of the CSC have already been replaced, while the replacement of the DT electronics is planned for LS3. There are no plans to replace the sensors of the DT, RPC, or CSC systems, which are expected to cope effectively with the radiation levels anticipated at the HL-LHC. A pictorial representation of the upgraded muon system is shown in Figure 4.5.

4.3 The addition of the fourth dimension in CMS event reconstruction at the HL-LHC

In addition to the upgrade expected for the sub-detectors of CMS, a completely new concept of detector measuring the time of arrival of charged particles will be integrated into CMS to mitigate the extreme pileup conditions envisaged, while sustaining the harsh radiation environment. The MIP Timing Detector is designed to provide precise timing information for minimum ionizing particles (MIPs). This information is crucial for distinguishing among the up to 200 pileup interactions expected at the HL-LHC. The MTD will offer timing data across a pseudorapidity range of up to ± 3, with a time resolution of 30 ps at the start of HL-LHC operation. However, performance is expected to degrade to about 60 ps in the barrel by the end of its operations due to radiation damage to the sensors [93].

4.3.1 Motivation for an additional dimension in event reconstruction

The concept behind this detector relies on the fact that the beams have a longitudinal extent along the z-axis, with collisions distributed according to an RMS of approximately 4.5 cm, and do not occur simultaneously but exhibit a time spread characterized by an RMS of 180–200 ps, due to the crossing time of the particle bunches, as illustrated in Figure 4.6. Consequently, associating a precise time information

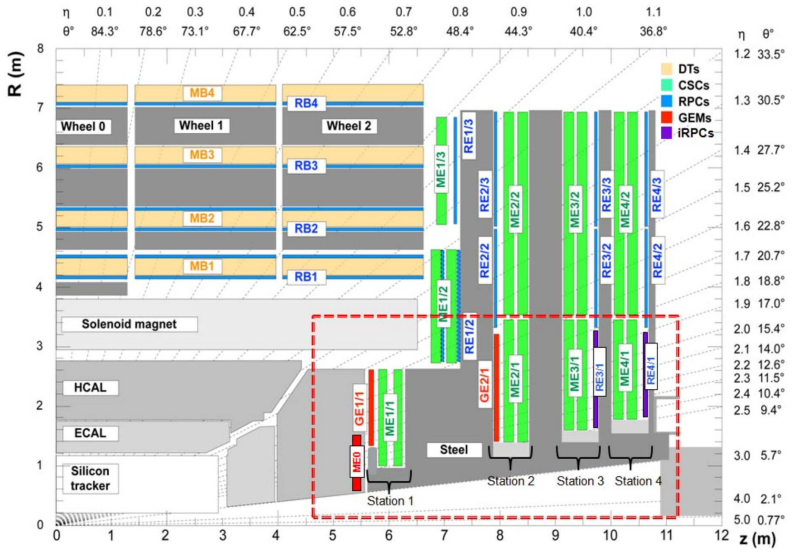

Figure 4.5: Schematic view of the muon system, with DT, RPC, and CSC detectors shown in yellow, blue, and green, respectively. The new forward detectors for the Phase-II upgrades are highlighted with a dashed red box: GEM stations are shown in red, and RPC stations in blue [90].

to charged particles is expected to disentangle vertices that would otherwise be overlapped in space, enabling better reconstruction of the primary vertex. The time of a collision vertex can be reconstructed by associating tracks from the vertex with hits and their MTD times. More details on vertex and track reconstruction using MTD information are provided in Section 4.4.1.

When moving from a nominal scenario of 140 collisions per bunch crossing (with a stable leveled luminosity of $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$) to an ultimate scenario of 200 collisions per bunch crossing (at a leveled luminosity of $7.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$), the vertex density ($\text{dN}_\text{V}/\text{dz}$) increases from 1.2 to 1.9 vertices per millimeter, as shown in the left panel of Figure 4.7, making it very challenging to distinguish different vertices. For example, to associate a track with the primary vertex, specific requirements must be placed on the distance of closest approach to the vertex along the beam axis. However, tracks from displaced sources (such as secondary interactions, particle decays in flight) require a wider selection window, approximately on the order of 1 mm. This broad selection leads to a high level of contamination from pileup tracks into the primary vertex, resulting in poorer resolution, efficiency, and identification purity. The incorporation of timing information for each track will significantly reduce the number of pileup tracks incorrectly associated with the hard-interaction vertex. This reduction can be quantitatively seen in the right panel of Figure 4.7, which shows the number of pileup tracks incorrectly associated with the primary vertex as a function of line density under different operating conditions. It is evident that adding timing information with 30–40 ps precision reduces incorrect associations by more than a factor of two, demonstrating the efficacy of four-dimensional vertex reconstruction under conditions with 200 pileup interactions.

The reduction of incorrect associations between primary vertices and tracks will enhance the reconstruction purity of various physics objects. For instance, leptons and photons will exhibit higher identification efficiency. Additionally, the identification of b-jets, which relies on secondary vertex reconstruction, will also improve, along with objects that require the reconstruction of global event quantities, such as jets and missing transverse momentum. Timing will provide CMS with new capabilities, allowing for the identification of charged hadrons as pions, kaons, and protons based on their time of flight. Furthermore, the ability to reconstruct the timing of displaced vertices will enable searches for long-lived particles (LLPs) by measuring their relativistic velocity parameters. A discussion on studies on the impact of timing on physics analyses is reported in Section 4.4.

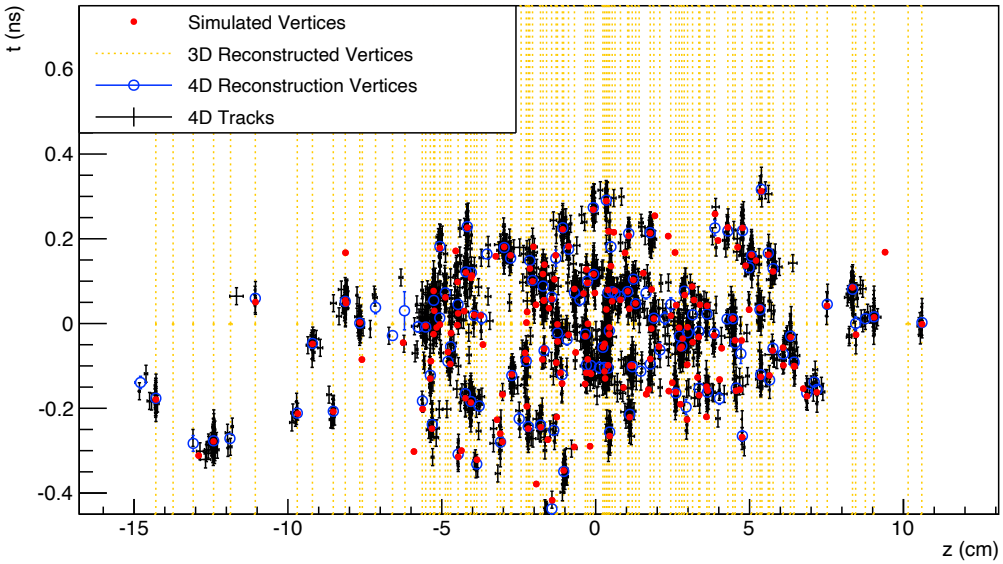

Figure 4.6: Vertices from a simulated bunch crossing, reconstructed using different approaches, are shown in the time–z plane under conditions of 200 pileup interactions. A time resolution of 30 ps for MIP measurements is assumed in both the barrel and endcap regions. The interaction region is located at $z = 0$ cm, where z is the coordinate along the beam direction (see Figure 2.2). The origin of the time axis corresponds to the moment when the beams overlap in z. Simulated vertices are represented by red dots; vertices reconstructed using 3D spatial information only are shown in yellow; tracks and vertices reconstructed using 4D information (including time) are shown in black and blue, respectively. It is clearly visible that the addition of precise timing information allows one to disentangle vertices that would otherwise overlap spatially [93].

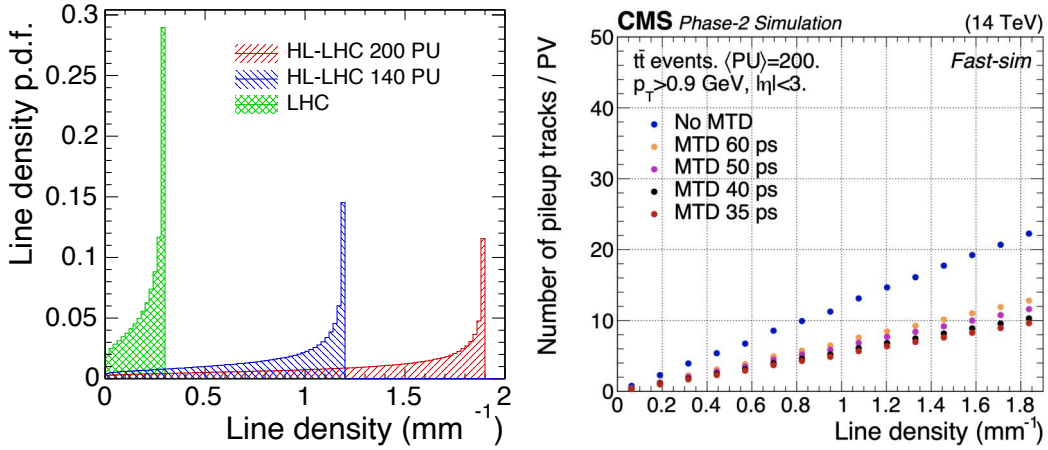

Figure 4.7: Left: The probability density functions of the line density along the beam axis for different pileup scenarios are shown. LHC conditions with about 30 pileup interactions are represented in green, featuring a mode of 0.3 mm^{-1} and a mean of 0.2 mm^{-1}. HL-LHC conditions with 140 pileup interactions are shown in blue, with a mode of 1.2 mm^{-1} and a mean of 0.9 mm^{-1}. The ultimate HL-LHC scenario with 200 pileup interactions is shown in red, with a mode of 1.9 mm^{-1} and a mean of 1.4 mm^{-1}. Right: The number of pileup tracks incorrectly associated with the primary vertex (PV) is shown as a function of the line density for different time resolution scenarios, illustrating the significant reduction achieved by the MTD in mitigating pileup (from the blue points to the others) [93].

4.3.2 The MTD design

The MTD will be integrated into the existing layers of CMS and must operate efficiently under the extreme conditions expected at the HL-LHC. It is required to achieve timing precision of 30–60 ps to effectively mitigate pileup during Phase-2 operations. These requirements impose stringent mechanical constraints and demand for tolerance to magnetic fields and high radiation exposure. The physics studies on the impact of timing on analyses, which is discussed later in Section 4.4, support the need for nearly hermetic coverage of the pseudorapidity region. The only available and functional allocation is between the outer tracker and the calorimeters. The barrel region will be equipped with the BTL, which will cover up to $|\eta| < 1.5$, while the endcap subsystem will cover between 1.6 and 3.0. A graphical representation of the MTD is provided in Figure 4.8. This structure aligns with the existing sections of the CMS detector, motivated by the necessity for service access and maintenance, as well as the deployment of different technologies in regions with varying radiation exposure.

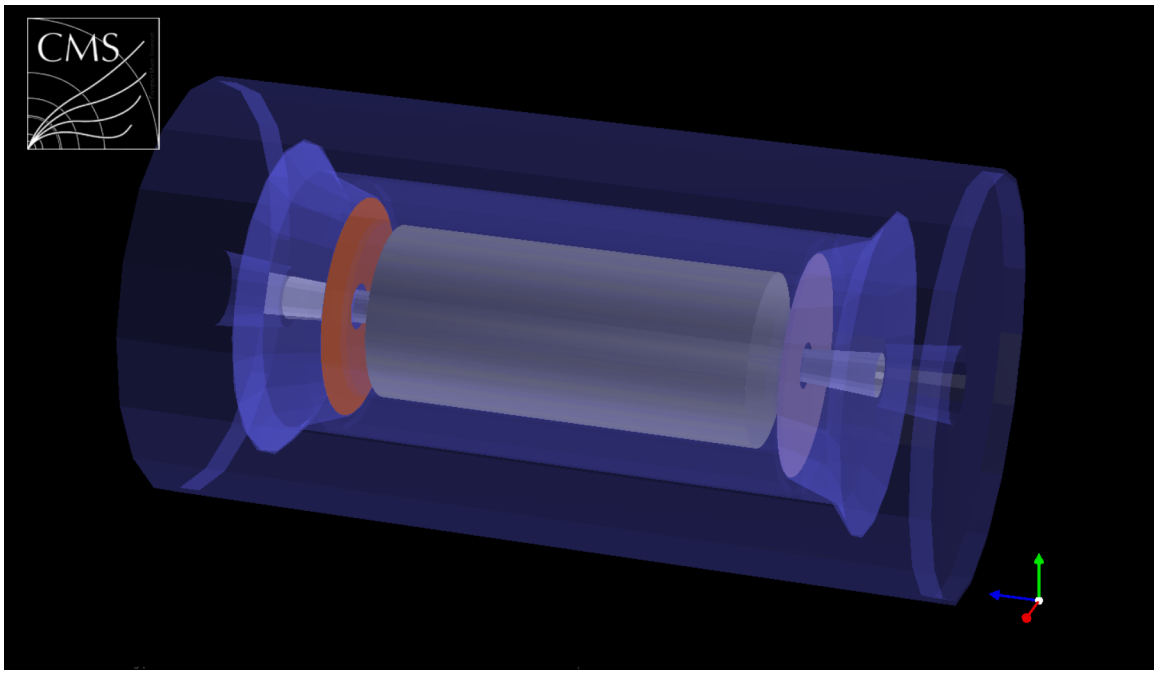

Figure 4.8: A graphical representation of the MTD as reproduced with the GEANT4 software is shown. The grey cylinder corresponds to the barrel section of the MTD, placed between the tracker and the ECAL, while the endcap sections are shown in orange and light violet, located in front of the endcap calorimeter [93].

4.3.2.1 Technical requirements

Stringent requirements are set for both the barrel and endcap regions of the MTD, resulting in specific constraints on the detector design. The main aspects are detailed below for each system in terms of mechanical limits, impact on the performance of other detectors, and radiation tolerance. Additionally, both systems must satisfy specific needs regarding occupancy and operating temperature. The radiation dose expected for the subsystems has been simulated using the FLUKA Monte Carlo multi-particle transport code with the available Phase-2 CMS geometry, assuming an integrated luminosity of at least 3000 fb^{-1}. At this level of integrated luminosity, detectors are expected to be exposed to a radiation dose of approximately 30 kGy and a 1 MeV neutron equivalent fluence, $\text{n}_{\text{eq}}/\text{cm}^2$, of up to 2×10^{14} $\text{n}_{\text{eq}}/\text{cm}^2$ in the barrel, and up to 450 kGy and 1.6×10^{15} $\text{n}_{\text{eq}}/\text{cm}^2$ in the endcap at high pseudorapidity. Hence, sensors have been developed and tested to demonstrate their ability to operate reliably under such irradiation levels. In addition, the sensor sizes must ensure a few percent occupancy, i.e. the fraction of hit channels over the total, and a manageable data volume. The MTD design requires high granularity and minimized dead regions that do not detect particles, in order to provide reliable timing information, while at the

same time the data volume must remain transferable without overwhelming CMS capabilities. Additional requirements such as power consumption, electronics, and cost must be considered in the design of the individual active elements constituting the MTD, guiding the choice of $\sim 1-2\,\mathrm{cm}^2$ for the BTL sensor cell size and $\sim 2\,\mathrm{mm}^2$ at $|\eta| \sim 3$ in the ETL. In terms of operating temperature, both systems must operate close to -30°C to reduce time jitter due to noise and leakage currents, which are correlated with temperature. The temperature control and cooling system will allow operation at this temperature, while still enabling maintenance at room temperature during technical stops or shutdown periods of the LHC. Operating across this range of temperatures requires the sensors to remain fully functional under several thermal cycles.

The BTL will be installed in the BTL Tracker Support Tube (BTST), i.e. a carbon fiber cylinder 2 m in diameter and 5 m long, shown in Figure 4.9, occupying a total radial space of 40 mm. The outer tracker, which will be entirely replaced, and the BTL services will be integrated together. The integration of the BTL must be completed before that of the tracker, imposing strict schedule constraints. The addition of the BTL reduces the outermost barrel layer of the tracker, but simulations have demonstrated that this has a negligible impact on track reconstruction performance. Tracker performance must also be preserved from mechanical and thermal perspectives, since the BTL and tracker will share the same volume and cooling services. The ECAL also requires specific temperature control. Therefore, the BTL must be designed so that it does not affect the ECAL's thermal conditions in any way. The BTST will be integrated into CMS, making it impossible to remove or modify its services. Thus, the BTL will not be accessible during its lifetime for repairs or maintenance, requiring it to be robust and fully reliable throughout the entire operation.

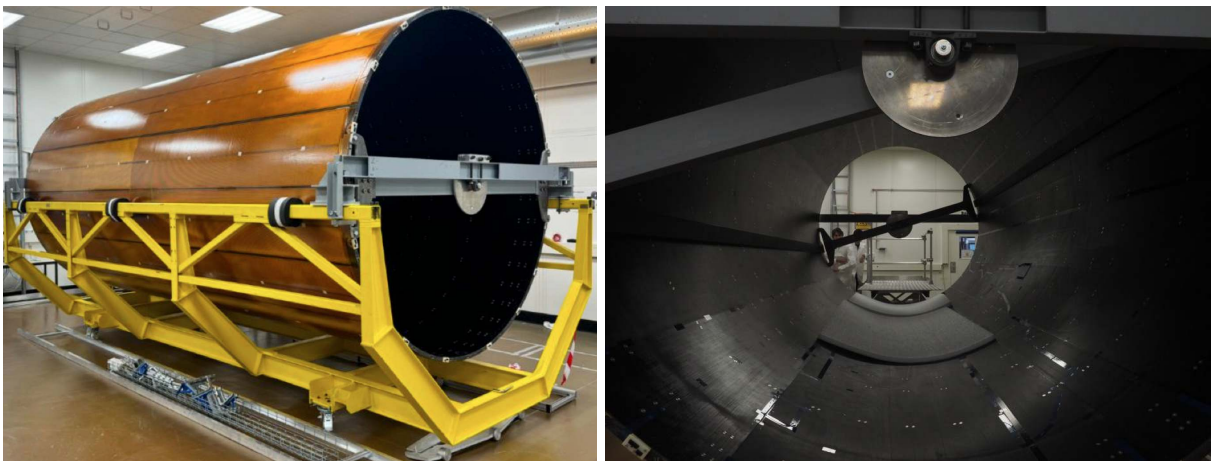

Figure 4.9: Two pictures of the actual BTST, currently located and under study at the Tracker Integration Facility (TIF) at CERN, are shown in this figure: the lateral view (left) and the internal view (right). The installation of the BTL trays will take place on the innermost external surface of this hollow cylinder. After its integration, the tracker, sharing this structure, will be integrated as the next step.

The ETL will be installed on the Calorimeter Endcap (CE) "nose" cone, between the tracker and the CE, with an allocated space of 45 mm. Due to the nose cone geometry, the ETL will not extend to $|\eta|$ =4 as the tracker does. Unlike the BTL and tracker, the ETL and CE will be integrated separately, allowing for a more flexible schedule and potential further sensor developments. The ETL and CE will occupy different cold and dry volumes, enabling separate operation. Nevertheless, the ETL will share the same cooling plant as the CE, therefore, it must not affect the CE's performance. The ETL will be accessible for repairs during operation, and it will be possible to remove segments and reinstall them during technical stops.

The set of constraints described above, combined with cost and schedule considerations, has led to the current MTD design: a thin layer integrated between the tracker and the electromagnetic calorimeter, covering both the barrel and endcap regions up to a pseudorapidity of $|\eta| = 3.0$. A detailed campaign

of research and development has been conducted for both ETL and BTL sensor technologies in order to meet all requirements. In particular, the different radiation exposure, the different surface coverage (the BTL features a surface approximately 2.5 times larger than that of the ETL), and the cost and schedule constraints have led to the following choice of sensor technologies: crystal scintillator bars coupled to Silicon Photomultipliers in the barrel, and Low Gain Avalanche Diodes (LGADs) in the endcaps. The option of crystal coupled to SiPMs offers reduced cost and a more suitable development timeline for the BTL, while LGADs provide superior performance in terms of time resolution at the expected radiation dose and can be developed with a more flexible schedule. The two subsystems will share a common clock, backend systems, detector slow control, and safety system.

4.3.3 BTL design overview

The barrel section of the MTD consists of a cylindrical detector, covering an active surface of 38 m^2, 40 mm thick, and extending 5200 mm along the z-axis. It will be installed in the BTST, covering a radial envelope between 1148 and 1188 mm from the beam. Considering the requirements on the BTL design, scintillating crystals coupled to SiPMs have proven to be the most accessible, production ready, and well-studied technology. The basic active element consists of a scintillating LYSO:Ce crystal bar, 54.7 mm long along the ϕ direction, 3.12 mm in the z direction, and 3.75 mm in radial thickness. Each end of the bar is coupled to a SiPM, and their time measurements are combined. A BTL Sensor Module (BTL SM) is composed of 16 crystal bars coupled to SiPMs at both ends, and its readout is performed through a dedicated ASIC, the Time-of-Flight High Rate (TOFHIR) [94]. The TOFHIR extracts timing information from the input current provided by 32 channels (i.e. 32 SiPMs, one SiPM per bar end) by discriminating signal pulses on the leading edge and converting them via a time-to-digital converter (TDC). Two TOFHIR chips are installed on a Front-End (FE) board, providing signal extraction for two BTL SMs, which together constitute a Detector Module (DM). Twelve DMs form a Readout Unit (RU), and the twelve FE boards are collectively read out by a Concentrator Card (CC). The CC provides low-voltage power, bias voltage, TEC bias voltage, and houses three low-power Gigabit Transceivers (lpGBTs) transmitting data, control, and monitoring information from the FE boards to the backend electronics in the CMS Underground Service Cavern (USC). Six RUs in a row form a tray, which is installed on cooling plates where a dedicated CO_2 system maintains a temperature of T = -35°C. In total, the BTL will be covered by 72 trays, as shown in the graphical representation in Figure 4.10, comprising a total of 331,776 channels.

Thorough studies have been conducted since the proposed design in the Technical Design Report (TDR) proposal [93] in order to achieve the desired time resolution performance, ranging from 30 to 60 ps at the end of operation. The main challenge lies in coping with radiation damage induced in SiPMs. The expected fluence will increase spurious counts within the sensors, i.e. the dark count rate (DCR), up to a few tens of GHz per SiPM, a technological challenge never previously faced in a large-scale experiment. Several key innovations and modifications have been introduced to finally reach the target performance, and they will be extensively described in Chapter 5, while a brief overview of the principles of operation of both BTL and ETL is given below.

4.3.3.1 BTL principles of operation

The basic interaction between particles and the BTL occurs through scintillation within crystal scintillators. In general, the crystal lattice of a scintillator features three energy bands: the valence band, the conduction band, and an intermediate forbidden band. Electrons in the valence band are normally bound at lattice sites, while electrons in the conduction band have sufficient energy to migrate throughout the crystal. When the material absorbs energy through particle interactions, an electron can be excited from the valence band to the conduction band, leaving behind a hole. Its return to the valence band generates the emission of a photon, which is usually too high in energy to be visible. In scintillators, this process is enhanced by the addition of activator impurities. These activators increase the probability of visible photon emission during de-excitation by providing sites in the forbidden band where electrons can return to the valence band. Since the energy released is lower, this transition produces a visible photon, and thus scintillation occurs. [95]. For the BTL, an inorganic material consisting of Lutetium Yttrium Orthosilicate crystals doped with Cerium (LYSO:Ce) was selected. In these scintillators, the emission centers are due

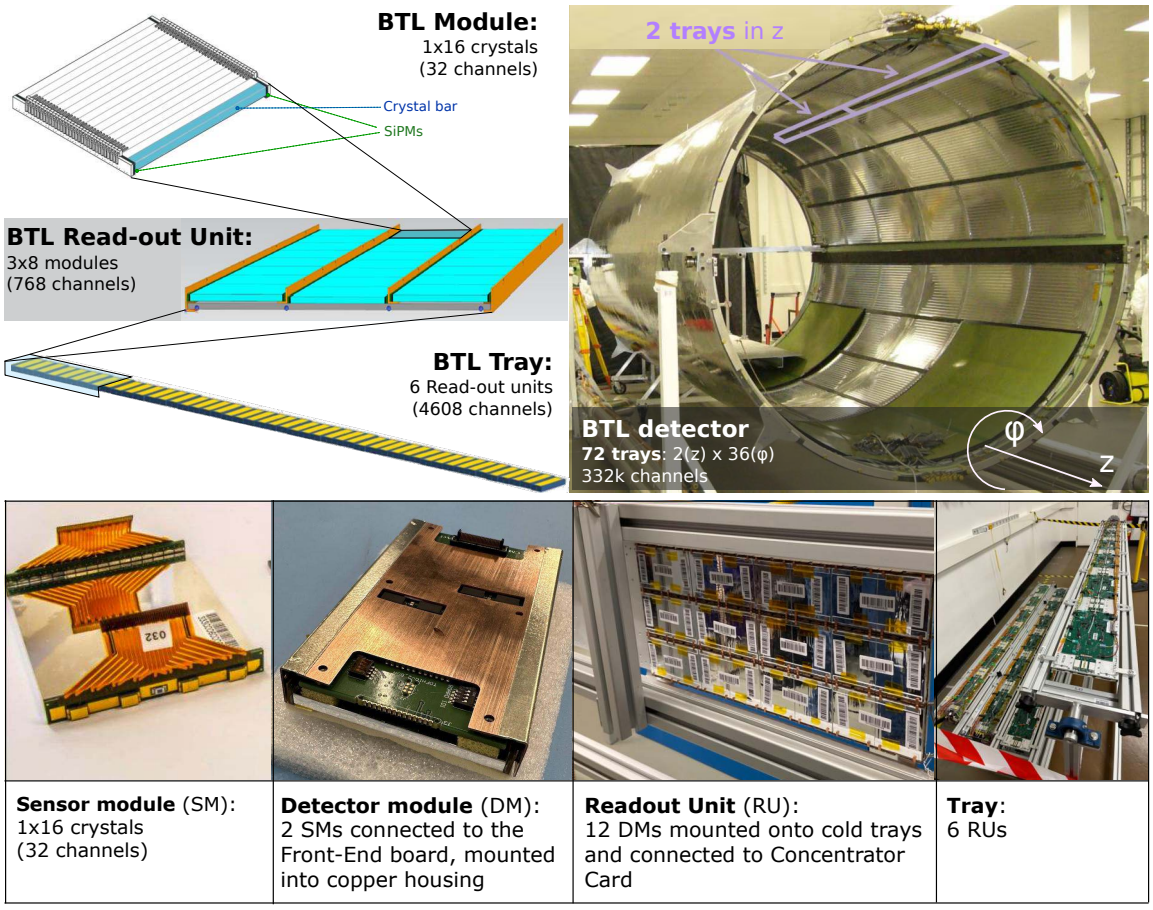

Figure 4.10: Top: Schematic representation of a BTL tray structure. The entire BTL surface will be covered by 72 trays, arranged in two layers along the z direction (the figure shows an initial mockup of the BTST, to be compared with the actual product in Figure 4.9) [93]. Bottom: The fundamental element is the sensor module, composed of 16 crystal bars coupled to SiPMs. Two sensor modules are connected to a front-end board to form a detector module. Twelve detector modules are mounted on a cooling plate and a Control Card, forming a Readout Unit. Finally, six RUs make up one tray. Right:

to the Cerium dopant: holes are trapped by Ce^{3+} ions, temporarily forming Ce^{4+}, which then captures a free electron and transitions to an excited state. This excited state subsequently de-excites, emitting a blue photon. Scintillation therefore consists of absorbing incoming radiation, converting its energy into electron-hole pairs, and transferring this energy to luminescent centers that cause the emission of light. Two main features characterize the performance of LYSO:Ce: the high light yield (LY), representing the amount of light produced per unit of energy deposited ($\sim 40,000$ photons per MeV), and the fast scintillation kinetics, with a decay time in the range of 40–45 ns depending on the manufacturer. These properties have attracted interest both in medical applications, such as Positron Emission Tomography (PET), and in High Energy Physics applications.

The light produced by the LYSO:Ce scintillators is read out by two SiPMs, one coupled to each end of the crystal bar, with active areas matching the cross-section of the crystal end faces to maximize light collection efficiency. The SiPM detects the light produced by the scintillator and via internal avalanche it amplifies the produced signal: above a breakdown voltage (V_{br}), both electrons and holes produced from the light detection can generate secondary e-h pairs. SiPMs provide fast response, making them particularly suitable for timing applications, as well as low operating voltage, compactness, and robustness. SiPMs coupled to a bar record two different times at which the MIP crosses the detector,

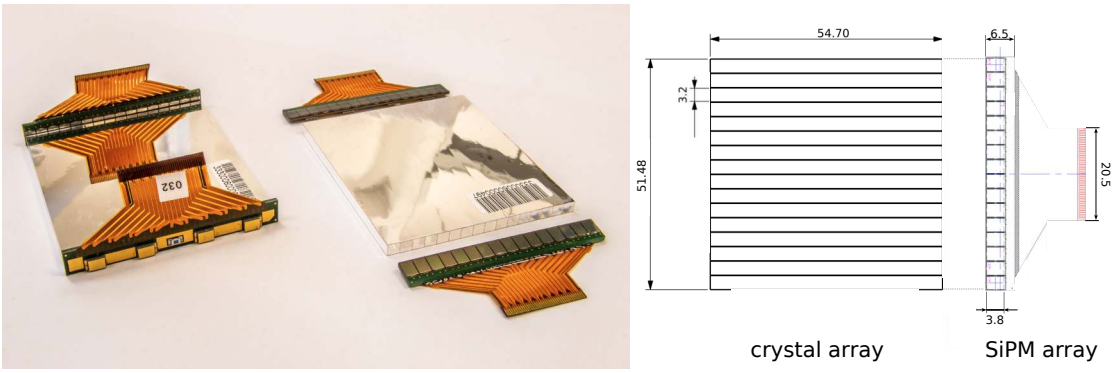

Figure 4.11: Left: Picture of a LYSO crystal array with SiPM arrays before and after gluing them together. Right: A schematic drawing of the crystal and SiPM arrays with dimensions given in millimeters.

referred to as the left and right time stamps, t_L and t_R. The jitter of these terms is assumed to be uncorrelated and to fluctuate independently following a Gaussian distribution. The time stamp measured at each SiPM is dependent on the impact position of the particle, due to the propagation time of optical photons in the crystal. The time information computed as the average of the two time stamps is, instead, independent on the impact position. It provides a uniform estimate of the time of arrival of the MIP along the entire bar, with a time resolution improvement by a factor $\sqrt{2}$ compared to the single readout per crystal. The coupling of LYSO:Ce crystals to SiPMs ultimately produces an electrical signal, and the timing information is extracted by discriminating the generated pulse. The time pick-off method employed in this technology is leading-edge discrimination, where the time is defined as the point at which the pulse crosses a specific threshold. Sources of uncertainty in this method may introduce time jitter due to random fluctuations in the pulse amplitude and shape. Indeed, the discrimination threshold may be crossed at slightly different times depending on these fluctuations, as shown in the left panel of Figure 4.12. However, a large slope of the pulse significantly reduces errors in time extraction with this method. As shown in the right panel of Figure 4.12, another source of uncertainty is the amplitude walk associated with the discrimination threshold, which can be corrected using an amplitude-related variable, such as the integrated charge.

4.3.3.2 BTL time resolution contributions

Along the detection chain, several effects introduce stochastic and systematic fluctuations that degrade the overall timing performance of the detector. The time resolution achieved with the BTL is given by the sum in quadrature of several contributions, as reported in Equation 4.1.

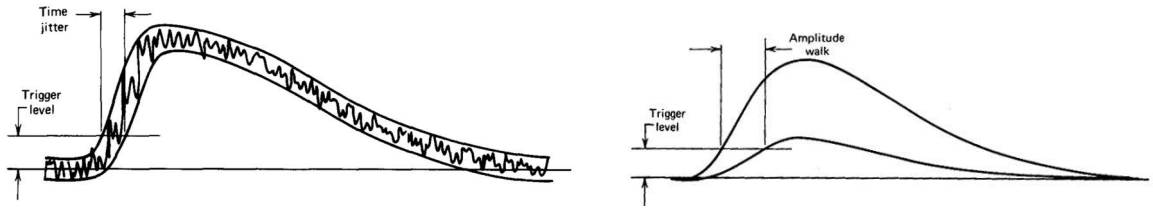

Figure 4.12: Left: An envelope of multiple signals with random noise contributions, illustrating the effect of trigger discrimination in terms of time jitter. Right: Pulses with identical shapes but different amplitudes cross the trigger threshold at different times, causing the so-called amplitude walk in leading-edge triggering [95].

$$\sigma_t^{BTL} = \sigma_{clock} \oplus \sigma_{electronic} \oplus \sigma_{photo-statistic} \oplus \sigma_{DCR} \tag{4.1}$$

All these uncertainty are largely dependent on the amount of photoelectrons (N_{pe}) produced in the detection of a charged particle. The number of photoelectrons is given by the product of the LY of the crystal, the light collection efficiency (LCE) of the detection system, the photon detection efficiency (PDE) of the SiPM, and the deposited energy. A more detailed discussion of these parameters is reported in Chapter 5.

- σ_{clock}: the CMS readout electronics are synchronized with the LHC bunch crossing. To satisfy the time resolution requirements of the MTD, the clock distribution system must provide less than 15 ps RMS link-to-link jitter across all distribution links.

- $\sigma_{electronic}$: time jitter due to electronics noise depends on the slope of the pulse (dI/dt) at the discrimination threshold, which increases approximately linearly with larger signals, i.e. with a higher number of photoelectrons N_{pe}. The electronics term is assumed to follow the relation $\sigma_{ele} \propto \frac{\sigma_{noise}}{dI/dt} \oplus \sigma_{TDC}$. The steepness of the pulse follows $\frac{dI}{dt} \propto N_{pe} \cdot G \cdot f_{SiPM}$, with G being the SiPM gain and f_{SiPM} a factor depending on the electrical properties of the SiPM, such as quenching resistance, cell capacitance, and grid capacitance [96]. The contributions from the TDC (σ_{TDC}) and from noise (σ_{noise}) have been evaluated through laboratory measurements, as discussed in Section 5.8.

- $\sigma_{photo-statistic}$: stochastic fluctuations in the time of arrival of photons detected by the SiPMs generate an additional uncertainty, expressed as $\sigma_{photo-statistic} \propto \sqrt{\frac{\tau_d}{N_{pe}}}$. The time constant τ_d correspond to the decay time of the scintillation pulse, amounting to approximately 43 ns for the LYSO:Ce crystals developed for the BTL. This contribution is expected to account for 25–35 ps.

- σ_{DCR}: radiation damage to the SiPMs generates an additional jitter that can be expressed as $\sigma_{DCR} \propto \frac{\sqrt{DCR}}{N_{pe}}$. The increase in integrated luminosity during HL-LHC operation will induce radiation damage in the silicon junction, creating defects in the material lattice and thereby generating dark counts. The magnitude of the DCR increases linearly with the fluence and it depends on several factors such as SiPM technology, operating temperature, annealing history, and operating voltage, as discussed in detail in Section 5.3.1. This contribution is expected to be negligible at the beginning of operation but may reach up to approximately 40 ps by the end of the detector's lifetime.

The sensor design has been carefully studied and optimized to minimize the impact of each contribution on the overall time resolution, with the goal of achieving the design target of 30-60 ps. Understanding the origin of each contribution and developing new strategies to reduce its impact has been a major focus of this work. This effort spanned several years and ultimately led to the validation of the final design's performance. Following the confirmation that the target performance could be achieved with this design, large-scale production of the detector began and is now nearing completion, with integration expected by February 2026 according to the current schedule. Most of this work has been dedicated to the development and validation of the target performance of this detector, along with the definition of the production protocols and quality procedures, thus a detailed discussion is presented in Chapter 5.

4.3.4 ETL design overview

The BTL will cover pseudorapidity up to $|\eta| = 1.48$, while the endcap region will be covered up to $|\eta| = 3.0$ by the ETL. The ETL consists of two disks placed on each side of the interaction region (IR), providing two measurements per track with a time resolution of 35 ps per track and 50 ps per hit. The disks will be positioned between the Tracker and the CE, at 2.98 m from the IR, covering a pseudorapidity region between 1.6 and 3.0, with a thickness of about 45 mm, plus 20 mm of additional space for thermal isolation from the other detectors. The space allotted to the ETL comes from a reduction of the neutron moderator of the CE, which protects the Tracker from neutrons produced in hadronic interactions within the CE. This reduction was proved to have only a marginal effect on the increase of particle flux in the Tracker. Since the ETL will be exposed to extremely harsh radiation conditions, the detector must remain accessible during technical stops for maintenance and repairs. For this reason, it is not placed inside the BTST or the CE. Instead, the disks are operated in a cold, dry volume, isolated from the other sub-detectors to ensure independent cooling and access. The disks are split in two to form a clamshell around the beam axis, facilitating installation and maintenance. The sensors will be mounted on structures called wedges, which can be installed and removed even with the beam pipe in place, simplifying repair procedures. The use of two disks is motivated by the need to cover areas occupied by cables or other infrastructure: by staggering the placement of sensors, hermetic coverage is achieved, providing on average 1.7 hits per track. The total sensor surface amounts to 7.9 m^2 per endcap, corresponding to approximately 8.5 million channels overall.

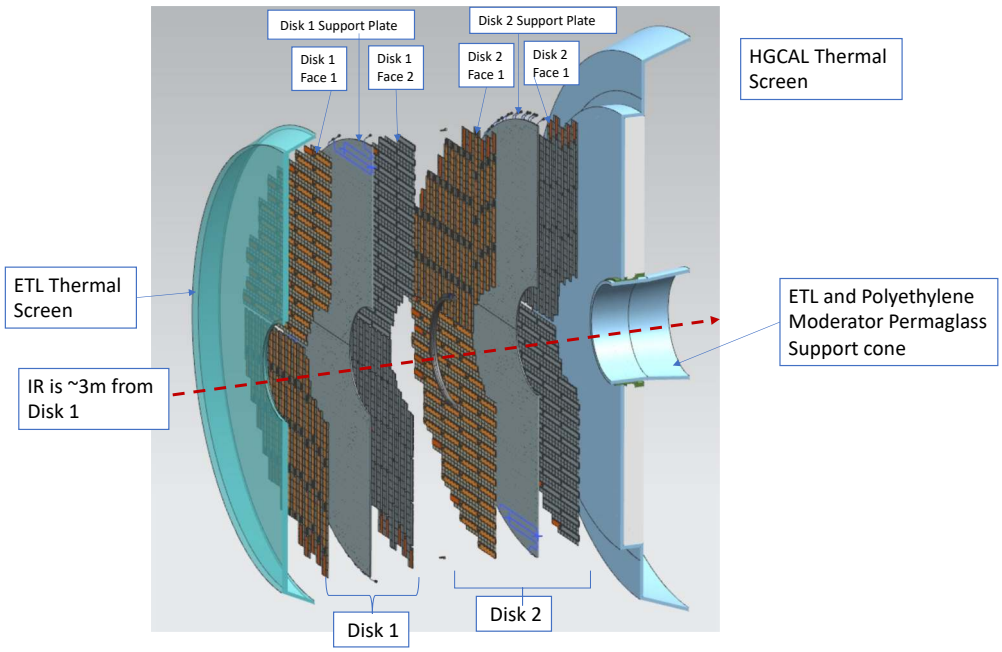

Figure 4.13: A cross-sectional view of the ETL along the beam axis (red dashed line) is shown, with the interaction point located on the left. The two ETL disks, instrumented with sensors on both faces, are visible as Disk 1 and Disk 2. Each disk consists of a support plate and two active faces with LGADs (in grey) and services (in orange). A thermal screen separates the ETL from the tracker (on the left), while a neutron moderator and another thermal screen isolate the ETL from the endcap calorimeter, protecting the tracker from particles backscattered from the HGCal [93].

The main motivation driving the choice of sensor technology for this sub-system is the radiation tolerance required in this region. The fluence is expected to reach up to 1.6×10^{15} n_{eq}/cm^2 in the innermost 12% of the detector, while 70% will be exposed to less than 8×10^{14} n_{eq}/cm^2. Therefore, the choice has been oriented toward silicon sensors, specifically Low Gain Avalanche Diodes: thin planar silicon sensors with an internal gain of about 10–30, capable of providing precise time measurements even after high radiation doses. The desired performance is expected with a 50 μm thick active region on a 300 μm thick silicon wafer, for a cell size smaller than 2 mm^2, balancing sensor capacitance. Each module

will consist of a 16 × 16 square LGAD array, with individual cells of size 1.3 × 1.3 mm^2, for a total module surface of 21 × 21 mm^2. The LGAD array requires a good uniformity across the channels and a dedicated design of the inter-pad regions to ensure stable and precise timing. A dedicated ASIC, the Endcap Timing Readout Chip (ETROC), has been developed. It is bump-bonded to the sensor matrix and provides the readout of a module. Timing information is extracted through leading-edge discrimination of the LGAD pulse, followed by circuits that measure the Time-of-Arrival (ToA) and the ToT, the latter enabling amplitude measurement and time-walk correction. The ETROCs then communicate with service hybrids, consisting of a power board and a readout board, which read the signals and send data to the USC, while also providing DC power, bias voltage, slow control, and monitoring.

4.3.4.1 ETL principles of operation

The chosen technology for the ETL is the LGAD, which is a silicon planar sensor based on the PIN diode concept: when fully depleted by a sufficient bias voltage, ionizing particles traversing the sensor generate electron-hole pairs, which move toward the electrodes following the electric field. The movement of the charge carriers induces a charge change on the electrodes, producing a measurable current. The energy loss of an ionizing particle follows a Landau distribution, and the induced current is described by the Shockley-Ramo theorem [97]. The main improvement in LGAD technology is the addition of an internal gain layer with a high Boron acceptor concentration ($\sim 10^{16}$ atoms/cm^3), where the electric field is high enough ($>$ 300 kV/cm) to allow electron-hole pairs to generate secondary pairs via impact ionization. This gain layer is implemented as a narrow p$^+$ doped layer of about 0.5–1 μm close to the n^{++} electrode, creating a drift region of 0.5–2 μm. Local depletion in this region generates the high electric field, while preventing electrical breakdown elsewhere in the sensor through a deep n-doped implant lateral spread, called the Junction Termination Extension (JTE). The gain layer depth can vary within the mentioned range, with shallow or deep implants achieving the same gain at different electric fields, the latter being lower than the former. The operating gain for LGADs is chosen in the range 10–30 to optimize the signal-to-noise ratio (SNR). Increasing gain improves the SNR as long as shot noise remains subdominant and electronics noise is constant. When shot noise becomes dominant, further increasing the gain amplifies the noise more than the signal, making the SNR worse. Therefore, a gain between 10 and 30 achieves the optimal balance. The excellent timing performance of LGADs arises from additional design features: thin sensors for fast signals, a uniform weighting field, a sufficiently high electric field to saturate carrier drift velocity, fine segmentation, and a high fill factor [97]. A comparison between a traditional silicon sensor and LGAD is shown in Figure 4.14.

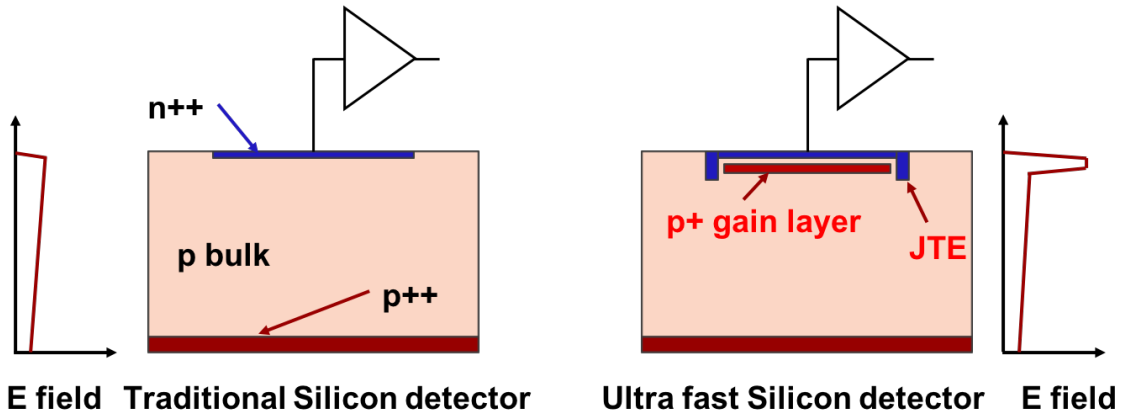

Figure 4.14: A schematic cross section of a traditional silicon sensor (left) and an ultra-fast silicon detector (right) is shown. The LGAD design includes an additional p$^+$ implant beneath the p–n junction. The electric fields in both structures are shown on the side, highlighting the effect of the p$^+$ gain layer on charge multiplication. The LGAD may also include a JTE to control the electric field at the sensor edges [98].

4.3.4.2 ETL time resolution contributions

The timing resolution of LGADs can be expressed as in Equation 4.2.

$$\sigma_t^2 = \sigma_{jitter}^2 \ + \ (\sigma_{total\ ionization} \ + \ \sigma_{local\ ionization})^2 \ + \sigma_{distortion}^2 \ + \ \sigma_{TDC}^2 \tag{4.2}$$

The individual contributions are described as follows:

- σ_{jitter}: contribution from early or late firing of the comparator due to electronics noise (shot noise is typically negligible [2]). The slope of the pulse can be expressed in terms of the total signal charge Q_{in}, the detector capacitance C_d, and the signal rise time t_{rise}, as $\sigma_{jitter} \approx \frac{NC_d}{Q_{in}}\sqrt{t_{rise}}$.
- $\sigma_{total\,ionization+local\,ionization}$: stochastic fluctuations in the ionization process. The total ionization fluctuation ($\sigma_{total\,ionization}$) is corrected using the ToT measurement. Local fluctuations ($\sigma_{local\ ionization}$), due to non-uniform creation of electron-hole pairs along the particle path (Landau noise), contribute irreducibly to the time resolution and depend on sensor thickness. For 50 μm thick sensors, this term is approximately 30 ps.
- $\sigma_{distortion}$: arises from non-uniform weighting fields or non-saturated drift velocity, both of which can be mitigated. Weighting field non-uniformities are minimized by adopting an almost ideal parallel-plate geometry. Drift velocity non-saturation is minimized by applying a sufficiently high bias voltage to ensure carriers reach saturated velocity.
- σ_{TDC}: contribution from TDC binning, assumed to be below 10 ps.

