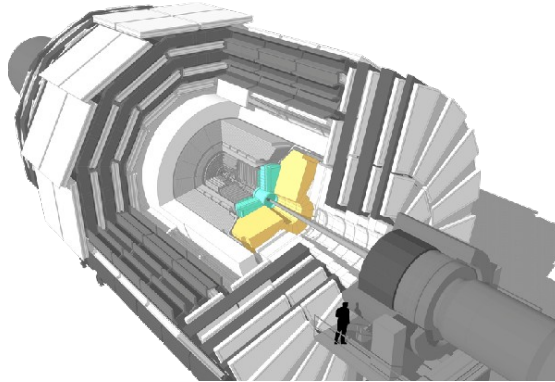




Optimization of the local reconstruction in a high granular calorimeter using a heterogeneous computing model

Daniela Beatriz Cabral Cardoso

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisor(s): Prof. Pedro Manuel Vieira de Castro Ferreira da Silva
Prof. Michele Gallinaro

Examination Committee

Chairperson: Prof. José Guilherme Teixeira de Almeida Milhano
Supervisor: Prof. Pedro Manuel Vieira de Castro Ferreira da Silva
Member of the Committee: Prof. João Carlos Carvalho de Sá Seixas
Prof. Tiziano Camporesi

May 2025

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I would like to sincerely thank my supervisors, Pedro and Michele, for their invaluable support and guidance throughout this project. Their insight and encouragement were essential at every stage of the work. Pedro, thank you for guiding me from the very beginning, for giving me the space to learn and grow, while always being available to answer every question, even the silly ones. Michele, thank you for accepting to supervise this project from LIP. Your constant encouragement helped keep me motivated and allowed me to deliver this work to the best of my ability.

I am also grateful to LIP and the CMS collaboration for their support during my visit to CERN in September, which allowed me to participate in the test beam and gain valuable experience.

Lastly, to my friends and family, thank you for your constant support throughout this journey. A special mention goes to Diogo, you have been the best friend I could have asked for, always present and encouraging. I would also like to especially thank João, my brother, mother, and grandmother, whose unwavering support has been a foundation throughout my academic path.

Resumo

A fase de melhoramento do *High-Luminosity Large Hadron Collider* (HL-LHC) no CERN irá gerar dados de colisões próton-próton com as mais altas energias de centro de massa registadas até hoje. Com mais de 140 interações simultâneas por cruzamento de feixes e uma luminosidade de $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, o detetor *CMS* e a infraestrutura de processamento enfrentam desafios significativos, incluindo radiação intensa, altos fluxos de partículas e a necessidade de reconstrução rápida e precisa para obter medições fiáveis.

Para dar resposta a estes desafios, o *CMS* será amplamente atualizado, incluindo a substituição dos seus *endcaps* por um novo calorímetro de alta granularidade (*High Granularity Calorimeter* HGCal). Este sistema baseado em sensores de silício irá fornecer medições de energia e tempo em cerca de 6 milhões de canais, permitindo uma reconstrução precisa do fluxo de partículas mesmo sob condições extremas de sobreposição de eventos (*pileup*).

Esta tese apresenta o desenvolvimento e validação de uma estrutura local de reconstrução e calibração usando dados de testes de feixe do HGCal. Destaca-se a criação de um método automatizado para extrair o valor mais provável (MPV) dos depósitos de energia de partículas de mínima ionização (MIP). O processamento inclui subtração de pedestal, correção de ruído, alinhamento temporal. O método de calibração foi depois testado e validado num teste de feixe do HGCal. O método proposto permite uma calibração por canal robusta, essencial para a operação precisa do HGCal na era do HL-LHC.

Palavras-chave: Detetores de Silício, Intercalibração, *High Granularity Calorimeter*, CMS, HL-LHC, Detetores

Abstract

The High-Luminosity phase of the Large Hadron Collider (HL-LHC) at the European Organisation for Nuclear Research (CERN) will deliver an unprecedented dataset of proton-proton collisions at the highest centre-of-mass energies to date. With over 140 simultaneous interactions per bunch crossing and luminosity of around $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, both the CMS detector and its computing infrastructure face significant challenges. These include severe radiation damage, high particle fluxes, and the need for fast and precise reconstruction to isolate physics signals of interest, particularly for precision measurements. To meet these demands, the CMS experiment is undergoing major upgrades, including the replacement of the endcap calorimeters with a high-granularity sampling calorimeter (HGCAL). HGCAL will provide combined energy and time measurements from approximately six million channels, enabling accurate particle flow reconstruction under extreme pileup. The upgrade features a silicon-based sampling structure with fine spatial segmentation, enhancing both spatial and temporal resolution.

This thesis explores the development and validation of a local reconstruction and calibration framework using data from HGCAL beam tests. A key contribution is an automated method to extract the most probable value (MPV) of energy deposits from minimum ionising particles (MIP) across individual silicon sensors, a critical step in calorimeter energy calibration. Essential processing steps include pedestal subtraction, noise correction, and timing alignment. The calibration algorithms were tested and validated under real test beam conditions. This work contributes to a per-channel calibration method that will be essential for the accurate calibration of HGCAL in preparation for its deployment in the HL-LHC era.

Keywords: Silicon detectors, Intercalibration, High Granularity Calorimeter, CMS, HL-LHC, Detectors

Contents

Acknowledgments	v
Resumo	vii
Abstract	ix
List of Tables	xiii
List of Figures	xv
Nomenclature	xxi
Glossary	xxiii
1 Motivation and outline	1
2 Introduction	3
2.1 The Standard Model of particle physics	3
2.2 From collision to detection	6
2.3 Large Hadron Collider	8
2.4 CMS Experiment and the High Luminosity LHC	11
3 Energy Loss in Matter and Calorimetry	15
3.1 Interactions of particles in matter	15
3.2 Development of particle showers	18
3.3 Calorimetry	19
3.3.1 Electromagnetic calorimeter	21
3.3.2 Hadronic calorimeter	22
3.3.3 Electromagnetic Calorimetry: From Homogeneous to Sampling Designs	23
4 CMS Endcap HGCALE	26
4.1 Detector Design	26
4.1.1 Electronics	30
4.2 Detector Readout	32
4.2.1 DAQ	32
4.3 Event Reconstruction	34

5	Test Beam and Calibration	37
5.1	Experimental Apparatus	37
5.1.1	SPS and Beam line	37
5.1.2	Test beams in 2023	39
5.1.3	Test beams in 2024	42
5.2	Calibration	44
5.2.1	MIP Signal	46
5.2.2	Corrections	47
5.2.3	Fit	48
5.2.4	Bethe-Bloch and correction factor f	48
5.2.5	Quality criteria	49
6	Results	51
6.1	Results from Test Beam 2023	52
6.1.1	Pedestal subtraction	52
6.1.2	Data-driven estimation of the pedestal in physics runs	52
6.1.3	Common mode channels	54
6.2	Automatising the MIP Calibration	59
6.3	Results from Test Beam 2024	60
6.3.1	Pedestal and noise subtraction	61
6.3.2	Trigger phase analysis	64
6.3.3	MIP Scale	65
6.3.4	Injecting the constants needed for local reconstruction in CMSSW	70
7	Conclusions	73
7.1	Conclusions and outlook	73
7.2	Future Work	74
	Bibliography	75
A	Other results	79

List of Tables

2.1	Main Standard Model branching ratios $\mathcal{B}$ of the Higgs boson and their relative theoretical uncertainties for $m_H = 125$ GeV [12, 13]. The last column shows the experimental signal-strength parameters, which quantify the deviations from the SM predictions for both production modes and decay channels. The uncertainties comprise statistical and systematic components [9].	5
3.1	Properties of selected materials characterising the energy loss and evolution of particle showers. Z and A are the atomic number and mass number, respectively. ρ is the mass density, and $(dE/dx)_{\min}$ is the minimum energy loss per distance dx due to ionisation. X_0 is the radiation length, R_M the Molière radius, and λ_n the nuclear interaction length. All values are taken from [29].	20
3.2	Energy resolution parameters for homogeneous (ECAL and sampling calorimeters (HGCAL)). ECAL performance is shown for various particles and conditions in both barrel and endcap regions. HGCAL performance is based on test beam measurements of the electromagnetic section with positrons in the 20–300 GeV range.	25
4.1	Comparative overview of the silicon and scintillator-based components of the HGCAL endcap calorimeters, highlighting absorber composition, detector depth, module coverage, and segmentation for HGCAL [39].	28
6.1	Fit parameters for channels 121 and 122 of Run 14230133.	57
6.3	Performance metrics from a single-threaded CPU reconstruction job.	70
6.4	Comparison of reconstruction performance across CPU and GPU configurations.	71

List of Figures

2.1	Elementary particles in the Standard Model, showing the spin, charge and approximate values for the mass of each particle, obtained experimentally [7].	4
2.2	Schematic overview of the CERN accelerator complex (not to scale). The complex chains the production of H^+ ions with the beam-beam collisions at each of the four interaction points. The production-to-collision sequence follows the following order: proton source - RFQ - LINAC2 - PSB - PS - SPS - LHC [17].	9
2.3	The High-Luminosity Upgrade plan for the LHC at CERN, including the different past runs and corresponding $\sqrt{s}$ [21].	10
2.4	Schematic of a transverse slice of the CMS detector showing specific particle interactions, from the beam interaction region to the muon detector [22].	12
3.1	Mean rate of energy loss, also called mass stopping power, of positive muons in copper. The vertical blue bands indicate the boundaries between different approximation regions [26].	16
3.2	Most probable energy loss in silicon, scaled to the mean loss of a minimum ionizing particle, $388 \text{ eV}/\mu\text{m}$ ($1.66 \text{ MeV g}^{-1} \text{ cm}^2$) [26].	17
3.3	Development of an electromagnetic shower (a) and a hadronic shower starting by neutron-nucleus interaction after a distance of λ in the matter(b) [28].	19
3.4	Comparison of a homogeneous calorimeter (left) based on scintillating crystals and a sampling calorimeter (right) like HGCal, using alternating layers of absorber and silicon sensors. Granularity, longitudinal sampling, and absorber/active layer structures are highlighted.	24
4.1	Left: Longitudinal view of the HGCal. The different regions are made in the figure with CE-E corresponding to the electromagnetic calorimeter and CE-H to the hadronic calorimeter. The 47 layers of active elements are represented within the pink envelope [36]. Right: Layout of a silicon-only (top) and mixed silicon-scintillator (bottom, not to scale) layer indicating the different silicon wafer thickness regions. [37].	27
4.2	Schematic illustration of an 8" silicon module for the HGCal, which is divided into large-cells of 1.18 cm^2 for the Low Density (LD) (left) and the small-cells of 0.52 cm^2 High Density (HD) (right) [38].	29

4.3	From left to right: 1. 3D view of the cassette structure of the HGCal. The Si hexagon wafers are represented on the very first plane, with the different colours showing the cassettes in which they are assembled. The beam travels in the inner radius of the apparatus 2. Representation of layer 3 of HGCal. The Colours are used to represent the different cassettes and thicknesses of the 8" Si wafers. The bars display the orientation of the DCDC power converters [41] 3. Representation of a cassette with the readout trains employed in the trigger and data acquisition of the detector, with the location of HD and LD Hexaboards 4. An exploded view of a hexmodule displaying (bottom to top) the baseplate (CuW or C), the Kapton TM foil, the Si sensor and the PCB [42].	30
4.4	Side by side comparison of High Density (HD) and Low Density (LD) hexaboard design [43].	31
4.5	Schematic illustration of the three-fold diamond configuration of trigger cell division (TC and STC) for Low Density (LD) (left) and High Density (HD) (right) silicon modules. Formation of TCs by a clustering procedure with four large cells of 1.18 cm ² (left) and nine small cells of 0.52 cm ² (right) [38].	31
4.6	DAQ Readout (upstream) architectures for the HGCal from the FE to the BE.	32
4.7	Fast control commands path for the system.	33
4.8	Slow control commands path for the system.	33
4.9	Offline data flow in HGCal test beam operations. The diagram illustrates the complete chain of offline data processing and calibration for HGCal beam tests, from frontend configuration to locally reconstructed output in the NANO format. Data acquisition begins at the experimental site, where frontend electronics configurations (stored in YAML files) are loaded by the <i>Serenity backend</i> , which interfaces with the HGCal detector and stores raw output to disk. The <i>Run control</i> and <i>Run registry servers</i> manage and log metadata about the run. The O2O (online-to-offline) system extracts configuration constants from YAML files and converts them into JSON files used by reconstruction. Raw data are unpacked and passed through digitization (RAW → DIGI), quality monitoring (DQM), and local reconstruction, all managed by the CMSSW framework. The <i>HTCondor</i> batch system handles job distribution for digitization and reconstruction. Calibrations are extracted from DIGI-level data during the prompt calibration phase, stored in JSON format, and fed back into the pipeline. Final output datasets, including DIGI, RECO, and NANO formats, are stored on disk in a flat tree structure for downstream analysis. Figure by I. Neutelings.	35
5.1	SPS beam status monitoring screen. The yellow curve is the relative particle luminosity, while the white curve is the magnetic field upstream. The perpendicular line is the updating bar, which goes from left to right and completes a cycle [47].	39
5.2	Photograph of the September 2023 test beam setup at H4, showing the integration of absorbers in front of the silicon modules. The absorbers were removed for a subset of runs to explore alternative configurations.	40

5.3	Close-up photograph of the silicon modules used in 2023 test beams for single module testing. From left to right: Low Density module 200 μm Si sensor thickness, High Density module 120 μm Si sensor thickness, Low Density partial module 300 μm Si sensor thickness.	40
5.4	Top: Photograph of the H4 beamline setup used during the 2023 test beam. Bottom: Silicon modules LD-V3 module, with 300 μm thick Si sensors and bias voltage of -270V connected to wagons (red), these modules were used in the full electronics chain testing.	41
5.5	Test beam setups at H4 in 2023. The first configuration (a) shows the simplified scheme used for single-module testing, while the second (b) represents the full electronic chain [48].	41
5.6	Layers and scintillator tiles mounted inside the 3T H2 superconducting magnet during the 2024 test beam.	43
5.7	Scheme of the multi-layer setup used in September 2024 alongside the identification of each module in each layer, showing three-module trains assembled on the trains (3 modules per train except on the additional train with a partial hexaboard). The rotation of each module is also shown schematically. Figure by C. Amendola.	43
5.8	Illustrative diagram of the September 2024 test beam setup. The system includes cooling (chiller), powering (low voltage power supply - 11V and bias voltage supply -280 V and leakage currents on the order of 0.5 μA) and the 3T magnet. Before the beam hits the setup, it goes by the trigger scintillators and the lead brick. Scheme by Arnaud Steen. . .	44
5.9	Photographs of different layers of the setup. Left: front/back train (3 low-density modules) with electronic components (concentrator mezzanine, low voltage power supply, DCDC bias voltage power supply and busbar). Middle: additional train featuring a high-density module with a partial bare hexaboard and partial concentrator mezzanine. Right: System inside the magnet with all the components from Figure 5.8 connected.	45
6.1	Difference between ADC counts of current bunch crossing and ADC counts of previous bunch crossing for all 222 channels of the module for the muon Run 14230133.	53
6.2	MIP Spectrum for channel 121 and 122 for the muon Run 14230133.	53
6.3	2D histogram of trigger time versus ADC counts, with a superimposed profile showing the mean ADC counts at each trigger time. The colour map represents the density of data points, while the profile highlights the central trend across the distribution.	54
6.4	MIP Spectrum after TrigTime selection for channels 121 and 122 for the muon Run 14230133.	55

6.5	Left: 2D histogram showing the correlation between the ADC counts of a given channel and the common mode (CM) ADC value (the average across a group of channels). Each bin represents how often a pair of ADC and CM values occurs. The red line is a linear fit showing a strong positive correlation: when the CM value increases, so does the individual channel's ADC value. This indicates that common mode noise is present, contributing a shared offset to all channels. Right: The same histogram after CM corrections have been applied. Now, the distribution is centred around zero, with no visible correlation. This confirms that the CM noise has been successfully removed from the signal.	56
6.6	MIP Spectrum for channels of Run 14230133 before and after CM corrections.	57
6.7	Top: MIP Spectrum superimposed by the Landau Gaussian convolution function fit to the data points. Bottom: MIP Spectrum superimposed by the Landau Gaussian convolution function fit to the data points (Log Scale). The second pad in the bottom plot shows the ratio between the function and experimental points.	58
6.8	Signal over noise (S/N) for data from several muon runs of the 2023 test beam data. . . .	59
6.9	Distribution of the relative uncertainty of the most probable value (MPV) of the MIP signal for the 2023 test beam	59
6.10	(a) DQM inspection plot of a layer of three modules, yellow channels are channels with the highest signal. Figure compiled by C.Amendola. (b) DQM inspection plot of two layers of three modules, yellow channels with the highest signal, run 1726942607 200GeV electrons 3T. Figure compiled by A.Zghiche.	61
6.11	Monitoring plots showing the pedestal mean (top), noise (middle) and common mode slope (bottom) for the six modules from the first two layers in the 2024 beam test. Each point corresponds to an individual channel, with different colours representing different modules. These quantities are obtained from dedicated pedestal runs. The CM slope indicates the degree of correlation with common mode noise and is particularly relevant for precise signal extraction in MIP analysis.	63
6.12	Monitoring plots showing the pedestal mean (top), noise (middle) and common mode slope (bottom) for the same module 14 in several pedestal runs of the 2024 beam test. Each point corresponds to an individual channel, with different colours representing different runs.	63
6.13	Noise distribution as a function of the cell area, normalised to the full cell area.	64
6.14	RecHit energy in ADC counts as a function of the trigger phase (ns) for Relay 1726518879 and Module 16. The filtering and projection steps isolate a clean and unbiased MIP signal.	66
6.15	RecHit energy in ADC counts as a function of the trigger phase (ns) for Relay 1726347199 Module 14.	66
6.16	MIP Spectrum and Langauss fit for Relay 1726347199 with channel, erx and roc comparison.	68
6.17	Distribution of the signal-to-noise ratio, defined as the most probable value (MPV) of the MIP signal divided by the corresponding noise (σ).	69

6.18 Distribution of the relative uncertainty of the most probable value (MPV) of the MIP signal for the 2024 test beam	69
6.19 Comparison of event throughput (events per second) across CPU and GPU configurations using CMSSW reconstruction workflows. Each coloured line corresponds setup: CPU and GPU jobs running with 4 and 8 threads, respectively (previous results from Izaak Neutelings) and CPU result with 1 thread from data of the current test beam.	71
A.1 MIP Spectrum for 8 channels. Each row displays two plots, showing the results for different channels (ML14 channels 10 and 25, ML16 channels 149 and 151, ML17 channels 36 and 9, ML19 channel 153 and ML20 channel 64) in several modules for Relay 1726518879.	80
A.2 MIP Spectrum for 8 channels. Each row displays two plots, showing the results for different channels(ML14 channels 27 and 34, ML16 channels 153 and 210, ML17 channels 29 and 33, ML18 channel 24 channel 33) in several modules for Relay 1726347199.	81
A.3 MIP Spectrum for all the eRx (from 0 to 5 in order) for Module 16 from Relay 1726347199.	82

Nomenclature

Greek symbols

β	$\frac{v}{c}$
η	Pseudorapidity
γ	Lorentz factor $\frac{1}{\sqrt{1-\beta^2}}$
λ	Interaction length
ρ	Pearson correction factor
σ	Standard deviation
θ	Angle between the particle $\mathbf{p}$ and the positive direction of the beam axis

Roman symbols

$m_e c^2$	Electron mass $\times c^2$
N_A	Avogadro number
p_T	Transverse momentum
r_e	classical electron radius
T_{max}	Maximum possible energy transfer to an electron in a single collision
X_0	Radiation length
$\mathbf{p}$	Three-momentum
A	Atomic mass of absorber [g mol ⁻¹]
I	Mean excitation energy [eV]
K	Coefficient $4\pi N_A r_e^2 m_e c^2$
Z	Atomic number of absorber
z	Charge number of incident particle

Glossary

$\mathcal{B}$	branching ratios
ADC	Analog to Digital Converter
ASIC	Application Specific Integrated Circuit
BE	Back-end
BSM	Beyond the Standard Model
BX	Bunch crossing
CE	Endcap Calorimeter
CE-E	Electromagnetic CE
CE-H	Hadronic CE
CERN	European Organization for Nuclear Research
CM	Common Mode
CMS	Compact Muon Solenoid
CMSSW	CMS Software
CP violation	violation of charge-parity symmetry
DAQ	Data Acquisition
DQM	Data Quality Monitoring
ECAL	electromagnetic calorimeter
ECON-D	DAQ e-link Concentrator
ECON-T	Trigger e-link Concentrator
eRx	ePort line receivers
EW	Electroweak
FC	Fast Control
FE	Front-end
HCAL	hadronic calorimeter
HD	High Density
HGCAL	High-Granularity Calorimeter
HGCROC	HGCAL ROC
HL-LHC	High Luminosity LHC
LD	Low Density
LHC	Large Hadron Collider

IpGBT	Low-Power GigaBit Transceiver
MIP	Minimum Ionising Particle
MPV	Most Probable Value
PCB	Printed Circuit Boards
PCC	Pearson correlation coefficient
PDFs	parton distribution functions
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
RFQ	radio frequency quadrupole
SC	Slow Control
SiPM	Silicon Photomultipliers
SM	Standard Model
SPS	Super Proton Synchrotron
TC	Trigger Cells
TOA	Time of Arrival
TOT	Time over Threshold
TPG	Trigger Primitive Generator
VTRX+	Versatile Link Plus Transceiver

Chapter 1

Motivation and outline

High-energy particle physics is responsible for advancing our understanding of the fundamental constituents of the universe. The quest to probe the smallest building blocks of matter has led to the development of particle detectors with remarkable discoveries in the past century. Developed in the early 1970s, the Standard Model of Particle Physics has successfully explained nearly all experimental results and made precise predictions across a wide range of phenomena. The progress and confirmation of the Standard Model would not have been possible without continuous advancements in particle detector technology, which play a crucial role in observing and measuring the fundamental particles and interactions that the theory describes.

As part of the Compact Muon Solenoid (CMS) experiment upgrades for the High-Luminosity Large Hadron Collider (HL-LHC), high-granularity sampling calorimeters are being implemented to enhance spatial resolution and particle identification capabilities. In addition, precision timing capabilities are being integrated directly into the calorimeters. These upgrades will allow the detector to measure not only the energy and position of particles, but also their time of arrival with a resolution of tens of picoseconds. As part of this strategy, the upgraded endcap calorimeter will provide measurements of energy and time for particles using approximately 6 million channels. Traditional three-dimensional vertex reconstruction can be enhanced by incorporating timing information, resulting in a four-dimensional fit that explicitly accounts for the spatial position of vertices along the beamline and their timing within the bunch crossing. This advancement is particularly valuable in high-pileup conditions, as it allows for the precise identification of the primary vertex among multiple reconstructed pileup vertices. The integration of timing data from calorimeters and the Minimum Ionising Particle (MIP) timing detector further improves primary vertex identification, enhancing the accuracy of particle flow reconstruction. Additionally, the use of a highly granular calorimeter contributes not only to mitigating pileup effects but also to achieving finer resolution in particle flow measurements, significantly advancing the capabilities of vertexing and event reconstruction.

According to this, the project has the goal of exploring the algorithm calibration of the energy and time measurements for the High-Granularity Calorimeter (HGCal) using data from test beams to establish a fast local reconstruction algorithm, using MIP signals. The MIP peak serves as a fundamental

reference point for calorimeter energy calibration, accurately identifying and fitting this peak is essential for translating detector signals into meaningful energy measurements. Precise calibration is essential for achieving the detector's design performance. Some of the results shown in this work have been included in a detector performance summary note approved by the HGCAL Project and the CMS Collaboration [1]. There are several challenges faced when automating a Minimum Ionising Particle (MIP) scale determination since it comes at the end of a chain of processes including the identification of channels impinged by a particle, the pedestal and noise mitigation and the selection of the best timing phase. In test beams, this is not easily automated. There is a need to have different approaches for focused beams or defocused beams.

This thesis is structured to guide the reader through the theoretical and experimental foundations that support the development of a calibration method for HGCAL. In Chapter 2, key concepts relevant to the work are introduced, including the Standard Model of Particle Physics, the main features of the LHC, and the current design of the CMS experiment. Chapter 3 then explores the principles of calorimetry and the mechanisms of energy loss in matter, with a focus on the distinction between homogeneous and sampling calorimeters, setting the stage for understanding the technology behind HGCAL. Chapter 3 is dedicated to the detailed design of HGCAL itself, describing how the detector modules are built, how signals are read out, and how data flows through the system. Building on this, Chapter 5 presents the test beam campaigns that were essential to this work. The setup of the 2023 test beam is described, which enabled the initial development of the calibration method, as well as the 2024 campaign at CERN, in which I actively participated. This chapter also details the full calibration procedure, including the method for extracting the Most Probable Value (MPV) of Minimum Ionising Particle (MIP) signals. In Chapter 6, the results of applying the calibration method and its integration into the local reconstruction algorithm are presented, evaluating its performance under realistic conditions. Finally, Chapter 7 concludes the thesis with a summary of the main contributions and a discussion of the relevance of this work for the successful operation of HGCAL in the high pileup environment at the HL-LHC.

Chapter 2

Introduction

2.1 The Standard Model of particle physics

The Standard Model (SM), a cornerstone of modern physics, was developed through decades of theoretical and experimental effort since the mid-20th century [2]. It provides a, so far, exact description of different generations of particles and their interactions through the carriers of the electromagnetic, weak and strong forces. The fundamental particles can be divided into matter particles (fermions), the mediators of interactions (gauge bosons), and the Higgs field (scalar boson). These are represented in Figure 2.1. Fermions have half-integer spin and are known as quarks or leptons depending on whether they carry or not a strong color charge. Gauge bosons have integer spin (vector bosons) and mediate three fundamental interactions. Gluons mediate the strong force, described by Quantum Chromodynamics (QCD), photons mediate the electromagnetic interaction, and the W and Z bosons mediate the weak force. The latter interactions have been demonstrated early to be unifiable as a single Electroweak (EW) interaction, which can be modelled as a gauge interaction, in a $SU(2) \times U(1)$ group (see e.g. [3]).

In this picture, the Higgs mechanism is responsible for giving mass to the EW bosons: W and Z, confining their interactions to short ranges. In addition, its Yukawa-type of interactions with the fermions gives rise to the mass of the particles. This mechanism was proposed in 1964 [4], and was experimentally confirmed recently in 2012, at the LHC [5, 6].

Several experiments at lepton and hadron colliders have confirmed the reliability of the SM predictions. Recent results of the LHC extend this confirmation with unprecedented statistics at the highest energy scale achieved [8]. However, despite the SM success and precision of experimentally measured parameters, it can not be a complete theory of fundamental interactions since it lacks explanation and reasoning for:

- Gravitation, one of the fundamental forces. Due to the weak nature of gravity in comparison to other fundamental forces and the incompatibility with general relativity, gravity can't, so far, be incorporated into the SM;
- The realisation that the violation of charge-parity symmetry (CP violation) in the SM is insufficient to explain the observed matter-antimatter asymmetry of the Universe;

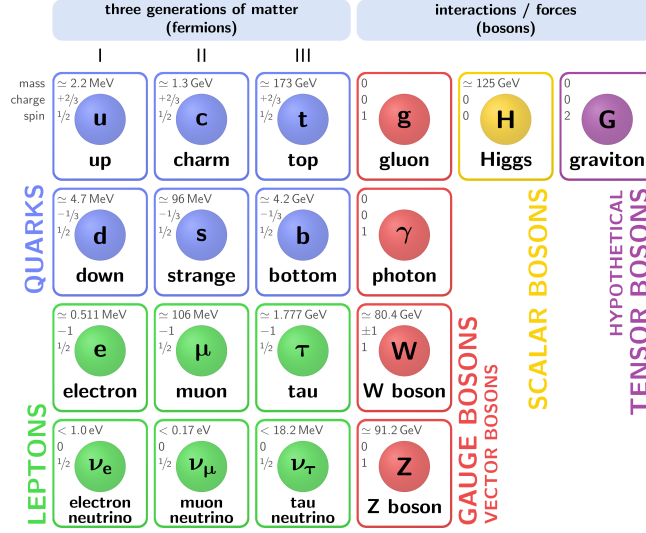


Figure 2.1: Elementary particles in the Standard Model, showing the spin, charge and approximate values for the mass of each particle, obtained experimentally [7].

- The massless conception of neutrinos in the standard model and the observation of their oscillation, implying a contradiction with this assumption;
- Evidence from astronomical measurements for the existence of an additional type of weakly-interacting matter: dark matter.

In addition to these puzzles, there are several questions about the recently found Higgs boson and the precise understanding of the electroweak symmetry breaking. At the LHC, most Higgs boson coupling constants to other particles have, so far, been measured with a precision of 5–10%, leaving a significant leeway for deviations from the SM prediction, as well as the possibility of coupling to new, yet unobserved particles [9].

So far, the efforts made to measure the Higgs boson properties experimentally have been limited by the availability of data [10] (approximately 360 fb^{-1} accumulated by the CMS experiment since the LHC start in 2010).

Given the complex structure of protons, whose constituents are described by parton distribution functions (PDFs), different production modes are possible for a Higgs boson, they depend on both the center-of-mass energy ($\sqrt{s}$) and the proton constituents involved in the interaction. While the majority of the Higgs is expected to be produced through gluon fusion, a significant part is produced through the fusion of vector bosons, or in association with vector bosons. At a lower rate, the Higgs can be produced with a pair of top or bottom quarks. The experimental measurements of the Higgs boson depend on both the SM cross sections, which quantify an effective interaction area for the production of the Higgs after a collision of two protons [11], and the branching ratios ($\mathcal{B}$) of the Higgs decays. Experimentally, these quantities are implicit when selecting a pre-determined event topology and contain intrinsic information on different couplings, which may contribute to the loops. The experimental status of the measurements in different final states is summarised in Table 2.1. For a given initial (i) and final (f) states, the signal strength parameter, μ , for individual production channels $\mu_i = \sigma_i/(\sigma_i)_{\text{SM}}$ and for decay

modes is defined as $\mu^f = \mathcal{B}^f / (\mathcal{B}^f)_{\text{SM}}$, where σ is the production cross-section and $\mathcal{B}$ is the branching fraction. Agreement with SM expectations yields $\mu = 1$. As can be observed, the results have, in most cases, large uncertainties and would benefit from a significant increase in data.

