

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF PHYSICS AND ASTRONOMY ”A. RIGHI”

SECOND CYCLE DEGREE

PHYSICS

Study of the evolution of radiation-induced damage in Silicon Photomultipliers with Emission Microscopy

Supervisor

Prof. Pietro Antonioli

Co-supervisor

Dr. Roberto Preghenella
Dr. Antonio Paladino

Defended by

Ludovica Rainero

Graduation Session October 2026

Academic Year 2025/2026

Sono stato fortunatissimo nelle cose materiali; molti hanno lavorato più di me e non hanno ottenuto metà dei miei risultati; ma io non avrei mai potuto fare quello che ho fatto senza le abitudini formate in quel periodo, abitudini di puntialità, d'ordine, di diligenza, senza la determinazione di concentrarmi su un oggetto alla volta, anche se seguito a brevissima distanza da un altro. Sa il cielo che non scrivo questo con spirito di autoesaltazione. Voglio soltanto dire che, qualsiasi cosa io abbia cercato di fare in vita mia, ho cercato di farla bene, con tutta l'anima; a qualunque cosa mi sia dedicato, mi sono dedicato interamente; che nelle cose grandi e nelle piccole, ho sempre fatto veramente sul serio. Non ho mai creduto che una capacità naturale o acquisita possa pretendere di sussistere e sperare di raggiungere il suo fine, indipendentemente da qualità di semplice applicazione, costanza e tenacia. Non mettere mai mano a niente a cui non potessi dedicare tutto me stesso, e non atteggiarmi mai a deprezzare il mio lavoro, qualunque fosse: queste, me ne accorgo ora, sono state le mie regole auree.

Charles Dickens, *David Copperfield*

Abstract

Silicon Photomultipliers (SiPMs) are compact solid-state photon detectors combining single-photon sensitivity, high photon-detection efficiency, fast response and insensitivity to magnetic fields. These characteristics make them suitable candidates for the photon-detection system of the dual-radiator Ring Imaging Cherenkov detector of the ePIC experiment at the future Electron-Ion Collider. Their operation in a radiation environment, however, is affected by radiation damage, which increases the dark-count rate and dark current and consequently degrades single-photon detection performance. Characterizing the evolution of this damage and its recovery through thermal annealing is therefore important for the long-term operation of SiPMs, and for fundamental studies of radiation-induced defects and their evolution in SiPMs. This thesis investigates radiation-induced modifications in Hamamatsu SiPMs through Emission Microscopy (EMMI), a spatially resolved optical technique sensitive to the secondary photons produced during Geiger avalanches. Devices with different microcell pitches were characterized before irradiation, after irradiation with approximately $140\ \mathrm{MeV}$ protons corresponding to a $1\ MeV$ neutron-equivalent (n_{eq}) fluence of $10^9\ \mathrm{cm}^{-2}$, and throughout a sequence of thermal annealing spanning temperatures between $75^{\circ}\mathrm{C}$ and $150^{\circ}\mathrm{C}$. Spatially localized regions of avalanche-associated optical emission, referred to as *hotspots*, were individuated. The analysis combines whole-device hotspot counting, the characterization of luminosity distributions and population-averaged trends, measurements of the optical intensity of individual emission sites and its dependence on temperature and overvoltage, and the tracking of individual hotspots across successive annealing stages.

Irradiation produces a strong increase in the number of detectable emission hotspots. Four SiPMs were investigated after irradiation, while one of them had also been characterized before irradiation, providing a direct reference for the comparison. For this device, the number of detected hotspots increases from 23 ± 2 before irradiation to 562 ± 29 after irradiation. Post-irradiation populations in the range of 500–600 hotspots are observed in all four sensors. Irradiation gives rise to a new population of detectable hotspots characterized by a luminosity distribution that differs significantly from that observed before irradiation. Under the same operating conditions, $T = 20^{\circ}\mathrm{C}$ and $V_{\mathrm{over}} = 7\ \mathrm{V}$, the luminosities measured before irradiation range within [20, 480] a.u., with a median of L_{median}(before irradiation) $= 56\ a.u.$, whereas after irradiation the distribution extends to substantially higher values, with a range within [40, 2600] a.u. and a median of L_{median}(after irradiation) $= 210\ a.u.$. The measured hotspot luminosity before irradiation increases exponentially with temperature and follows an approximately quadratic dependence on overvoltage. After irradiation, the hotspot luminosity retains an exponential dependence on temperature, although the temperature coefficient λ changes, while during annealing it remains of the same order of magnitude for most stages. The overvoltage dependence remains close to quadratic under most investigated conditions. Thermal annealing progressively reduces radiation-induced optical activity: after the complete treatment, the whole-device hotspot population is reduced by approximately 55–60%, while the mean optical intensity and dark current exhibit even larger relative decreases. Individual hotspot tracking nevertheless reveals a heterogeneous microscopic evolution, including gradual

decreases, disappearance and reappearance, transient variations and, more rarely, increasing emission. These results show that EMMI provides a valuable spatial probe of radiation-induced avalanche activity in SiPMs, complementing electrical measurements integrated over the whole device and providing insight into the spatial distribution and evolution of radiation-induced defects. They also indicate that thermal annealing acts both by reducing the hotspot population and, more prominently, by decreasing the luminosity of the hotspots that remain detectable. EMMI therefore opens a new window onto the development of radiation-induced defects in SiPMs and their evolution during thermal annealing.

Contents

Abstract i

Introduction 1

1 The Electron Ion Collider 4

1.1 Quantum Chromodynamics and hadronic matter . . . 5

1.1.1 Strong interaction . . . 5

1.1.2 The QCD Lagrangian . . . 5

1.1.3 Confinement and asymptotic freedom . . . 7

1.1.4 Confinement . . . 8

1.1.5 Origin of nucleon mass and spin . . . 10

1.1.6 Non-perturbative versus perturbative QCD . . . 12

1.2 Proton structure . . . 14

1.2.1 DIS kinematic variables . . . 14

Virtuality of the photon . . . 14

Bjorken scaling variable . . . 15

Inelasticity . . . 15

ν energy variable . . . 15

1.2.2 Structure Functions . . . 15

1.2.3 The quark parton model and structure functions . . . 17

Static quark model . . . 17

Dynamic quark model . . . 17

1.2.4 Collinear approximation and PDFs . . . 19

1.2.5 Experimental Access to PDFs . . . 20

Definition . . . 21

1.2.6 Transverse momentum structure and TMDs . . . 21

1.2.6.1 Wigner distributions . . . 21

1.2.6.2 TMDs factorization . . . 22

1.2.7 Experimental access to TMDs and FFs . . . 24

Definition . . . 24

Polarization . . . 25

1.2.8 Spatial imaging and GPDs . . . 25

1.2.9 Experimental Access to GPDs . . . 25

Definition . . . 25

Deeply Virtual Compton Scattering . . . 25

Exclusive Vector Meson Production . . . 25

1.3 Nuclear structure and QCD in nuclei . . . 26

1.3.1 Nuclear modifications of parton distributions . . . 26

1.3.2 Shadowing and the small-x region . . . 28

1.3.3 Increasing gluon density and possible saturation . . . 29

1.3.4 Coherent and diffractive processes . . . 31
1.3.5 Propagation and hadronization of partons in nuclear matter . . . 32
1.3.6 Comparison of ep and eA collisions . . . 34
1.4 The EIC physics programme . . . 35
1.4.1 Origin of nucleon mass and spin . . . 35
1.4.2 Multidimensional imaging of nucleons . . . 36
1.4.3 Gluon distributions and saturation . . . 38
1.4.4 Emergence of hadrons and interaction with nuclear matter . . . 39
1.5 Experimental requirements to access to EIC physics . . . 40
1.5.1 Inclusive measurements . . . 41
Scattered electron . . . 42
Hadronic recoil . . . 42
1.5.2 Semi-inclusive measurements . . . 43
Electron and charged hadron identification . . . 43
1.5.3 Exclusive measurements . . . 44
1.6 The ePIC experiment . . . 44
1.6.1 Overall detector concept . . . 44
1.6.2 Asymmetric detector layout . . . 45
1.6.3 Tracking and vertexing . . . 46
1.6.4 Calorimetry . . . 47
1.6.5 Far-forward and far-backward systems . . . 49
1.6.6 General PID role . . . 50

2 Particle Identification at the EIC **52**
2.1 Particle identification . . . 52
2.2 Role of particle identification in EIC physics . . . 54
2.2.1 Flavour separation in SIDIS . . . 54
2.2.2 Fragmentation functions . . . 55
2.2.3 Transverse-spin and TMD measurements . . . 57
2.2.4 Strange-hadron and heavy-flavour studies . . . 58
2.2.5 Exclusive final states . . . 60
2.3 Particle-identification requirements . . . 61
2.3.1 Momentum and pseudorapidity distributions . . . 62
2.3.2 Required $\pi/K/p$ separation . . . 64
2.3.3 Different requirements in the electron endcap, barrel and hadron endcap 65
2.3.4 Complementarity of timing and dRICH . . . 65
2.4 ePIC PID systems . . . 67
2.4.1 Backward proximity-focusing RICH . . . 67
2.4.2 Barrel DIRC . . . 69
2.4.3 Barrel and forward TOF . . . 70
2.4.4 Forward dRICH . . . 71
2.5 Dual-radiator RICH detector . . . 72
2.5.1 Detector principle . . . 72

2.5.1.1 Aerogel and gas radiators . . . 72
Aerogel radiator . . . 72
Gas radiator . . . 73
2.5.2 Common focusing system . . . 75
2.5.3 Photon detection plane . . . 75
ALCOR readout . . . 76
Streaming readout . . . 77
2.5.4 Magnetic field and geometrical constraints . . . 78
2.6 dRICH SiPM-based photon detection . . . 79
2.6.1 Need for single-photon sensitivity . . . 80
2.6.2 Spatial and timing requirements . . . 81
2.6.3 Choice of SiPMs . . . 82
2.6.4 Magnetic-field insensitivity . . . 83
2.6.5 Dark-count background . . . 84
2.6.6 Expected radiation damage . . . 85
2.6.7 Cooling and annealing as mitigation strategies . . . 87

3 Silicon Photomultipliers **89**
3.1 Silicon Photomultipliers operating principle . . . 89
3.1.1 Diode working principle: the photoconversion . . . 89
3.1.2 SPAD working principle: the multiplication . . . 91
3.1.3 SPAD fabrication . . . 92
3.2 Main SiPM parameters . . . 94
3.2.1 SPAD electrical model and Gain . . . 94
3.2.2 Photodetection probability . . . 96
3.2.3 SiPM structure . . . 97
3.2.4 Photodetection efficiency . . . 98
3.2.5 Break-down voltage and overvoltage . . . 98
3.2.6 Dynamic range . . . 100
3.2.7 Timing resolution . . . 101
3.3 Noise and dark-count rate . . . 104
3.3.1 Dark Count Rate . . . 104
3.3.2 Afterpulse . . . 105
3.3.3 Prompt, delayed and external crosstalk . . . 106
3.3.4 Dark current and leakage current . . . 107
3.4 Radiation damage of SiPM . . . 108
3.4.1 Ionizing and Non-Ionizing Energy Loss . . . 109
3.4.2 Point and cluster defects . . . 110
3.4.3 Radiation-induced energy levels in the band gap . . . 110
3.4.4 Increase of generation current and DCR . . . 112
3.4.5 Loss of single-photon resolution . . . 114
3.4.6 Fluence and temperature dependence . . . 115
3.5 Radiation-damage mitigation . . . 117

3.5.1 Operation at reduced temperature . . . 117
3.5.2 Narrow timing windows . . . 119
3.5.3 Sensor and electronics optimization . . . 120
3.5.4 Thermal annealing . . . 121
3.5.5 Conventional oven annealing . . . 121
3.5.6 In-situ annealing . . . 122
3.6 Defect characterization . . . 125
3.6.1 Light emission during avalanche processes . . . 125
3.6.2 Emission hotspots and electrically active defects . . . 126
3.6.3 Emission microscopy for SiPM characterization . . . 127
3.6.4 Observables and limitations of EMMI . . . 129
3.6.5 Motivation for studying irradiation and annealing . . . 130

4 Emission Microscopy measurements **131**
4.1 Sensors under study . . . 131
4.2 EMMI Experimental Set-Up . . . 132
4.2.1 Peltier cooler and temperature sensor . . . 134
4.2.2 EMMI acquisition system . . . 134
4.2.3 SiPM biasing and I-V characterization . . . 135
4.2.4 Oven for annealing . . . 136
4.3 Image acquisition . . . 136
4.4 Image processing . . . 140
4.4.1 Image clean-up . . . 141
4.4.2 Image rotation and shift . . . 143
4.5 Hotspot characterization . . . 145
4.5.1 Identification and localisation . . . 145
Background estimation and subtraction . . . 145
Detection threshold . . . 145
Source detection and deblending . . . 146
Measurement of source properties . . . 147
Inspection of an individual source . . . 147
4.5.2 Luminosity calculation . . . 147
Luminosity calculation . . . 147
Statistical uncertainty calculation . . . 148
4.5.3 Hotspot identification across annealing phases . . . 148
4.6 Evaluation of systematic uncertainties . . . 149
4.6.1 Study on isolation cut . . . 149
4.6.2 Study on radius . . . 152
4.6.3 Calculation of systematic uncertainties . . . 155

5 Results and Discussion **157**
5.1 Reference measurements on non-irradiated SiPMs . . . 157
5.1.1 Hotspots counting and distribution . . . 157
5.1.2 Hotspots intensity . . . 158

5.1.3 Dependence on temperature . . . 159
5.1.4 Dependence on overvoltage . . . 161
5.2 Temperature dependence of the emission intensity . . . 163
5.2.1 Parametrization of the temperature dependence . . . 164
5.2.2 Fit quality and hotspot-to-hotspot variability . . . 164
5.2.3 Comparison between new and irradiated sensor . . . 166
5.2.4 Evolution of the temperature coefficient with annealing . . . 169
5.3 Overvoltage dependence of the emission intensity . . . 172
5.3.1 Parametrization of the overvoltage dependence . . . 173
5.3.2 Variation among individual emission hotspots . . . 176
5.3.3 Quality of the mean exponent description . . . 176
5.3.4 Comparison between new and irradiated sensors . . . 177
5.3.5 Evolution during annealing . . . 178
5.4 Effects of irradiation . . . 180
5.4.1 Number of the emission hotspots . . . 181
5.4.2 Distribution of the emission intensities . . . 183
Comparison with the non-irradiated sensor. . . . 184
5.5 Evolution with annealing . . . 186
5.5.1 Hotspot counting – whole device . . . 186
5.5.2 Hotspot counting – central region . . . 190
5.5.3 Luminosity distributions . . . 194
Sensor A1 . . . 194
Sensor B1 . . . 196
Considerations . . . 196
5.5.4 Individual hotspots behaviour . . . 197
Decreasing and disappearing . . . 199
Decreasing . . . 199
Peaks . . . 201
Constant . . . 201
Increasing . . . 202
Appeared . . . 202
5.6 Discussion . . . 203