4.3.4.3 ETL sensors development

A thorough campaign of R&D and optimization of ETL sensors has been conducted to evaluate their performance and compliance with technical requirements. In recent years, both Hamamatsu Photonics (HPK) and Fondazione Bruno Kessler (FBK) 16 × 16 LGAD pad sensors have been tested at different irradiation levels. HPK sensors were shown to operate within the allowed voltage range up to 550 V, maintaining good performance up to a fluence of 1.5×10^{15} n_{eq}/cm^2, while FBK sensors delivered the required performance across all expected fluences. A short review of the results achieved so far by the ETL team is reported below, with samples produced in 2024, consisting of single-pad devices with carbon implantation to enhance radiation tolerance. Single-pad devices were characterized both in beam tests and laboratory measurements to evaluate performance in terms of collected charge, time resolution, and leakage current, under conditions equivalent to the beginning and end of ETL operation at the HL-LHC. Results on non-irradiated sensors are shown in Figure 4.15, reporting collected charge and achieved time resolution as a function of bias voltage. All samples exhibit collected charge above 8 fC at 175 V, as required, and time resolution values between 55 and 60 ps at the voltage corresponding to 8 fC (V(8fC)), meeting the requirement of being below 70 ps per hit.

Radiation damage in silicon sensors can be divided into surface and bulk damage. Surface damage consists of trapped positive charges at the Si/SiO_2 interfaces. Bulk damage arises from silicon interstitials and vacancies (Frenkel pairs) [97]. Some pairs may react with impurities to produce defects and clusters. Irradiation also reduces the LGAD gain via the initial acceptor removal mechanism, which deactivates the dopants forming the gain layer [98]. The microscopic origin is not fully established, but likely involves radiation generating atoms that deactivate doping elements through kick-out reactions, forming ion-acceptor complexes and reducing the electric field in the layer. To improve radiation tolerance, carbonated gain layers were introduced, forming ion-carbon complexes instead of ion-acceptor ones, improving radiation resistance by at least a factor of two compared to non-carbonated sensors [98, 100]. Also, increasing the bias voltage could mitigate this effect but it progressively increases the risk of Signal Event Burnouts (SEB) [101], which occur when highly ionizing particles create a conductive channel, locally burning the sensor. Figure 4.16 shows the crater formed after SEB. Safe operation is ensured for electric fields below about 12 V/μm [101]. Two samples were irradiated to low fluence (1×10^{15} n_{eq}/cm^2) and high fluence (1.5×10^{15} n_{eq}/cm^2) to emulate ETL end-of-operation conditions. Measurements, extensively described in [99], show time resolutions of 37 ps and 38 ps for low and high fluence, respectively,

[2] The LGADs operate with relatively low gain, so shot noise is typically subdominant compared to electronics noise. However, in heavily irradiated sensors, the leakage current may increase significantly, impacting timing performance.

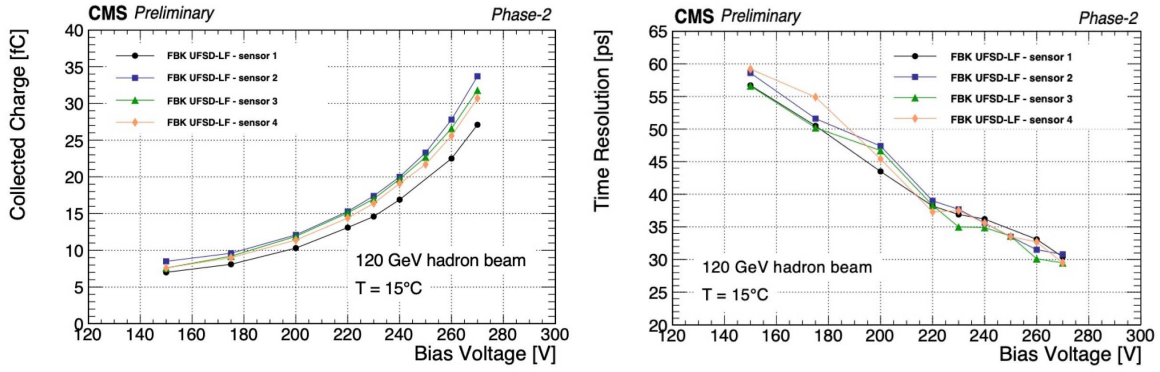

Figure 4.15: Four non-irradiated single-pad LGADs were tested simultaneously at a temperature of 15°C using a 120 GeV hadron beam. The performance is reported in terms of collected charge (left) and time resolution (right) as functions of the bias voltage for the four LGADs, each represented by a different color [99].

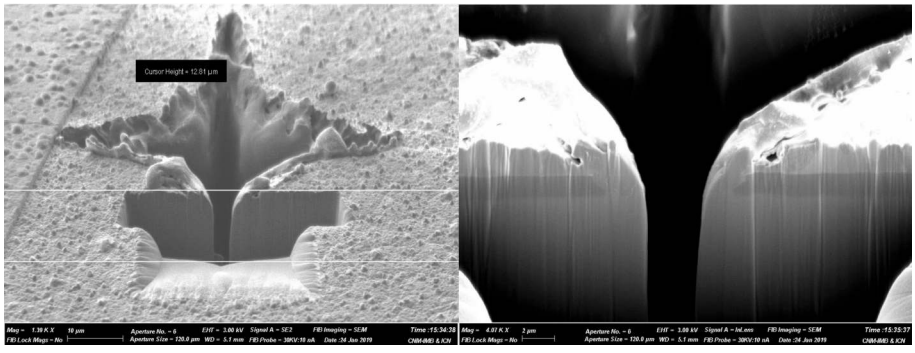

Figure 4.16: Microscope images of LGAD sensors showing the effects of Single Event Burnout. A visible crater is formed on the surface, leading to complete device failure [102].

at V(8fC). These results, shown in Figure 4.17 are well within the 50 ps requirement, demonstrating excellent performance after the full expected radiation dose.

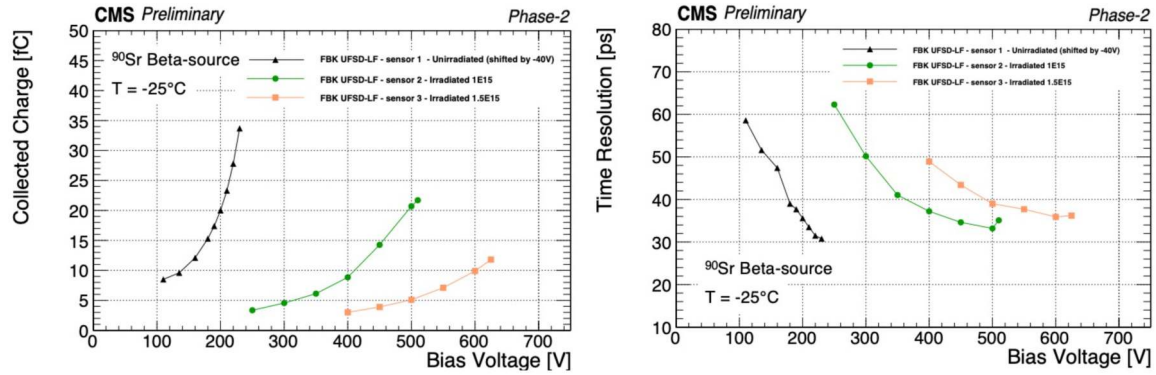

Figure 4.17: The performance of three LGAD devices is compared in terms of collected charge and time resolution as functions of the bias voltage. One unirradiated sensor (shifted by –40 V for clarity) is shown in black, while two irradiated sensors (at 1 and 1.5 n_{eq}/cm^2) are shown in green and orange, respectively. All measurements were performed at –25°C using a ^{90}Sr beta source [99].

Arrays of 16 × 16 pads were tested for breakdown voltage and leakage current. Breakdown voltages ranged between 245–270 V, compliant with the required range of 130–280 V. The leakage current at V(8fC) was 7–9 μA, below the 10 μA limit, but close to the threshold. Analysis attributed this to substrate material, prompting optimization of wafer selection by the vendor [103]. Interpad resistance was evaluated to ensure charge is collected only by the hit pad. The JTE implant provides high interpad resistance and pad isolation. Tests on a 1 × 2 pad array confirmed good isolation, verifying the quality of the production, as shown in Figure 4.18.

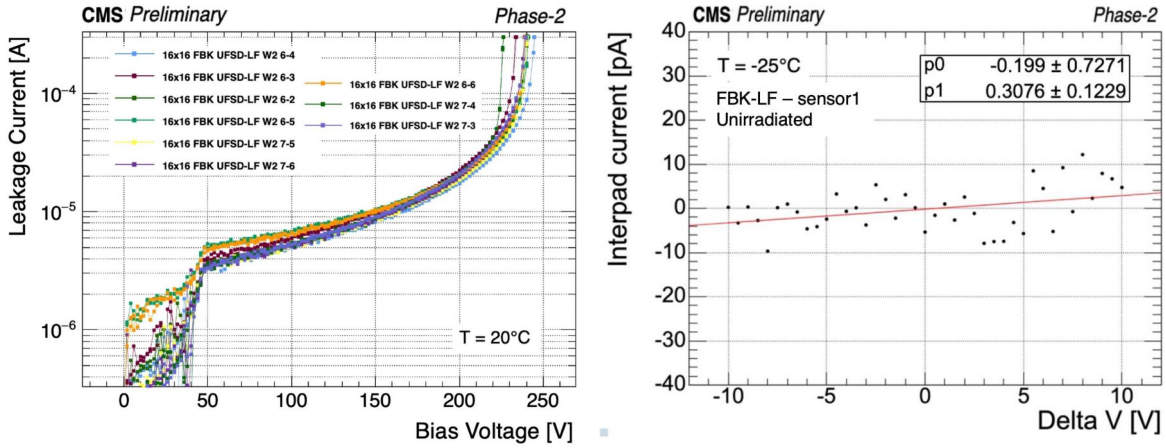

Figure 4.18: Left: The performance of nine unirradiated 16 × 16 LGAD matrices is compared in terms of leakage current as a function of bias voltage, measured at 20°C. The initial change in slope around 50 V corresponds to depletion of the gain layer, after which internal multiplication starts, producing the exponential trend. Right: A 1 × 2 LGAD array was tested at –25°C to evaluate inter-pad resistance using needle probes (two to bias the pads and one to ground the guard ring). The inter-pad resistance was measured as a function of the voltage difference between pads when applying 180 V, yielding a result of 2.85×10^{12} Ω for this device [99].

4.3.4.4 ETL module

The elementary unit of the ETL is the module, which consists of four LGAD arrays coupled to ETROCs (referred to as hybrids), a mechanical connection to the cooling system, and a PCB that interfaces with the readout board. Once the sensors demonstrate the desired performance, it is crucial to test the complete module. The ETROC ASIC is designed with 16 × 16 channels, each including a pre-amplifier, discriminator, and TDC, plus a buffer for data storage [104]. The ASIC is required to provide low noise and power consumption, radiation hardness, uniformity, and sensitivity to LGAD signals throughout the detector's lifetime. Beam tests on several hybrids demonstrated that ETROC+LGAD array modules achieve a time resolution of approximately 35 ps across a 70 V bias range, as shown in Figure 4.19, confirming the target of time resolution.

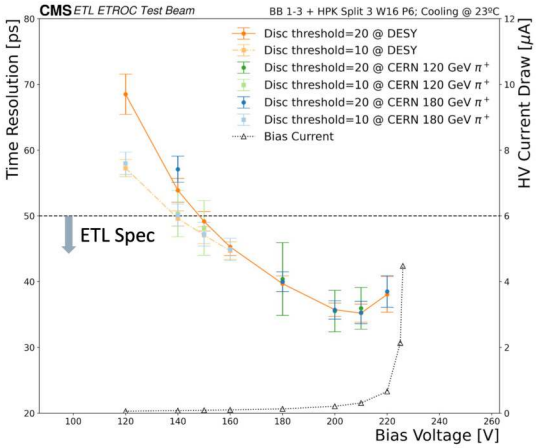

Figure 4.19: ETROC has been coupled with LGADs and tested both with a beam of 120 GeV pions and with a beam of 4 GeV electrons and positrons. The system demonstrated the ability to achieve the target resolution below 50 ps within a wide range of 70 V. The different lines correspond to different data-taking conditions with the different beams [105].

The ETL is approaching the production stage, with four assembly centers planned: Fermilab and Boston University in the USA, IFCA in Spain, INFN in Italy, and KNU in Korea. These centers are expected to produce approximately 8000 modules. Figure 4.20 shows the module design, including the ceramic baseplate with thermal pathway, PET thermal adhesive films, LGAD arrays, ETROC, and module PCB with power and readout interface.

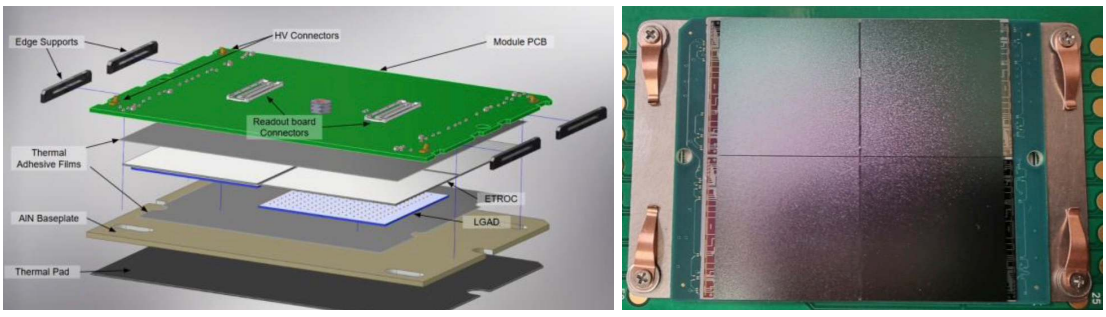

Figure 4.20: The ETL module design is shown in a sketch on the left and in a picture on the right. A ceramic baseplate with a thermal pathway is placed at the bottom; it is covered with PET thermal adhesive films, and on top are located the LGAD arrays, which are bump-bonded to the ETROC and finally read out through the top module PCB with power and readout interfaces [105].

4.3.5 The MTD common systems

4.3.5.1 DAQ

The MTD DAQ system is designed to collect data from the readout chips, reconstruct timing information, and transmit it to the central CMS DAQ. The system is shared between BTL and ETL. Data are first collected at the Front-End level, where devices operate close to the sensors and are subjected to high radiation doses. FEs must reliably collect signals, digitize them, and transmit them to the back-end electronics. The back-end consists of electronic boards with Field Programmable Gate Arrays (FPGA) that receive data from the FEs via high-speed bidirectional optical links and transmit them to the central CMS DAQ. The DAQ operates in response to L1 trigger requests, providing data only for signals above a configurable threshold (approximately 0.25 MIP). In the BTL, FEs receive the trigger and transmit data via high-speed optical links, with 874 links in BTL and 1600 in ETL. Data from the FEs are transmitted to the back-end via Versatile Links+, which include radiation-tolerant multi-gigabit communication ASICs, low-power Giga-Bit Transceivers [106], and optical transceivers (VTRx+). To handle the expected MTD data rates, lpGBTs operate in 5.12 or 10.24 Gb/s modes, depending on occupancy. Data are finally received in the USC by Firefly transceivers on MTD DAQ boards and sent to the event builder using a dedicated protocol. The back-end is managed by dedicated boards named Serenity [107], each equipped with two FPGAs and dedicated software and firmware for MTD operation. Additionally, the DAQ is responsible for distributing fast and slow control signals and for slow monitoring, including temperature and voltage control.

4.3.5.2 Clock

The LHC bunch crossings are synchronized with the CMS readout electronics at a rate of 40.078 MHz. To meet the MTD timing resolution goals, the precision clock must not introduce a jitter larger than 15 ps. Two alternative clock distribution schemes are under consideration: the baseline approach uses bidirectional DAQ links that send an encoded clock to the readout electronics, while the alternative approach samples the clock and distributes it without encoding, requiring additional optical fibers and transceivers for each electronics block. The performance of the embedded clock distribution system has been evaluated using the time interval error method, where a second clock distribution link serves as reference. The clock instabilities were measured to be within the design specifications, being less than 10 ps, and an RMS jitter of 3.2 ps was measured, well within the required limits, demonstrating that the embedded system is suitable for final deployment [108].

4.3.5.3 Cooling

The MTD will share the cooling system with the Tracker and the ECAL, with a total envisaged cooling power of approximately 600 kW. The system consists of two-phase, accumulator-controlled loop cooling plants installed in the USC cavern, sharing a common primary chiller on the P5 surface. CO_2 is pumped from the USC to the detectors through manifolds and patch panels. CO_2 is stored on the P5 surface and is used to either supply or remove cooling from the underground systems depending on operational conditions. The BTL cooling is fed by two manifolds located at the X0 floor and near the X4 floor of the UXC cavern, while the ETL is fed through two manifolds per YE1, integrated with the CE manifolds. The cooling system will be commissioned in standalone mode without detectors connected. During integration, the BTL cooling loop will be tested at the TIF, and the ETL wedges will be tested at Point 5.

4.4 Precision timing impact on CMS analyses

The precise time information provided by both the ETL and BTL will be implemented for CMS track and vertex reconstruction within the existing CMS reconstruction framework, ParticleFlow (see Section 2.3.1 for more details). Studies of the MTD physics impact rely on simulations of the Phase II CMS detector using GEANT4, with digitization emulated based on the assumptions of the detector design described in Section 4.3.2.

This section provides a general overview of how precision timing information from the MTD is integrated into track and vertex reconstruction, and its impact on pileup mitigation at the HL-LHC. The

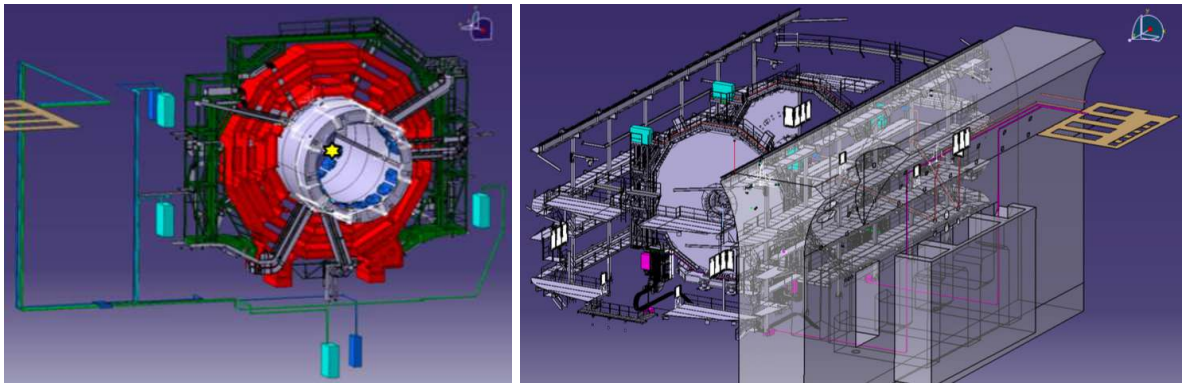

Figure 4.21: A graphical representation of the transfer lines and the CO_2 manifolds is shown, illustrating how they serve both the barrel section of the MTD (left) and the endcap (right). The BTL manifolds are expected to be located on the X0 and X4 floors, which are shown in blue, while on the right the ETL manifolds are represented as being expected at the top of the YE1 structure. [93]

subsequent discussion covers its effect on object reconstruction and the consequences for physics analyses. The purpose is to present the physics motivation and expected benefits of precision timing, which provide the framework for the detector development and performance studies described in this thesis. The first part summarizes the reconstruction strategy to include timing information into tracking and vertexing. The following subsections focus on the resulting improvements in object reconstruction and on the physics-level impact enabled by the inclusion of timing information.

4.4.1 Impact of timing on track and vertex reconstruction

The MTD time measurement is associated with a track and back-propagated to the time of the track at the collision vertex, i.e. the track time at the point of closest approach, using the particle's time of flight, as described in Equation 4.3.

$$t_{track} = t_{MTD} - TOF \tag{4.3}$$

The corresponding uncertainty is given by Equation 4.4.

$$\sigma_{t_{track}} = \sqrt{\sigma^2_{MTD} + \sigma^2_{TOF}} \tag{4.4}$$

The time of flight (TOF) of a particle with momentum p traveling a distance L is computed as:

$$TOF = \frac{L}{\beta(p,m)\,c} = \frac{L\sqrt{p^2+m^2}}{p} \tag{4.5}$$

For both subsystems, track reconstruction starts with energy reconstruction. A particle crossing the detector deposits energy in multiple cells, which are then clustered in adjacent cells associated with the same track. Tracks reconstructed in the tracker need to be matched to hits in the MTD. Studying the matching efficiency is crucial for improving resolution and reducing fake associations. The energy deposition in the ETL is uniform due to the constant sensor thickness, while in the BTL it varies according to the different slant thicknesses traversed by particles. After energy clustering, the spatial and temporal coordinates of the cluster are extracted as the barycenter weighted by the energy of individual hits. Tracks reconstructed by the inner detectors are propagated to the MTD surface. If a compatible cluster in position is found, the track is refitted, including MTD measurements. Specifically, the refitted track is back-propagated to the vertex and corrected for the particle's time of flight. Assuming the particle to be a pion, the time of flight is computed from the actual path length and velocity, allowing the estimation of the time at the beam line.

To evaluate the association efficiency between a track and an MTD cluster, matching with simulated particles is performed. Tracks are correctly associated with an MTD cluster approximately 80–85% of the time for muons, even at very high pileup. The efficiency decreases for pions due to nuclear interactions

in the tracker that can degrade the track. The efficiency does not depend on p_T, but it depends on the detector geometry, with BTL coverage around 82% and ETL exceeding 99% [109]. It is important to note that subtracting the time of flight from the MTD measurement introduces a small additional uncertainty. To evaluate this, the time measured by the MTD cluster and corrected for the TOF back to the beam line is compared with the simulated true time for a sample of pions produced in $t\bar{t}$ events at 14 TeV. The distributions of the difference between reconstructed and simulated times show compatibility with the expected MTD time resolution of 30–40 ps in both the BTL and ETL regions, indicating a negligible impact of the time-of-flight correction on the overall resolution [93]. Recent studies using an updated strategy, including the latest vertex reconstruction algorithm [110], were performed with $t\bar{t}$ events at a pileup of 200, considering tracks with $p_T > 0.7$ GeV. As shown in Equation 4.5, the uncertainty on track momentum affects the uncertainty on the estimated time of flight. This uncertainty was evaluated as a function of track momentum for three mass hypotheses: pion, kaon, and proton. The results, shown in Figure 4.22 (left), indicate that the TOF uncertainty is comparable to the MTD time uncertainty only for low-momentum protons ($p_T \leq 2$ GeV), and is much smaller otherwise. Consequently, the overall track time resolution is dominated by the MTD measurement, except for low-momentum protons, leading to a larger total track time uncertainty in that case (Figure 4.22, right).

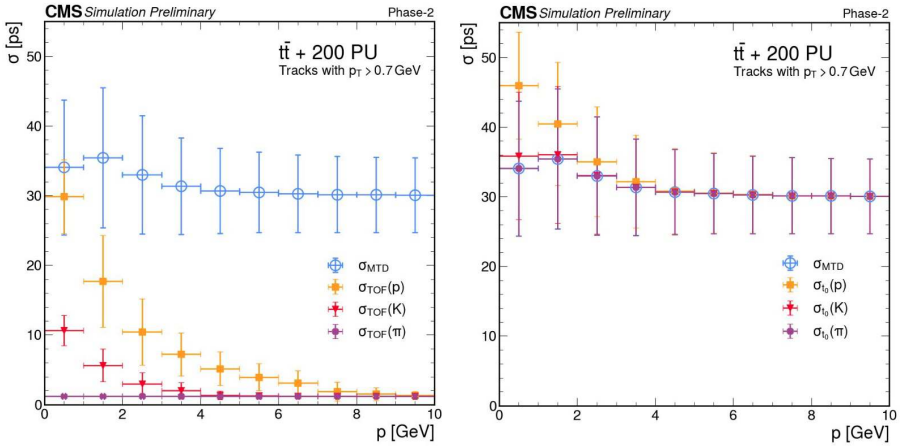

Figure 4.22: Studies performed using trakcs with a transverse momentum pt¿0.7 GeV from ttbar events in a condition of 200 pileup. Left: The resolution on the time of flight is impacted by the recontructed track momentum uncertainty. Its performance is evaluated as a function of the track momentum for three different mass hyportheses and it is compared to the estimated MTD hit time uncertainty. Right: The track time is also reported as a funciton of the transverse momentum under differen mass hypothese [110].

The extracted time information at the beam line can be used for reconstructing vertices in both space and time. A natural approach to integrate timing in vertex reconstruction is the 4D extension of the 3D deterministic annealing (DA) [63] currently used in CMS. Several strategies have been explored since the TDR studies [93]: the "*4DLegacy*" method [93], the "*4D*" method, and the "*3Dt*" algorithm [111]. Using the *4DLegacy* method, the inclusion of the time dimension reduces vertex merging from 15% to 1% and makes splitting negligible in a 200-pileup environment. This significantly improves object reconstruction, including pileup jet identification, MET reconstruction, b-jet tagging, and lepton isolation. The *4DLegacy* reconstruction is performed iteratively: first, a pion mass hypothesis is assumed and the uncertainty is inflated to account for differences among other mass hypotheses. The compatibility between the estimated vertex and the track is then assessed under alternative mass hypotheses. If a different mass hypothesis fits the track better, this information is used to refine the reconstruction. The vertex time is computed as a weighted average of all tracks associated with the vertex, using Equations 4.6 and 4.7.

$$t_{vertex} = \frac{\sum_i \frac{1}{\sigma^2_{t,i}} t_i}{\sum_i \frac{1}{\sigma^2_{t,i}}} \tag{4.6}$$

$$\sigma_{t_{vertex}} = \frac{1}{\sqrt{\sum_i \frac{1}{\sigma_{t,i}^2}}} \tag{4.7}$$

The *4D* reconstruction is an updated strategy compared to *4DLegacy*, replacing the first iteration with a 3D vertex reconstruction. The main limitation of *4DLegacy* is the initial mass hypothesis. By using the time computed on clustered 3D vertices as the initial guess, CPU time is reduced by approximately 30%. The *3Dt* method starts from a 3D vertex reconstruction, with the vertex time computed via the DA algorithm, which minimizes a cost function, based on track times under the three mass hypotheses (pion, kaon, proton).

A dedicated matching procedure was developed to compare reconstructed vertices to MC truth. Tracks from the same space-time point are associated with the same vertex. Each reconstructed track is matched to a simulated particle, which defines a true primary vertex. For each reconstructed vertex, simulated vertices are evaluated using a weight based on adaptive fit weight, spatial and time resolution, and the beam-spot time width. The simulated vertex with the highest weight is chosen. If a simulated vertex dominates multiple reconstructed vertices, it is assigned to the one with the highest weight. Matched vertices are removed from the set, and the iteration continues up to eight times. Reconstructed vertices matched within these iterations are labeled "real", otherwise, they are "fake". Figure 4.23 shows the number of real and fake vertices as a function of pileup using a 200-pileup $t\bar{t}$ sample for the three methods. The *3Dt* method yields an increased number of real vertices compared to the others, while also showing the highest number of fake vertices.

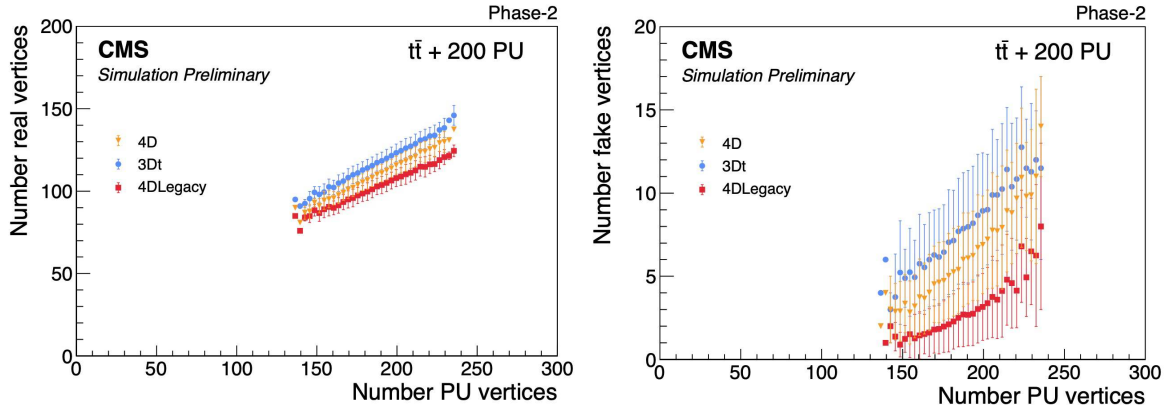

Figure 4.23: The number of real (left) and fake (right) reconstructed primary vertices is shown as a function of the number of pileup vertices, using different vertex reconstruction methods: in orange the *4D*, in blue the *3Dt*, and in red the *4DLegacy* [111].

The precision of the time measurement is assessed by computing the resolution as the difference between the reconstructed vertex time and the true simulation time. Time resolution distributions are fitted with a double Gaussian, and the parameters shown in Figure 4.24 correspond to the narrowest Gaussian component. The *4D* and *3Dt* methods improve the time resolution by about 7 ps compared to the *4DLegacy* method. A negative bias is observed for the *4DLegacy* method due to overestimation of the track mass. The *4D* and *3Dt* methods correct this bias by simultaneously considering all available mass hypotheses.

Using timing improves both track and vertex reconstruction and reduces pileup contamination. In CMS, pileup rejection is currently performed by requiring that the hard interaction vertex has a distance in z to the track smaller than 1 mm, as expressed in Equation 4.8.

$$|\Delta z(\text{track}, \text{PV})| < 1\,\text{mm} \tag{4.8}$$

At interaction densities above approximately 100 pileup events, this criterion may not sufficiently reduce pileup track contamination. The MTD time measurement adds a constraint on the time difference, as in

Equation 4.9:

$$|\Delta t(\text{track}, \text{PV})| < N \times \sigma_t^{\text{track}} \tag{4.9}$$

The three track and vertex reconstruction algorithms have been evaluated for pileup rejection. Different physics variables are used to quantify contamination from pileup tracks and jets in the leading vertex (the primary vertex with the highest p_T^2 sum). Tracks are categorized as associated with the primary vertex, originating from secondary vertices of the same pp collision, coming from pileup interactions, or fake tracks unmatched to any true particle. Vertex sorting is jet-based, so both track- and jet-level variables are considered. Tracks belonging to the same vertex are clustered using FastJet as AK4 jets and two track sets are considered: including and excluding pileup tracks. The difference between these sets quantifies pileup contamination, i.e. the fraction of pileup tracks over the total number of tracks. These variables measure the impact of pileup on the choice and quality of the primary vertex. Reducing these fractions improves vertex isolation and mitigates pileup. Results on track and jet variables are shown in Figure 4.24, indicating a pileup reduction of 10–15% with the *4DLegacy* and *4D* methods compared to *3Dt*.

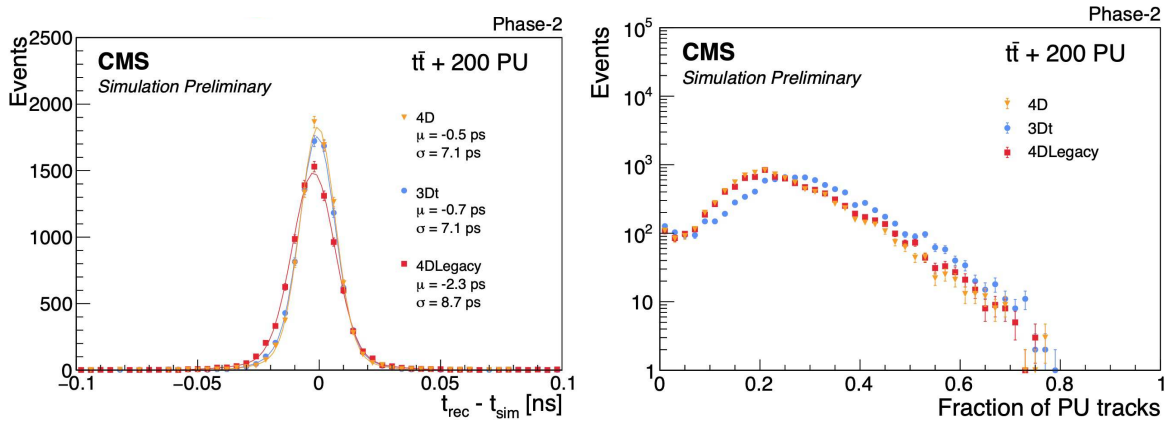

Figure 4.24: Left: Time resolution distributions for different vertex reconstruction methods are shown: in orange the 4D, in blue the 3Dt, and in red the 4DLegacy. The time resolution is computed as the difference between the reconstructed vertex time and the simulated vertex time, with a double-Gaussian fit being performed on the distributions. Right: Pileup contamination in track-level variable is shown for different vertex reconstruction methods [111].

The updated *3Dt* and *4D* methods outperform *4DLegacy* in vertex reconstruction and pileup mitigation, enabling more robust algorithms to fully exploit the precise MTD time measurements. Pileup reduction directly benefits physics analyses. In CMS Phase-2, pileup jets are expected to increase by up to 30% in the endcap regions, reducing signal efficiency. The MTD compensates for this degradation, enhancing the purity of selected jets.

4.4.2 Reconstruction of physics objects with the integration of time information

The impact of incorporating timing information into track and vertex reconstruction is reflected in many final-state observables. Preliminary studies are extensively discussed in Ref. [93]. Some examples of the improvements for various objects, such as those relevant in the $HH \to bb\tau\tau$ searches described in Chapter 3, are presented below.

4.4.2.1 Missing transverse energy reconstruction

MET reconstruction strongly depends on the correct reconstruction of all other objects, as described in Section 2.3.1. Pileup tracks erroneously associated with the primary vertex can bias the MET, especially at high pileup. Timing information helps mitigate this effect. For pileup increasing from 140 to 200 interactions, p_T^{miss} is expected to degrade by about 15%, affecting analyses that rely on its reconstruction.

The impact of timing on MET is evaluated using $Z \to \mu\mu$ events, associating tracks to the primary vertex with Equations 4.8 and 4.9, using $\mathrm{N} \times \sigma_{\mathrm{t}}^{\mathrm{track}} = 90$ ps. The p_T^{miss} performance is quantified via the scale and resolution of the hadronic recoil against the Z boson. Figure 4.25 shows that resolution worsens with increasing pileup. Including timing reduces this degradation, improving resolution by 10–15% at 200 pileup, effectively recovering performance equivalent to 140 pileup. High-p_T^{miss} tails (p_T^{miss}> 110 GeV), where real MET events are rare, also show reduced event counts with timing, decreasing background contamination in searches for new physics.

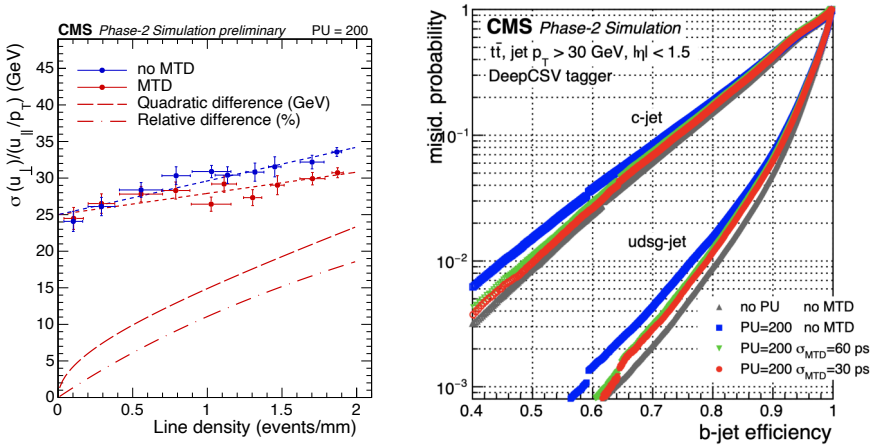

Figure 4.25: Left: The resolution of the hadronic recoil component perpendicular to the Z boson momentum is shown as a function of pileup density, with timing information (red) and without timing information (blue). Studies are conducted in a pileup scenario of 200 interactions. Right: The performance in tagging secondary vertices is shown in terms of ROC curves, illustrating heavy-flavor tagging performance with MTD timing in different scenarios in the barrel region [93].

4.4.2.2 Heavy flavor tagging

Secondary vertex identification is crucial for correctly tagging jet flavor, as discussed in Section 2.3.1. Pileup-induced spurious secondary vertices can contaminate signal events. Timing information from the MTD helps remove tracks inconsistent in time with the primary vertex, reducing spurious secondary vertices by 30% and improving the separation between b-jets and light-flavor jets, as illustrated by the ROC curves in Figure 4.25. The recovered performance in heavy flavor tagging at high pileup thanks to timing has direct implications for several physics analyses, being particularly valuable for di-Higgs searches, since the $b\bar{b}$ decay is the one with the highest branching fraction among Higgs decays.

4.4.2.3 Lepton isolation

Timing information also improves lepton isolation efficiency. Muons provide a clean test sample for performance studies. A sample of $Z \to \mu\mu$ events is used to evaluate isolation for prompt muons (produced at the interaction vertex from Z decay), and a $t\bar{t}$ sample is used for non-prompt muons (mostly from semi-leptonic heavy-flavor decays). Prompt muons are identified via generator matching. Isolation is quantified as the sum of transverse momentum of particles within a cone of $\Delta R < 0.3$ around the muon. Charged particles are required to have $p_T > 0.7$ GeV in the BTL and $p_T > 0.4$ GeV in the ETL and must be compatible with the PV using Equation 4.8 (and Equation 4.9 when timing is included). The ratio of the isolation sum to the muon p_T is compared to a threshold to define isolation. Efficiency is the fraction of muons passing this requirement. Figure 4.26 shows efficiency as a function of line vertex density for barrel and endcap. At 200 pileup, timing increases efficiency by 3–4% in the barrel and 6–7% in the endcap. Studies with electrons show similar efficiency gains, confirming that the improvements hold for single isolated tracks independently of flavor. Hadronic tau decays, discussed in Section 2.3.1.5, involve more complex signatures, making background rejection more challenging. Timing information is therefore even more beneficial, providing an efficiency gain of 10–15% at fixed jet misidentification probabilities typical of many Higgs analyses.

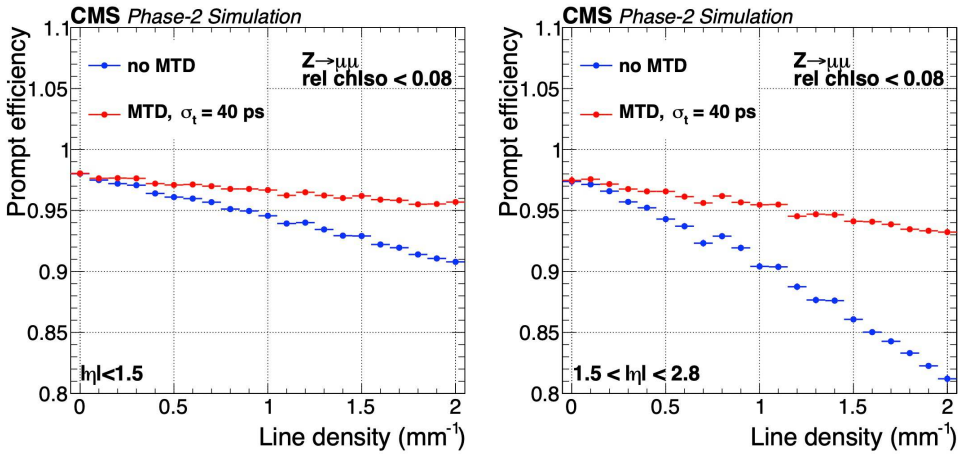

Figure 4.26: Muon isolation efficiency, with a lepton isolation of 0.08, is shown as a function of line vertex density for the BTL (left) and ETL (right). Two different scenarios are compared: in red, the case with MTD providing a time resolution of 40 ps, and in blue, the case without time information is reported [93].

4.4.2.4 Electron identification

The MTD can also be exploited to improve the identification purity of electrons. Single electrons and single charged pions reconstructed as electrons interact differently in the MTD, providing a useful handle for discrimination. Electrons typically deposit more energy in the MTD compared with pions, which behave as MIPs, as shown in Figure 4.27. Studies conducted for the TDR [93] trained a BDT to distinguish single electrons from single pions. A significant improvement in efficiency was observed in the barrel region, while a more modest gain was reported in the endcap [93].

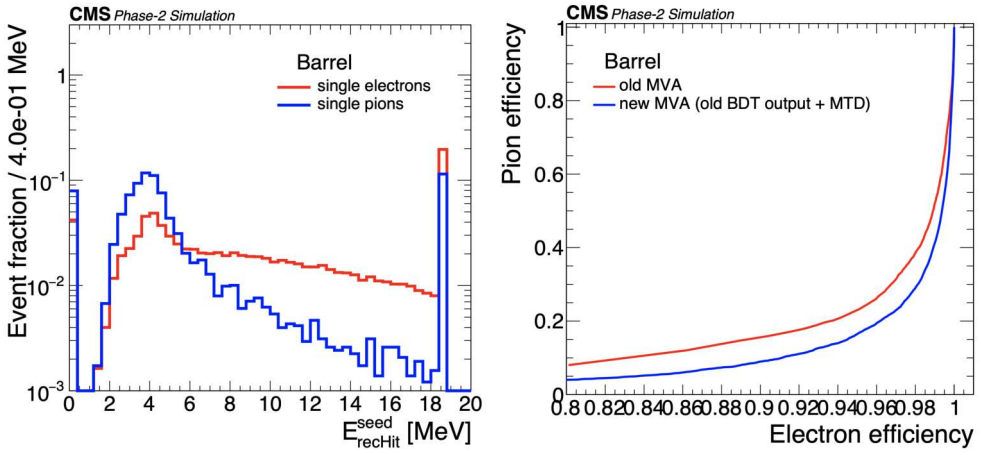

Figure 4.27: Left: The spectrum of energy deposited in the BTL by an isolated electron and an isolated pion is represented in red and blue, respectively. Right: ROC curves demonstrate the identification efficiency for the two particles in the barrel, using two different techniques: the default MVA discriminant used for Phase-2 studies is shown in red, while in blue the one involving the observables related to the MTD is shown [93].

4.4.2.5 Time-of-flight particle identification

MTD timing provides the CMS experiment with the capability to measure particle time of flight, enhancing particle identification precision. This is particularly valuable for Heavy Ion physics and studies focusing on low-mass QCD and flavor physics. Particle identification is performed by measuring particle velocities from the time of flight. The velocity is computed as in Equation 4.10, where L is the path

length between the vertex and the MTD, and Δt is the difference between the vertex time t^{vertex} and the MTD hit time t^{MTD}.

$$\frac{1}{\beta} = \frac{c(t^{MTD} - t^{vertex})}{L} \tag{4.10}$$

The expected performance as a function of p_T^{miss} and particle rapidity was studied. Figure 4.28 shows the separation between charged pions and kaons in the BTL and ETL. The discrimination improves at high rapidity due to increased momentum, which reduces the time-of-flight difference.

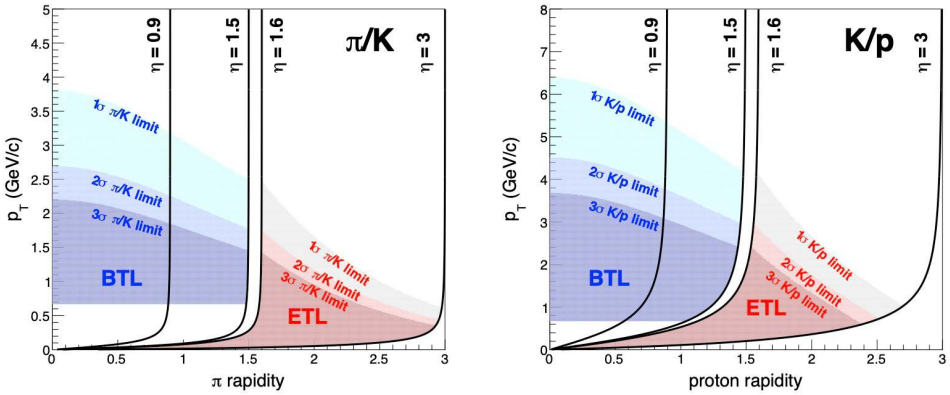

Figure 4.28: Separation between charged pions and kaons using MTD time-of-flight measurements is shown, with the expected performance for charged particle identification presented as a function of p_T and rapidity, assuming MTD timing of about 30 ps. Different colors correspond to different numbers of sigmas of separation. On the left, the π/K separation is shown, while on the right, the K/p separation is shown [93].

4.4.3 Physics analyses

The combined improvements in object reconstruction and pileup mitigation increase the efficiency of many analyses involving complex final states. Pileup mitigation alone can provide gains equivalent to 15–30% more integrated luminosity, corresponding to roughly three additional years of data taking without the MTD. For example, improved b-jet tagging enhances analyses of Higgs bosons decaying to $b\bar{b}$, which has the highest branching ratio. Enhanced MET reconstruction improves analyses involving $H \to \tau\tau$ decays. Together, these improvements directly impact di-Higgs searches, which require precise reconstruction of multiple objects. Improved lepton isolation from the MTD further contributes to di-Higgs analyses. The MTD's ability to distinguish displaced vertices also enables searches for long-lived particles in CMS. By measuring LLP velocities using Equation 4.10, their masses can be inferred, improving sensitivity to various theoretical models (e.g. Split-SUSY, GMSB).

4.4.3.1 Di-Higgs searches

The importance of di-Higgs searches was highlighted in Chapter 1 and showcased in Chapter 3. The increased integrated luminosity at the HL-LHC allows for a more precise measurement of the Higgs pair production. Current searches focus on different final states, balancing high branching ratios with clean experimental signatures to reduce background contamination. To evaluate the impact of MTD timing, studies were performed [93, 109] for the most sensitive final states: $bbbb$, $bb\tau\tau$, and $bb\gamma\gamma$. The cumulative improvements in object reconstruction provided by the MTD, described in the previous section, result in increases of 3–4% in lepton/photon efficiency, 4–6% in b-jet identification, and 10–15% in tau reconstruction. Consequently, the overall signal sensitivity improves by 15–35%, depending on the channel. The impact is especially significant for the $bb\tau\tau$ final state, where improvements in both b-jet and tau reconstruction enhance sensitivity. In this channel, the inclusion of MTD timing improves the expected significance from 1.30 (no MTD) to 1.60 for early-operation conditions, corresponding to an increase of about 23%. Studies considered different MTD time resolutions, corresponding to early and late operation conditions, accounting for BTL performance degradation due to radiation damage on SiPMs. The

expected significance improvements provided by the MTD are reported in Table 4.1. The MTD signifi-

Channel	MTD @ $\sigma_t = 35$ ps	ETL @ $\sigma_t = 35$ ps and BTL @ $\sigma_t = 70$ ps	No MTD
bbbb	0.95	0.94	0.88
$bb\tau\tau$	1.60	1.44	1.30
$bb\gamma\gamma$	1.90	1.81	1.70
Combined	2.66	2.51	2.31

Table 4.1: Projected significance for the SM HH signal decaying into different final states, i.e. the three main channels driving the sensitivity of the search, when exploiting MTD timing. Values are given in units of σ for an integrated luminosity of 3000 fb^{-1}, and different time resolution scenarios are considered, corresponding to no time information, time information at the beginning of operation, and at the end-of-operation conditions [109].

cantly improves all HH search channels, with cumulative gains when multiple object reconstructions are combined. These improvements correspond to roughly 2–3 additional years of data taking, even when accounting for performance degradation.