Particle Type	$\mathcal{B}$	Rel. Uncertainty %	Signal strength
Quarks			
$H \rightarrow b\bar{b}$	0.582	± 0.65	$1.05^{+0.22}_{-0.21}$
$H \rightarrow c\bar{c}$	2.89×10^{-2}	± 1.2	
$H \rightarrow t\bar{t}H$	-	-	$0.94^{+0.08}_{-0.07}$
Leptons			
$H \rightarrow \tau^+\tau^-$	6.27×10^{-2}	$+1.17, -1.16$	$0.85^{+0.10}_{-0.10}$
$H \rightarrow \mu^+\mu^-$	2.18×10^{-4}	± 1.23	$1.21^{+0.45}_{-0.42}$
Vector Bosons			
$H \rightarrow WW^*$	0.214	± 0.99	$0.97^{+0.09}_{-0.09}$
$H \rightarrow ZZ^*$	2.62×10^{-2}	± 0.99	$0.97^{+0.12}_{-0.11}$
Photons			
$H \rightarrow \gamma\gamma$	2.27×10^{-3}	$+1.73, -1.72$	$1.13^{+0.09}_{-0.09}$
$H \rightarrow Z\gamma$	1.53×10^{-3}	± 5.71	$2.59^{+1.07}_{-0.96}$
Gluons			
$H \rightarrow gg$	8.19×10^{-2}	$+3.40, -3.41$	$0.97^{+0.08}_{-0.07}$

Table 2.1: Main Standard Model branching ratios $\mathcal{B}$ of the Higgs boson and their relative theoretical uncertainties for $m_H = 125 \text{ GeV}$ [12, 13]. The last column shows the experimental signal-strength parameters, which quantify the deviations from the SM predictions for both production modes and decay channels. The uncertainties comprise statistical and systematic components [9].

In this context, the High Luminosity phase of the LHC (HL-LHC) is expected to refine significantly the program of measuring the Higgs couplings to the order of 1% uncertainty. Additionally, it is expected to provide the first observations of double Higgs production may become possible at the LHC with an anticipated precision of approximately 50% for the coupling parameter [14]. Some final states of the Higgs boson will remain, however, entirely unexplored at the HL-LHC, for example, the decay of the Higgs boson to gluons or electrons.

Precision studies of the Higgs sector are crucial for advancing our understanding of particle physics and potentially uncovering new physics beyond the SM. Key objectives include measuring the Higgs boson's interactions with all SM particles, particularly the top quark due to its strong coupling to the Higgs; determining the Higgs boson's self-coupling to map the shape of the Higgs potential, which influences early-universe cosmology; investigating potential interactions between the Higgs boson and hypothetical particles, such as dark matter candidates or a hidden sector, through rare decays and precise measurements and leveraging Higgs studies to test Standard Model predictions involving fundamental parameters (e.g., Fermi constant, Weinberg angle, W/Z boson masses), where discrepancies could signal new physics [15]. This comprehensive exploration of the Higgs sector has the potential to reshape the understanding of the universe at its most fundamental level.

The study of the Higgs boson is inherently multidisciplinary, requiring knowledge of a full detector system and the reconstruction of a wide variety of final states. This involves an almost complete coverage of the solid angle, allowing identification and measurement of numerous decay products. In this con-

text, calorimeters play a central role, as they are responsible for the reconstruction of many key physics objects, such as jets, electrons, photons, hadronically decaying taus and, indirectly, missing transverse energy. Their performance is therefore crucial to the success of the Higgs program. In the HL-LHC, the upgrade of the calorimeters is therefore essential as they'll provide the necessary means to accomplish the ambitious physics programs. This makes the development and deployment of high-granularity calorimeters like the HGCal a strategic asset for future precision Higgs measurements.

2.2 From collision to detection

Accelerator physics experiments are generically categorised in two types: fixed-target experiments, where a beam of particles is directed at a static target, and beam-beam collisions, where two particle beams are accelerated in opposite directions and made to collide near an experimental apparatus. Two parameters are of relevance. The first one is the center-of-mass energy $\sqrt{s}$ achieved in the collisions. For a beam-beam collision, this can be calculated as the square root of the Lorentz invariant s , from the momentum and energy of the colliding particles [16]:

$$s = (E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2 \quad , \quad (2.1)$$

where natural units where $c = 1$ are implicit. Simultaneously, to make the generation of events of interest from the collision effective, a particle accelerator must generate a large number of collisions per unit of time to provide sufficient data. This rate depends on the cross sections of the processes of interest and a parameter known as instantaneous luminosity $\mathcal{L}$. At the CERN LHC, the beam particles are grouped in bunches that cross at a fixed frequency of 40 MHz in each of the four interaction points. Surrounding or capturing partially the outcome of the interactions lie the four detectors: ATLAS, CMS, ALICE and LHCb. The instantaneous luminosity can be calculated as being proportional to the number of particles in each bunch (N_1 and N_2) and inversely proportional to the transverse beam sizes across x and y directions (σ_x^* and σ_y^*). Furthermore $\mathcal{L}$ scales proportionally to the crossing frequency f as:

$$\mathcal{L} = \frac{f}{4\pi} \frac{N_1 N_2}{\sigma_x^* \sigma_y^*} \quad . \quad (2.2)$$

The expected number of events is estimated by the product of the luminosity integrated over the run time of the experiment and the cross section,

$$N_{\text{exp}} = \sigma \int \mathcal{L}(t) dt \quad . \quad (2.3)$$

Taking into consideration that detectors have an associated acceptance A , that is the ratio of events in which the final state of a particle reaches said detector and an efficiency for detection ε , inverting equation 2.3,

$$\sigma = \frac{N_{\text{obs}}}{\varepsilon A \int \mathcal{L}(t) dt} \quad , \quad (2.4)$$

thus making a cross-section measurement from the number of experimentally observed events.

To accurately measure the energy and properties of the resulting collision particles, detectors are meticulously designed to exploit the different ways particles interact with matter. In the context of this thesis, particular attention is paid to calorimeters. These detectors measure the energy of electromagnetic and hadronic showers through interactions with dense absorber materials. Three key parameters taken into consideration in the calorimeter design are the radiation length (X_0), the Molière radius R_M and the interaction length λ . These parameters determine the longitudinal and transverse depth required for a calorimeter to fully absorb the energy of particle showers.

The radiation length provides a measure of how electromagnetic showers develop in a material longitudinally. It is defined as the mean distance over which an electron loses all but $1/e$ of its initial energy. For high-energy electrons ($> \text{GeV}$), this loss occurs via Bremsstrahlung ($e \rightarrow e\gamma$). Based on the same definition, a photon beam reduces its energy by $1/e$ after $9/7 X_0$. For the same energy range of interest for the LHC detectors, this occurs primarily through pair conversions in the material ($\gamma \rightarrow e^+e^-$). The radiation length (X_0) can be approximated by the following formula,

$$X_0 = \frac{716.4 \times A}{Z(Z+1) \cdot \ln \frac{287}{\sqrt{Z}}} \cdot \frac{1}{\rho} \quad , \quad (2.5)$$

where A and Z is the atomic number and weight and ρ is the density of the material.

The charged particles of the showers that traverse the material of the calorimeters are affected by multiple Coulomb scatterings. This leads to an opening of the showers in the lateral plane. The so-called Molière radius (R_M) is used to quantify this spread. It represents the radius of a cylinder around the shower that contains 90% of the energy deposited by the shower in the absorbers. A smaller R_M indicates a more compact shower, which is desirable for achieving fine spatial resolution and minimising overlap between nearby particles in high-density environments. High- Z materials and air gaps are typically responsible for larger R_M . This quantity can be approximated by:

$$R_M = 21.2 \frac{X_0}{E_c} \quad , \quad (2.6)$$

where the critical energy E_c is defined as the energy where ionisation takes over.

Finally, the nuclear interaction length (λ) is used to estimate the behaviour of the hadronic shower. This quantity corresponds to the average distance that a high-energy hadron travels in a material before undergoing an inelastic nuclear interaction. This parameter guides the required depth of hadronic calorimeters to fully contain the more diffuse and extended hadronic showers, compared to the electromagnetic showers. Typically, at least 10λ are required to contain a hadronic shower.

$$\lambda = \frac{A}{N_A \sigma_{\text{total}}} \quad . \quad (2.7)$$

These material-dependent parameters are critical for calorimeter design and influence the choice of absorber materials and segmentation. Materials with short radiation lengths and small Molière radii, such as lead or tungsten, are preferred in electromagnetic calorimeters for compact and efficient shower

containment, while hadronic calorimeters prioritise sufficient depth using materials with suitable interaction lengths.

The choice of the technology for the calorimeter usually follows from the physics targets to be achieved. For the LHC, these comprise precision measurements using photons, electrons and jets but also searches for new physics (e.g. with delayed decays in the detector volume).

2.3 Large Hadron Collider

The Large Hadron Collider is the world's largest and most powerful particle accelerator. It has been running since 2008 at European Organization for Nuclear Research (CERN) and it is located in Geneva, Switzerland. In the 27 km circumference, beams of different species (protons, heavy ions) collide every 25 ns. In Run 3 (2023-26) proton beams collide at 13.6 TeV of energy with an average of 60 simultaneous collisions (pileup).

Particle accelerators increase the energy and the velocity of particles typically through radio frequency (RF) cavities and control their trajectory by electromagnetic fields created by dipole and higher-order magnets. In the LHC, 16 superconducting RF cavities operating at 400 MHz deliver finely tuned electromagnetic impulses, accelerating protons from 450 GeV to 6.8 TeV over millions of revolutions. The field in an RF cavity is made to oscillate at a given frequency, so timing the arrival of the bunched particles is important.

Before being accelerated in the main LHC ring, protons pass through a series of preparatory accelerators, as seen in Figure 2.2. The process begins with the ionization of hydrogen gas using high-energy electrons, followed by extraction via a 90 kV voltage. The protons are then sent to a radio frequency quadrupole (RFQ), where they are accelerated to 750 keV, focused, and organized into separate bunches. This term refers to the groups of protons that undergo multiple stages of acceleration to reach the desired energy. A RFQ is a type of linear accelerator that makes use of oscillating electric fields within a quadrupole electrode structure to both focus and accelerate charged particle beams. It operates at radio frequencies and is effective at capturing and bunching low-energy ions, preparing them for injection into subsequent accelerator stages.

The bunches are further accelerated to 50 MeV by LINAC2, a linear accelerator, and then to 1.4 GeV by the Proton Synchrotron Booster (PSB). The PSB is made up of four superimposed synchrotron rings that receive beams of negative hydrogen ions (H^-) from the linear accelerator, then H^- ions are stripped of their two electrons during injection from the linear accelerator into the PSB to leave only protons, which are accelerated for injection into the Proton Synchrotron (PS). When the PS takes over, boosting the energy to 25 GeV. Finally, the Super Proton Synchrotron (SPS), a 6.9 km synchrotron, accelerates the beam to 450 GeV. At this stage, the beam is split into two, with the opposing beams transferred into the LHC for final acceleration.

The protons are separated into a maximum of 2556 bunches, each containing 1.15×10^{11} protons. Due to the density of the beam and the collision rate, the average instantaneous luminosity at each interaction point is typically $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. When the bunches intersect at the detector points, several

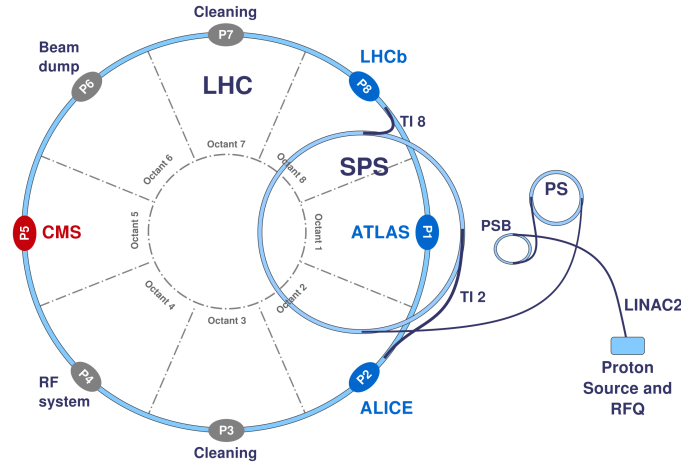


Figure 2.2: Schematic overview of the CERN accelerator complex (not to scale). The complex chains the production of H^+ ions with the beam-beam collisions at each of the four interaction points. The production-to-collision sequence follows the following order: proton source - RFQ - LINAC2 - PSB - PS - SPS - LHC [17].

protons may end up colliding, resulting in simultaneous collisions or vertices. This phenomenon is called pileup. Different processes may occur and generate different particles which emerge from the primary vertex of each collision. As the collision happens, particle detectors measure direction, time of arrival, electric charge and energy deposits. For this to happen, detectors combine different kinds of components such as trackers, electromagnetic calorimeters, hadronic calorimeters and particle identification detectors based on velocity.

To ensure the safe and efficient operation of the LHC, several complementary systems are crucial beyond the chain of accelerators. Once protons reach their peak energy, they are either directed into collisions or safely absorbed by the beam dump system, which is a massive graphite and metal block designed to capture the energy of the beam without damage. Given the intensity of the beam, uncontrolled losses could be catastrophic, which is why precise beam cleaning and collimation systems are employed throughout the ring. These remove stray particles that deviate from the core beam path, protecting both the LHC components and the detectors. The performance of the systems relies on an ultra-clean environment inside the beam pipes, maintained by a rigorous vacuum and a procedure known as scrubbing. During scrubbing runs, low-energy beams are intentionally circulated to dislodge microscopic gas molecules and generate electron clouds, which in turn condition the beam pipe surfaces. This self-cleaning effect improves vacuum quality and stabilises beam dynamics for high-intensity physics runs.

The LHC began its first operational run (Run 1) in 2009, achieving collision energies of 7 TeV [18]. A short shutdown occurred in 2012, after the discovery of the Higgs boson, before planned upgrades resumed in 2013. During this period, lead-ion collisions were conducted, providing valuable data on extreme conditions shortly after the Big Bang, 10^{-6} s to be precise. The LHC restarted operations in 2015 after a shutdown, during which major upgrades were completed to enhance beam energy and improve detector capabilities, achieving a collision energy of 13 TeV. The focus during Run 2 was on increasing

luminosity, leading to a significant boost in collision rates and data collection [19]. The next phase, Long Shutdown 2 (LS2), took place from 2018 to 2022, aimed at further upgrades for the HL-LHC. These upgrades were designed to boost luminosity by a factor of 10, improving the LHC's capacity for scientific discovery. The shutdown concluded in 2022, preparing the LHC for Run 3, which began with a new maximum collision energy of 13.6 TeV [20]. Run 3 focused on further increasing luminosity, expected to continue until 2026, with subsequent upgrades planned for the HL-LHC to enhance performance even further. The LHC Run plan and the HL-LHC schedule can be observed in Figure 2.3. The HL-LHC is set to amplify LHC's capabilities, increasing the luminosity by a factor of 10, and commencing operations in 2030. This upgrade will provide proton-proton collisions at $\sqrt{s} = 14$ TeV, enabling precision studies on the Higgs Boson within the SM and looking Beyond the Standard Model (BSM). The HL-LHC's advanced measurements will include those of rare events like the di-Higgs production, which will provide a direct measurement of the Higgs triple self-coupling. Moreover, the project aims to enhance the understanding of the Higgs mechanism and how the mass of the heaviest contributes to the stability of the Higgs potential.



Figure 2.3: The High-Luminosity Upgrade plan for the LHC at CERN, including the different past runs and corresponding $\sqrt{s}$ [21].

To increase the number of events produced the LHC will employ more densely packed proton bunches, resulting in a higher luminosity (up to $7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$), increasing the number of simultaneous collisions, i.e. pileup. It is expected that each bunch crossing at the HL-LHC will produce 140-200 pileup collisions, which poses challenges in accurately reconstructing particle tracks and identifying relevant interactions. The CMS detector, originally designed for a maximum pileup of around 60 interactions per crossing (proton bunches crossing every 25 ns), will not be able to operate under the new luminosity conditions.

The proposed solution involves replacing, where needed, the detectors with improved radiation-hard

elements. Increasing the readout granularity both in the sensitive detectors and in the readout electronics provides better discrimination of nearby particles from different collisions. Furthermore, deploying precision timing information will leverage the non-simultaneity of the collisions within a proton bunch crossing.

2.4 CMS Experiment and the High Luminosity LHC

The CMS detector is built around a solenoid magnet that takes the form of a cylindrical coil of superconducting cable that generates a magnetic field of 4 T, around 100,000 times greater than Earth's magnetic field. CMS has already contributed to outstanding confirmations, such as the experimental confirmation of a particle consistent with the Higgs Boson. As described previously in section 2.3, CMS has undergone several LHC runs since 2010, approaching the limit for performance at this energy level.

The innermost subdetector, the Silicon Tracker, is composed of two main technologies: pixel detectors near the beam pipe, which provide high-resolution position measurements critical for identifying primary and secondary vertices, and strip detectors farther out, which extend tracking over a larger volume. The fine granularity and coverage of the tracker allow efficient reconstruction of charged particles in dense environments, such as those created in high-luminosity collisions. Surrounding the tracker, the electromagnetic calorimeter (ECAL) is made of lead tungstate crystals, chosen for their fast response and high density. The hadronic calorimeter (HCAL), placed outside the ECAL, uses layers of brass absorber and plastic scintillator to measure the energy of hadrons. It plays a key role in the overall energy balance of each event and is essential for identifying jets and calculating missing transverse energy. Surrounding the calorimeters and embedded in the steel return yoke of the magnet, the Muon System consists of various technologies, including Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). These provide spatial and timing resolution, allowing efficient identification and momentum measurement of muons even at high energies and in forward regions. Together, these subdetectors work in a coordinated way to deliver the comprehensive reconstruction of events.

Starting from the beam interaction region, particles first pass through a Si tracker, where the trajectories, tracks of charged particles and their vertices are reconstructed using signals, hits from the tracker's sensitive layers. The tracker operates within a magnetic field, which bends the trajectories, enabling the determination of the electric charges and momenta of the charged particles. Electrons and photons are then absorbed in the ECAL, where their resulting electromagnetic showers are detected as clusters of energy in neighboring cells. These clusters allow for the measurement of the particles' energy and direction. Charged and neutral hadrons may also produce hadronic showers in the ECAL, which are subsequently fully absorbed in the HCAL. The resulting energy clusters in the HCAL help estimate the energies and directions of these hadrons. Muons and neutrinos, on the other hand, pass through the calorimeters with minimal or no interaction. While neutrinos escape detection, muons produce signals in additional tracking layers, known as muon detectors, situated outside the calorimeters. This process is illustrated in Figure 2.4, which provides a schematic transverse view of the CMS detector.

In the CMS detector, which has cylindrical geometry symmetry along the z-axis (beam axis), it is

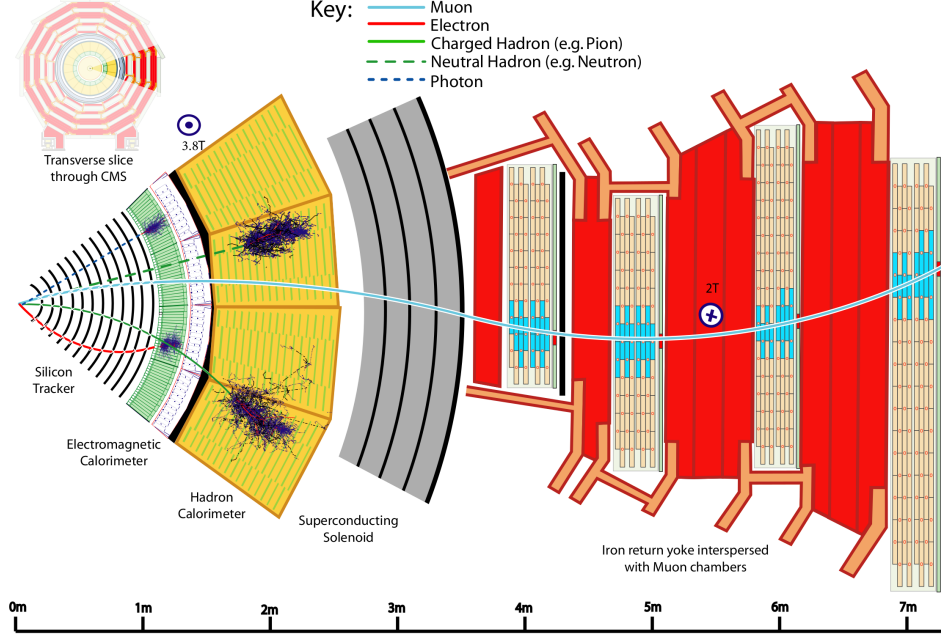


Figure 2.4: Schematic of a transverse slice of the CMS detector showing specific particle interactions, from the beam interaction region to the muon detector [22].

essential to accurately describe the angles of particles from high-energy collisions. Considering the physics events generated from the interaction point, it is reasonable to use spherical coordinates (r, θ, ϕ) to describe physics events and the detector system. However, at the TeV scale, the center-of-mass frame is often boosted along the beam axis, making the polar angle θ less convenient for comparing particle trajectories between events since the $\Delta\theta$ measurement between the two fragments that originate from the interaction point is no longer Lorentz invariant. The rapidity is introduced since it remains stable under Lorentz boosts along the beam direction. The problem with rapidity is that it can be hard to measure for highly relativistic particles. Pseudorapidity is an approximation to rapidity, equivalent to rapidity for massless particles, derived from the θ and it is defined as,

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] , \quad (2.8)$$

and it is particularly useful in the LHC, where the composite nature of the colliding protons means that the interactions rarely have their center-of-mass frame coincident with the detector rest frame, and where the complexity of the physics means that η is easier to estimate than the rapidity. The CMS detector layout and subcomponents: tracker, muon systems, calorimeters are segmented in η instead of θ , allowing for a more consistent and meaningful way to map and analyse particle trajectories and detector coverage, especially in the forward and backward regions.

The CMS layout and architecture will change significantly with the HL-LHC. The Tracker will be completely replaced with new pixel and strip tracking detectors that offer higher granularity and extend coverage in pseudorapidity from $\eta = 2.5$ to $\eta = 3.8$ [23]. Surrounding the outer tracker, a new timing detector will measure the time-of-arrival of tracks with a resolution of 35 ps, providing additional discrimination

against pileup [24]. The Front-end (FE) of the ECAL will be upgraded to allow access to calorimetric information already at the level-1 trigger stage, while new readout electronics will be installed for the barrel calorimeters. The muon system will also undergo significant improvements, including the addition of advanced Resistive Plate Chambers (RPCs) and Gas Electron Multipliers (GEMs), which will enhance triggering efficiency, improve momentum resolution in the forward region, and extend coverage up to $\eta \approx 3$ [25]. The endcap calorimeters will be fully replaced to meet the demands of high radiation and event rates with the High-Granularity Calorimeter, which will be explained in detail in Chapter 4. Particular emphasis will be placed on the inter-calibration of the HGCAL, which constitutes the central focus of this thesis.

Chapter 3

Energy Loss in Matter and Calorimetry

Calorimeters play a crucial role in particle physics experiments in particle energy measurements. Calorimeters are destructive components of the detectors, because when an incident particle enters the calorimeter, it interacts with the material and produces a shower of secondary particles, which produce detectable signals, and their energies can be several orders of magnitude smaller than the incident particle.

The evolution of the particle showers depends on the type of incident particle, but the calorimeter's goal remains the same: shower containment in the calorimeter's components for accurate measurements. In general, one distinguishes showers that are initiated by electromagnetically interacting particles and showers due to an incident hadron. While the fundamental principles of calorimetry are conceptually straightforward, their practical implementation in experiments is highly challenging due to the complex interplay of multiple concurrent interactions within the material.

3.1 Interactions of particles in matter

The most abundant secondary particles in a typical calorimeter used in high-energy particle physics are electrons, photons, positrons, charged and neutral hadrons. Those particles interact with the calorimeter material in various ways. In this section, the underlying interaction processes are summarised.

Charged particles interact with the electromagnetic fields of the atoms in the material, they can interact in various ways depending on their momentum. The most important processes are the elastic scattering with shell electrons (low energies), excitation or ionisation (intermediate energies) and interaction of the Coulomb field of a nucleus and emission of Bremsstrahlung (high energy). The mean rate of energy loss by moderately relativistic charged heavy particles, also referred to as mass stopping power in units of $\text{MeV cm}^2 \text{g}^{-1}$, can be described by the Bethe-Bloch equation

$$\left\langle -\frac{dE}{dx} \right\rangle = Kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e \beta^2 \gamma^2 T_{\max}}{I^2} \right) - \beta^2 - \delta(\beta\gamma) \right] . \quad (3.1)$$

This expression is valid for heavy particles in the range of $0.1 \leq \beta\gamma \leq 100$, where energy loss occurs

primarily through interactions with atomic electrons, leading to ionisation of the material. However, at sufficiently high energies, radiative processes become more important than ionisation for all charged particles, as can be seen in Figure 3.1. Figure 3.1 shows the stopping power of heavy charged particles in different materials in the Bethe region, the red curve depicts the mass stopping power for positive muons in copper as a function of their momentum.

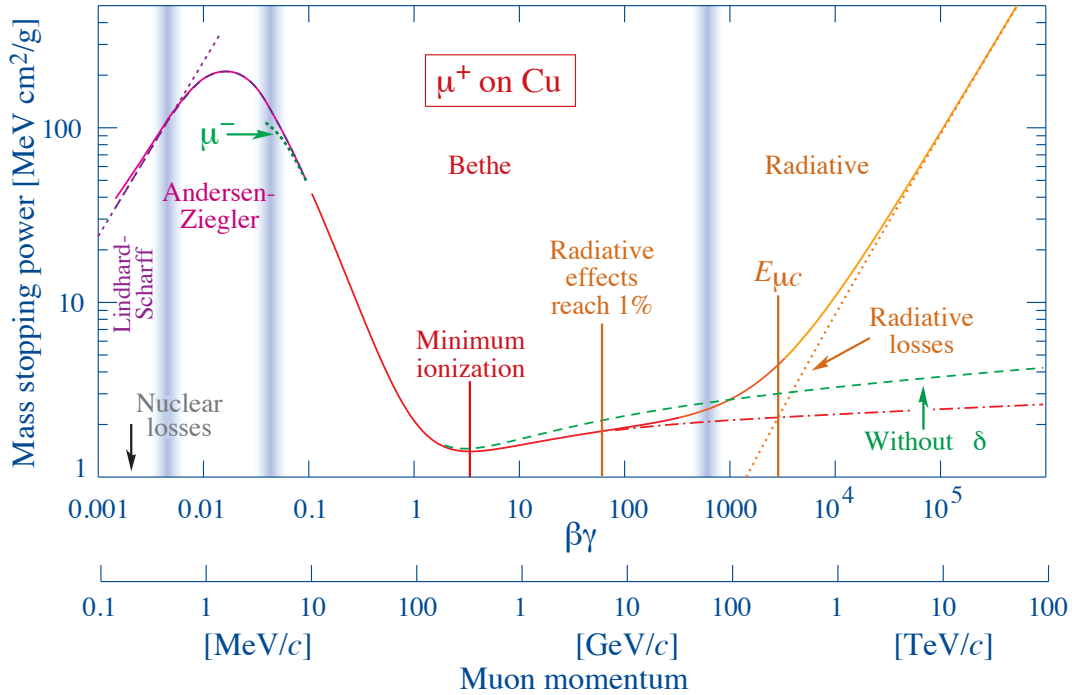


Figure 3.1: Mean rate of energy loss, also called mass stopping power, of positive muons in copper. The vertical blue bands indicate the boundaries between different approximation regions [26].

The qualitative behaviour of the stopping power is similar for all solid materials. Gases have a slightly different stopping power curve at higher energies (like hydrogen and helium) [26], although there is a slight decrease in the rate of energy loss with increasing Z . The qualitative behaviour difference at high energies between gas and the other materials is due to the density-effect correction $\delta(\beta\gamma)$.

$\left\langle -\frac{dE}{dx} \right\rangle$ represents the average energy loss of a particle. However, the distribution of energy deposition is not symmetric and is biased due to rare high-energy deposition events. As a result, the most probable energy loss is lower than the average energy loss predicted by the Bethe equation.

Above the minimum ionisation point, around $\beta\gamma \approx 3$, the most probable energy loss remains nearly constant over a wide energy range. Particles with this behaviour are known as MIPs. Some relativistic particles can be considered MIPs because their most probable energy loss remains near the minimum ionisation level even at momenta above $\beta\gamma \approx 3$. For this reason, the most probable energy loss is frequently used for detector calibration. In a thin layer of material, the incident particles undergo few collisions. As a result, the amount of deposited energy still follows a biased distribution with its peak at the most probable energy deposition and a longer tail to higher values of energy, for high-energy single events.

Figure 3.2 shows the energy loss rate for muons as a function of momentum, with a minimum, the

minimum ionising point, corresponding to $\beta\gamma \approx 3$. While muons are often treated as MIPs due to their high momentum and nearly constant stopping power, their energy loss is slightly above the exact minimum. This small offset must be taken into account to avoid systematic bias in MIP calibration procedures.

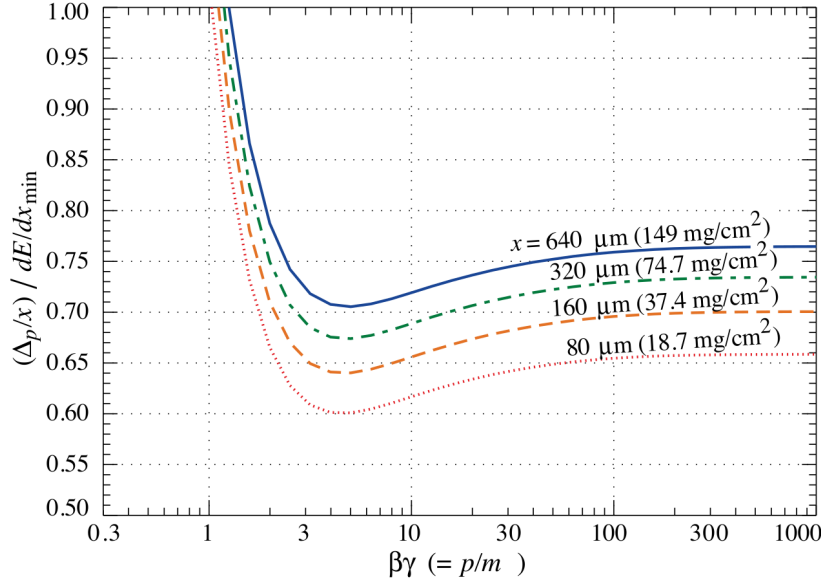


Figure 3.2: Most probable energy loss in silicon, scaled to the mean loss of a minimum ionizing particle, $388 \text{ eV}/\mu\text{m}$ ($1.66 \text{ MeV g}^{-1} \text{ cm}^2$) [26].

The Landau distribution is commonly used as a parametrisation for the energy deposition by ionisation in a thin layer [26]. This distribution has a location parameter that determines the position of the peak and a scale parameter that determines how broad this distribution is. As the thickness of the detector increases, more individual collisions contribute to the total ionisation, and the sum of the deposited energy approaches a Gaussian. The convolution of a Landau and a Gaussian distribution is often referred to as “Languass” function, and it is used to model a MIP signal. The signal area corresponding to one MIP can be determined by fitting such a “Languass” function to the histogram of the signal areas. The position of the peak is extracted from the fit result and serves as the calibration constant and is used as a reference or standard for calibrating and measuring the performance of silicon sensors [24].