Conclusions **208**

List of Figures **210**

List of Tables **223**

References **224**

Acknowledgement **230**

Introduction

The Electron-Ion Collider (EIC) will provide a new experimental environment for the investigation of Quantum Chromodynamics and of the internal structure of hadronic matter. Through collisions of polarized electrons with polarized protons and nuclei over a broad kinematic range, the EIC physics programme is designed to address fundamental questions concerning the emergence of nucleon mass and spin, the multidimensional distribution of quarks and gluons, the behaviour of gluons at high densities, and the formation and propagation of hadrons in nuclear matter. Extracting this physics requires not only high-precision tracking and calorimetry, but also efficient identification of the particles produced in the collision. Particle identification is particularly important for semi-inclusive and exclusive measurements, in which the flavour and kinematic information carried by the final-state hadrons must be reconstructed. In the ePIC detector, different particle-identification technologies are combined to provide the required momentum and pseudorapidity coverage. In the forward region, an important role is fulfilled by the dual-radiator Ring Imaging Cherenkov detector (dRICH), in which Cherenkov photons produced in two different radiators are focused onto a dedicated photon-detection plane. The detection of these photons requires sensors capable of operating at the single-photon level with high spatial and temporal resolution and within the magnetic field of the experiment. Silicon Photomultipliers represent a solution for this application. Their high photon-detection efficiency, compactness, fast response and insensitivity to magnetic fields make them suitable for the dRICH photon-detection system. At the same time, their operation in the radiation environment expected at the EIC introduces an important limitation. Non-ionizing radiation can produce damages in the silicon lattice that act as generation centres, increasing the dark-count rate and dark current and progressively degrading the ability of the sensor to resolve individual photons. Radiation damage therefore represents one of the relevant aspects that must be understood and mitigated in view of long-term SiPM operation. Several strategies can be adopted to reduce its impact. Cooling suppresses thermally generated carriers during detector operation, while thermal annealing can partially recover radiation-induced degradation. Electrical measurements such as dark-current and dark-count-rate measurements provide an essential macroscopic characterization of this recovery. However, they integrate the response of the complete device and do not directly reveal how the underlying radiation-induced activity is spatially distributed or how individual active regions evolve during annealing. The work presented in this thesis addresses this aspect through Emission Microscopy (EMMI). During a Geiger avalanche, secondary photons, in the optical spectrum, are emitted by active sites. By imaging this light with a sufficiently sensitive camera, spatially localized regions characterized by enhanced avalanche activity can be observed and be referred to as emission hotspots. EMMI therefore provides a complementary approach to conventional electrical measurements: rather than measuring only the integrated response of the complete SiPM, it allows the spatial distribution and optical intensity of individual light-emitting hotspots to be investigated. The aim of this thesis is to use this technique to characterize the modifications induced by irradiation and, in particular, to follow the evolution of radiation-induced emission centres throughout a controlled thermal-annealing sequence. The thesis is organized as follows. Chapter 1, The Electron Ion Collider, introduces the

physics framework motivating the experiment. It begins with the basic aspects of Quantum Chromodynamics, including the strong interaction, confinement and asymptotic freedom, and the unresolved questions concerning the origin of nucleon mass and spin. The discussion then moves to the structure of the proton, introducing deep-inelastic scattering, structure functions, parton distribution functions, transverse-momentum-dependent distributions and generalized parton distributions. Nuclear modifications and QCD phenomena in nuclei are subsequently discussed, leading to the principal elements of the EIC physics programme. The chapter concludes by translating these physics goals into experimental requirements and by introducing the overall concept of the ePIC detector, including its tracking, calorimetric and particle-identification systems. Chapter 2, Particle Identification at the EIC, focuses on the role of particle identification in accessing the EIC physics programme. The connection between particle identification and flavour separation, fragmentation functions, transverse-spin and TMD measurements, strange- and heavy-flavour physics, and exclusive final states is discussed. The momentum and pseudorapidity ranges relevant to ePIC are then used to motivate the required pion, kaon and proton separation and the complementarity of the different PID systems. Particular attention is devoted to the forward dRICH, its dual-radiator concept and common focusing system, and to the requirements imposed on its photon-detection plane. The chapter finally introduces SiPM-based photon detection, discussing the need for single-photon sensitivity and the implications of dark-count background and radiation damage, as well as cooling and annealing as possible mitigation strategies. Chapter 3, Silicon Photomultipliers, provides the detector-physics background required for the experimental study. The operating principle of semiconductor photodiodes, Single-Photon Avalanche Diodes and complete SiPM devices is reviewed, followed by the main parameters governing their response, including gain, photodetection probability and efficiency, breakdown voltage, overvoltage, dynamic range and timing resolution. Particular attention is then given to the different sources of SiPM noise, including dark counts, afterpulsing and optical cross-talk, and to their relation with dark and leakage currents. The effects of radiation on silicon are subsequently introduced, from energy deposition and displacement damage to the formation of point and cluster defects and radiation-induced levels in the band gap. The corresponding increase in dark current and dark-count rate and the degradation of single-photon resolution are discussed together with possible mitigation strategies, including thermal annealing. The final part of the chapter introduces defect characterization through avalanche-induced light emission and establishes the physical basis, observables and limitations of EMMI. Chapter 4, Emission Microscopy measurements, describes the experimental procedure developed for this thesis. The investigated SiPMs and the EMMI apparatus are presented first, including temperature control, SiPM biasing, current-voltage characterization and the oven used for the annealing treatments. The acquisition procedure is then described together with the image-processing chain required to remove camera artefacts and geometrically align measurements acquired at different annealing stages. Particular emphasis is placed on the identification and localization of emission hotspots, the determination of their optical intensity and the procedure used to associate the same spatial position across successive annealing phases. The chapter concludes with the evaluation of the systematic uncertainties associated with the hotspot-selection and luminosity-integration procedures. Finally, Chapter 5, Results and Discussion, presents the experimental results. Reference measurements on non-irradiated SiPMs first establish the initial

hotspot population and its optical properties. The dependence of the emission intensity on temperature and overvoltage is then characterized, providing empirical parameters with which the different sensor conditions can be compared. The effects of irradiation are studied through changes in the number, spatial distribution and luminosity of the detected emission hotspots. The central part of the chapter is devoted to their evolution during annealing, combining whole-device counting with spatial tracking of individual hotspots, normalized luminosity distributions and forced measurements at fixed hotspot coordinates. The EMMI observables are finally compared with the evolution of the dark current in order to identify common trends and differences between the spatially resolved optical response and the macroscopic electrical behaviour of the irradiated SiPMs. Through this progression, the thesis moves from the physics motivation of the EIC to a specific detector challenge and finally to a microscopic experimental investigation of radiation damage and annealing in SiPMs.

1 The Electron Ion Collider

Nuclei are composed of protons and neutrons, collectively referred to as nucleons, which are themselves not fundamental particles. Rather, nucleons are composite systems made of quarks and gluons, the fundamental degrees of freedom of the strong interaction, described by Quantum Chromodynamics (QCD). Despite the remarkable success of QCD, several fundamental questions concerning the structure and dynamics of strongly interacting matter remain unanswered. The Electron-Ion Collider (EIC) [1], shown in Figure 1, is designed to address these open questions and provide unprecedented insight into the quark and gluon structure of matter.

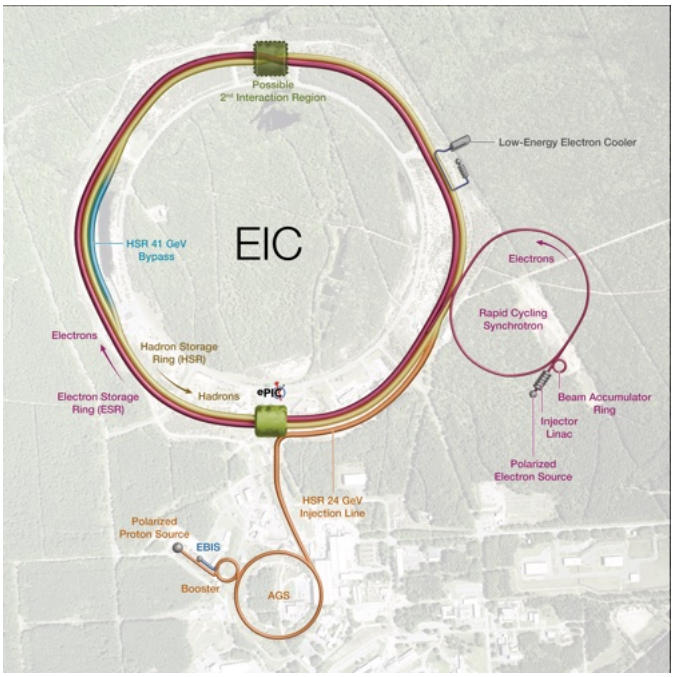


Figure 1: Schematic representing the planned Electron Ion Collider [1].

The EIC is a future electron-proton and electron-ion collider to be constructed at Brookhaven National Laboratory (BNL) on Long Island, New York by the U.S. Department of Energy in this decade. It will be constructed upgrading the existing Relativistic Heavy Ion Collider complex [2] and it is foreseen to start operation in the mid 2030. The EIC will be the world's first collider for polarised electron-proton (and light ions) and for electron-nucleus collisions, enabling multidimensional measurements of the spatial and momentum distributions of quarks and gluons that are essential for reconstructing a three-dimensional picture of nucleon and nuclear structure. It will have a variable centre-of-mass energy, running from $\approx 20\,\mathrm{GeV}$ to $\approx 140\,\mathrm{GeV}$ [3], and a luminosity in excess of $10^{33} cm^{-2} s^{-1}$. In order to probe the structure of the nucleons, high energy lepton-hadron inelastic scattering, where the hadron breaks up, are performed. The only collision point fully planned at the moment will host the ePIC (electron-Proton/Ion Collider) experiment. ePIC will have the classical barrel end-cap structure, with a series of complementary

sub-detectors dedicated to charged-particle tracking and vertex reconstruction, particle identification, electromagnetic and hadronic calorimetry, and precise measurements of particle energies and momenta.

This Chapter aims to address the physics case of the experiment. A preliminary overview of the fundamental physics will be provided, covering QCD (Sec. 1.1), proton structure (Sec. 1.2) and nuclear structure (Sec. 1.3). Sec. 1.4 will be devoted to a detailed description of the fundamental questions that EIC aims to answer. The subsequent sections are dedicated to the experimental architecture, describing how these fundamental questions are investigated experimentally (Sec. 1.5) and how the ePIC experiment is structured (Sec. 1.6).

1.1 Quantum Chromodynamics and hadronic matter

1.1.1 Strong interaction

The Standard Model (SM) of elementary Particle Physics is a theory that describes how the fundamental building blocks of matter interact through the fundamental forces. The Standard Model is composed by three of the four fundamental interaction in nature: Electromagnetic, Weak and Strong (the gravitational force is not yet included in the Standard Model) and by its "building blocks"

- 6 leptons (+6 antileptons)
- 6 quarks (+6 antiquarks)
- 12 gauge bosons

The Figure 2 represents a schematic of the model.

The strong force, also known as the strong interaction, is the fundamental interaction that binds quarks together to form protons, neutrons, and all other hadrons. The nuclear force that holds protons and neutrons together within atomic nuclei is a residual manifestation of this underlying strong interaction. At the most fundamental level, the strong force is described by the quantum field theory called Quantum Chromodynamics (QCD), which is a non-Abelian theory. In QCD, the strong interaction arises from the exchange of massless particles called gluons, which mediate the force between particles that carry a property known as colour charge. Apart from gluons, only quarks posses the colour charge. The symmetry group governing these interactions is SU(3) colour. To describe the dynamics within this framework, we require an invariant Lagrangian under local transformations of this symmetry group. Quantum Chromodynamics is a gauge theory, therefore is renormalizable, and its coupling constant, representing the interaction strength at a given energy scale, is a running parameters that varies with energy. Colour charge is not observable in isolation due to a property known as *colour confinement*, which prevents quarks and gluons from existing freely outside of hadrons. This property will be deeply discussed in Sec. 1.1.4.

1.1.2 The QCD Lagrangian

The complete Lagrangian describing QCD is

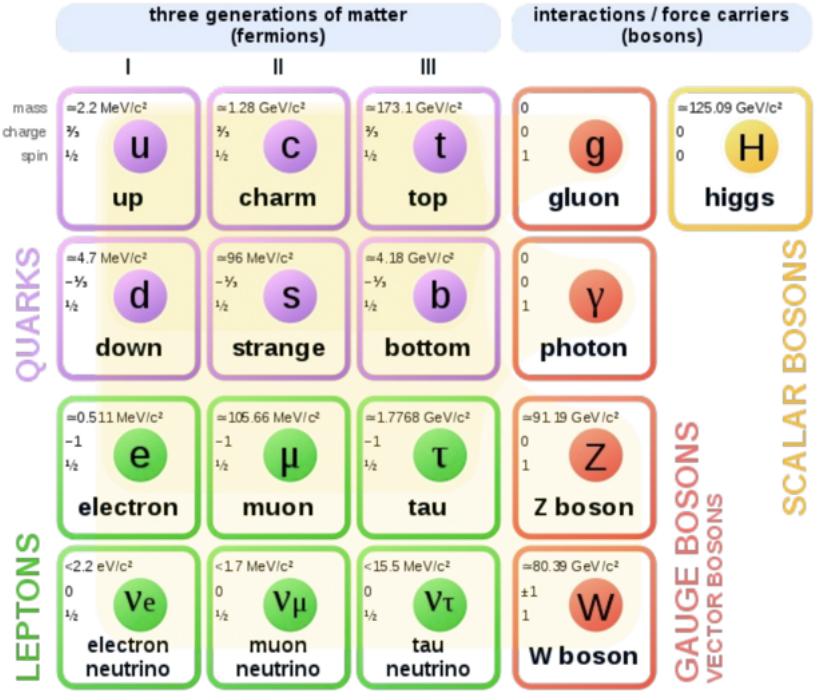


Figure 2: The Standard Model of Particle Physics displays the fundamental particles of matter and the force that govern their interactions [4].

$$\mathcal{L} = \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} A^C_\mu - m_q \delta_{ab} \right) \psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu} \tag{1}$$

where the field strength tensor is given by:

$$F^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu - g_s f_{ABC} A^B_\mu A^C_\nu \tag{2}$$

and the single pieces mean

- $\psi_{q,a(b)} =$ quark field spinors of colour index $a(b) = 1, 2, 3$ with $N_c = 3$
- $m_q =$ quark mass
- $\gamma^\mu =$ Dirac matrices
- $A^C_\mu =$ gluon fields of index $C = 1, \ldots, N_c^2 - 1 = 8$
- $t^C = \frac{1}{2}\lambda^C$
- $\lambda^C =$ Gell-Mann matrices
- $f_{ABC} =$ structure constants of QCD

- $g_s =$ QCD coupling, with $g_s^2 = 4\pi\alpha_s$
- $\alpha_s =$ the strong coupling

The commutator of SU(3) generators satisfies:

$$[t^A, t^B] = if_{ABC}t^C$$

We can see the Lagrangian is composed by the sum of four terms:

1. The free propagation of a Dirac spinor;
2. The quark gluon interaction term;
3. The mass term;
4. The gluon-gluon interaction term; this last term comes from the fact that QCD is a non Abelian theory.