4.4.3.2 Long-Lived Particles

Many BSM models predict unstable particles with long lifetimes, capable of traveling measurable distances in the detector before decaying. Long lifetimes can arise from weak interactions, high masses, or suppressed decay phase space. The MTD time-of-flight measurement allows the time between LLP production and its decay (secondary vertex) to be measured. This timing enables the determination of the particle's velocity, Lorentz factor, and, if the energy is known, its mass. Different scenarios of long-lived particles have been evaluated in the context of the impact of MTD timing information on physics analyses, as presented in the TDR studies [93]. A relevant case study has been recently updated, considering the decay of a neutralino into a photon and a gravitino. Since the photon is produced at the neutralino decay vertex, which does not coincide with the primary vertex, its arrival time at the detector can be exploited to discriminate signal from background. The neutralino time of flight can then be inferred from the photon timing. In this context, an accurate MTD time resolution is crucial to determine both the primary vertex time and the photon arrival time. An analysis was performed to assess the impact of different MTD performance scenarios on the sensitivity, as shown in Figure 4.29. The results clearly demonstrate a significant improvement in sensitivity for short lifetimes and high neutralino masses, achievable only thanks to the inclusion of MTD time-of-flight information.

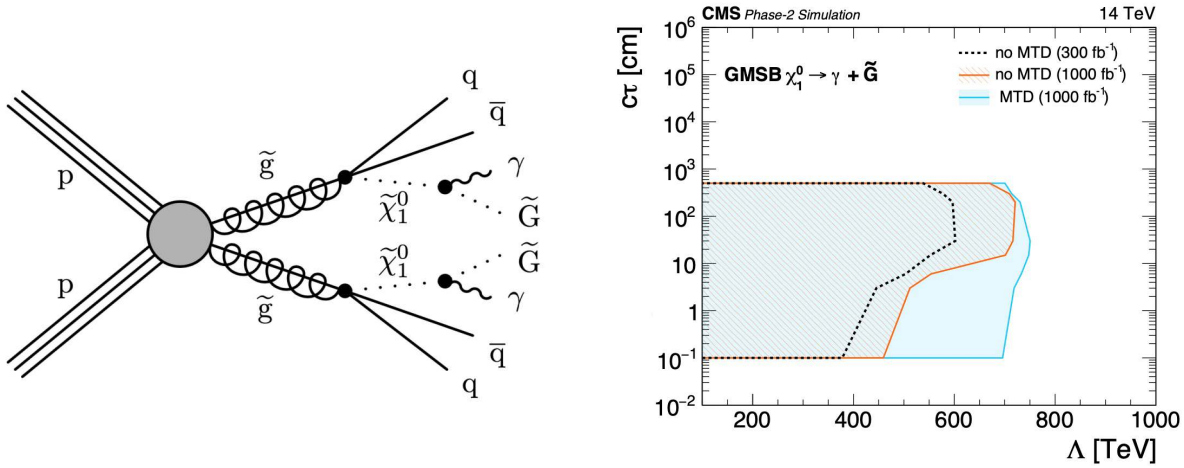

Figure 4.29: Left: The Feynman diagram for the SUSY process resulting in a diphoton final state via gluino production is shown. Right: Sensitivity to GMSB $\tilde{\chi}_1^0 \to \tilde{G} + \gamma$ signals, expressed in terms of neutralino lifetimes, for 180 ps and 30 ps timing resolution, corresponding to the Phase-2 detector without MTD and with MTD, respectively. The impact of MTD timing on neutralino decays into photon + gravitino is illustrated. Studies show increased sensitivity at short lifetimes and high masses [93]. Similar results are obtained for degraded BTL performance [109].

Chapter 5

The MTD Barrel Timing Layer: from design optimization to detector construction

The MTD is designed to cope with the high pileup conditions expected at the HL-LHC, having a significant impact on the reconstruction efficiency of vertices and tracks, and thereby improving the overall performance of object reconstruction. Without such timing information, pileup degrades the reconstruction of physics objects and, consequently, the sensitivity of many analyses, especially those involving complex multi-object final states, such as di-Higgs searches. The expected improvements for many physics cases require a precise time resolution of 30–60 ps, imposing stringent requirements on the sensor performance.

The present work has been specifically dedicated to R&D and characterization of the barrel section of the MTD, in order to meet the target performance. The BTL is expected to cover a pseudorapidity region up to $|\eta| = 1.48$ in CMS. Since it will be housed within the BTST, which will remain installed during the entire HL-LHC phase, its long-term reliability is a major priority, requiring a robust design and extensive testing. Schedule and cost constraints guided the choice of sensors toward mature technologies suitable for mass production within the project timeline, as described in Sections 5.2, 5.3. The sensor technology choice of LYSO:Ce crystals and SiPMs was motivated by the fact that they are well-established technologies with industrial-scale production capabilities and are widely used in Positron Emission Tomography. Moreover, due to the space constraints discussed in Section 4.3.2.1, the BTL must be integrated into the BTST without delaying the tracker installation.

Since the TDR, several sensor characteristics have been refined following the outcomes of dedicated development studies, as detailed in Sections 5.5, 5.6, 5.7. A comprehensive study of the time performance evolution throughout the BTL lifetime has been carried out, focusing on understanding and optimizing all factors affecting the overall performance. The experience gained during the R&D phase provided a solid foundation for developing procedures, protocols for large-scale assembly, and quality assurance and quality control (QA/QC), ensuring the best possible performance both at the sensor module and detector module level. Following the validation of the final sensor design, qualification and assembly procedures have been defined and applied to BTL mass production, as discussed in Section 5.9 and Section 5.10. A perspective on the upcoming integration and commissioning phases at CMS is presented in Section 5.11.

A large part of this work has focused on the R&D of the sensors constituting the BTL, with the goal of optimizing their performance to achieve the target resolution under the expected operating conditions within the MTD at the HL-LHC. Within the following discussion, the chapter integrates both general context and the specific contributions of this thesis. In particular, this work addresses the optimization of SiPM performance, the maximization of light output, the execution and analysis of several test beam campaigns, the definition and implementation of QA/QC procedures for sensor modules, detector modules, and trays. The qualification of individual LYSO crystals and SiPMs during large-scale production, as well as the development of the readout electronics, are part of the broader BTL effort and were carried

out by dedicated groups. These aspects are therefore discussed only at the level required for completeness and are referenced appropriately.

5.1 Time resolution drivers in BTL

The Barrel Timing Layer is designed to measure the time of arrival of minimum ionizing particles with a resolution of 30 ps, which is expected to degrade to 60 ps during HL-LHC operation due to radiation damage. The sensitive element of the BTL is a LYSO scintillating crystal coupled to SiPMs, operating as described in Section 4.3.3.1. A BTL sensor module consists of an array of 16 LYSO:Ce crystal bars, coupled to two SiPM arrays. The crystals are designed with a bar-like geometry exploiting total internal reflection to enhance light collection efficiency while minimizing the SiPM active area, thereby reducing both costs and power consumption. The SiPM active area is chosen to closely match the crystal bar end face, maximizing light collection efficiency. The designed granularity ensures an average occupancy of about 8% at a pileup level of 200, reducing the probability of double hits in the same cell during a bunch crossing and reducing the impact of pileup effects from previous crossings on the time measurement. Each crystal bar measures 54.7 mm in length, 3.12 mm in width, and 3.75 mm in thickness[1]. The longitudinal axis of the crystal bars is oriented along the ϕ direction (see Figure 2.2 for the CMS coordinate system). The double readout guarantees uniform response across the crystal surface, along with tracking capability and redundancy in time measurement. A MIP crossing the scintillating material produces scintillation light proportional to the light yield of the crystal. A fraction of these photons (defining the LCE) reaches the SiPM active area. Only a fraction of those photons is actually detected by the SiPM, quantified by the SiPM PDE. The detected photons are converted into electrons, usually referred to as photoelectrons, which are then amplified with a gain of the order of $10^5 - 10^6$. The resulting current signal is discriminated and digitized by the TOFHIR chip to extract the timing information related to the MIP arrival, also referred to as the "time stamp".

The overall time resolution is described in Equation 4.1 and is dominated by electronics noise (Equation 5.1), the photo-statistics term (Equation 5.2), and the DCR contribution (Equation 5.3), which becomes more relevant towards the end of the detector lifetime due to radiation damage.

$$\sigma_t^{ele} = \frac{\sigma_{noise}}{dI/dt} \oplus \sigma_{TDC} \tag{5.1}$$

$$\sigma_t^{phot} \propto \frac{1}{N_{pe}^{\alpha}} \quad \text{with} \quad \alpha \approx 0.5 \tag{5.2}$$

$$\sigma_t^{DCR} \propto \frac{DCR^{\beta}}{N_{pe}} \quad \text{with} \quad \beta \approx 0.5 \tag{5.3}$$

Each of these terms strongly depends on the signal amplitude, i.e. the number of photoelectrons detected at the SiPM, which is defined in Equation 5.4.

$$N_{pe} = LO \cdot PDE \cdot E_{dep} = LY \cdot LCE \cdot PDE \cdot E_{dep} \tag{5.4}$$

This quantity is determined by the intrinsic light yield (LY) of the crystal, the light collection efficiency (LCE) of the sensor module, the photon detection efficiency (PDE) of the SiPM, and the energy deposited by the MIP in the crystal (E_{dep}). The light yield depends on the specific scintillator choice, which in the case of the LYSO:Ce crystals is expected to be on the order of 40,000 photons/MeV. More details on the crystal specifications are provided in Section 5.2. The light collection efficiency accounts for several effects related to losses in light transport between the LYSO crystals and the SiPMs, and can be expressed as in Equation 5.5:

$$LCE = \epsilon_{\text{coll}} \cdot \epsilon_{\text{det}} \tag{5.5}$$

where ϵ_{coll} is the efficiency of collecting scintillation light at the end of the crystal bar, which depends on self-absorption, scattering, reflection at the crystal surface, and reflection from the ESR covering. ϵ_{det} accounts for the inefficiencies of photons reaching the active area of the SiPM, depending on the refractive

[1]The studies discussed in Section 5.5 guided the definition of the final sensor specifications, which are here reported and differ from the initial design reported in the TDR [93].

index and thickness of the medium between the crystal and the SiPM (i.e. the glue), as well as on the thickness and refractive index of the SiPM protective window and the geometry of the crystal end face and the SiPM active area. Studies focusing on the maximization of such efficiencies were conducted and are described in Section 5.4. An overall light collection efficiency of 10% is expected per BTL SiPM, resulting in a light output (LO) of approximately 4,000 photons/MeV per readout end.

Overall, the number of photoelectrons is expected to vary throughout the BTL lifetime due to different operating conditions, in particular changes in the SiPM bias voltage that affect the PDE, and irradiation effects that induce a further PDE loss in the SiPMs (more details on the SiPM parameters are given in Section 5.3). The LY is instead expected to remain unaffected by the irradiation levels foreseen for the HL-LHC, while a small impact is expected on LCE due to transparency loss (as discussed in Section 5.2). Additionally, the signal amplitude is expected to vary across the pseudorapidity region of the detector because of the different energy deposition per MIP as a function of η, resulting in a varying number of detected photoelectrons. The final sensor design is expected to produce approximately 12,500 photoelectrons for each MIP crossing the crystal at the beginning of operation, degrading to about 6,500 by the end of operation due to SiPM irradiation and the reduction of the operating bias voltage. These values are obtained by considering a light output of about 2,400 photoelectrons per MeV of energy deposit, and overvoltages of 3.5 and 1.0 V, which correspond to a PDE of 57% and 30% and a gain of 10^6 and 3.6×10^5 at the beginning and end of operation, respectively.

Reaching this final configuration required a thorough optimization campaign aimed at achieving the design performance, All the mentioned effects contribute to the overall time resolution measured at the BTL, therefore requiring a thorough evaluation of each component. Models and measurements were performed to evaluate the final performance with different sensor specifications and operating conditions. A detailed discussion of such studies is provided in the following sections, particularly addressing strategies for:

- Increasing the light output by optimizing the efficiency of light collection and augmenting the energy deposit, as reported in Section 5.4, together with the corresponding evaluation on beam in Section 5.6.2.
- Increasing the signal amplitude by optimizing the SiPM parameters that define the pulse shape, as outlined in Section 5.3.2, together with the corresponding evaluation on beam in Section 5.6.1.
- Reducing the dark count rate in SiPMs caused by radiation damage, as discussed in Section 5.3.1.

5.2 LYSO:Ce crystals characteristics

The basic detection principle of the BTL relies on the interaction of particles with the scintillating crystals. The passage of charged particles through the scintillator excites the crystal lattice. As the material de-excites, photons are emitted in an amount proportional to the energy deposited in the scintillator. Impurities or activator ions (dopants) may be introduced into the lattice to enhance light emission and reduce self-absorption. In the case of LYSO:Ce, the activator ions are Ce^{3+}. The emitted light then propagates through the crystal, with its transmission determined by the material's optical properties, such as absorption and scattering.

Cerium-doped Lutetium Yttrium Orthosilicate (LYSO:Ce – $[Lu_{(1-x)}Y_x]_2SiO_5 : Ce$) crystals were selected as the optimal scintillator for the BTL due to their high intrinsic light yield of about 40,000 photons/MeV, fast scintillation rise time (< 100 ps), and short decay time (~ 40 ns) [112]. In general, Yttrium and Cerium content might vary in crystals, depending on the production process. The stoichiometry group ($[Lu_{(1-x)}Y_x]$) is not fixed and can vary slightly depending on the crystal growth, typically with $x < 0.1$. The Yttrium fraction is expected to affect mass density and thus MIP energy deposition, whereas the Cerium content is expected to correlate with the decay time. In BTL LYSO:Ce crystals, a high density of 7.1 g/cm^3 is measured, corresponding to an energy deposit by a MIP with a Landau most probable value (MPV) of 0.86 MeV/mm. Precision timing performance is typically evaluated based on the number of photons produced within a $\sim$500 ps window after the onset of scintillation. For LYSO:Ce

crystals read out at both ends, this corresponds to a light output of approximately 2,000 photons. The high density also enables smaller, more compact crystals that fit within the BTL's radial envelope in CMS. LYSO:Ce crystals are non-hygroscopic, meaning they do not absorb humidity, and they emit at 420 nm, well matched to the SiPM sensitivity range. They are also widely used and well-studied in PET and in high-precision calorimeters such as those developed for Mu2e and KLOE-2 [113]. Alternative scintillators were investigated, but LYSO:Ce offered the best balance of performance, radiation tolerance, cost, and scalability for mass production. The bar-like geometry minimizes the SiPM area compared to the crystal volume, thus limiting the power consumption and channel count. Also, it is strategic for achieving optimal LCE, which relies strongly on photons undergoing total internal reflection. This geometry results in an average photon path length equal to the distance from the impact point to the SiPM. Each of the 16 bars in the array is individually wrapped to ensure optical isolation and suppress light cross-talk. ESR (Enhanced Specular Reflector) foils by 3M are used, providing a thin and radiation-tolerant solution.

For the large-scale production of BTL crystals, detailed technical specifications are defined to ensure uniform properties across all scintillating crystals and are reported in Table 5.1. The RMS spread of key parameters is required to be less than 5%. The light output, when crystals are coupled in air with a photosensor on one end, to SiPMs, has to exceed 4,000 photons/MeV, and the decay time of the scintillation process must have a value smaller than 45 ns. The packaged scintillator arrays consist of bars with 80 μm ESR foils placed between adjacent bars to better isolate each channel, with a reflectivity greater than 98.5% for light at 420 nm and good radiation tolerance. An additional reflective foil wraps the entire array with a thickness of about 0.2 mm, and a glue layer ensures stability and mechanical uniformity along the module. The pitch between adjacent bars is 3.2 mm, and the total size of a crystal array in the $\phi \times z$ plane is about 54.70 $\times$ 51.48 mm, with a flatness of the non-wrapped end faces better than 0.05 mm to ensure optimal coupling to the SiPM array.

LYSO:Ce crystal parameter	**Specification**
Density	> 7.05 g/cm^3
Light output (single end readout)	> 4000 photons/MeV
Decay time	< 45 ns
Time resolution (air coupling)	< 140 ps
Optical cross talk	$< 25\%$
Light output loss after irradiation	$< 20\%$
Array dimensions	
Length [mm]	54.7 ± 0.03
Width [mm]	51.38 ± 0.15
Thickness [mm]	4.12 ± 0.10
Surface polishing	< 15 nm

Table 5.1: Technical specifications of the BTL crystal arrays. The requirements for scintillation properties, radiation tolerance, and dimensions are reported.

Over the past years, the crystal properties from twelve vendors were measured by a dedicated team at INFN Rome to evaluate their suitability for mass production. Relevant results, which are useful for the discussion of the crystals, are reported below, however, these measurements were not performed as part of the present work and are detailed in Ref. [113]. Mechanical, optical, and scintillation properties of the samples were systematically assessed. Each manufacturer produces crystals with slight variations in the Yttrium and Cerium content, which affect specific material properties. The Yttrium concentration was evaluated through density measurements, obtained from mass and volume, and independently verified via Inductively Coupled Plasma Mass Spectroscopy (ICP-MS) at the Gran Sasso National Laboratory. The measured densities ranged between 7.0 and 7.4 g/cm^3, with a standard deviation below 1%, overall compatible with the requirement of being above 7.05 g/cm^3. The Yttrium content measured by ICP-MS is found to correlate linearly with the crystal density for each vendor. Although density is not a crucial parameter in characterizing timing performance, it remains an important parameter for the overall ma-

terial characterization. Factors such as crystal size, transmission, self-absorption, and photoluminescence play a crucial role [114] in determining the overall quality of a scintillator. Therefore, these parameters were thoroughly characterized. Transmission spectra were measured in the 300–800 nm range along the three crystal axes: longitudinal (L), width (w), and thickness (t). These measurements were performed at CERN and NIMP (Bucharest) using double-beam spectrophotometers. The results demonstrate good reproducibility, ensuring measurement consistency. The transmission threshold in the UV region, determined by the Cerium content, is shown in the left panel of Figure 5.1. The w and t spectra show the $5d^1$ absorbance band of Ce^{3+}. Cerium content can be estimated via optical absorption, defined as in Equation 5.6, where d is the transverse thickness and α is the absorption coefficient, which accounts for both Cerium and other absorption centers.

$$A \sim \alpha \cdot d \sim (\alpha_{Ce^{3+}} + \alpha_{other}) \tag{5.6}$$

Cerium content is modeled with a Gaussian function, while other absorption contributions are described empirically using the Urbach approximation (i.e. an exponential tail). Fitting the absorption data with this model allows estimating the Cerium content, which is found to have values in the range between 0.3 and 2.0 depending on the producer. The overall measurement reproducibility is estimated to be around 6%. Photoluminescence is measured in the 370-550 nm emission range, and the resulting spectrum shows

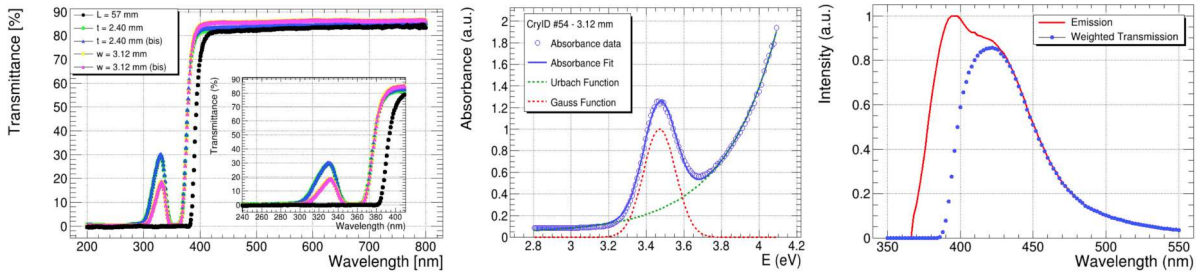

Figure 5.1: Optical properties of the BTL LYSO crystals are reported in terms of transmission, absorbance, and emission spectra. The left panel shows the optical transmission spectra measured along the three orientations of the crystal bar: one longitudinal (black) and two transversal directions (yellow/violet and green/blue), where two sets of measurements are presented. The middle panel displays the absorbance spectra with the overlaid fit and the individual components contributing to the total fit. The right panel shows the emission spectrum (red) together with the emission spectrum weighted by the transmittance (blue) [113].

the scintillator emission as a function of wavelength [115]. All vendors' crystals exhibited characteristic peaks at 420 nm and 396 nm. Since transmittance and absorption vary with wavelength, not all emitted light reaches the photosensor. To estimate the signal effectively detected, the photoluminescence spectrum is weighted by the transmittance, thereby accounting for the spectral sensitivity of the photosensor, which is a crucial information for an optimal light detection.

The light output and decay time of crystals from each manufacturer were measured in a dedicated setup, and results are averaged over 15 samples per producer. One end face of each crystal bar was placed in contact[2] with a PMT window, while the opposite end was left free in contact with air. The entire setup was kept in a controlled environment at a stable temperature of 20°C. The PMT single-photoelectron response was calibrated using a pulsed fast blue LED. Light output measurements are performed by evaluating the light produced by the interaction of annihilation photons emitted from a ^{22}Na radioactive source. By dividing the total charge detected from the crystals by the estimated single-photoelectron charge and the photon energy, the light output of the crystal bars is determined in terms of photoelectrons per MeV. Correcting this value by the quantum efficiency of the PMT, weighted over the LYSO emission spectrum, provides an estimate in terms of photons per MeV. For each vendor, the average over 15 samples is computed, producing values from 4,200 to 5,400 photons/MeV. The associated uncertainty of the order of 4% reflects the uniformity of performance among samples. The scintillation decay time is another crucial parameter, as it strongly impacts timing precision. It is extracted from the acquired

[2]The coupling was done in air, without any optical grease, in order to optimize reproducibility of the measurement.

waveforms using a fit function consisting of a single exponential decay and a Gaussian turn-on. Results are averaged across samples from each vendor, yielding decay times between 38 and 45 ns, with deviations of about 1% per vendor, consistent with the reproducibility of the measurement. A key figure of merit for assessing the timing capabilities of a crystal is the ratio between the light output and the decay time, which is therefore evaluated for each vendor. The results of these characterization measurements are summarized in Figure 5.2.

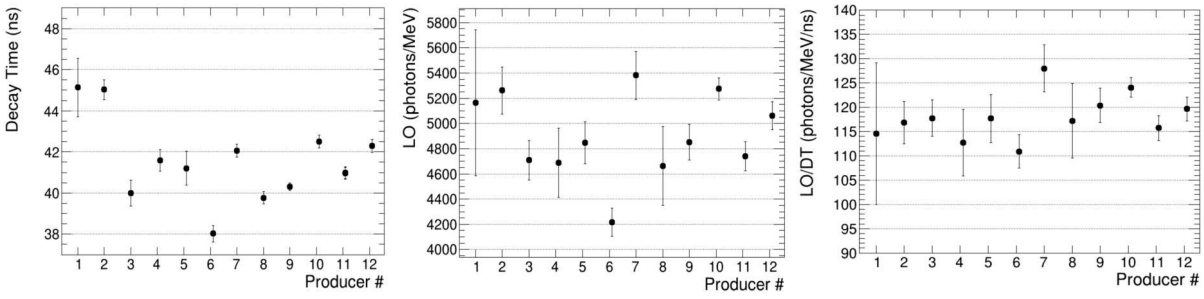

Figure 5.2: The light output and decay time of samples from each producer were evaluated and combined to characterize the timing performance of the crystals. From left to right, the decay time, light output, and the ratio between light output and decay time are shown for each vendor [113].

The BTL crystals will be exposed to high radiation fluences, leading to damage from electromagnetic energy deposition (γ-rays), neutrons, and charged hadrons [116]. Radiation-induced absorption can degrade optical transmittance and reduce the light yield, thereby affecting detector performance. In literature, studies [117] show that LYSO crystals exhibit a decreased light output and an unchanged scintillation decay time after γ irradiation. To assess radiation effects, several first BTL prototype samples were irradiated with 24 GeV protons to a fluence of 2×10^{13} cm^{-2}, which is above the integrated level expected for BTL. Results are reported in Figure 5.3 and show negligible losses in transmission and time resolution, consistent with expectations from hadron fluences and ionizing radiation. The transparency loss T along a crystal of length L is quantified by the induced absorption coefficient expressed in Equation 5.7 and it was measured as 0.5 m^{-1} at 420 nm.

$$\mu_{ind} = \frac{1}{L} \ln \left(\frac{T_{before}}{T_{after}} \right), \tag{5.7}$$

Moreover, the radiation hardness of BTL crystal samples against γ-rays was evaluated for several samples provided by the twelve LYSO manufacturers. Measurements were carried out at the ENEA-Casaccia Calliope facility [113] with an integrated dose of 50 kGy, corresponding to the expected BTL lifetime exposure with an additional 50% safety margin. After irradiation, average light output decreased by 9% ($\pm$3%), likely due to transparency loss from absorption centers, while scintillation decay time remained unchanged within 2% experimental uncertainty, confirming what expected from literature studies. Results are summarized in Figure 5.4.

Finally, since the BTL crystals will operate over a wide temperature range, it is also important to characterize their performance at low temperatures, particularly close to the operating point. The figure of merit was observed to increase consistently at lower temperatures, with variations between 3% and 8% depending on the crystal manufacturer [113].

Overall, the characterization studies demonstrated good performance for all samples and provided a solid basis for defining reliable quality assurance and quality control protocols in preparation for crystal mass production, which is discussed in Section 5.10.1.1. A single manufacturer was selected as the best bidder among a restricted list of manufacturers that were proven capable to satisfy the crystal specifications required by the BTL in terms of light output, decay time, company reliability, and production capacity: Sichuan Tianle Photonics.

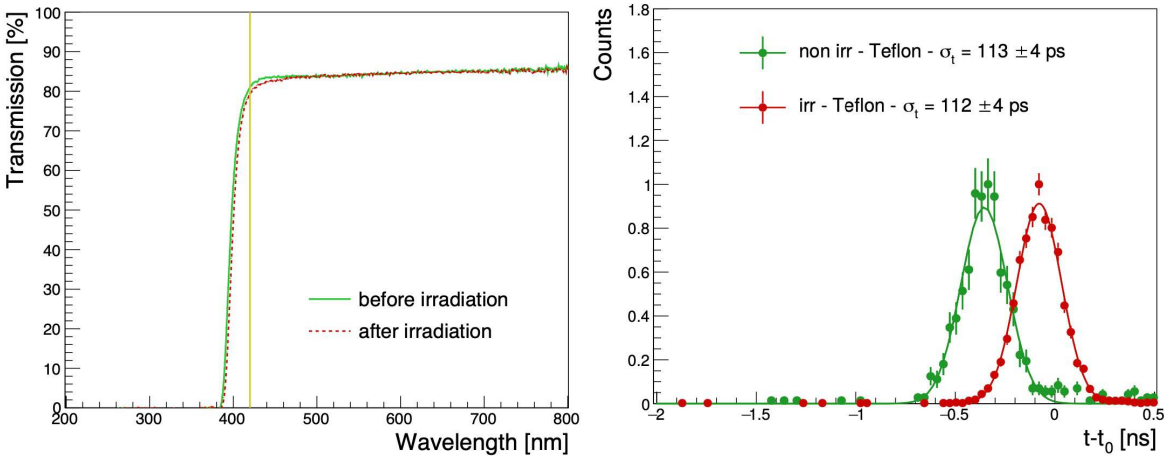

Figure 5.3: Preliminary studies with the first prototypes of LYSO:Ce crystal bars were conducted to investigate the effect of irradiation with 24 GeV protons up to a fluence of 2×10^{13} cm^{-2} on the bars' performance in terms of light yield and time resolution. Left: Transparency curves as a function of wavelength before and after irradiation are shown. Right: Time resolution measured with 511 keV γ-rays before and after irradiation is shown. The relative comparison of the values highlights the almost negligible effect of irradiation on the timing performance [93].

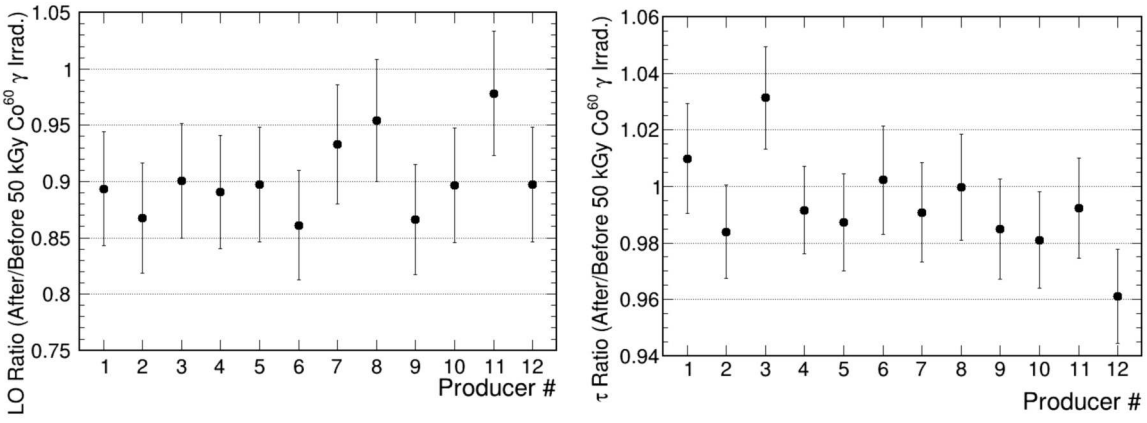

Figure 5.4: Samples were irradiated with an integrated dose of 50 kGy using γ-rays. Their performance in terms of light output and decay time was measured before and after irradiation, and the corresponding ratios are reported, with light output shown on the left and decay time on the right [113].

5.3 Optimization of SiPM performance

Each crystal end face is coupled to a Silicon Photomultiplier for detecting scintillation photons. Thanks to favorable characteristics such as low operating voltage, compactness, and robustness, SiPMs are widely used in fields such as optical spectroscopy, fluorescence detection, quantum physics, medical imaging, gamma spectroscopy, and high-energy physics. The SiPM detects the light produced by the scintillator through hundreds or thousands of single-photon avalanche diodes (SPADs), also referred to as microcells or cells, which are integrated on a common substrate and connected in parallel. The output signals from all fired SPADs are summed to form the total SiPM output signal. Thus, the SiPM operation fundamentally relies on the SPAD response: a single charge carrier entering the high-electric-field region initiates a self-sustaining avalanche, generating multiple electron–hole (e-h) pairs until the avalanche is quenched. When reverse biased, the p-side of the SPAD is fully depleted, creating an electric field. The high-field region serves as the multiplication zone, while the remaining region acts as a drift region. The SPAD operates in Geiger mode when biased above the breakdown voltage V_{br}, at which self-sustaining avalanche multiplication occurs. Below V_{br}, two regimes exist: the photodiode regime, where no multiplication occurs and current is proportional to incident light, and the avalanche photodiode regime, where moderate avalanche multiplication occurs, providing a gain of $\mathcal{O}(10–100)$. Above V_{br}, both electrons and holes can generate secondary e-h pairs, and the avalanche continues until the quenching resistor stops the avalanche and restores the diode (Figure 5.5).

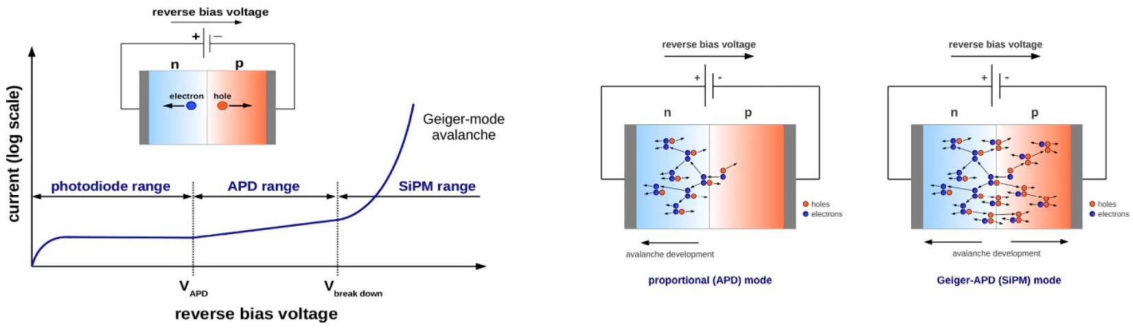

Figure 5.5: A p–n junction can operate under different reverse-bias conditions. The left plot shows the current as a function of the reverse bias voltage, illustrating the different multiplication regimes for photodiodes, APDs, and SiPMs. The right diagram shows the charge multiplication processes under APD and SiPM operating modes: in the APD, only electrons undergo avalanche multiplication, while in the SiPM, both electrons and holes contribute to ionization [118].

The equivalent circuit model of the SiPM can be considered for understanding its functioning, and it is shown in Figure 5.6. Each SPAD is represented as a diode capacitance C_d in series with an internal resistance R_d, and a quenching resistor R_q in parallel with a parasitic capacitance C_q. When a photon triggers an avalanche, C_d discharges while C_q recharges, generating the output signal. The avalanche stops when the current through R_d reaches a minimum, defining the cell recovery time. The figure illustrates the scenario in which N_f cells are fired out of a total N_{tot}. Each of the electrical parameters described can be measured through IV and impedance measurements [118, 119]. The resulting signal shape is influenced by both fast and slow components, determined by the quenching capacitance and resistance of each cell, respectively.
SPADs close to the fired one can lead to correlated noise, known as crosstalk. Prompt crosstalk occurs when secondary photons trigger neighboring SPADs within a few hundred ps, inflating counts. Delayed crosstalk originates from carriers in non-depleted regions, producing avalanches with delays of hundreds of ns. External crosstalk arises from photons reflected back by protective layers. These effects are mitigated through sensor design, e.g. by reducing the depletion region or protective layer thickness.

Individual SPAD signals can be read out separately, either through dedicated electronics or by measuring the amplitude of the summed SPAD signals [118]. This provides single-photon sensitivity and fast avalanche amplification, making SiPMs highly suitable for timing applications. The number of secondary

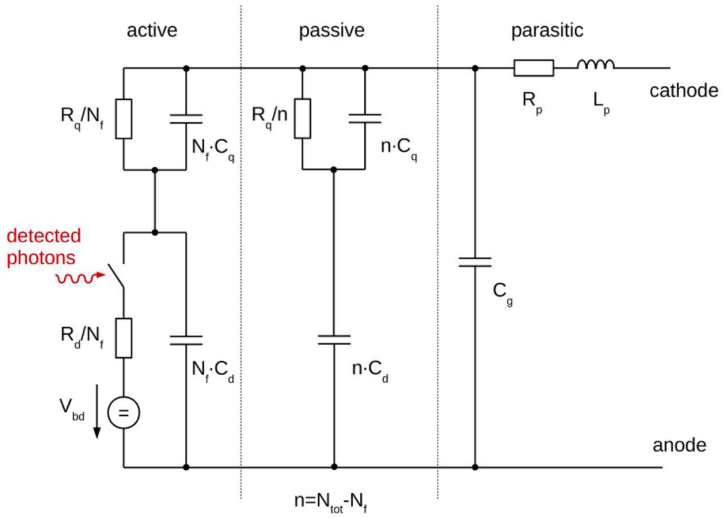

Figure 5.6: The equivalent electrical circuit of a SiPM is shown. The impinging photon acts as a switch in the circuit. Each diode has a resistance R_d and a capacitance C_d, along with a quenching resistor R_q and its associated capacitance C_q. The fired cell is represented as active (on the left), while the non-triggered cells are represented as passive components in parallel, together with parasitic elements. N_f denotes the number of fired cells, and N_{tot} is the total number of cells [118].

electron-hole pairs produced per primary carrier determines the multiplication factor, which diverges above the breakdown voltage V_{br} and depends on the diode structure and temperature. This operation mode provides a high gain (G) on the order of 10^5–10^7, expressed as in Equation 5.8 where q is the elementary charge, C_q and C_d are the quenching capacitance and diode capacitance (see Figure 5.6) and V_{ov} is the overvoltage, which is expressed in Equation 5.9.

$$G = \frac{avalanche\ charge}{q} = \frac{V_{ov} \cdot (C_q + C_d)}{q} \tag{5.8}$$

$$V_{ov} = V_{bias} - V_{br}. \tag{5.9}$$

SiPMs must be biased above V_{br} to enable carrier multiplication. The breakdown voltage is estimated from the IV curve by computing the derivative of the current with respect to voltage, and its value is temperature dependent. SiPMs operate at relatively low bias voltages (typically 22–80 V), allowing limited power consumption. Thermal generation of carriers can produce dark count events that produce a primary Dark Count Rate (DCR), modeled using Shockley–Read–Hall statistics. DCR decreases by a factor of roughly 2 for every 10°C decrease in operating temperature and depends on the active area and overvoltage. Thermally generated electrons trigger avalanches identical to photon signals, representing noise, which becomes significant under radiation exposure. Charge trapping and delayed release can generate afterpulses, with amplitudes smaller than primary events. The probability of afterpulsing depends on trap density, release times, and SPAD recharge time.

The SiPM performance can be assessed in terms of different parameters, including Excess Charge Factor (ECF), PDE, and gain [120]. The ECF accounts for additional charge due to crosstalk and afterpulsing and it is expressed in Equation 5.10.

$$ECF = \frac{\langle Q \rangle}{\langle Q_N \rangle} \tag{5.10}$$

PDE represents the probability that an incident photon is detected and can be expressed as in Equation 5.11, where QE is the quantum efficiency, P_T the avalanche probability, and FF_{eff} the effective fill factor.

$$PDE(V_{ov}, \lambda) = QE(\lambda) \cdot P_T(V_{ov}, \lambda) \cdot FF_{eff}(V_{ov}, \lambda) \tag{5.11}$$

The SiPM gain provides information regarding the multiplication factor of primary carriers and it depends on the overvoltage applied and on the size of the individual cells. All the discussed parameters are

fundamental for evaluating SiPM performance and optimizing the signal pulse shape. Extensive studies were carried out after the TDR publication to select the optimal SiPM design for BTL, achieving a timing resolution of 30–60 ps under the expected radiation fluence of 2×10^{14} $n_{eq}/cm^2$and a dose up to 50 kGy. A detailed tuning of the electrical parameters was conducted, and the final design of the BTL SiPMs is summarized in Table 5.2. The active area is matched to the LYSO crystal, with an end face of 2.91×3.80 mm^2, aligned with the crystal center. Each SiPM is covered by a silicone resin layer with a thickness of 300 μm, which is the minimum achievable, and a refractive index of approximately 1.5 at 420 nm, serving as a protective window. The resin is made of radiation-tolerant material, with less than 5% transparency loss (crucial for maintaining optimal LCE), making it preferable to epoxy. The cell dimensions are 25×25 μm^2, providing an optimal balance between PDE, gain, and DCR (extensive discussion on these parameters is presented in Section 5.6.1). The sensors are optimized to achieve a PDE greater than 50% and a gain of the order of 10^6 at $V_{ov} = 3$ V. It should also be noted that both the PDE and the gain have been shown to be affected by radiation damage, resulting in lower values. For BTL SiPMs, the expected losses in PDE and gain at the end-of-life operating conditions are approximately $\mathcal{O}(15\%)$ and $\mathcal{O}(5\%)$, respectively [121].

Each of the 16 crystal bars is coupled to a SiPM, forming a SiPM array. They are mounted on a 0.7 mm-thick custom substrate that combines PCB and AlN materials, ensuring temperature uniformity (less than 0.5°C variation) across the package[3] and providing efficient heat dissipation from the SiPMs, with a thermal resistance below 7 W/K per SiPM. This substantially reduces self-heating effects due to increased current after radiation damage. The SiPM substrate was designed to ensure a uniform temperature across the SiPM array, resulting in a time spread due to temperature variations smaller than 5 ps. The package dimensions are 51.5×6.5 mm^2, with an overall thickness below 1.6 mm, including the protective resin. Arrays include integrated low-inductance flex cables for SiPM biasing and signal routing to the front-end electronics where the TOFHIR ASIC is integrated. The flex cables are designed with low inductance (less than 4 nH/SiPM channel), are compatible with 0.5 mm pitch 40 ch Panasonic AYF534035, and operate over a temperature range from -45°C to +60°C. The flex cable thickness is less than 200 μm, excluding the stiffener, which provides additional robustness, and they are designed for a bending radius of 2.0 mm. Four Thermoelectric Coolers (TECs) with dimensions of 3×4 mm^2, 0.95 mm thick, are soldered on the back of each array, ensuring optimal heat extraction. TECs are reflow-soldered at 220°C to the SiPM package, transferring heat to the copper housing and cooling plates (this thermal coupling required thoughtful evaluation and it is discussed in Section 5.9). TECs are biased in series via the SiPM flex traces and demonstrate mechanical reliability under thermal cycling (-50°C to +60°C). Additionally, each sensor is monitored with a radiation tolerant Resistive Temperature Detector (0805 RTD PT1000), which is is placed on the rear side of the SiPM array package to provide precise temperature measurement. A picture showing both sides of the SiPM array is reported in Figure 5.7.

5.3.1 Mitigation of radiation damage

As discussed for crystals, it is crucial to evaluate the radiation tolerance of BTL SiPMs, given the expected fluences to which they will be exposed. SiPMs, being sensitive to single charge carriers, are particularly sensitive to spurious counts induced by radiation damage. Radiation introduces two main types of defects in silicon sensors: surface damage, due to Ionizing Energy Loss (IEL), and bulk damage, caused by Non-Ionizing Energy Loss (NIEL) [122]. IEL defects arise from charges generated and trapped in the silicon dioxide (SiO_2) layer or at its interface with the silicon substrate, while NIEL defects occur when energetic particles (neutrons, protons, pions) transfer an amount of energy to atom larger than the binding energy of a silicon atom, causing the displacement of silicon atoms from the lattice.

NIEL defects manifest as vacancies (missing atoms) and interstitials (atoms displaced within the lattice). Interstitials and vacancies may diffuse and annihilate at regular lattice sites, partially mitigating the damage. This process can be enhanced by heating, known as annealing. Alternatively, defects may combine with impurities, forming stable defects that alter the silicon doping. Radiation-induced displacement damage is proportional to the energy imparted to the lattice. A hardness factor κ allows comparison of damage from different particles and energy spectra $\Phi(E)$ [123]. To provide a common comparison among effects caused by different particles and energies, it is assumed that radiation damage

[3] A 1°C increase in operating temperature has been observed to raise V_{br} by several tens of mV.

Figure 5.7: A picture of both sides of the SiPM array is shown: the back side, where the four TECs and the RTD are visible, and the front side, showing the active area of the 16 SiPMs at the bottom.

SiPM characteristic at 20°C	**Value**
Cell size	25 x 25 μm^2
Operating T	-50 to +70 °C
Breakdown Voltage	$34 < V_{br} < 41$ V
Max total V_{br} spread across arrays	< 2 V
V_{br} spread within 16 ch array	< 150 mV
dV_{br}/dT	< 40 mV/°C
Dark current at 3V V_{ov}	< 50 nA/mm^2
PDE at 420 nm (3V)	$> 50\%$
Gain (3V)	$0.8 \times 10^6 < \text{Gain} < 1 \times 10^6$
ECF (3V) after pulsing	< 1.06
Afterpulsing	$< 5\%$
Capacitance	< 75 pF/mm^2
R_q/N	$20 < R_q/\text{N} < 25\ \Omega$
Array dimensions	
Length [mm]	51.5 ± 0.1 mm
Height [mm]	6.5 ± 0.1 mm
Thickness (with resin, w/o TECs)	1.4 ± 0.1 mm
Thickness (with resin and TECs)	2.45 ± 0.1 mm
Max resin thickness	0.35 ± 0.05 mm

Table 5.2: Technical specifications for non-irradiated BTL SiPMs are provided for various parameters, including both mechanical and performance characteristics, for individual SiPMs as well as for the SiPM arrays.

scales with the NIEL factors (NIEL scaling hypothesis), with the reference defined as 1 MeV neutrons. Under this assumption, every fluence is converted to the equivalent 1 MeV neutron fluence, reported as n_{eq}/cm^2. An example of different NIEL factors in silicon for neutrons, protons, pions, and electrons is shown in the left panel of Figure 5.8: protons produce significant displacement damage due to their high energy deposit per unit length, and above 50 MeV the NIEL values of neutrons and protons become comparable. Macroscopically, radiation defects introduce energy levels within the silicon bandgap ($E_G \sim$ 1.12 eV). Mid-gap levels increase leakage current through enhanced thermal carrier generation, while levels near the band edges act as trapping centers, reducing signal formation. Radiation can also modify the effective doping, thereby changing the depletion and breakdown voltages. Nuclear interactions may

remove dopant atoms, further affecting device operation. The right panel of Figure 5.8 summarizes these macroscopic effects. The increase in leakage current is proportional to fluence up to $\sim 5 \times 10^{13}$ n_{eq}/cm^2,

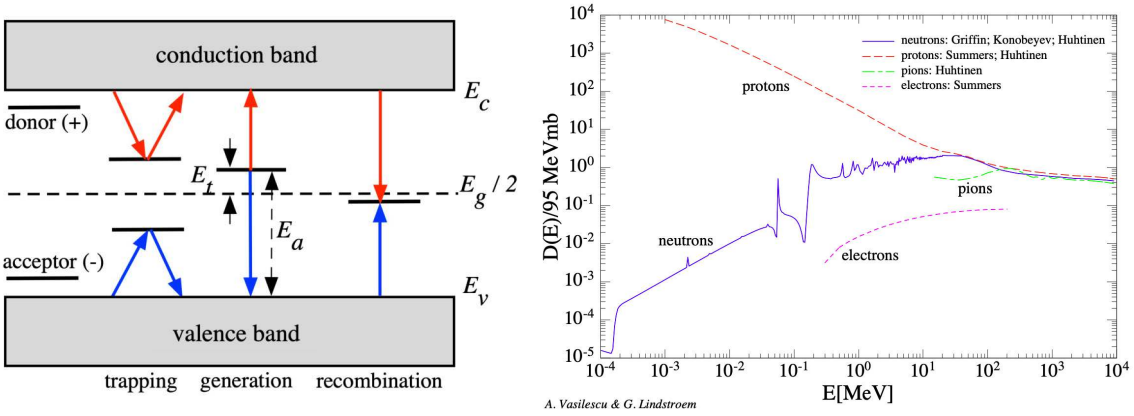

Figure 5.8: Left: Pictorial representation of possible defects caused by radiation damage in a silicon sensor. The displacement of electrons (red lines) and holes (blue lines) is illustrated [122]. Right: The displacement damage cross section in silicon, normalized to the value for 1 MeV neutrons, is shown for different particles as a function of their energy [124].

with a radiation damage coefficient α defined as in Equation 5.12 [122], where ΔI is the increase in current due to irradiation, $A \cdot w$ is the depleted volume, and Φ_{eq} is the 1 MeV neutron equivalent fluence.

$$\alpha = \frac{\Delta I}{A \cdot w \cdot \Phi_{eq}} \tag{5.12}$$

The total leakage current I produced by radiation damage depends on temperature: at high temperatures, thermal generation dominates with Arrhenius behavior (slope ~ 0.6 eV), while at low temperatures (i.e. temperatures smaller than $\mathcal{O}(100\ \mathrm{K})$) tunneling through defects becomes significant, and DCR becomes nearly temperature-independent. The DCR arises from spurious charge generated by radiation effects within the sensors and can be computed from the total dark current I, the SiPM gain G, and the elementary charge e as in Equation 5.13:

$$DCR = \frac{I}{G \cdot e}. \tag{5.13}$$

IEL effects also contribute to the increase of the total dark current by creating additional defects induced by photons and low-energy charged particles (<300 keV) that affect the SiPM surface. Radiation-induced DCR increases with the integrated luminosity, thereby degrading timing performance. Part of this DCR can be mitigated through annealing, as elevated temperatures accelerate defect recombination in the silicon lattice. For instance, annealing at 40°C for 100 hours halves the DCR, whereas at 15°C the same reduction requires approximately ten times longer. Detailed simulation models [123, 122] successfully describe DCR evolution under irradiation and thermal annealing with an accuracy of ∼10%.