The mechanisms by which particles interact with and lose energy in matter depend strongly on the particle type and energy. While all particles are subject to electromagnetic interactions to some degree, the dominant energy loss processes vary significantly between electrons, photons, and hadrons due to differences in mass, charge, and interaction cross sections.

- **Electrons:** The Bethe formula is not directly applicable to electrons because it assumes that the incident particle is much heavier than the atomic electrons, allowing simplifications such as negligible deflection of the projectile. In the case of electrons, this assumption breaks down the cross section must account for indistinguishability and large-angle scattering. Nonetheless, the qualitative energy loss behaviour of electrons is similar to that of heavy charged particles: they exhibit a minimum in the ionisation energy loss, followed by a rapid increase at higher energies. How-

ever, for electrons, the increase is dominated by radiative processes, primarily Bremsstrahlung, at much lower energies than heavier particles like muons. Bremsstrahlung, the emission of photons by charged particles in the electromagnetic fields of atoms, becomes the primary energy loss mechanism at relativistic energies.

- **Photons:** At low energies (up to approximately 100 keV), the dominant interaction of photons with matter is the photoelectric effect, in which the photon is absorbed by an atom and an electron is ejected, resulting in ionisation. At intermediate energies, Compton scattering becomes significant, wherein the photon transfers a fraction of its energy to an atomic electron and is deflected with a correspondingly longer wavelength. At photon energies above a few MeV, pair production becomes the primary interaction mechanism. In this process, the photon is converted into an electron-positron pair in the Coulomb field of a nearby nucleus, which absorbs the recoil momentum to satisfy conservation laws.
- **Hadrons:** Charged hadrons primarily lose energy through ionisation, similar to other charged particles, while radiative losses via Bremsstrahlung become significant only at extremely high energies and are negligible in most practical cases. In addition, hadrons can undergo nuclear interactions mediated by the strong force, leading to complex secondary processes. Neutral hadrons, such as neutral pions, decay rapidly; for instance, neutral pions have a mean lifetime of approximately 10^{-16} s and decay into two photons in about 98.8% of cases, initiating electromagnetic cascades [26]. Slow neutrons, with a mean lifetime of around 880 seconds, are typically captured by nuclei before decaying, releasing additional photons during the capture process. Overall, hadronic interactions with matter result in energy transfers through nuclear recoil or the production of neutrinos. The associated energy loss mechanisms, particularly those leading to non-ionising energy deposition, contribute to the so-called "invisible energy" in calorimetric measurements.

As discussed previously in section 2.2, for processes involving electrons and photons, high-energy electrons that lose energy by the radiation of a photon and high-energy photons that lose energy by pair production, the radiation length and mean free path are important quantities to take into account.

3.2 Development of particle showers

Electromagnetic showers are induced by incidences of photons, electrons or positrons, whereas hadronic showers are caused by incidences of hadrons. Due to the distinct interactions of the respective shower constituents with the material, the evolution of both shower types is fundamentally different.

The two processes for electromagnetic showers are the pair production (or photon conversion into an e^-e^+ pair) and the bremsstrahlung (emission of a photon from an e^- or e^+). The photon conversion into an e^-e^+ pair can be observed in Figure 3.3a, even though electron and photon induced showers are different in the way they start developing. Upon entering an absorber medium, electrons start to radiate immediately; on the other hand, photons travel, on average, $9/7$ radiation lengths (X_0) in the absorbing medium before their first interaction [27]. These showers grow by electromagnetic interactions with

atomic electromagnetic fields, which give increasingly more particles of lower energy, until the energy carried by individual particles cannot sustain the production processes.

The showering cascade due to an incident hadron is initiated by a nuclear interaction of the projectile particle with an atomic nucleus. Following the primary interaction, the subsequent cascade of shower particles consists of an electromagnetically and a hadronically interacting component, as it can be observed in Figure 3.3b. The characteristic length scale for the development of showers with these interactions is the nuclear interaction length λ_I . Secondary hadrons initiate further nuclear interactions that cause sub-showers at various locations in the material, rendering hadronically induced showers more sparse, wider and deeper than their electromagnetic counterpart. The hadronic shower shape is driven by shower-by-shower fluctuations, resulting in a level of complexity of showers much higher compared to electromagnetic showers.

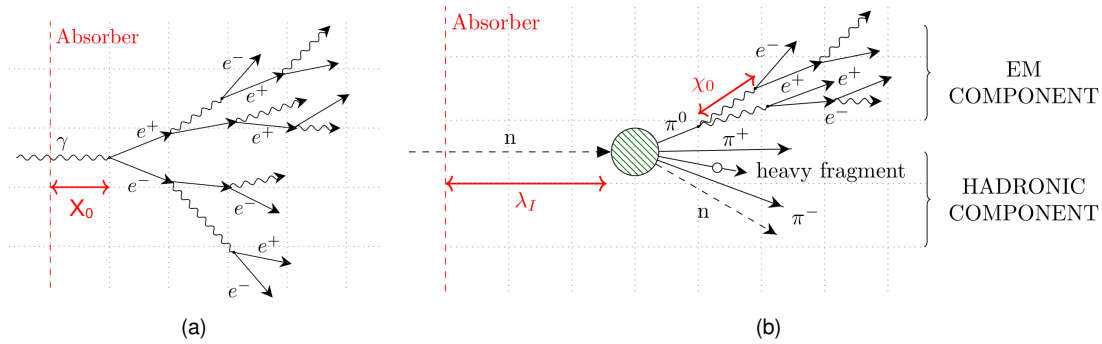


Figure 3.3: Development of an electromagnetic shower (a) and a hadronic shower starting by neutron-nucleus interaction after a distance of λ in the matter(b) [28].

3.3 Calorimetry

The study of particle showers provides the foundation for calorimetric techniques, which rely on measuring the energy deposited during these cascades. Calorimeters are responsible for absorbing the particles and measuring their energies. With the energies of the particles at the LHC, the interactions in the calorimeter cause a cascade of secondary particles commonly referred to as particle showers. These showers cease once the particles do not possess enough energy to create more particles. Ideally, calorimeters are built to contain the full shower and measure its energy to obtain the incident particle energy. Calorimeters typically consist of two subsystems: the electromagnetic and the hadronic calorimeters. The materials chosen to build these detectors are purposefully chosen to take into account the desired aspects mentioned above in section 2.2 and 3.1, the materials chosen for HGCal and other materials mentioned thoroughly are shown in Table 3.1.

Calorimeters can be broadly categorised into two types according to their readout implementation: homogeneous and sampling calorimeters.

In homogeneous calorimeters, the same material performs both the absorption of particle energy and the signal generation. The entire detector is sensitive to particles since it acts simultaneously as an ab-

Table 3.1: Properties of selected materials characterising the energy loss and evolution of particle showers. Z and A are the atomic number and mass number, respectively. ρ is the mass density, and $(dE/dx)_{\min}$ is the minimum energy loss per distance dx due to ionisation. X_0 is the radiation length, R_M the Molière radius, and λ_n the nuclear interaction length. All values are taken from [29].

Material	Z	A	ρ [g/cm ³]	$(\frac{dE}{dx})_{\min}$ [MeV/cm]	X_0 [cm]	R_M [cm]	λ_n [cm]
Carbon (C)	6	12.01	2.27	3.952	18.85	4.894	37.89
Silicon (Si)	14	28.09	2.33	3.876	9.370	4.944	46.52
Copper (Cu)	29	63.55	8.96	12.57	1.436	1.568	15.32
Tungsten (W)	74	183.84	19.3	22.10	0.3504	0.9327	9.946
Lead (Pb)	82	207.2	11.4	12.74	0.5612	1.602	17.59

sorbing material and a signal generator. The materials used are dense scintillating crystals (lead-loaded glass or liquefied noble gases) and offer high energy resolution due to the complete containment of the shower within the sensitive medium. Typically, measurements of photons produced in the interactions within the material are used to evaluate the energy of the initial particle. These photons, generated via a process known as scintillation, are then detected by photo-detectors located at the far end of the material. The CMS ECAL is a notable example. Since the full shower is contained in the material, the number of particles that create a signal is maximised and leading to overall good energy resolutions, from 3.3% to 4.5% from the latest data [30], however, a possible disadvantage can be the limited knowledge on the location of the energy deposits in the calorimeter itself.

In contrast, sampling calorimeters alternate layers of passive absorber material and active sensor material. The sensitive/active material samples the shower energy density between the passive layers, where the absorber initiates and sustains the shower. The active materials can be plastic scintillators, silicon, and liquified noble gases, while the passive materials have to be dense, depending on the shower type, lead for electromagnetic showers and iron for hadronic showers. The passive absorber layers are chosen to have a high density and high atomic number to produce more compact showers. This design allows for more detailed spatial reconstruction (including longitudinal shower profiling) and radiation hardness, but generally suffers from reduced energy resolution due to only a fraction of the shower being sampled. HGCAL belongs to this category, leveraging silicon sensors as active elements between absorber plates. A defining feature of HGCAL is its exceptional granularity: the energy deposits are read out in a fine grid transverse to the shower axis, with individual cell sizes much smaller than the Molière radius, which improves the separation of nearby showers and the reconstruction of complex events with high particle density.

With over six million readout channels and precise longitudinal segmentation, HGCAL not only measures the three-dimensional structure of showers, but also their time evolution and energy, enabling full five-dimensional (5D) reconstruction. This capability significantly enhances particle identification, pileup rejection, and overall calorimeter performance in the challenging high-radiation environment expected at the HL-LHC.

The electromagnetic and hadronic showers extend to different scales due to the difference between the nuclear interaction length and the radiation length in matter, therefore, calorimeters are often divided into electromagnetic and hadronic sections. The separation of the calorimeters allows for independent

optimisation of each system according to its specific requirements, leading to improved overall performance, particularly for the detection of electromagnetic showers. However, this design also introduces additional complexity in the reconstruction of hadronic energy, due to transition effects between the calorimeters.

The performance of a calorimeter is typically evaluated through metrics such as energy resolution, linearity of response, granularity, radiation tolerance, and power consumption. For electromagnetic calorimeters, achieving excellent resolution across a wide energy range is critical. In homogeneous designs, linearity is generally straightforward due to the direct correlation between energy deposit and signal, however, sampling calorimeters require careful calibration.

3.3.1 Electromagnetic calorimeter

The simple development of electromagnetic showers is described in section 3.1. The electromagnetic calorimeter is optimised to measure electromagnetic showers produced by electrons (e^-), positrons (e^+) and photons (γ).

The energy of an incoming particle is evaluated by the total signal generated by the showering particles. Ideally, the signal amplitude is proportional to the energy of the primary particle. The concept of calorimeter linearity refers to the level at which this proportionality is achieved in practice. The precision with which the energy can be measured is evaluated by the relative width of the signal distribution $\frac{\sigma}{E}$, which defines the energy resolution of the calorimeter. The calorimeter response for electromagnetic and hadronic showers, evaluated with the linearity and the energy resolution of the calorimeter.

The development of electromagnetic showers is a stochastic process, ruled by successive pair production and Bremsstrahlung events, and subject to inherent fluctuations. These fluctuations follow Poisson statistics. In general, the number of particles that constitute the calorimeter signal is proportional to the total measured signal. For higher energies, a more precise measurement is expected.

In homogeneous calorimeters, the energy resolution is primarily limited by fluctuations in the number of secondary particles produced in the shower, with $\sigma \propto \sqrt{N_{\text{shower}}} \propto \sqrt{E}$. In sampling calorimeters like HGCAL, this behaviour is further affected by sampling fluctuations due to the discrete structure of the detector, where only a fraction of the shower energy is measured directly.

In sampling calorimeters, the showers are sampled at a finite number of points, where the active medium is located, rather than recorded for their entire length. For these calorimeters, the number of particles that constitute the calorimeter signal is described by the number of different shower particles in the active calorimeter layers, $N_{\text{sampled}} \approx f_{\text{sampling}} \times N_{\text{shower}}$, where f_{sampling} represents the fraction of the total absorbed energy deposited in the active material. The fluctuations in the number of different shower particles in the active calorimeter layers, convoluted with fluctuations in the amount of energy deposited by individual particles, are referred to as sampling fluctuations and depend on both f_{sampling} and on the sampling frequency, which is determined by the number of different sampling elements in the region where the showers develop. The energy resolution of a calorimeter can be parameterised as

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus C \oplus \frac{N}{E} \quad (3.2)$$

where $\oplus$ represents addition in quadrature and E and σ are in GeV.

All the statistics-related contributions to the energy resolution are described by the stochastic component of the energy resolution, given by the first term in equation 3.2. S is called the stochastic term and is commonly used with the energy E given in units of GeV. Other influences on the signal distribution are non-uniformities in the detector, such as dead regions or inhomogeneities in the detector material, and the calibration uncertainty. The signal uncertainties introduced by these factors scale with the energy and result in a constant energy resolution, defined with the constant term C in the middle term of the equation 3.2. Precise calibration of the detector's response to MIPs helps establish a uniform energy scale across the calorimeter, thereby reducing cell-to-cell response variations. This directly contributes to a smaller constant term in the energy resolution, improving performance, especially at high energies. Additional consideration for a full description of the energy resolution of a calorimeter is the energy-independent noise contribution arising from the electronic noise of the readout chain. This is described with the noise term N [26].

3.3.2 Hadronic calorimeter

The hadronic section of the calorimeter follows the electromagnetic section. Hadrons will have already lost some of their energy in the EM section when reaching the hadronic section; however, they start later than EM showers and last longer. The distinction between homogeneous and sampling calorimeters is even more pronounced for hadrons. Homogeneous hadronic calorimeters are impractical due to their size and cost, so all modern hadronic calorimeters, including the CMS HCAL and HGCAL's hadronic section, use the sampling design. Iron or brass is often used as an absorber, while plastic scintillators, gaseous detectors, or silicon layers serve as active materials. In HGCAL, the transition from plastic scintillators to silicon in the hadronic endcap aims to maintain excellent spatial resolution and timing in high-radiation zones. Many interesting events in high-energy physics have hadronic final states that form particle jets. Therefore, the precise measurement of jet energy is an important task for particle detectors. Hadronic calorimeters are always sampling calorimeters and have a worse energy resolution than electromagnetic calorimeters because fractions of electromagnetic sub-showers fluctuate from shower to shower. Since the detector often responds differently to electromagnetic and hadronic shower components, these fluctuations compromise the overall energy resolution.

Furthermore, the energy resolution for hadrons is intrinsically limited by large event-to-event fluctuations of the shower evolution and the presence of "invisible" components that do not contribute to the detector signal, such as nuclear binding energy losses and neutrino production. These invisible energy losses vary significantly from one event to another and are a fundamental limitation for hadron calorimetry. As a result, the stochastic term in the energy resolution is typically larger for hadrons than for electromagnetic particles, and this remains the limiting factor not only for the measurement of single hadrons but also for the high-level performance in jet reconstruction and in determining global event

properties such as missing transverse energy.

3.3.3 Electromagnetic Calorimetry: From Homogeneous to Sampling Designs

Electromagnetic showers are stochastic processes, characterised by fluctuations in their development that scale with energy. As the number of particles in a shower increases with energy, the relative statistical fluctuations decrease, improving the precision of the energy measurement. The current electromagnetic calorimeter in CMS is a homogeneous calorimeter composed of 61200 lead tungstate (PbWO_4) crystals. This material was chosen for its very high density, fast response, and short radiation length (0.89 cm), which allows compactness and precise energy measurements. The scintillating crystals in ECAL produce scintillation light proportional to the energy deposited by the particle shower. PbWO_4 scintillates with a fast, predominantly blue emission, and is read out by avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps. However, PbWO_4 is relatively inefficient in terms of light yield and is sensitive to radiation-induced transparency loss, particularly in the high-radiation environment of the LHC endcaps. Over time, this degradation affects calibration and compromises performance under increasing luminosity conditions.

To address these limitations, HGCal represents a significant evolution. Instead of using a homogeneous design, HGCal adopts a sampling calorimeter architecture, alternating layers of passive absorber with active silicon sensors. This fundamental shift trades some raw energy resolution for unprecedented granularity and radiation hardness. Each silicon layer samples the shower with high spatial and temporal precision, allowing for a detailed five-dimensional reconstruction: three spatial coordinates, time, and energy. Granularity, in the transverse and longitudinal directions, has emerged as a critical feature for detectors. A calorimeter is considered highly granular when the lateral segmentation of its readout is smaller than the Molière radius.

The schematic in Figure 3.4 contrasts the compact, uniform energy deposition in a crystal calorimeter with a sampling calorimeter's layered, point-sampled structure. In HGCal, fine transverse granularity separates nearby showers, critical for particle identification in dense environments. Longitudinal granularity ensures that the shower profile can be reconstructed layer-by-layer, supporting precise time and depth measurements.

Silicon sensors do not scintillate but instead operate as solid-state detectors: charged particles generate electron-hole pairs as they pass through the silicon, producing a current pulse. The energy deposited is proportional to the signal amplitude. However, variations in sensor properties, such as thickness, doping, and readout electronics, need a careful, cell-by-cell calibration to ensure a uniform response across the detector. Because of this, precise calibration of the signal response becomes essential, particularly using MIPs as a standard candle.

MIP signals correspond to the most probable energy deposition by typically high-energy muons traversing the sensor without showering. They are used to intercalibrate the response across all silicon cells. Fitting the signal spectra from MIPs provides the key calibration constants for each channel. Without MIP calibration, systematic variations between cells would compromise both the energy res-

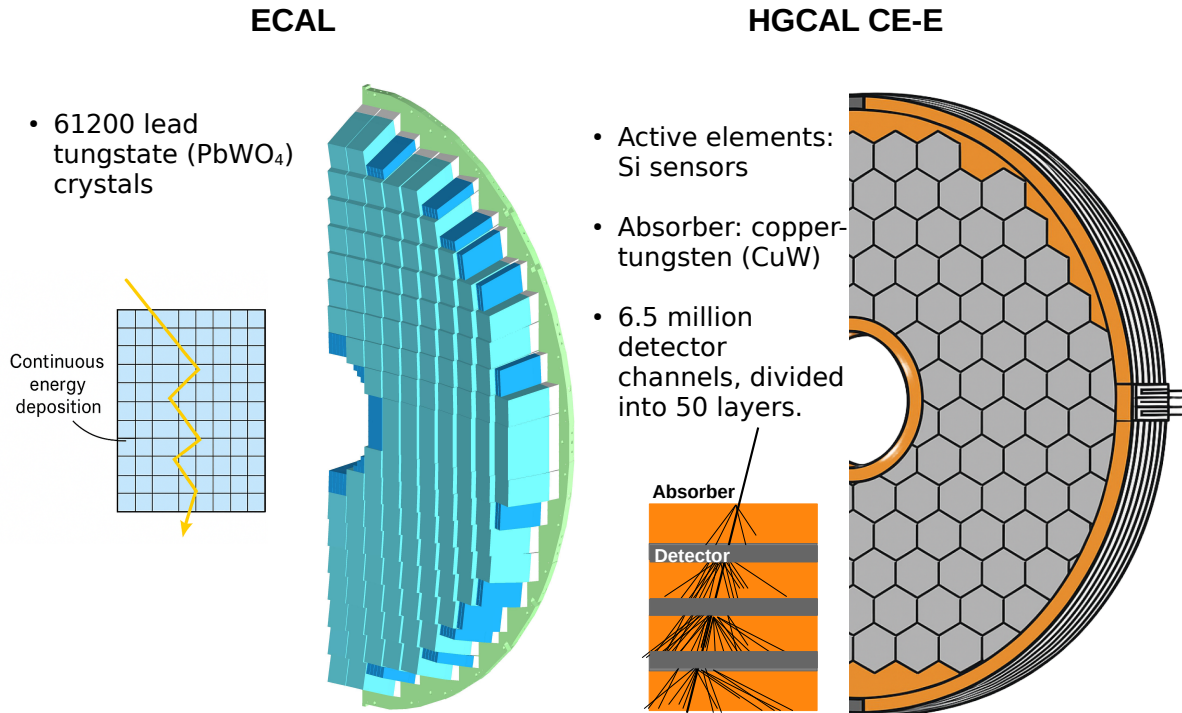


Figure 3.4: Comparison of a homogeneous calorimeter (left) based on scintillating crystals and a sampling calorimeter (right) like HGCAL, using alternating layers of absorber and silicon sensors. Granularity, longitudinal sampling, and absorber/active layer structures are highlighted.

olution and the spatial precision critical for particle flow reconstruction. This step is a cornerstone of HGCAL operation, ensuring that the detector fully exploits its high granularity and 5D capabilities even under the extreme conditions of the HL-LHC.

Another difference between ECAL and HGCAL lies in their material composition and long-term performance under radiation. ECAL offers excellent energy resolution but suffers from radiation-induced transparency loss over time. To maintain its performance, a continuous laser monitoring system is required to track and correct for changes in crystal transparency. In contrast, HGCAL, which employs radiation-hard silicon sensors and passive absorber layers, is inherently more robust in the high-radiation environment expected at the HL-LHC. This increased radiation tolerance allows for more stable operation and reduced performance degradation over time.

As explored previously, a key performance metric of a calorimeter is its energy resolution, which is closely tied to the detector's architectural design, described in Equation 3.2, with three main components: the stochastic term, the constant term, and the noise term.

In ECAL, the same dense scintillating material is used for both energy absorption and signal generation, resulting in a compact design with excellent energy containment. In contrast, since HGCAL is composed of alternating absorber and active sensor layers, this leads to additional fluctuations due to the sampling process, which are reflected in a larger stochastic term. While the stochastic term dom-

inates at low energies and reflects fluctuations in the shower development and signal generation, the constant term becomes increasingly relevant at higher energies. This term encapsulates detector non-uniformities, calibration errors, and geometric effects. In sampling calorimeters, such as the HGCAL, precise intercalibration is essential to mitigate these variations.

Table 3.2 summarises the energy resolution for both calorimeter types, and when possible, the comparison between the stochastic and constant terms, derived from test beam studies and simulation results. Homogeneous calorimeters exhibit low stochastic terms of only a few per cent and constant terms typically $\leq 1\%$, exploiting materials like PbWO_4 crystals for contained showers. On the other hand, sampling calorimeters introduce additional fluctuations due to passive layers; their stochastic term typically ranges from 5–20% $/\sqrt{E}$, and only with finely segmented absorber structures can one approach the performance of homogeneous designs [31].

Table 3.2: Energy resolution parameters for homogeneous (ECAL and sampling calorimeters (HGCAL). ECAL performance is shown for various particles and conditions in both barrel and endcap regions. HGCAL performance is based on test beam measurements of the electromagnetic section with positrons in the 20–300 GeV range.

Region	Condition / Particle Type	Resolution	
ECAL			
Test Beam (Barrel crystal)		0.5% ($S = 2.8\%$, $C = 0.3\%$)	[32]
Barrel ($ \eta < 1.5$)	Electrons from $Z \rightarrow ee$	$< 2\%$	[33]
Barrel	Photons from $H \rightarrow \gamma\gamma$ (125 GeV)	1.1% to 2.6%	[33]
Endcap ($1.5 < \eta < 2.5$)	Electrons from $Z \rightarrow ee$ (Run 2)	$\sim 4.5\%$	[30]
Endcap ($1.5 < \eta < 2.5$)	Electrons (2022–2023 calibrations)	$\sim 3.3\%$	[30]
HGCAL			
EM section	20–300 GeV positrons (test beam)	$S = \sim 22\%/\sqrt{E}$, $C = 0.4\text{--}0.6\%$	[34]

Chapter 4

CMS Endcap HGCAL

The design of the High-Granularity Calorimeter (HGCAL) for the CMS endcaps represents a significant leap forward in calorimetry, introducing a fine-grained, high-resolution approach to meet the challenges of the HL-LHC era. With the increased luminosity and resulting pileup conditions, the existing detector systems in CMS are no longer optimal for precise energy measurement and particle identification, especially in the forward regions where radiation damage and overlapping collisions are most severe since luminosity will be ten times larger than originally planned for the current detector subsystem's of the CMS experiment. HGCAL addresses these challenges by incorporating an unprecedented level of granularity in both lateral and longitudinal dimensions, alongside robust timing capabilities. These features are pivotal for isolating signals from individual particle interactions in pileup environments and for maintaining performance in high-radiation zones throughout the HL-LHC's lifetime.

A cornerstone of HGCAL's architecture is its modular design based on wedge-shaped cassettes, which are composed of tightly integrated silicon modules, referred to as hexmodules. This modularity facilitates both the mechanical assembly and the thermal management of the detector, as well as ensuring uniform coverage in the pseudorapidity range targeted by the endcap. The following section provides a detailed examination of the structural composition of the electromagnetic section, focusing on the layered cassette design, the internal arrangement of hexmodules, and the role of each subcomponent in achieving the calorimeter's performance goals. These improvements collectively ensure that the HGCAL not only matches but surpasses the capabilities of the existing CMS endcap calorimeters, providing the precision and resilience necessary for the physics program of the HL-LHC.

4.1 Detector Design

The HGCAL is divided into two primary sections: the electromagnetic section CE-E and the hadronic section CE-H, each designed with distinct technologies suited to their respective roles. The electromagnetic section is optimised for the detection and measurement of electrons and photons, where energy is deposited through electromagnetic showers. Layers of silicon sensors interleaved with dense absorber materials achieve fine sampling and high radiation tolerance. Meanwhile, the hadronic section extends

the calorimeter's depth and sampling capacity for hadrons, incorporating a mix of silicon sensors and scintillator tiles read out with Silicon Photomultipliers (SiPMs). Together, these two sections form a calorimeter that enables precise energy and position reconstruction, extending the tracker-level particle flow paradigm into the calorimeters.

The absorber layers, essential in both CE-E and CE-H sections, employ a mix of materials, including lead, stainless steel, copper, and copper-tungsten plates. In Figure 4.1, a longitudinal view of HGCal is shown, illustrating the structural layout of the electromagnetic and hadronic sections in the endcap calorimeters. On the left, the full detector geometry is shown, where the CE-E and CE-H are visibly segmented into 47 layers enclosed within the detector envelope. The figure also indicates the type of active material used in each section: silicon sensors in CE-E and the inner CE-H layers, and a combination of silicon and scintillator tiles with SiPM readout in the outer CE-H layers. On the right, two schematic depictions detail the constitution of the detector layers: the top layout shows a silicon-only layer, highlighting the different regions of silicon wafer thickness, while the bottom layout represents a mixed silicon–scintillator configuration, used in the CE-H. These schematics help visualise the varying granularity and composition of the detector. Both electromagnetic and a portion of the hadronic section will be based on silicon sensors of $0.5 - 1 \text{ cm}^2$ cell size, providing good two-shower separation because the lateral cell size is smaller than the Molière radius, which is 2.8 cm in the CE-E. The remainder of the hadronic section is based on highly-segmented scintillators with SiPM readout [35]. Electrons and photons deposit almost all of their energy in the electromagnetic region of the HGCal, depending on their pseudorapidity, whereas hadrons are expected to deposit most of their energy in the hadronic region of the HGCal. In addition, electrons, being charged particles, produce hits in the tracker layers.

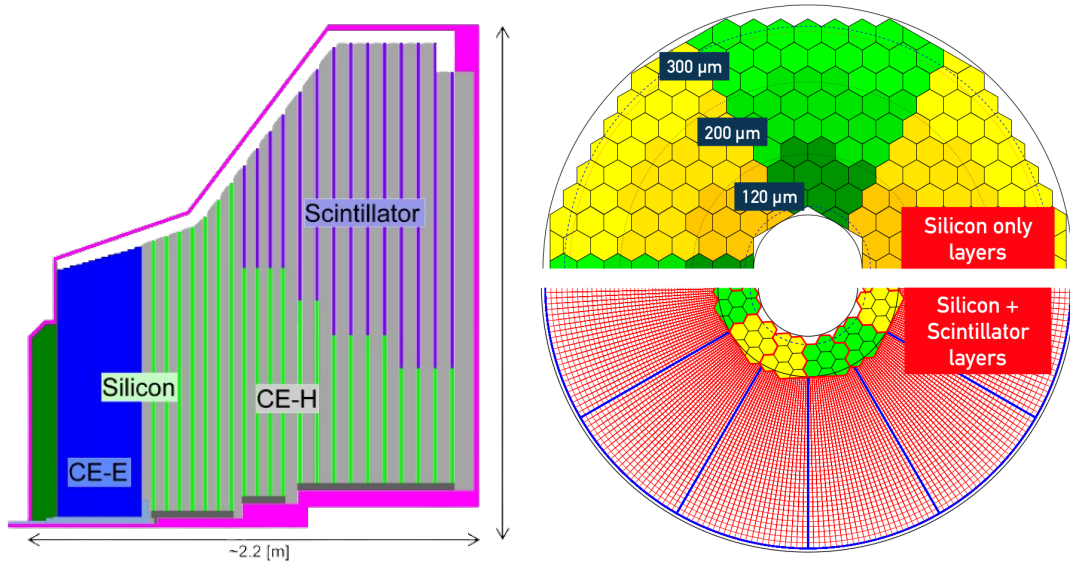


Figure 4.1: **Left:** Longitudinal view of the HGCal. The different regions are made in the figure with CE-E corresponding to the electromagnetic calorimeter and CE-H to the hadronic calorimeter. The 47 layers of active elements are represented within the pink envelope [36]. **Right:** Layout of a silicon-only (top) and mixed silicon-scintillator (bottom, not to scale) layer indicating the different silicon wafer thickness regions. [37].

The whole system will be brought down to a temperature of -30°C so that the increase in leakage

current caused by radiation damage does not result in high noise and so that the detector can still collect enough charge from particle hits to reliably record events even after irradiation. This design also introduces significant benefits for the reconstruction of physics objects, while maintaining resilience to radiation-induced damage. The high granularity of the detector enables precise particle-flow measurements, extending from the tracker into the calorimeter. This, coupled with its timing capabilities (better than 100 ps for charge deposits above 12 fC [38]), facilitates the accurate subtraction of energy from pileup events, ensuring excellent energy resolution even in high pileup scenarios.

Table 4.1 presents a comparative summary of the key design parameters and characteristics of the HGCal detector components across both endcaps. The CE-E and CE-H sections cover a combined area of 600 m² and consist of 27 000 modules. These modules contribute to a total of 6 million channels, featuring cell sizes in the range of 0.5-1 cm². The CE-H scintillator section, on the other hand, relies on scintillator tiles with SiPM readout, covering 500 m² with coarser granularity, cell sizes ranging from 4-30 cm², and about 2500 modules resulting in 400 000 readout channels. Each endcap integrates dedicated absorber materials tailored to the expected radiation environment: lead, copper and tungsten in the CE-E and CE-H to support electromagnetic and early hadronic shower development, and stainless steel in the CE-H to complete the hadronic energy containment. The calorimeter depth and weight per endcap further reflect these functional distinctions, the electromagnetic section extending to 24.94 X₀ radiation lengths and the hadronic 7.79 interaction lengths. Together with the CE-H section, it will constitute a total mass of over 200 tons per endcap.