1.1.3 Confinement and asymptotic freedom

The strong coupling is one of the fundamental parameters of the Standard Model and is defined in relation to the g_s, the coupling constant that appears in the QCD Lagrangian at the vertex interactions:

$$\alpha_s = g_s^2/4\pi \tag{3}$$

In the QCD sector of the SM, once the quark masses are taken as input parameters, the strong coupling constant α_s governs the strength of the strong interaction. Its numerical value is not predicted by the SM and must therefore be determined experimentally. In this respect, α_s is analogous to other free parameters of the theory, such as the Yukawa couplings, whose values determine the fermion masses after electroweak symmetry breaking. Once α_s is fixed at a given energy scale, however, QCD predicts its evolution with the energy scale through the renormalization group equations. An important feature of the strong coupling is its energy dependence, as it can be seen from Figure 3.
The strength of the interaction varies with the energy scale, meaning α_s is not constant but "runs" with the energy. This running of the coupling constant is a direct consequence of the structure of QCD in the vacuum and is influenced by the dynamics of gluons and quarks. The behaviour of the α_s can be derived from renormalization procedures [5], obtaining

$$\alpha_s(\mu^2) = \frac{1}{\frac{(33-2n_f)}{12\pi}\ln\left(\frac{\mu^2}{\Lambda^2_{QCD}}\right)} \tag{4}$$

where μ is the renormalization scale. Or, in term of the four-momentum transfer:

$$\alpha_s(Q^2) = \frac{1}{\frac{(33-2n_f)}{12\pi}\ln\left(\frac{Q^2}{\Lambda^2_{QCD}}\right)} \tag{5}$$

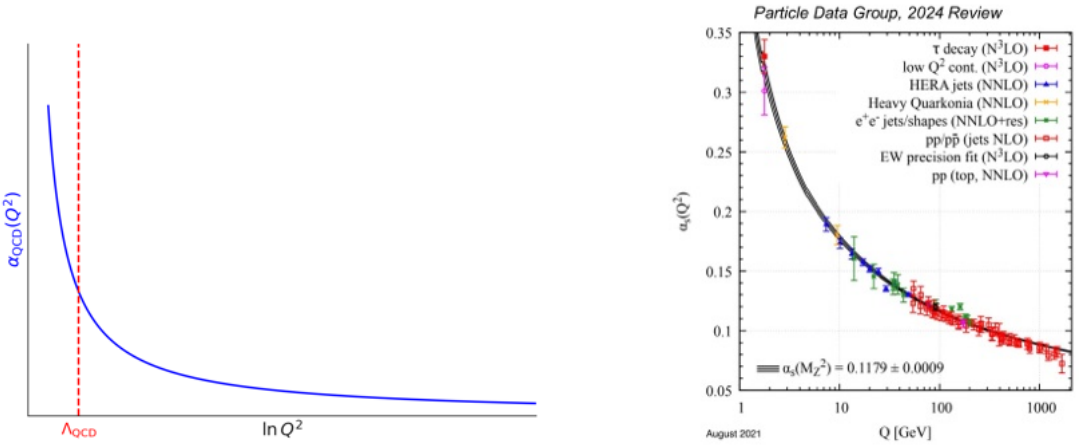


Figure 3: On the left, a schematic draw of the behaviour of α_s. Summary of determinations of α_s as a function of the energy scale Q compared to the running of the coupling computed at five loops taking as an input the current PDG average, $\alpha_s(m_z^2) = 0.11800.0009$. [4].

n_f is the number of active flavours. Λ_{QCD} is the threshold value below which the theory can no longer be treated perturbatively; its exact value depends on the order of the perturbation theory, the renormalization scheme and the number of active flavour n_f at the scale where α_s is probed, but its range is between $210\,\mathrm{MeV}$ and $300\,\mathrm{MeV}$ [4].

Studying the equation 5, one can observe two behaviours:

- For low energy four-momentum transfer, or large distances, the coupling becomes stronger; we enter the regime of non-perturbative QCD and arises the phenomena of **confinement**, preventing quarks and gluons from existing as free particles and restricting them to bound, colour-neutral hadronic states. The topic will be more precisely discussed in the Section 1.1.4

- For high energy four-momentum transfer, or short distances, the coupling becomes asymptotically smaller; we enter the regime of perturbative QCD and the phenomena of **asymptotic freedom** arises, where quarks can be described as quasi-free particles.

1.1.4 Confinement

Confinement explains the phenomenon where quarks and gluons are never observed as free particles but are always bound inside hadrons. This fundamental property dictates the behaviour of strongly interacting matter and is a direct consequence of the non-Abelian nature of QCD. Quark confinement can be understood through the concept of flux tubes. When a quark-antiquark pair is confined inside a hadron, the strong interaction is effectively concentrated within a flux tube, leading to a characteristic potential. A widely used phenomenological model to describe this behaviour is the Cornell potential [6], given by:

$$V(r) = -\frac{a}{r} + \sigma r \tag{6}$$

here the first term represents a short-distance Coulomb-like attraction due to one-gluon exchange, and the second term describes the long-range linear confinement potential arising from the formation of a flux tube between the quark and antiquark. This behaviour is illustrated in Figure 4, where at short distances, the Coulombian term dominates and at long distances the linear term dominates.

$$V(r) = -\frac{a}{r} + \sigma r$$

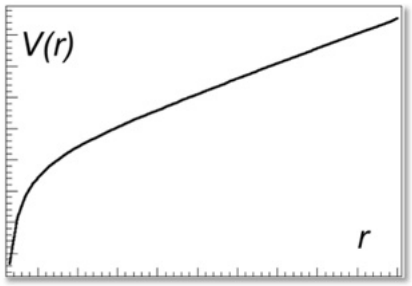


Figure 4: Schematic trend of the Cornell potential [5]

If an external force attempts to separate the quark-antiquark pair, the energy stored in the flux tube increases. However, instead of allowing infinite separation, the system undergoes **string breaking**: once the energy in the flux tube reaches a critical threshold, a new quark-antiquark pair is created from the vacuum, leading to the formation of new hadrons rather than isolated quarks. This process is schematically illustrated in Figure 5

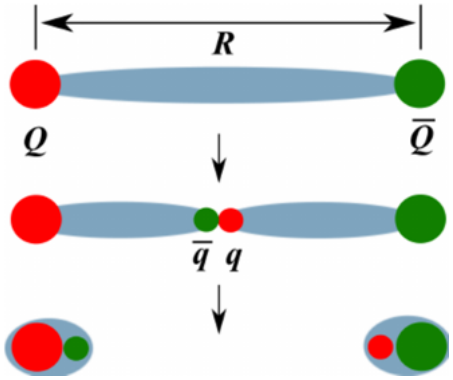


Figure 5: The conventional string breaking: the QCD string spanned between the static quark Q and the static antiquark $\bar{Q}$ breaks due to light $\bar{q}q$ pair creation [7]

This mechanism explains why isolated quarks are never observed in nature. Instead, they always form colourless bound states due to the increasing confinement potential. This process is crucial in high-energy particle collisions, such as those studied in experiments at the Large Hadron Collider (LHC) and at the future Electron Ion Collider (EIC), where quarks and gluons produced in high-energy interactions quickly hadronize into observable particles. The study of hadronization and confinement remains an active area of research, critically modelling a wide range of phenomena like heavy-ion collisions and the physics of neutron stars.

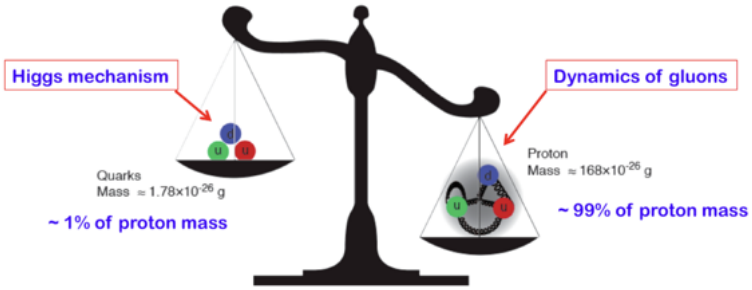


Figure 6: Illustrative representation of the different contributions to the proton mass, emphasizing that only a small fraction originates from the quark masses, while the dominant contribution emerges from QCD dynamics.

1.1.5 Origin of nucleon mass and spin

A direct consequence of the strongly interacting nature of QCD is that the properties of a nucleon cannot be understood simply as the sum of the properties of its constituent valence quarks. This is particularly evident for its mass. Quark masses arise after electroweak symmetry breaking through their Yukawa interactions with the Higgs field, although the corresponding Yukawa couplings, and therefore the numerical values of the quark masses, are free parameters of the Standard Model. Nevertheless, the sum of the masses of the three valence quarks accounts only for a small fraction of the proton mass. As schematically illustrated in Figure 6, the dominant contribution instead emerges from QCD dynamics, namely from the energy associated with quark and gluon fields and with their strong interactions inside the confined system.

The nucleon mass must therefore be regarded as an emergent property of QCD rather than as the simple sum of the masses of its constituents. In particular, the strong coupling characterising QCD at hadronic energy scales makes the calculation of the nucleon mass an intrinsically non-perturbative problem.

A similarly non-trivial picture arises for the nucleon spin. Nucleons are spin-$1/2$ particles and, in a naive quark-parton description, one might expect their spin to originate predominantly from the intrinsic spins of the three valence quarks. Polarized deep-inelastic scattering measurements demonstrated instead that quark helicity accounts for only a fraction of the total nucleon spin, giving rise to the so-called *proton spin puzzle* [3]. This result implies that the angular momentum of the nucleon must also contain substantial contributions associated with gluons and with the orbital motion of its constituents.

A precise decomposition of the nucleon spin requires some care because QCD is a gauge theory. Following the gauge-invariant formulation introduced by Ji [8], the total angular momentum operator can first be separated into quark and gluon contributions,

$$\mathbf{J}_{\mathrm{QCD}} = \mathbf{J}_q + \mathbf{J}_g. \tag{7}$$

The corresponding operators can be written as

$$\mathbf{J}_q = \int d^3x\, \psi^\dagger \left[\frac{\mathbf{\Sigma}}{2} + \mathbf{x} \times (-i\mathbf{D}) \right] \psi, \tag{8}$$

and

$$\mathbf{J}_g = \int d^3x\, \mathbf{x} \times (\mathbf{E} \times \mathbf{B})\,, \tag{9}$$

where $\mathbf{D}$ is the gauge-covariant derivative, while $\mathbf{E}$ and $\mathbf{B}$ denote the chromoelectric and chromomagnetic fields, respectively [8]. In this decomposition, both $\mathbf{J}_q$ and $\mathbf{J}_g$ are gauge invariant.

For a longitudinally polarized nucleon, this leads to the spin sum rule

$$\frac{1}{2} = J_q + J_g = \frac{1}{2}\Delta\Sigma + L_q + J_g, \tag{10}$$

where $\Delta\Sigma/2$ represents the contribution from quark helicity, L_q the gauge-invariant quark orbital angular momentum, and J_g the total gluon angular momentum [8]. The quark contribution can therefore be separated in a gauge-invariant way as

$$J_q = \frac{1}{2}\Delta\Sigma + L_q. \tag{11}$$

An important feature of this decomposition is that the total gluon contribution J_g does not, in general, admit a further gauge-invariant separation into gluon helicity and gluon orbital angular momentum. A decomposition of the form

$$J_g = \Delta G + L_g \tag{12}$$

can nevertheless be given a useful partonic interpretation in specific frameworks, in particular in the light-cone gauge and in the infinite-momentum frame [8]. Thus, while the schematic representation in Figure 7 is useful for illustrating the possible physical sources of nucleon spin, the separation of the gluon contribution into intrinsic and orbital components requires a more careful theoretical definition.

A particularly important aspect of the formulation is that the total angular momentum carried by quarks and gluons can be related to measurable quantities. Introducing the form factors of the quark and gluon energy-momentum tensors, their total angular-momentum contributions are given by

$$J_{q,g} = \frac{1}{2}\left[A_{q,g}(0) + B_{q,g}(0)\right], \tag{13}$$

where $A_{q,g}$ and $B_{q,g}$ are the corresponding energy–momentum-tensor form factors [8]. In particular, these quantities can be accessed through deeply virtual Compton scattering and are closely related to generalized parton distributions. This connection provides an important experimental route towards determining the orbital contribution of quarks to the nucleon spin.

The nucleon mass and spin therefore provide two complementary manifestations of the rich internal dynamics generated by QCD. Although the fundamental degrees of freedom are quarks and gluons, the observable properties of hadrons emerge from their collective interactions. Their quantitative description consequently requires perturbative methods where the strong coupling is sufficiently small, together with non-perturbative approaches in the low-energy regime relevant to hadron structure.

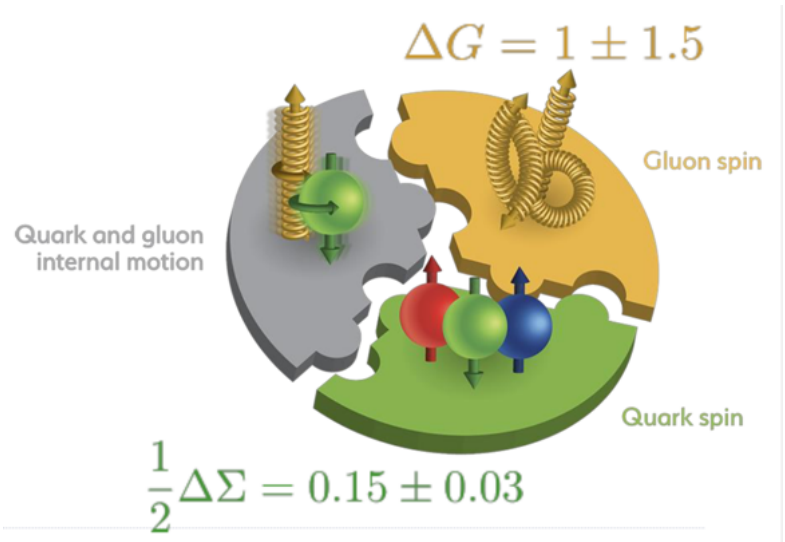


Figure 7: Schematic representation of the different contributions to the nucleon spin, including quark and gluon helicities and orbital angular momentum. The precise separation of these terms depends on the adopted gauge-invariant decomposition.