In the BTL context, radiation effects impact timing by introducing additional noise and reducing signal amplitude. The NIEL scaling hypothesis has been verified for BTL SiPMs with 24 GeV protons, confirming that proton fluences can be expressed as equivalent neutron fluences, hereafter indicated simply as n_{eq}/cm^2 [122]. In BTL SiPMs, radiation exposure and the consequent increase in DCR lead to a rise in power consumption (up to 30 mW per SiPM channel) and may induce self-heating, which can degrade performance. Since DCR strongly depends on operating temperature, careful management of this parameter provides an effective means to mitigate time resolution degradation. SiPMs will be operated at low temperatures and will require close monitoring to ensure proper DCR control. To further reduce DCR increases due to radiation exposure, Thermoelectric Coolers, shown in Figure 5.9, are integrated into the SiPM package design. TECs enable both a reduction of the operating temperature and an increase of the annealing temperature. Indeed TECs will lower the SiPM temperature to -45°C during operation, reducing the DCR, and can raise it to ~ 60°C for in situ annealing [125] during technical stops. Thermal

annealing is expected to reduce the DCR, thereby mitigating the detrimental effects of radiation damage on timing performance.

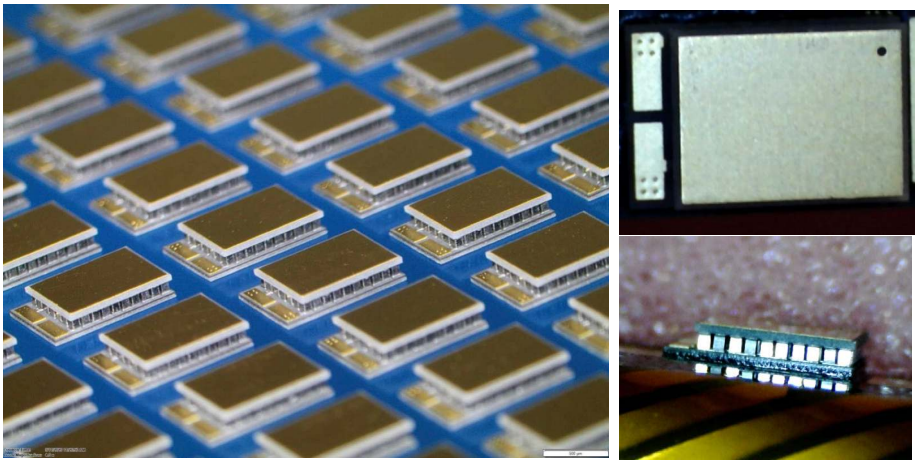

Figure 5.9: Left: Picture of a batch of TECs [125]. Right: Microscope images of a TEC installed on the SiPM packaging, shown in frontal (top) and lateral (bottom) views. In the frontal view, the TEC area and the two electrical contacts are clearly visible, while in the lateral view the TEC pillars can be seen.

As TECs are mounted directly on the SiPM package, they are exposed to the same radiation dose as the sensors themselves. TECs have been demonstrated to be radiation tolerant up to the projected BTL irradiation level of 2×10^{14} $\mathrm{n_{eq}/cm^2}$. Additionally, the impact of the use of TECs on the DCR was systematically studied and detailed report is given in Ref. [125], under different annealing and operating temperature scenarios. DCR as a function of temperature is described by Equation 5.14, where E_g is the silicon bandgap and Γ accounts for tunneling effects.

$$\frac{DCR}{DCR(-30°C)} = N(1+\Gamma)T^2 \exp\left\{-\frac{E_g/2+\Delta}{kT}\right\} \tag{5.14}$$

Figure 5.10 shows, in the left panel, the dark current measured for irradiated SiPMs subjected to different annealing scenarios, and, in the right panel, the ratio of the measured DCR to the DCR measured at 1 V overvoltage, with the data fitted using the model described in Equation 5.14. The dual benefit of reducing DCR and enabling further recovery via high-temperature annealing has been evaluated under various temperature scenarios, as shown in Figure 5.11. The DCR of a SiPM biased at 1 V OV is modeled as a function of BTL operation years during HL-LHC for different operating and annealing temperatures, showing significant performance improvement at low operating temperatures and with high-temperature annealing. These studies led to the choice of operating SiPMs at $T_{op} = -45$°C and performing periodic in situ annealing at $T_a = 60$°C, which was also demonstrated to be within power budget.

To reduce the contribution to the time resolution degradation caused by the DCR generated in SiPMs due to radiation exposure, several strategies are envisaged for the BTL. One is the discussed smart thermal management provided by TECs: during HL-LHC technical stops, the BTL cold plate temperature can be increased from -35°C to 10°C, while TECs operate in reverse to create a temperature gradient up to 50°C, reaching 60°C at the SiPM surface. This annealing reduces DCR and improves timing performance. Additionally, to mitigate higher DCR, the SiPM overvoltage will be reduced from 3.5 V at the beginning of operation (non-irradiated sensors) to 1 V at the end of operation (sensors irradiated to full fluence). The decrease of SiPM overvoltage is necessary both to mitigate the increase of DCR with overvoltage and to limit the large power dissipation that would otherwise drive the SiPM into a self-heating regime, significantly degrading performance due to the DCR contribution. Also, baseline fluctuations due to increased DCR are partially corrected by dedicated ASIC circuits (see Section 5.8).

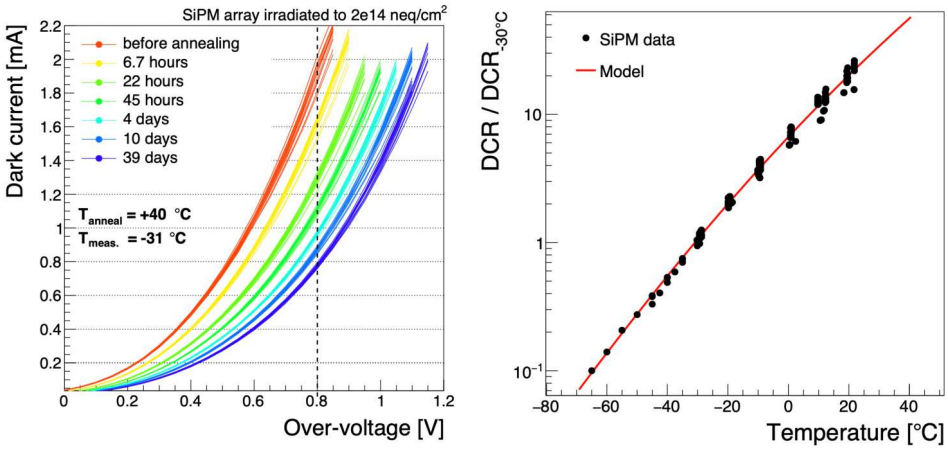

Figure 5.10: Left: The dark current is shown as a function of overvoltage for SiPMs irradiated to 2×10^{14} n_{eq}/cm^2, measured at a temperature of -31 °Cand annealed at 40 °Cfor different periods of time, corresponding to the different colors in the plot. Right: The ratio of the DCR measured at 1 V overvoltage at a given temperature to the DCR measured at -30 °Cis shown as a function of temperature. The data values are fitted using Equation 5.14 [125].

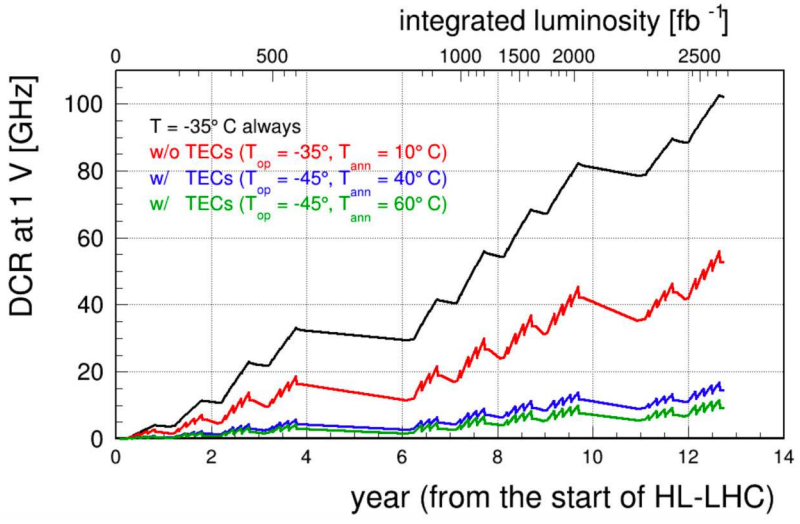

Figure 5.11: The evolution of the DCR measured at 1 V overvoltage during BTL operation is shown as a function of integrated luminosity and years since the start of the HL-LHC. Different lines correspond to different scenarios, evaluating various operating temperatures (T_{op}) and annealing conditions (T_{ann}) [125].

5.3.2 Signal pulse shape enhancement

The produced SiPM signal is critical for the overall performance, given that the steepness of the current pulse (dI/dt) critically affects the electronics term of the time resolution, as expressed in Equation 5.1. This work contributed to detailed studies of the SiPM signal pulse shapes and electrical parameters, which specifically assessed the impact of different cell sizes on the signal shape. While the original choice was oriented towards a 15 μm pixel size, which seemed to represent an optimal balance between a reduced DCR and a decent PDE, a thorough study was then dedicated to evaluating the impact of other dimensions on the SiPM pulse shape. The results demonstrate better performance with SiPMs having a cell size of 25 μm. The effects of cell size on time resolution were evaluated through both modeling and laboratory measurements of SiPMs with cell sizes of 15, 20, 25, and 30 μm. Samples from two different vendors were evaluated: Hamamatsu Photonics (HPK) and Fondazione Bruno Kessler (FBK). The two manufacturers employed different wafer technologies, which bring different SiPM properties. As HPK sensors demonstrated superior performance across the range of radiation scenarios expected at the HL-LHC, as widely discussed in Section 5.6, the final choice of vendor for the BTL was HPK.

To precisely evaluate the functioning of the SiPM, an electrical model of SiPM operation in the BTL context was used. As mentioned before, the SiPM is made of many SPADs in parallel (see Figure 5.6), producing together a current that is the sum of the single microcells fired. Each cell operates in Geiger mode producing a gain that depends on the applied overvoltage V_{ov} and on the SiPM capacitances, as in Equation 5.8. Several studies in the current literature have already explored the definition of an electrical model of SiPM functioning, such as in Refs. [126, 127, 128, 129], among others. The solution obtained in Ref.[129] was used for the following studies and it is reported in Equation 5.15.

$$I_L(t) = Ge\left[a_1 \exp\left(-\frac{t}{\tau_{cd1}}\right) + a_2 \exp\left(-\frac{t}{\tau_{cd2}}\right) + a_3 \exp\left(-\frac{t}{\tau_{md}}\right) + a_4 \exp\left(-\frac{t}{\tau_{mr}}\right)\right], \tag{5.15}$$

with

$$\tau_{cd1} = \frac{\tau_1+\tau_3+\tau_4}{2} + \frac{\sqrt{(\tau_1+\tau_3+\tau_4)^2 - 4(\tau_1\tau_4+\tau_2\tau_3)}}{2} \qquad \tau_1 = R_q(C_d+C_q)$$

$$\tau_{cd2} = \frac{\tau_1+\tau_3+\tau_4}{2} - \frac{\sqrt{(\tau_1+\tau_3+\tau_4)^2 - 4(\tau_1\tau_4+\tau_2\tau_3)}}{2} \qquad \tau_2 = R_qC_q$$

$$\tau_{md} = C_d\frac{R_dR_q}{R_d+R_q} \qquad \tau_3 = N_cR_LC_d$$

$$\tau_{mr} = R_d(C_d+C_q) \qquad \tau_4 = R_LC_g$$

$$a_1 = \frac{\tau_{cd1}(\tau_{cd1}-\tau_2)}{(\tau_{cd1}-\tau_{cd2})(\tau_{cd1}-\tau_{md})(\tau_{cd1}-\tau_{mr})}$$

$$a_2 = \frac{\tau_{cd2}(\tau_{cd2}-\tau_2)}{(\tau_{cd2}-\tau_{cd1})(\tau_{cd2}-\tau_{md})(\tau_{cd2}-\tau_{mr})}$$

$$a_3 = \frac{\tau_{md}(\tau_{md}-\tau_2)}{(\tau_{md}-\tau_{cd1})(\tau_{md}-\tau_{cd2})(\tau_{md}-\tau_{mr})}$$

$$a_4 = \frac{\tau_{mr}(\tau_{mr}-\tau_2)}{(\tau_{mr}-\tau_{cd1})(\tau_{mr}-\tau_{cd2})(\tau_{mr}-\tau_{md})}$$

The analytical expression reported in Equation 5.15 describes the current response of a single cell in a SiPM following the detection of a single photon. In this model, the cell is represented as a current source in parallel with its intrinsic diode capacitance C_d, while the quenching resistor R_q and the capacitance C_q model the avalanche quenching and cell recovery dynamics. The remaining cell connected in parallel, along with the readout resistance R_L and the parasitic capacitance C_g, contribute additional circuit time constants, leading to a response that is a superposition of four exponentially decaying terms. Each time constant τ_{cd1}, τ_{cd2}, τ_{md}, and τ_{mr} arises from a combination of the cell RC elements and the readout

circuitry, and the corresponding coefficients a_1–a_4 ensure that the total integrated current equals the total charge delivered by the avalanche, $Q = Ge$. This expression includes both the fast initial discharge of the avalanche and the slower recovery of the cell, providing a description of the single-photoelectron response that can be used to model the SiPM output for multiple incident photons by linear superposition.

The described SiPM electrical model was validated through measurements conducted in laboratory. A UV laser was shined towards naked SiPMs that were readout with an AdvanSiD board [130]. A variety of different 15 μm SiPMs were measured, with different active areas and, in general, different parameters such as grid capacitance, diode capacitance, quenching capacitance, and quenching resistance. The expected pulse shape is computed from the model inserting the different parameters of the SiPM. The measured pulse shapes were found to be consistent with the formulated model, for the varying parameters, as shown in Figure 5.12.

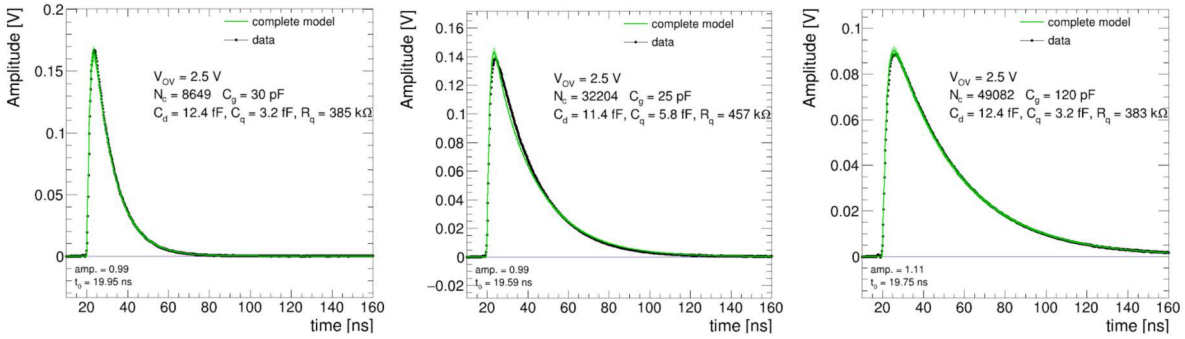

Figure 5.12: Pulse shapes obtained from laboratory measurements (black dots) and from model calculations (green line) are compared for SiPMs featuring different electrical parameters. Examples of three SiPMs are shown, with the corresponding parameters reported in the figure, measured at an overvoltage of 2.5 V. An overall good agreement is observed between data and expectations.

These studies highlighted the potential for an optimized SiPM design. The quenching capacitance depends on the resistive paths and therefore on the cell perimeter, while the grid capacitance can be expressed as a function of the cell area. The overall gain depends on the sum of these capacitances, leading to a strong dependence of the device performance on the microcell size. Increasing the cell dimension is expected to enhance the gain and steepness of the SiPM response, thereby reducing the contribution of the electronics to the time resolution. Furthermore, a higher signal amplitude allows for a lower discriminator threshold in terms of the number of photoelectrons produced.[4] A larger signal amplitude also improves the fill factor: for a given total area, increasing the cell size reduces the number of cells, decreasing the dead area and thus increasing the photon detection efficiency. To estimate the expected signal pulse shapes for different SiPM cell sizes, the expected SiPM parameters were inserted into the model and the resulting signals, shown in Figure 5.13, exhibit a steeper pulse slope for increasing SiPM cell sizes.

These studies motivated the production of SiPMs with various SiPM cell sizes. Properties of SiPM with different cell sizes in terms of gain and effective PDE (i.e. the measured PDE weighted by the emission spectrum of LYSO:Ce crystals) were measured as a function of overvoltage for the two vendors, as reported in Figures 5.14. A significant improvement in both PDE and gain is observed for larger cell sizes compared to the nominal 15 μm pixels. While these effects are beneficial, increased capacitances also slow down the signals and induce a higher DCR. For a given DCR level, larger cells result in a higher fraction of cells being busy due to avalanches triggered by dark counts, making them unavailable for detecting real photons and effectively reducing the PDE. Additionally, higher gain requires increased power consumption and enhances SiPM self-heating. The overall balance of these effects is non-trivial and requires a comprehensive evaluation of the final timing performance, as further discussed in Section 5.5.

[4]The choice of the discriminator threshold involves a trade-off between the main contributions to the time resolution. At low thresholds, stochastic fluctuations in the photon arrival time dominate, while at high thresholds, the impact of electronic noise is reduced. Therefore, an optimal threshold must be determined to balance these effects.

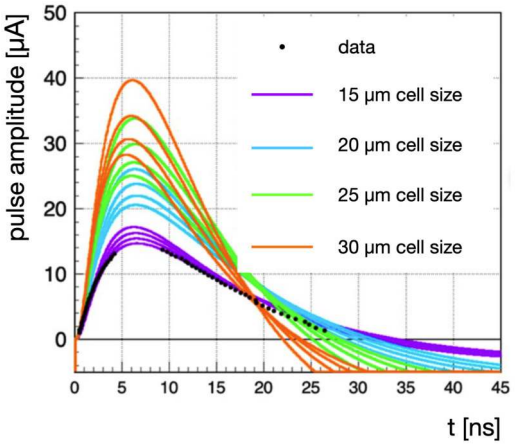

Figure 5.13: The pulse shapes produced with the described model are shown in the figure, with different colors corresponding to different cell sizes: purple for 15 μm, light blue for 20 μm, green for 25 μm, and orange for 30 μm. For each cell size, multiple curves are displayed corresponding to different values of the quenching resistor (200, 300, 400, and 500 Ω), with lower resistor values resulting in steeper pulse slopes. Measured pulse shapes for SiPMs with 15 μm cell size are also shown and are in good agreement with the modeled curves.

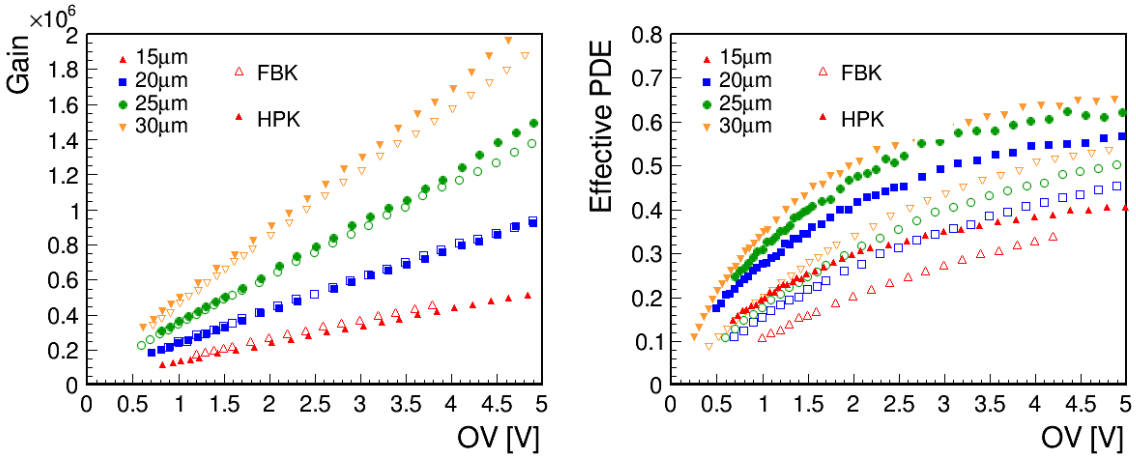

Figure 5.14: Measured SiPM gain and PDE values as a function of overvoltage for both vendors and different cell sizes. The PDE values are reported as effective PDE, i.e. the measured PDE weighted by the emission spectrum of LYSO:Ce crystals [96].

5.4 Factors affecting the light output at sensor module level

Crystal and SiPMs with the defined technical specifications are coupled to form a sensor module. Each BTL sensor module consists of a LYSO matrix of 16×1 crystal bars coupled at both ends to SiPM arrays of 16 SiPMs. Coupling LYSO crystals and SiPMs is particularly critical because it must maximize the LCE. Many studies were carried out on different aspects before identifying the best practice. An aspect affecting the LCE is, for instance, the glue layer used between the crystals and the wrapping. Studies of different approaches show that a reduced amount of glue, schematically represented in Figure 5.15, enhances the LCE, with an impact on the LO of about 10%. The amount of glue used to couple LYSO

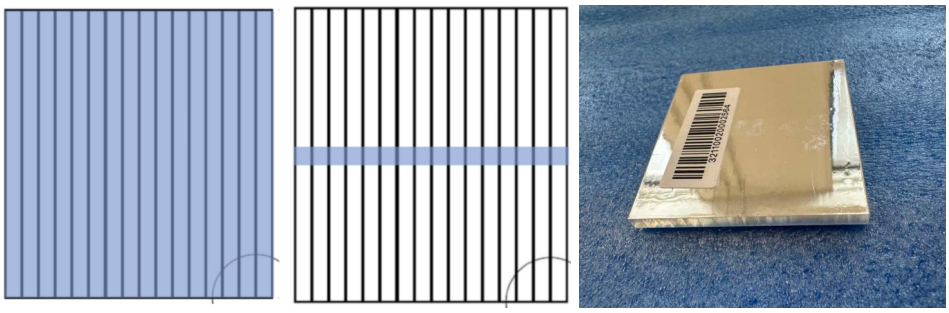

Figure 5.15: Left: Sketch of two configurations in terms of glue coverage between crystals and wrapping, with the full-coverage configuration on the left and a reduced-glue configuration in the middle. Right: Picture of a production BTL LYSO:Ce array consisting of 16 crystal bars.

crystals to SiPMs also affects the LCE: when photons cross an interface between different media, some can be reflected and reach a SiPM corresponding to neighboring bars. Even in the absence of reflections, photons exit the crystal with a wide angular distribution, so that the farther the SiPM from the crystal, the more photons (including direct ones) may reach adjacent SiPMs due to geometric effects. This phenomenon increases cross talk and reduces the light output, requiring minimization of glue thickness while maintaining enough coverage to avoid air gaps. Thus, SiPM arrays are glued to the LYSO matrix with a 100 μm layer of RTV3145 glue, a material already used for coupling CMS ECAL crystals to photodetectors. This glue provides good optical matching to the SiPM protective window, it is radiation tolerant to the levels expected in the BTL, and can operate at low temperatures. Dedicated assembly procedures and QA/QC protocols were defined for large-scale production and are discussed in Section 5.10.2.

Another important parameter affecting the overall amount of photoelectrons produced is the thickness of the sensor module. The design proposed in the MTD TDR [93] envisaged three thicknesses along the pseudorapidity direction to maintain an approximately constant slant thickness crossed by particles: type 1 (T1) modules with 3.75 mm thickness at low pseudorapidity, type 2 (T2) with 3.00 mm thickness at medium pseudorapidity, and type 3 (T3) with 2.4 mm thickness at high pseudorapidity. The sensor module thickness determines both the SiPM active area and the LYSO crystal bar end-face dimensions. To evaluate the impact of the sensor module thickness on performance, light output was measured in laboratory tests, as documented in Ref. [131]. Results demonstrate an increase in performance for thicker crystals: the LCE of the thickest sensors is up to 15% higher than that of the thinnest ones, as shown in Figure 5.16. Since scintillation photons are emitted isotropically inside the crystal volume, crystals with a larger end face allow a larger solid angle of photons to reach the SiPM without reflections. Additionally, the average number of reflections at the crystal lateral surface is smaller, increasing the probability of detection. This effect provides an additional boost of the light output and it is expected to further improve the timing performance. In addition to an improved LCE, the N_{pe} is also higher for thicker crystals, due to the larger energy deposit resulting from the greater slant thickness traversed by the particles. This comes at the cost of a higher DCR associated with larger SiPMs, and a slightly slower pulse shape. The combined impact of these effects on time performance therefore required dedicated measurements on beam, which are reported in Section 5.6.2.

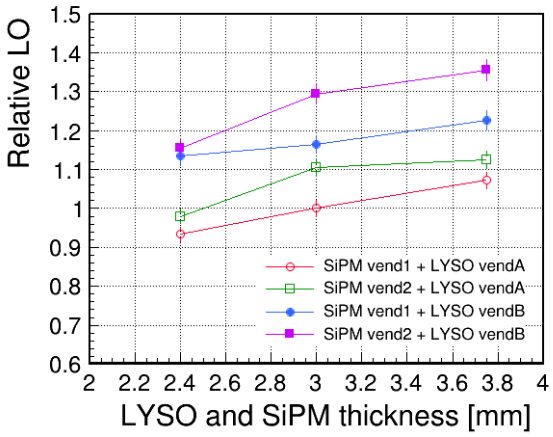

Figure 5.16: The light output values for modules consisting of different combinations of SiPM and LYSO manufacturers are shown in different colors as a function of the module thickness. Each light output value is normalized to the value measured for a specific type 2 module [131].

5.5 Optimization of sensor parameters and performance confirmation with test beam

The studies on sensor specifications discussed in the previous Sections motivated the evaluation of the performance of sensors with different features in beam tests, in order to obtain results under conditions as close as possible to those of the final detector. Specifically, the impact of the sensor module thickness and the SiPM cell size on the timing performance was evaluated under different operating conditions expected in the BTL. Modules with final specifications were subsequently measured under varying conditions in terms of DCR level and energy deposition. Their performance was evaluated for each configuration, also assessing the expected time resolution contributions.

Both non-irradiated and irradiated modules were measured in these studies, in order to evaluate the performance across the entire BTL lifetime at HL-LHC. Sensors will be exposed to different levels of radiation during BTL operation, up to an expected fluence of 2×10^{14} n_{eq}/cm^2 after 3000 fb^{-1}. The radiation damage caused by neutrons and charged hadrons at the HL-LHC, under the NIEL hypothesis, can be expressed in terms of 1 MeV n_{eq}/cm^2 and is estimated via simulations. Irradiation of SiPMs[5] was conducted at the JSI neutron reactor in Ljubljana up to the expected total fluence in BTL at the end of operation. Three different irradiation levels were produced: SiPMs were irradiated to 1×10^{13} n_{eq}/cm^2, 1×10^{14} n_{eq}/cm^2, and 2×10^{14} n_{eq}/cm^2, corresponding to a tenth, half, and end of operation, respectively. In addition to the integrated fluences expected during BTL operation, the annealing scheme discussed in Section 5.3.1 must also be evaluated. The effect of in-situ annealing is emulated with an accelerated annealing procedure: SiPM arrays were annealed at 70°C for 40 minutes, three days at 110°C, and four days at 120°C. This annealing sequence is expected to provide an annealing equivalent to about twice what is expected in BTL operation with an annealing temperature of 60°C during the technical stops envisaged at the HL-LHC [123]. It must be noted that the irradiation level irreversibly affects the SiPM parameters, including gain loss, PDE loss, breakdown voltage drift, and the change of DCR versus overvoltage behavior, while the annealing procedure is not expected to affect any of these parameters. The operating temperature during the test beam measurements was chosen to emulate the expected DCR in the BTL at a given point in time, corresponding to a specific integrated luminosity, based on the annealing that the SiPMs would have undergone in situ at that time, assuming operation at −45°C. Changing the temperature only affects the DCR and cannot be used to emulate radiation-induced variations in gain or PDE. Therefore, each irradiated SiPM can only properly emulate one point in the BTL lifetime, even if measurements at different temperatures are performed. The DCR levels are expected to be negligible

[5]The effect of radiation exposure on the timing performance is expected to be dominated by radiation damage to the SiPMs, with a negligible effect on LYSO crystals.

at the beginning of operation. For irradiated SiPMs, their values can be estimated using Equation 5.13. More specifically, by measuring the SiPM array current (I_{array}), the DCR can be computed as in Equation 5.16, assuming that the current is equally distributed among the 16 channels of the array.

$$DCR = \frac{I_{array} \ / \ 16}{G \cdot e} \tag{5.16}$$

A detailed study of the impact of different parameters on the time resolution requires a thorough understanding of the contributing terms, which are discussed in Section 5.1. In particular, the electronics contribution arises from random fluctuations in the signal amplitude and shape caused by electronics effects. To evaluate it, a system in which the photo-statistics term is negligible is considered: a coincidence setup consisting of two SiPMs not coupled to any scintillator, illuminated with laser light at a wavelength of $\lambda = 370$ nm and a pulse width < 45 ps, and read out with tester boards using the TOFHIR2 ASIC. Different laser intensities and SiPM overvoltages are used to probe a wide range of signal amplitudes. In this configuration, the fluctuations affecting the time resolution are expected to originate solely from the electronics contribution. The measured time resolution is studied as a function of the slope of the measured pulse. While TOFHIR2 does not provide a measurement of the full pulse, the average pulse shape can still be reconstructed by evaluating the times corresponding to specific discriminator thresholds. By performing a threshold scan, the corresponding times are extracted, allowing the average pulse shape to be obtained. Fitting the rising edge of the pulse with a linear function at the discriminator threshold enables the evaluation of its slope (dI/dt). Results of the coincidence time resolution measured per channel as a function of the pulse slope are reported in Figure 5.17, where data are modeled using Equation 5.1. The parameters extracted from the fit are reported in Equation 5.18 and are used in the

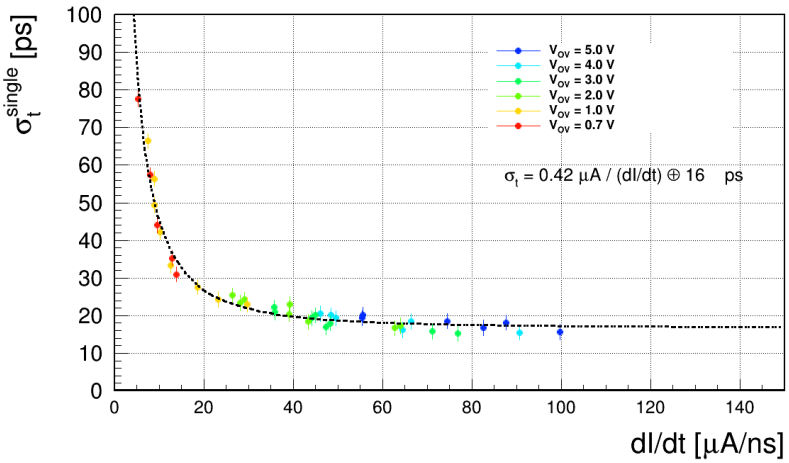

Figure 5.17: The coincidence time resolution per channel is shown as a function of the measured pulse slope. Data at different overvoltages are shown in different colors. All data are fitted with Equation 5.1, and the resulting parameters are reported.

following studies to evaluate the electronics noise, given a measured slope of the pulse.

$$\sigma_{noise} = (0.420 \pm 0.035)\,\mu\text{A}, \tag{5.17}$$

$$\sigma_{TDC} = (16 \pm 2)\,\text{ps} \tag{5.18}$$

5.5.1 Test beam experimental setup

A significant part of this work has been dedicated to the collection and analysis of data at various test beam campaigns in which modules were positioned along a beam line, and the time of arrival of a charged particle was recorded. Beam test campaigns were conducted at the Fermilab Test Beam Facility and at the CERN SPS H8 beam line, using 120 GeV protons and 180 GeV pions, respectively, in 2022 and 2023. Results from the different test beam campaigns are equivalent, as demonstrated in Figure 5.18, confirming a good reproducibility of the measurements. Therefore, in the following, results are presented collectively rather than separated by campaign. In both facilities, accelerators deliver spills of particles to the exper-

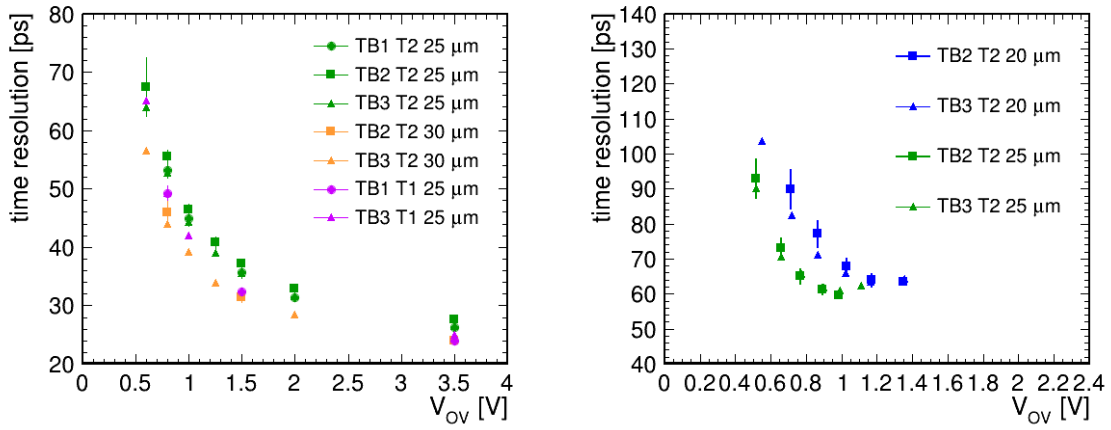

Figure 5.18: The time resolution is shown for the same module using the same color, and for the same test beam campaign using the same marker. The performance of various modules tested in three different campaigns (TB1, TB2, TB3) is evaluated as a function of the SiPM overvoltage, both for the non-irradiated modules (left) and for the modules irradiated to the levels of DCR expected at the end of operation of BTL. The comparison across different test beam campaigns highlights the good reproducibility of the measurements within the uncertainties.

imental setup with varying durations and beam characteristics. Nevertheless, both 120 GeV protons and 180 GeV pions produce an energy deposit similar to that expected from MIPs at the HL-LHC under BTL operating conditions, thus providing a realistic environment for evaluating sensor performance. In BTL, a charged particle crossing a crystal is expected to deposit an energy spectrum following a Landau distribution, due to stochastic fluctuations in a thin absorber, with the MPV of approximately 0.86 MeV/mm. The signal produced by the interaction of the particle within the sensor module is used to extract the timing information, and the corresponding time resolution is evaluated.

In the test beam facility, the beam first traverses wire chambers used for tracking and then reaches the experimental setup, which is housed inside a light-tight cold box that provides controlled and stable temperature conditions for the sensors. The reference frame of the setup is defined with the z-axis running along the beam direction, while the x-axis and y-axis are aligned with the crystal's short and long axes, respectively. A sketch of the sensor module reference frame compared to the beam is shown in Figure 5.19. Inside the box, a reference module is rotated by 90° relative to the beam direction so that it is oriented perpendicular to the axis of the crystal bar. The reference module provides a time resolution of about 30 ps with a position resolution of about 3 mm for incoming particles in the horizontal plane. The sensor module under test is placed downstream of the reference module, with the crystal bar axis being parallel to the (x-z) plane. The module orientation relative to the beam is varied by an angle θ, controlled remotely via a table that allows positioning with a precision of 1°. Specifically, the module is always tilted such that a MIP crosses a single crystal. Finally, a plastic scintillator of 7×7 cm^2, whose light is detected by a SiPM, acts as the acquisition trigger. The temperature inside the box can be varied between -30°C and 15°C, while humidity is controlled using nitrogen gas. The TECs provide an additional temperature difference of 5-10°C, and the operating temperature is measured with the PT1000 sensors installed on the rear side of the SiPM array. The set temperature demonstrated to remain stable within ± 1°C. The reference and test modules are mounted on a dedicated metal stand equipped with fans (shown in Figure 5.20), which ensures optimal thermal coupling to the TECs. The TECs contact the metal side via a specific thermal grease to enhance thermal conductivity. These supports are designed for the test beam setup and are not intended for final BTL operation, for which a dedicated mechanical design will be used, as discussed in Section 5.10.3. Sensor modules were tested with tester boards housing TOFHIR2 ASICs of either the 2X or 2C versions, both of which demonstrated similar performance. Signals are routed via Samtec HQCD cables from the tester boards to a Front-End Board type D (FEB-D) [132] (shown in Figure 5.20), developed by PETsys Electronics for SiPM readout in PET applications. The FEB-D supplies power to the ALDOs for biasing the SiPMs of the sensor modules and to the TOFHIR2 ASICs

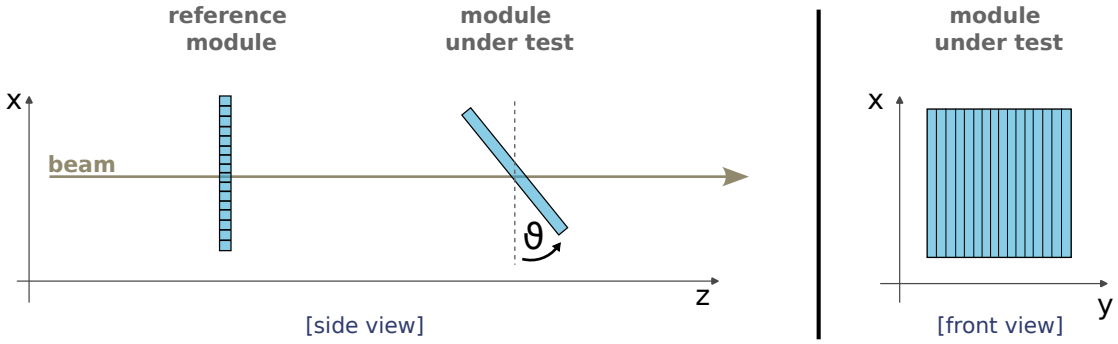

Figure 5.19: Schematic drawing of the configuration of modules and their rotation in the experimental setup. The reference module is placed upstream, followed by the module under test along the z direction. A view of the $x-y$ plane for the module under test is also shown on the right. These two views correspond to the side view (left) and front view (right) of the module under test.

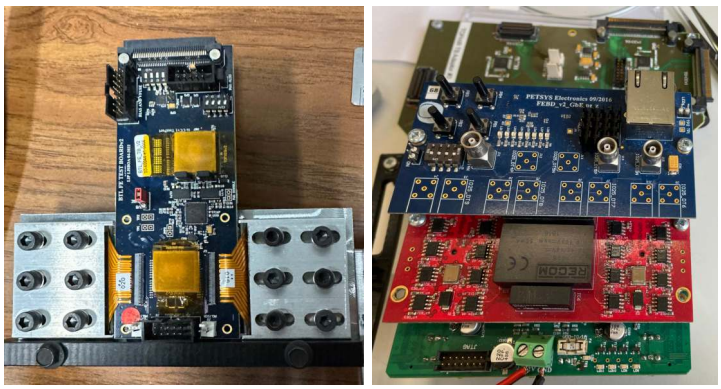

Figure 5.20: Left: Picture of a tester board mounted on the heat sink and mechanical support. Right: Picture of the FEB-D board used to provide bias to the electronics components, such as ALDO and TOFHIR, and to read out the SiPM signals.

on the tester board for proper signal processing. It also handles readout and transmits the data to a PC for event reconstruction. Separate cables are connected between the tester board and power supply to bias the TECs via the corresponding pins, using the appropriate paths in the flex cables. Pictures of the experimental setups used at the facilities, along with a schematic, are shown in Figure 5.21.

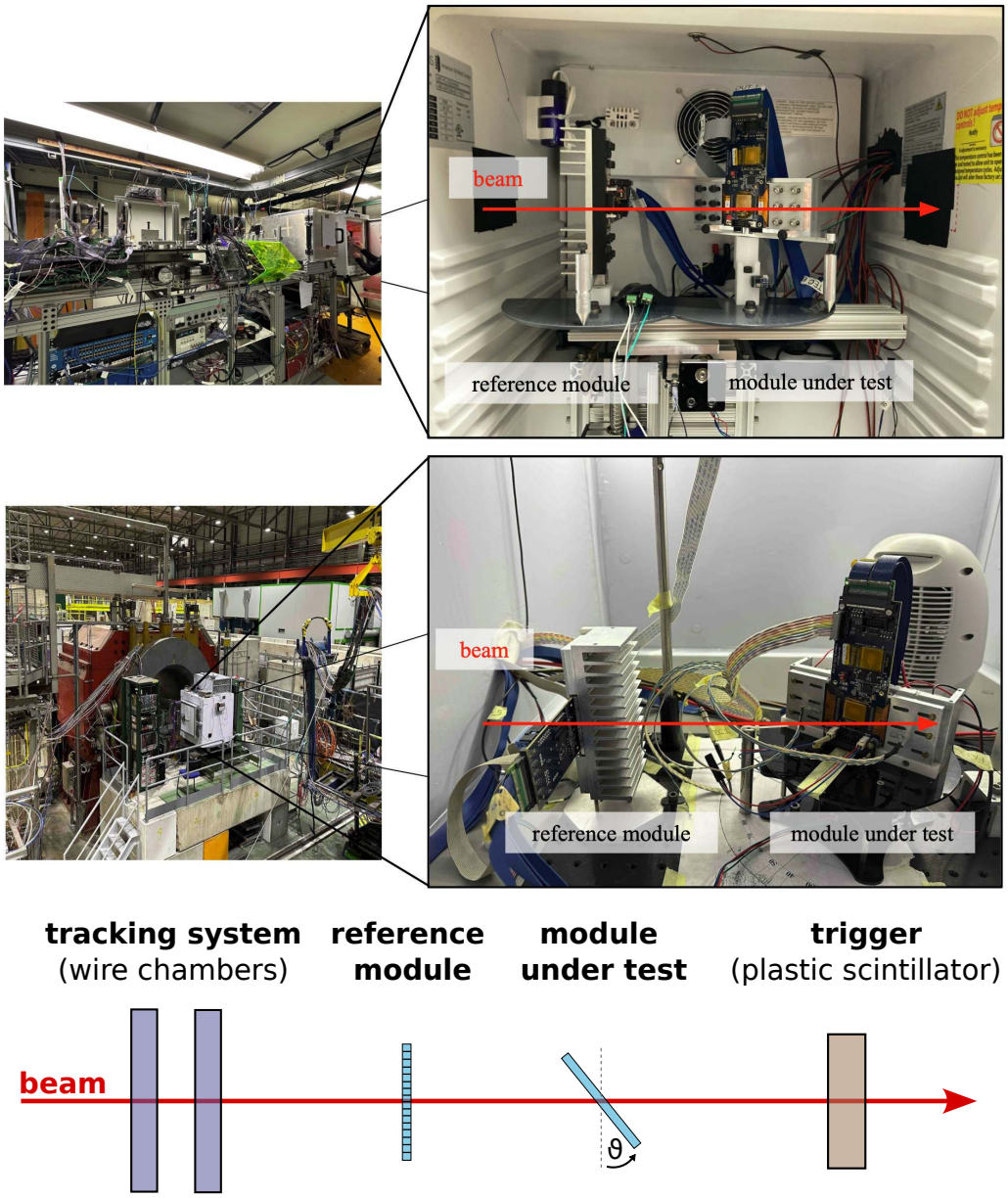

Figure 5.21: Pictures of the experimental setup are shown with labels indicating the various components, both for the setup at the Fermilab facility (top row) and for the CERN setup (middle row). The setup is located inside a cold, light-tight box and consists of a reference module, a module under test, and a trigger module, listed in order along the beam direction. A schematic representation of the beam line traversing the setup, also showing the tilt of the module under test with respect to the beam direction is shown in the bottom row.

5.5.2 Analysis methods

The cold box is aligned with respect to the beam using an adjustable table that can be moved vertically. The desired beam shape is tuned together with the beam line experts to cover the width of the sensor

module under test. The beam shape is verified using the wire chambers, which serve as a tracking system. An example of a well-shaped beam is shown in Figure 5.22. Since the sensor module under test is rotated at different tilt angles, it is necessary to calibrate the angular reference corresponding to the perpendicular to the beam. This can be checked by performing energy measurement at different tilting angles. The MPV of the Landau distribution, used to model the energy distribution, can be evaluated as a function of the angle set by the rotating table. This provides an estimate of the table's angle offset, which can then be corrected accordingly. An example of a calibration measurement of the energy deposit as a function of the tilting angle is shown in Figure 5.22.