Table 4.1: Comparative overview of the silicon and scintillator-based components of the HGCal endcap calorimeters, highlighting absorber composition, detector depth, module coverage, and segmentation for HGCal [39].

	Silicon sensors		Scintillators
	CE-E	CE-H	CE-H
<i>per endcap</i>			
Absorber	Pb, CuW, Cu		Stainless steel, Cu
Depth	24.94 X ₀ , 2 λ		7.79 λ
Weight	23t		205t
<i>both endcaps</i>			
Area	600 m ²		500 m ²
# Modules	27000		2500
Channel size	0.5-1 cm ²		4-30 cm ²
# Channels	6M		0.4M

Semiconductor, or solid-state, detectors have been an essential component of particle physics experiments for about 25 years [40]. Among the various semiconductor materials, silicon stands out as particularly well-suited for tracking and calorimetry applications, due to its excellent spatial resolution and well-understood behaviour under irradiation. The principle of particle detection in silicon differs from other materials. Doped silicon requires an externally applied bias voltage to create a depletion region, an area free of mobile charge carriers. The depth of this depletion region directly impacts the charge

collection efficiency, which increases with the applied voltage until the full depletion point is reached. In practice, a slightly higher bias voltage is used to ensure a uniform and stable electric field. When a charged particle traverses the silicon, it creates electron-hole pairs through ionisation. These charge carriers are then collected by the electric field, producing a measurable current. In the context of the HL-LHC, where the main experimental challenges are high pileup and intense radiation levels, silicon's fast timing and fine segmentation capabilities are essential for maintaining excellent detector performance.

Silicon is chosen due to its radiation tolerance, granularity and precise timing. The radiation-induced leakage current can be mitigated by operating the detector at a low temperature, and the signal loss is minimised by using thin sensors and a high operation voltage. Si's photolithographic production process allows any desired lateral granularity, and Si can reach intrinsic time resolution for high-energy showers of 12 ps [35]. The thickness of Si will vary depending on the region, in the highest radiation region $120\mu\text{m}$, in the medium radiation region $200\mu\text{m}$ and in the lowest radiation region $300\mu\text{m}$. Silicon thickness is related to radiation, and the use of lower thickness sensors in the innermost part of the detector is due to their resilience to irradiation. Thicker sensors have improved charge collection before irradiation, although they are more affected by radiation since the charge collection efficiency is reduced when irradiated. The choice of the sensor in hexagon shape is driven by cost since a reduction of the silicon material waste is obtained when produced from a circular wafer compared to rectangles or squares. Figure 4.2 presents a drawing of the final choice on the selected sensor modules for HGCal with large and small SCs. The size of the single cell is optimised for physics performance considerations, such as the lateral spread of electromagnetic showers, and given the signal-over-noise ratio.

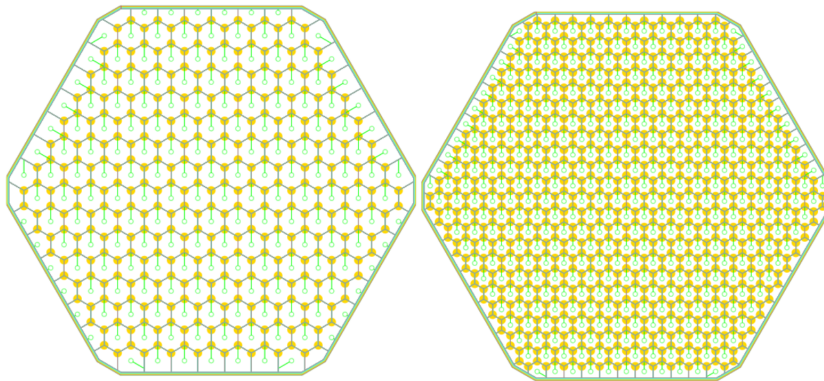


Figure 4.2: Schematic illustration of an 8" silicon module for the HGCal, which is divided into large-cells of 1.18 cm^2 for the Low Density (LD) (left) and the small-cells of 0.52 cm^2 High Density (HD) (right) [38].

The CE-E is composed of multiple layers, each subdivided into wedge-shaped structures known as cassettes. These cassettes are built by a combination of silicon modules, which are the fundamental building blocks of the electromagnetic calorimeter. The hexmodule comprises a thermally-conductive copper tungsten (CuW) baseplate, KaptonTM foil for electrical insulation, Si sensor, and Hexaboard Printed Circuit Boards (PCB). To support the modules and dissipate the heat from the electronics, they are placed on copper cooling plates. The cassette structure alongside the hexmodule is shown in Figure 4.3. The Hexaboard is a complex eight-layer readout PCB equipped with multiple HGCROC ASICs

reading the Si diode pads with low noise and large dynamic range. These chips perform on-board signal digitisation, measuring both charge and timing of the incident signals, and transfer the data to the motherboard. The silicon sensor is electrically connected to the PCB via wire bonds, which also provide the bias voltage necessary for sensor operation. The baseplate contributes to rigidity and acts as part of the absorbing material for the electromagnetic showers.

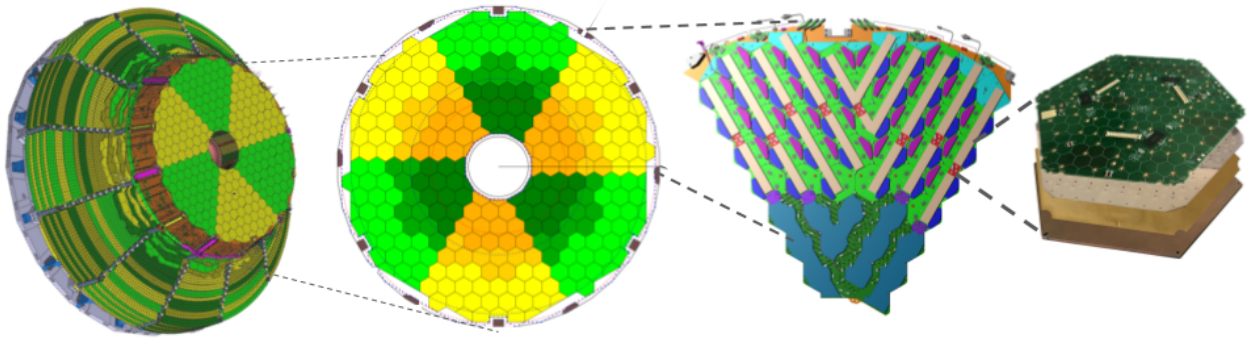
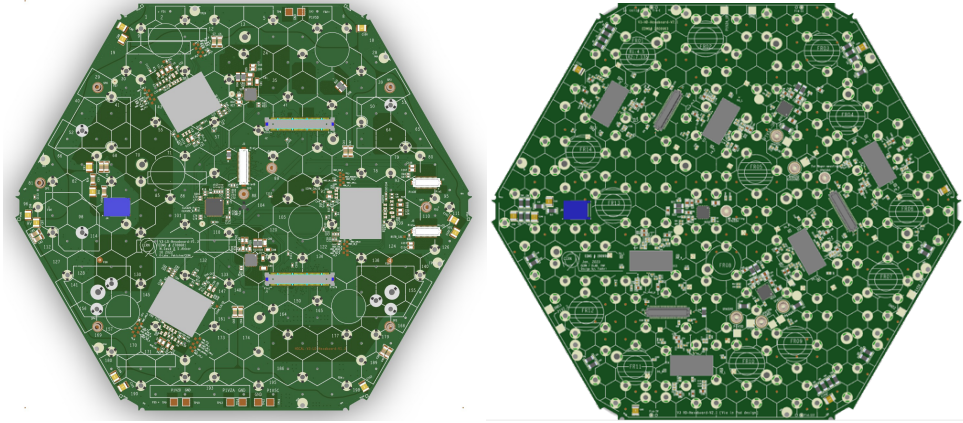


Figure 4.3: From left to right: **1.** 3D view of the cassette structure of the HGCAL. The Si hexagon wafers are represented on the very first plane, with the different colours showing the cassettes in which they are assembled. The beam travels in the inner radius of the apparatus **2.** Representation of layer 3 of HGCAL. The Colours are used to represent the different cassettes and thicknesses of the 8" Si wafers. The bars display the orientation of the DCDC power converters [41] **3.** Representation of a cassette with the readout trains employed in the trigger and data acquisition of the detector, with the location of HD and LD Hexaboards **4.** An exploded view of a hexmodule displaying (bottom to top) the baseplate (CuW or C), the KaptonTM foil, the Si sensor and the PCB [42].

4.1.1 Electronics

Depending on their radial placement within the cassettes, there are High Density (HD) and Low Density (LD) modules, which refer to the granularity of their silicon sensors and associated electronics. Both module prototypes are shown in Figure 4.4. The modules are grouped in so-called trains, including engines and wagons. On one hand, engines are complex, small, identical, and costly PCBs. Wagons, on the other hand, are simple, large, varied, and inexpensive PCBs which route the readout and control lines between the different modules and the engine. Engines also come in LD and HD varieties, associated with the respective hexaboards. Hexaboards are connected to the engines via a wagon. In the LD region, the hexaboards and wagons are connected via interposers: mezzanine boards, for powering, control, data and trigger-primitive generation, and off-detector transfer. The interposer connects to the DAQ e-link Concentrator (ECON-D). Engines convert electrical signals to optical data. The engines and modules are connected via PCB boards called wagons, and the finished assembly with all the trains will be the cassette.

In the LD Hexaboard, there are 3 HGCROCs with 78 channels for readout (72 normal channels, 4 common mode ones and 2 calibration channels), however, only 64 of these are used for readout in the LD board. For the Trigger Primitive Generator (TPG), the hexaboards are divided into supercells with a different number of trigger cells each. The silicon sensors have different cell sizes. For LD sensors with 200 μm and 300 μm active thickness will have 192 cells each with 1.18 cm^2 . For the HD section



(a) Top view of Low Density (LD) Hexaboard v1.3. (b) Top view of High Density (HD) Hexaboard v3.

Figure 4.4: Side by side comparison of High Density (HD) and Low Density (LD) hexaboard design [43].

of the detector, $120\ \mu\text{m}$ active thickness will have 432 cells, each with $0.52\ \text{cm}^2$ size. The HGCROC samples each cell from the sensor (charge and timing), and these are read out using the DAQ path. The HGCROC also computes (addition of charge) a coarser unit, i.e. collection of 2×2 or 3×3 cells to form a Trigger Cells (TC), and these propagate through the trigger path. The concentrator ASIC calculates the sum of energies for a whole module (Module Sum) and Super Trigger Cells, i.e. cluster of TCs when set up accordingly, as can be observed in Figure 4.5.

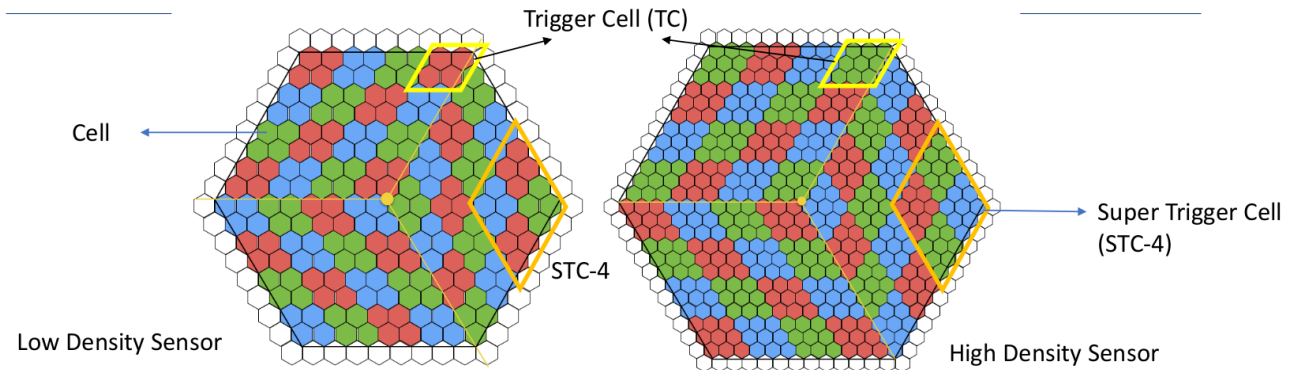


Figure 4.5: Schematic illustration of the three-fold diamond configuration of trigger cell division (TC and STC) for Low Density (LD) (left) and High Density (HD) (right) silicon modules. Formation of TCs by a clustering procedure with four large cells of $1.18\ \text{cm}^2$ (left) and nine small cells of $0.52\ \text{cm}^2$ (right) [38].

HGCAL was designed to capture detailed collision information at 40 MHz, matching the rate of proton bunch collisions. However, each event generates a large amount of data, between 2.5 to 3.5 MB. Storing all of this data is not feasible. To manage this, a quick analysis is done on a lower-resolution stream of data. This determines whether to keep the high-resolution data, which can be read out from the detector at a rate of up to 750 kHz. Only around 7500 events per second are stored in permanent storage after further processing by a software system. There are two separate paths for reading out the different resolution data: the DAQ path handles high-resolution data, and the TPG path deals with lower-resolution data and decides when to acquire the high-resolution data. Due to its volume, some processing and selection are necessary early in the chain. While part of this process occurs on the

detector, where radiation constraints are crucial, some of the electronics are shielded from radiation. The readout process is divided into on-detector Front-end (FE) and off-detector Back-end (BE) electronics. The detector comprises about 6 million channels, all read by the HGCROC ASICs. These ASICs perform three types of signal measurements: Analog to Digital Converter (ADC) for small charge deposits, Time over Threshold (TOT) for substantial deposits that saturate the pre-amplifier, and Time of Arrival (TOA), which measures the timing of the charge deposit [44].

4.2 Detector Readout

4.2.1 DAQ

The data flow in HGCAL begins with the finely segmented silicon sensors, mounted on hexaboards, which convert particle energy deposits into analog signals. These are processed by custom FE electronics, responsible not only for amplification and digitisation, but also for implementing slow and fast control functionalities, such as configuration settings, threshold updates, and synchronisation signals. The digitised information is then transmitted through the Data Acquisition (DAQ) chain, from the FE to the BE electronics. This full readout and control infrastructure lays the foundation for subsequent stages in the data pipeline.

The HGCROC high-resolution data are sent to the concentrator ASIC ECON-Ds, mounted on the wagons in HD regions. In the LD regions, the mezzanine has the ECON-D, responsible for processing the data from the HGCROCs on that LD hexaboard. ECON-D output is communicated to the DAQ BE. The FE interface with the BE is done via the Low-Power GigaBit Transceiver (lpGBT) [45], which provides high-speed bidirectional communication: the downlink (BE to FE, used for control and configuration) and the uplink (FE to BE, used for ECON-D data. The Versatile Link Plus Transceiver (VTRX+) ASIC converts the electrical signals from the Low-Power GigaBit Transceiver (lpGBT) into optical signals, completing the FE readout chain. The lpGBTs and VTRX+s are mounted on an engine to ensure communication. This full readout process is shown schematically in Figure 4.6.

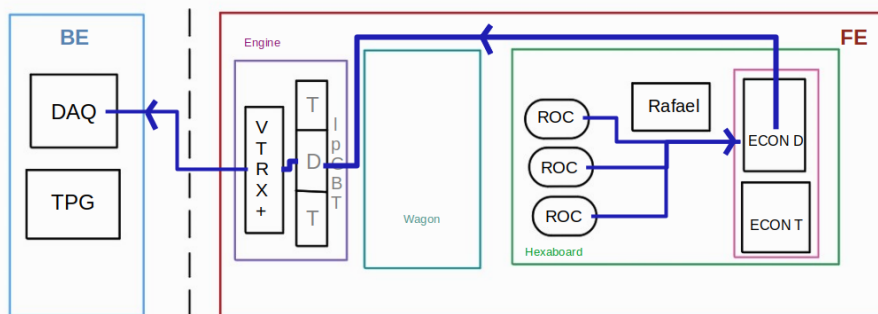


Figure 4.6: DAQ Readout (upstream) architectures for the HGCAL from the FE to the BE.

The ASICs need to be configured, and these commands can be Fast Control (FC) Commands or Slow Control (SC) Commands. The FC commands are synchronous commands concerning the bunch

crossings; thus, they control the data flow in the detector. The SC commands are asynchronous concerning the bunch crossings and they are used to configure and monitor the FE ASICs parameters [46]. The fast control and slow control commands path for the electronics system is shown in Figures 4.7 and 4.8, following the same logic as the readout but downstream.

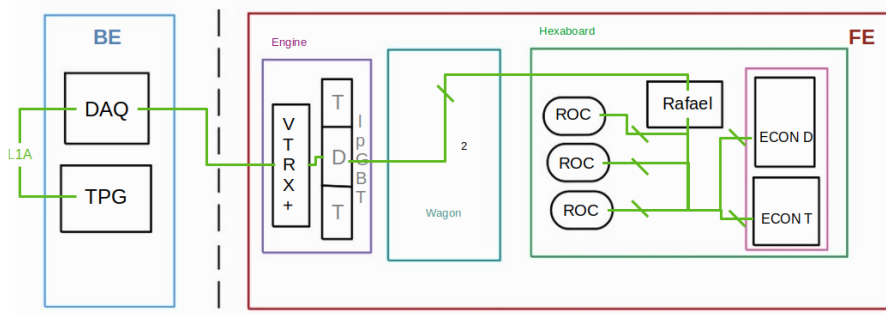


Figure 4.7: Fast control commands path for the system.

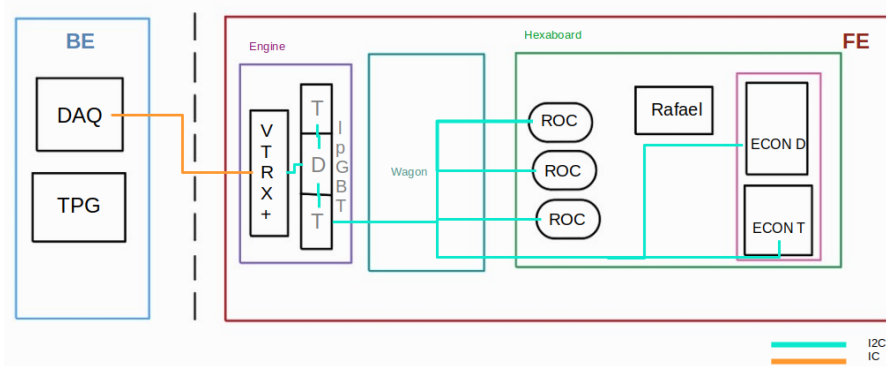


Figure 4.8: Slow control commands path for the system.

The correct configuration and operation of the DAQ system is a fundamental prerequisite for successful data collection in test beam campaigns. Accurate pedestal subtraction, common-mode noise correction, and precise timing alignment, all essential for reliable MIP spectrum extraction, depend critically on the performance and synchronisation of the front-end electronics. The proper transmission of FC and SC commands ensures that the ASICs operate within optimal parameters and that the data captured by the system is complete and well-structured.

The Trigger Primitive Generator (TPG) in HGCal is a critical step that transforms the raw data from front-end electronics into compact, low-latency summaries used for real-time trigger decisions. Once the analog signals from the silicon sensors are digitised and transmitted to the BE electronics, the TPG logic combines energy deposits over configurable regions and time windows to compute quantities such as transverse energy and cluster shapes. These primitives represent coarse-grained summaries of detector activity and are designed to capture the essential features of high-energy deposits while suppressing noise and background.

4.3 Event Reconstruction

Following the previous sections, the HGCAL readout chain begins with finely segmented silicon sensors mounted on hexagon modules connected to front-end ASICs (HGCROC), which continuously digitise charge signals at the bunch-crossing frequency of 40 MHz. These digitised signals are transmitted via radiation-hard data links to the back-end electronics system, where both detector configuration (Slow Control) and real-time data handling (Fast Control and acquisition) are managed. At this stage, one branch of the data stream is routed into the TPG, which is implemented on FPGAs and performs low-latency processing to identify localised energy deposits, summarise them into coarse-grained quantities (such as transverse energy), and forward them to the Level-1 (L1) trigger logic, where it is used to identify regions of interest and make fast accept/reject decisions within a fixed latency. . Meanwhile, the full-resolution data for events passing the L1 trigger is buffered and passed on to the high-level reconstruction chain. The design of the TPG in HGCAL is tightly coupled with the calorimeter's fine granularity and timing resolution, enabling early discrimination of electromagnetic and hadronic showers. The reconstruction chain includes digitisation (DIGI), clustering and hit association (RECO), and ultimately the production of analysis-ready objects (NANO). Throughout this chain, a consistent calibration, including the MIP response, ensures that energy measurements remain accurate and comparable across all processing stages. Thus, the MIP scale established at the earliest stage of the front-end readout propagates through to final physics objects, linking hardware calibration directly to event-level reconstruction and analysis. Following data acquisition and initial readout, the recorded signals enter a structured processing chain that transforms raw detector output into analysis-ready data products. Figure 4.9 illustrates this offline data flow, including both the configuration and output stages described previously.

The first step in offline processing is local reconstruction, which interprets the raw data from the front-end electronics. This includes energy and time measurements collected at the LHC bunch-crossing frequency. Specifically, the system records energy information from the current and previous bunch crossings (Bunch crossing BX and BX_{-1}) and time measurements. This requires several calibrations, such as pedestal and common mode subtraction, linearisation, and energy scaling (converting charge to MIPs and then to MeV). The procedures may vary depending on the sensor type (silicon or scintillator tiles) and the electronics configuration.

Next, the Unpacker reads the binary output files, discards signals from non-connected channels, and stores the data in structured formats that include Analog to Digital Converter (ADC), Time over Threshold (TOT) and Time of Arrival (TOA). This stage produces what are known as DIGI objects, which serve as the foundation for downstream processing. At this stage, Data Quality Monitoring (DQM) is also performed: summary histograms are automatically generated every million events and recorded via Run Control into the Run Registry, allowing for real-time validation of data quality during test beam operations.

The RecHit Producer builds on DIGI by applying low-level corrections and assigning quality flags to each hit, resulting in calibrated energy deposits, or RecHits. These are used to form higher-level

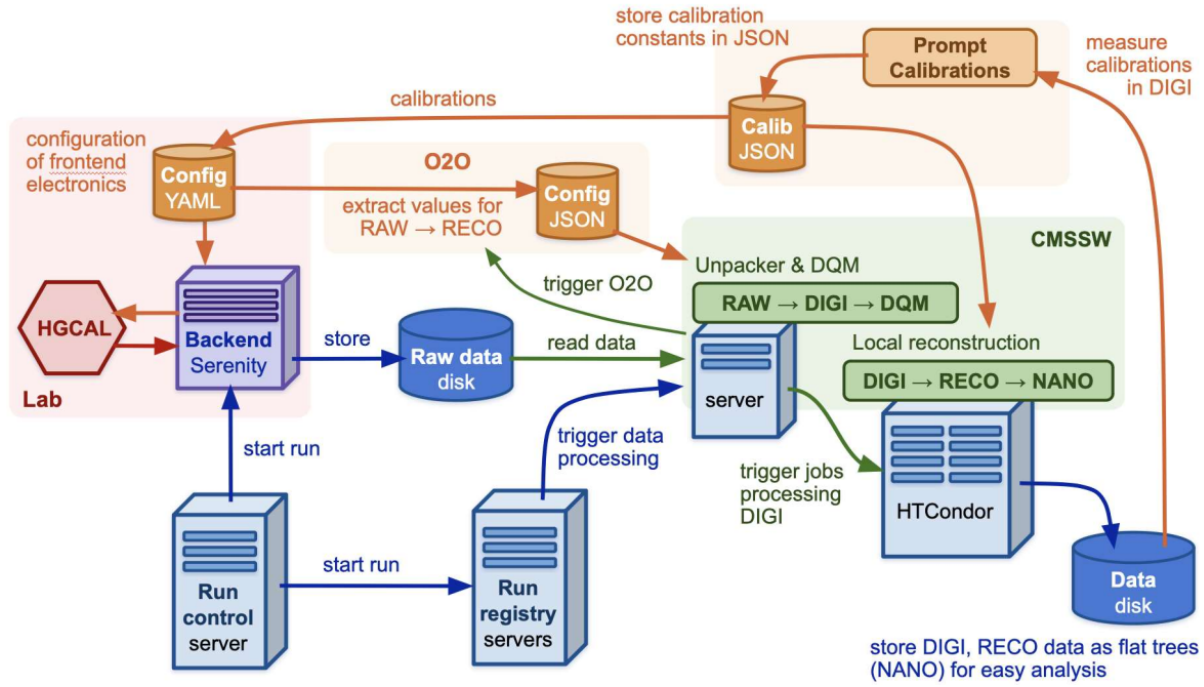


Figure 4.9: Offline data flow in HGCAL test beam operations. The diagram illustrates the complete chain of offline data processing and calibration for HGCAL beam tests, from frontend configuration to locally reconstructed output in the NANO format. Data acquisition begins at the experimental site, where frontend electronics configurations (stored in YAML files) are loaded by the *Serenity backend*, which interfaces with the HGCAL detector and stores raw output to disk. The *Run control* and *Run registry servers* manage and log metadata about the run. The O2O (online-to-offline) system extracts configuration constants from YAML files and converts them into JSON files used by reconstruction. Raw data are unpacked and passed through digitization (RAW → DIGI), quality monitoring (DQM), and local reconstruction, all managed by the CMSSW framework. The *HTCondor* batch system handles job distribution for digitization and reconstruction. Calibrations are extracted from DIGI-level data during the prompt calibration phase, stored in JSON format, and fed back into the pipeline. Final output datasets, including DIGI, RECO, and NANO formats, are stored on disk in a flat tree structure for downstream analysis. Figure by I. Neutelings.

observables. For streamlined analysis and visualisation, RDataFrame-based routines process these outputs into compact NANO files. These routines support parallel and distributed computing and can run independently of the CMSSW software framework, providing flexibility during analysis.

Finally, updated calibration constants (such as pedestal levels or MIP conversion factors) obtained from this processing loop are written back into the configuration files, ensuring consistency with the front-end electronics setup for future runs.

Chapter 5

Test Beam and Calibration

The precise calibration of detector modules is essential for translating recorded signals into meaningful physical quantities. A key step in this calibration is setting the Minimum Ionising Particle (MIP) scale, which defines the response of a given channel to a known reference signal, typically a muon traversing the sensor. To address this, test beam campaigns play a crucial role in this process. They allow the study of the detector response under controlled conditions using well-characterised particle beams. This chapter presents an overview of the test beam campaigns conducted in 2023 and 2024. It begins with a description of the beam area where these tests were carried out, followed by a detailed account of the experimental setups used in each campaign. The 2023 test beam campaign marked a significant milestone, focusing on the characterisation of a single module and, for the first time, testing the complete data acquisition chain. This provided an invaluable opportunity to evaluate the end-to-end system performance, from sensor readout to final data reconstruction. By the 2024 test beam campaign, the setup had evolved substantially. Larger prototypes were introduced, featuring multiple layers with multiple modules, including a configuration with three modules on a wagon and another combining full and semi-modules. The campaign also explored integration with tile-scintillator modules and included performance tests under a magnetic field. The chapter concludes with a comprehensive guide to the calibration procedure using MIP data, highlighting how test beam results are crucial to the refinement of the detector response and accurate data reconstruction. This work was developed in the context of participation in both test beam campaigns: in August 2023 and again in September 2024, contributing to the data acquisition and analysis efforts.

5.1 Experimental Apparatus

5.1.1 SPS and Beam line

Test beams are particle beams specifically produced for experimental studies, typically composed of long-lived particles such as electrons, muons, or pions, with well-defined energies and high particle-type purity. They are used to test and calibrate detector components under controlled conditions. These beams are made to provide a reference to study the performance of the new apparatus and test devel-

opments as they are being made. This section 5.1 will provide a basic introduction to the main beam facility used by the team testing HGCal modules at the H2 and H4 beam line. The Super Proton Synchrotron (SPS) at the CERN's accelerator complex mentioned in section 2.3 relating to the other LHC components. With a 7 km circumference, the particles that came from the Proton Synchrotron are accelerated in the SPS. Besides the proton, particles like electrons and positrons can also be provided, depending on the demand of the experiments downstream.

The extracted 400 GeV proton beam from the SPS is transported over 1 km with the help of magnets and then split into three parts, each of which is directed towards specific primary targets where secondary particles are created. There are 4 fixed targets: T2 and T4, T6, T10, while the first two produce beams for the EHN1 (Experimental Hall North 1) and are received by H2 and H4 (T2), H6 and H8 (T4) beam lines. The North Area target cavern, where the targets are located, is about 11 m underground to contain the radiation produced from the impact of the high-intensity extracted beam from the SPS. In the proton mode, the extracted intensity is typically a 10^{13} protons per cycle at 400 GeV from which only a fraction of about 40% interacts in the targets, and the rest is dumped in a controlled way in the TCC2 cavern. The surrounding earth acts as natural shielding for radiation, and the height difference to the experimental hall is sufficient to reduce the muon background to the experiments. The EHN1 is where the test beams took place, and the H2 beam line is located downstream of the target T2. After the secondary beam is generated in the target, the beam line transports the momentum-selected secondary particles to the EHN1. Users can further optimise the beam status by modifying the beam line's magnets and collimators. The target station has several different lengths of beryllium (Be) plates. The target plate is chosen based on the requested secondary particle momentum and type. The beam line features a high-energy and high-resolution beam. There are various particles with different momentum that can be selected: protons, charged pions, muons and electrons.

Electron/positron, muon and charged pion are selected to test the performance of the HGCal module layers. These long-lived particles are generated from different processes.

- Electron: A secondary e^+ or e^- beam can be produced by producing π_0 in a thin target. The decay chain is $\pi_0 \rightarrow 2\gamma \rightarrow 2e^+ + 2e^-$.
- Charged pion: Generated from the incident proton into the fixed target.
- Muon: Produced by the charged pion decay that is generated when a high-energy proton beam collides with a target.

The general beam status in SPS can be explained by a 45 ~ 60 second "cycle". Within a cycle, there are usually 2 spills with a duration of approximately 5 seconds, as shown in the Figure 5.1. This status is an important reference during the data taking. Test beam users can judge the information from the monitor to see if the beam is received in the upstream.