1.1.6 Non-perturbative versus perturbative QCD

The discussion of nucleon mass and spin highlights a general feature of QCD: although its fundamental Lagrangian is known, obtaining quantitative predictions from it strongly depends on the energy scale of the process under consideration. As discussed previously, this originates from the running of the strong coupling constant α_s. At sufficiently large momentum transfer Q, corresponding to short distances, α_s becomes small as a consequence of asymptotic freedom. In this regime, QCD can be treated using perturbation theory and is commonly referred to as *perturbative QCD* (pQCD). Conversely, when the characteristic energy scale approaches Λ_{QCD}, the coupling becomes large and an expansion in powers of α_s is no longer feasible. This defines the *non-perturbative* regime of QCD, where phenomena such as confinement, hadron formation and the emergence of hadronic properties become dominant.

In perturbative QCD, physical observables are calculated as expansions in powers of the strong coupling. For a generic partonic cross section, one can schematically write

$$d\hat{\sigma} = d\hat{\sigma}^{(0)} + \alpha_s d\hat{\sigma}^{(1)} + \alpha_s^2 d\hat{\sigma}^{(2)} + \alpha_s^3 d\hat{\sigma}^{(3)} + \cdots , \tag{14}$$

where the successive terms correspond to increasingly higher orders in the perturbative expansion. The first contribution is referred to as *Leading Order* (LO), while the following terms are denoted as *Next-to-Leading Order* (NLO), *Next-to-Next-to-Leading Order* (NNLO), and so on. Higher-order terms account for additional real and virtual parton emissions and progressively improve the theoretical description of the process. When α_s is sufficiently small, these corrections generally become smaller with increasing perturbative order, allowing a controlled approximation of the physical observable [9].

The natural domain of pQCD is therefore represented by *hard processes*, namely reactions characterised by a momentum transfer much larger than the QCD confinement scale. Important examples include the Drell–Yan process, production of high-transverse-momentum jets and photons, heavy-quark production and, of particular relevance in the present context, Deep Inelastic Scattering (DIS). In such processes, the short-distance interaction occurs on a scale where quarks and gluons can be treated as the appropriate degrees of freedom and the corresponding partonic subprocess can be calculated perturbatively. This is precisely the regime exploited in high-energy lepton-hadron collisions to probe the internal structure of nucleons. However, the particles entering and leaving an experiment are not free quarks and gluons, but colour-neutral hadrons. Therefore, even when the elementary hard interaction can be described perturbatively, the complete process necessarily contains physics associated with the non-perturbative regime. The internal structure of the initial hadron, as well as the formation of hadrons from final-state partons, involves long-distance QCD dynamics that cannot in general be calculated through a finite perturbative expansion. This represents the fundamental connection between the partonic description provided by QCD and the hadronic states that are prepared and detected experimentally.

A central tool for studying QCD in this strongly coupled regime is *lattice QCD*. In this approach, originally developed to investigate confinement, continuous space-time is replaced by a discrete four-dimensional lattice with spacing a. Quark fields are defined on lattice sites, while the gluon degrees of freedom are associated with the links connecting neighbouring sites. The QCD path integral can then be evaluated numerically, typically using Monte Carlo methods to sample the most relevant field configurations [10]. Lattice QCD therefore provides a first-principles method for extracting non-perturbative predictions directly from the QCD Lagrangian, although at the price of a substantial computational effort. One of the major successes of lattice QCD is the calculation of static hadronic properties. In particular, the experimentally observed hadron mass spectrum can be reproduced with high accuracy, providing an important validation of QCD in the low-energy regime. This is especially significant in view of the discussion of the nucleon mass given above: the agreement between lattice calculations and measured hadron masses demonstrates quantitatively that QCD dynamics can generate the observed masses of hadrons even though the bare masses of the light quarks represent only a small contribution. Lattice calculations can also be used to study the thermodynamic properties of strongly interacting matter and the transition between confined hadronic matter and the deconfined quark-gluon plasma. Perturbative and non-perturbative QCD should therefore not be regarded as two independent descriptions, but rather as complementary tools applicable to different aspects of the same theory. In realistic high-energy processes, the short distance interaction may be perturbatively calculable, while the long distance structure of the incoming hadron and the transition from partons to observable hadrons remain intrinsically non-perturbative. The separation of these different scales is formalised through *factorisation*, which allows a measurable cross section to be expressed in terms of a perturbatively calculable hard component and universal non-perturbative quantities describing hadron structure or hadron formation. The factorised structure of a semi-inclusive cross section can be schematically written as [9]:

$$d\sigma = \sum_{i,j} f_{i/H}(x) \otimes d\hat{\sigma}_{i\to j}(Q) \otimes D_{h/j}(z) \tag{15}$$

Where $f_{i/H}$ is the PDF describing the non-perturbative structure of the incoming hadron, $d\hat{\sigma}$ is the short-distance perturbative hard interaction, and $D_{h/j}$ is the fragmentation function describing the non-perturbative transition from the outgoing parton to the observed hadron. These quantities will be explained in the following Sections.

1.2 Proton structure

In order to probe the nucleons structure, high energy **lepton-hadron inelastic scattering**, where the hadron breaks up, are performed.

1.2.1 DIS kinematic variables

For $e + p/A \rightarrow e' + X$ inelastic scattering, shown in Figure 8, the hadronic final state resulting from the break-up of the hadron usually consist of many particles.

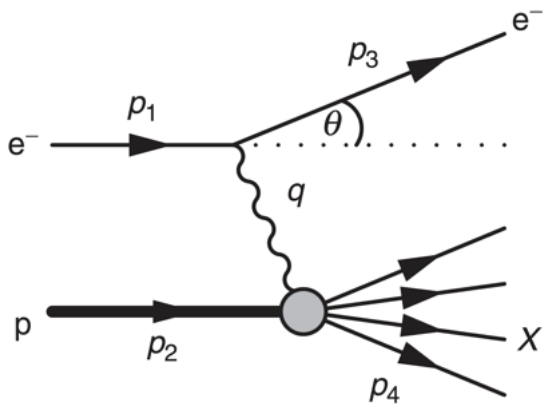


Figure 8: Example of lepton-hadron inelastic scattering with an electron and a proton [5].

The invariant mass of this hadronic system, denoted W, depends on the four-momentum of the virtual photon, $W^2 = p_4^2 = (p_2 + q)^2$, and therefore can take a range of values. Compared to the elastic electron-proton scattering process, where the invariant mass of the final state is always the mass of the proton, this additional degree of freedom in the inelastic scattering process means that the event kinematics must be specified by two quantities. The two kinematic variables used to describe inelastic scattering are usually chosen from the Lorentz invariant quantities W, x, y, ν and Q^2, defined below. Figure 8 will be taken as a reference for the 4-momentum quantities.

Virtuality of the photon The virtuality of the exchanged gauge boson is defined as the negative four momentum squared of the virtual photon, $Q^2 = -q^2$. When written in terms of the four-momenta of the initial and final state electrons,

$$Q^2 = -(p_1 - p_3)^2 = -2m_e^2 + 2p_1 \cdot p_3 = -2m_e^2 + 2E_1E_3 - 2p_1p_3cos\theta \tag{16}$$

In inelastic scattering, the energies are sufficiently high that the electron mass can be neglected and therefore, to a very good approximation

$$Q^2 \approx 2E_1E_3(1 - cos\theta) = 4E_1E_3sin^2\frac{\theta}{2} \tag{17}$$

implying that Q^2 is always positive.

Bjorken scaling variable The Lorentz-invariant dimensionless quantity

$$x \equiv \frac{Q^2}{2p_2q} \tag{18}$$

is a very important kinematic variable in the discussion of the quark model of deep inelastic scattering, since it can be interpreted as the fraction of hadron momentum carried by the interacting parton. The range of possible values of x can be found by writing the four-momentum of the hadronic system in terms of that of the virtual photon. One can find out that $0 \leq x \leq 1$. The variable x can also be seen as the "elasticity" of the scattering process, since $x = 1$ corresponds to an elastic scattering.

Inelasticity A second dimensionless Lorentz-invariant quantity, the inelasticity y, is defined as

$$y \equiv \frac{p_2 \cdot q}{p_2 \cdot p_1} \tag{19}$$

It can be identified as the fractional energy lost by the electron in the scattering process in the frame where the hadron is initially at rest. In this frame, the energy of the final-state hadronic system in always larger than the energy of the initial-state proton, $E_4 \geq m_h$, which implies the electron must lose energy. Consequently, y is constrained to be in the range $0 \leq y \leq 1$.

ν energy variable Sometimes it is more convenient to work in terms of energies, rather than the fractional energy loss described by y. In this case the related quantity

$$\nu \equiv \frac{p_2 \cdot q}{m_p} \tag{20}$$

is often used. In the frame where the inital-state proton is at rest, $\nu = E_1 - E3$, is simply the energy lost by the electron.

1.2.2 Structure Functions

Structure functions are introduced to parametrize the structure of the target as "seen" by the virtual photon and are defined through the lepton scattering cross section. Since the independent kinematic variables are two, as explained in Sec.1.2.1, the cross section is double differential. It can be shown that the most general (parity conserving) Lorentz-invariant expression for

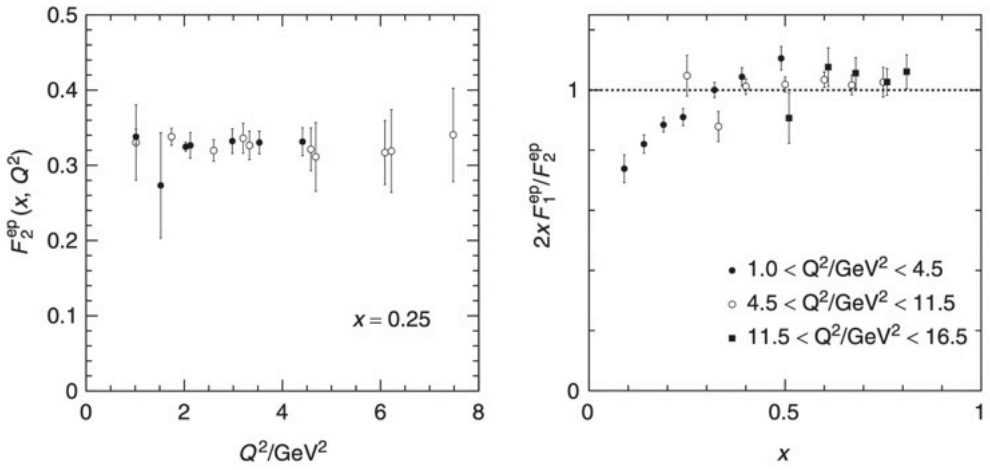


Figure 9: Early structure function measurements from fixed-target electron-proton inelastic scattering at SLAC. Left: measurements of $F_2(x, Q^2)$ showing Bjorken scaling. Right:measurements of $2xF_1/F_2$ showing the Callan-Gross relation. Image from [5]. Adapted from [11] and [12].

the cross section for $ep \to eX$ inelastic scattering, mediated by the exchange of a single virtual photon, is [5]:

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^2}[(1 - y - \frac{m_p^2 y^2}{Q^2})\frac{F_2(x, Q^2)}{x} + y^2 F_1(x, Q^2)] \tag{21}$$

Where $F_1(x, Q^2)$ and $F_2(x, Q^2)$ emerge from the coupling between the quarks electric charge and the photon; they describe the momentum distribution of the quarks within the proton. Considering data obtained from the SLAC experiment, with electrons of energies between $5\,\mathrm{GeV}$ and $20\,\mathrm{GeV}$, two important properties of the structure functions emerged:

- Both $F_1(x, Q^2)$ and $F_2(x, Q^2)$ are almost independent of Q^2, allowing the structure functions to be written as $F_1(x)$ and $F_2(x)$. This observation is called **Bjorken scaling**. This result is strongly suggestive of scattering from point-like constituents within the proton [5];

- In the deep inelastic scattering regime, the structure functions $F_1(x)$ and $F_2(x)$ are not independent, but they satisfy the *Callan-Gross relation* $F_2(x) = 2xF_1(x)$. This observation can be explained by assuming that the underlying process is electron-proton inelastic scattering is the *elastic* scattering of electrons from point-like spin-half constituent particles within the proton, namely the quarks. In this case the electric and magnetic contributions to the scattering process are related by the fixed magnetic moment of a Dirac particle.

1.2.3 The quark parton model and structure functions

Following the seminal work of Gell-Mann and Zweig on the quark model and of Bjorken on the scaling behaviour of deep-inelastic scattering, Feynman formulated the parton model, describing the proton in terms of point-like constituents, which he named *partons* [13]. The main assumptions regarding ep DIS in the quark-parton model are that the basic interaction is the elastic scattering from a quark within the proton, the parton is considered as quasi-free, the momentum of the proton is sufficiently high to neglect the quark mass and to neglect any momentum transverse to the direction of the proton (infinite momentum frame). In this framework, the Bjorken-x can be identified as the fraction of the proton momentum carried by the struck quark [5]. The quarks inside the proton will interact with each other through the exchange of gluons. The dynamics of this interacting system will result in a distribution of quark momenta within the proton. To account for this distribution, we introduce **parton distribution functions** $f_i(x)$ such that $f_i(x)$ is the number of quarks of type i within a proton with momenta between x and $x + dx$. In practice, the functional forms of the PDFs depend on the detailed dynamics of the proton; they are not a priori known and have to be obtained from experiments. Correlating Eq.21 with the expression of the electron-proton cross section, summed over all types of quarks, we can relate the measured structure functions to the underlying quark distributions:

$$F_2(x) = \sum_i e_i^2 x f_i(x) \qquad F_1(x) = \frac{1}{2x} F_2(x) \tag{22}$$

Static quark model In the static quark model, one assumes that the proton is composed only by 2 up quarks and 1 down quark. This enters in contrast with the measurements obtained regarding the structure functions. Defining $f_u = \int_0^1 [xu(x) + x\bar{u}(x)]dx$ the fraction of the proton momentum carried by the up quark, and similarly for the down quark f_d, the measurements obtained during the SLAC experiment [14] are:

$$\int F_2^{ep}(x)dx \approx 0.18 \qquad \int F_2^{en}(x)dx \approx 0.12 \tag{23}$$

and therefore

$$f_u \approx 0.36 \quad f_d \approx 0.18 \tag{24}$$

So, in the proton, the up quarks carry twice the momentum of the down quarks, as expected; but quarks carry just over 50% of the total proton momentum. The rest is carried by gluons. Because the gluons are electrically neutral, they do not contribute to the QED process of electron-proton or electron-neutron deep inelastic scattering.