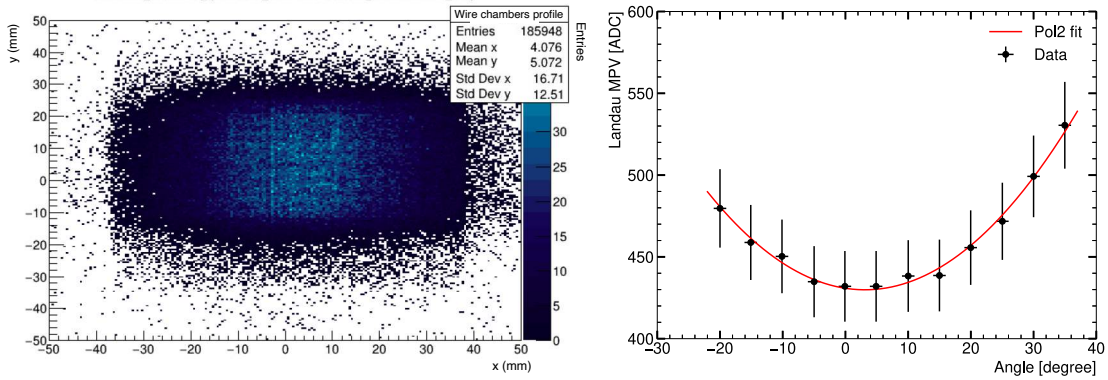

Figure 5.22: Left: Beam density recorded by the wire chamber in the x and y directions (see Figure 5.19), highlighting good coverage of the entire sensor module. Right: Example of angle calibration, showing the Landau MPV of the energy deposit as a function of the tilting angle of the sensor module under test. A fit is overlaid to estimate the angle offset of the configuration.

TOFHIR2 provides measurements of the time at which a pulse crosses a fixed threshold and of the integrated pulse charge, which is related to the energy deposited inside a crystal by a particle. Events in a crystal bar are selected by requiring energy deposits compatible with a MIP, whose typical energy spectrum is shown in Figure 5.23. The energy deposited in the bar is measured as the average of the charges integrated by the two SiPMs at the ends of the bar. The energy spectrum can be described by a Landau distribution, and events are selected within a range between 0.80·MPV and a cut-off on the maximum amplitude to exclude saturated signals at high ADC channels. To reduce the beam spot size to the order of millimeters along the longitudinal axis of the test bar, a selection on the impact point position is applied, requiring coincidence with a MIP event in one central bar of the reference module.

In BTL, the time of arrival of a MIP in a single bar is computed from the average of the two time measurements, t_L and t_R, corresponding to the measurement at the two sides of the bar, as shown in Equation 5.19.

$$t_{bar} = 0.5 \cdot (t_L + t_R) \tag{5.19}$$

When no correlated uncertainties exist between the two time stamps, and for a fixed impact point along the bar, the bar time resolution can be estimated as half of the spread of the difference $t_{\text{diff}} = t_L - t_R$ between the two SiPM measurements, as described in Equation 5.20. This is particularly useful when a precise external reference is not available (for instance, an MCP).

$$\sigma_{t_{bar}} = \frac{1}{2}\sigma_{t_{diff}} \tag{5.20}$$

The time difference method was already used in Ref. [133], where it was shown to provide a time resolution of the crystal bar comparable to that obtained from the average time, using a Micro-Channel Plate detector placed upstream as a time reference. It must be considered that, in the analysis, events are required to be in coincidence with a bar of the reference module, which introduces a spread of a few millimeters in the impact point. Therefore, a small residual contribution to the time resolution remains in the time difference estimation, which can be evaluated as in Equation 5.21, with k being a coefficient

measured to be 9 ps/mm, approximately corresponding to the reciprocal of the speed of light in LYSO:Ce.

$$\Delta t = 2k\,\Delta x \tag{5.21}$$

The impact on time resolution of the residual contribution given by the spread of the selected impact point can be computed from the illuminated portion of the bar Δx, which increases with the tilting angle θ of the bar, as in Equation 5.22. This additional contribution is estimated to be approximately 10 ps for $\theta = 32^o$, 13 ps for $\theta = 52^o$, and 18 ps for $\theta = 64^o$.

$$\sigma_{pos} \;=\; \sigma_{\Delta t/2} \;=\; k\sigma_x \;=\; k\,\Delta x\,/\,\sqrt{12} \tag{5.22}$$

Also, a varying signal amplitude might affect the time measurement, as described in Section 4.3.3.1, hence requiring amplitude walk corrections evaluation. In this work, they are evaluated empirically by studying the dependence of t_{diff} on the ratio of the energies measured at the two bar ends. This dependence is visible in the example in the right panel of Figure 5.23, where the measured time is shown as a function of the energy ratio. An empirical polynomial function is used to fit the data and provide a correction for the raw time measurement in each event. This method is used to estimate the time measurement in every bar of a sensor module.

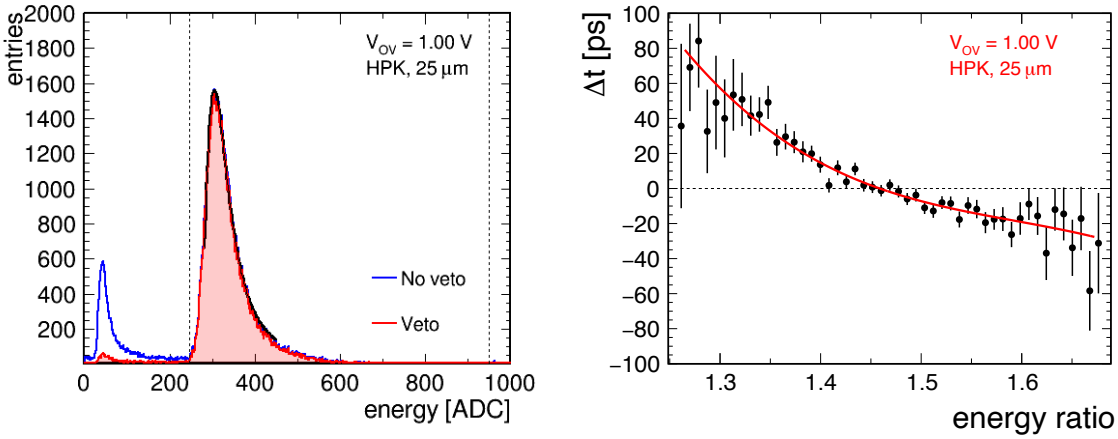

Figure 5.23: Left: An example of a MIP energy deposit distribution in a single bar is shown, computed as the average of the energies measured by the two SiPMs. The distribution is fitted with a Landau function, and events are selected within a range defined by the dashed lines. At low energies, events from cross-talk could contaminate the signal (blue distribution, "no veto"). By requiring that the neighboring bars do not record energies above $0.80 \cdot$ MPV, these events are significantly reduced (red distribution, "veto"). Additionally, applying a lower cut at $0.80 \cdot$ MPV on the energy deposited in the bar further limits this effect. Right: An example of the dependence of t_{diff} on the energy ratio measured at the two ends of the bar is shown, with an overlaid polynomial function used to interpolate the data and correct them.

The time measurement is extracted from leading-edge discrimination, as mentioned in Section 4.3.3.1. For every configuration tested during the test beam campaigns, different discriminator thresholds are set and the time performance is evaluated for each. Eventually, the timing threshold is determined as the one providing the best time resolution. An example of a performed threshold scan is shown in Figure 5.24, where the time resolution is reported as a function of the discriminator threshold. Additionally, measurements are taken for varying ranges of overvoltages, depending on the irradiation level of the sensors. For non-irradiated sensors, a range of approximately 0.5–3.5 V overvoltage can be measured, while for irradiated sensors, due to radiation damage, the overvoltage scan is limited to 0.5–1.0 V. In Figure 5.24, the threshold scans for each bias voltage used are reported, both for non-irradiated sensors and for sensors

irradiated to the full fluence expected at the end of BTL operation. The used thresholds correspond, in terms of single photoelectron amplitude, to a range varying from approximately 1 photoelectron (for 1 μA, with a SiPM having a cell size of 25 μm operating at $V_{OV} = 3.5$ V) to 32 photoelectrons (for 8 μA, with the same SiPM operating at $V_{OV} = 1.0$ V). The specific threshold at which the time resolution is extracted has a non-trivial effect on the contributions determining the overall performance: the photo-statistics term increases with the threshold, while the electronics noise decreases as the threshold increases. Therefore, an optimal balance must be found, usually in the 2–3 μA range. For irradiated sensors, the additional contribution from the DCR term makes it necessary to further adjust the threshold.

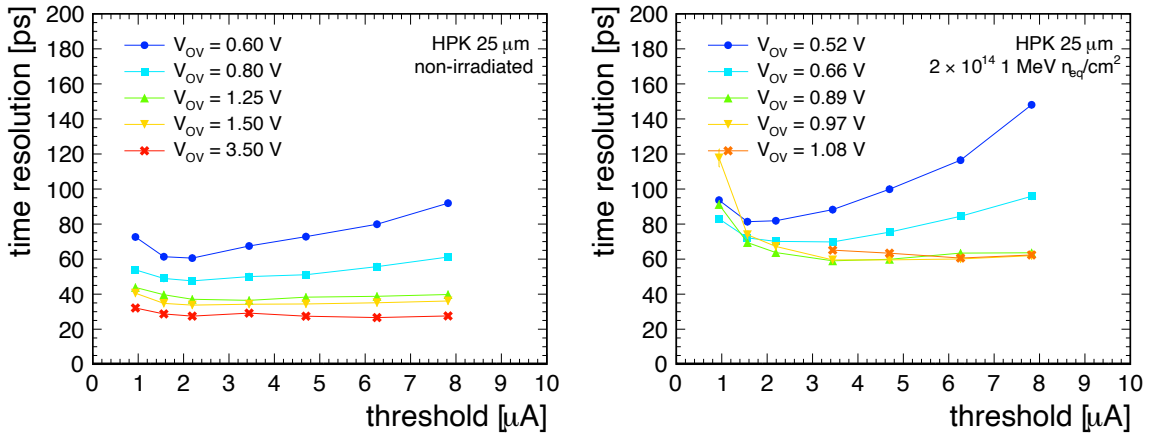

Figure 5.24: The time resolution as a function of the leading-edge discrimination threshold is shown for different SiPM overvoltages, both for non-irradiated sensors (left) and for SiPMs irradiated to 2×10^{14} n_{eq}/cm^2, operating at a temperature of $-35°C$ to emulate the DCR level expected at the end of BTL operation.

For each configuration, the slope of the pulse is measured for every bar via a linear fit performed at the discrimination threshold, and the average slope across bars is taken as an estimate of the pulse slope for the module. The slope estimation carries an uncertainty of about 10%, arising from fluctuations in the linear fit range and variations observed between channels. An example of the rising edge of the average measured pulse shape measured in a bar of a non-irradiated module and in a module irradiated to the expected fluence at the end of operation is shown in Figure 5.25.

5.6 Results of comparative studies

The first phase of the test beam campaigns was dedicated to evaluating the impact of various parameters on the time performance. These included the SiPM cell size and the sensor module thickness, as motivated by the studies presented in Section 5.3.2 and Section 5.4, as well as different operating conditions such as irradiation fluences and temperatures. The final sensor design, featuring 25 μm cell size SiPMs and 3.75 mm-thick sensor modules, is the result of the studies outlined below. The results discussed in this Section were obtained before the vendor selection and were therefore conducted on sensor modules composed of crystals and SiPMs from different producers. Crystal matrices were produced by Sichuan Tianle Photonics and Suzhou JT Crystal Technology, which demonstrated similar performance in terms of light output. SiPMs were produced by HPK and FBK using different silicon wafer technologies, which affect parameters such as PDE, gain, and ECF differently. The protective layer on the active area also differs: HPK SiPMs are covered with 300 μm of silicone resin, while FBK SiPMs are covered with a 50 μm layer of sputtered glass, which is expected to provide greater light collection efficiency. The different technologies are also affected differently by radiation damage; studies of these SiPMs before and after irradiation can be found in Refs. [125, 134].

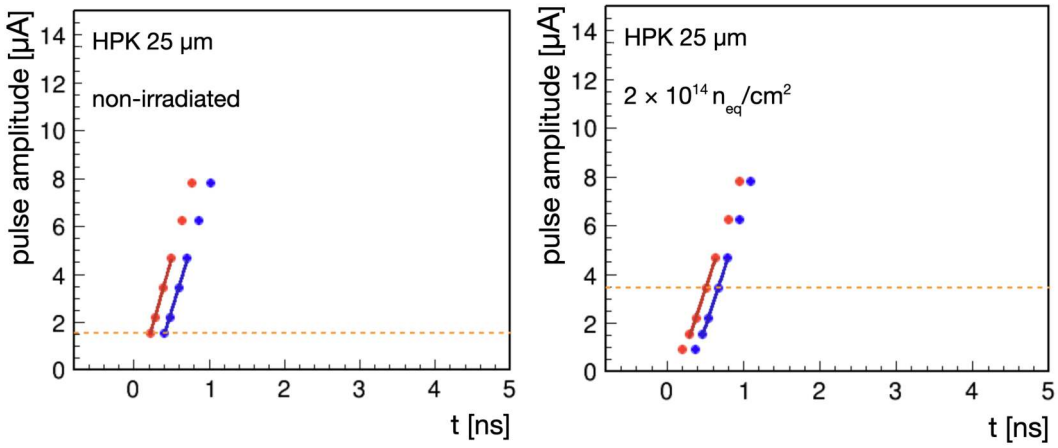

Figure 5.25: An example of average pulse shape measured in one bar averaged over events, for each SiPM, is reported both for a non-irradiated module (left) and for a module irradiated to the fluence expected at the end of operation (2×10^{14} n_{eq}/cm^2). The rising edge for both the left and right SiPM is shown and the slope of the pulse is computed via a linear fit at the timing threshold for every SiPM. The slope of the pulse associated to a bar is then computed as the weighted average of the measurements extracted for the two bar ends.

5.6.1 SiPM cell size impact on time resolution

The studies discussed in Section 5.3 motivated a further investigation of the impact of cell size on time performance, to validate how the increased signal amplitude and the increased DCR associated with larger SiPM cell sizes balance in the overall performance. Modules consisting of SiPMs with varying cell sizes were evaluated in beam tests. Sensor modules with a thickness of 3.00 mm (T2), consisting of HPK and FBK SiPMs were tested. Cell sizes of 15, 20, 25, and 30 μm were assessed for non-irradiated HPK and FBK SiPMs. At the time of the beam test campaign, the only SiPMs available for irradiation were HPK SiPMs with cell sizes of 15, 20, 25, and 30 μm, and FBK SiPMs with cell sizes of 15 and 25 μm. The irradiated SiPMs were exposed to a fluence of 2×10^{14} n_{eq}/cm^2. The operating temperature was set to 5°C for non-irradiated sensors, since the DCR levels are negligible and the light output performance is not influenced by temperature. While, irradiated sensors were measured at a temperature of −35°C, corresponding to a DCR level equivalent to that expected at the end of BTL operation at the HL-LHC. Non-irradiated sensor modules and modules irradiated to the fluence expected at the end of operation were measured with a tilting angle of 52° with respect to the beam, to emulate an average energy deposit expected in a medium pseudorapidity region. The expected operating voltages are around 3.5 V overvoltage at the beginning of operation (non-irradiated sensors) and approximately 1 V overvoltage at the end of operation (sensors irradiated to full fluence).

Time resolution measured for modules with SiPMs of different cell sizes is measured for each bar following the methods described in Section 5.5.2, and the overall performance is computed as the average over the bars of the module, with an associated uncertainty given by the spread of the measured time resolution among bars. As mentioned before, the time resolution is extracted for every configuration at the discrimination threshold, or timing threshold, providing the best performance. The timing thresholds can be studied as a function of the SiPM overvoltage, as shown in Figure 5.26. For larger cell sizes, the optimal balance between the aforementioned effects is found at higher thresholds compared to smaller cell sizes. This is enabled by an increased signal amplitude in larger cell sizes and it is expected to reduce the impact of electronics noise and the DCR term in irradiated modules.

5.6.1.1 Non-irradiated modules

Results for the sensor modules consisting of HPK and FBK non-irradiated SiPMs are reported in Figure 5.27. A time resolution of 25-30 ps for an overvoltage of 3.5 V is observed, with comparable results between the two SiPM vendors. The performance observed in modules with varying SiPM cell sizes is

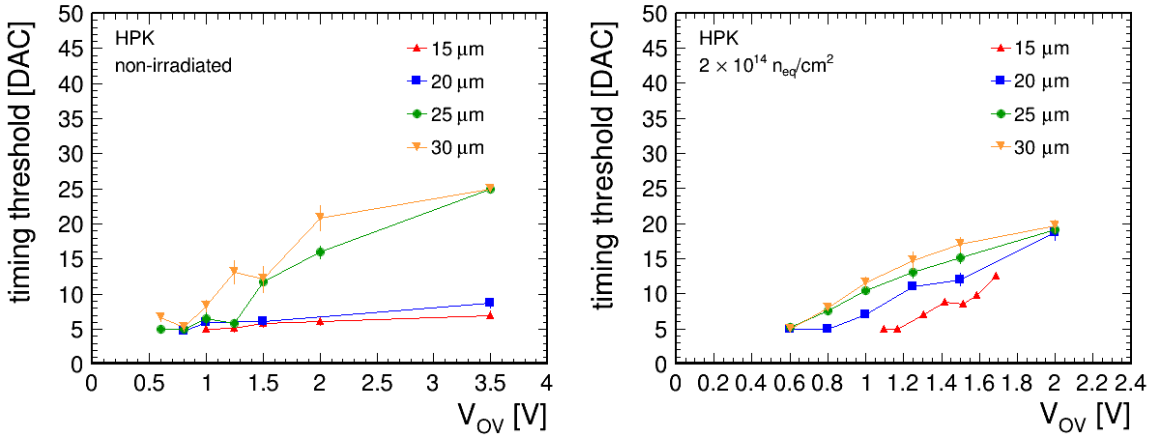

Figure 5.26: The timing threshold corresponding to the best time resolution for each module as a function of the SiPM overvoltage is shown. Results for modules with HPK SiPMs of different cell sizes are reported for both non-irradiated modules (left) and modules irradiated to full fluence (right). For non-irradiated modules, higher overvoltage and larger cell sizes produce higher signal amplitudes, allowing an increase of the timing threshold to reduce the contribution of the electronics. This improvement continues until it is limited by the worsening of the photo-statistics term. For irradiated modules, the effect of the DCR on the threshold is visible: all cell sizes tend to require a higher threshold for increasing overvoltage, as the DCR increases.

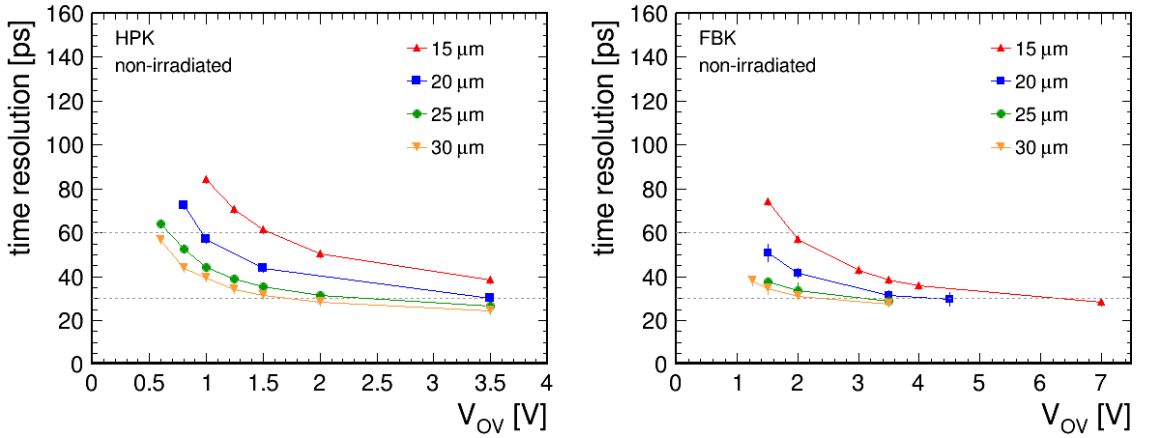

Figure 5.27: The time resolution as a function of SiPM overvoltage is shown for modules with different non-irradiated SiPM cell sizes, for SiPMs produced by HPK (left) and FBK (right).

further investigated by breaking down the total time resolution measured on beam into the expected contributing terms (as defined in Equation 4.1). The photo-statistics term is computed as the quadrature difference between the total time resolution and the electronics noise term, assuming them to be the only two contributions in this configuration. Results for the different HPK SiPM cell sizes are reported in Figure 5.28.

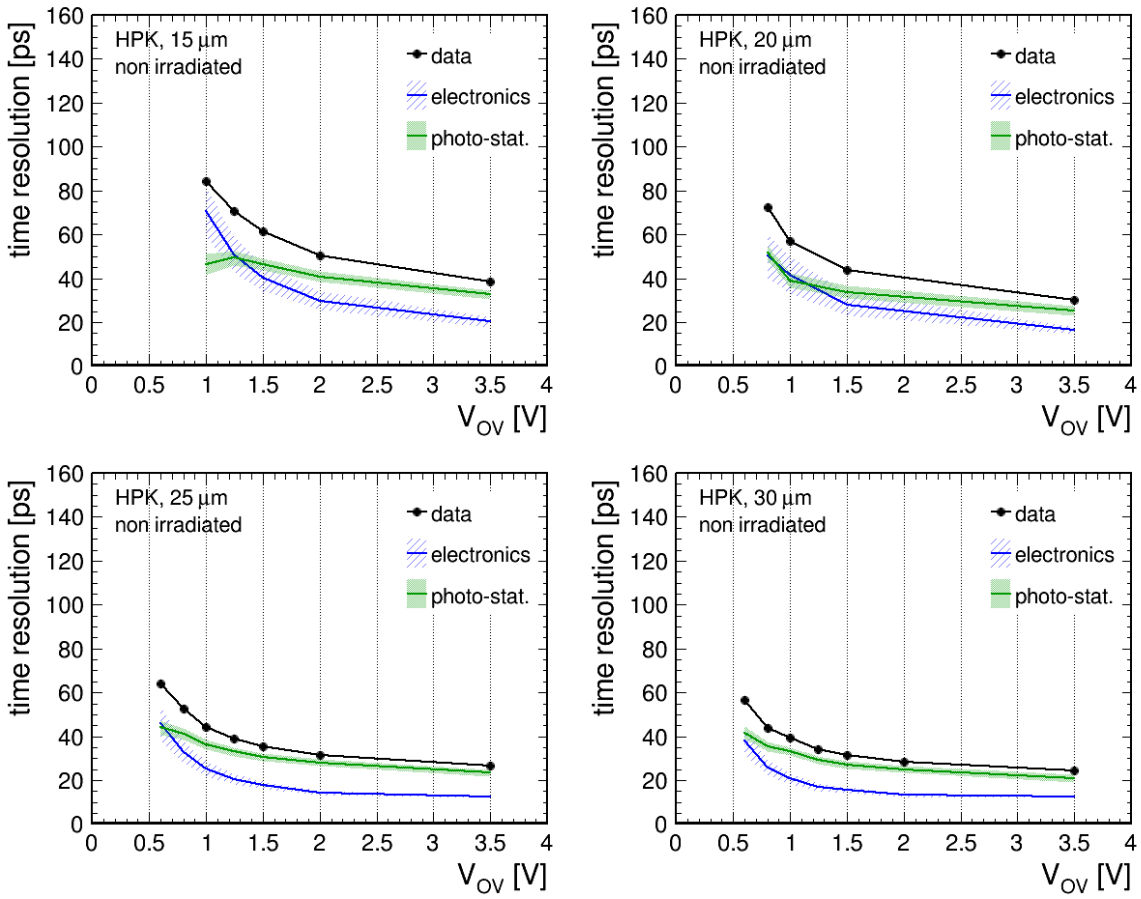

Figure 5.28: The time resolution as a function of the overvoltage is shown, along with the breakdown of its contributions, for T2 non-irradiated modules with SiPMs of 15, 20, 25 and 30 μm cell size. Data points are marked in black, whereas the photo-statistics and electronics contributions are shown in green and blue, respectively. The bands represent the associated uncertainties. In the beginning of operation conditions, the dominant term is expected to be the photo-statistics term for every SiPM cell size.

In non-irradiated conditions, the larger PDE provided by increased cell sizes increases the number of detected photoelectrons, resulting in a reduction of the photo-statistics contribution. At the same time, the increased SiPM gain leads to a steeper signal slope, as shown in Figure 5.29, thereby improving the electronics-related contribution to the time resolution. An approximately linear increase of the signal slope with increasing overvoltage is observed, consistent with the linear dependence of the gain on the overvoltage, which is the dominant factor affecting the pulse slope. Overall, for all investigated cell sizes, the dominant contribution to the total time resolution is the photo-statistics term. The dependence of the photo-statistics contribution on the SiPM PDE is investigated in more detail, and the calculated photo-statistics terms under varying SiPM characteristics are used to derive an empirical and predictive model of this dependence. To this end, the photo-statistics values as a function of the PDE are fitted with a global empirical function given in Equation 5.23. Data and fit are reported in Figure 5.29, along with the resulting fit parameters yielding an exponent of $\alpha = -0.73 \pm 0.04$. This fitted parameter slightly differs from the expected value in Equation 5.2, likely due to experimental threshold effects, which are not accounted for in the model.

$$\sigma_{photo} = p_0 \cdot PDE^{\alpha} \tag{5.23}$$

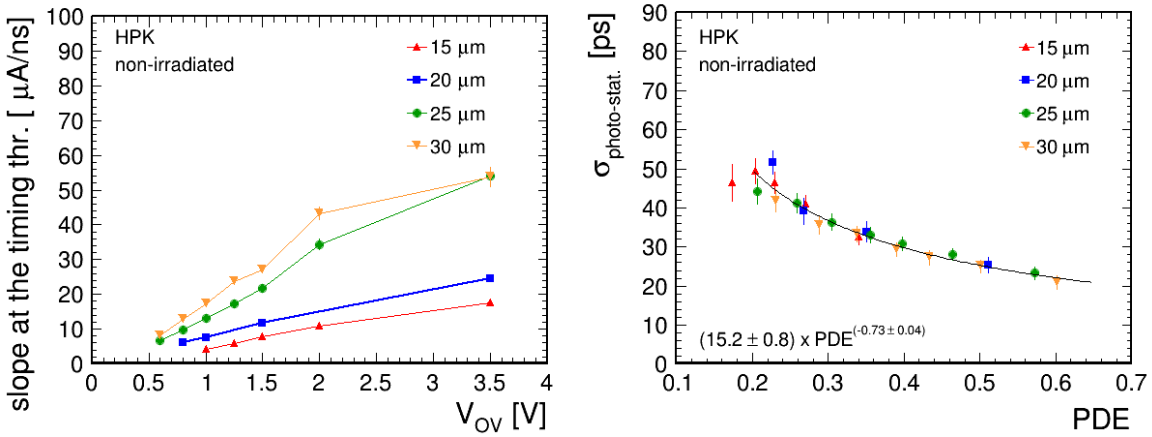

Figure 5.29: Left: The slope of the pulse measured at the timing threshold is shown as a function of the SiPM overvoltage for modules with non-irradiated HPK SiPMs of varying cell sizes. Right: The photo-statistics term is shown as a function of the PDE, with values obtained for different HPK SiPM cell sizes indicated by different colors. All data points are fitted using the function in Equation 5.23, and the resulting fit parameters are reported in the figure.

5.6.1.2 Modules irradiated to end of BTL operation conditions

While the results discussed so far refer to non-irradiated conditions, radiation damage is expected to significantly modify the timing performance during the BTL lifetime at the HL-LHC. In particular, the increase in dark counts plays a major role in determining the time resolution after irradiation. DCR levels are estimated for each irradiated SiPM array from measurements of the SiPM array current, as described in Equation 5.16. Aiming to investigate the different dependence of DCR on overvoltage and gain for the different cell sizes, its values are studied and compared in Figure 5.30. At fixed overvoltage

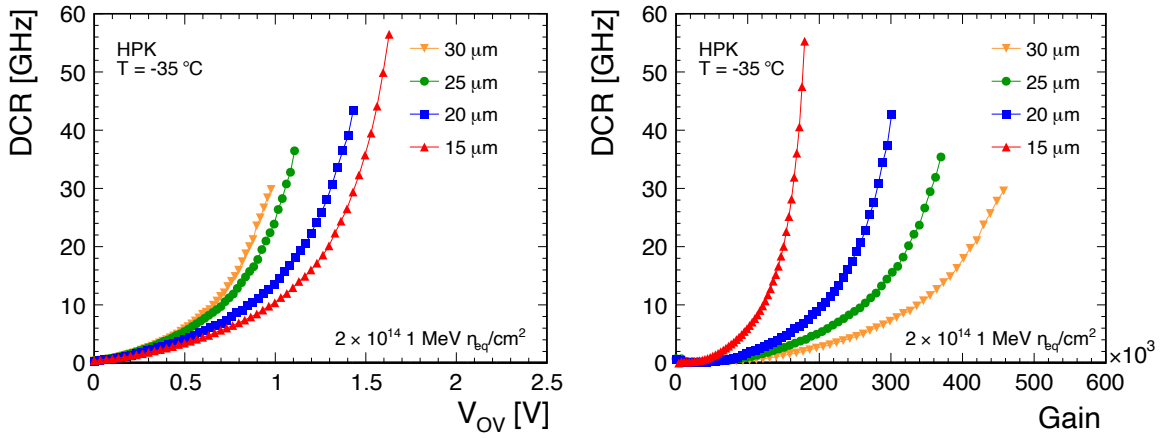

Figure 5.30: The DCR measured in a SiPM is shown as a function of overvoltage and as a function of gain for different configurations. The comparison of DCR for T2 (3.00 mm thick) HPK SiPMs with different cell sizes (15, 20, 25, and 30 μm) is shown, both as a function of overvoltage (left) and gain (right).

and operating conditions (same irradiation level, annealing, and operating temperature), a DCR more than twice as high is observed, for instance, between 15 μm and 25 μm cell size HPK SiPMs. SiPMs with larger cell sizes have a higher fill factor, corresponding to more active area, which increases the probability of producing spurious counts, i.e. increases the DCR levels. DCR is also evaluated as a function of the gain, where up to a factor of four improvement is observed in gain at the same DCR level from 15 μm to 30 μm cell size. This is beneficial for increasing the signal amplitude and therefore reducing

the electronics contribution.

After evaluating the DCR from currents measurement, the time resolution is assessed with the methods described in the previous Sections. Results are shown as a function of the overvoltge in Figure 5.31, where the optimal condition is found to be in the range between 0.8 and 1.4 V. A large improvement in time resolution is observed for HPK SiPMs with cell sizes of 20, 25 and 30 μm compared to the 15 μm SiPMs, yielding almost a factor of two better performance. While, an almost negligible performance variation is observed for FBK SiPMs between 25 μm and 15 μm cell size.

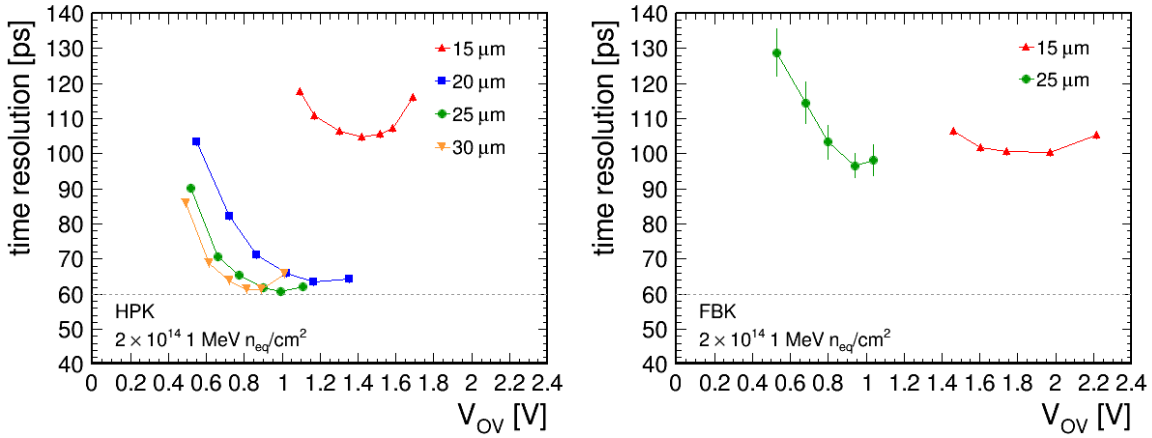

Figure 5.31: The time resolution as a function of SiPM overvoltage is shown for T2 modules with different SiPM cell sizes irradiated to 2×10^{14} $n_{eq}/cm^2$and operated at $-35°C$ to emulate the end-of-operation DCR levels. Results are shown for SiPMs produced by HPK (left) and FBK (right).

In order to interpret these results and quantify the relative impact of the different noise sources after irradiation, the total time resolution is decomposed into its individual terms. The photo-statistics term is estimated from the parametrization reported in Equation 5.23, taking into account the observed signal amplitude (gain × PDE) loss in irradiated SiPMs. The uncertainty associated with this estimate is obtained by propagating the errors of both the fit parameters and the measured PDE. The DCR contribution is then derived as the quadrature difference between the total time resolution and the contributions from electronic noise and photo-statistics, with the corresponding uncertainty evaluated through error propagation. These terms are computed for each module, and a summary of the individual contributions for all modules with HPK SiPMs is shown in Figure 5.32. The photo-statistics term measured for larger cells shows a slight improvement due to the increased PDE, despite the lower operating voltage, while both the electronics noise and DCR contributions are significantly reduced around the optimal overvoltage. Although the larger cell size leads to higher DCR and PDE values, the overall effect results in a net performance gain, primarily due to the improved signal-to-noise ratio. The increased SiPM gain and PDE enhance the signal amplitude, thereby reducing the relative contribution of the electronics noise, which is particularly critical at low operating voltages. The impact of the increased gain can be evaluated through the pulse slope, as shown in Figure 5.33. Similarly to what is observed for non-irradiated SiPMs, the slope shows an approximately linear increase with increasing overvoltage. However, in irradiated SiPMs the DCR levels significantly affect the overall performance, requiring a more detailed evaluation of this effect. The behavior of the time resolution term due to DCR was studied as a function of the DCR itself, normalized to the signal amplitude, using the ratio of the measured PDE to a reference PDE obtained for 25 μm HPK SiPMs at 1 V overvoltage. The data were collectively fitted with a global function, as expressed in Equation 5.24. The resulting fit parameter is $\beta = 0.39 \pm 0.06$, which is compatible with the expected value of 0.5 within 1.8σ. Data and the corresponding fit are shown in Figure 5.33.

$$\sigma_{DCR} = p_1 \cdot DCR^{\beta} \tag{5.24}$$

Another important consideration for irradiated SiPMs is that the BTL power budget limits the power consumption per SiPM to approximately 30 mW. The increased dark counts observed in irradiated SiPMs

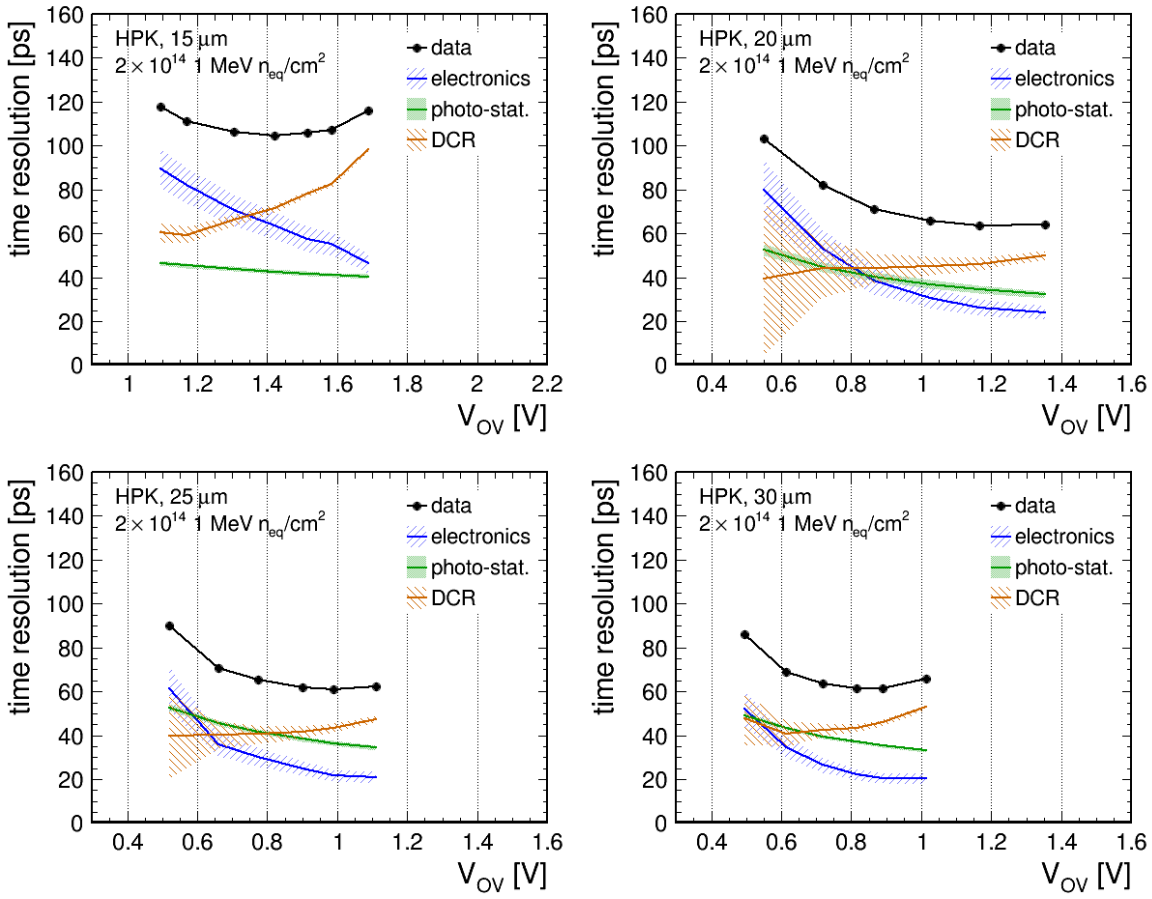

Figure 5.32: The time resolution as a function of the overvoltage is shown, together with the breakdown of its contributions, for T2 modules (3.00 mm thick) with HPK SiPMs of 15, 20, 25, and 30 μm cell size, irradiated to 2×10^{14} n_{eq}/cm^2. Data points are shown in black, while the photo-statistics, electronics, and DCR contributions are represented in green, blue, and orange, respectively. The bands indicate the associated uncertainties. A significant improvement in the DCR contribution is observed when moving from the 15 μm to larger cell sizes, which can be attributed to the beneficial effect of an overall higher signal-to-noise ratio.

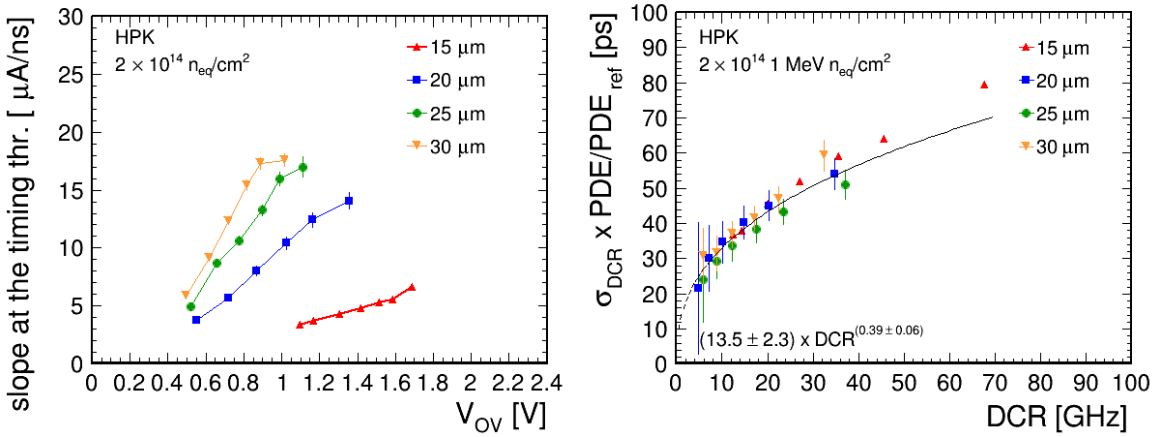

Figure 5.33: Left: Studies for the different T2 modules with HPK SiPMs of varying cell sizes, irradiated to full fluence, are reported. The slope of the pulse measured at the timing threshold is shown as a function of the SiPM overvoltage. Right: The DCR term is normalized to the signal amplitude computed as the ratio between the measured PDE and a reference PDE (PDE_{ref}), i.e. the PDE measured for 25 μm HPK SiPMs at 1 V OV. The resulting values are shown as a function of the DCR for different SiPM cell sizes. The data are modeled with the fit function in Equation 5.24, with the fit results reported in the figure.

also lead to higher dark currents and, consequently, higher static power consumption. To evaluate the power consumption of the configurations tested on beam, the static power is calculated as the product of the SiPM bias voltage and its dark current. The results, expressed as the time resolution as a function of static power, are shown in the left panel of Figure 5.34. These measurements demonstrate that a time resolution of approximately 60 ps can be achieved with T2 sensor modules featuring 25 μm cells, remaining within the power budget.

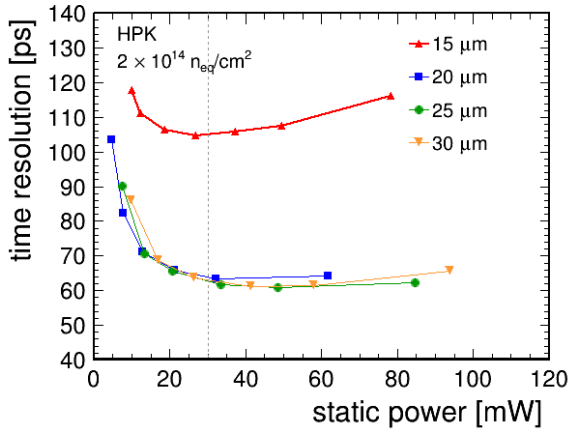

Figure 5.34: The time resolution measured for different irradiated HPK SiPM cell sizes is shown as a function of the static power.

From the results obtained with irradiated SiPMs, it is evident that the 15 μm HPK SiPM exhibits significantly worse performance. This can be explained by several factors. First, when moving from 15 μm to larger cell sizes, HPK confirmed that a different wafer design was implemented to enhance the PDE, and direct measurements showed that the improvement achieved exceeds what would be expected solely from the increased cell size. Moreover, the signal amplitude loss due to radiation damage was estimated to be about 30% for the 15 μm sensors, compared to only 20% for the larger cell sizes. In

contrast, for FBK sensors, only a limited improvement is observed with increasing cell size. This can be attributed to the fact that no signal amplitude loss was observed for the 15 μm sensors, while a 20% reduction was measured for the 25 μm devices, counterbalancing the improvement expected from the larger cell size and thus explaining the observed timing performance. The SiPMs also exhibited a shift in breakdown voltage after irradiation, suggesting structural damage that affects the electric field. Such effects may vary between the center and the edges of the cells, and therefore depend on the cell size. Effects of radiation-induced damage in SiPMs are discussed in Section 5.3.1, but further investigation studies could be performed to fully understand these effects. Within the scope of this work, the observed degradation of FBK SiPM performance after irradiation led to the selection of HPK SiPMs for BTL production. By comparing the performance of modules equipped with HPK SiPMs of different cell sizes, a clear improvement in the signal-to-noise ratio is observed for larger cells. In particular, the 25 μm HPK SiPMs were identified as the optimal choice for the BTL detector, meeting the target timing performance while remaining within the allowed power budget of 30 mW per SiPM.

5.6.2 Impact of sensor thickness on timing performance

The impact of sensor module thickness on performance was also assessed in beam tests, by measuring the time resolution obtained with modules with thicknesses of 3.75 mm (T1), 3.00 mm (T2), and 2.40 mm (T3), as in the original TDR design and discussed in Section 5.4. The sensor thickness refers to both the crystal matrix thickness and the active area of the SiPM coupled to the bar end face. For this study, the final choice of SiPM cell size was used, employing HPK SiPMs with 25 μm. Additionally, to evaluate performance over the full BTL lifetime, both non-irradiated and irradiated sensors (to the final nominal fluence) were measured.

Modules with different thicknesses were tested in the experimental setup described in Section 5.5.1, following the analysis methods discussed in Section 5.5.2. All modules were tilted by an angle of $\theta = 64°$ to emulate the energy deposit of a MIP in the high pseudorapidity region of the BTL with $|\eta| > 1.15$, where the T3 geometry was envisaged. Their performance is assessed in terms of time resolution as a function of operating voltage, for both non-irradiated and irradiated sensors, and the results are shown in Figure 5.35. To evaluate these results within a more consistent and detailed context, the contributions

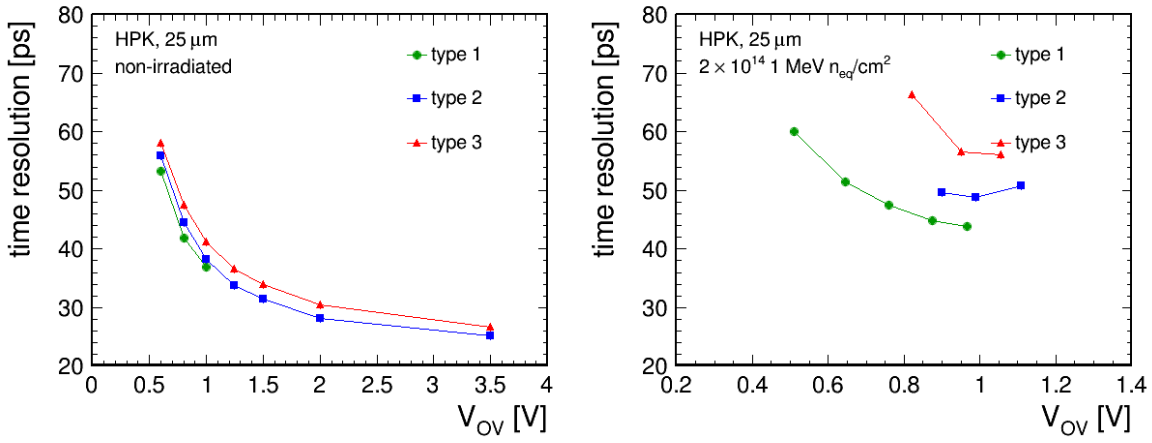

Figure 5.35: The time resolution as a function of overvoltage is shown for three types of modules: T1 (green), T2 (blue), and T3 (red). Results are presented for non-irradiated modules (left) and for modules irradiated to 2×10^{14} n_{eq}/cm^2 (right), representing the beginning and end of operation, respectively. The modules were tilted to emulate a pseudorapidity region of $|\eta| \sim 1.15$.

to the time resolution are extracted following the methods described in the previous section. The results, shown at the optimal overvoltage for each case, are presented in Figure 5.36 as a function of sensor thickness.