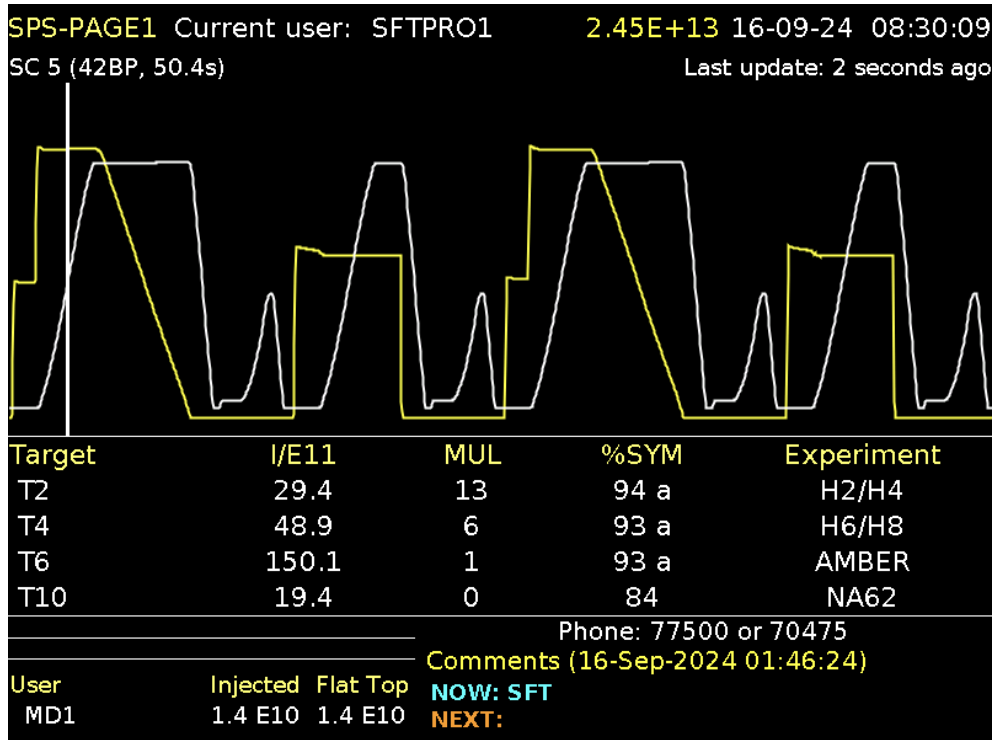


Figure 5.1: SPS beam status monitoring screen. The yellow curve is the relative particle luminosity, while the white curve is the magnetic field upstream. The perpendicular line is the updating bar, which goes from left to right and completes a cycle [47].

5.1.2 Test beams in 2023

In 2023, test beams HGCal were conducted on the H4 beamline in two phases: from August 2nd to 9th and from September 13th to 28th. The tests used various particle beams to evaluate the system's behaviour under different conditions, including electrons, muons and pions. For both configurations from August and September, an external trigger system based on multiple scintillators and PMTs with a precision of less than 1 ns was used, with absorbers of lead plates in front of the modules, as shown in Figure 5.2.

The August 2023 test beam focused on characterising a single silicon module using a standalone setup managed by a hexaboard controller for data acquisition and control, as can be observed in Figure 5.5a. Four module types were tested: High Density full module with a 120 μm silicon sensor, a Low Density full module with a 200 μm sensor, and two Low Density partial modules with 300 μm sensors, as can be observed in Figure 5.3. The different configurations are necessary to assess the different sensor thicknesses and module designs, as they are being developed, improved and manufactured.

In September, the focus shifted to testing the full electronics chain, which included components from the front-end to the back-end: ECON-D/ECON-T, IpGBT, VTRX+, and the data acquisition system, as can be observed in detail in Figure 5.5b. Two LD modules were used, one equipped with a 300 μm sensor and the other with a 200 μm sensor, seen in Figure 5.4 alongside a picture of H4 with the HGCal setup. These tests aimed to validate the functionality of the full readout chain, with both Data Acquisition (DAQ) and trigger paths successfully demonstrated. Significant progress was made in sys-

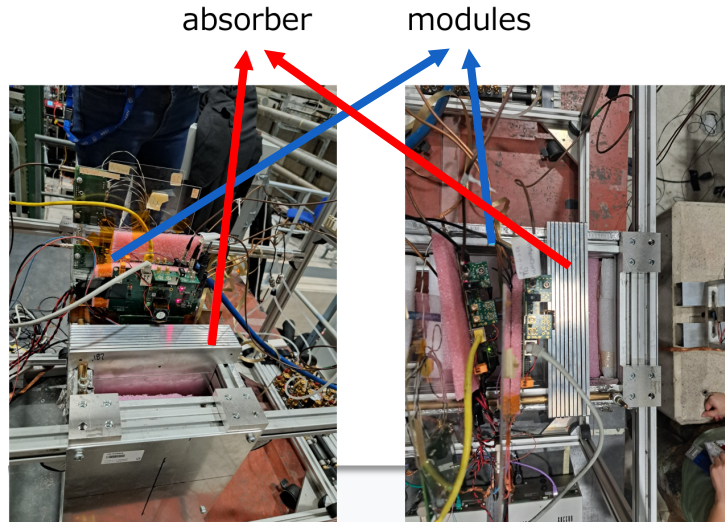


Figure 5.2: Photograph of the September 2023 test beam setup at H4, showing the integration of absorbers in front of the silicon modules. The absorbers were removed for a subset of runs to explore alternative configurations.

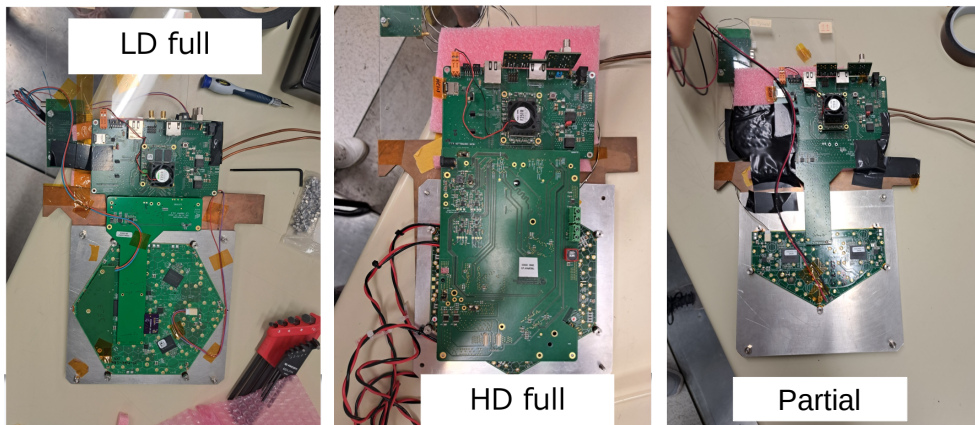


Figure 5.3: Close-up photograph of the silicon modules used in 2023 test beams for single module testing. From left to right: Low Density module 200 μm Si sensor thickness, High Density module 120 μm Si sensor thickness, Low Density partial module 300 μm Si sensor thickness.

tem improvements between August and September. The data processing and Data Quality Monitoring (DQM) pipeline was simplified to increase efficiency. A link was established between Run Control and the Run Registry server, enabling online inspection of the Run Registry and DQM servers. The complete data reconstruction workflow, including RECO-DQM-NANO, was implemented in the computing environment at CERN with processing triggered based on a curated list of valid runs. These tests provided valuable insights into the performance of HGCal silicon modules and the full electronics chain, setting the stage for future advancements in data acquisition and signal processing.

5.1.3 Test beams in 2024

The 2023 test beam campaigns marked a milestone in validating the functionality of individual silicon modules and the full readout chain for the detector. The August tests focused on single-module performance, while the September campaign successfully demonstrated the integration of front-end and back-end electronics, including the ECON-D/ECON-T, IpGBT, VTRX+, and both the data acquisition and trigger paths. These efforts laid the groundwork for reliable end-to-end data processing and reconstruction within a controlled beam environment.

Building on these foundations, the 2024 test beam campaign represented a significant step forward in system integration and complexity. Instead of isolated modules, the focus shifted to constructing realistic detector layers composed of three interconnected silicon modules, assembled on a shared mechanical structure (wagon). For the first time, these layers were tested in conjunction with scintillator modules, simulating the hybrid configuration of the final HGCal. The setup was also placed in a magnetic field of up to 3T, providing critical performance data under more realistic operating conditions. This progression from component-level validation to subsystem-scale integration marks a key transition toward full detector readiness. The tests were conducted on the SPS H2 beam line in two main phases: July/August and September, each addressing specific integration and performance goals.

July/August Phase: During this phase, tests were conducted using a 3T superconducting magnet, providing a controlled magnetic field environment for the first time. During the first week, the magnet was powered and commissioned without a beam, focusing on establishing safe and stable operation. Activities included ramp-up procedures, power and slow control monitoring, and thermal regulation via a water chiller set at 20°C. In the second week, beam delivery resumed, and ROCv3A modules with prototype ECON concentrators were tested using electron and muon beams in the 120–200 GeV range. Alongside the silicon modules, a scintillator tile module from KIT was also tested in parallel to evaluate its performance under similar conditions. Figure 5.6 shows the experimental configuration with silicon modules mounted inside the H2 magnet. Modules were attached to copper plates from the 2018 prototypes and water-cooled at 15°C for thermal stability. This phase laid the groundwork for full integration testing, validating thermal and mechanical interfaces under field conditions.

September Phase: The September phase introduced a significantly more advanced setup. For the first time, three-layer assemblies of silicon modules were tested, enabling a more realistic evaluation of the calorimeter system. Initially, two identical layers, each comprising three full LD modules with 300 μm silicon sensors, were mounted on T-shaped wagons and powered using a combination of busbar and wire connections. Later, a third layer was added to the system, a “partial train” featured one full module and one partial hexaboard, all using ROCv3B ASICs and production-grade ECON-D/ECON-T concentrators, as can be observed in the experimental scheme of Figure 5.7. A bigger and improved scintillator tile module from KIT continued to be tested concurrently. The focus remained on fine-tuning the system for optimal readout performance, particularly in the presence of the magnetic field.

All modules operated under beam and were later subjected to a magnetic field during the final five days. This enabled the evaluation of timing and signal performance under realistic electromagnetic conditions, including a study of the time-of-arrival (ToA) resolution after trimming the time-to-digital converter

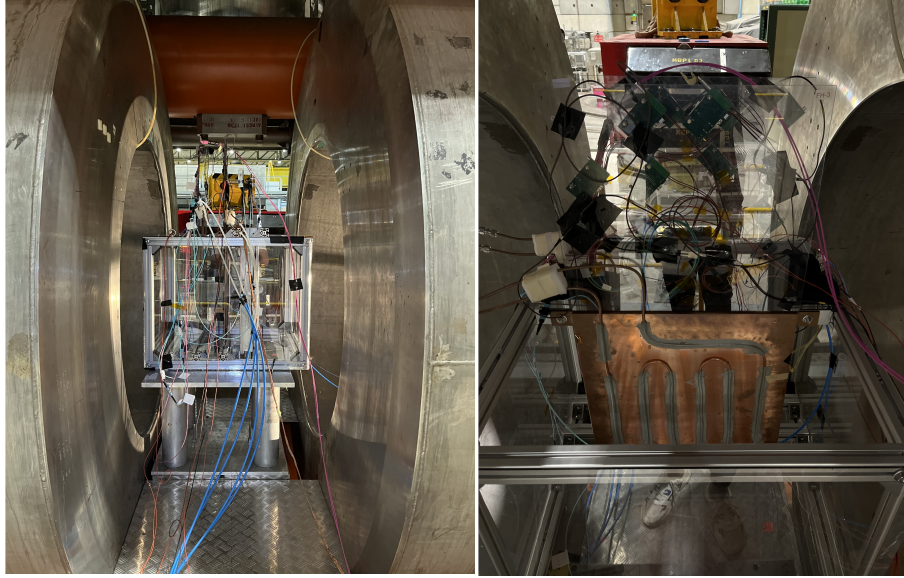


Figure 5.6: Layers and scintillator tiles mounted inside the 3T H2 superconducting magnet during the 2024 test beam.

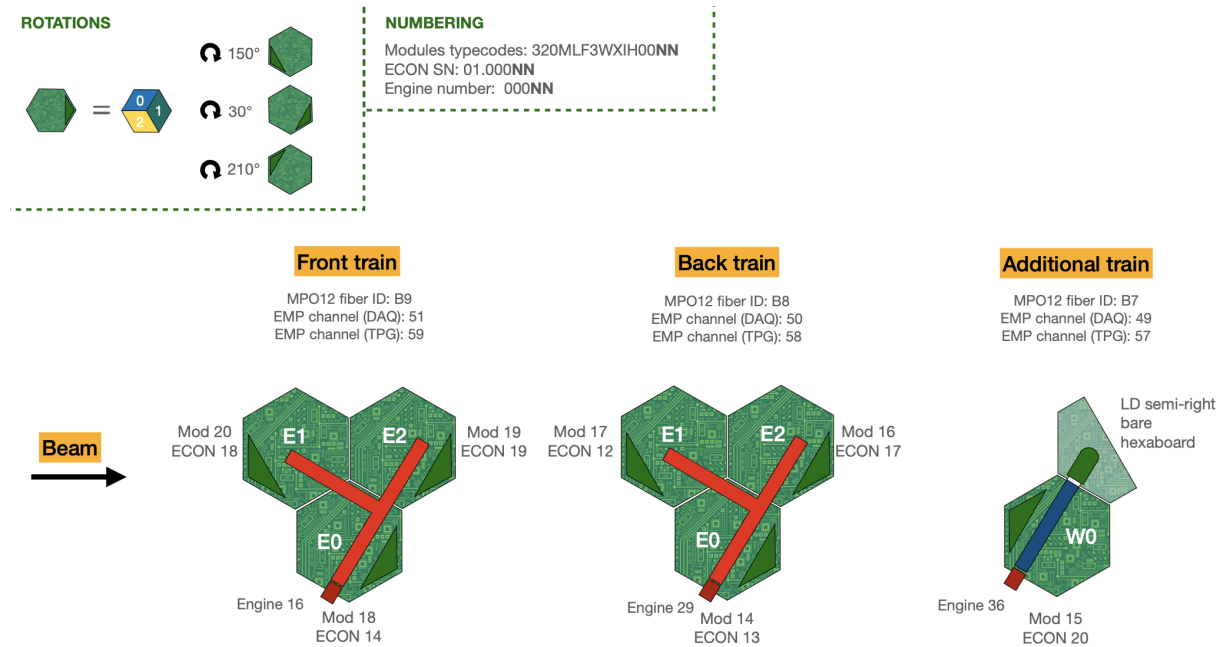


Figure 5.7: Scheme of the multi-layer setup used in September 2024 alongside the identification of each module in each layer, showing three-module trains assembled on the trains (3 modules per train except on the additional train with a partial hexaboard). The rotation of each module is also shown schematically. Figure by C. Amendola.

(TDC). Modules were powered with the latest DC-DC converters, with individual bias voltages of -280 V and leakage currents on the order of $0.5\text{ }\mu\text{A}$. The full system configuration, including cooling, powering, and readout architecture, is illustrated in Figure 5.8. It includes details such as the chiller, beamline elements (trigger scintillators, PMTs, lead brick), and power distribution. A photograph of the layer system is shown in Figure 5.9, giving a direct view of the physical arrangement.

The data acquisition and monitoring framework was significantly more automated than in 2023. On-line reconstruction and DQM were continuously performed, enabling prompt feedback on performance and guiding adjustments to thresholds and timing configurations. Across both phases, more than 3.3 billion events were collected, generating around 5 TB of data. This rich dataset then went under analysis for studies in noise, ADC response, ToT/ToA calibration, and trigger signal handling.

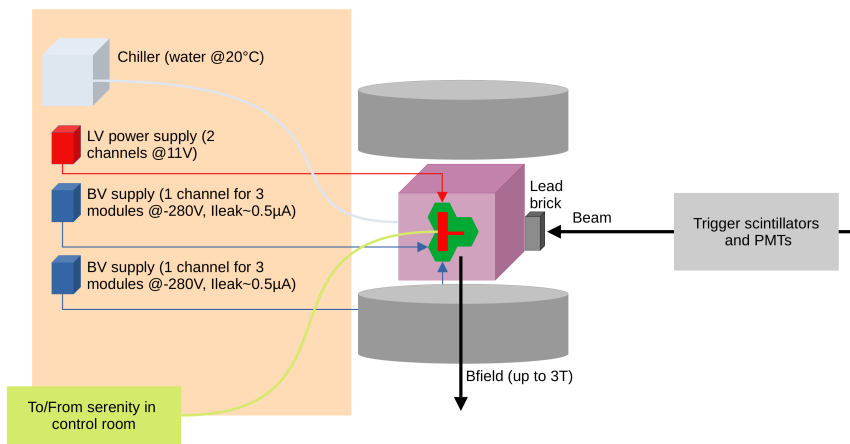


Figure 5.8: Illustrative diagram of the September 2024 test beam setup. The system includes cooling (chiller), powering (low voltage power supply - 11V and bias voltage supply -280 V and leakage currents on the order of $0.5\text{ }\mu\text{A}$) and the 3T magnet. Before the beam hits the setup, it goes by the trigger scintillators and the lead brick. Scheme by Arnaud Steen.

5.2 Calibration

The test beams conducted in 2023 and 2024 were instrumental in developing and validating the procedure for calibrating the HGCal silicon modules using minimum ionising particles. In high-precision calorimetry, accurately determining the MIP scale is essential for converting raw detector signals into meaningful energy measurements. The 2023 test beam provided a simplified environment featuring individual modules, a functional readout chain and a focused beam, which enabled the initial development and prototyping of the calibration method presented below. In contrast, the 2024 test beam offered a more realistic setup, with multi-module layers and operation within a magnetic field under defocused beams. This allowed for the practical application and refinement of the MIP calibration technique under conditions closer to those expected in the next steps (testing of cassettes). One of the outcomes of the 2023 test beam campaign was the collection of muon data. This dataset formed the foundation for designing and testing an automated calibration workflow, aimed at minimising manual intervention



Figure 5.9: Photographs of different layers of the setup. **Left:** front/back train (3 low-density modules) with electronic components (concentrator mezzanine, low voltage power supply, DCDC bias voltage power supply and busbar). **Middle:** additional train featuring a high-density module with a partial bare hexaboard and partial concentrator mezzanine. **Right:** System inside the magnet with all the components from Figure 5.8 connected.

while ensuring uniform performance across the readout channels. The combination of these two test campaigns thus marked the full progression from method development to automation of the MIP scale into a calibration routine. Such calibration cannot be reliably developed without test beams, as these provide the only opportunity to measure the detector response to known, minimum ionising particles in a clean and controlled setting. Attempting this directly would be complicated by high pileup and a lack of isolated MIP signals. Without this calibration, the energy reconstruction would be inconsistent across channels, affecting detector performance and limiting the physics reach of HGCal.

The automated MIP scale determination process consists of the following key steps:

- Identification of the beam spot using the average signal above noise (for focused beams)
- Selection of the optimal time phase in the channel(s) of interest
- Fitting of the MIP signal spectra after event selection
- Application of quality criteria to validate fit performance
- Derivation of calibration constants based on the extracted MIP peak position

This procedure provides a robust framework adaptable to a large-scale module calibration, supporting future system commissioning and performance monitoring. To apply it effectively, it is essential to understand the expected detector response to a MIP signal and how this response is digitised by the readout electronics. In particular, converting the raw ADC counts into physical energy units requires knowledge of both the sensor properties and the front-end electronics configuration. The following section provides the context for this conversion. In a $300\ \mu\text{m}$ sensor, it is expected that a MIP deposit produces an ionisation charge of approximately 3.5 fC. In practice, the particles used in the test beam

had relativistic energies and are expected to have a mean energy loss rate which is close to the minimum predicted by the Bethe equation. Thus, muons and pions in the energies provided by the CERN north area beam lines, can be considered as MIPs, modulo a second-order correction [26].

The energy can be measured up to a dynamic range of 320 fC with 10-bit ADC. Higher pre-amplifier gains can be used, reducing the dynamical ranges by factors of 2 to 160 fC and 80 fC. Thus a 3.5 fC MIP deposit is expected to correspond to approximately 11, 22 or 33 ADC counts above the pedestal, depending on the gain configured. Higher energy deposits make use of time over threshold (TOT) measurements using time-to-digital (TDC) converters. Furthermore, the necessity for calibration arises to establish a correlation between ADC counts and the corresponding energy values. Calibration is imperative for accurately interpreting the recorded data, ensuring that the ADC counts can be effectively matched with specific energy levels, thereby enhancing the precision and reliability of the electronic system's measurements. ADC to MIP energy conversion factors are needed.

Since the test beams represented the first full-scale evaluation of the electronic readout chain (in 2023) and complete detector layers (in 2024), a significant portion of the data collected does not carry direct physics meaning. Consequently, a careful selection process is required to filter the data and extract meaningful information. Although data quality monitoring was in place during the 2024 test beam, the system was still undergoing substantial tuning, including adjustments to the backend (BE) electronics and software parameters, so a critical analysis of the recorded data is essential.

The calibration of the readout channels using MIP follows a defined sequence after producing the RECO files: first, pedestal subtraction is applied to centre the Gaussian noise peak around zero; then, a study and selection based on the trigtime is performed to assess whether the time window was valid within the non-synchronous acquisition system; finally, a Langauss fit is applied to the resulting MIP spectra.

In the current sections on Calibration, the procedure developed for MIP fitting using the 2023 test beam data is described in detail, and used for the 2024 test beam data processing. The method begins with the identification of channels traversed by a MIP, followed by the application of several corrections to the raw spectra in preparation for fitting. The MIP signal is then fitted using a Landau distribution convoluted with a Gaussian ("Langauss"). Finally, the post-fit corrections applied to the most probable value (MPV) are described, along with a justification for their necessity.

5.2.1 MIP Signal

To identify which channels contain a signal, we compare the digitised charge measurements (ADC counts) from two consecutive 25 ns time intervals: the current bunch crossing ADC and the preceding one ($ADC-1$). The rationale is that if a particle crosses the detector during the current bunch crossing, the corresponding channel will show a sudden increase in signal compared to the previous one. This difference, $ADC - ADC-1$, serves as a simple yet effective discriminator for identifying active channels. In focused beam conditions, where a high signal-to-noise ratio is expected, channels with MIP signals will typically show a significant increase in ADC counts relative to noise. Even in the absence of pedestal

subtraction, this differential approach enables the construction of a list of candidate channels with potential signals. These selected channels are then used for further analysis steps, such as timing validation and MIP spectral fitting.

5.2.2 Corrections

Pedestals

Before any meaningful signal calculation can be performed, pedestal subtraction is applied to each channel to remove the baseline electronic noise inherent to the readout system. The pedestal represents the average ADC value in the absence of a signal and varies slightly across channels, ROC and modules as discussed previously. By subtracting this pedestal value from each channel, the Gaussian distributed noise is centered around zero. This step is essential for improving the accuracy of downstream analyses, such as identifying signal peaks and afterwards fitting MIP spectra, as it standardises the noise floor across all channels and enhances the contrast between signal and noise.

Trigger timing

To determine the optimal time window for signal selection, a scan of ADC counts versus trigtme (or trigphase) is performed. This scan reveals the appearance and disappearance of the MIP signal as a function of time, allowing us to identify the region where genuine signals are most likely to occur. The optimal trigtme window is then defined as the interval corresponding to the highest number of counts above the noise threshold. Although the system operates synchronously with the 25 ns clock, it may be out of phase with the arrival of signals from the SPS beam (Figure 5.1). As a result, the data acquisition window might not always align with the exact peak of the energy deposition. This can lead to distortions in the observed energy distribution. For example, when plotting the reconstructed hit energy as a function of time, the distribution may resemble a peak but not necessarily capture its maximum. This makes trigtme selection a critical step to ensure that the recorded signals reflect the true MIP response as closely as possible. To address this, an algorithm is used to define a trigtme window that captures a clean MIP signal with sufficient statistics for a stable fit, while remaining unbiased.

Pedestal and Common Mode

After the TrigTime corrections are applied, noise mitigation is still necessary. In addition to the stochastic noise of each channel, there may be coherent noise sources affecting multiple channels simultaneously, such as fluctuations in the power supply or grounding. To address this, Common Mode (CM) corrections are applied before the final MIP fit.

Each readout chip includes 4 CM channels connected to 33 pF capacitors that mimic the input capacitance of a typical sensor. These channels are not connected to active sensor pads, allowing them to serve as dedicated noise monitors. The fluctuations in the pedestal of the CM channels are expected to be correlated with the fluctuations of the full sensor.

After deriving the pedestals for the regular channels and for the CM channels, the Pearson correlation coefficient (PCC) is derived. The PCC is a statistical metric that measures the strength and direction of a linear relationship between two variables. In this particular case, the correlation between the ADC and ADC_{CM} . After this is derived the MIP Spectrum is corrected following

$$ADC - Pedestal - \rho \times (ADC_{CM} - Pedestal_{CM}) \quad , \quad (5.1)$$

where ρ is the Pearson coefficient. This correction helps suppress coherent noise and reduces the width of the MIP peak, leading to improved fit quality and more accurate extraction of calibration constants.

5.2.3 Fit

To extract the MIP calibration constant, the corrected signal spectrum is fitted using the Langauss model. This composite model reflects the physical nature of the signal: the Landau component accounts for the energy loss distribution of MIPs traversing a thin silicon sensor, while the Gaussian models the electronic noise present in the system.

The fit yields several parameters: the amplitude, mean, and standard deviation of the Gaussian noise component; the Landau width (scale parameter), the most probable value (MPV) of the Landau, which corresponds to the MIP signal, and the total area under the Landau curve, which reflects the number of MIP-like events.

To stabilise the fit and reflect the physical assumptions of the system, the standard deviation for the Gaussian and the convoluted function is forced to be the same in the fitting method.. This ensures that the same noise term contributes to both the overall signal shape and the baseline fluctuations. The resulting MPV value is used as the calibration point for converting ADC counts into physical units of deposited energy.

5.2.4 Bethe-Bloch and correction factor f

After the MPV is extracted from the fit parameters, a final correction is necessary. This requires the beam energy to be available when creating the calibration dataset. As described in Section 3.1, high-energy muons interact predominantly via ionisation, and their mean energy loss per unit length in matter follows the Bethe-Bloch equation 3.1. However, the Bethe-Bloch equation provides the average energy loss, while detector signals from thin layers, such as our 300 μm silicon sensors, exhibit Landau-like fluctuations due to the small number of interactions. As such, the MPV extracted from the signal spectrum is a better calibration anchor for MIP-like particles.

The correction factor is derived from the beam energy, muons with 150 GeV, and considering the silicon sensor's 300 μm thickness. The mean energy loss per length value for a Minimum Ionising Particle is 388 eV/ μm . Since the main interest is in relative calibration, this energy can remain in arbitrary units ("ADC units") without converting to MeV. This scaling procedure is physically grounded in the Bethe-Bloch formalism introduced earlier in section 3.1.

$$\text{Corrected MPV} = \text{MPV} \times f \times 388 \times \text{thickness}. \quad (5.2)$$

Following the Bethe-Bloch equation [26] for moderately relativistic charged heavy particles, the correction factor can be obtained by taking the quotient of the minimum energy loss and the energy loss for muons, using their momentum,

$$f = \frac{\frac{dE}{dx} \min}{\frac{dE}{dx} \text{ muon}}, \quad (5.3)$$

with the muon energy for the particles being used in the Test Beam.

This analysis's calibration is based on 150 GeV muons from the SPS beamline. Although such muons are relativistic and close to minimum ionising, their energy loss in silicon is slightly above the theoretical minimum. Using stopping power data from [26], the correction factor is $f = 0.9693$.

This factor ensures that the calibrated MPV corresponds to the response expected from a true Minimum Ionising Particle. Although the correction is relatively small (3%), it becomes important when comparing results across different beam energies, sensor types, or in simulation-to-data comparisons. The correction is especially relevant in high-granularity calorimeters, where precise energy scaling is critical for maintaining linearity and resolution.

5.2.5 Quality criteria

The simplest way to control and check the quality of the fit results is to inspect parameter distributions and impose quality criteria.

The most probable energy loss for detectors with thickness x in g cm^{-2} is [49]

$$\Delta p = \xi \left[\ln \frac{2mc^2 \beta^2 \gamma^2}{I} + \ln \frac{\xi}{I} + j - \beta^2 - \delta(\beta\gamma) \right] \quad (5.4)$$

where $j=0.020$ [26] and

$$\xi = \frac{K}{2} \left\langle \frac{Z}{A} \right\rangle z^2 \frac{x}{\beta^2}. \quad (5.5)$$

The Landau width being $\omega = 4\xi$,

$$\frac{\Delta p}{\omega} = \frac{1}{4} \left[\ln \frac{2mc^2 \beta^2 \gamma^2}{I} + \ln \frac{\xi}{I} + j - \beta^2 - \delta(\beta\gamma) \right]. \quad (5.6)$$

Assuming the density effect correction to ionization energy loss term $\delta(\beta\gamma) \approx 0$, for 150 GeV muons, $\frac{\Delta p}{\omega} = 10.08 + \ln x/4$.

This quality criterion serves as a fundamental step in selecting data that yields meaningful and physically reliable results. By applying a threshold based on statistical significance, it is ensured that the extracted parameters, like the MPV or noise width, are not artefacts of fluctuations or insufficient sampling. Additionally, this criterion informs the strategy for handling channels with limited statistics. In some cases, it may be appropriate to exclude them from quantitative fits or normalisation procedures. In others, they might be treated separately, grouped by eRx (half ROC) or HGCROC, to identify systematic

effects related to readout or geometry. Ultimately, these decisions aim to preserve the integrity of the calibration and ensure that conclusions are not driven by statistical outliers or undersampled channels.

Chapter 6

Results

This chapter presents the analysis and findings obtained from the 2023 beam test data, with a focus on the development and implementation of an automatised MIP analysis. Why the focus on the MIP data? Being a sampling calorimeter, HGCAL measures the total energy of a shower, scaling the energy deposited in the sensors. This measured energy results mostly from ionising energy deposits. Hence the for N charged particles traversing the sensitive detectors, the total ionising energy is approximately $N \times \text{MIP}$. MIPs deposit a small but quantifiable amount of energy in the Si sensors, which then can be modelled as it was explored in the previous sections, and fit parameters extracted for the inter-calibration of HGCAL sensors. The goal was thus to extract meaningful performance metrics, such as the energy scale (most probable value - MPV) of the MIP signal and the associated signal-to-noise ratio (S/N). This analysis was then applied to the 2024 beam test data, contributing both to understanding the sensor's behaviour in the presence of a beam. This contributes to the advancement of local reconstruction techniques for HGCAL.