Dynamic quark model The picture of a proton as a bound state consisting of three "valence" quarks is overly simplistic. The proton also contains a sea of virtual gluons that give rise to an antiquark component through $g \to q\bar{q}$ pair production. To reflect these two distinct components, the up/down-quark PDF can be split into the contribution from the two valence quarks, written as $u_V(x)/d_V(x)$, and a contribution from the sea quarks $u_S(x)/d_S(x)$. In the case of the

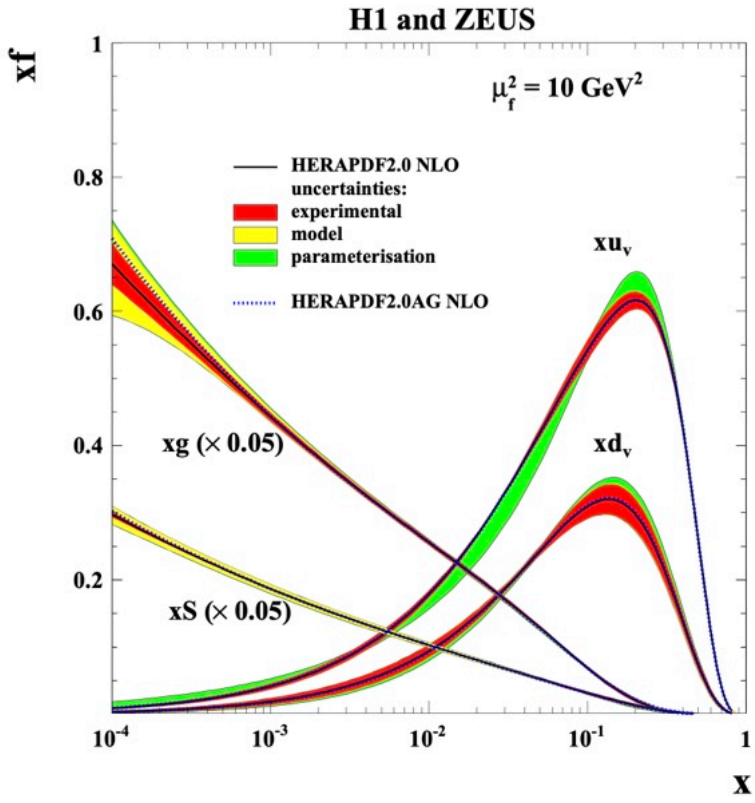


Figure 10: Proton parton distribution functions obtained from the combined H1 and ZEUS deep-inelastic scattering data in the HERAPDF2.0 NLO analysis at $Q^2 = 10\ \mathrm{GeV}^2$. The distributions xu_v, xd_v, xS and xg are shown as functions of Bjorken-x; the sea and gluon distributions are scaled down by a factor of 20 for visualization. Adapted from [15].

antiquarks PDFs, there's only the sea quark contributions. The parton distributions are determined experimentally through global QCD analyses of scattering data. A particularly important contribution was provided by the HERA collider, where the H1 and ZEUS experiments measured inclusive $e^{\pm}p$ deep-inelastic scattering over a broad kinematic range. The combined HERA data were used to obtain the HERAPDF2.0 set of proton PDFs [15]. The resulting distributions, shown in Figure 10, provide a direct illustration of the changing partonic composition of the proton as a function of Bjorken-x. At large x the valence-quark distributions are important, whereas towards small x the sea-quark and gluon components become increasingly prominent. In particular, the gluon distribution is constrained indirectly by the measured Q^2 dependence of the inclusive DIS cross section, through DGLAP evolution, and by the longitudinal structure function F_L [15].

Another interesting effect, not considered in the non-interacting quark parton model, emerged from the HERA data [16] (See Figure 11). Conducting studies with a wider energy range ($Q^2 = [10^{-1}, 10^5]GeV$) it was shown that F_2 does depend on Q^2 (scale breaking effect) and in particular F_2 increases strongly with Q^2 at low Bjorken-x (sea quark region) and F_2 decreases with Q^2

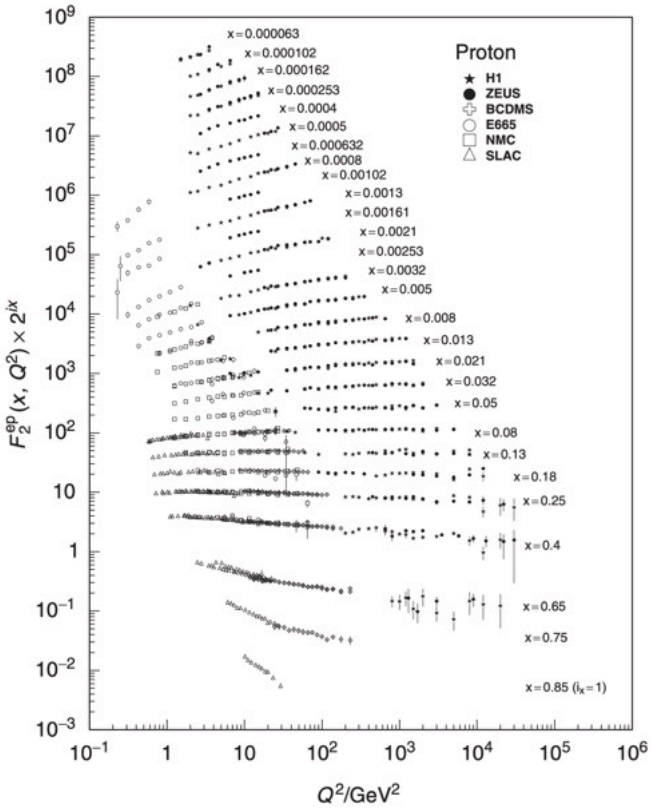


Figure 11: Measurements of $F_2^{ep}(x, Q^2)$ at HERA. Results from both the H1 and ZEUS experiments are shown. Image from [5].

for high Bjorken-x (valence quark region). This effect is called **Bjorken-x scaling violation** and is due to QCD effect not considered by QPM. Observation of Bjorken scaling violation at HERA was the evidence for QCD effects whitin the nucleon, so an improved QCD parton model was necessary.

A key observation is the role of the energy scale in distinguishing perturbative from non-perturbative QCD. Virtual quark production inside the proton primarily occurs in the low-momentum fraction regime, where interactions are non-perturbative. Since partons share the total proton momentum, the accessible energy scale remains close to QCD, reinforcing the need for non-perturbative techniques.

1.2.4 Collinear approximation and PDFs

Although the inclusive DIS cross section is invariant under boosts along the collision axis (longitudinal direction), the scattering is best pictured in the infinite momentum frame in which the proton moves very fast and all point-like particles inside it, the partons, move in parallel as an

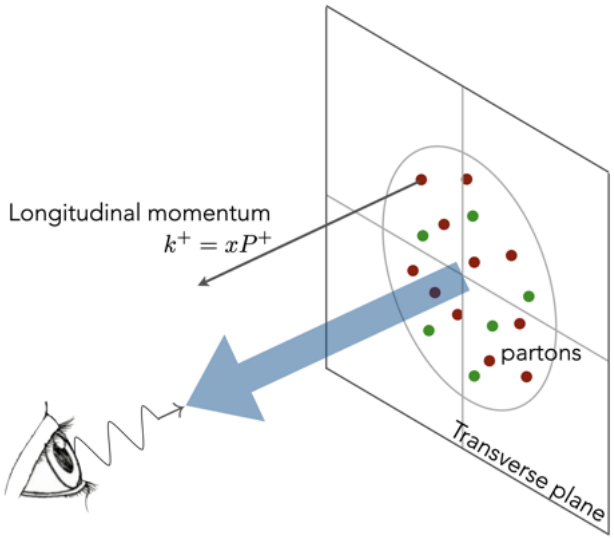


Figure 12: Schematic representing the collinear approximation, used to obtain 1 dimensional PDFs [17].

incoherent beam. Mathematically, the longitudinal momentum of a parton is expressed as:

$$k^+ = xP^+ \tag{25}$$

where k^+ is the parton momentum and P^+ is the total momentum of the nucleon. In this approximation we consider the parton momentum to be only longitudinal, since $k^+ \sim Q$ and the transverse momentum $Q_T \sim 1/R \ll Q$. This approximation, schematically represented in Figure 12, simplifies the theoretical treatment of parton interactions and constitutes the basis of collinear factorization in high-energy QCD calculations.

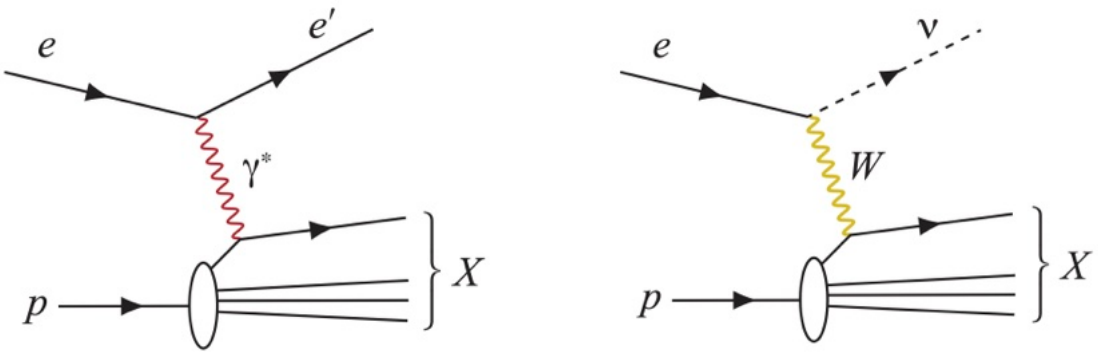
In the QCD collinear factorization framework, the proton structure functions are expressed as a convolution of coefficient functions and parton distribution functions (PDFs):

$$\mathscr{F}_a^k(x, Q^2) = \sum_i C_{a,i}^k(\alpha_s, \frac{Q^2}{\mu_F^2}, m_c, m_b, \frac{x}{z}) \otimes {}_i(\alpha_s, z) \tag{26}$$

where $C_{a,i}^k$ are the *coefficient functions*, which are process-dependent and can be expanded perturbatively in terms of powers of α_s. ${}_i$ are the *parton distribution functions*, which, as said above, account for the non-perturbative effects of the hadron structure and are assumed to be process-independent (universal).

1.2.5 Experimental Access to PDFs

PDFs are accessed experimentally via Inclusive Deep Inelastic Scattering.



(a) (b)

Figure 13: Schematic of the Inclusive Deep Inelastic Scattering process, for (a) Neutral Current and (b) Charged Current categories.

Definition Inclusive Deep Inelastic Scattering (Inclusive DIS) refers to a process in which a lepton scatters off a hadron, and only the scattered lepton or the nuclear recoil are detected in the final state. Inclusive DIS is divided into two categories depending on the exchanged boson:

- **Neutral current Inclusive DIS**: $l + p/A \rightarrow l' + X$, where the electron interacts with the hadron exchanging a virtual photon (Figure 13a);
- **Charged-current Inclusive DIS**: $l+p/A \rightarrow \nu_l+X$, with sufficient high energy the electron-quark interaction is mediated by the exchange of a $W^{\pm}$ or Z^0 gauge boson (Figure 13a);

By definition, inclusive reactions do not place any constraints on the flavour of the interaction that occurs at the boson-quark vertex or the type of particles produced in the final state.

1.2.6 Transverse momentum structure and TMDs

The Standard collinear PDFs used so far describe the distribution of partons in **one dimension** (longitudinal) in momentum space. To obtain a more complete picture of the nucleon, one should also consider the two dimensions transverse to the beam. For this reason **Transverse Momentum Dependent (TMDs)** parton distributions are introduced [17]. TMDs are probability distribution densities to find a parton (quark or gluon) carrying a longitudinal momentum fraction $k^+ = xP^+$ and a transverse momentum k_T inside a colliding nucleon. These functions allow one to describe with better precision many nucleon properties such as spin–momentum correlations, orbital motion, and spatial asymmetries inside the nucleon.

1.2.6.1 Wigner distributions

A conceptually powerful starting point for defining TMDs are the Wigner distributions. They represents quantum phase-space distributions exploiting the correlation between the parton momenta $(x, \vec{k}_T)$ and their spatial coordinates $\vec{b}_T$ (See Figure 14). Through this correlation we can study rotational motion and, in particular, angular momentum as well as other mechanical

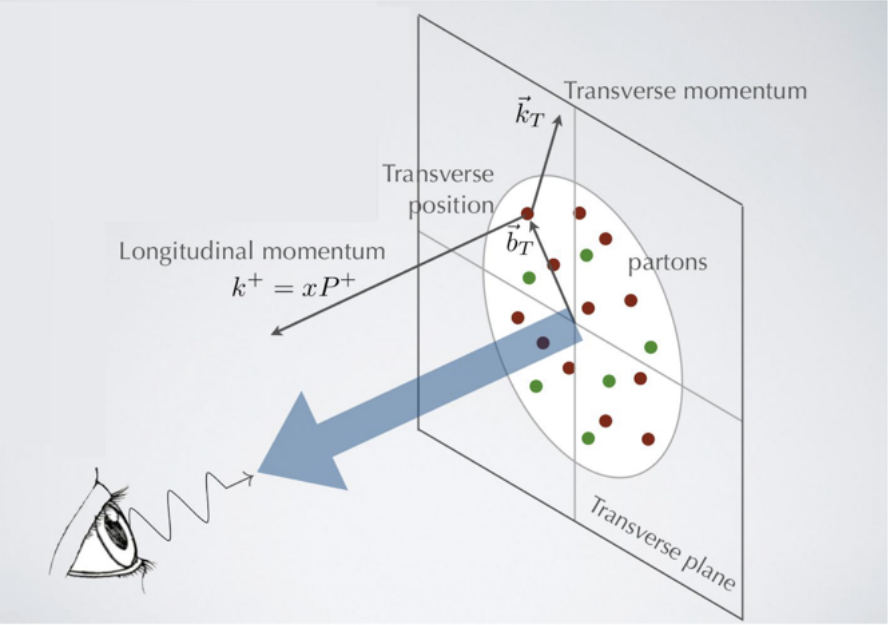


Figure 14: Schematic representing the key variables for Wigner distributions and TMDs: $x, \vec{k}_T, \vec{b}_T$ [17].

properties of the proton [17]. Although it cannot be interpreted as a classical probability density due to the uncertainty principle, it encodes maximal information about the partonic structure of hadrons. TMD parton distribution functions (TMD PDFs) emerge as specific projections of the Wigner distribution, obtained by integrating over transverse spatial coordinates while keeping the transverse momentum dependence. Wigner distributions are also used to obtain other important QCD variables. For example, by taking the integral of the Wigner distribution over transverse momentum, one obtains a so-called impact parameter distribution (IPD) describing the longitudinal momentum fraction x distribution of partons located at a given transverse distance from the hadron center of momentum [17]. Also, performing a Fourier transform with respect to the transverse coordinate variable, one obtains a generalized parton distribution (GPD). GPDs bring the study of momentum-coordinate-space correlations inside the proton within experimental grasp since they are key observables parametrizing the matrix elements of deeply virtual exclusive scattering experiments. A schematic of Wigner distribution employments can be seen in Figure 15. A complete dissertation regarding Wigner distribution will not be performed in this thesis, for reference see [17].