For the non-irradiated case, only a marginal improvement is observed when increasing thickness from T3 to T1, with two main effects at play. Increasing the sensor thickness increases the energy deposited in the crystal, which, as deduced from Equation 5.4, increases the number of photoelectrons. Additionally,

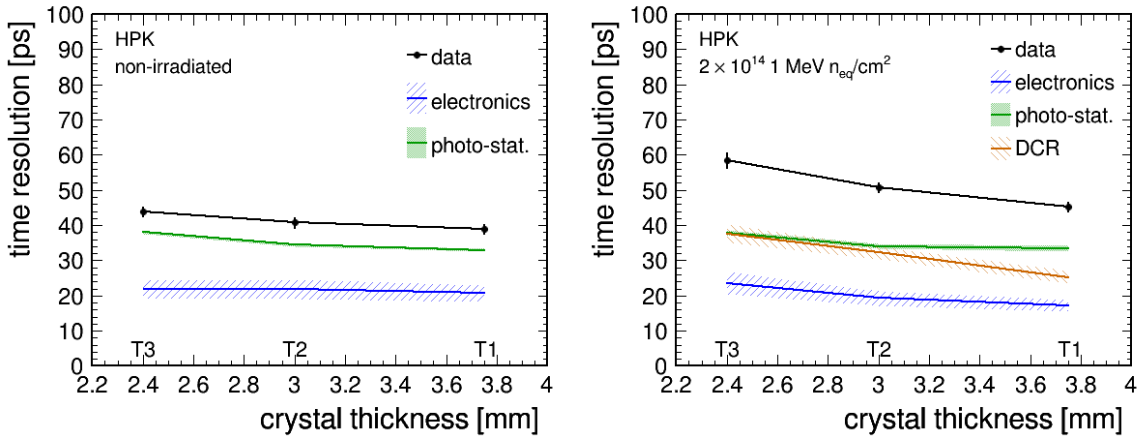

Figure 5.36: Time resolution and its breakdown into contributing terms (i.e., photo-statistics, electronics, and DCR) are shown for the non-irradiated (left) and irradiated (right) cases as a function of the sensor module thickness. The modules were tilted to emulate a pseudorapidity region of $|\eta| \sim 1.15$ and results correspond to an overvoltage of $V_{OV} = 0.95$ V.

a better light collection efficiency is expected, as described in Section 5.4, due to improved geometrical acceptance in thicker crystals. However, the larger active area of the SiPM also results in a slower response, as discussed in Section 5.3.2, because of the larger capacitance associated with a larger active area. The balance of these effects leads to the electronics noise being nearly independent of thickness (remaining around 20–22 ps), while the photo-statistics term shows only a small improvement of a few ps from the increased thickness (improving from about 38 ps to 34 ps). For irradiated sensors, a larger improvement is observed when comparing T3 to T1 modules, which can be attributed to the mitigating effect that increased thickness has on the DCR term. The total time resolution measured for irradiated sensors as a function of thickness exhibits a trend approximately scaling as the inverse square root of thickness, consistent with the beneficial impact of the increased number of photoelectrons in Equation 5.3, confirming the underlying assumptions. Although the DCR worsens due to the larger SiPM active area, the increased number of photoelectrons results in a net improvement. Based on these observations, modules of constant thickness at all pseudorapidities (T1 modules, 3.75 mm thick) were selected as providing the best overall performance.

5.6.3 Impact of fluence and operating temperature on performance

Throughout the BTL lifetime, the SiPMs will be exposed to varying neutron fluences, ranging from non-irradiated conditions up to the end-of-operation level of 2×10^{14} n_{eq}/cm^2. Since some radiation-damage effects do not scale linearly with fluence, it is also important to study intermediate fluence levels. To this end, two additional points along the expected BTL lifetime were emulated by irradiating SiPMs to fluences of 1×10^{14} n_{eq}/cm^2 and 1×10^{13} n_{eq}/cm^2, applying the same annealing scheme described before, and operating them at different temperatures to emulate various DCR levels.

The DCR was first studied as a function of overvoltage under different irradiation levels and operating temperatures. The results are shown in Figure 5.37. Modules with a thickness of 3.75 mm (T1) and a SiPM cell size of 25 μm, irradiated to fluences of 1×10^{13}, 1×10^{14}, and 2×10^{14} n_{eq}/cm^2, emulating different points along the BTL lifetime, are compared. For a relative comparison, T2 modules consisting of SiPM arrays from HPK and FBK were irradiated to full fluence. The measured DCR was then compared, with the HPK SiPM arrays also operated at different temperatures to evaluate the corresponding effect on DCR. The effect of different irradiation levels on the DCR consists in an increase faster than linear. Also, the DCR is expected to depend on temperature, where an increase by a factor of 2 every 10°C is observed, confirming previous observations (see Section 5.8). Due to different wafer technologies employed by HPK and FBK in SiPM production, radiation exposure is expected to affect the two sensor samples differently. A reduction in DCR of about 20% is observed in FBK SiPMs compared to HPK

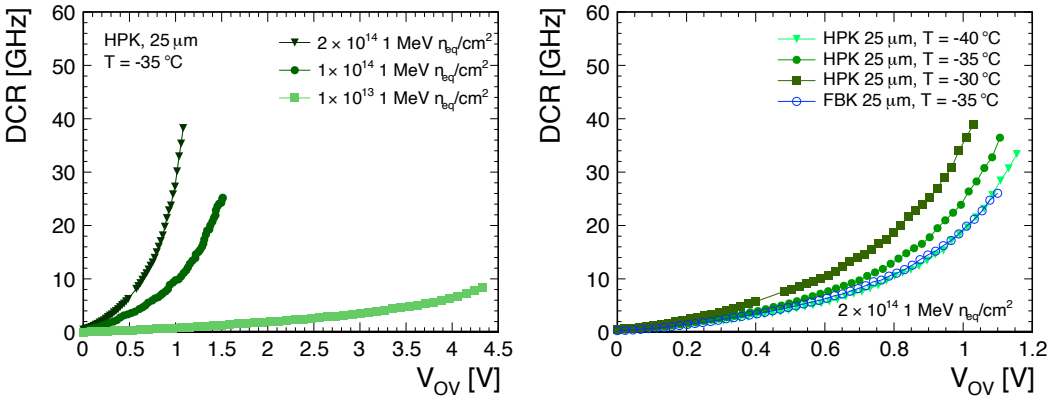

Figure 5.37: The DCR measured in a SiPM is shown as a function of overvoltage and as a function of gain for different configurations. Left: T1 modules (3.75 mm thick) with 25 μm cell-size HPK SiPMs irradiated to different fluences (1×10^{13}, 1×10^{14}, and 2×10^{14} n_{eq}/cm^2) are compared. Right: T2 modules (3.00 mm thick) with 25 μm cell-size SiPMs from HPK and FBK irradiated to 2×10^{14} n_{eq}/cm^2 were compared under different operating temperatures.

SiPMs. However, this did not result in an improvement in the time resolution due to the effects of PDE loss, as discussed in Section 5.6.1.2.

The time resolution of T1 modules irradiated to the three different fluence levels was measured at various operating temperatures, emulating different DCR conditions. The analysis of the time resolution follows the procedure described in the previous section. Its values are reported as a function of both overvoltage and static power per SiPM for each configuration in Figure 5.38. The observed time resolution exhibits a stronger dependence on temperature for more heavily irradiated SiPMs compared to less irradiated ones, with differences of up to 5 ps for a 5°C change. Temperature is expected to affect only the DCR, as demonstrated in Figure 5.39, where the measured slope of the signal pulse is shown as a function of overvoltage for modules irradiated to 1×10^{14} n_{eq}/cm^2 and 2×10^{14} n_{eq}/cm^2. The slope is found to be nearly independent of operating temperature. Since both the noise and photo-statistics contributions are determined by the pulse slope, as discussed in the previous sections, it can be concluded that the only temperature-dependent contribution to the time resolution originates from the DCR term. Given that the DCR is the only parameter affected by temperature, and that its contribution to the time resolution is negligible at low fluences, it becomes evident why the temperature dependence becomes more pronounced at higher DCR levels. Evaluating the performance in terms of static power further emphasizes the advantages of operating SiPMs at lower temperatures, which simultaneously reduces power consumption and mitigates self-heating effects. The results shown in Figure 5.38 demonstrate that time resolutions of 30, 44, and 58 ps can be achieved under operating conditions defined in terms of DCR levels, and thus irradiation and operating temperature, corresponding to exposure equivalent to half a year, five years, and ten years of HL-LHC operation, respectively, all within the allowed BTL power budget.

This first stage of test beam campaigns was dedicated to comparative studies evaluating performance variation under different conditions in terms of different parameters, such as the SiPM cell sizes, the sensor module thickness, the irradiation levels and the operating temperatures. The performance observed with different SiPM cell sizes oriented the final sensor design towards a cell dimension of 25 μm, motivated by the superior performance compared to smaller pixels and also by the demonstrated capability of operating within the power budget. In addition, only HPK was selected as the SiPM vendor for the BTL, due to the better radiation tolerance and overall timing performance. Indeed, the 25 μm HPK SiPMs provide an optimal balance between the increased PDE and gain (raising the signal amplitude) and the increased DCR (due to larger cells) and power dissipation. An additional potential for improvement has been offered by increasing the sensor thickness, bringing an overall net increase in the number of

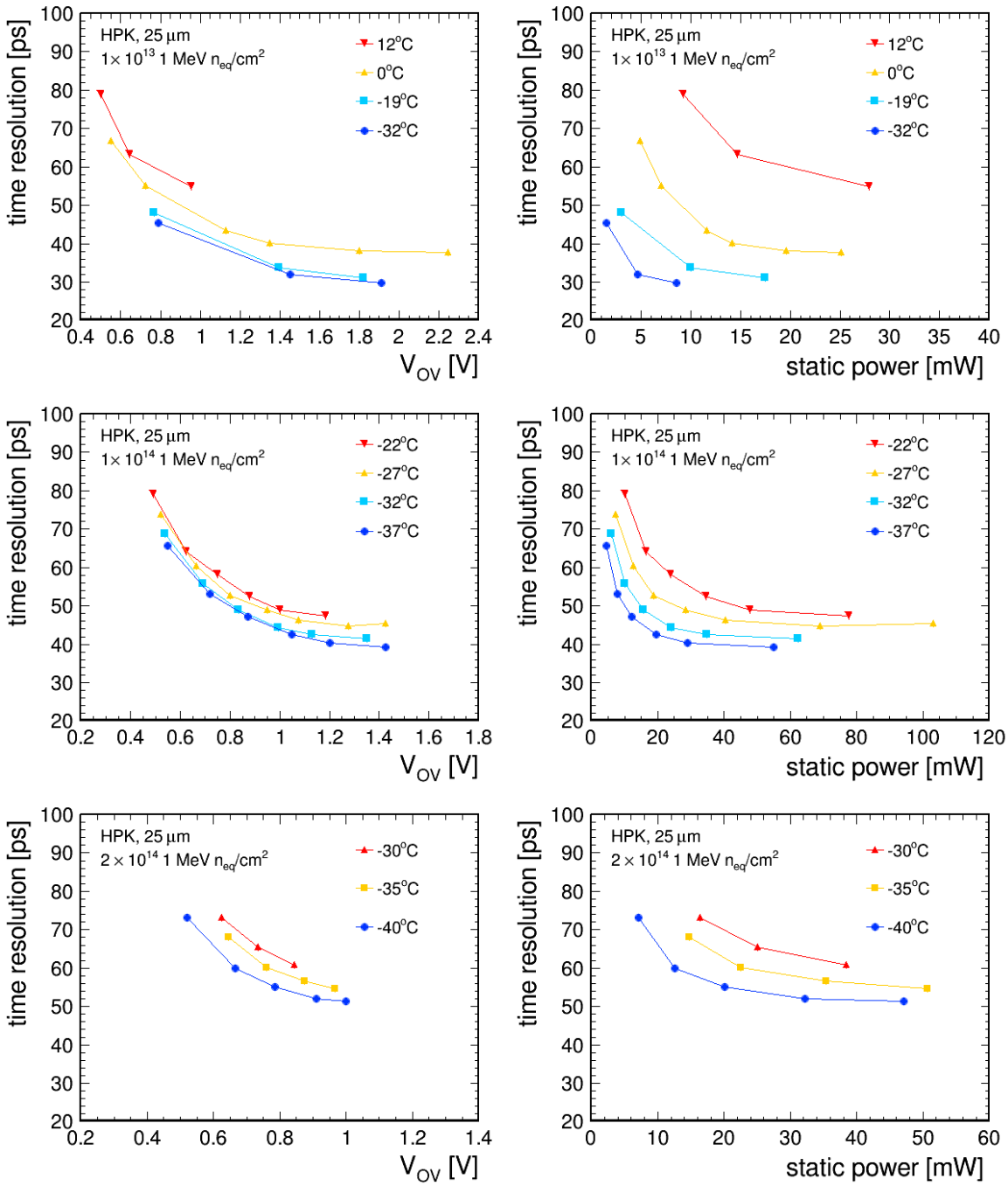

Figure 5.38: The effect of different operating temperatures at various radiation levels was evaluated for T1 modules (3.75 mm thick) with 25 μm HPK SiPM cell size. Modules were irradiated to 1×10^{13} n_{eq}/cm^2 (top), 1×10^{14} n_{eq}/cm^2 (middle), and 2×10^{14} n_{eq}/cm^2 (bottom), with all SiPMs undergoing the same annealing procedure. Time resolution is reported as a function of overvoltage in the left column and as a function of static power in the right column.

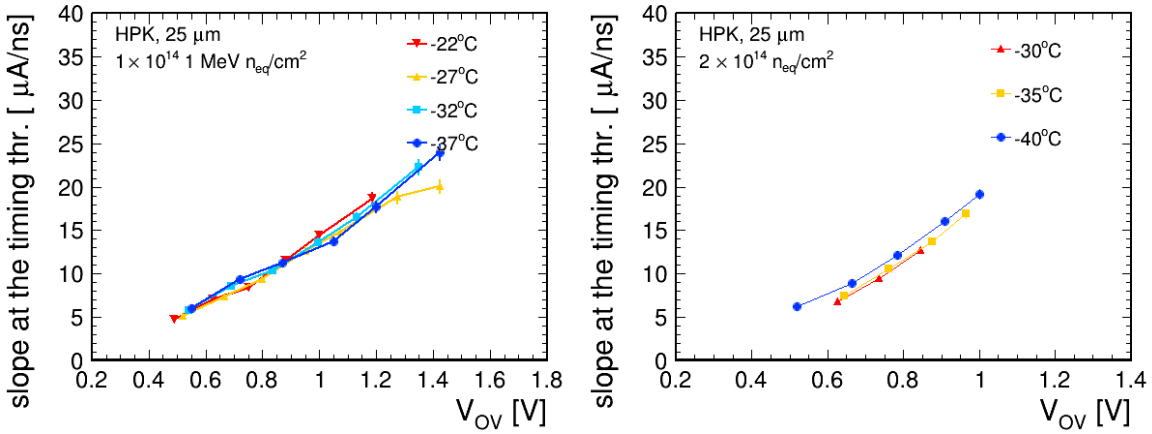

Figure 5.39: The slope of the pulse at the timing threshold is shown as a function of SiPM overvoltage for T1 modules (3.75 mm thick) with 25 μm HPK SiPM cell size irradiated to 1×10^{14} n_{eq}/cm^2 (left) and 2×10^{14} n_{eq}/cm^2 (right) at different operating temperatures. The results indicate that operating temperature and DCR levels have an almost negligible impact on the pulse slope.

photoelectrons produced, which is particularly beneficial in mitigating the DCR term degradation. The improvement in performance has finally driven the choice of instrumenting the entire BTL surface with T1 sensors, with an affordable cost increase of 10%.

5.7 Confirming timing performance of sensors with final specifications

The best-performing technical specifications identified in the previous studies are adopted for the final sensor design. Sensor modules with a thickness of 3.75 mm and a SiPM cell size of 25 μm were tested in beam to assess their performance. The evaluation of the BTL sensor modules under near-real conditions focused on the uniformity of performance across the module, the effects of different irradiation levels corresponding to various stages of the BTL lifetime, and the performance across the pseudorapidity region covered by the BTL. To assess performance under BTL lifetime conditions, SiPMs were exposed to neutron fluences of 1×10^{13}, 1×10^{14}, and 2×10^{14} n_{eq}/cm^2, corresponding to integrated luminosities of 150, 1500, and 3000 fb^{-1}, respectively. All devices underwent the same annealing procedure described in the previous studies, and the operating conditions were tuned according to the DCR level expected in realistic BTL conditions. These three irradiated configurations, together with a sensor module consisting of non-irradiated SiPMs, were tested in beam using the experimental setup described in Section 5.5.1. The results were analyzed following the methods described in Section 5.5.2, and the contributing terms were estimated as detailed in Section 5.6.1.

The timing performance was evaluated as a function of the SiPM overvoltage for both non-irradiated and irradiated modules. Results with modules tilted at an angle $\theta = 52°$, corresponding to an energy deposit of 5.2 MeV expected in the central pseudorapidity region of BTL, are shown in Figure 5.40. The total measured time resolution is further broken down into the individual terms contributing to the overall performance. As seen in Figure 5.35, the electronics noise is dominant at low overvoltage, where the SiPM gain and PDE are small, therefore making the signal amplitude limited and the overall contribution to time resolution quite high (Equation 5.1). With increasing overvoltage, as visible in Figure 5.14, both PDE and gain increase, generating a larger signal pulse that in turn reduces the electronics contribution, and at this point the dominant contribution becomes the photo-statistics term in the non-irradiated case, where there is no radiation effect producing DCR. At the optimal operating conditions of approximately 3.5 V a time resolution of 25 ps is observed, emulating the performance at the start of operation conditions at the HL-LHC. In the irradiated case, depending on the radiation exposure, the DCR starts playing a role and generates corresponding fluctuations in the DCR term that become more and more relevant.

For the end of operation irradiation condition, an optimal overvoltage of approximately 1 V is identified, among which the operations would be degraded by self-heating effects on the SiPMs and they would anyway not allowed by power budget constraints. At this optimal overvoltage, a DCR of approximately 20 GHz is measured, to which a term of about 35 ps is associated. The total time resolution observed at this optimal bias voltage is 55 ps.

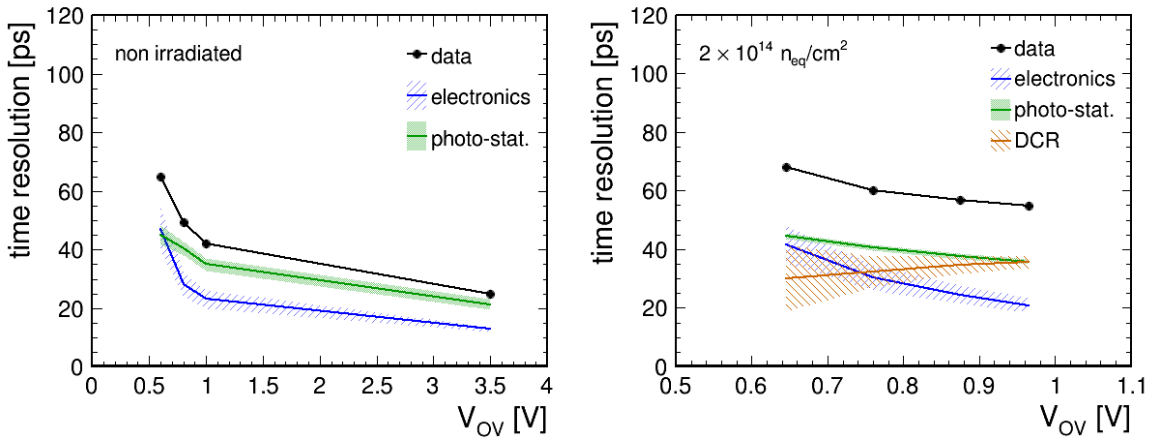

Figure 5.40: Time resolution and its contributing terms are shown for T1 modules with 25 μm SiPM cell size in the non-irradiated (left) and irradiated to 2×10^{14} n_{eq}/cm^2 (right) cases. As in the previous figures, the colored lines indicate the contributions from electronics (blue), photostatistics (green), and DCR (orange). Uncertainties in the data points and bands representing the contributions are included.

With sensor module dimensions of 52×55 mm^2, an important aspect to evaluate is the response uniformity within the module. This was studied by measuring the time resolution of the 16 crystal bars for different MIP impact points along the longitudinal axis of the bars. Measurements were performed both for non-irradiated sensors and for sensors irradiated to 2×10^{14} n_{eq}/cm^2. A spread smaller than 2 ps RMS is observed in both cases, as shown in Figure 5.41, demonstrating[6] a uniform light output across the bars and confirming an effective optical coupling of all SiPMs to the crystal bars. This also indicates a small spread in breakdown voltages within a single array, which must be smaller than 150 mV by technical specifications. The uniformity along the x direction (see Figure 5.19) was evaluated by selecting events where the MIP interacted in different bars of the reference module. With a tilting angle $\theta = 52°$, this corresponds to scanning in steps of approximately 5 mm. Also in this case, the uniformity along the bar is measured to be better than 2 ps for both the non-irradiated an irradiated sensors.

5.7.1 Expected BTL performance across pseudorapidity

Sensor modules will cover the entire surface of the BTL detector, spanning a pseudorapidity region up to $|\eta| = 1.48$. This results in a performance variation along the barrel axis, since the energy deposition depends on the angle at which particles cross the modules. At higher pseudorapidity values, particles traverse a larger slant thickness and thus deposit more energy, which in turn translates into better time resolution[7]. To evaluate this effect, non-irradiated modules and modules irradiated to 2×10^{14} n_{eq}/cm^2 were tilted at three different angles, $\theta = 32°$, $52°$, and $64°$, selected to emulate the MPV energy deposit in the low, medium, and high pseudorapidity regions of the BTL. Their performance was evaluated as a function of overvoltage, with the operational range constrained by the power budget. Results are reported in Figure 5.42 and demonstrate the improvement provided by the increased energy deposit, which

[6] Potential variations within a single module across the bars can be attributed to either the light output or the SiPM breakdown voltage (and therefore the PDE), with the latter being constrained by design to ensure a small spread across SiPM channels.

[7] This improvement is marginally reduced by the increased radiation level expected at high $|\eta|$ values, ranging from about 1.65×10^{14} to 1.90×10^{14} n_{eq}/cm^2 [93] when going from $|\eta| = 0$ to $|\eta| = 1.45$

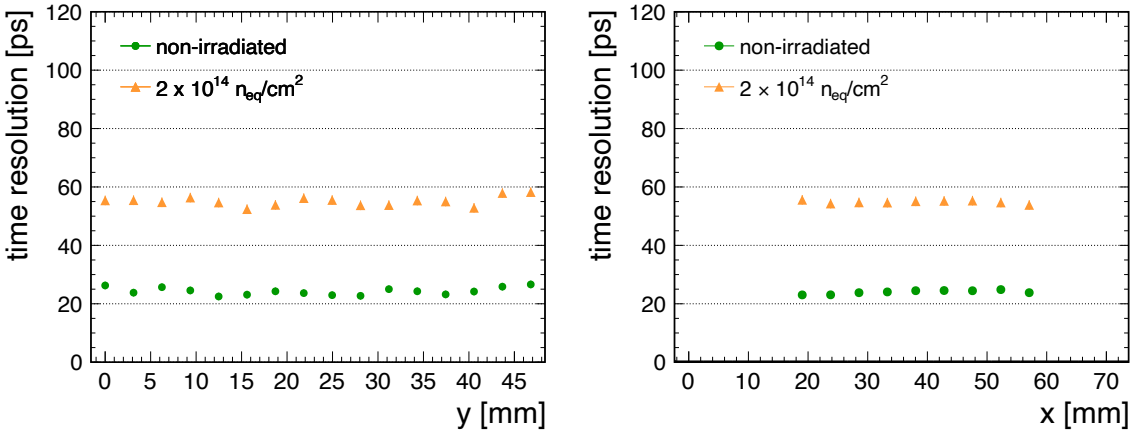

Figure 5.41: The time resolution along the y-axis (left) and x-axis (right) is reported for the non-irradiated case (green markers) operated at 3.5 V overvoltage and for the irradiated case to 2×10^{14} n_{eq}/cm^2 (orange markers) operated at 0.96 V overvoltage. For the uniformity along the x-axis, the time resolution at each impact point is averaged over all bars of the module. Due to the coarse determination of the x position, measurements do not cover the first and last 5 mm of the bar.

directly enhances the number of photoelectrons and therefore the overall performance. Furthermore, the improvement is more pronounced for irradiated SiPMs, where the DCR scaling as $1/N_{pe}$ is a significant contribution (as shown in Figure 5.40).

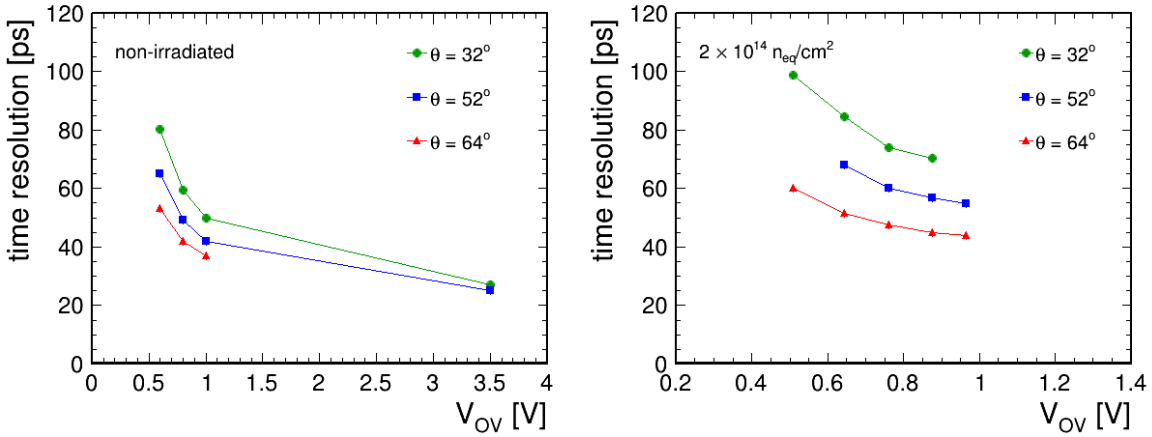

Figure 5.42: Time resolution as a function of overvoltage for modules with non-irradiated (left) and irradiated SiPMs (right). Modules were tilted at different angles to emulate the energy deposition at various pseudorapidities. The missing point at high overvoltage for non-irradiated SiPMs at $\theta = 64°$ is due to poor data quality.

The performance across the BTL pseudorapidity can be assessed by mapping the measurements taken at different test beam angles θ to the corresponding pseudorapidity points. To this end, the energy deposited at each angle is evaluated and converted into the corresponding η value. This is done by estimating the energy deposit at the test beam and determining the expected η at which the same energy would be deposited in the BTL, based on minimum-bias simulated events, as shown in the left panel of Figure 5.43. The energy deposited at the test beam is expressed in Equation 5.25, where the energy deposited by a MIP per unit length, MPV_{MIP}, is multiplied by the slant thickness traversed by a particle

in a crystal tilted by an angle θ with respect to the beam direction.

$$E_{dep} = \frac{MPV_{MIP} \text{ thickness}}{\cos\theta}. \tag{5.25}$$

The time resolution was extracted at the optimal operating point within the expected power budget of 30 mW per channel, and the corresponding η value was determined by evaluating its energy deposit as described above. The resulting time resolution measured at the test beam as a function of pseudorapidity is shown in the right panel of Figure 5.43 for different irradiation conditions: non-irradiated SiPMs, and SiPMs irradiated to 1×10^{14} n_{eq}/cm^2 and 2×10^{14} n_{eq}/cm^2, corresponding to the beginning, middle, and end of detector operation, respectively. These data can be compared with expectations obtained by scaling the time resolution terms with the energy deposit, which affects the number of photoelectrons as described in Equation 5.4. The expectation values as a function of η are normalized to the experimental measurement at $\eta = 0.2$ and are represented by the solid line in the figure. However, these expectations do not account for test-beam-specific experimental conditions, in which t_{diff} has a small residual dependence on the impact point, as discussed in Section 5.5.2. This effect is represented by the dashed line, and including it restores the agreement between data and expectation.

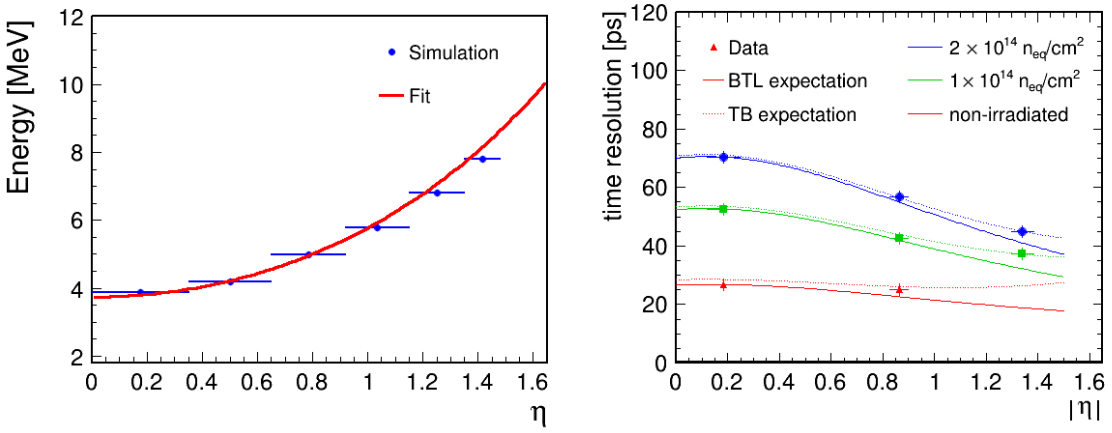

Figure 5.43: Left: Energy deposit is shown as a function of the pseudorapidity region for simulated events with minimum bias events (blue markers) and fitted with a functional form considering the variation of energy expected at varying angle in the BTL, with a MPV of 0.86 MeV/mm. Right: The time resolution is shown as a function of pseudorapidity for three different irradiation conditions, indicated with different colors. The expected BTL performance is represented by a solid line, while the test beam (TB) expectation is shown with a dashed line and the TB data with markers. The TB expectation corresponds to the BTL expectation with additional experimental effects, including the residual dependence of the time resolution estimated using the t_{diff} method on the impact point (see text for details).

5.7.2 Expected BTL performance under cumulative irradiation at the HL-LHC

The performance across the BTL lifetime was then estimated by associating different irradiation levels with the expected stages of the BTL operation. The results are shown in Figure 5.44. The uncertainty on the luminosity accounts for both the 10% uncertainty in the irradiation level and the 10% uncertainty from the annealing model (see Section 5.3.1), yielding a total uncertainty of approximately 15%. The test beam results are compared with the MTD TDR expectations, which were used to calibrate the physics studies, and are found to be within the target. The optimization campaigns described in this Chapter were shown to be effective in significantly reducing the DCR contribution by the end of operation, resulting in comparable contributions from the three terms, with the photo-statistics term dominating. Overall, these results demonstrate that, with the final design and technical specifications, the BTL sensor modules can achieve the target time resolution required for the detector to meet its physics goals during

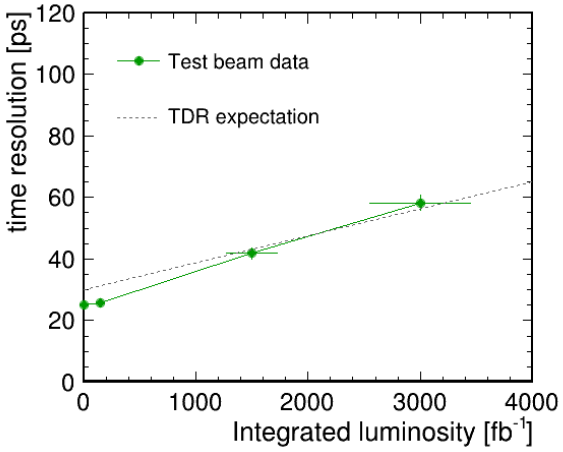

Figure 5.44: Time resolution as a function of the integrated luminosity at the HL-LHC. Data are represented by markers, and the TDR expectation by a dashed line. Uncertainties in the data points are included as error bars.

HL-LHC operation. They provide a cornerstone validation of the sensor design, confirming that the performance needed to enable the analysis improvements described in Section 4.4 is reliably met. These results also represent the first demonstration of such sensors operating under extreme radiation damage while achieving a time resolution below 60 ps. This concludes an intensive and demanding optimization campaign and marks the beginning of the production phase, which is presented in Section 5.10.

5.8 BTL electronics

The signal produced by the sensor modules is processed and routed through a complex system of electronic components. When a particle interacts with a sensor module, the resulting signal is transmitted via the SiPM flex cables to the Front-End board, where dedicated signal processing is performed. Two BTL sensor modules are connected to each FE, which houses two TOFHIR2 chips [94] that extract timing information from the current signals produced by the modules. Twelve FEs are connected to a concentrator card, which interfaces the system readout with the FEs. The CC houses two low-power gigabit transceivers associated with two Versatile Link Plus chips, responsible for sending data to and from the DAQ. It also hosts two Giga-Bit Transceiver Slow Control Adapter (GBT-SCA) chips, which monitor the low voltage, temperature, and currents of the SiPMs. Two power converter cards are connected to the CC, providing power from a low-voltage distribution system via DC-DC converters. The DC-DC output voltages are tuned and filtered by the ALDO2 voltage regulator [135], located on the FE boards (one per TOFHIR). Data produced by the TOFHIR2 ASICs are transmitted via two E-links to the two lpGBTs on the CC. The lpGBTs transmit configuration data and fast control signals to the TOFHIR2 chips and distribute 160 MHz clocks to the ASICs.

5.8.0.1 Front-End board

The BTL FE board, shown in Figure 5.45, houses two 32-channel TOFHIR2 ASICs (TSMC 130 nm) and two Adjustable Low-dropout linear regulator (ALDO) ASICs (ONSEMI 350 nm) for low-voltage and bias-voltage regulation, as well as four SiPM array input connectors. Both ASIC types have proven to be tolerant to the radiation levels expected for BTL operation. The FE is connected to the CC via an FT5 SAMTEC controller port (named J1 in the design), which routes the board power, temperature of two SiPM arrays (the ones connected to the J4 and J7 connectors), serial feeding of the TECs, the clock, fast controls, and high-speed communication links. The SiPM array input connectors, denoted as J4, J5, J6, and J7, are Zero Insertion Force connectors from Hirose, where the flex cables are inserted and clipped. The FE design also allows probing and monitoring of several signals via dedicated probing contacts on J3 (e.g. ALDO2 output bias, TOFHIR2 power supply, temperatures of SiPM arrays connected to J5 and

J7, ground, etc.). Each FE board in BTL will be identified via a four-position DIP switch, which can be set to values from 0 to 15, corresponding to a specific position in the CC. This identification number must be set during the integration stage of the DM to the tray, as discussed in Section 5.10.4.

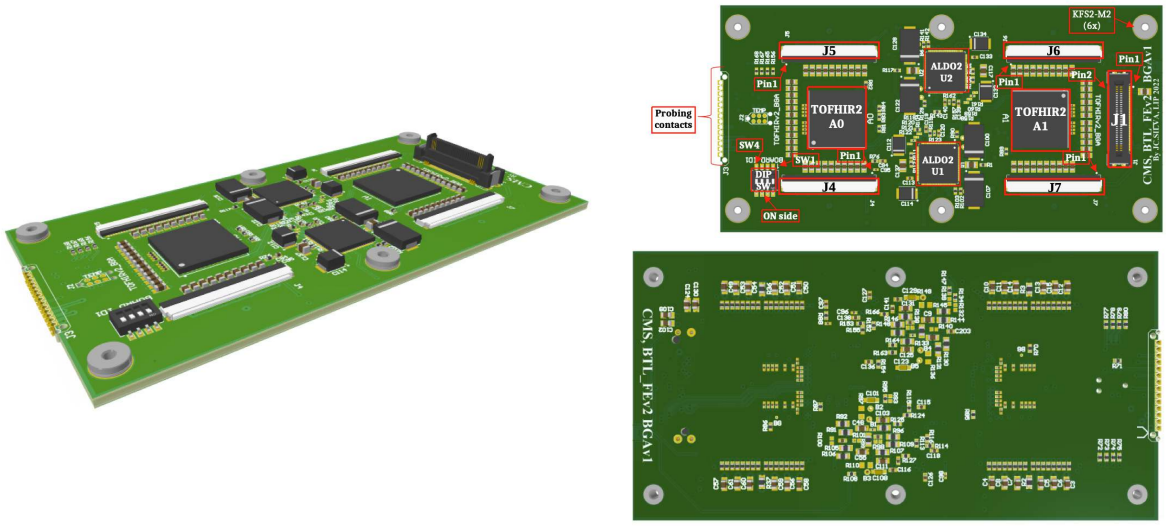

Figure 5.45: Front-end drawings are shown with a side view on the left and top and bottom views on the right, where all components are visible.

For timing applications, it is crucial to use readout amplifiers capable of handling high-bandwidth, fast signals while maintaining a good signal-to-noise ratio. When a high channel density is required, it is also important to balance performance and power consumption. All these factors must be taken into account when designing the chip responsible for the readout. To address these requirements, the TOFHIR2 chip has been developed: it is a low-power, low-noise chip that provides readout and digitization for the SiPMs. It also ensures signal amplification, shaping, and discrimination for every independent SiPM channel. Each TOFHIR2 has dimensions of 8.5 $\times$ 5.4 mm^2, assembled in a Ball Grid Array (BGA) package of 14 $\times$ 14 mm^2. It can operate with a maximum channel hit rate of 7.5 MHz, and its maximum output data rate is 320 Mbit/s. It has configurable timing, trigger, and ToT thresholds. The TOFHIR2 ASIC is required to measure MIP timing with a precision of 30–60 ps during BTL operation at the HL-LHC, mitigate the effect of DCR via a dedicated circuit, and provide a measurement of the signal amplitude with a precision better than 5% to enable time-walk corrections. The expected signal rate depends on the energy threshold: it must operate with a high MIP signal rate of 2.5 Mhit/s per channel and a rate of low-energy hits (<1 MeV) up to 5 Mhit/s per channel. It is also expected to handle the CMS Level-1 trigger and synchronization signals while consuming less than 20 mW/channel, with radiation tolerance up to the expected ionization dose (TID of 2.9 Mrad) and an integrated fluence of 2×10^{14} n$_{eq}$/cm^2.

The chip development was performed by PETsys Electronics [136], with the 10-bit SAR ADC design provided by AGH [137]. The first version, TOFHIR1, was adapted from the existing TOFPET chip [138] used for PET applications. Several developments [139, 140] led to the final production version, TOFHIR2C, which meets all design goals. The chip consists of 32 independent channels with time and charge digitizers and digital control for each SiPM channel, a service block providing bias currents, reference and monitoring features, logical control, and trigger logic that manages the chip configuration, digital data flow, and CMS L1 trigger filtering. A schematic of a single ASIC channel is shown in Figure 5.46. Each channel includes a pre-amplifier, two post-amplifiers (T and E), three leading-edge discriminators (T1, T2, and E), two ToA converters (TAC), one Charge-to-Amplitude Converter (QAC), one 40 MHz 10-bit SAR ADC, and local control logic. The input current is replicated into three branches for timing, energy discrimination, and charge integration. A dedicated pulse-filtering stage is applied after the amplifier to mitigate the effect of DCR. A Differential Leading Edge Discriminator (DLED) [141] is used to reduce DCR effects and baseline fluctuations generated by the pile-up of residual pulse tails at high

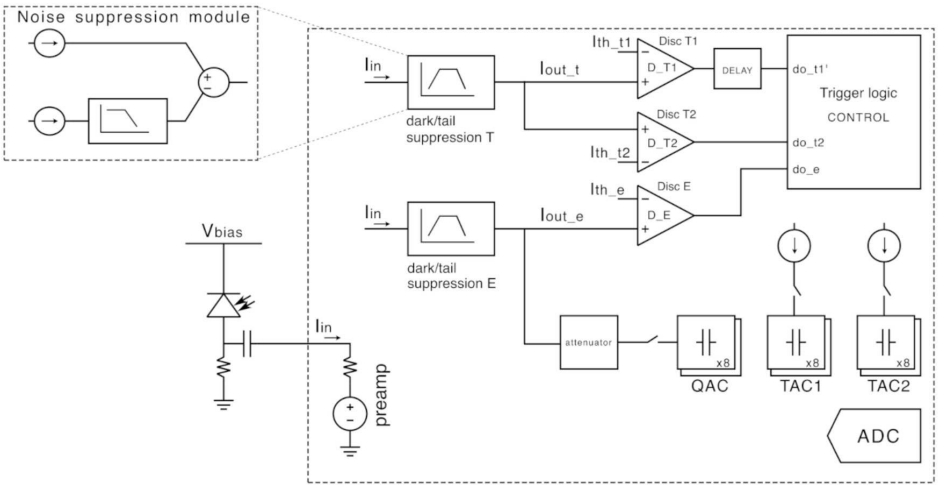

Figure 5.46: TOFHIR2 ASIC channel block diagram [94].

event rates. The technique consists of subtracting the delayed input from the original signal, resulting in a bipolar signal, as shown in Figure 5.47. Its impact was evaluated in simulations [94], demonstrating that for a DCR of 55 GHz and the signal yield expected at the end of operation of BTL, the DCR contribution to time resolution is improved by a factor of 3.5. The DLED output current is fed to the T1 and T2 comparators, with threshold currents set by Digital-to-Analog Converters (DACs) configurable with different LSB settings [0.156, 0.313, 0.47, 0.63 μA] for T1 and [0.313, 0.63, 0.94, 1.25 μA] for T2. The T2 DAC thresholds are also used for the energy branch, and charge is integrated via an active integration block (QAC). Time measurement is extracted with a TDC consisting of a coarse counter measuring the rising edge of T1 and a finer counter measuring the rising or falling edge of T2, providing the signal width. Each channel can store 8 TAC and QAC analog buffers before digitization. When a channel remains in the ready state for 100 μs, an invalid event is triggered to rearm the buffers, limiting drift due to leakage current. The TOFHIR digital design does not use the CMS clock, it has only an intrinsic coarse time tag counting clock cycles since the last Resync signal. Therefore, the backend is responsible for mapping the TOFHIR time tag to CMS event identifiers. The chip's digital design also includes an input for an external digital test pulse, which can be either externally provided or internally generated.

Extensive studies have been conducted on the chip and are reported in Ref. [94]. These studies demonstrated that the estimated power consumption of the ASIC per channel is 18.4 mW. The electronics term (defined in Equation 5.1) was measured by shining blue laser light on two bare SiPMs with a beam splitter. The coincidence time resolution was evaluated between the two as a function of the slope of the signal pulse, and by fitting the data with the expected analytical function, the noise term was measured to be $\sigma_{noise} = 0.36 \pm 0.03\,\mu A$ and $\sigma_{TDC} = 12 \pm 1\,ps$. The TDC performance was also evaluated: a binning of approximately 11.3 ps was estimated with low dispersion, overall linearity was verified, and a resolution of about 13.3 ps was measured with a dispersion of 5.3% RMS. The impact of channel rate on the time resolution was assessed via UV laser measurements, and the results showed that the performance remains stable as a function of the average pulse rate. In addition, the response in terms of signal amplitude proved to be linear both in the ranges typical of the beginning and end of operation. TOFHIR2 demonstrated stable operation under the high-rate conditions expected in BTL, with a maximum channel occupancy of about 8% for energies greater than 1 MeV. The extreme radiation conditions in which the sensors must operate also require a detailed evaluation of the impact of radiation exposure on the chips. Radiation exposure in silicon devices provokes the generation of trapped charges at oxide/silicon interfaces, causing device shifts and potentially leading to chip failure. This performance degradation depends on chip technology, process, and bias conditions, and has been studied in the same technology used in TOFHIR2 [142, 143]. Studies in Ref. [94] exposed chips to different total ionizing dose and evaluated Single Event Effects (SEEs). Irradiation campaigns exposed the chips up to 7 Mrad, and the results highlighted a decrease of about 5% in the ALDO DAC range, which was almost fully

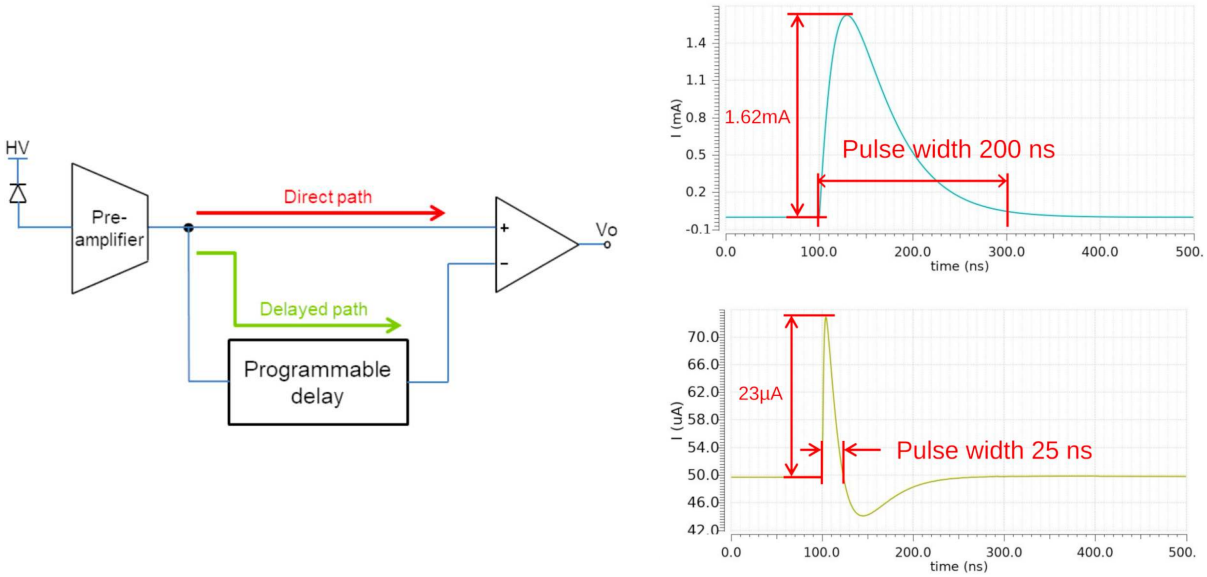

Figure 5.47: Left: The DLED block diagram is shown. Right: An example of the waveforms before (top) and after (bottom) DLED pulse shaping is shown [94].

recovered after annealing at room temperature. This recovery is representative of the pauses expected during HL-LHC technical stops, resulting in a net effect within 1%. The SEE radiation tests showed that configuration corruption with loss of synchronization and lockup of data transmission is expected to be small ($\sim 10^{-9}$) but not negligible. Therefore, BTL operation will require periodic chip reconfiguration, including reloading and resynchronization.

The TOFHIR2 supply voltage is provided by the ALDO2 ASIC, which receives it from the CC connected to J1. The ALDO ASIC has also been developed to provide the SiPM voltage bias. ALDO2 [135] is a multifunctional, adjustable, low-dropout linear regulator based on 0.35 μm HV CMOS technology from ONSEMI. It has four independent regulators: two low-voltage channels (with a maximum value of 3.3 V) used to filter and stabilize the power supply of front-end chips, and two HV channels (with a maximum voltage of 50 V), dedicated to the biasing of the SiPMs. The regulators are independent and can be turned off as a protection against over-current or over-temperature. They also allow monitoring of the bias current and performing current-voltage curves, which are crucial measurements for evaluating the performance of the SiPM arrays over the BTL lifetime. Studies demonstrated that ALDO2 maintains optimal performance across a wide operating temperature range, from -40 to 50 °C, and is not significantly affected by radiation damage. The single-event transients remain below 2×10^{-7} cm^{-2}, corresponding to less than 1 mHz under BTL operating conditions, with very small amplitude and duration of such transients. The bias voltage of SiPM arrays can be configured via the DAQ, and each ALDO2 ASIC can provide a different bias voltage per SiPM array, ensuring better handling of effects that might cause breakdown voltage variations.