The chapter is organised into three sections. Section 6.1 presents the results obtained in the 2023 Test Beam, where different datasets were analysed to extract MIP spectra and fit the characteristic peaks observed in the charge distributions. This section includes a description of the fitting procedure, the criteria used to select valid channels for focused beam runs, and the challenges encountered during data processing. This analysis of the initial dataset was instrumental in shaping the final version of the MIP analysis method. It has also helped identify specific improvements in the data handling and fitting strategies. Once the methodology and the initial results were validated, they were adapted in a chain of procedures used for "Local Calibration" described in Section 6.2. The MIP analysis was integrated into the sequence of procedures used to calibrate individual channels at a local level. This involved automating the extraction of MIP peaks, standardising the fit parameters, and ensuring consistency across different modules and configurations. The calibration constants derived from this process have been used to correct channel-to-channel variations in response, improving the uniformity of the reconstructed energy. The third section (Section 6.3) reports the results obtained using the 2024 beam test data. The results validate the previously developed tools and can be used as an example for future campaigns. Comparisons between the 2023 and 2024 results are also made, namely comparing response stability

and the local calibration results. Finally, in Section 6.3.4, a study is made using the central CMS software (CMSSW) of the computing performance of local reconstruction kernels, which make use of the constants derived in the test beams.

This chapter, therefore, serves as both a technical report on the data analysis conducted using the beam test campaigns and as documentation of the methods developed to support the HGCAL local reconstruction efforts.

6.1 Results from Test Beam 2023

6.1.1 Pedestal subtraction

Even when the HGCROCs are not recording any signal, they have a baseline positive signal, designated as pedestal (P). The pedestal does not convey any signal-related information and fluctuates from HGCROC to HGCROC, module to module. Its amplitude and fluctuation can be affected by operational (e.g. voltage) and environmental (e.g. temperature, humidity) constraints. The pedestal needs to be estimated before a physics run and subtracted to correct baseline differences that could impact the physical interpretation of the results. To correctly subtract the pedestals, specific runs with no beam are used, called pedestal runs. These are calibration runs performed to measure the baseline values for each channel in each hexamodule. After this correction, the pedestal peak should now be centred around 0 ADC counts.

6.1.2 Data-driven estimation of the pedestal in physics runs

Thanks to the possibility of the HGCROC to provide energy measurements for two consecutive 25 ns intervals, an in-situ subtraction technique of the pedestal can be used. The difference between the ADC counts of the current bunch crossing and the previous bunch crossing is shown in Figure 6.1 for all 222 channels of the module. These measurements are independent of each other and hence, they have the same expected value $E[\text{ADC}] = P$. Thus $E[\text{ADC} - \text{ADC}_{-1}] = E[\text{ADC}] - E[\text{ADC}_{-1}] = 0$, cancelling effectively the pedestal if no beam is hitting the sensors. This method is effective provided that the intrinsic fluctuation of the pedestal is small compared to the real energy deposits (i.e. large S/N). This is due to the fact that $\text{Var}[\text{ADC} - \text{ADC}_{-1}] = 2 \cdot \text{Var}[\text{ADC}] = 2N^2$. In the presence of a beam, the difference between the two measurements will leave the pedestal-subtracted deposited energy. When this technique is applied to focused beam data, channels that have a larger difference ($\text{ADC} - \text{ADC}_{-1}$) compared to others are identified. This is illustrated in Figure 6.1. The large fluctuations in the ADC differences suggest the presence of signals, some channels detect significant energy deposits, while others remain around 0. From the distribution, one can identify and compile a list of channels of interest, since it is where the signal from the particles is being detected.

The plots shown from this point concern two channels flagged as the channels with the highest $\text{ADC} - \text{ADC}_{-1}$ for this particular run to illustrate the methodology employed in developing the algorithm. Figure 6.2 displays the resulting MIP spectrum, plotted for two selected channels (121 and 122). One

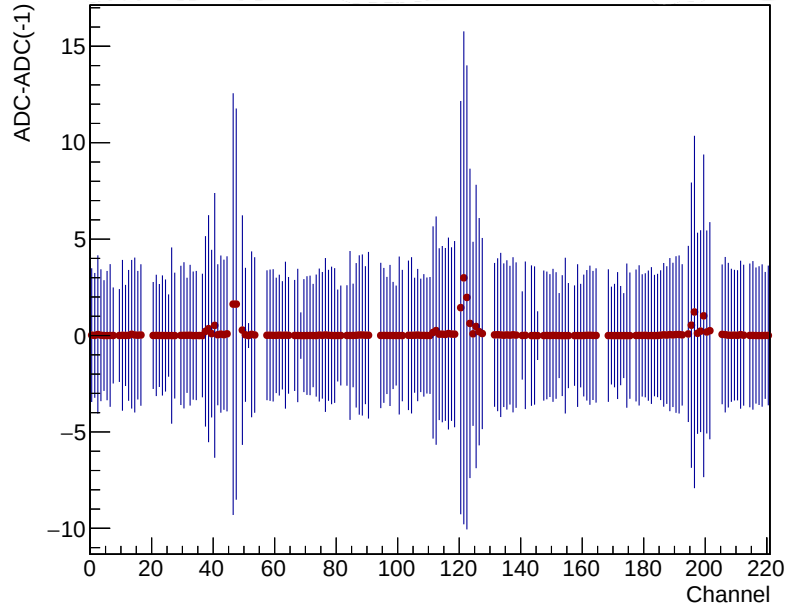


Figure 6.1: Difference between ADC counts of current bunch crossing and ADC counts of previous bunch crossing for all 222 channels of the module for the muon Run 14230133.

can observe two populations: A large population centred at zero, which can be identified as the intrinsic noise and the MIP spectrum (peaking around 10 ADC counts), which represents the distribution of energy deposits in the module. The characteristic peak in the spectrum indicates the typical response of the sensors to such particles and provides a reference point for understanding energy measurements and providing calibration.

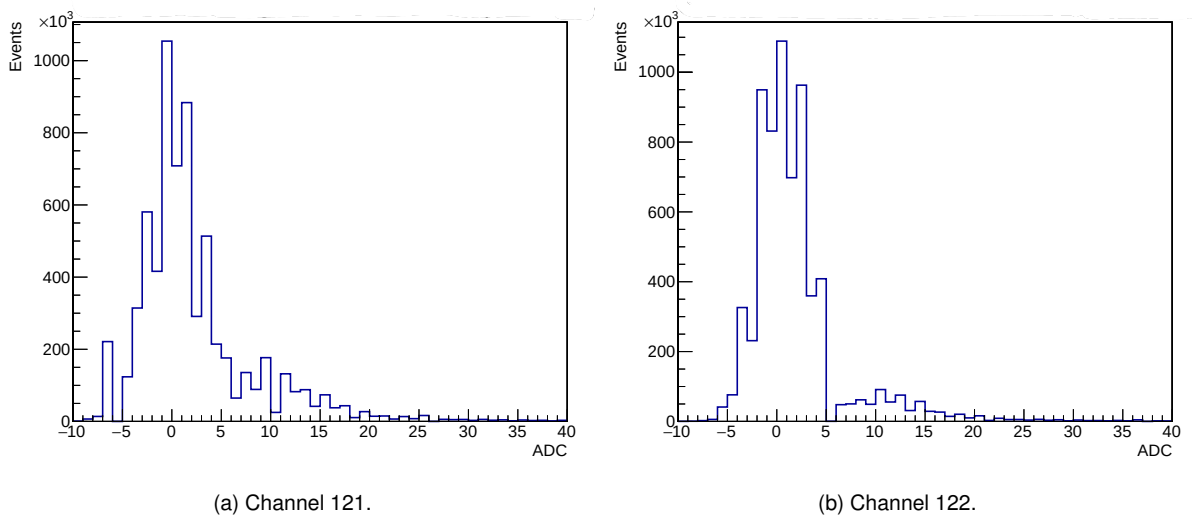


Figure 6.2: MIP Spectrum for channel 121 and 122 for the muon Run 14230133.

Next, the corrections discussed in section 5 are implemented. In the first place, the trigger phase (TrigTime) needs to be appropriately selected, given that the beam is asynchronous and may lead to out-

of-phase sampling of the signals. At the LHC, the 40 MHz clock defines the bunch crossings, meaning that the timing of data acquisition is aligned with the collisions happening at fixed intervals (every 25 ns). In test beam experiments, however, the situation differs because the SPS beam is continuous during each ≈ 5 s spill. In the DAQ setup used the TrigTime serves as a measure of the delay between an external trigger signal, indicating when an event of interest occurs, and the signal reaching the modules. It is not an absolute time value but represents the relative timing between the trigger and the module readout. A selection of an appropriate TrigTime window is required to ensure that the recorded signals correspond accurately to the wanted events.

Figure 6.3 shows how such selection can be done; the histogram shows the correlation between TrigTime and ADC counts, showcasing a colour map that represents the density of acquired events in each category. In the figure, a profile (average ADC measured for a fixed TrigTime bin) is superimposed and represented with red points. It can be seen that there is an optimal region which maximises the response to the muon beam. Hence, the TrigTime between 115 and 125 maximises the amplitude of the signal and is used in the following. This selection is critical for accurately correlating ADC counts with the corresponding trigger events, improving the MIP spectra. As illustrated in Figure 6.4, the separation between the signal and noise populations clearly improves after TrigTime selection.

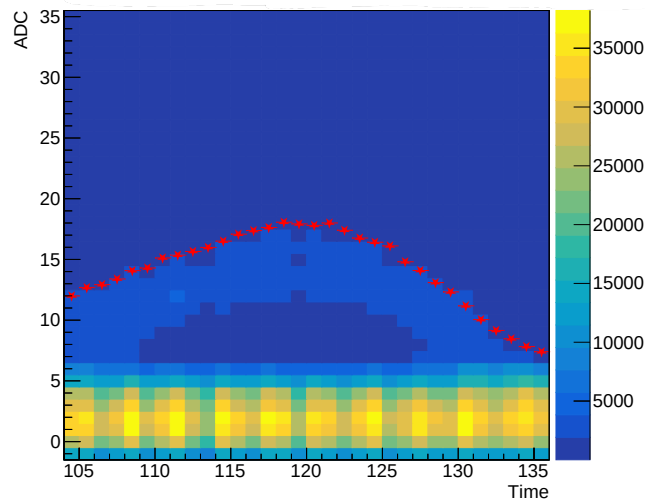


Figure 6.3: 2D histogram of trigger time versus ADC counts, with a superimposed profile showing the mean ADC counts at each trigger time. The colour map represents the density of data points, while the profile highlights the central trend across the distribution.

6.1.3 Common mode channels

Common mode noise may affect all channels of a readout system in a correlated manner, unlike statistical or random noise, which is uncorrelated between channels. CM noise has typically a low frequency; thus, when sampled each 40 MHz, it appears as a simultaneous offset in the signals of several or all channels. The origin of CM noise can be traced to several sources commonly found in electronic systems, such as ground loops, external electromagnetic interference, and capacitive or

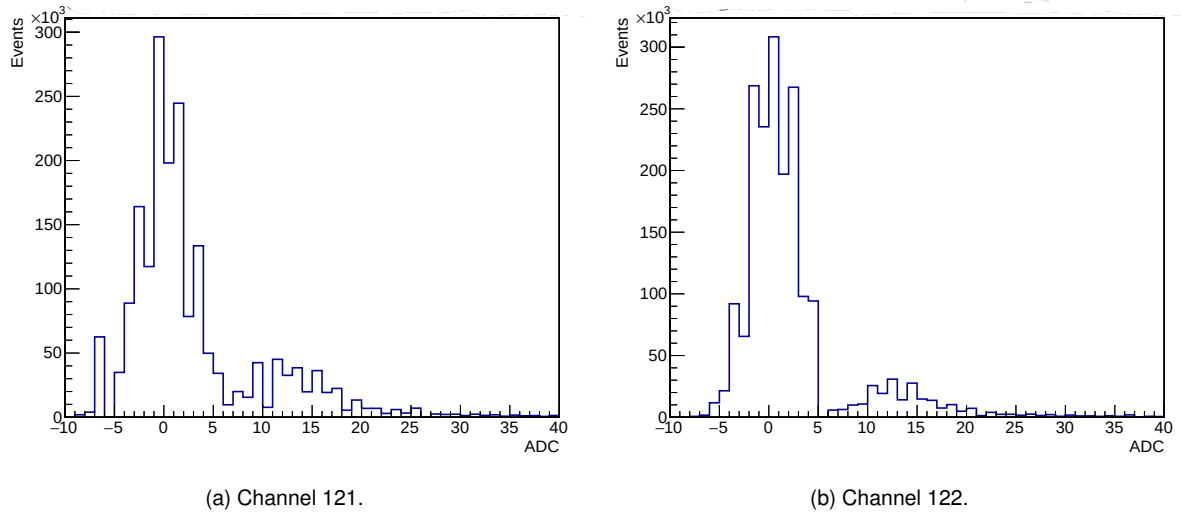


Figure 6.4: MIP Spectrum after TrigTime selection for channels 121 and 122 for the muon Run 14230133.

inductive coupling. For example, parasitic capacitance can allow external disturbances to couple into the readout circuit, introducing a coherent noise component across channels. Si sensors, being themselves capacitors after depletion, are thus sensitive to CM by nature. Since CM noise affects the accuracy of the signal measurements, especially for small signals, it must be taken into account in the corrections. The modules have dedicated common mode channels present, which can be used for this purpose. Common mode channels are associated to groups of 37 channels (half-ROC) and are not connected to a sensor but to a grounded capacitor.

The left plot in Figure 6.5 shows a correlation between a channel's ADC counts and the corresponding CM value, indicated by the linear trend of the red fit. The plot is obtained from a run without a beam. This observation confirms that fluctuations in the CM value systematically affect individual channel readings, as expected. The subtraction of the CM contribution from each channel's ADC value can be made following Equation 5.1. To quantify the correlation, the Pearson correlation coefficient ρ is computed from the original distribution (left plot), and is used as input in Equation 5.1 to get the corrected ADC values. After such correction, the CM pedestals and regular pedestals are subtracted as the right plot demonstrates. The correlated component has been successfully removed, and the distribution becomes centred around zero with no obvious structure. It is indicative that the signals are now independent of the CM noise. This correction is thus crucial for isolating genuine signals from correlated noise, leaving mostly the contributions from stochastic noise.

With the CM at hand, the effectiveness of the correction in the MIP spectrum can be verified. Figures 6.6 show the improvement of the spectra for both selected channels after the CM corrections are applied. The quality of the spectra is significantly improved, with the signal and the intrinsic aliasing mostly removed in the final spectra.

The MIP spectrum can be fitted with the following physics model:

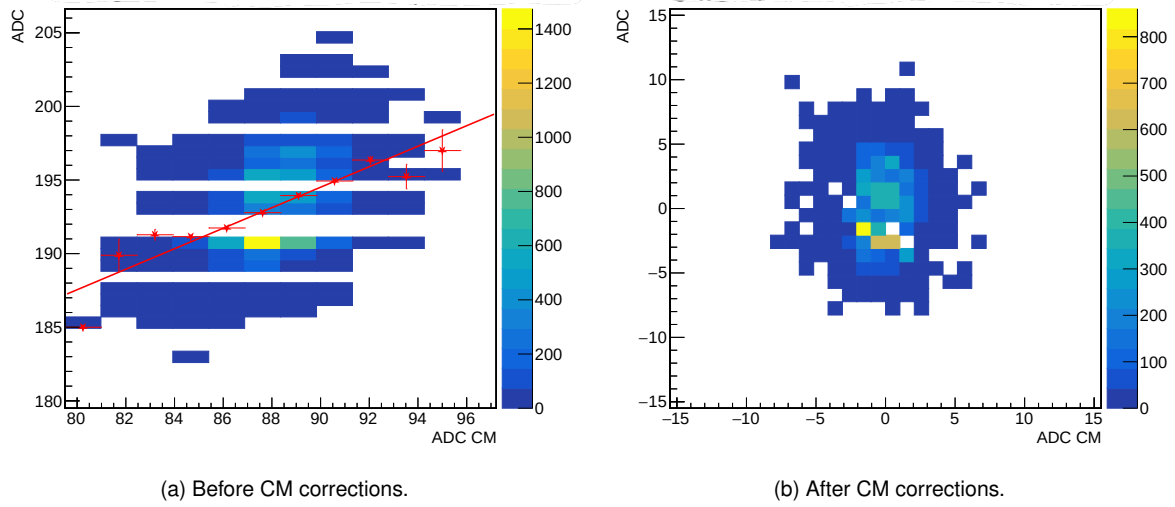


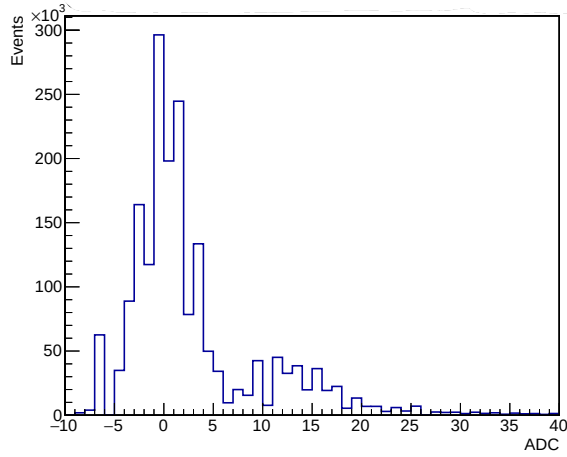
Figure 6.5: **Left:** 2D histogram showing the correlation between the ADC counts of a given channel and the common mode (CM) ADC value (the average across a group of channels). Each bin represents how often a pair of ADC and CM values occurs. The red line is a linear fit showing a strong positive correlation: when the CM value increases, so does the individual channel's ADC value. This indicates that common mode noise is present, contributing a shared offset to all channels. **Right:** The same histogram after CM corrections have been applied. Now, the distribution is centred around zero, with no visible correlation. This confirms that the CM noise has been successfully removed from the signal.

$$f(q, \mu_L, c_L, \mu_p, \sigma_p, f_b) = f_b \mathcal{G}(q, \mu_p, \sigma_p) + (1 - f_b) \cdot [\mathcal{L}(q, \mu_L + \mu_p, c_L) \otimes \mathcal{G}(0, \sigma_p)] \quad , \quad (6.1)$$

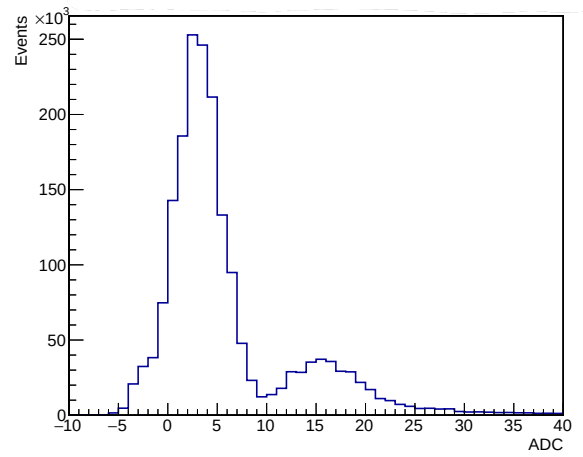
which implements a fitting function that combines the Landau distribution (signal) with a Gaussian function (noise), using six fitting parameters:

- f_B : Amplitude of the Gaussian noise component.
- μ_P : Mean of the Gaussian noise component.
- σ of the Gaussian noise component, also referred to as N in several places in this document.
- c_L Width (scale) parameter of the Landau density.
- μ_L : Most Probable Value (MPV) parameter of the Landau density.

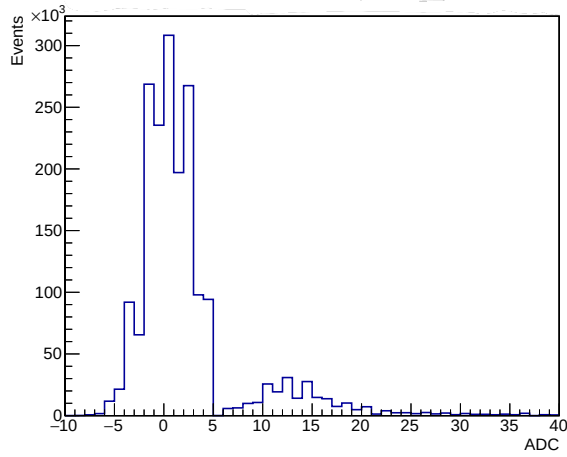
The fitting is performed using a binned maximum likelihood method, and the result is shown in Figures 6.7a 6.7b. The logarithmic plot of the same data, including also the ratio between the fitted function and experimental data, is shown in Figures 6.7c 6.7d. The ratio is correctly centred around 1, indicating the appropriateness of the physics model. In Figure 6.7d, which corresponds to channel 122, a secondary bump can be observed 37 ADC. This feature is not captured by the primary Landau–Gaussian convolution and may indicate the presence of a second MIP. However, the location of this bump does not align precisely with $2 \times MPV$ of the fitted MIP signal, suggesting that it could also be due to overlapping events, Landau fluctuations, or other detector effects. Further investigation would be required to



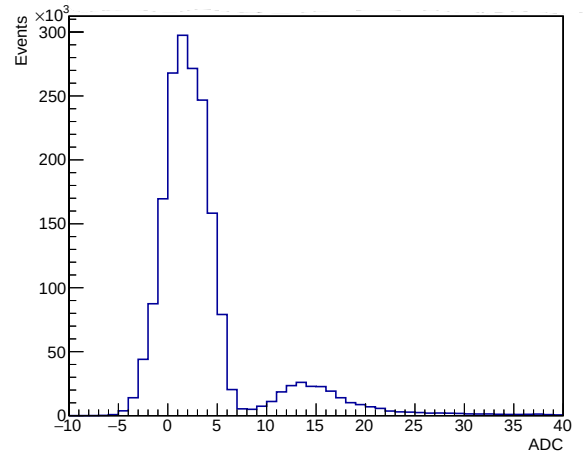
(a) Channel 121:Before CM corrections.



(b) Channel 121:After CM corrections.



(c) Channel 122:Before CM corrections.



(d) Channel 122:After CM corrections.

Figure 6.6: MIP Spectrum for channels of Run 14230133 before and after CM corrections.

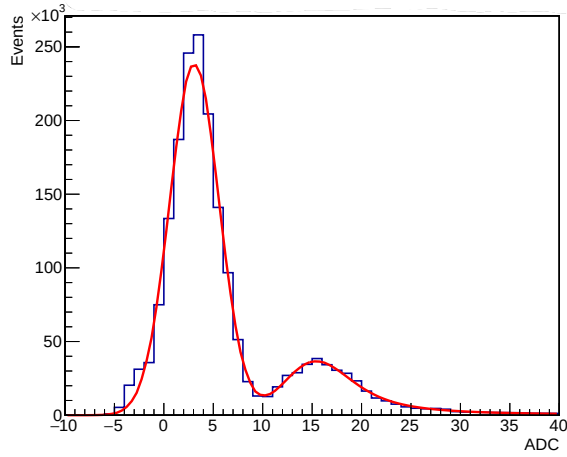
determine whether this is a statistical fluctuation or a recurrent physical signature. Not taking a second MIP into account does not influence the fit model since the fit is usually performed to 30 ADC.

The fit parameters obtained for both channels can be found in Table 6.1, with the corrected MPV in ADC units and MeV.

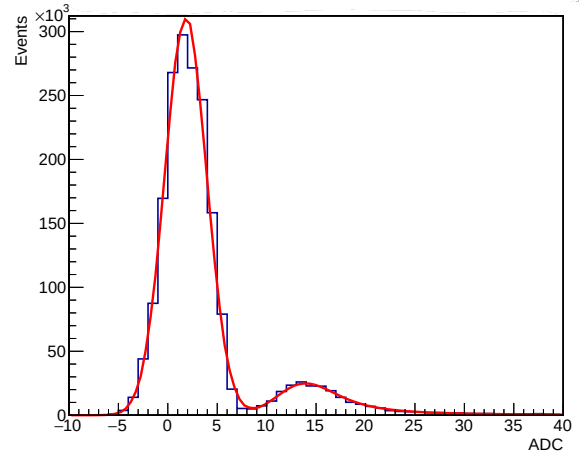
Table 6.1: Fit parameters for channels 121 and 122 of Run 14230133.

Parameter	Channel 121	Channel 122
σ	2.481	2.102
MPV	14.410	13.086
Bethe-Bloch corrected MPV (ADC)	13.968	12.684
Bethe-Bloch corrected MPV (MeV)	1.6230	1.4765

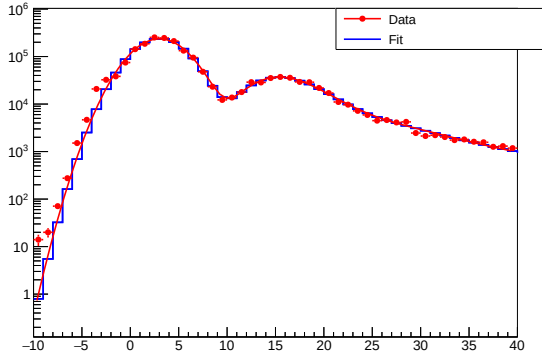
Repeating the procedure for other channels, one can extract the Signal (MPV) and Noise (N) for each one and derive the signal-to-noise ratio (S/N). The result is shown in Figure 6.8 and confirms the



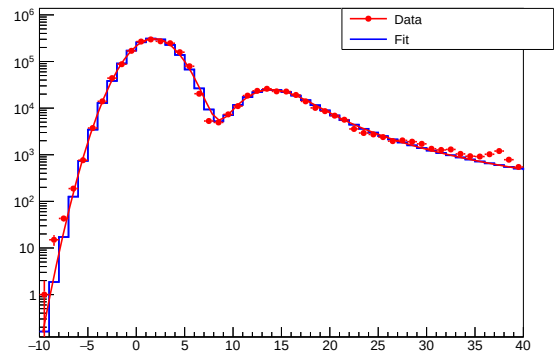
(a) Channel 121.



(b) Channel 122.



(c) Channel 121.



(d) Channel 122.

Figure 6.7: **Top:** MIP Spectrum superimposed by the Landau Gaussian convolution function fit to the data points. **Bottom:** MIP Spectrum superimposed by the Landau Gaussian convolution function fit to the data points (Log Scale). The second pad in the bottom plot shows the ratio between the function and experimental points.

excellent quality of the data and sensors. The values are within the expected range ($S/N \sim 9$). Besides the S/N , the relative uncertainty of the most probable value (MPV) can assess how precisely the MPV was measured. Lower values usually mean more reliable MPV values and MIP fits as a consequence. Figure 6.9 shows that in the 2023 dataset, the distribution of relative MPV uncertainty peaks around 1%, indicating generally precise fits. A small tail above 5% suggests that some relays and channels had unstable or noisy MIP fits.

The result also validates the consistency of section 5.2.5's checks with the result obtained from Equation 5.6. Therefore, the model is deemed adequate and ready for integration in an automatised sequence, which is described next.

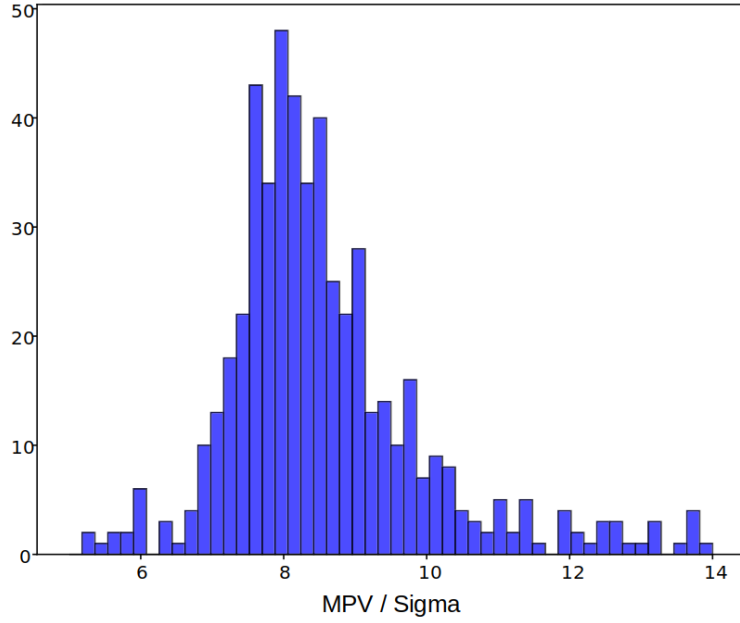


Figure 6.8: Signal over noise (S/N) for data from several muon runs of the 2023 test beam data.

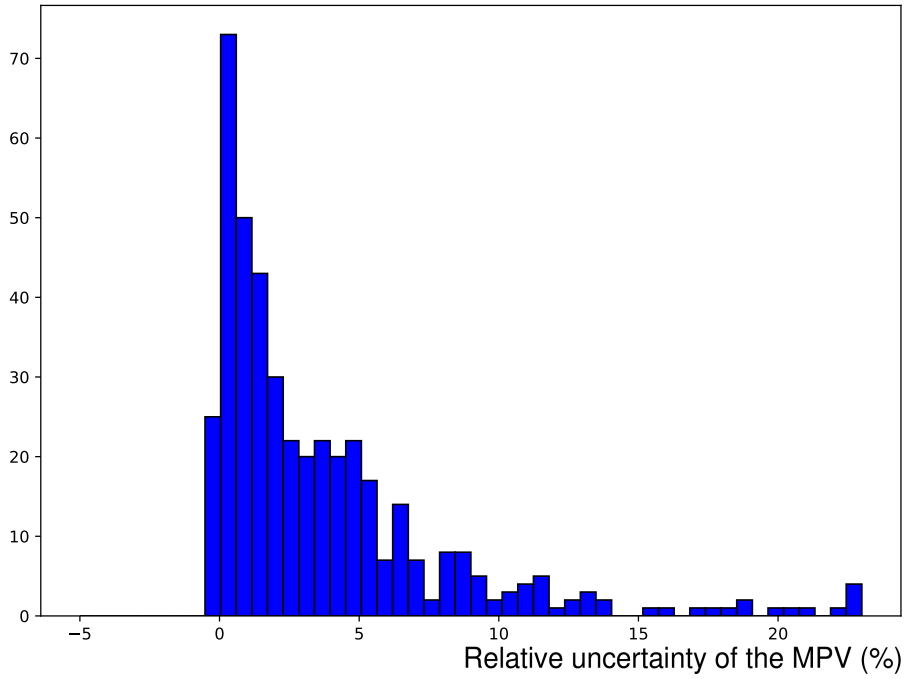


Figure 6.9: Distribution of the relative uncertainty of the most probable value (MPV) of the MIP signal for the 2023 test beam

6.2 Automatising the MIP Calibration

Between the Test Beam campaigns of 2023 and 2024, work was done in generalising the procedure and implementing it in the CMS HGCal DPG software repository for central use in future system tests. The `HGCal LocalCalibration` class aims to facilitate the calibration workflow for the HGCal system, including testing hexaboards in local labs and integrating calibration into the larger software ecosystem

for test beams and larger-scale systems (cassettes and eventually a full endcap). Its objectives are to expand the software framework, streamline the workflow, and scale it for larger systems. The calibration procedures include pedestal and de-noising, trimming (uniformising response, linearising charge and time measurements), among others.