1.2.6.2 TMDs factorization

TMD factorization provides the theoretical framework to describe hard scattering processes in which multiple scales are present, corresponding to the two observed momenta in the process: Q indicating the longitudinal momentum transfer and q_T the transverse momentum, where $Q \gg q_T$; Q is large to ensure the use of perturbative QCD, q_T is much smaller such that it

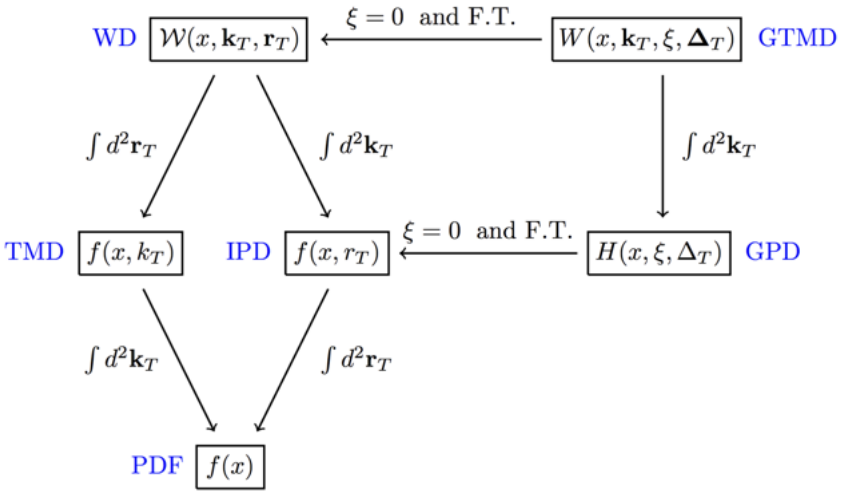


Figure 15: Quantities characterizing the multi-dimensional parton structure of hadrons and the relations between them. **WD**: Wigner Distribution; **TMD**: Trasnverse Momentum Distribution; **IPD**: Impact Parameter Distribution; **GPD**: Generalized Parton Distribution; **GTMD**: Generalized TMD; **GPD**: Generalized Parton Distribution [17]. The variable ξ, known as the *skewness*, quantifies the longitudinal momentum transferred to the nucleon. In a commonly used symmetric parametrization, the active parton enters the hard scattering carrying a longitudinal momentum fraction $x + \xi$ and leaves it with a fraction $x - \xi$.

is sensitive to parton's transverse momentum. Within TMD factorization, cross sections are expressed as convolutions of [9]:

- TMD PDFs: A parton TMD PDF describes the probability to find a parton with longitudinal momentum fraction x and transverse momentum k_T inside the incoming hadron;
- TMD fragmentation functions (TMD FFs): A TMD fragmentation function describes how a parton hadronizes into a final-state hadron with longitudinal momentum fraction z and transverse momentum p_T relative to the fragmenting parton;

TMD parton distribution functions and fragmentation functions are not directly calculable in the non-perturbative regime of QCD, and must therefore be extracted from experimental data or computed using lattice QCD methods in limited cases.

Experimentally, TMDs are accessed through processes that are sensitive to the transverse momentum of final-state particles. Semi-inclusive deep inelastic scattering (SIDIS), Drell–Yan production at low transverse momentum, and electron–positron annihilation into hadron pairs provide complementary channels for isolating different combinations of TMD PDFs and FFs.

Global fits of TMDs rely on combining data from these processes under the assumption of TMD factorization and universality (with known sign changes in certain functions). The transverse momentum distributions observed experimentally are then matched to theoretical pre-

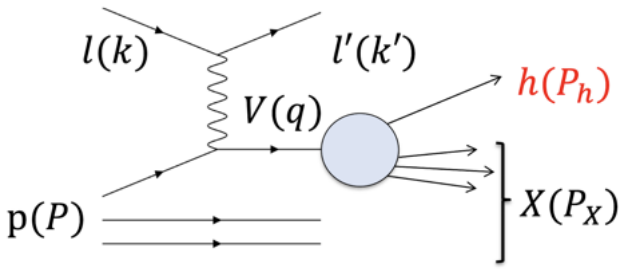


Figure 16: Schematic representing the Semi-Inclusive Deep Inelastic Scattering process [3]. $V(q)$ denotes the virtual gauge boson exchanged between the incoming lepton and the nucleon, carrying four-momentum $q = k - k'$. In electromagnetic SIDIS it corresponds to a virtual photon, γ^*, with virtuality $Q^2 = -q^2$, which probes the internal partonic structure of the proton.

dictions through convolution integrals, allowing the extraction of intrinsic momentum widths, spin correlations, and flavour dependence.

1.2.7 Experimental access to TMDs and FFs

TMD PDFs and FFs are accessed via Semi-Inclusive DIS.

Definition Semi-inclusive DIS, $e + p/A \rightarrow e' + h^{\pm,0} + X$, mediated by a virtual photon, is a process where the incoming lepton (an electron, a positron or muon) with momentum k scatters off a proton with momentum P, both of which can be polarized; this process requires the measurement of at least one identified hadron in coincidence with the scattered lepton. See Figure 16 for a schematic of the process.

When the nucleon target is polarized, SIDIS becomes sensitive to a rich set of spin–momentum correlations encoded in polarized TMD PDFs.

In addition to DIS variables (1.2.1), two new SIDIS variables are introduced to describe the kinematic of the process:

- z_h, is the ratio of the energy of the hadron to that of the virtual photon in the proton rest frame and is defined as $z_h = \frac{P \cdot P_h}{P \cdot q}$;
- $\vec{P}_{hT} = \vec{P}_h - \frac{\vec{P}_h \cdot \vec{q}}{|\vec{q}|}\hat{q}$

In the parton model picture, the lepton-proton SIDIS cross section can be factorized when $Q \gg P_{hT}/z_h$ in term of both PDFs $q(x)$(initial state) and FFs $D_q^h(z)$(hadronic final state) [10]:

$$\sigma^{SIDIS} \sim q(x) \otimes \hat{\sigma}^{\gamma q \rightarrow} \otimes D_q^h(z) \tag{27}$$

Polarization When a spin 1/2 quark is probed in a scattering experiment with polarized nucleons, we can access 4 possible polarization states ($\leftarrow \rightarrow \uparrow \downarrow$). In case of an unpolarized nucleon, we can extract $f_{q/h}$, the unpolarized collinear quark distribution of an unpolarized nucleon. For a *longitudinally polarized nucleon*, we can extract $\Delta f_{q/h}$, the collinear quark helicity distribution of a longitudinally polarized nucleon. For a transversely polarized nucleon, we can extract $\delta f_{q/h}$, the collinear quark transversity distribution of a transversely polarized nucleon.

1.2.8 Spatial imaging and GPDs

To obtain more properties of the nucleon, one could study **Generalized Parton Distributions (GPDs)**. GPDs are probability distribution densities, obtained integrating Wigner distributions over the momentum phase space, describing the 2D spin-dependent coordinate space (transverse) and the 1D longitudinal momentum [17]. Also GPDs are non-perturbative functions. The nucleon has four quark and gluon GPDs: H, E and their polarized version; so polarized lepton and hadron beams are necessary to access all the GPDs. These functions allow one to describe with better precision many nucleon properties such as spin–momentum correlations, orbital motion, and spatial asymmetries inside the nucleon.

1.2.9 Experimental Access to GPDs

GPDs are probed with Exclusive DIS.

Definition In exclusive DIS, $l + p/A \rightarrow l' + p'/A' + \gamma/h^{'0}/VM$, the boson emitted by the lepton probes the internal structure of the nucleon while the target remains intact in the final state. It requires the measurement of all particles in the final state with high precision.

Exclusive processed are mainly used to probe GPDs (Generalized Parton Distributions). Two of the most important exclusive DIS modes are Deeply Virtual Compton Scattering (DVCS) and Exclusive Vector Meson Production (VM)[3].

Deeply Virtual Compton Scattering In DVCS, a virtual photon exchanged between the lepton and nucleon scatters off the nucleon and emerges as a real photon, while the nucleon remains intact : $l + N \rightarrow l' + N' + \gamma$. Since the initial and final nucleon momenta differ, the process is sensitive to off-forward parton correlations encoded in GPDs. DVCS is theoretically very clean because the final state contains a photon rather than a hadron, avoiding uncertainties related to hadronization or meson wave functions. It is therefore considered one of the golden channels for extracting GPDs and studying the three-dimensional structure of the nucleon.

Exclusive Vector Meson Production VM occurs when the energy is sufficient to "convert" the virtual photon into a vector meson, while leaving the target intact. Here the virtual photon first fluctuates into a quark–antiquark pair, which interacts coherently with the nucleon through a colour-singlet exchange (often described effectively as two-gluon exchange at high energy) and finally hadronizes into the observed vector meson. VM production is more sensitive to specific partonic degrees of freedom depending on the meson species. Heavy vector mesons such as J/ψ consent a direct access to the gluon content of the nucleon because charm anti-charm

production is dominated by gluon interactions. Meanwhile light mesons probe different flavour combinations of quarks and gluons.

1.3 Nuclear structure and QCD in nuclei

The study of nuclei at the partonic level represents one of the central components of the Electron-Ion Collider physics program. While a nucleus is conventionally described as a bound system of protons and neutrons, deep-inelastic scattering measurements on nuclear targets have shown that the partonic structure of a bound nucleon differs from that of a free nucleon. If the nucleus behaved as an incoherent superposition of independent nucleons, the parton distributions per nucleon would be expected to coincide with those of free nucleons. Instead, nuclear-to-nucleon DIS ratios exhibit a characteristic dependence on Bjorken-x, including shadowing at low x, antishadowing at intermediate x, and the EMC suppression at larger x [18, 3]. The nuclear environment therefore modifies the momentum distributions of partons, while at sufficiently small values of x the increasing density of gluons may give rise to non-linear QCD dynamics. In addition, nuclei provide a controlled environment in which the propagation, energy loss, and hadronization of coloured partons can be studied. The possibility of comparing electron-proton (ep) and electron-nucleus (eA) collisions over a broad range of collision energies and nuclear species therefore makes the EIC a unique laboratory for investigating the emergence of nuclear phenomena from the underlying degrees of freedom of QCD [3, 18].

1.3.1 Nuclear modifications of parton distributions

Parton distribution functions (PDFs) describe the longitudinal momentum structure of a fast-moving hadron. For a given resolution scale Q^2, they characterize the probability density of finding a quark, antiquark, or gluon carrying a fraction x of the longitudinal momentum of the parent hadron. In the case of a nucleus, the corresponding quantities are referred to as nuclear parton distribution functions (nPDFs). As for free-proton PDFs, nPDFs are assumed to be universal quantities within QCD factorization and are extracted from global analyses of experimental data. Their present determination is, however, considerably less precise than that of free-proton PDFs, mainly because the available nuclear data cover a more restricted region of the (x, Q^2) phase space and are affected by larger experimental uncertainties [3].

A convenient way to quantify nuclear modifications is to compare the parton distribution inside a nucleus with the corresponding distribution in a free proton, after normalizing the nuclear result to the number of nucleons. If a nucleus behaved as an incoherent collection of independent protons and neutrons, such a ratio would be expected to remain close to unity. Experimental measurements instead show a characteristic dependence on x: the distribution is suppressed at small x, enhanced at intermediate x, and again suppressed at larger x. These regions are conventionally identified as *shadowing*, *antishadowing*, and the *EMC region*, respectively [18].

The latter takes its name from the European Muon Collaboration (EMC), which first observed an unexpected modification of the nucleon structure function in deep-inelastic muon scattering on nuclear targets. In particular, the ratio of the structure function per nucleon measured in iron to that measured in deuterium, $F_2^N(\mathrm{Fe})/F_2^N(\mathrm{D})$, was found to decrease significantly below

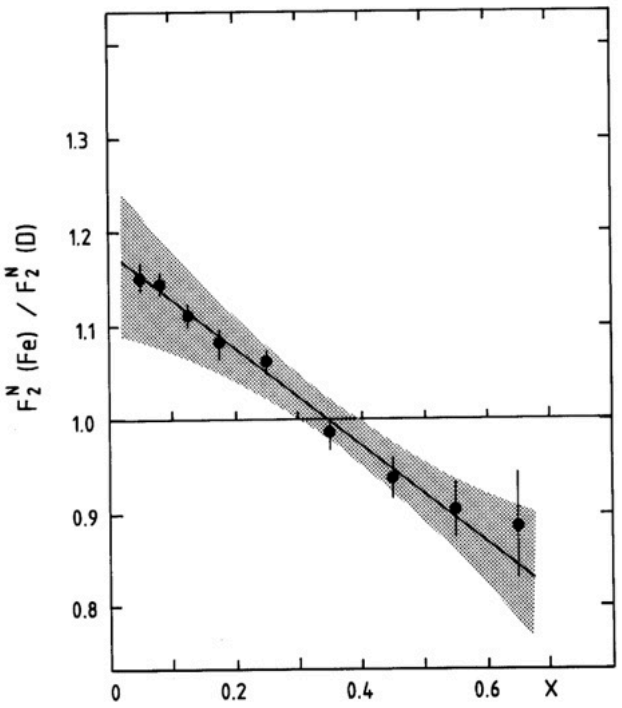


Figure 17: Ratio of the nucleon structure functions measured in iron and deuterium as a function of Bjorken-x. The departure of the ratio from unity, and in particular its decrease in the region $0.3 \lesssim x \lesssim 0.7$, provided the original experimental evidence for what became known as the EMC effect. Adapted from [19].

unity for approximately $0.3 \lesssim x \lesssim 0.7$, as shown in Figure 17. This observation demonstrated that the partonic structure of a bound nucleon cannot be described simply as that of an isolated nucleon embedded in a nucleus [19]. At still larger x, the ratio rises again, with an increasing contribution from the Fermi motion of nucleons inside the nucleus.