5.8.0.2 Concentrator Card

The CC is designed to provide communication between the sensor readout and the FE boards. The CC houses two RAFAEL ASICs for clock distribution, lpGBTs that provide an interface between the E-links and the VTRX+ opto-module, two GBT-SCA chips, and connections for two PCCs (shown in the right panel of Figure 5.48). The CC features 12 connectors for the J1 connectors of the FE boards, connecting the DMs. It also has two TEC connectors providing power (up to 45 V) to the TEC network biasing all the DMs (blue connectors in Figure 5.48), two input bias voltage connectors for CC bias (black connectors in Figure 5.48), four bias voltage connectors providing up to 42 V for ALDO bias (green connectors in Figure 5.48), and two PCC input voltage connectors (10–12 V, black connectors in Figure 5.48). The

functionalities are explicitly written on the Printed Circuit Board (PCB) silkscreen near the connectors, and labeling is also present for the two PCCs expected to be connected to the CC: one powered at 1.2 V and the other at 2.5 V.

Figure 5.48: Left: A picture of a production CC mounted on a cold plate is shown. The connectors for cables providing TEC power, board power, and chip biasing (colored connectors) are visible. The fibers to be connected for readout via Serenity are protected with Kapton tape before connection, and the connectors on the sides, where the PCCs will later be attached, are also visible. Right: A picture of a PCC is reported.

Configurations for the FE, trigger and resync are sent from the lpGBT [106, 144] at a rate of 80 Mb/s. Clock distribution to the TOFHIR2 ASICs is also managed by the lpGBTs, with a frequency of 160 MHz. It features three paths for triggering, DAQ, and slow control information, ensuring that a single bidirectional link can be used simultaneously for data readout, trigger data, timing, experiment control, and monitoring (e.g. LV control, LV monitoring, temperature monitoring, and SiPM current monitoring). Two optical transceivers route data from the lpGBTs to the back-end with 70 m multi-mode optical fibers. Additional control and monitoring is provided in the CC via the GBT-SCA ASIC, a radiation-tolerant chip used to distribute control and monitoring signals to the FE components, along with transmitting information needed for setting environmental parameters. It also serves as a monitor for voltage, current, and SiPM temperature.

Data are processed at the TOFHIR2 level, sent to the lpGBTs in the CC, and routed to the MTD DAQ nodes. There, data are collected and sent to the CMS event builders. The MTD DAQ is expected to operate with an 8% channel occupancy, 128-bit event size, and an L1 rate of 750 kHz, corresponding to 5.5 Gb/s per RU in the BTL. The MTD DAQ will be based on the Serenity platform, an Advanced Telecommunications Architecture (ATCA) compliant platform. It provides power, clocking, optical interfaces, and electrical connections. The data coming from the CCs are managed by daughter cards with integrated FPGAs. The correct transmission and reconstruction of data sent from the TOFHIR2 to the Serenity are ensured by a set of firmware and software tools.

5.9 BTL thermal management

Section 5.3.1 highlighted the importance of smart thermal management, with operation at $T_{op} = -45^{\circ}$C and annealing at $T_a = 60^{\circ}$C to efficiently reduce the DCR. The BTL cooling is provided by the MTD cooling system, with a demonstrated thermal gradient of 2°C between the CO_2 outer pipe temperature and the SiPM. The presence of thermal gradients would reduce the benefits of cooling and negate the efforts put into performance optimization. As shown in Figure 5.38, higher temperatures are expected to degrade performance. For instance, in the full fluence scenario, a temperature increase of 2°C results in an approximate 5% deterioration of the time resolution, increasing up to 10% for a 5°C rise.

This motivates an efficient thermal coupling of all components, particularly referring to the following interfaces: the cooling plate to the cooling pipe, the CC and PPCs to the cooling plate, the copper housing to the cooling plate, the FE ASICs to the copper housing, and the TECs to the copper housing. A pictorial sketch of these interfaces is shown in Figure 5.49. In the cooling plate-to-cooling pipe interface, the

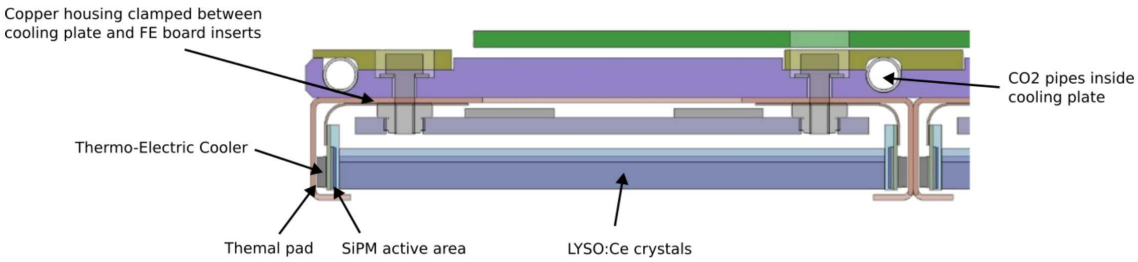

Figure 5.49: Lateral view of a BTL detector module mounted on the cooling plate. The contributions of each component are indicated in the sketch [125].

difference in thermal expansion between the stainless steel pipes and the aluminum plate is large enough to require a flexible thermal interface material, thin enough to minimize thermal resistance but thick enough to withstand thermal cycles. The best material was identified as the grease Laird Tgrease 2500. Each "cold tray", consisting of cooling pipes installed within the cold plates, requires proper greasing of the grooves where the pipes are placed during assembly. The gap between the CC and PCC and the cooling plate is 1.1 mm, with thermal pads needed under the areas where integrated circuits are located (e.g., PCCs, GBT-SCAs, lpGBTs, RTDs, etc.), as shown in Figure 5.50. The optimal material was identified as a thermal pad made of filled silicone ceramic (Laird A17916-05 [145]), with a thickness of 1.27 mm (subject to compression) and a thermal conductivity of 2 W/mK. For the copper housing-to-cooling plate interface, the best material was identified as Parker Chomerics THERMFLOW T558 [146]. T558 is a thermally enhanced phase-change material designed to completely fill interfacial air gaps and voids within electronics assemblies, covered with an aluminum foil, with a thickness of 0.115 mm and a thermal conductivity of 1.5 W/mK. Simulations of the heat generated by the ASICs in the FE and the TECs [147] demonstrated that placing two stripes of thermal material along the long lateral edges is sufficient to dissipate the heat efficiently. Different thermal pads are used for thermal coupling between the ALDO and TOFHIR ASICs and the copper housing, depending on the available thickness. Specifically, Laird A10092-02 [148] thermal material, matching the dimensions of the TOFHIR2 encapsulation, is intended to provide the required interface, while Laird A17747-03 [149] thermal pads for the ALDO2 encapsulation are intended to provide the interface. The application of these pads on the ASICs is shown in Figure 5.50.

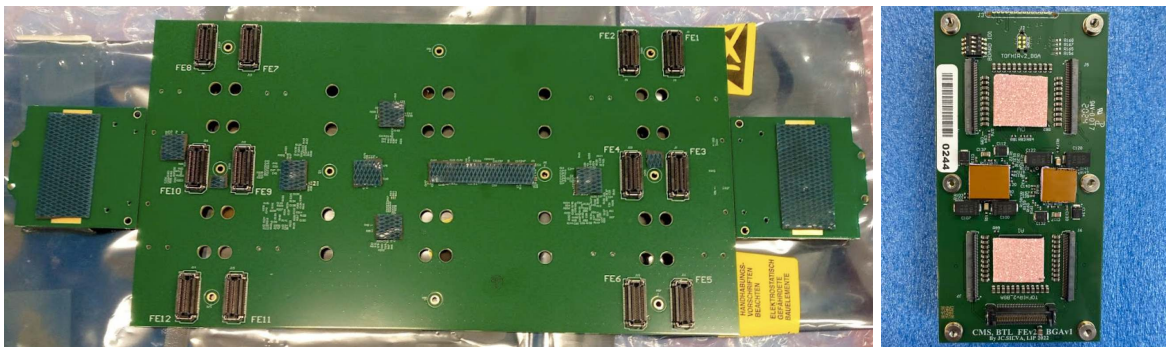

Figure 5.50: Left: Picture of the rear side of the CC (the side going in contact with the cooling plate) showing the placement of all thermal pads. Right: Picture of a Front-End showing where the thermal materials are applied to couple the chips to the copper housing.

Part of this work focused on identifying an optimal thermal interface material between the TECs and the copper housing. To guarantee proper operation of both the TECs and the SiPM, parasitic heat must be minimized. Studies were carried out on detector modules assembled with different thermal materials placed between the TECs and the copper housing wall. The initial choice, T558, revealed a risk of failure: during assembly, part of the material could be scraped into the TEC pillars (see Figure 5.51), causing electrical shorts. This occurred because the aluminum foil covering the grease could come into contact

with the electrical pads of the TECs. The short manifested as anomalous thermal behavior: in some regions of the detector module, the temperature difference under a fixed bias current was excessively high, while in others no response was observed at all (Figure 5.51), thus confirming the short in the TEC network. This failure made the TECs non-functional. To address this, an experimental procedure was defined to assess the thermal coupling of the detector module under realistic conditions. A preliminary setup is shown in Figure 5.51. The measurement consisted of mounting a detector module onto a mock-up cooling plate, emulating final conditions. The four TECs were biased with incremental current steps, each expected to produce a specific temperature difference. A reference temperature sensor was placed on the cooling plate mock-up to provide a reference temperature measurement. Any deviation from the expected response of a TEC indicated poor thermal coupling. An example of such a failure is shown in Figure 5.51. This procedure later became the basis of a large-scale protocol for quality assessment of detector modules in mass production, as it is discussed in Section 5.10.3. Alternative materials were evaluated to overcome the issue presented by T558, and among all the potential thermal materials that were evaluated, the T557 [150] was selected. It provides the same grease-based thermal material as T558 but without the aluminum foil. Approximately 70 modules were assembled and tested, and all tested modules demonstrated excellent thermal conductivity.

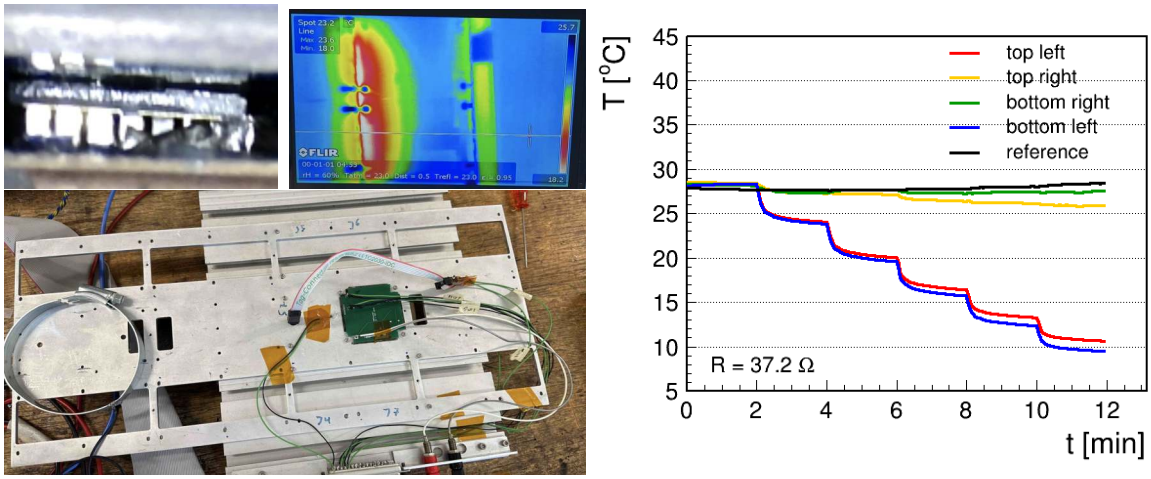

Figure 5.51: A scratch in the thermal material inside the TEC (top-left, microscope image) caused the aluminum layer covering the thermal material to contact the TEC electrical pads, creating an electrical short. As a result, shown in the top-middle thermal camera image, one side of the DM experienced the full voltage difference, leading to excessive cooling on that side while the opposite side remained unaffected. The temperature difference produced by the four SiPM arrays in the DM was measured as a function of the current applied to the TECs using the setup shown in the bottom-left, where the module was mounted on a cooling plate mockup and properly powered and read out. The plot on the right shows that, as expected, the temperature of one side changes with the current, while the other side remains essentially unchanged.

5.10 The large mass-production of the BTL

Following the successful demonstration of the detector performance targets with the optimal design, full mass production began in July 2024 and is currently ongoing. Four BTL Assembly Centers (BACs) were identified for mass production: the University of Beijing (PKU), the California Institute of Technology (Caltech), the University of Milano-Bicocca (MIB), and the University of Virginia (UVA). Each center is responsible for producing approximately one quarter of the total number of sensor modules, detector modules, and trays. Trays are first validated at each BAC, then shipped to CERN for further testing, and eventually integrated into the BTST.

Each step of the production process requires specific protocols to ensure uniformity across the assembly centers and, consequently, throughout the BTL system. QA/QC procedures were defined during

dedicated workshops at CERN, to which I also contributed, where members and experts from all BACs established common standards for production and qualification. Individual sensors are produced by selected vendors according to the specifications described in Section **??**, with the vendors also responsible for their qualification. In addition, further QA/QC is performed on a sampling basis by dedicated teams: the University of Debrecen and the University of Notre Dame are responsible for SiPM QA/QC, while the INFN-Rome team performs measurements on the crystals. Sensor modules are assembled in the BACs, where well-defined procedures are applied for gluing LYSO matrices to SiPM arrays and for verifying their performance, as discussed in Section 5.10.2. Two sensor modules are then connected to the front-end board and housed in a copper housing, which provides both mechanical support and optimal thermal coupling with the cooling plates, forming a detector module. DM assembly requires the application of several thermal interface pads to ensure a minimal thermal gradient between the SiPMs and the cooling plates, as discussed in Section 5.9. The efficiency of such thermal coupling is assessed via specific tests, as described in Section 5.10.3. Finally, twelve detector modules are mounted on one side of a cold tray, consisting of cooling plates with embedded cooling pipes, with the CCs and PPCs mounted on the opposite side, forming a RU. Each tray is composed of six RUs, and the procedures for tray assembly and QA/QC are described in Section 5.10.4.

5.10.1 Sensors mass production and qualification

5.10.1.1 LYSO:Ce crystals

Full production of LYSO:Ce crystals started in December 2023 and continued until December 2024. QA/QC was performed by the INFN-Rome team on individual crystal bars and arrays, sampling two bars from the top, middle, and bottom of each ingot, representing 5% of the total produced arrays. Building upon their experience with characterization studies on samples from various vendors prior to the final selection, they defined robust and specific protocols for the QA/QC measurements of production samples. Measurements included dimensions, light output, decay time, optical cross-talk, and time resolution using 511 keV photons from a ^{22}Na source in a dedicated experimental setup, as described in Ref. [151]. QA/QC of all batches has been successfully completed, showing overall good uniformity of performance across all samples. Measurement reproducibility on bars is 5.5% for light output and 0.22 ns for decay time, with uniformity within 1.2 ns, meeting the specifications. Results are shown in Figure 5.52. Dimensional tolerances of the LYSO arrays were also verified with a Coordinate Measuring

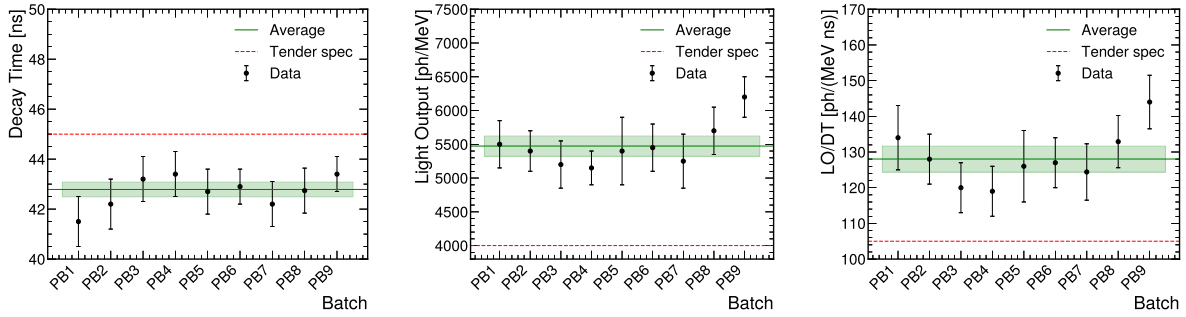

Figure 5.52: Measurements were conducted by the LYSO QA/QC team and results of QA/QC of LYSO:Ce bars from different production batches (x-axis) are shown in terms of decay time (left), light output (middle), and the ratio of light output to decay time (right). For each batch, the data are presented as the average value, while the associated RMS is used to represent the error bar. All batch results fall within the technical specifications outlined in tender process, represented by the red line, while the average is indicated by the green band.

Machine (precision 4 μm). All batches demonstrate good compliance with the specifications. Optical properties of the LYSO arrays were measured with a ^{22}Na source. Results show an average light yield uniformity of approximately 3.5%. Also, irradiation tests with 50 kGy from a ^{60}Co source show an 8% decrease in light output with 2% spread, well within the <20% specification. Results are shown in Figure 5.53. Production of all 10,368 LYSO arrays (165,888 crystal bars) and their qualification

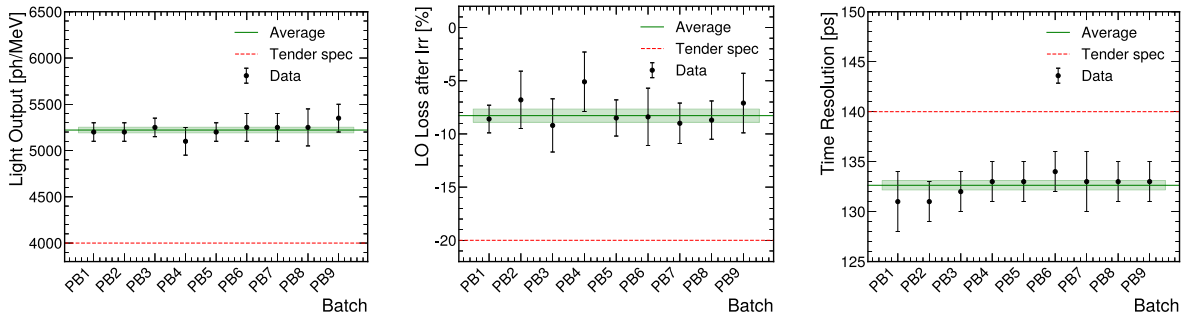

Figure 5.53: Measurements were conducted by the LYSO QA/QC team and results of QA/QC of LYSO:Ce arrays from different production batches (x-axis) are shown in terms light output (left), light loss after irradiation (middle), and time resolution (right). For each batch, the average value is shown with the RMS as the error bar. All batches meet the technical specifications defined in the tender process (red line), while the overall average is indicated by the green band.

successfully met the BTL requirements, with a timely campaign of qualification.

5.10.1.2 SiPMs

The QA/QC measurements of the SiPMs were carried out by two independent institutes, the University of Debrecen and the University of Notre Dame, with their setups shown in Figure 5.54. The measurements were cross-calibrated using samples tested by both teams. Good correlation was observed for all samples. While QA/QC for LYSO crystals was performed on selected samples from each ingot, several

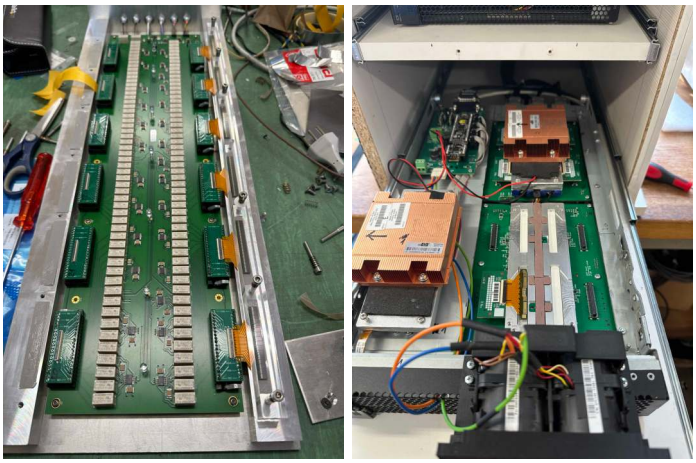

Figure 5.54: Pictures of the setups used by the two teams dedicated to SiPM QA/QC, in which a dozen SiPM arrays are loaded and a dedicated automation system performs all the necessary characterization measurements for each array [152].

measurements were carried out on every SiPM array: the breakdown voltage measurement at 20°C, the measurement of the breakdown voltage spread within each array, the TEC resistance for each TEC, and the dark current at different breakdown voltages. Other properties were measured on a sampling basis: breakdown voltage at -35°C, PDE, gain, ECF, single photoelectron charge, performance after irradiation and accelerated annealing, and performance after 100 thermal cycles between -50°C and 70°C. Results of the breakdown voltage and its spread within the array are reported in Figure 5.55. Production of all 20,736 SiPM arrays (331,776 channels) and their qualification successfully met the BTL requirements, with an achieved yield larger than 99%.

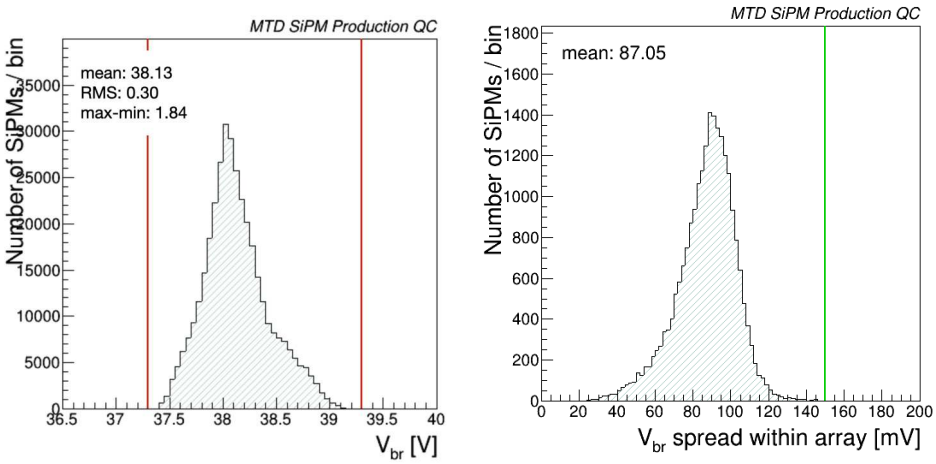

Figure 5.55: Measurements were conducted by the SiPM QA/QC teams and results of all the production SiPM arrays in terms of the measured breakdown voltage of each array and the spread of breakdown voltages of the SiPMs within the arrays. Both results demonstrate that they lie within specifications and exhibit good uniformity across all production [152].

5.10.2 Sensor Modules assembly and QA/QC

Figure 5.56: Pictures of LYSO:Ce crystals (left) and SiPM arrays (right) used for production at the Milano-Bicocca assembly center.

The gluing procedure required careful optimization, as it is critical for the light output performance (see Equation 5.5). Both the alignment between LYSO crystals and SiPMs and the amount of glue used are particularly important. A dedicated mechanical tool was designed to position the LYSO matrix on a horizontal plane. Lateral rotating walls allow precise alignment of the SiPM packages, via a system of springs and a rotating stage, the walls can be brought into proximity with the LYSO crystal end faces. The full assembly procedure is documented in Figure 5.57. The LYSO matrix is positioned on the horizontal plane, while the lateral walls, initially at rest, ensure that the SiPMs are correctly placed in their slots. Dedicated metallic masks, with holes corresponding to the active surface of the SiPM arrays and a thickness of 100 μm, a value that determines the amount of glue used between SiPMs and crystals and has been optimized based on LCE studies, as reported in Section 5.4, are placed on top of the SiPM arrays and held in place by magnets on the lateral walls. A syringe with a fixed volume of RTV glue is used, and a plastic head ensures that the drop of glue is consistent, avoiding bubbles or air pockets. Three drops of glue are applied at equidistant positions along the SiPM array. A dedicated plastic applicator is then clipped to the lateral walls, and a soft plastic lever spreads the glue uniformly across

the entire SiPM surface. The lateral walls are rotated to face the crystal bars and locked in position to prevent further rotation. A system of springs pushes the SiPM arrays against the LYSO matrix. This setup required extensive development, with each step optimized to maximize reproducibility. The gluing tool was designed with high precision to ensure optimal alignment and good coupling between the SiPMs and the crystals. The sensor module is then left under spring tension for one day to allow the glue to cure.

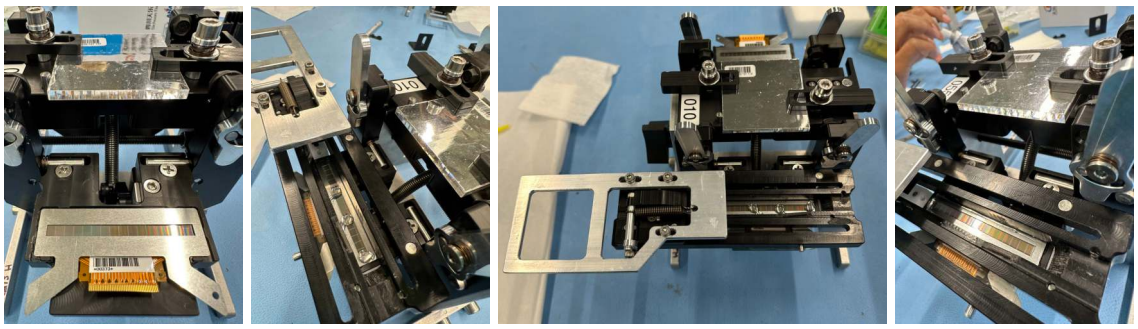

Figure 5.57: Four illustrative pictures of the gluing procedure are shown. From left to right: the positioning of the LYSO array and the SiPM array in their respective locations is shown first, followed by the placement of the metal mask on top of the SiPM. In the second picture, the tool used to apply the glue is shown, with the three glue spots already deposited on the SiPM to be spread. In the third picture, the lever is shown starting to move, and finally, in the last picture, the applied glue layer is visible.

Qualification of the sensor modules is aimed to verify the number of photoelectrons measured at each SiPMs, which directly determines time resolution (Section 4.1). Uniformity across the module is also verified to guarantee consistent performance across channels. As discussed in Equation 5.4, the number of photoelectrons depends on the light yield of the crystals (verified during crystal qualification), the PDE of the SiPMs (verified during SiPM qualification), and the light collection efficiency, which mainly depends on the gluing process (alignment, glue amount, etc.). Measuring the number of photoelectrons under controlled conditions provides a probe to evaluate the quality of the gluing step. A ^{22}Na source is used for the light output measurements, evaluating the charge produced by the interaction of the 511 keV photons. The total charge produced is normalized by the charge corresponding to a single photoelectron, which can be measured from the distance between peaks in the charge integral spectrum. For QA/QC, dark single photoelectrons, arising from thermal effects or random fluctuations, are used. To perform these measurements, a dedicated tester board, the BTL QA/QC jig, was developed. It can test up to 12 LYSO matrices simultaneously, as shown in Figure 5.58. Each sensor module is connected via flex cables to Panasonic connectors. The signal is routed to an amplifier board, where a transimpedance amplifier converts the current pulse into a voltage signal. The voltage is then amplified by a factor of 4, followed by a 4× gain reduction stage and an unattenuated path, allowing measurements of both single photons and MeV-scale energy deposits. Signals are finally processed by a CAEN DT5742 digitizer with a fixed dynamic range. The same jig, acquisition system, and analysis software are used across all BACs, ensuring consistency.

QA/QC results classify modules into two main categories: those considered suitable for production and those corresponding to rejected BTL SMs. A module is considered to fail if more than two bars meet any of the following criteria: a channel has a light output less than 0.8 times the reference average light output, where the reference average is derived from measurements of the first production modules; a bar has a light output less than 0.85 times the reference average light output; a bar exhibits an absolute left-right asymmetry greater than 0.2; or a bar has an energy resolution greater than 6%. In addition, a module is considered unacceptable if either of the following global criteria is violated: the measured single photon electron charge deviates by more than 10% from the module-average value, or the mean asymmetry of all bars in the module exceeds 0.08. Comparisons among the four BACs demonstrated compatible results. As of writing, the Milano-Bicocca BAC has produced and qualified 2536 BTL SMs, as reported in Figure 5.59, representing the 97.8%. The average light output amounts to approximately 3230 photoelectrons/MeV at 3 V overvoltage, with a spread of about 5%. The average

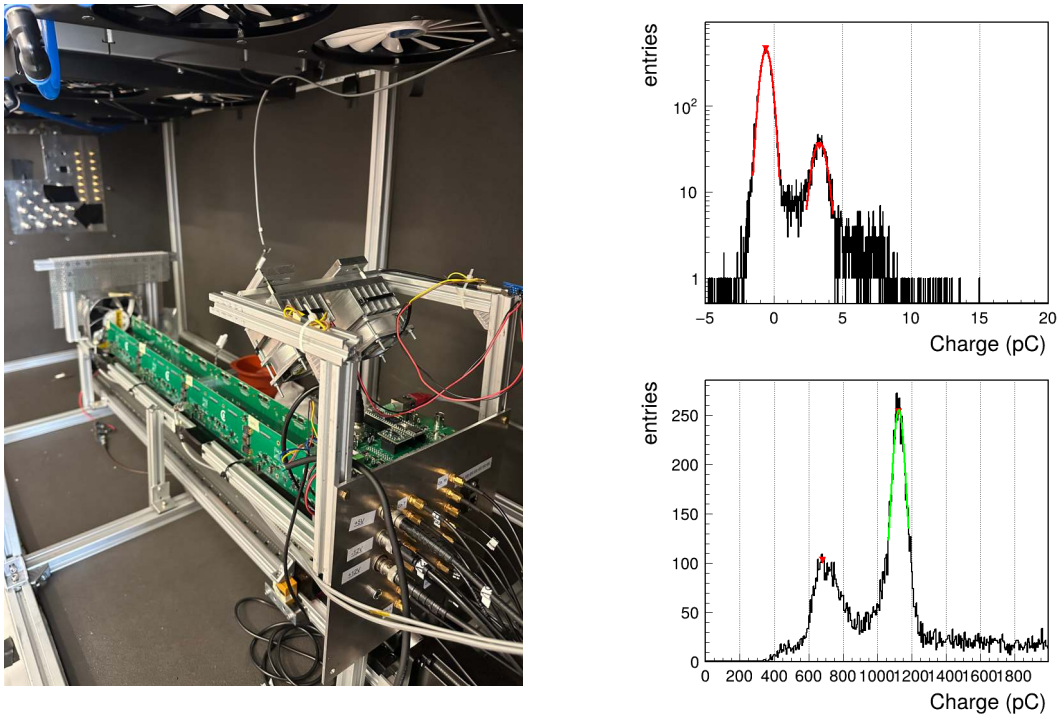

Figure 5.58: Left: The BTL QA/QC jig used for sensor module measurements is shown. The setup is housed inside a temperature- and humidity-controlled, light-tight box to ensure stable and reproducible conditions. Right: Examples of the single photoelectron charge spectrum and the source deposition energy spectrum are shown, with the peaks fitted using Gaussian functions.

single-photoelectron charge is measured to be 3.95 pC, with an uncertainty of 1.5%. Comparable results have been obtained at the other assembly centers, which have so far produced 1962 SMs at Caltech, 2651 at PKU, and 2601 at UVA. Overall, about 95% of the qualified modules pass the selections defined for being considered suitable for production.

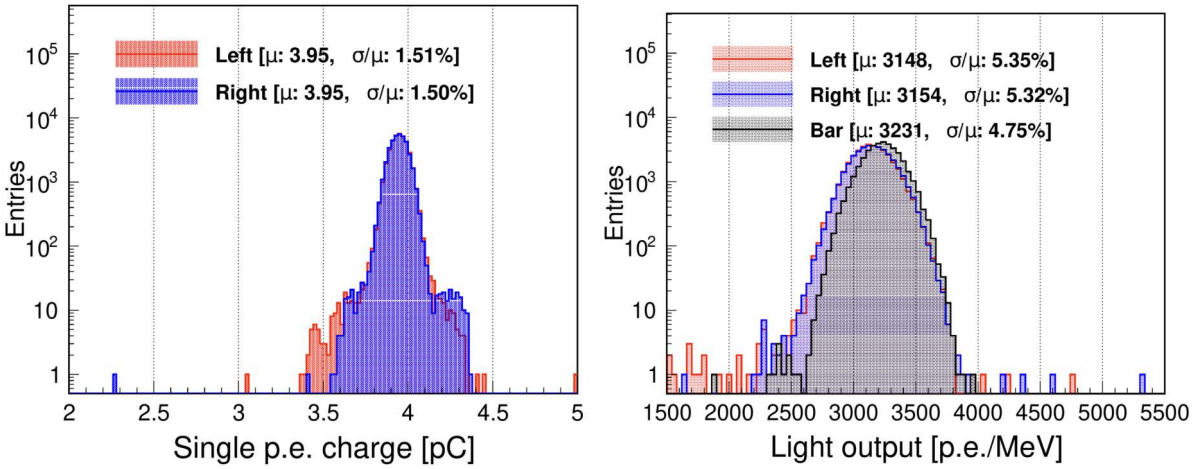

Figure 5.59: Single photoelectron charge and light output results for all the sensor modules produced at the Milano-Bicocca BAC are shown on the left and right, respectively. The results are split into the left and right sides of the bar (red and blue, respectively), with the average of the bar represented in black. The mean and uncertainty are also reported for each distribution.

5.10.3 Detector Modules assembly and QA/QC

Qualified sensor modules with comparable performance are assembled into a detector module: BTL SMs are connected to the FE and inserted into the copper housing, whose main purpose is to ensure good thermal coupling and mechanical stability. The functionality of the TECs in each BTL sensor module is verified using an LCR meter, which measures frequency-dependent electrical parameters, with an expected resistance of approximately 13 Ω for TECs. The assembly procedure involves several steps, with an example of assembled DM illustrated in Figure 5.60. The FE board is first equipped with dedicated thermal pads: two for the TOFHIR2 ASICs and two for the ALDO2 ASICs. Additionally, two foam strips are glued beneath the FE for thermal and mechanical support. Two sensor modules are then connected to the FE via flex cables. The internal walls of the copper housing, which are in contact with the TECs, are covered with two strips of T557, cut to the length of the housing and wide enough to cover the entire TEC surface. A dedicated tool was designed to slightly bend the copper housing, allowing insertion of the FE and SM system while ensuring proper alignment and reducing the risk of scraping the TECs against the copper or thermal material. Once the sensor modules are inserted, the tool is released, closing the six feet that press against the LYSO matrices. This secures the LYSO and SiPM arrays inside the mechanical support, minimizing any stress on the TECs[8].

Figure 5.60: Each detector module is assembled from two pre-assembled and qualified sensor modules, which are subsequently connected to the front-end and inserted into the copper housing. A picture of a sensor module and of the two sides of a detector module are shown from left to right, respectively.

As discussed in Section 5.9, ensuring good thermal coupling at all interfaces is crucial to reach the lowest possible SiPM operating temperature and thus the smallest DCR level. To verify thermal connection quality, each detector module is mounted onto a cold plate and connected to a customized board capable of reading the temperatures of the four SiPM arrays and supplying current to the TECs. The TECs are biased with a fixed current of 0.25 A and allowed to reach thermal equilibrium within about 4 minutes. Both the absolute temperature drop and uniformity among the four sides are evaluated to assess thermal coupling quality. The threshold required to pass these tests is defined in terms of temperature drop, which is required to be smaller than -18°C. As of writing, the Milano-Bicocca BAC has produced and qualified 1256 DMs, as reported in Figure 5.62, corresponding to 89% of the total production. An average temperature difference of -19.2°C is measured, with an uncertainty of 1.4%, and a spread of 0.21°C among the four sides of the DM is observed. Similar results are obtained at other assembly centers, where the current production amounts to 810 DMs (62.5%) at Caltech, 519 (40%) at PKU, and 1210 (93%) at UVA.

5.10.4 Tray assembly and QA/QC

Prior to handling any electrical components, strict electrostatic discharge precautions were observed, ensuring that both the operator and all parts were properly grounded. The assembly of the BTL trays begins with the preparation of the cold trays, consisting of qualified cold plates and cooling loops, onto which the CCs and PCCs are mounted. Thermal pads were placed under the integrated circuits, and

[8]TECs can withstand forces up to a maximum of 10 N.

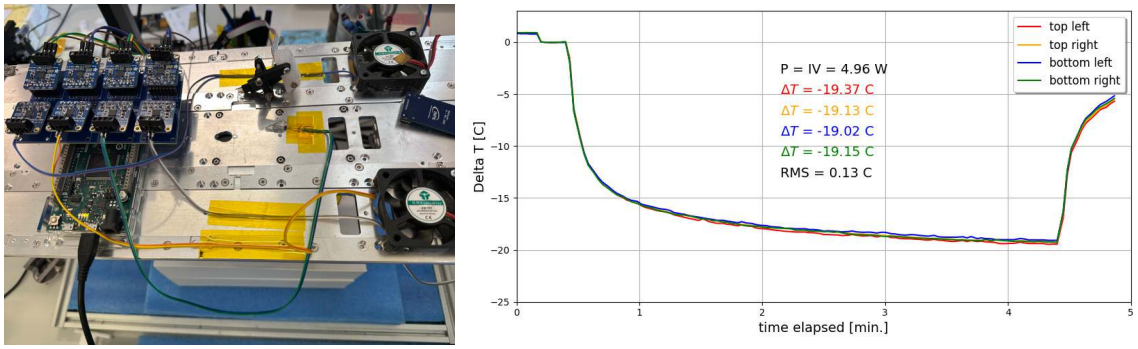

Figure 5.61: A picture of the detector module QA/QC setup used at the Milano-Bicocca assembly center is shown on the left. The DM is mounted on the underside of the cooling plate, which is not visible, while above it the various cables connected to the board are visible; this board attaches to the front-end and allows temperature readout. A fixed current is then applied to the TECs to verify the uniformity of the four SiPM arrays and the absolute value reached in terms of temperature difference after a defined period.

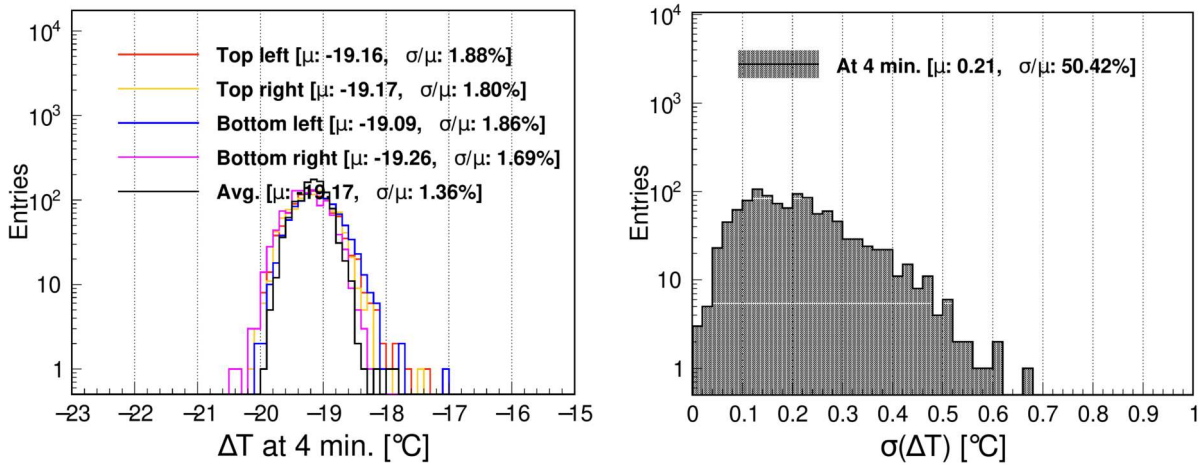

Figure 5.62: QA/QC results from the Milano-Bicocca BAC on assembled detector modules are shown. On the left, the results of the temperature difference measured at each of the four sides are presented, along with the average. For each distribution, both the mean and the uncertainty are reported. On the right, the spread of the average temperature difference is shown for all the assembled detector modules.

specific spacers were positioned between the boards and the cold plates, as illustrated in Figure 5.50. The boards were secured using washers and screws to guarantee stable mechanical and thermal contact. All steps of the assembly process, from sensor qualification to the integration of DMs and CC mounting, were systematically documented in the MTD database, ensuring full traceability of each component throughout production. The precise positioning of each CC along the tray was recorded and DMs were mounted on the opposite side of the cold plate using dedicated insertion pins to guarantee proper alignment with the front-end connectors. The configuration of the front-end within each CC, including switch settings, was adjusted and logged, establishing the calibration and communication parameters for the final DAQ system. Cabling was performed as the final stage of the assembly. Custom-designed plastic holders were used to route power, low-voltage, high-voltage, and sensing lines for each readout unit along the tray. Additional plastic encapsulation secured the fiber pig-tails connecting the DAQ and lpGBT fibers, ensuring mechanical stability and protection. An example of a fully assembled tray at the Milano-Bicocca assembly center is shown in Figure 5.63.

Figure 5.63: A photograph of a detector module being installed in the tray is shown on the top left, the completed installation with electronics and cables in place is shown on the top right, while the final assembly with all detector modules mounted on one side of the tray is shown in the bottom,

At the time of writing, the tray QA/QC relies on a preliminary version of the BTL DAQ, which provides communication between the Serenity system and the CCs. This setup enables slow-control functionalities, including current and temperature monitoring, TDC resolution measurements, evaluation of energy peaks using the intrinsic radioactivity of LYSO crystals, and assessment of TOFHIR noise. Calibrations of the front-end ASICs are performed for each tray, including ALDO calibrations that provide the bias voltage to the SiPMs and enable measurement of their IV curves. The breakdown voltage of SiPM arrays can thus be determined in situ, allowing correlations with the QA/QC measurements performed on individual sensor modules. An example of such a correlation is shown in Figure 5.64, demonstrating good agreement between the two datasets. Similarly, energy spectra measured at the tray level can be compared with those obtained during the BTL SM QA/QC, providing additional validation of the assembly quality. A convenient source of radiation for these studies is the intrinsic radioactivity of LYSO crystals. The ^{176}Lu isotope undergoes β^- decay with an endpoint of 593 keV to ^{176}Hf, which subsequently de-excites through emission of γ-rays at 88, 202, and 307 keV. Analysis of these intrinsic peaks offers a reliable measure of light output, enabling direct comparison with the ^{22}Na-based measurements performed during the sensor module QA/QC stage. An example of this correlation is reported in Figure 5.64, showing consistent

agreement between the two sets of measurements. Finally, the digital test pulse integrated in the front-end (see Section 5.8) provides a measurement of the TDC resolution, allowing a thorough assessment of front-end performance.

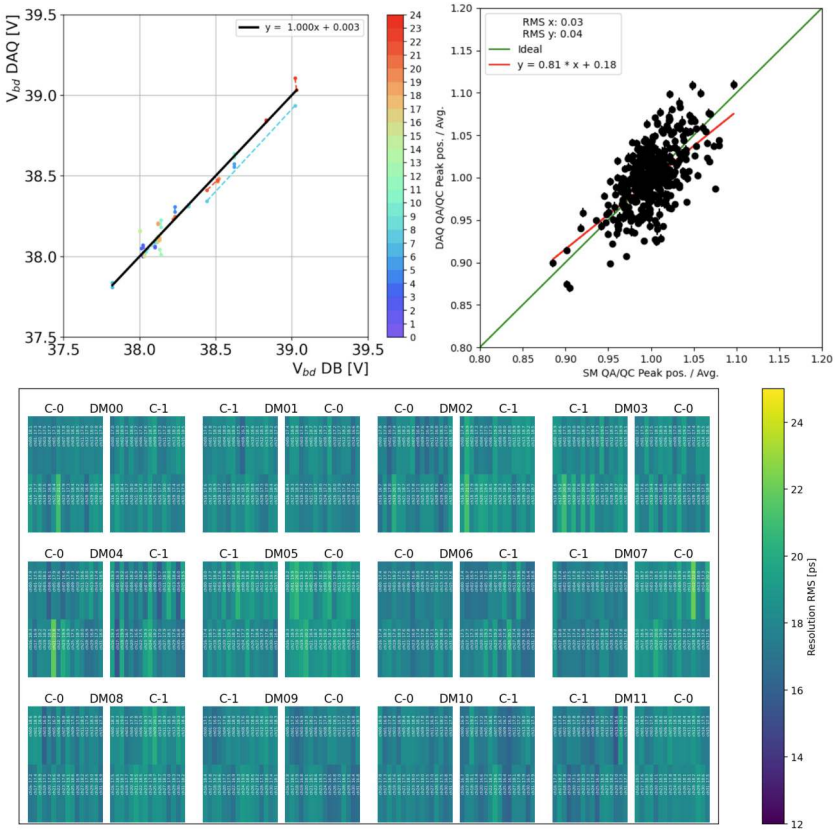

Figure 5.64: An example of the correlation between the breakdown voltage measured during the SiPM QA/QC phase and the breakdown voltage measured in situ with the DAQ readout is shown on the left, highlighting a good agreement between the two measurements. In the middle, an example of the correlation between the energy measurement performed during the BTL SM QA/QC and the energy measurement performed in situ on the tray using intrinsic radioactivity is shown, also demonstrating a good level of correlation with an RMS of 5%. Finally, the RMS resolution of the TDC measured in situ with the DAQ is shown for a readout unit as an example, exhibiting good uniformity across all channels.

5.11 Plans for integration and calibration in CMS

At the time of writing, the BACs have completed assembly and QA/QC for 90% of the sensor modules, 70% of the detector modules, and 25% of the trays. Tray insertion into the BTST is scheduled to begin by the end of January 2026, with full insertion expected to take approximately one year, at which point BTL access to the BTST will end, to allow for tracker integration. Each assembled tray is shipped to the Tracker Integration Facility at CERN, where it is tested in a test stand with CO_2 cooling before being inserted into the BTST, when tray QA/QC measurements will be re-performed to ensure nothing got damaged during shipment. After confirming proper operation in the CO_2 test stand, each tray is covered on the LYSO side with an Airex insulator. Tray insertion follows a procedure extensively validated using a BTST mock-up: tooling has been designed and produced to allow rotation of the tray cradle within the specified slot using the worm gear unit, followed by fine adjustment of the rail alignment. The cradle is then slid along z to couple the cradle rail to the I-beams inside the BTST, locked in z, and the tray is inserted. Cable management and BTL services (e.g. CO_2, power, and data link connections) have also been defined, reproducing the final setup at LHC P5. However, their development is beyond the scope

of this work, further details can be found in Ref. [93].