A Python-based HGICAL base class (`HGICALCalibration`) has been created to serve as a foundation for diverse calibration tasks. The base class is designed to be flexible, allowing methods to be overridden for different analyses and online calibrations. It makes use of tools like `RDataFrame` for efficient data processing, enabling multithreading or potentially integrating with distributed computing frameworks for more intensive tasks. The focus of the work was therefore expanding the software to accommodate test beam requirements while maintaining generalisation without compromising generalisation for future HGICAL systems.

The first class implemented was `HGICALPedestal` , which inherits from the base class and provides the following functionalities:

- Profiling pedestal and CM for different modules and gains.
- Exporting the results to a JSON configuration file as input for the RAW-DIGI-RECO-NANO calibration workflow in CMSSW.
- Providing the base values needed to configure the trigger sums of the HGCROC and the zero-suppression block of the ECON-D.

The second class implemented was `HGICALMIPScaleAnalysis` , which handles MIP analysis and calibration. It also generates JSON configurations for inclusion in the final calibration following the algorithm explained in the previous section. Other calibration tasks, including TOT and TOA linearization, are similarly structured, each producing an output contributing to the final comprehensive calibration file. They have been implemented by other contributors to the HGICAL software.

Developing the `LocalCalibration` class is a team effort, with contributions from multiple calibration areas culminating in a single JSON file that integrates all necessary calibration data for the HGICAL system. The structured architecture, usage of common methods shared by all classes and the production of commonly agreed data reports ensure a consistent and efficient calibration process across different components of the HGICAL system. The 2024 September beam test data were analysed using the `LocalCalibration` class. The results are reported next.

6.3 Results from Test Beam 2024

Figure 6.10 compiles a representation of the average energy deposited in each channel of the modules used in 2024. It allows, through visual inspection, to identify anomalies with the beam, the signal, or with the hardware that could be fixed and results from a prompt data quality analysis (DQM). The modules in each layer are combined, as previously shown in Figure 5.7. It can be observed that some channels in both left and right figures have no signal (all dark blue channels) while others have high energy deposited (yellow), corresponding to the beam spot.

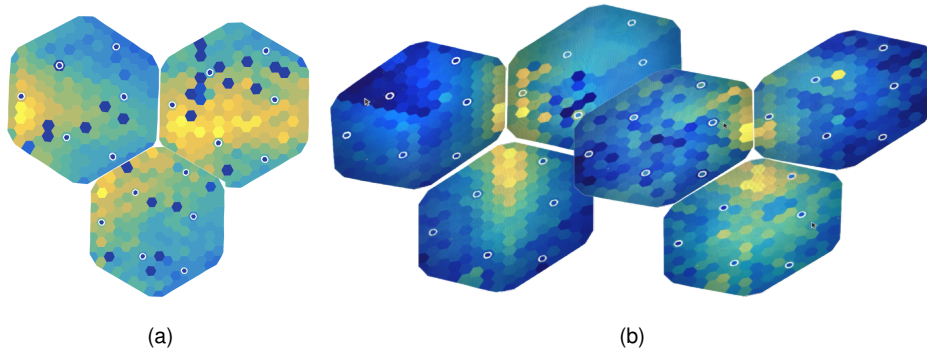


Figure 6.10: (a) DQM inspection plot of a layer of three modules, yellow channels are channels with the highest signal. Figure compiled by C.Amendola. (b) DQM inspection plot of two layers of three modules, yellow channels with the highest signal, run 1726942607 200GeV electrons 3T. Figure compiled by A.Zghiche.

Data were taken in passthrough mode (all channels readout) and with Zero Suppression (ZS) mode. In the latter case, only hits above a threshold are readout. The ZS is applied by the ECON-D and was configured with values of σ and 3σ above the pedestal. The immediate result is that with Non-Zero Suppressed (NZS) mode, one expects to find a pedestal noise peaking population and a MIP Landau distribution, while with ZS applied, the pedestal region is highly suppressed and truncated at higher noise fluctuations. This distinction has implications for how the signal and noise are extracted from the energy spectra and needs to be accounted for in downstream analyses.

A set of beam test runs was selected to perform the MIP signal extraction and evaluate the behaviour of the prototype modules under different conditions, such as different beam energies, different beam types, magnetic fields and readout modes. Table 6.2 summarises the main runs used and mentioned in this analysis, including physics runs and pedestal runs. Pedestal runs are taken without beam particles, with a random trigger used to determine baseline and noise values, and are included for completeness since they are mentioned in some plots. Pedestal runs are typically a few seconds short with no beam, while physics runs involve exposing the detector to particle beams, particularly muons. In the 2024 beam test, defocused beams were used, requiring long exposure times of the sensors, unlike in 2023. In addition to the relatively small energy deposits from MIP, the defocused beam leads to a dilution of the beam area, requiring larger integration times to accumulate enough statistics. Thus, runs with muons usually were taken over night periods. Pion beams were also used. Electron runs on the other hand, produce a significantly high number of hits with large energy and don't require large periods of acquisition time to accumulate enough statistics.

6.3.1 Pedestal and noise subtraction

Following what is discussed in 6.2, Figures 6.11 and 6.12 display the output of the online monitoring tool for pedestal mean values, pedestal widths, and the correlation slope between regular channels and common mode channels, accessible via the HGCal viewer, derived from dedicated pedestal runs. These plots result directly from the usage of the `HGCalPedestal` class. During the test beam, these

Table 6.2: Summary of beam test runs used throughout this manuscript. Each entry includes the relay number, primary particle type, beam energy (in GeV), magnetic field configuration (in Tesla), and data acquisition settings. “NZS” refers to No Zero Suppression mode, while “ZS+1N” and “ZS+3N” correspond to Zero Suppression with thresholds set to one and three times the noise level, respectively.

Run	Particle	Energy	B	
1726518879	Muon	150	0	NZS
1726347199	Muon	150	0	NZS
1726994513	Muon	250	0	ZS+1N
1727219172	Muon	250	0	ZS+3N
1727153297	Muon	250	3	NZS
1727132796	Muon	250	3	NZS
1726954855	Muon	250	0	NZS
1727204018	Pion	150	0	NZS
1727211076	No beam	-	-	-
1727217678	No beam	-	-	-
1727209838	No beam	-	-	-
1727210663	No beam	-	-	-

values have been regularly monitored and were used as input data for zero suppression thresholds applied in the ECON-D processor. Both figures display three panels, each showing a pedestal, width and CM slope for the channels in a module. As a reference, with standard gain settings, 1 ADC unit corresponds to approximately 0.19 fC, determined from dedicated calibration runs.

The pedestal mean represents the baseline ADC value in the absence of signal and is subtracted from raw data to isolate the actual charge deposition. Values vary between 140 and 190 ADC and, apart from the abnormal channel recording a higher value for the pedestal of a single channel in Module 16 of Figure 6.11, everything looks as expected. The same range of values is recorded in different Pedestal Runs as observed in Figure 6.12.

The width of the pedestal distribution (standard deviation) quantifies the electronic noise. Lower noises translate naturally to better sensitivity. Channels with significantly higher noise need to be analysed individually, like one channel in Module 18 of Figure 6.11. Most channels have pedestal width values between 1 and 2 ADC counts, with some spikes likely due to out-of-the-ordinary noisy channels. For different runs, we see a similar behaviour with 1 and 2 ADC counts. If a group of channels is noisy, this can be detected and corrected, or when doing data analysis, the groups of channels can be disregarded for physical results. As a reference, for the 300 μm sensors used in the test beam, a MIP signal of 3.5 fC=18 ADC counts, well above the measured noises.

The bottom panel provides a measure of how sensitive each channel is to common-mode noise; a higher slope indicates a stronger coupling between the channel and the global noise level, which can affect the precision of the pedestal subtraction and the shape of the extracted signal. This becomes particularly relevant in analyses involving MIP signals, where subtle differences in the low-charge region can bias the fit if CM effects are not properly accounted for.

Figure 6.13 shows the noise level measured as a function of the relative cell area, the quotient between the cell area and the full cell area. From figure 4.2, it can be observed that depending on



Figure 6.11: Monitoring plots showing the pedestal mean (top), noise (middle) and common mode slope (bottom) for the six modules from the first two layers in the 2024 beam test. Each point corresponds to an individual channel, with different colours representing different modules. These quantities are obtained from dedicated pedestal runs. The CM slope indicates the degree of correlation with common mode noise and is particularly relevant for precise signal extraction in MIP analysis.

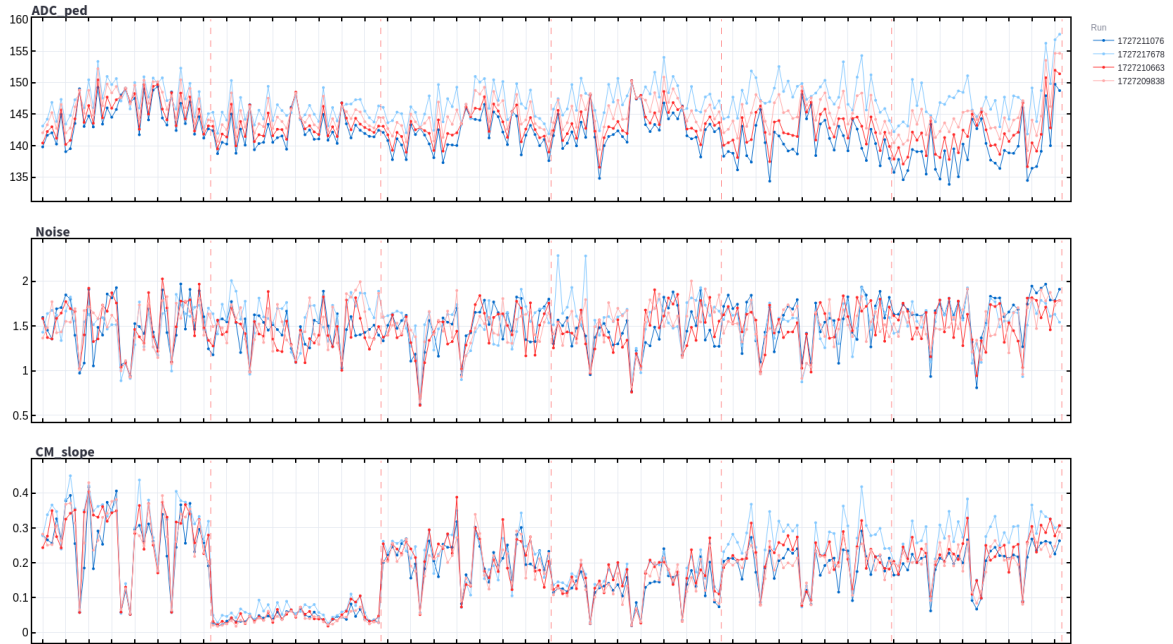


Figure 6.12: Monitoring plots showing the pedestal mean (top), noise (middle) and common mode slope (bottom) for the same module 14 in several pedestal runs of the 2024 beam test. Each point corresponds to an individual channel, with different colours representing different runs.

their location in the wafer, cells have different sizes, so they are compared relatively to the cell that is considered full. The whole cell is most abundant in the centre part of the wafer. The 2D histogram shows how noise is distributed across different cell sizes. The red line represents the unbinned polynomial fit to the data, while the dashed grey line shows the expected values from the HGCROC simulation prediction. The polynomial fit is consistent with simulations in how the noise scales with increasing cell size; the simulation line follows the same general trends as the data, and small deviations indicate areas where the real noise is slightly lower than predicted. Understanding how the noise scales with cell size is crucial to optimise detector performance. If noise becomes too large in certain areas, a calibration adjustment is necessary. Noise properties depend on detector geometry and electronics specifications, and comparing measurements with simulations ensures that detector models accurately predict performance before full deployment in the detector's assembly.

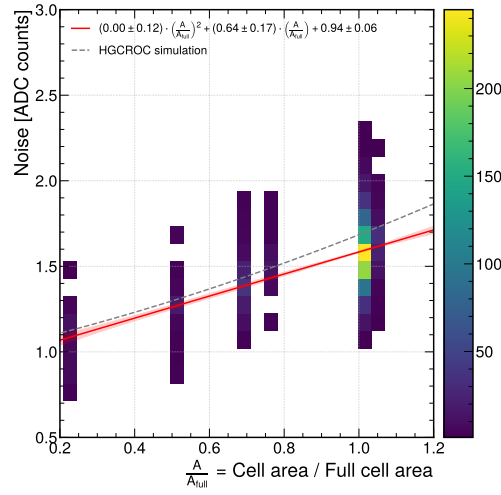


Figure 6.13: Noise distribution as a function of the cell area, normalised to the full cell area.

6.3.2 Trigger phase analysis

As explained back in Section 6.1, a quality-based selection is applied based on statistical and temporal criteria to ensure the most unbiased extraction of the MIP signal across all channels for an asynchronous beam. While in the test beam (TB) setup, the timing adjustment is relatively coarse, with a resolution of about 1 ns, the LHC provides a finer adjustment capability, allowing for more precise phase alignment. The procedure has been automatised such that for each channel, a TrigTime (or Trig-Phase) window is found that isolates a clean and stable signal, free from low-statistics fluctuations or noise-dominated regions.

An algorithm is employed to determine this window based on the energy profile as a function of TrigTime and a minimum acceptable energy threshold, which is defined as:

$$E_{\min} = \text{loc} + \text{RMS} \times k_{\text{RMS}} \quad , \quad (6.2)$$

where loc corresponds to the location of the pedestal (typically approximated by the mean or mode), RMS is the root mean square of the energy distribution (minimally affected by the presence of signal), and k_{RMS} is a configurable multiplicative factor that determines the strictness of the threshold. This ensures that the signal to be selected in phase is not too impacted by the noise.

The algorithm scans the energy profile over TrigTime values and identifies the window where the RecHit energy consistently exceeds E_{min} , constrained further by a parameter that prevents inclusion of unstable or rapidly falling signal regions. This selection is performed independently for each channel, adapting to per-channel fluctuations in timing and signal shape.

This procedure helps avoid bias introduced by manually selecting time slices or averaging over regions where the signal is absent. It also ensures that the fit to the MIP peak is based on a sufficiently populated and representative signal region.

Illustrating the impact of this selection:

- Figure 6.14a shows the rechit energy as a function of TrigTime for a representative module;
- Figure 6.14b shows the same plot, but filtered to include only energy values exceeding E_{min} , highlighting the stability and sharpness of the selected window;
- Figure 6.14c shows the trigttime window selection applied to this specific Relay and Module;
- Comparing 6.14a and 6.15a, one can observe that there is a Relay with well-aligned timing, where the selected TrigTime window coincides with the signal peak, and another with poor alignment, where the signal is not fully captured within the window.

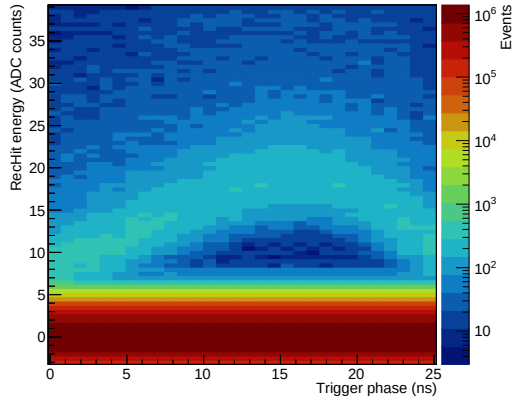
These plots demonstrate the effectiveness of the algorithm in identifying the valid calibration region and motivate its use in downstream analyses, including signal-to-noise estimation and fit stability.

Figures 6.15a and 6.15b further illustrate the challenge in cases of poor beam alignment or timing mismatches. Although a window is selected, the maximum energy might lie outside of it, introducing uncertainty in the extracted signal. In these cases, the runs have been discarded and the electronics reconfigured to have a better in-phase signal.

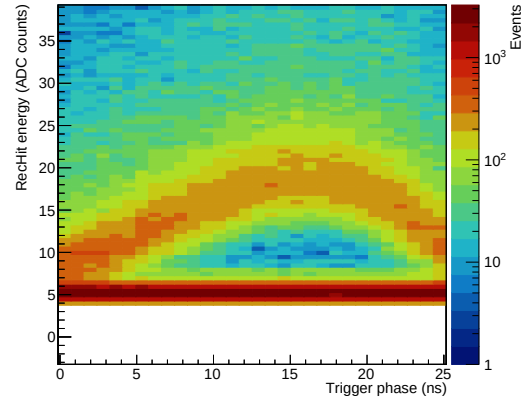
It is essential to verify a posteriori that the automatised window selection truly contains the global maximum of the energy distribution. This has been achieved by inspecting the plots automatically produced by the calibration routines. The algorithm may occasionally select a region that does not correspond to the true MIP peak, e.g. if the electronics are not adequately configured, leading to biased estimates of the signal amplitude. The plots discussed above helped identify cases where the algorithm performs robustly and those where additional timing corrections or manual intervention are necessary to recover the correct signal shape. This further reinforces the importance of complementing automated procedures with dedicated diagnostic tools and expert validation when dealing with complex data.

6.3.3 MIP Scale

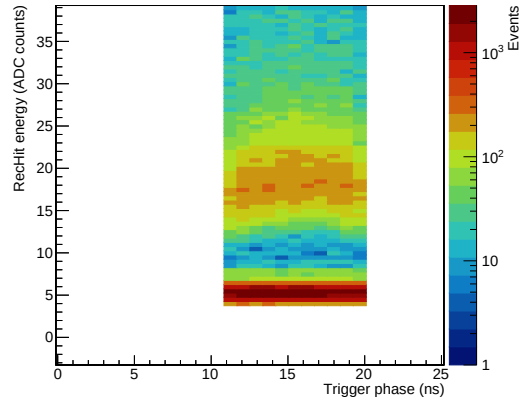
Unlike focused beams, which concentrate particles along a narrow trajectory, resulting in energy deposits mainly in a small group of channels (recall Figure 6.1), defocused beams produce a broad



(a) ReHit energy vs TrigTime.

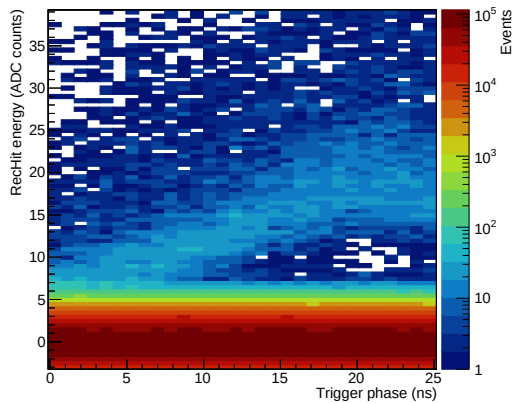


(b) Same data, filtered for $E > E_{\min}$.

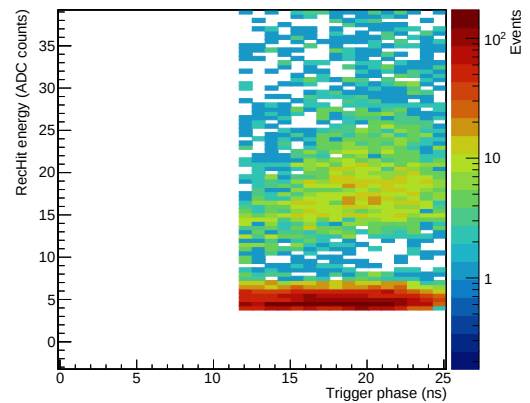


(c) Selected TrigTime window showing the energy profile used for the MPV fit.

Figure 6.14: ReHit energy in ADC counts as a function of the trigger phase (ns) for Relay 1726518879 and Module 16. The filtering and projection steps isolate a clean and unbiased MIP signal.



(a) ReHit energy vs TrigTime - Relay 1726347199.



(b) Selected TrigTime window showing the energy profile used for the fit.

Figure 6.15: ReHit energy in ADC counts as a function of the trigger phase (ns) for Relay 1726347199 Module 14.

spread of energy deposits across the layers of the modules, which can be observed in 6.10. The defocused beam 2024 configuration eliminates the need to identify the channel with the highest signal; however, it adds a few challenges as well. The wide spatial distribution of the energy deposits allowed for an inclusive analysis approach, where all channels were analysed and considered. This strategy was particularly beneficial for system-level studies of the modules' response and for characterising average behaviour across groups of channels, allowing the identification of abnormal behaviour in channels, such as a lack of signal. This spread also came with a cost, adding to the nature of muons, fewer hits per channel, especially given the time constraints of the beam test campaign, which resulted in low statistics that limited the ability to perform channel-by-channel fits.

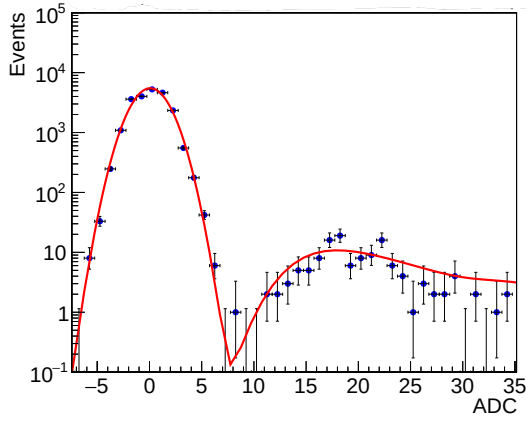
Strategy for Low-Statistics MIP Fitting

Given the limited number of muons per channel due to the defocused beam and short data-taking periods, individual channel spectra often suffered from low statistics. To extract meaningful MIP parameters, a hierarchical aggregation approach was implemented: individual channels were first fit with the physics model of Equation 6.1. The compatibility of the spectra with MIP signals has been inspected. After this first analysis, it was decided to aggregate the data at the eRx (37 channels) and ROC (74 channels) levels, and the fits were repeated. This approach increased the statistical precision while preserving resolution within modules and enabled both localised and global performance evaluations.

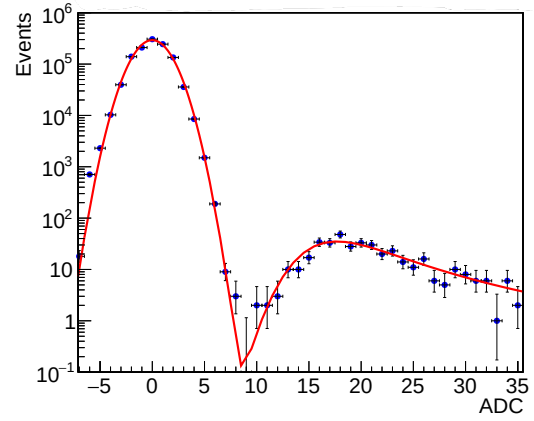
This workflow demanded careful judgment in selecting the aggregation level for fitting. While channel-level fits provided spatial resolution, many had insufficient statistics to extract the MPV reliably. Aggregation at the eRx or ROC level introduced statistical stability but required careful validation to avoid bias from malfunctioning or inactive channels.

As described in Section 6.1, the parameters of interest are extracted from a fit to the data. The goal of the fit is to identify and quantify the most probable energy deposition by a minimum ionising particle (MIP), modelled by the Gaussian and Landau convolution. Figure 6.16 (a) shows the fit performed in a single channel from Module 16 of Run 1726347199, where it can be observed that the expected shape was obtained, a prominent pedestal followed by a MIP signal peak. It is evident from this figure the limited statistics in each channel. Additional examples are shown in App. A. To overcome the low statistics in each channel, signals from multiple channels were added, first at the eRx level, in Figure 6.16 (b). This aggregation improves statistics, which is noticeable in a smoother shape of the distribution. Finally, Figure 6.16 (c) shows the final signal, added at a ROC level, grouping 74 channels at a time, with improved statistics. The specific run and module alongside the MPV and σ are concentrated in Table A.1.

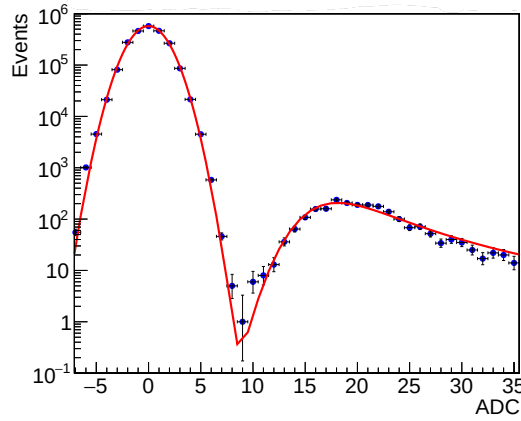
Figure 6.17 shows the distribution of the signal-to-noise ratio, defined as the MPV of the MIP signal divided by the noise σ , for a selected set of channels where the statistics were judged to be sufficient to do the measurement in 2024. To ensure the reliability of this result, a careful selection of the data was performed. Only runs deemed physically meaningful were included, those for which the TrigTime window reliably captured the MIP signal, and where the channels exhibited valid energy deposits. This avoids bias from channels with poor statistics or ill-defined timing. Furthermore, quality checks on the



(a) Fit to the ADC distribution of a single channel using a Langauss convolution - Relay 1726347199 Module 16 channel 189



(b) Fit to the ADC distribution of a single channel using a Langauss convolution - Relay 1726347199 Module 16 erx0.



(c) ADC spectrum aggregated at the HGCROC level (74 channels) - Relay 1726347199 Module 16 roc0.

Figure 6.16: MIP Spectrum and Langauss fit for Relay 1726347199 with channel, erx and roc comparison.

stability and convergence of the spectrum fits were applied to ensure that both the MPV and σ values were extracted from well-behaved distributions. The S/N distribution of Figure 6.17 peaks around values between 10 and 11, with a relatively narrow and symmetric shape, indicating that the signal is typically 10- 11 times larger than the noise across most channels. This reflects a consistent and high-quality signal reconstruction performance within the silicon modules. The small spread observed can arise from natural variations in cell geometry, sensor area, or electronic characteristics, which influence the noise level. The signal-to-noise ratio plot thus provides a robust summary of the system's response during the test beam and serves as a benchmark for identifying any outlier channels or modules that may require further inspection or calibration tuning.

Figure 6.18 shows the 2024 distribution of the relative uncertainty, which peaks around 1%, with fewer high-uncertainty outliers above 5% compared to the 2023 test beam (Figure 6.9). This suggests improvements in detector stability or fitting procedure. The reduced tail indicates more consistent fit quality across all channels, and it is also a reflection of the criteria imposed on the fits.

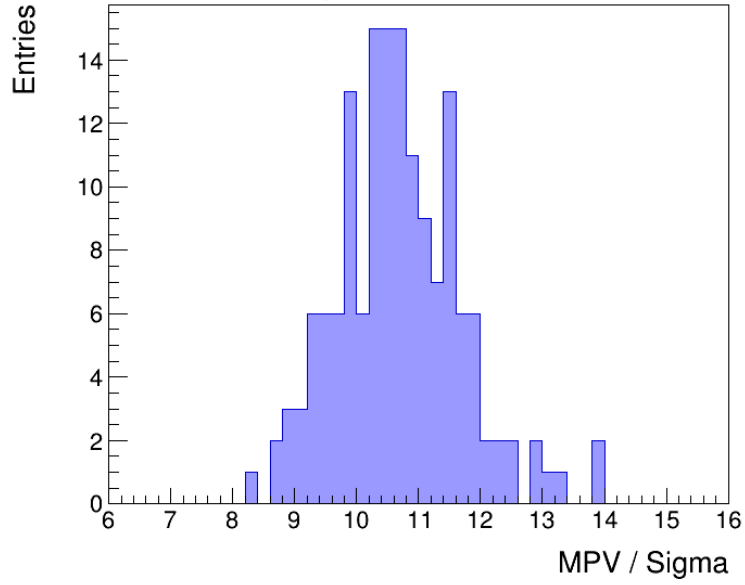


Figure 6.17: Distribution of the signal-to-noise ratio, defined as the most probable value (MPV) of the MIP signal divided by the corresponding noise (σ).

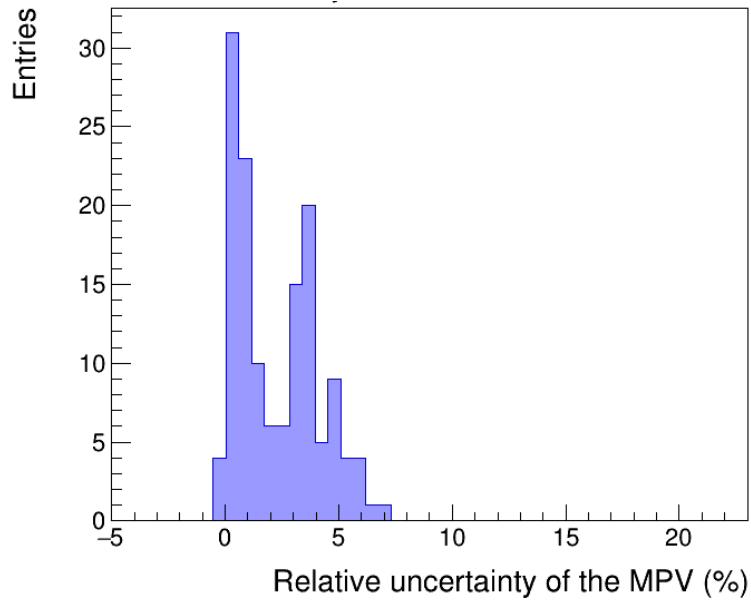


Figure 6.18: Distribution of the relative uncertainty of the most probable value (MPV) of the MIP signal for the 2024 test beam

A comparison of the two test beams reveals a clear evolution in performance and setup. Between 2023 and 2024, the ROC design changed from version v3a to v3b, and the 2024 configuration introduced DC-DC powering and an engine interconnecting three modules, electronics that had not previously been operated with the modules. These upgrades influenced the noise characteristics and had a direct impact on the signal-to-noise ratio. The noise level (σ) improved from a ~ 2.48 in 2023 ~ 2.0 in 2024. Correspondingly, the signal-to-noise ratio increased from roughly 8.2 to 10.2, an improvement consistent with a 25% increase in signal relative to noise. This enhancement reflects not only the refined hardware configuration but also a more stable and characterised response in 2024.

6.3.4 Injecting the constants needed for local reconstruction in CMSSW

The results obtained above provide the necessary constants for the local reconstruction of the energy of each hit. CMS uses a centralised software framework (CMSSW) which can handle simulation, reconstruction and analysis routines for the online selection at the High Level Trigger (HLT) and the offline data processing and analysis.

For the local reconstruction of the energy of hits in HGCal, the implementation is made using the ALPAKA library. This library abstracts the protocol specifics required to run on either a CPU or GPU, and thus makes transparent the process of offloading routines in a heterogeneous computing system. This flexibility is crucial for the efficient analysis of large data such as that expected to be produced by HGCal. In CMSSW, the routines that utilise the pedestal, common-mode correlation, and MIP scale have been implemented, and a test of their performance has been conducted.

Using a single-threaded CPU, the production of calibrated RecHits achieves a throughput of 615 events per second, with an average event time of 1.38 milliseconds, as shown in Table 6.3.

Table 6.3: Performance metrics from a single-threaded CPU reconstruction job.