The characteristic behaviour of the nuclear modification factor is illustrated in Figure 18. The deviation from unity directly shows that the internal partonic structure of bound nucleons is modified by the surrounding nuclear environment. Consequently, precise measurements of nPDFs provide information on the extent to which a nucleus can be regarded as a simple ensemble of independent nucleons and, more generally, on how nuclear binding emerges from QCD degrees of freedom [18].

Inclusive deep-inelastic scattering constitutes one of the primary tools for constraining these distributions. The inclusive DIS cross section can be written in terms of the structure functions F_2 and F_L. The former is predominantly sensitive to quark and antiquark distributions, while gluons contribute indirectly through the Q^2 evolution of the structure functions. The longitudinal structure function F_L instead receives a more direct contribution from the gluon distribution and is therefore especially important at low x, where the sea-quark and gluon components become increasingly important, as illustrated for the proton PDFs in Figure 10. Its extraction

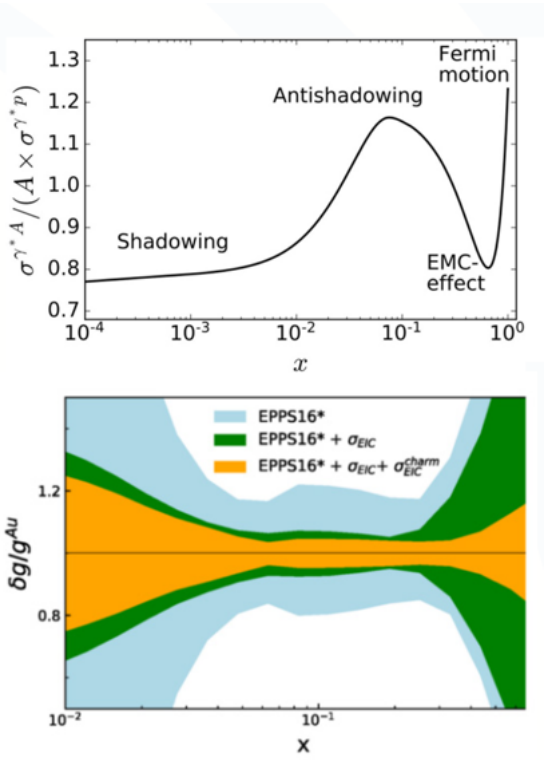


Figure 18: Nuclear modification of parton distributions. The upper panel shows the ratio between the virtual-photon cross section on a nucleus and that on a proton, normalized by the nuclear mass number, highlighting the shadowing, antishadowing, and EMC regions. The lower panel illustrates the expected reduction of the uncertainty on the gluon distribution in a gold nucleus after the inclusion of projected EIC inclusive and charm measurements. Adapted from [18].

requires measurements performed at different centre of mass energies. Additional direct sensitivity to the nuclear gluon distribution is provided by heavy flavour production, since at leading order heavy quark pairs can be generated through photon-gluon fusion. Combining inclusive and charm measurements at the EIC is expected to substantially reduce the present uncertainty on the gluon nPDF over a broad range of x [3].

1.3.2 Shadowing and the small-x region

The low-x region is of particular interest because it corresponds to partons carrying only a small fraction of the longitudinal momentum of the parent nucleon or nucleus. In QCD, the radiation of low-momentum gluons is strongly favoured. A gluon can split into two gluons, which may subsequently undergo further splittings, while gluons may also produce quark-antiquark pairs. As this cascade develops, an increasingly large number of partons populate the low-x region. For this reason, the internal structure of a hadron becomes progressively dominated by gluons

when sufficiently small momentum fractions are probed [18, 3].

In nuclei, the measured parton density per nucleon at low x is found to be smaller than the density that would be obtained by simply combining free-nucleon distributions. This suppression is referred to as *nuclear shadowing*. In terms of the nuclear modification ratio introduced in the previous subsection, shadowing corresponds to a value smaller than unity at low x, as visible in Figure 18. The phenomenon indicates that, at sufficiently high energies, the interaction of the virtual photon with the nucleus cannot always be regarded as an incoherent interaction with an isolated nucleon [18].

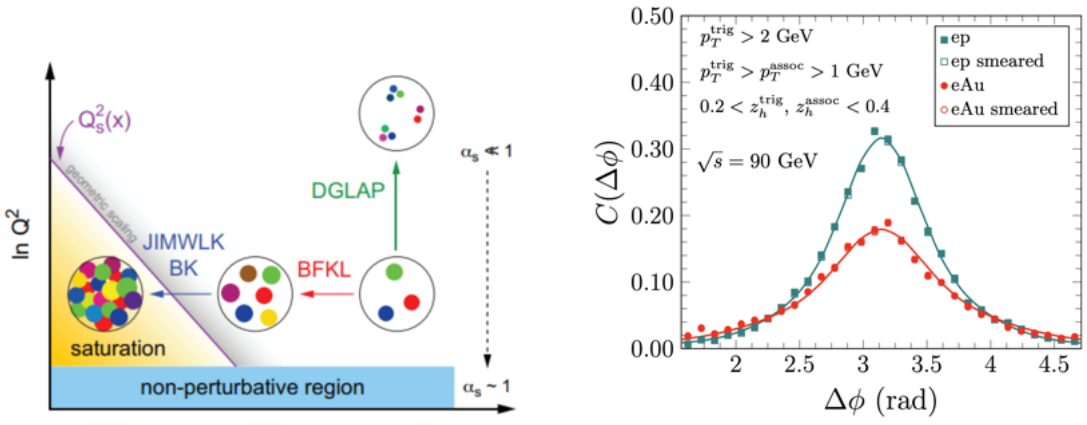
From the experimental point of view, the EIC will extend the nuclear DIS program to values of x significantly smaller than those presently constrained by fixed-target measurements. This is particularly important for the gluon distribution, which remains weakly constrained in nuclei. The large lever arm in Q^2 will allow the gluon content to be investigated through scaling violations, through measurements of F_L, and through heavy-flavour production. According to the projected EIC studies [3] the inclusion of precise inclusive DIS measurements can strongly reduce the uncertainty bands of present nPDF parametrizations, particularly below $x \sim 10^{-2}$, where existing data are scarce or absent [3].

It is important to distinguish nuclear shadowing from gluon saturation. Shadowing refers phenomenologically to the suppression of nuclear parton distributions or nuclear cross sections relative to the corresponding free-nucleon expectation at low x. Saturation, instead, refers to a specific dynamical regime in which the gluon density becomes sufficiently large for non-linear interactions among gluons to become important. The two phenomena occur in overlapping regions of phase space and are therefore closely related experimentally, but they are not equivalent concepts. Measurements of inclusive, semi-inclusive, and diffractive processes at the EIC are expected to provide complementary information needed to distinguish between different theoretical descriptions of the small-x nuclear regime [3].

1.3.3 Increasing gluon density and possible saturation

The build-up of a large gluon population toward small x, qualitatively visible in the proton PDF extraction shown in Figure 10, follows from a distinctive property of the strong interaction: gluons themselves carry colour charge and can therefore interact with one another. A gluon can radiate additional gluons, which can in turn undergo further splittings, while gluons can also generate sea quarks through $g \to q\bar{q}$. Repeated emissions consequently populate progressively smaller momentum fractions as the collision energy increases [15, 18].

Such a growth cannot continue indefinitely. When the density of gluons becomes sufficiently large, the probability for gluon recombination processes also increases. In particular, the process $gg \to g$ counteracts gluon radiation. When the rate of gluon recombination becomes comparable to the rate of gluon production, the growth of the gluon density is expected to slow down, leading to the regime known as *gluon saturation*. The corresponding high-density small-x dynamics are commonly described within the Colour Glass Condensate (CGC), an effective field theory of high-energy QCD designed to describe systems with large gluon occupation numbers [20]. Linear small-x evolution is associated with gluon emission, while non-linear evolution equations, such as the Balitsky-Kovchegov and JIMWLK equations, include gluon recombination effects and describe the transition toward the saturated regime [3].



(a) (b)

Figure 19: Illustration of the gluon saturation regime and of one of its possible experimental signatures. (a) shows the expected transition between the dilute and saturated regions in the (x, Q^2) plane. (b) shows the predicted dihadron azimuthal correlation for electron-proton and electron-+-gold collisions, where the suppression of the away-side peak at $\Delta\phi \simeq \pi$ reflects the larger saturation scale of the nuclear target. Adapted from [18].

A characteristic momentum scale, the saturation scale Q_s, is introduced to distinguish the dilute and dense regimes. Schematically, for $Q \gg Q_s$ the system can be treated as a dilute partonic system, whereas for $Q \leq Q_s$ non-linear interactions become important, as indicated in the Figure 19a. A central prediction is that the saturation scale increases both as x decreases and as the atomic mass number A increases. For a large nucleus, it is expected [3] that approximately

$$Q^2_{s,A} \propto A^{1/3}, \tag{28}$$

reflecting the larger number of nucleons that overlap at a given transverse position. Heavy nuclei therefore provide an effective enhancement of the gluon density and make the onset of non-linear QCD dynamics accessible at larger values of x than in a proton and at lower energies [3].

A convincing identification of saturation requires several independent experimental signatures. One proposed observable is the azimuthal correlation between two hadrons or jets produced approximately back-to-back. In a dilute system, hard scattering gives rise to a pronounced away-side peak near $\Delta\phi = \pi$. In a dense gluonic system, additional transverse momentum acquired from the nuclear gluon field broadens the correlation and can suppress the away-side peak. This behaviour is illustrated in Figure 19b, where the predicted correlation in electron-gold collisions is significantly reduced relative to the electron-proton case. Inclusive structure functions, dijet and dihadron correlations, and diffractive processes therefore form a complementary set of observables for investigating the possible onset of saturation at the EIC [18, 3].

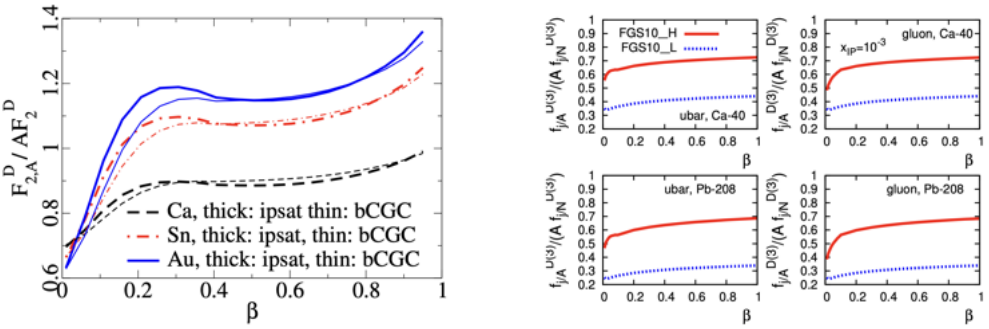


Figure 20: Nuclear modifications in diffractive scattering. The left panel shows ratios of nuclear to proton diffractive structure functions for different nuclei within saturation-based dipole models, while the right panels show corresponding nuclear modifications of diffractive sea-quark and gluon distributions in a leading-twist shadowing description. The different predictions illustrate the strong sensitivity of diffraction to the underlying small-x dynamics. Adapted from [3].

1.3.4 Coherent and diffractive processes

Diffractive scattering provides an especially sensitive probe of the gluonic structure of nuclei. In a diffractive DIS process, the interaction between the virtual photon and the target occurs through an overall colour-neutral exchange. As a consequence, the final state contains a region in rapidity with no produced particles, conventionally referred to as a *rapidity gap*. Diffractive processes are particularly interesting at small x because their cross sections exhibit a strong sensitivity to the gluon density and therefore to the onset of non-linear QCD effects [18, 3].

In nuclear diffraction it is useful to distinguish between coherent and incoherent interactions. In *coherent diffraction*, the nucleus remains intact and the virtual photon effectively probes the nucleus as a whole. The momentum-transfer dependence of the cross section therefore contains information on the global spatial distribution of partons within the nucleus and is particularly sensitive to the nuclear radius and density profile. In *incoherent diffraction*, instead, the nucleus breaks up or one of its constituent nucleons dissociates. Such processes are sensitive to fluctuations in the positions of nucleons and, at a more microscopic level, to fluctuations of the partonic substructure of individual nucleons [3].

The relevance of nuclear diffraction to saturation is enhanced by its stronger dependence on the gluon distribution than that of inclusive observables. In simplified terms, the diffractive cross section is sensitive to approximately the square of the gluonic scattering amplitude. Consequently, modifications associated with a dense gluon field can be amplified. Coherent diffraction constitutes a larger fraction of the diffractive cross section in eA than in ep scattering [3]. In the large-A dipole picture, before reaching the black-disk limit, the coherent diffractive cross section scales parametrically as

$$\sigma_{\mathrm{diff}} \sim A^{4/3}, \tag{29}$$

whereas the total cross section grows approximately as A [3].

Diffractive measurements are also valuable because different theoretical pictures can lead to qualitatively different nuclear modifications. Saturation-based dipole calculations may pre-

dict an enhanced diffractive fraction in nuclei, while leading-twist shadowing approaches can instead lead to a suppression associated with multiple scattering and the survival probability of the rapidity gap. This contrast is illustrated in Figure 20. Precise measurements of coherent and incoherent diffraction, combined with forward detection of the nuclear remnants, can therefore provide both structural information on nuclei and a discriminating test of the dynamics governing the high-density gluon regime [3].

1.3.5 Propagation and hadronization of partons in nuclear matter

Electron–nucleus collisions also provide a controlled environment in which to study the propagation of a high-energy colour charge through cold nuclear matter. In DIS, the exchanged virtual photon interacts with a quark inside one of the nucleons. The struck quark then propagates through the remaining nuclear medium, undergoing interactions with its colour fields and developing a partonic shower before eventually forming colour-neutral hadrons. The nuclear target can therefore be used as a finite QCD medium whose size can be varied by changing the nuclear species [3].

The interaction of a propagating parton with the medium can lead to energy loss and to a broadening of its transverse-momentum distribution. An important quantity used to characterize this process is the transport coefficient $\hat{q}$, defined as the average transverse momentum squared transferred from the medium to the propagating parton per unit path length. Existing determinations from nuclear DIS span a relatively broad range of approximately

$$\hat{q} \simeq 0.02 - 0.14\ \mathrm{GeV}^2/\mathrm{fm}, \tag{30}$$

illustrating the present uncertainty in the transport properties of cold nuclear matter. Precision jet and hadron measurements at the EIC are expected to considerably improve these constraints [3].