Qualification of the detector at every production stage ensures the expected performance, as evaluated in the test beam campaigns presented in Section 5.7, where results demonstrate that the time-resolution target can be achieved. To maintain the time resolution, it will be crucial during LHC operation to synchronize thousands of channels with high precision In CMS, synchronization of the relative time response across the detector and time-walk corrections will be derived in situ using an inclusive sample of tracks. Preliminary studies reported in [93] indicate that frequent, granular monitoring and calibration will be feasible and are not expected to limit MTD performance. Synchronization is required to ensure coherent timing across all channels, along with the ability to optimize detector response under evolving operational conditions. Two types of calibrations can be identified: low-level calibrations involving DAQ and electronics, and high-level calibrations addressing channel synchronization, and alignment. Low-level calibrations consist of front-end procedures, such as discriminator threshold calibrations to evaluate baseline and noise, TDC calibrations to determine the phase of digital test pulses relative to the clock, QDC calibrations for charge integration, and IV calibrations[9]. High-level calibrations require precise channel synchronization at the few-ps level, both among MTD channels and with the calorimeters. For this purpose, all tracks collected at the HLT can be used for inter-calibration. The mean of the distribution of reconstructed track times at the vertex can serve as a calibration reference. Preliminary studies [93] have demonstrated that with an allocation of 1 kHz of HLT rate and an average occupancy of 5% for 200 PU events, the required 1,000 or 10,000 tracks can be collected in about 20 or 200 s per channel, enabling frequent and granular calibrations.

Currently, alignment algorithms are being developed, building upon the software already used for other CMS systems, with the goal of establishing an alignment procedure with respect to the tracker using muon tracks. Spatial alignment is not considered critical, given that the average hit density is below one track per 10 cm^2 and the resolution of the track extrapolated to the MTD is at most of the order of 1 mm. The difference between tracker tracks extrapolated to the MTD and actual hits in the MTD will be fitted with alignment constants that best describe this difference. The precision achievable with this method is much better than the spatial resolution of the MTD sensors, and therefore not a limiting factor. This procedure can be frequently repeated to monitor movements relative to the tracker.

While many developments on the software side are still ongoing, production is expected to be completed next year. The next challenge will be to bring this detector into full operation within CMS, efficiently collecting timing information with a resolution shown to remain within 60 ps, with uniform performance across the 38 m^2 sensor surface. The BTL represents a scientific revolution, not only in terms of technologies never before deployed at this scale, with this level of precision and radiation tolerance, but also in terms of concept: it introduces, for the first time in a HEP detector, the "fourth dimension", bringing a paradigm shift in the way we reconstruct and interpret events in CMS.

[9]The SiPM breakdown voltage is expected to change with operating temperature, and setting the correct overvoltage is critical for performance, as has been extensively demonstrated.

Conclusions

The study of the Higgs sector remains one of the most powerful probes of the Standard Model, offering direct sensitivity to the shape of the Higgs potential and to possible extensions beyond the Standard Model. In particular, the measurement of the Higgs boson self-coupling, accessible only through non-resonant di-Higgs production, represents a cornerstone of Higgs measurements and a major experimental challenge for current and future high-energy colliders.

This work presented two complementary efforts toward this goal. The first focused on the analysis of di-Higgs production in the $HH \to bb\tau\tau$ final state using LHC data as they became progressively available. Starting from the Run 2 dataset, corresponding to an integrated luminosity of 138 fb^{-1} collected at a center-of-mass energy of 13 TeV, a new boosted category was developed and optimized, leveraging advanced machine learning based taggers to identify highly collimated Higgs boson decays in $b\bar{b}$ pairs within the resonant search for $HH \to bb\tau\tau$. This analysis targets the resonant production of HH coming from the decay of heavy spin-0 and spin-2 resonances predicted by BSM theories, with masses ranging from 250 to 3000 GeV. At large resonance masses, where the decay products of the Higgs bosons become increasingly collimated, the boosted category provides a valuable gain in sensitivity compared to resolved topologies. This improvement is visible in the expected 95% CL upper limits, which show that above 1.5 TeV the boosted category becomes the dominant contributor to the overall sensitivity, demonstrating the effectiveness of the dedicated tagger in this challenging region of phase space. With the first data from Run 3 becoming available, the non-resonant search for $HH \to bb\tau\tau$ was developed, together with the optimization of a dedicated boosted reconstruction, building on the methods and experience developed in the resonant search. This analysis uses 62.4 fb^{-1} of proton–proton collisions collected at a center-of-mass energy of 13.6 TeV with the CMS detector. It includes several improvements introduced with Run 3, such as new triggers, updated tagging algorithms, and implementation of an optimized boosted $H \to bb$ category. A dedicated working point for the tagger addressing boosted topologies of the $H \to bb$ decays was defined by evaluating the optimal operating point that maximize the overall sensitivity of this search. Moreover, a deep neural network specifically trained for the boosted topology was optimized to provide the final discriminant for the statistical analysis, achieving strong separation between signal and background processes. The expected upper limit at 95% CL on the inclusive $HH \to bb\tau\tau$ cross section results to be 5.7 times the Standard Model prediction, compared to 5.2 obtained in Run 2 with twice the integrated luminosity. Constraints at 95% CL on the Higgs self-coupling modifier κ_λ and HHVV coupling κ_{2V} were obtained as $-3.75 < \kappa_\lambda < 10.5$ and $-0.75 < \kappa_{2V} < 2.95$, respectively, both consistent with SM expectations. Although the boosted category is currently limited by statistics, its inclusion strengthens the robustness and reach of the analysis in the high-momentum phase space. The developments presented here lay the foundation for future searches at the HL-LHC, where the increased dataset will fully exploit the potential of the boosted topology in the searches for di-Higgs production.

The second component of this work focused extensively on the technological developments required for the CMS upgrades at the HL-LHC, in particular on the design and production of the barrel section of the MTD. This detector is designed to face one of the main experimental challenges of the HL-LHC, i.e. the reconstruction of events under extreme pileup conditions, by providing precise timing information for charged particles. The integration of timing measurements with a resolution of about 30–60 ps in CMS is expected to significantly enhance vertex reconstruction and pileup mitigation, with an expected corresponding improvement in the sensitivity of physics analyses such as di-Higgs searches, equivalent to an effective gain of 2–3 additional years of data-taking. This thesis presented a comprehensive development and characterization of the sensors constituting the Barrel Timing Layer. The R&D, validation, and

production of the BTL sensor modules, made of LYSO:Ce crystal bars coupled to SiPMs, were carried out with the goal of achieving the target time resolution required by the MTD design. A critical challenge of the design arises from the performance degradation expected at high fluences up to 2×10^{14} n_{eq}/cm^2, which induces time jitter due to dark counts in the SiPM. To mitigate this, two complementary strategies were implemented: an optimized thermal management based on thermoelectric coolers, and an improved sensor design to increase the signal amplitude. The TECs allow SiPM operation at a temperature of -45°C, reducing the DCR during operation by approximately a factor of two, and provide 60°C during technical stops for thermal annealing of radiation-induced damage. The presented work was dedicated to comprehensive R&D studies aimed at optimizing sensor performance compared to the baseline design proposed in the MTD TDR. Studies based on laboratory and test beam measurements explored the impact of SiPM cell size, sensor thickness, operating temperature, and irradiation levels on signal amplitude and time resolution. The main contributions to the overall time resolution were identified and modeled, including photo-statistics, electronics noise, and DCR noise, the latter becoming increasingly dominant with higher irradiation. Performance with SiPM cell sizes of 15, 20, 25, and 30 μm was systematically studied, demonstrating that larger cell sizes improve PDE and gain, resulting in a larger number of photoelectrons and steeper signal slope, improving all the terms contributing to the time resolution. Also, results from irradiated SiPMs demonstrate that, despite the DCR increase with cell size, the net effect of the increased signal-to-noise ratio leads to overall improved performance. A detailed analytical model was developed to quantify the dependence of the time resolution on PDE and DCR, across different configurations, resulting in a deep understanding of the factors defining the final performance and representing a crucial validation of the characterization campaign. Additionally, thicker sensor modules were evaluated in beam tests, showing an improvement in total time resolution that becomes particularly significant for irradiated SiPMs, where it scales approximately with the square root of thickness, consistent with the expected increase in the number of photoelectrons. Comparative studies of sensors irradiated to different fluences and operated at varying temperatures quantified the DCR dependence on both irradiation and temperature. These studies confirmed that DCR increases more than linearly with fluence and roughly doubles every 10°C, with a stronger temperature dependence observed for highly irradiated SiPMs, where a 5 ps change in time resolution was observed for a temperature variation of about 5°C. This highlights the critical importance of thermal management as a strategy to mitigate DCR, the only term in the time resolution explicitly dependent on temperature. This intensive R&D campaign provided a deep understanding of the factors influencing timing performance under HL-LHC conditions. The results guided the final design choice for the BTL: sensor modules 3.75 mm thick with 25 μm SiPM cell size.
This final configuration demonstrated a time resolution of 25 ps for non-irradiated modules operated at 3.5 V overvoltage, degrading to 55 ps for irradiated modules operated at 1 V, consistent with end-of-lifetime conditions. This configuration achieves superior performance compared to the TDR proposal, meeting the target time resolution within the allowed power budget for SiPM operation. The response was found to be uniform to within 2 ps RMS across the entire active surface of approximately 52×55 mm^2 for both irradiated and non-irradiated modules. Furthermore, dedicated studies of the module performance as a function of the particle impact angle were performed to emulate different pseudorapidity regions of the final detector, with corresponding results confirming that the time resolution scales consistently with the energy deposited in the crystals. Beyond its direct impact on the final design of the BTL, this work provides a valuable framework for the optimization of large-area scintillator–SiPM–based timing detectors. The developed methods and models can be applied to future systems targeting few–tens–of–picosecond resolution in both low and high-radiation environments up to fluences of 2×10^{14} n_{eq}/cm^2. Overall, these results demonstrated that the final BTL module design meets the stringent timing requirements for operation at the HL-LHC, marking a fundamental step in the validation of the detector technology and allowing the beginning of the BTL production.

Thorough studies were carried out to define common and reproducible procedures for the assembly and QA/QC of BTL components, which are currently being produced in four assembly centers. Dedicated gluing methods have been developed to maximize the light collection efficiency through precisely tuned tools for coupling the LYSO:Ce crystal arrays to the SiPM arrays. The QA/QC procedures build upon the methods employed during the characterization studies for determining the light output. In particular, the uniformity within each module and the overall light output are verified through measurements based on the detection of photons emitted in the decay chain of ^{22}Na. The current yield achieved in sensor module assembly and QA/QC is about 95%. The assembly of detector modules required extensive studies

to identify the optimal materials and methods ensuring efficient thermal coupling between the TECs and the mechanical housing, in which two sensor modules and their front-end board are integrated, forming the detector module. A dedicated investigation revealed limitations in the baseline choice of materials and identified an improved solution, ensuring proper thermal coupling between the TECs and the cooling plate via the copper housing. This optimization is particularly crucial given the strong dependence of performance on temperature. These studies guided the development of QA/QC procedures aimed at verifying the functioning of the TECs, the temperature uniformity and the cooling capability. The procedures for tray assembly, and testing were defined collectively among the assembly centers and validated through a series of dedicated workshops, ensuring a common approach to production and testing. Currently, 25% of the trays have been produced and the completion of the BTL integration is expected by the end of next year. The BTL will achieve a time resolution of 30–60 ps up to fluences of 2×10^{14} n_{eq}/cm^2, with uniform performance ensured by the standardized assembly and QA/QC procedures. This represents a major technological milestone in high-energy physics, introducing four-dimensional event reconstruction as a new paradigm at the HL-LHC.

Bibliography

[1] Charles Byrne. *A Brief History of Electromagnetism*. New York: Springer, 2013 (cit. on p. 2).

[2] Steven Weinberg. "The Search for Unity: Notes for a History of Quantum Field Theory". In: *Daedalus* 106 (1977) (cit. on p. 3).

[3] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Physical Review* 96 (1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191 (cit. on p. 3).

[4] Steven Weinberg. "The Making of the Standard Model". In: *European Physical Journal C* 34 (2004), pp. 5–13. DOI: 10.1140/epjc/s2004-01761-1. eprint: hep-ph/0401010 (cit. on p. 3).

[5] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physics Letters* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9 (cit. on p. 3).

[6] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons II". In: *Physics Letters* 13 (1964), pp. 508–509. DOI: 10.1016/0031-9163(64)90736-9 (cit. on p. 3).

[7] P. W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Physical Review* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156 (cit. on p. 3).

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Physical Review Letters* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cit. on p. 3).

[9] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physical Review Letters* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cit. on p. 3).

[10] F. J. Hasert et al. "Search for elastic muon-neutrino electron scattering". In: *Physics Letters B* 46 (1973), pp. 121–138 (cit. on p. 4).

[11] P. Musset et al. "Search for neutral currents in neutrino interactions". In: *Journal de Physique* 34.Suppl. 11-12 (1973), T34 (cit. on p. 4).

[12] O. W. Greenberg. "Spin and unitary-spin independence in a paraquark model of baryons and mesons". In: *Physical Review Letters* 13 (1964), p. 598 (cit. on p. 4).

[13] M. Y. Han and Y. Nambu. "Three-triplet model with double SU(3) symmetry". In: *Physical Review* 139 (1965), p. 1006 (cit. on p. 4).

[14] W. A. Bardeen, H. Fritzsch, and M. Gell-Mann. "Light-cone current algebra, π^0 decay, and e^+e^- annihilation". In: *Scale and Conformal Symmetry in Hadron Physics*. Ed. by R. Gatto. New York: Wiley, 1973, p. 139 (cit. on p. 4).

[15] J. J. Aubert et al. "Experimental observation of a heavy particle J". In: *Physical Review Letters* 33 (1974), p. 1404 (cit. on p. 4).

[16] J. E. Augustin et al. "Discovery of a narrow resonance in e^+e^- annihilation". In: *Physical Review Letters* 33 (1974), p. 1406 (cit. on p. 4).

[17] G. Arnison and others (UA1 Collaboration). "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540\,\text{GeV}$". In: *Physics Letters B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2 (cit. on p. 4).

[18] G. Arnison and others (UA1 Collaboration). "Observation of the muonic decay of the charged intermediate vector boson". In: *Physics Letters B* 147 (1984), pp. 241–246. DOI: 10.1016/0370-2693(84)91315-4 (cit. on p. 4).

[19] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001 (cit. on p. 4).

[20] Ezio Torassa. "The Standard Model and Higgs physics". In: *Progress in Particle and Nuclear Physics* 100 (2018), pp. 69–106. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2018.01.003. URL: https://www.sciencedirect.com/science/article/pii/S0146641018300036 (cit. on p. 4).

[21] Mark Thomson. *Modern Particle Physics*. 1st. Cambridge, UK: Cambridge University Press, 2013. ISBN: 978-1107002173 (cit. on p. 4).

[22] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. 1st. Reading, MA: Addison-Wesley, 1995 (cit. on p. 4).

[23] Steven Weinberg. *The Quantum Theory of Fields*. Vol. 2. Cambridge: Cambridge University Press, 1996 (cit. on p. 4).

[24] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Physical Review* 105.4 (1957), pp. 1413–1415 (cit. on p. 5).

[25] TikZ.net. *3D Axis example*. 2024. URL: https://tikz.net/axis3d_cms/ (cit. on pp. 5, 20, 31).

[26] University of Oxford Department of Physics. *Why two Higgs are better than one*. 2023. URL: https://www.physics.ox.ac.uk/news/why-two-higgs-are-better-one (cit. on p. 6).

[27] The CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. URL: https://doi.org/10.1038/s41586-022-04892-x (cit. on pp. 8–12, 39).

[28] CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Phys. Lett. B* 805 (2020), p. 135425. DOI: 10.1016/j.physletb.2020.135425. URL: https://doi.org/10.1016/j.physletb.2020.135425 (cit. on p. 8).

[29] "Stairway to discovery: A report on the CMS programme of cross section measurements from millibarns to femtobarns". In: *Physics Reports* 1115 (2025). CMS physics results from the first decade of LHC data, pp. 3–115. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2024.11.005. URL: https://www.sciencedirect.com/science/article/pii/S0370157324003867 (cit. on pp. 8, 9).

[30] The CMS Collaboration. "Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production". In: *Nature Physics* 18 (2022), pp. 1329–1334. DOI: 10.1038/s41567-022-01682-0. URL: https://doi.org/10.1038/s41567-022-01682-0 (cit. on p. 8).

[31] M. Grazzini et al. "Higgs boson pair production at NNLO with top quark mass effects". In: *Journal of High Energy Physics* 2018.5 (May 2018). ISSN: 1029-8479. DOI: 10.1007/jhep05(2018)059. URL: http://dx.doi.org/10.1007/JHEP05(2018)059 (cit. on pp. 10, 61).

[32] Frédéric A. Dreyer and Alexander Karlberg. "Vector-boson fusion Higgs pair production at ¡mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML" display="inline"¿¡mml:mrow¿¡mml:msup¿¡mml:mrow¿¡m mathvariant="normal"¿N¡/mml:mi¿¡/mml:mrow¿¡mml:mrow¿¡mml:mn¿3¡/mml:mn¿¡/mml:mrow¿¡/mml:msup¿¡m In: *Physical Review D* 98.11 (Dec. 2018). ISSN: 2470-0029. DOI: 10.1103/physrevd.98.114016. URL: http://dx.doi.org/10.1103/PhysRevD.98.114016 (cit. on pp. 10, 61).

[33] Frédéric A. Dreyer et al. "Precise predictions for double-Higgs production via vector-boson fusion". In: *The European Physical Journal C* 80.11 (Nov. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-08610-7. URL: http://dx.doi.org/10.1140/epjc/s10052-020-08610-7 (cit. on pp. 10, 61).

[34] Andre Sopczak. "Prospects for Higgs Boson Research at the LHC". In: (Sept. 2025). DOI: 10.48550/arXiv.2509.22455 (cit. on p. 11).

[35] ATLAS Collaboration and G. Aad et al. "Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Physical Review Letters* 133.10 (2024), p. 101801. DOI: 10.1103/PhysRevLett.133.101801 (cit. on p. 11).

[36] A L Read. "Presentation of search results: the CLs technique". In: 28.10 (Sept. 2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088/0954-3899/28/10/313 (cit. on pp. 12, 51).

[37] CMS Collaboration. *CMS Public Results: Luminosity Measurements*. 2025. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (cit. on p. 13).

[38] Benoît Hespel, David López-Val, and Eleni Vryonidou. "Higgs pair production via gluon fusion in the Two-Higgs-Doublet Model". In: *Journal of High Energy Physics* 2014.9 (Sept. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)124. URL: http://dx.doi.org/10.1007/JHEP09(2014)124 (cit. on p. 13).

[39] G.C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Physics Reports* 516.1 (2012). Theory and phenomenology of two-Higgs-doublet models, pp. 1–102. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2012.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157312000695 (cit. on p. 13).

[40] Hamed Sanati. "The Higgs spectrum of the Minimal Supersymmetric Standard Model (MSSM) at the tree-level and a review of the hMSSM as a key benchmark scenario". In: *Journal of Subatomic Particles and Cosmology* 4 (2025), p. 100081. ISSN: 3050-4805. DOI: https://doi.org/10.1016/j.jspc.2025.100081. URL: https://www.sciencedirect.com/science/article/pii/S3050480525000615 (cit. on p. 13).

[41] Mariana Frank, Beste Korutlu, and Manuel Toharia. "Higgs phenomenology in warped extra dimensions with a fourth generation". In: *Phys. Rev. D* 84 (7 Oct. 2011), p. 075009. DOI: 10.1103/PhysRevD.84.075009. URL: https://link.aps.org/doi/10.1103/PhysRevD.84.075009 (cit. on p. 13).

[42] CERN Yellow Report Working Group. *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*. Tech. rep. CERN, 2019. DOI: 10.23731/CYRM-2019-007. URL: https://cds.cern.ch/record/2703572 (cit. on pp. 14, 71).

[43] ATLAS and CMS Collaborations. *Highlights of the HL-LHC physics projections by ATLAS and CMS*. 2025. arXiv: 2504.00672 [hep-ex]. URL: https://arxiv.org/abs/2504.00672 (cit. on pp. 14, 15).

[44] ATLAS Collaboration. "Expected performance of the ATLAS detector at HL-LHC". In: (2018). in progress (cit. on p. 14).

[45] CMS Collaboration. *Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC*. Tech. rep. CMS-NOTE-2018-006. CERN, 2018 (cit. on p. 14).

[46] *Projected sensitivity of measurements of Higgs boson pair production with the ATLAS experiment at the HL-LHC*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2925853 (cit. on p. 14).

[47] CMS Collaboration. *Projection of CMS experimental reach on HH production at HL-LHC*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2928096 (cit. on p. 14).

[48] Albert De Roeck. "Dark matter searches at accelerators". In: *Nuclear Physics B* 1003 (2024). Special Issue of Nobel Symposium 182 on Dark Matter, p. 116480. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2024.116480. URL: https://www.sciencedirect.com/science/article/pii/S0550321324000464 (cit. on p. 17).

[49] S. et al. Acharya. "The ALICE experiment: a journey through QCD". In: *The European Physical Journal C* 84.8 (2024). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-12935-y. URL: http://dx.doi.org/10.1140/epjc/s10052-024-12935-y (cit. on p. 17).

[50] LHCb Collaboration. "Observation of charge–parity symmetry breaking in baryon decays". In: *Nature* 643 (2025), pp. 1223–1228. DOI: 10.1038/s41586-025-09119-3. URL: https://doi.org/10.1038/s41586-025-09119-3 (cit. on p. 17).

[51] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001. URL: https://doi.org/10.1088/1748-0221/3/08/S08001 (cit. on p. 18).

[52] Ewa Lopienska. "The CERN accelerator complex, layout in 2022". In: (2022). URL: https://cds.cern.ch/record/2800984 (cit. on p. 18).

[53] CMS Collaboration. *CMS Detector Design*. CERN. 2023. URL: https://cms.cern/news/cms-detector-design (cit. on p. 21).

[54] The CMS Collaboration et al. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://doi.org/10.1088/1748-0221/3/08/S08004 (cit. on p. 21).

[55] CMS Collaboration. "Development of the CMS detector for the CERN LHC Run 3". In: *Journal of Instrumentation* 19.05 (May 2024), P05064. ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/05/p05064. URL: http://dx.doi.org/10.1088/1748-0221/19/05/P05064 (cit. on pp. 21–23, 26).

[56] CMS Collaboration and V. Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (Jan. 2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. eprint: arXiv:1609.02366 (cit. on p. 24).

[57] "Enriching the physics program of the CMS experiment via data scouting and data parking". In: *Physics Reports* 1115 (2025). CMS physics results from the first decade of LHC data, pp. 678–772. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2024.09.006. URL: https://www.sciencedirect.com/science/article/pii/S0370157324003247 (cit. on p. 25).

[58] Juliette Alimena. *The CMS Trigger System for Run 3*. 2023. URL: https://ep-news.web.cern.ch/content/cms-trigger-system-run-3 (cit. on p. 25).

[59] "Novel strategy targeting HH and HHH production at High Level Trigger in Run 3". In: (2023). URL: https://cds.cern.ch/record/2868787 (cit. on p. 26).

[60] CMS Collaboration. In: 12.10 (Oct. 2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. URL: https://doi.org/10.1088/1748-0221/12/10/P10003 (cit. on p. 27).

[61] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. URL: https://doi.org/10.1088/1748-0221/9/10/P10009 (cit. on p. 28).

[62] P. Billoir and S. Qian. "Track fitting with multiple scattering and energy loss". In: *Nucl. Instrum. Meth. A* 294 (1990), p. 219 (cit. on p. 28).

[63] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788 (cit. on pp. 28, 92).

[64] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063 (cit. on p. 30).

[65] A.M. Sirunyan et al. "Pileup mitigation at CMS in 13 TeV data". In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09018. DOI: 10.1088/1748-0221/15/09/P09018. URL: https://doi.org/10.1088/1748-0221/15/09/P09018 (cit. on p. 30).

[66] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex] (cit. on pp. 31, 36).

[67] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph] (cit. on p. 31).

[68] CMS Collaboration. "Identification of hadronic tau lepton decays using a deep neural network". In: *JINST* 15.06 (2020), P06024. DOI: 10.1088/1748-0221/15/06/P06024. URL: https://doi.org/10.1088/1748-0221/15/06/P06024 (cit. on p. 33).

[69] Wiktor Matyszkiewicz and Artur Kalinowski. "TAU-PAIR INVARIANT MASS ESTIMATION USING MAXIMUM LIKELIHOOD ESTIMATION AND COLLINEAR APPROXIMATION". In: *Acta Physica Polonica B Proceedings Supplement* 18.5-A21 (2025). DOI: 10.5506/APhysPolBSupp.18.5-A21 (cit. on p. 34).

[70] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on p. 34).

[71] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *J. High Energ. Phys.* 2010.06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. URL: https://doi.org/10.1007/JHEP06(2010)043 (cit. on p. 34).

[72] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. URL: http://dx.doi.org/10.1016/j.cpc.2015.01.024 (cit. on p. 34).

[73] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 34).

[74] CMS Collaboration. *Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016–2018*. Tech. rep. CMS-DP-2020-019. CERN, Apr. 2020. URL: https://cds.cern.ch/record/2715872 (cit. on p. 36).

[75] "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector". In: (2018). URL: https://cds.cern.ch/record/2646773 (cit. on p. 36).

[76] R. L. Workman and others (Particle Data Group). "Review of Particle Physics, Chapter 40: Statistical Methods". In: *Prog. Theor. Exp. Phys.* 2022.8 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097. URL: https://pdg.lbl.gov/2022/reviews/rpp2022-rev-statistics.pdf (cit. on p. 37).

[77] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837 (cit. on p. 37).

[78] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *European Physical Journal C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on p. 37).

[79] A.M. Sirunyan, A. Tumasyan, and others (CMS Collaboration). "Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2021.3 (2021), p. 257. DOI: 10.1007/JHEP03(2021)257. URL: https://doi.org/10.1007/JHEP03(2021)257 (cit. on pp. 39, 40, 44, 50, 53, 55, 65, 67).

[80] I R Kenyon. "The Drell-Yan process". In: *Reports on Progress in Physics* 45.11 (Nov. 1982), p. 1261. DOI: 10.1088/0034-4885/45/11/002. URL: https://dx.doi.org/10.1088/0034-4885/45/11/002 (cit. on p. 40).

[81] "Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques". In: (2020). URL: https://cds.cern.ch/record/2707946 (cit. on p. 44).

[82] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. Tech. rep. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG. Geneva, 2017. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922. URL: https://cds.cern.ch/record/2227475 (cit. on p. 51).

[83] Konstantin Lehmann and Bernd Stelzer. "The Fake Factor Method and its relation to the Matrix Method". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1054 (2023), p. 168376. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2023.168376. URL: https://www.sciencedirect.com/science/article/pii/S0168900223003662 (cit. on p. 57).

[84] The ATLAS Collaboration et al. "Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector". In: *J. High Energ. Phys.* 2023 (2023), p. 40. DOI: 10.1007/JHEP07(2023)040. URL: https://doi.org/10.1007/JHEP07(2023)040 (cit. on p. 57).

[85] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980 (cit. on p. 60).

[86] J. Baglio et al. "$gg \to HH$: Combined uncertainties". In: *Phys. Rev. D* 103 (5 Mar. 2021), p. 056002. DOI: 10.1103/PhysRevD.103.056002. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.056002 (cit. on p. 61).

[87] Alexander Karlberg et al. *Ad interim recommendations for the Higgs boson production cross sections at* $\sqrt{s} = 13.6$ *TeV*. 2024. arXiv: 2402.09955 [hep-ph]. URL: https://arxiv.org/abs/2402.09955 (cit. on p. 61).

[88] I. Zurbano Fernández et al. *High-Luminosity Large Hadron Collider (HL-LHC)*. Tech. rep. CERN Yellow Reports: Monographs CERN-2020-010. Geneva, Switzerland: CERN, Dec. 2020. DOI: 10.23731/CYRM-2020-0010. URL: https://doi.org/10.23731/CYRM-2020-0010 (cit. on p. 70).

[89] CERN. *High Luminosity Large Hadron Collider (HL-LHC) Project*. 2025. URL: https://home.cern/science/accelerators/high-luminosity-lhc (cit. on pp. 70, 72).

[90] D. Contardo et al. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02. CERN, 2015. DOI: 10.17181/CERN.VU8I.D59J. URL: https://cds.cern.ch/record/2020886 (cit. on pp. 73, 76).

[91] CMS Collaboration. *CMS Phase II Upgrade Scope Document*. Tech. rep. CERN, Sept. 2015. URL: https://cds.cern.ch/record/2055167 (cit. on pp. 73, 74).

[92] Bora Akgün on behalf of the CMS HGCAL Collaboration. "An Overview of the CMS High Granularity Calorimeter". In: *Particles* 8.1 (2025), p. 4. DOI: 10.3390/particles8010004. URL: https://doi.org/10.3390/particles8010004 (cit. on p. 75).

[93] Collaboration CMS. *A MIP Timing Detector for the CMS Phase-2 Upgrade*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2667167 (cit. on pp. 75, 77, 78, 80, 81, 84, 91, 92, 94–99, 102, 107, 118, 139, 158).

[94] E. Albuquerque et al. "TOFHIR2: The readout ASIC of the CMS Barrel MIP Timing Detector". In: *Journal of Instrumentation* 19.5 (2024), P05048. DOI: 10.1088/1748-0221/19/05/P05048. URL: https://doi.org/10.1088/1748-0221/19/05/P05048 (cit. on pp. 80, 142, 144, 145).

[95] Glenn F Knoll. *Radiation detection and measurement*. New York, NY: Wiley, 2010. URL: https://cds.cern.ch/record/1300754 (cit. on pp. 80, 83).

[96] F. Addesa et al. "Optimization of LYSO crystals and SiPM parameters for the CMS MIP timing detector". In: *Journal of Instrumentation* 19.12 (Dec. 2024), P12020. DOI: 10.1088/1748-0221/19/12/P12020. URL: https://dx.doi.org/10.1088/1748-0221/19/12/P12020 (cit. on pp. 83, 117).

[97] Marco Ferrero et al. *An Introduction to Ultra-Fast Silicon Detectors*. CRC Press, July 2021. ISBN: 978-1-003-13194-6. DOI: 10.1201/9781003131946 (cit. on pp. 85, 86).

[98] M. Ferrero et al. "Radiation resistant LGAD design". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 919 (2019), pp. 16–26. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.11.121. URL: https://www.sciencedirect.com/science/article/pii/S0168900218317741 (cit. on pp. 85, 86).

[99] "Characterization of the FBK-LGAD devices manufactured by LFoundry for the CMS Endcap Timing Layer". In: (2025). URL: https://cds.cern.ch/record/2923948 (cit. on pp. 86–88).

[100] V. Sola et al. "A compensated design of the LGAD gain layer". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1040 (2022), p. 167232. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2022.167232. URL: https://www.sciencedirect.com/science/article/pii/S0168900222005861 (cit. on p. 86).

[101] Ronald J. Lipton. *LGAD Single Event Burnout Studies*. Tech. rep. Fermi National Accelerator Lab. (FNAL), Batavia, IL (United States), Aug. 2021. DOI: 10.2172/1841397. URL: https://www.osti.gov/biblio/1841397 (cit. on p. 86).

[102] M. Salva. *CMS Run 2b CNM Activities*. https://indico.cern.ch/event/803820/contributions/3343239/attachments/1808627/2952837/20190308_CmsRun2b_Cnm_Activities_Salva.pdf. 2019 (cit. on p. 87).

[103] L. Lanteri et al. "Characterization of the FBK-LGAD devices manufactured at an external foundry for large-volume productions". In: *Journal of Instrumentation* 20.07 (July 2025), p. C07039. DOI: 10.1088/1748-0221/20/07/C07039. URL: https://doi.org/10.1088/1748-0221/20/07/C07039 (cit. on p. 88).

[104] Wei Zhang et al. "A Low-Power Time-to-Digital Converter for the CMS Endcap Timing Layer (ETL) Upgrade". In: *IEEE Transactions on Nuclear Science* 68.8 (2021), pp. 1984–1992. DOI: 10.1109/TNS.2021.3085564 (cit. on p. 89).

[105] Valentina Sola. "The Endcap Timing Layer for the CMS MTD Upgrade". 2025. URL: https://indico.cern.ch/event/1445926/contributions/6487554/attachments/3072195/5435293/vs-ETL-FAST2025.pdf (cit. on p. 89).

[106] Paulo Moreira et al. *lpGBT documentation*. 2024. URL: https://cds.cern.ch/record/2809058 (cit. on pp. 90, 146).

[107] Andrew William Rose et al. "Serenity: An ATCA prototyping platform for CMS Phase-2". In: *PoS* TWEPP2018 (2019), p. 115. DOI: 10.22323/1.343.0115 (cit. on p. 90).

[108] Polina Simkina. "The DAQ and clock distribution system of CMS MIP Timing Detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1047 (2023), p. 167802. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2022.167802. URL: https://www.sciencedirect.com/science/article/pii/S0168900222010944 (cit. on p. 90).

[109] "Update of the MTD physics case". In: (2022). URL: https://cds.cern.ch/record/2815409 (cit. on pp. 92, 97–99).

[110] "Improved use of MTD time in vertex reconstruction". In: (2024). URL: https://cds.cern.ch/record/2904361 (cit. on p. 92).

[111] "Update of the vertex reconstruction using track time from MTD". In: (2024). URL: https://cds.cern.ch/record/2914583 (cit. on pp. 92–94).

[112] S. Gundacker et al. "Precise rise and decay time measurements of inorganic scintillators by means of X-ray and 511 keV excitation". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 891 (2018), pp. 42–52. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.02.074. URL: https://www.sciencedirect.com/science/article/pii/S0168900218302286 (cit. on p. 103).

[113] F.M. Addesa et al. "Comparative characterization study of LYSO:Ce crystals for timing applications". In: *Journal of Instrumentation* 17.08 (Aug. 2022), P08028. DOI: 10.1088/1748-0221/17/08/P08028. URL: https://dx.doi.org/10.1088/1748-0221/17/08/P08028 (cit. on pp. 104–107).

[114] M. Moszynski et al. "Absolute light output of scintillators". In: *IEEE Transactions on Nuclear Science* 44.3 (1997), pp. 1052–1061. DOI: 10.1109/23.603803 (cit. on p. 105).

[115] F. Addesa et al. *Optical spectroscopic characterization of LYSO crystals at the Calliope facility (ENEA Casaccia R.C.)* Tech. rep. ENEA-IRIS, 2020. URL: http://hdl.handle.net/20.500.12079/55987 (cit. on p. 105).

[116] Chen Hu et al. "Neutron-Induced Radiation Damage in BaF2, LYSO/LFS and PWO Crystals". In: *Journal of Physics: Conference Series* 1162.1 (Jan. 2019), p. 012020. DOI: 10.1088/1742-6596/1162/1/012020. URL: https://dx.doi.org/10.1088/1742-6596/1162/1/012020 (cit. on p. 106).

[117] Rihua Mao, Liyuan Zhang, and Ren-Yuan Zhu. "Gamma ray induced radiation damage in PWO and LSO/LYSO crystals". In: *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. 2009, pp. 2045–2049. DOI: 10.1109/NSSMIC.2009.5402126 (cit. on p. 106).

[118] Stefan Gundacker and Arjan Heering. "The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector". In: *Physics in Medicine and Biology* 65.17 (Aug. 2020), 17TR01. DOI: 10.1088/1361-6560/ab7b2d. URL: https://dx.doi.org/10.1088/1361-6560/ab7b2d (cit. on pp. 108, 109).

[119] S. Seifert et al. "Simulation of Silicon Photomultiplier Signals". In: *IEEE Transactions on Nuclear Science* 56.6 (2009), pp. 3726–3733. DOI: 10.1109/TNS.2009.2030728 (cit. on p. 108).

[120] Claudio Piemonte and Alberto Gola. "Overview on the main parameters and technology of modern Silicon Photomultipliers". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 926 (2019), pp. 2–15. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.11.119. URL: https://www.sciencedirect.com/science/article/pii/S0168900218317716 (cit. on p. 109).

[121] P. Musienko. *SiPM radiation damage*. Presentation at PD24 workshop, CERN. 2024. URL: https://indico.cern.ch/event/1404192/contributions/6059997/attachments/2970344/5227759/PD24-SiPM-rad-dam_Musienko-v2.pdf (cit. on p. 110).

[122] E. Garutti and Yu. Musienko. "Radiation damage of SiPMs". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 926 (May 2019), pp. 69–84. ISSN: 0168-9002. DOI: 10.1016/j.nima.2018.10.191. URL: http://dx.doi.org/10.1016/j.nima.2018.10.191 (cit. on pp. 110, 112).

[123] Michael Moll. "Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties". PhD thesis. Hamburg U., 1999 (cit. on pp. 110, 112, 119).

[124] A. Vasilescu and G. Lindstroem. *Displacement damage in silicon, on-line compilation*. URL: https://rd50.web.cern.ch/niel/ (cit. on p. 112).

[125] A. Bornheim et al. "Integration of thermo-electric coolers into the CMS MTD SiPM arrays for operation under high neutron fluence". In: *Journal of Instrumentation* 18.08 (Aug. 2023), P08020. DOI: 10.1088/1748-0221/18/08/P08020. URL: https://dx.doi.org/10.1088/1748-0221/18/08/P08020 (cit. on pp. 112–114, 126, 147).

[126] D. Marano et al. "Silicon Photomultipliers Electrical Model Extensive Analytical Analysis". In: *IEEE Transactions on Nuclear Science* 61.1 (2014), pp. 23–34. DOI: 10.1109/TNS.2013.2283231 (cit. on p. 115).

[127] Francesco Corsi et al. "Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 572 (Mar. 2007), pp. 416–418. DOI: 10.1016/j.nima.2006.10.219 (cit. on p. 115).

[128] T. Bretz, T. Hebbeker, and J. Kemp. *Extending the dynamic range of SiPMs by understanding their non-linear behavior*. 2020. DOI: 10.48550/ARXIV.2010.14886. URL: https://arxiv.org/abs/2010.14886 (cit. on p. 115).

[129] Abhinav Jha et al. "Simulating Silicon Photomultiplier Response to Scintillation Light". In: *IEEE Transactions on Nuclear Science* 60 (Feb. 2013), pp. 336–351. DOI: 10.1109/TNS.2012.2234135 (cit. on p. 115).

[130] Advansid. *Advansid — Home page*. 2025. URL: https://www.advansid.com (cit. on p. 116).

[131] Simona Palluotto for the CMS Collaboration. "Characterization of sensor modules for the CMS Barrel Timing Layer at HL-LHC". In: *Il Nuovo Cimento C* 46.1 (2023), p. 58. DOI: 10.1393/ncc/i2023-23058-1 (cit. on pp. 118, 119).

[132] S. A. PETsys Electronics Medical PET Detectors. *PETsys SiPM Readout Chain*. Technical Report v15. Taguspark, Lisbon Science and Technology Park, Oeiras, Portugal: PETsys Electronics, July 2023. URL: https://www.petsyselectronics.com/web/website/docs/products/product4/Long%20Flyer%20SiPM%20readout%20chain_v15.pdf (cit. on p. 121).

[133] R. Abbott et al. "Test beam characterization of sensor prototypes for the CMS Barrel MIP Timing Detector". In: *Journal of Instrumentation* 16.07 (July 2021), P07023. DOI: 10.1088/1748-0221/16/07/P07023. URL: https://dx.doi.org/10.1088/1748-0221/16/07/P07023 (cit. on p. 124).

[134] Yu. Musienko et al. "Radiation damage in silicon photomultipliers exposed to neutron radiation". In: *Journal of Instrumentation* 12.07 (July 2017), p. C07030. DOI: 10.1088/1748-0221/12/07/C07030. URL: https://dx.doi.org/10.1088/1748-0221/12/07/C07030 (cit. on p. 126).

[135] P. Carniti, C. Gotti, and G. Pessina. "ALDO2, a multi-function rad-hard linear regulator for SiPM-based HEP detectors". In: *Nuclear Instruments and Methods in Physics Research Section A* 1039 (2022), p. 167028. DOI: 10.1016/j.nima.2022.167028. arXiv: 2203.16098 [physics.ins-det] (cit. on pp. 142, 145).

[136] S.A. PETsys Electronics. *PETsys Electronics*. URL: https://www.petsyselectronics.com/web/ (cit. on p. 143).

[137] M. Firlej et al. "An ultra-low power 10-bit, 50 MSps SAR ADC for multi-channel readout ASICs". In: *JINST* 18 (2023), P11013. DOI: 10.1088/1748-0221/18/11/P11013 (cit. on p. 143).

[138] R. Bugalho et al. "Experimental characterization of the TOFPET2 ASIC". In: *Journal of Instrumentation* 14.3 (2019), P03029. DOI: 10.1088/1748-0221/14/03/P03029. URL: https://doi.org/10.1088/1748-0221/14/03/P03029 (cit. on p. 143).

[139] E. Albuquerque et al. "TOFHIR2: The readout ASIC of the CMS Barrel MIP Timing Detector". In: *Proceedings of the 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2020, pp. 1–7. DOI: 10.1109/NSS/MIC42677.2020.9507749. URL: https://doi.org/10.1109/NSS/MIC42677.2020.9507749 (cit. on p. 143).

[140] T. Niknejad et al. "Results with the TOFHIR2X Revision of the Front-end ASIC of the CMS MTD Barrel Timing Layer". In: *Proceedings of the 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. 2021, pp. 1–5. DOI: 10.1109/NSS/MIC44867.2021.9875751. URL: https://doi.org/10.1109/NSS/MIC44867.2021.9875751 (cit. on p. 143).

[141] A. Gola, C. Piemonte, and A. Tarolli. "The DLED Algorithm for Timing Measurements on Large Area SiPMs Coupled to Scintillators". In: *IEEE Transactions on Nuclear Science* 59.2 (2012), pp. 358–365. DOI: 10.1109/TNS.2011.2179654 (cit. on p. 143).

[142] L. Gonella et al. "Total ionizing dose effects in 130 nm commercial CMOS technologies for HEP experiments". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 582 (2007), pp. 750–754. DOI: 10.1016/j.nima.2007.07.002 (cit. on p. 144).

[143] F. Faccio et al. "Radiation-Induced Short Channel (RISCE) and Narrow Channel (RINCE) Effects in 65 and 130 nm MOSFETs". In: *IEEE Transactions on Nuclear Science* 62.6 (2015), pp. 2933–2940. DOI: 10.1109/TNS.2015.2491978 (cit. on p. 144).

[144] CERN lpGBT Project. *lpGBT Documentation and Manuals*. URL: https://lpgbt.web.cern.ch/lpgbt/ (cit. on p. 146).

[145] Laird Performance Materials. *Datasheet for Thermal Pad A17916-05*. 2025. URL: https://fr.farnell.com/laird/a17916-05/pad-thermique-228-6x228-6x1-27mm/dp/3677528 (cit. on p. 147).

[146] Parker Chomerics. *Datasheet for THERMFLOW® T558 Phase Change Material*. 2025. URL: https://www.digikey.it/it/product-highlight/p/parker-chomerics/thermflow-t558-phase-change-material (cit. on p. 147).

[147] Guillermo Reales Gutiérrez. *TOM Thermal Study (Draft, 2020-06-09)*. 2020. URL: https://cms-docdb.cern.ch/cgi-bin/DocDB/RetrieveFile?docid=14272&filename=TOM_Thermal_Study_DRAFT%202020.06.09%20%282%29.pdf&version=1 (cit. on p. 147).

[148] Laird Technologies. *Datasheet for Thermal Pad A10092-02*. 2025. URL: https://www.digikey.it/it/products/detail/laird-technologies-thermal-materials/A10092-02/4384095 (cit. on p. 147).

[149] Laird Performance Materials. *Datasheet for Thermal Pad A17916-05*. 2025. URL: https://fr.farnell.com/laird/a17916-05/pad-thermique-228-6x228-6x1-27mm/dp/3677528 (cit. on p. 147).

[150] Parker Chomerics. *Thermal Interface Materials for Electronics Cooling*. 2025. URL: https://www.parker.com/content/dam/Parker-com/Divisions-2011/Chomerics-Division/SupportAssets/Thermal_Interface_Materials_Catalog.pdf (cit. on p. 148).

[151] "Barrel Timing Layer crystals quality control plots". In: (2024). URL: https://cds.cern.ch/record/2909331 (cit. on p. 149).

[152] Mitchell Wayne. *Development, Characterization and Quality Control of Silicon Photomultipliers for the CMS Barrel Timing Layer*. Presented at Coordinating Panel on Advanced Detectors 2025. 2025. URL: https://indico.global/event/14966/contributions/133758/attachments/63077/122002/CPAD25%20Talk%20V4.pdf (cit. on pp. 150, 151).

List of Figures

List of Tables

Acknowledgments

It feels reductive to express my gratitude in just a few lines. Acknowledgments could easily fill many pages, yet I will keep them concise, hoping that a few words may still convey the depth of my appreciation at the end of this journey.

I am grateful to the reviewers, Luca Cadamuro and Indara Suarez, for the careful evaluation of this work and for the constructive and insightful discussions that followed.

I would like to sincerely acknowledge Tommaso, who has always welcomed every request for help and guidance with remarkable attentiveness and dedication. I am equally indebted to Alessio, who mentored me throughout these years and from whom I had the pleasure of learning everything from versors to the Higgs potential.

A special mention goes to the BTL group at Milano-Bicocca, a cohesive *gruppo di pallavolo* from whom I learned so much. My deepest gratitude goes to Martina, who has been a constant reference from my undergraduate years to the present work; her advice, patience, and practical mentorship have guided me throughout. I am equally thankful to Andrea, whose insight and hands-on help in the office and the lab were essential, and to Marco, whose guidance and readiness to assist have been invaluable. I would also like to acknowledge Federico and Flavia for their collaboration. Together, they all supported me from my very first `ssh` and `cmsenv`, and their presence has been fundamental.

I am grateful to the entire CMS Bicocca group for their collaboration and support, with a special mention to Simone and Andrea. I would like to give particular thanks to Raffaele and Francesco, whose patience, constant availability, 24/7, and readiness to tackle any problem have been invaluable to me. Working with all of you has taught me a great deal.

I would like to thank all the MTD colleagues I had the opportunity to work with, with special acknowledgment to Paolo for his guidance. Every test beam campaign and assembly workshop has been a valuable professional and personal experience.

Thanks also to all members of the $HH \to bb\tau\tau$ analysis, from the Klub to the CCLUB team, with a special mention to Jona for his constant support and kindness.

Ultimately, every journey is enriched by the people who walk alongside you. I have been fortunate to share these years with friends who brightened my offline moments, from Laveno to U2, Geneva, and R1. My wholehearted thanks go to all of you.

E per finire dal principio, un ringraziamento va alla mia famiglia, faro luminoso anche nelle notti di tormenta.