Metric	Value
CPU Model	AMD EPYC-Genoa
Number of Threads	1
Avg. Event Time [s]	0.001382
Event Throughput [events/s]	614.93
Max Event Time [s]	16.215
Min Event Time [s]	0.000111
Total Init Time [s]	62.72
Total Job Time [s]	177.68
Total Loop Time [s]	114.96
Total EventSetup Time [s]	6.13
Total Non-Module Time [s]	7.57
Total Job CPU Time [s]	85.64

If one considers, in addition, the fact that the raw data needs to be decoded (unpacked) and aligned before calibrating, the performance degrades if performed in series. The total throughput of the complete operation (DIGI to RECO) is shown in Figure 6.19 for different CPU and GPU benchmarks. One can observe a significant increase in throughput going from four to eight CPU threads and an additional jump when offloading the local reconstruction to the GPU. A total of 950 events/s has been achieved using a total of 900 modules read in NZS mode. Although the data size is comparable to that expected for the final HGCal system, the number of modules and their variety are not. The result is, however, promising in the sense that 1 ms/event seems achievable at the HLT level.

Further details are given in Table 6.4.

When scaling to a massive, high-throughput detector system like HGCal, understanding and optimising performance across CPU and GPU platforms is critical. Each event in HGCal may involve processing tens of thousands of hits across millions of channels, making per-event compute efficiency a

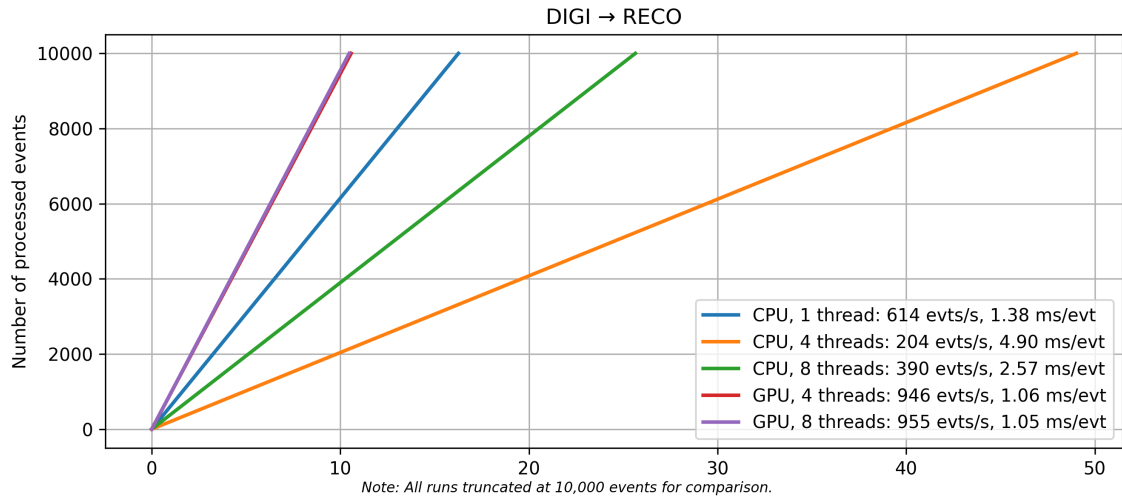


Figure 6.19: Comparison of event throughput (events per second) across CPU and GPU configurations using CMSSW reconstruction workflows. Each coloured line corresponds setup: CPU and GPU jobs running with 4 and 8 threads, respectively (previous results from Izaak Neutelings) and CPU result with 1 thread from data of the current test beam.

Table 6.4: Comparison of reconstruction performance across CPU and GPU configurations.

Configuration	Event Throughput [events/s]	Avg Event Time [ms]
CPU, 1 thread, 8 modules	615	1.38
CPU, 4 threads, 900 modules	204	4.90
CPU, 8 threads, 900 modules	390	2.57
GPU, 4 threads, 900 modules	946	1.06
GPU, 8 threads, 900 modules	955	1.05

bottleneck.

Performance assessments are important to guide the deployment of computing resources during HGCAL commissioning and data-taking periods such as Test Beams. These studies help determine the optimal balance between CPU and GPU resources, thread allocation, and module scheduling to ensure stable and efficient reconstruction performance. One example is the MIP calibration, which involves fitting response distributions across thousands of channels to precisely determine energy scale constants. This process can require processing millions of hits and fitting thousands of histograms, an operation that becomes slow without scalable compute resources. Efficient event throughput, parallelism, and GPU acceleration can substantially reduce the turnaround time of such calibration cycles, allowing for faster feedback during Test Beams.

Chapter 7

Conclusions

7.1 Conclusions and outlook

The development of the High-Granularity Calorimeter (HGCAL) is a central part of the CMS upgrade program in preparation for the High-Luminosity LHC (HL-LHC), which is expected to begin operation in 2029. With an anticipated increase in instantaneous luminosity and extreme pile-up conditions, current endcap calorimeters (ECAL and HCAL) will not be able to perform within the required resolution and radiation tolerance. HGCAL, as a highly segmented sampling calorimeter, is designed to address these limitations through precision shower reconstruction and time-tagging capabilities.

This work focused on the validation and performance of a local calibration framework using beam test data collected with HGCAL modules. It presented a comprehensive study on the automation and reliability of minimum ionising particle (MIP) scale calibration, which is a crucial step for interpreting energy deposits in silicon sensors. Emphasis was placed on ensuring the physical consistency of the MIP spectra, including critical preprocessing steps such as TrigTime synchronisation, ensuring the 25 ns integration window correctly captures the signal peak, pedestal corrections, and accurate common-mode noise subtraction. These steps proved essential for achieving a stable and meaningful signal-to-noise ratio across the majority of modules and relays considered.

The results of the calibration method testing on a test beam demonstrated that after the application of appropriate corrections and selection criteria, the extracted MIP peak spectra align with the prediction. The implementation of an automated, per-channel fitting procedure allowed for the efficient processing of large datasets. However, several challenges were encountered, among them the limited statistical power of the dataset, largely due to short muon beam runs, lack of physical reference in the data since the defocused beam limited the statistics per channel, leading to the grouping of data per half-ROC and ROC. The test setup was shared with multiple users conducting different studies under varying conditions, which made it challenging to keep up with the changes in conditions that might affect the results of the work being done with the MIP fits.

Over and above, this work provides a validated, automated method for per-channel calibration at the MIP scale, based on beam test data, which is suitable for integration into the broader HGCAL recon-

struction framework. It provides a reproducible calibration procedure along with quantitative performance metrics that can guide future decisions in the development of both online and offline calibration strategies. In addition, it demonstrates the feasibility of precise local signal reconstruction in highly granular calorimeters, with methods that can be adapted to similar high-rate, high-resolution detectors in other collider experiments.

Beyond calibration, preliminary benchmarking of the reconstruction workflow suggests promising potential for scalability, particularly with GPU-accelerated implementations. However, further time and dedicated effort are needed to fully optimise and validate the performance across different hardware configurations, especially in preparation for the demanding conditions expected at the HL-LHC.

7.2 Future Work

While this study establishes a solid foundation for calibration under beam test conditions, several key steps remain before the full deployment of the HGCal system. Expanded testing and data collection across groups of modules will be necessary. In particular, more hexaboards and HGCal modules must undergo both electrical validation and beam testing to improve statistical coverage and assess performance across different manufacturing batches. As time moves closer to the assembly of larger components, such as full cassettes, careful monitoring and refinement of test-beam conditions, especially beam profiles, trigger timing, and synchronisation, will be crucial to ensure the quality and usability of the resulting calibration data. Further development of the reconstruction algorithms is also required. Calibration must be extended to include additional readout modes, such as time-of-arrival (ToA) and time-over-threshold(ToT), and eventually be integrated into higher-level object reconstruction workflows as the complexity of the detector grows. All of the work will culminate in the successful construction, installation, and commissioning of the complete HGCal system at the CMS P5 site will depend on close coordination across multiple stages, including module production, cassette assembly, quality assurance, and real-time data validation.

In summary, this work represents a meaningful step toward ensuring that the HGCal detector will meet the demanding physics and performance requirements of the HL-LHC era. The calibration procedures, tools, and results presented here will continue to support ongoing commissioning efforts and contribute directly to the successful integration of HGCal into the CMS experiment.

Bibliography

- [1] CMS Collaboration. HGCAL system performance tests conducted in 2024 using pre-series components in electron, pion, and muon beams. *CMS Detector Performance Summary*, 2025. URL <https://cds.cern.ch/record/2932769>.
- [2] S. Weinberg. The Making of the standard model. *Eur. Phys. J. C*, 34:5–13, 2004. doi: 10.1140/epjc/s2004-01761-1.
- [3] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.
- [4] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [5] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [6] ATLAS Collaboration. Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2012.08.020>.
- [7] Cush. Standard model of elementary particles. https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, 2019.
- [8] The stairway to heaven. *Physics Reports*, 1115:1–2, 2025. ISSN 0370-1573. doi: <https://doi.org/10.1016/j.physrep.2025.01.004>. URL <https://www.sciencedirect.com/science/article/pii/S0370157325000249>. CMS physics results from the first decade of LHC data.
- [9] CMS Collaboration. A portrait of the higgs boson by the cms experiment ten years after the discovery. *Nature*, 607(7917):60–68, July 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04892-x. URL <http://dx.doi.org/10.1038/s41586-022-04892-x>.
- [10] CMS Collaboration. Full summary: proton-proton collisions since 2010 (Run 1 + Run 2 + Run 3) . https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Full_summary_proton_proton_colli, 2023.

- [11] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN Yellow Reports: Monographs. CERN, Geneva, 2011. doi: 10.5170/CERN-2011-002. URL <https://cds.cern.ch/record/1318996>. The authors dedicate this Report to the memory of Nicola Cabibbo and Georges Charpak.
- [12] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3.Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN Yellow Reports: Monographs. CERN, 2013. doi: 10.5170/CERN-2013-004. URL <https://cds.cern.ch/record/1559921>. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page:<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.
- [13] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. CERN Yellow Reports: Monographs. CERN, 2017. doi: 10.23731/CYRM-2017-002. URL <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32>.
- [14] B. Fleming, I. Shipsey, M. Demarteau, J. Fast, S. Golwala, Y.-K. Kim, A. Seiden, J. Hirschauer, G. Sciolla, O. Palamara, et al. Basic research needs for high energy physics detector research & development: Report of the office of science workshop on basic research needs for hep detector research and development: December 11-14, 2019. Technical report, USDOE Office of Science (SC) (United States), 12 2019. URL <https://www.osti.gov/biblio/1659761>.
- [15] R. K. Ellis, B. Heinemann, and e. a. de Blas. Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020. Technical report, European Strategy for Particle Physics Preparatory Group, Geneva, 2019. URL <http://cds.cern.ch/record/2691414>. 254 p.
- [16] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [17] S. Röcker. *Perspektiven zur Beobachtung von Einzel-Top-Quark-Produktion am CMS-Experiment mit Hilfe neuronaler Netze*. PhD thesis, KIT, 03 2011.
- [18] D. d’Enterria. CMS physics highlights in the LHC Run 1. *PoS*, Bormio2015:027, 2015. doi: 10.22323/1.238.0027. URL <https://cds.cern.ch/record/2012170>.
- [19] CMS Collaboration. Overview of the CMS Detector Performance at LHC Run 2. *Universe*, 5, 01 2019.
- [20] A. Hayrapetyan et al. Development of the CMS detector for the CERN LHC Run 3. *JINST*, 19(05): P05064, 2024. doi: 10.1088/1748-0221/19/05/P05064.
- [21] CERN. Schedule. <https://project-hl-lhc-industry.web.cern.ch/content/project-schedule>, 2025.
- [22] CMS. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12 (10):P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003. URL <http://cds.cern.ch/record/2270046>.

- [23] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger. Technical report, CERN, Geneva, 2021. URL <https://cds.cern.ch/record/2759072>. This is the final version of the document, approved by the LHCC.
- [24] C. CMS. A MIP Timing Detector for the CMS Phase-2 Upgrade. Technical report, CERN, Geneva, 2019. URL <https://cds.cern.ch/record/2667167>.
- [25] CMS Collaboration. The Phase-2 upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2283189>.
- [26] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024. doi: 10.1103/PhysRevD.110.030001.
- [27] M. Livan and R. Wigmans. *Shower Development*, pages 53–73. Springer International Publishing, Cham, 2019. ISBN 978-3-030-23653-3. doi: 10.1007/978-3-030-23653-3_3. URL https://doi.org/10.1007/978-3-030-23653-3_3.
- [28] Y. Israeli. *Energy Reconstruction in Highly Granular Calorimeters for Future Electron-Positron Colliders*. PhD thesis, TUM, 10 2018.
- [29] P. D. Group. Atomic and nuclear properties of materials for more than 300 materials. <https://pdg.lbl.gov/2012/AtomicNuclearProperties/>.
- [30] CMS Collaboration. Performance of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 13$ TeV. 2024. doi: 10.48550/ARXIV.2403.15518. URL <https://arxiv.org/abs/2403.15518>.
- [31] C. W. Fabjan and F. Gianotti. Calorimetry for Particle Physics. *Rev. Mod. Phys.*, 75:1243–1286, 2003. doi: 10.1103/RevModPhys.75.1243. URL <https://cds.cern.ch/record/692252>.
- [32] CMS Collaboration. Energy resolution of the barrel of the CMS electromagnetic calorimeter. Technical report, CERN, Geneva, 2007. URL <https://cds.cern.ch/record/1009081>.
- [33] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 8(09):P09009–P09009, Sept. 2013. ISSN 1748-0221. doi: 10.1088/1748-0221/8/09/p09009. URL <http://dx.doi.org/10.1088/1748-0221/8/09/P09009>.
- [34] CMS HGCal collaboration. Response of a CMS HGCal silicon-pad electromagnetic calorimeter prototype to 20–300 GeV positrons. *Journal of Instrumentation*, 17(05):P05022, may 2022. doi: 10.1088/1748-0221/17/05/P05022. URL <https://dx.doi.org/10.1088/1748-0221/17/05/P05022>.
- [35] P. Paulitsch. The silicon sensors for the high granularity calorimeter of cms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 978:164428, 2020. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2020.164428>. URL <https://www.sciencedirect.com/science/article/pii/S0168900220308251>.

- [36] D. Barney. Overview slide of ce with main parameters, 2022. URL <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=13251&version=9>.
- [37] A. Lobanov. Electronics and triggering challenges for the cms high granularity calorimeter. *Journal of Instrumentation*, 13(02):C02056–C02056, Feb. 2018. ISSN 1748-0221. doi: 10.1088/1748-0221/13/02/c02056. URL <http://dx.doi.org/10.1088/1748-0221/13/02/c02056>.
- [38] CMS Collaboration. The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2293646>.
- [39] M. Valentan. The cms high granularity calorimeter for the high luminosity lhc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 936:102–106, 2019. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2018.10.131>. URL <https://www.sciencedirect.com/science/article/pii/S016890021831444X>. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.
- [40] H.-G. Moser. Silicon detector systems in high energy physics. *Progress in Particle and Nuclear Physics*, 63(1):186–237, 2009. ISSN 0146-6410. doi: <https://doi.org/10.1016/j.pnpnp.2008.12.002>. URL <https://www.sciencedirect.com/science/article/pii/S0146641008000951>.
- [41] A. David. Performance of the cms hgcal for lhc phase 2. ICHEP 2022 Bologna, 2022.
- [42] M. Noy. The cms hgcal silicon region architecture specification and optimisation, 2021.
- [43] F. A. Khan. Design and Performance Optimisation of the Hexaboards for CMS HGCal On-Cassette Readout Electronics. Technical Report 03, CERN, Geneva, 2023. URL <https://cds.cern.ch/record/2839771>.
- [44] CMS Collaboration. HGCROC3 Spec Working Document 2.0, 2021.
- [45] lpGBT Team. lpgbtv1 manual. Technical report, CERN, 2024. URL <https://cds.cern.ch/record/2809058>.
- [46] P. M. Duarte Rosado. Scalable FPGA-based controller of a thousand-ASIC system in the CERN CMS HGCal detector. Master’s thesis, IST, 2023. URL <https://cds.cern.ch/record/2856363>. Presented Nov 2022.
- [47] CERN SBA Secondary Beam and Areas department. The H2 secondary beam line of EHN1/SPS. URL <http://sba.web.cern.ch/sba/BeamsAndAreas/h2/H2manual.html>.
- [48] G. Liu. Experience and first results of the hgcal beam tests in 2023, 2024. URL https://indico.cern.ch/event/1323113/contributions/5823577/attachments/2838827/4961559/20240417_BTTB2024_HGCALBT.pdf. Accessed: 2024-06-16.
- [49] H. Bichsel. Straggling in thin silicon detectors. *Rev. Mod. Phys.*, 60:663–699, Jul 1988. doi: 10.1103/RevModPhys.60.663. URL <https://link.aps.org/doi/10.1103/RevModPhys.60.663>.

Appendix A

Other results

This appendix contains supplementary plots illustrating the procedure used to extract the MIP response from the test beam data. Given the limited number of muon hits per channel in many runs, a large number of fits were carried out both at the individual channel level and in aggregated groups (eRx and ROC level). The examples shown here reflect the variability in fit quality depending on statistics, channel behaviour, and data-taking conditions. They serve to complement the representative fits shown in the main text and to document the robustness of the method used to determine the signal-to-noise performance of the silicon modules. ¹

¹Over 200 fits were performed and reviewed manually across different modules and aggregation levels. Each fit was inspected for convergence, peak structure, and alignment with expected MIP behaviour. This process was critical to ensure reliable signal extraction in the presence of statistical limitations.

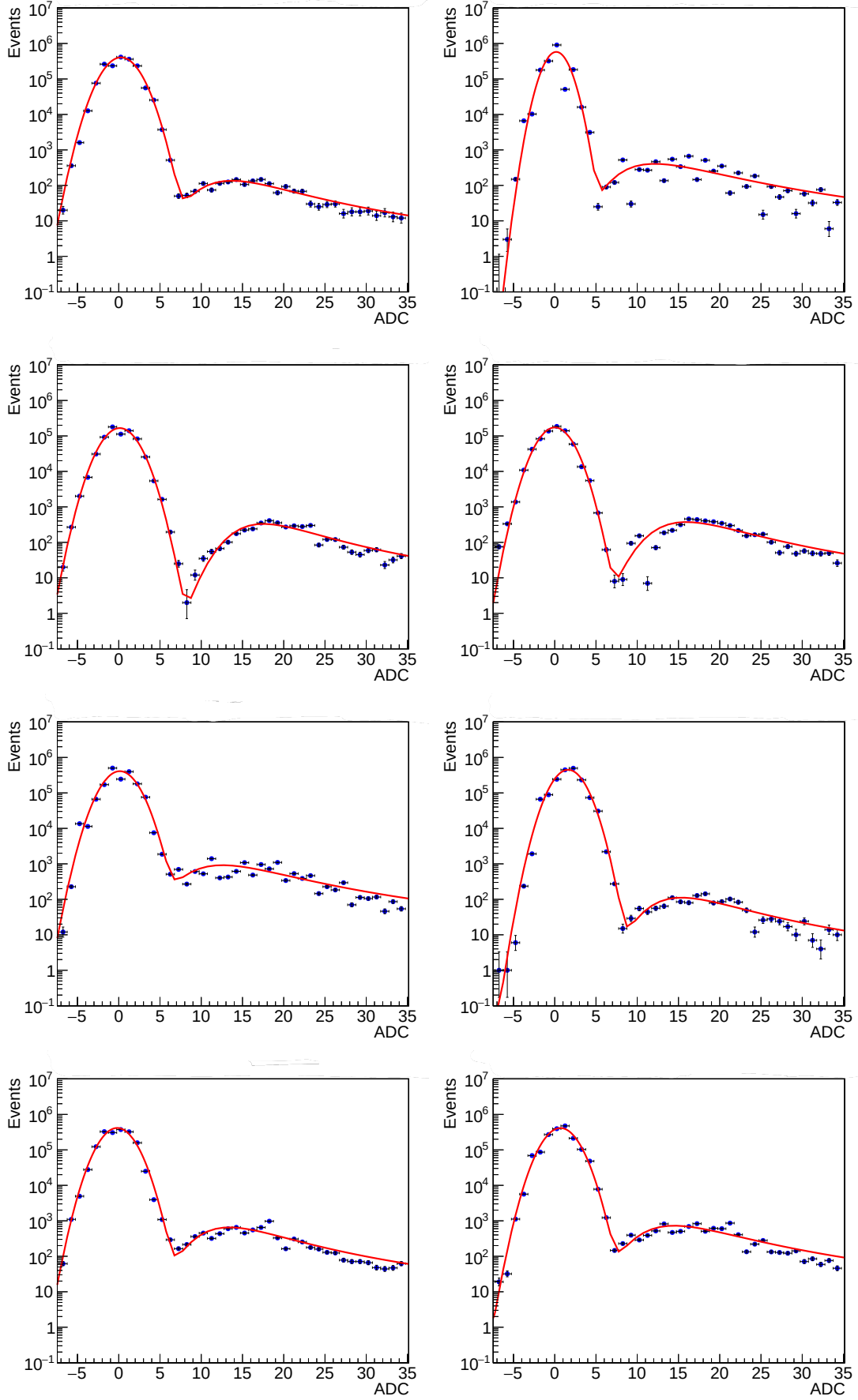


Figure A.1: MIP Spectrum for 8 channels. Each row displays two plots, showing the results for different channels (ML14 channels 10 and 25, ML16 channels 149 and 151, ML17 channels 36 and 9, ML19 channel 153 and ML20 channel 64) in several modules for Relay 1726518879.

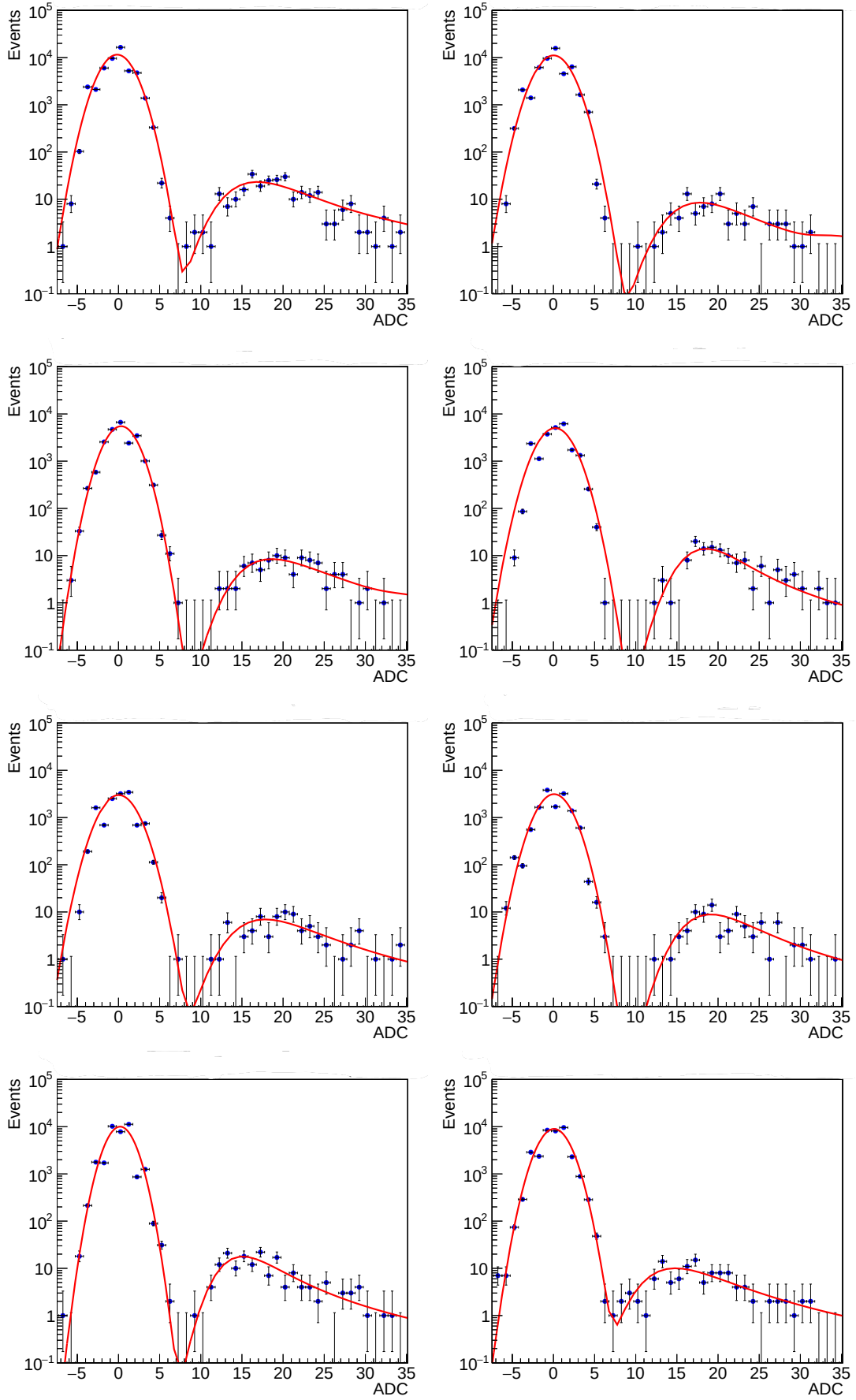


Figure A.2: MIP Spectrum for 8 channels. Each row displays two plots, showing the results for different channels(ML14 channels 27 and 34, ML16 channels 153 and 210, ML17 channels 29 and 33, ML18 channel 24 channel 33) in several modules for Relay 1726347199.

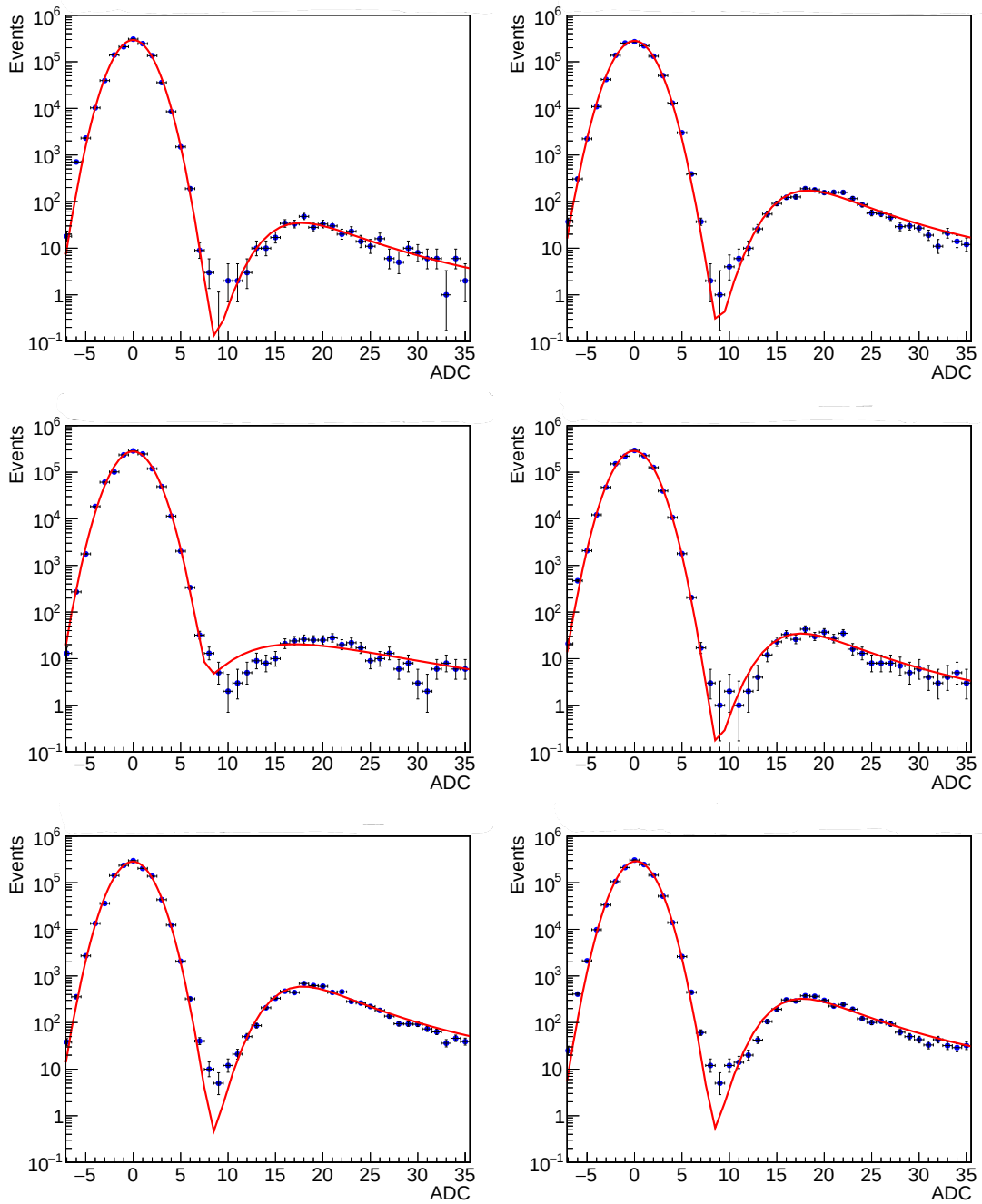


Figure A.3: MIP Spectrum for all the eRx (from 0 to 5 in order) for Module 16 from Relay 1726347199.

Table A.1: Summary of MIP fit parameters of interest of several runs. Each entry includes the relay number, module identification, Most Probable Value (MPV), noise σ

Run	Channel Identification	MPV	σ
1726518879	Module 14 Channel 10	13.33	1.65
1726518879	Module 14 Channel 25	12.25	1.15
1726518879	Module 16 Channel 149	17.39	1.63
1726518879	Module 16 Channel 151	16.12	1.56
1726518879	Module 17 Channel 9	13.61	1.43
1726518879	Module 17 Channel 36	12.61	1.63
1726518879	Module 19 Channel 153	13.94	1.43
1726518879	Module 20 Channel 64	13.37	1.62
1726347199	Module 16 eRx 0	17.77	1.54
1726347199	Module 16 eRx 1	18.25	1.61
1726347199	Module 16 eRx 2	17.76	1.62
1726347199	Module 16 eRx 3	17.72	1.58
1726347199	Module 16 eRx 4	17.91	1.59
1726347199	Module 16 eRx 5	17.39	1.56
1726347199	Module 14 channel 27	17.26	1.68
1726347199	Module 14 channel 34	17.75	1.74
1726347199	Module 16 channel 153	18.08	1.60
1726347199	Module 16 channel 189	18.17	1.60
1726347199	Module 16 channel 210	17.84	1.75
1726347199	Module 16 ROC 0	18.14	1.57
1726347199	Module 17 channel 29	17.79	1.76
1726347199	Module 17 channel 33	18.89	1.62
1726347199	Module 18 channel 24	14.63	1.40
1726347199	Module 18 channel 33	14.79	1.56