A closely related question concerns the space-time development of hadronization. If the formation time of a hadron is shorter than the time required for the propagating system to leave the nucleus, the newly formed hadron may interact with the nuclear medium and can be absorbed or rescattered. This mechanism can result in a suppression of hadron production in eA relative to ep collisions. At the larger centre of mass energies of the EIC, the formation times of light hadrons are generally increased, so that a significant part of their hadronization may occur outside the nucleus. Heavy-flavour hadrons, such as D and B mesons, provide a complementary probe because their characteristic formation times and mass scales differ substantially from those of light hadrons [3].

Identified-hadron and jet measurements provide complementary information on this process. Jets are closely related to the kinematics of the original scattered parton and encode the history of its interactions with the medium through changes in their energy, transverse profile, and substructure. Semi-inclusive hadron production, on the other hand, is directly sensitive to fragmentation. Nuclear fragmentation functions can be introduced to describe the modification of the transformation of a parton into observable hadrons in a nuclear environment. The use of SIDIS is particularly advantageous because it provides a cleaner separation between initial-state modifications, associated with nPDFs, and final state modifications related to propagation and fragmentation than is generally possible in hadron-nucleus or nucleus-nucleus collisions [3].

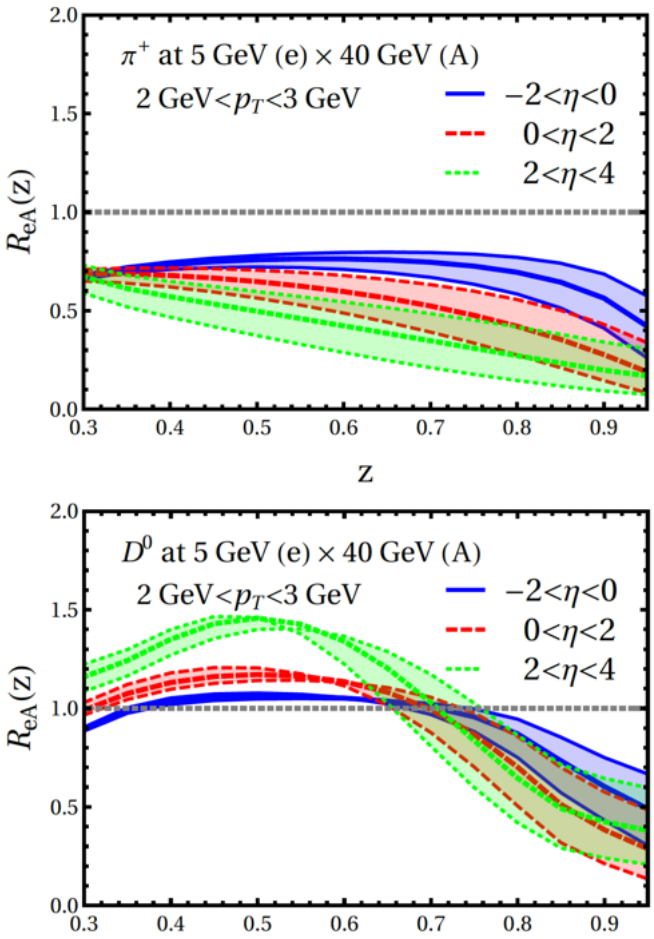


Figure 21: Predicted nuclear modification of hadron production in eA relative to ep scattering as a function of the fragmentation variable z. Light hadrons and heavy-flavour mesons exhibit different nuclear modifications, providing complementary sensitivity to the mechanisms governing parton propagation and hadron formation in nuclear matter. Adapted from [18].

Figure 21 illustrates the expected modification of particle production in nuclear collisions. Light pions and heavy-flavour mesons show substantially different behaviours, which provides the possibility of distinguishing among different descriptions of in-medium hadronization. A systematic program involving different nuclear species, collision energies, hadron flavours, and jet observables will therefore make it possible to investigate how coloured partons evolve into the colour-neutral particles observed experimentally [18, 3].

1.3.6 Comparison of ep and eA collisions

The direct comparison between ep and eA collisions is one of the most powerful elements of the EIC nuclear physics program. Electron-proton scattering provides a reference for the partonic structure and fragmentation processes associated with an approximately free proton. Replacing the proton with a nucleus introduces nuclear binding, coherent interactions among several nucleons, an enhanced gluon density, and a finite volume of cold nuclear matter. Ratios of observables measured in eA and ep collisions can therefore be constructed to isolate the effects induced by the nuclear environment [3, 18].

A generic nuclear modification ratio can be written schematically as

$$R_{eA} = \frac{1}{A}\frac{\mathrm{d}\sigma_{eA}}{\mathrm{d}\sigma_{ep}}, \tag{31}$$

where the precise definition depends on the observable under consideration. In the absence of nuclear effects, the ratio is expected to remain close to unity. Deviations from unity may instead originate from several physically distinct mechanisms. In inclusive DIS they reveal the modification of nuclear PDFs, including shadowing, antishadowing, and the EMC effect. At small x, differences between ep and eA can be amplified by the larger saturation scale of a heavy nucleus. In the final state, modifications of hadron and jet distributions can be related to energy loss, transverse momentum broadening, in-medium parton showers, and hadronization [3].

The comparison shown in Figure 19 provides a representative example of the additional sensitivity obtained from a nuclear target: the away-side dihadron correlation is substantially reduced in electron–gold collisions relative to electron– proton collisions. Similarly, Figure 18 shows how inclusive measurements quantify modifications of the initial-state parton distributions, while Figure 21 illustrates how final-state particle production can be modified by the passage through nuclear matter. The combined interpretation of these measurements is essential, since initial- and final-state nuclear effects may contribute to the same observable.

The possibility of changing the nuclear mass number A is particularly important. Light and heavy nuclei provide targets of different spatial size and gluon density, allowing the dependence of the observed modifications on the amount of nuclear matter to be studied systematically. For gluon saturation, the A dependence provides direct sensitivity to the predicted enhancement of Q_s. For parton propagation, it effectively changes the path length available for interactions with the medium. For nPDF measurements, a broad selection of nuclei makes it possible to determine experimentally how nuclear modifications evolve with A, rather than relying exclusively on assumed parameterizations [3].

For precision comparisons, it is important to collect ep and eA data at matching beam energies whenever possible. This minimizes uncertainties associated with extrapolations between

different kinematic regions and provides a cleaner reference for identifying genuine nuclear effects. The combination of proton and nuclear measurements within the same facility is therefore fundamental: ep collisions establish the elementary QCD baseline, while eA collisions introduce a tunable nuclear environment. By varying x, Q^2, the centre of mass energy, and the nuclear species, the EIC will allow the different manifestations of nuclear QCD to be investigated within a common and experimentally controlled framework [3].

1.4 The EIC physics programme

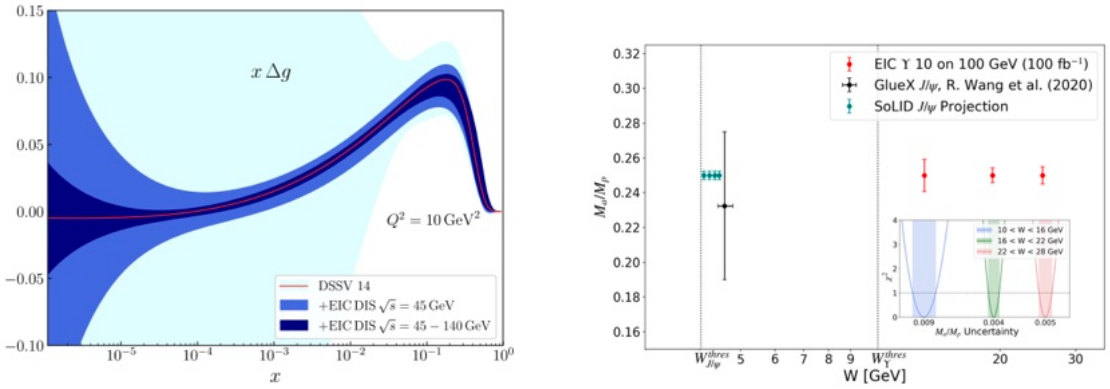
Some of the most relevant open questions to which the EIC will give an answer are [3]:

- How do the nucleonic properties such as mass and spin emerge from partons and their underlying interactions?
- How are partons inside the nucleon distributed in both momentum and position space?
- How do colour-charged quarks and gluons, and jets, interact with a nuclear medium? How do the confined hadronic states emerge from these quarks and gluons? How do the quark-gluon interactions create nuclear binding?
- How does a dense nuclear environment affect the dynamics of quarks and gluons, their correlations, and their interactions? What happens to the gluon density in nuclei? Does it saturate at high energy, giving rise to gluonic matter or a gluonic phase with universal properties in all nuclei and even in nucleons?

1.4.1 Origin of nucleon mass and spin

As discussed in Sec. 1.1, the mass and spin of the nucleon cannot be understood as simple sums of the corresponding properties of its valence quarks. They instead emerge from the collective dynamics of quarks and gluons governed by QCD. One of the main objectives of the EIC physics programme is therefore to move from this qualitative understanding towards a quantitative description of how the different partonic degrees of freedom generate the observed properties of the nucleon [3, 18].

In the case of the nucleon spin, the EIC will provide a substantial extension of the kinematic region explored by previous polarized scattering experiments. Present measurements constrain the quark and gluon helicity distributions only over a limited range in Bjorken-x, leaving the contribution of low-x partons, and in particular low-momentum gluons, comparatively poorly known. The combination of high luminosity, highly polarized electron and proton beams, and a broad centre-of-mass energy range will allow the helicity distributions to be determined over a significantly wider interval in x and Q^2. This will substantially improve the determination of the gluon contribution ΔG and, together with improved knowledge of the quark contribution, reduce the presently available phase space for the orbital angular momentum of quarks and gluons [3, 18]. In this sense, the EIC spin programme is directly connected to the multidimensional study of the nucleon discussed in the following subsection, since information on partonic orbital motion can also be obtained from spatial and momentum correlations.



(a) (b)

Figure 22: Examples of the expected EIC contribution to the understanding of the nucleon spin and mass. (a) Gluon helicity distribution obtained from a global analysis: the broader band represents the uncertainty from existing measurements, while the narrower band illustrates the expected improvement after the inclusion of EIC data. Image from [21] (b) Projected sensitivity to the trace-anomaly contribution to the proton mass through near-threshold quarkonium production. Adapted from [3].

The origin of the nucleon mass represents a complementary manifestation of the same strongly interacting dynamics. As introduced previously, only a small fraction of the proton mass can be associated directly with the Higgs-generated masses of its valence quarks, while the dominant contribution originates from QCD dynamics. The EIC programme aims to investigate this emergence of hadronic mass through observables particularly sensitive to the gluonic structure of the nucleon. Among the proposed measurements, it is identified the exclusive production of heavy quarkonia, such as J/ψ and Υ, close to their production thresholds as a possible source of information on the gluonic contribution associated with the QCD energy–momentum tensor and its trace anomaly [3]. Complementary information will be obtained through the multidimensional imaging programme, which connects the distributions of quarks and gluons not only to their momentum and spatial structure, but also to mechanical properties of the nucleon such as its internal pressure and force distributions [18, 3].

1.4.2 Multidimensional imaging of nucleons

The traditional description of nucleon structure through collinear PDFs provides a one-dimensional picture of the longitudinal momentum carried by quarks and gluons. As discussed in Sec. 1.2, TMDs and GPDs extend this description by introducing information on transverse momentum and transverse spatial coordinates, respectively. The objective of the EIC multidimensional imaging programme is not simply to determine these distributions with greater precision, but to combine complementary measurements into a tomographic description of the nucleon and, where possible, of nuclei [3, 18].

A major advance with respect to existing measurements will arise from the much broader

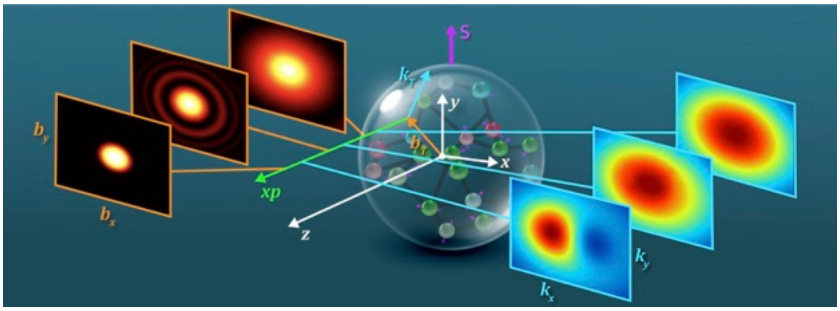


Figure 23: Schematic representation of the multidimensional partonic structure of a polarized proton for different values of the longitudinal momentum fraction x. The images on the left represent transverse spatial distributions, related to GPDs, while those on the right represent transverse-momentum distributions, related to TMDs. Image from [18].

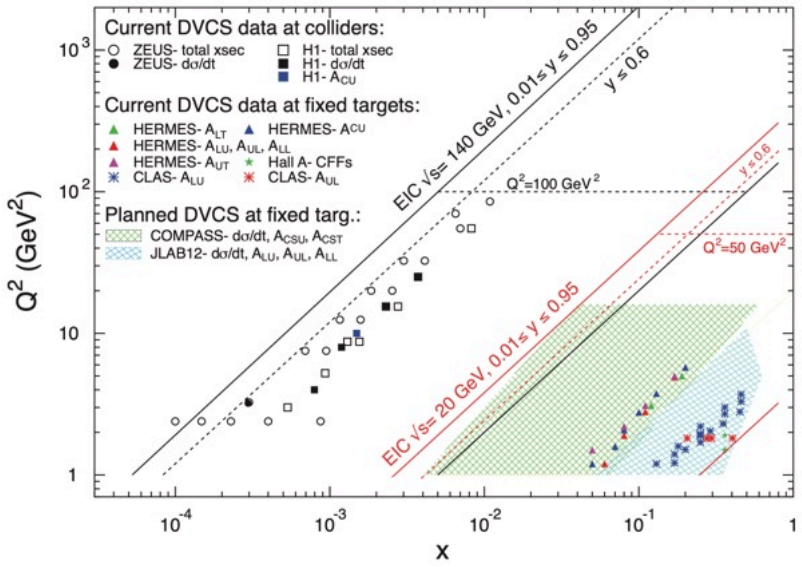


Figure 24: Kinematic coverage in the (x, Q^2) plane expected for deeply virtual Compton scattering at the EIC, compared with existing and planned DVCS measurements at fixed-target and collider experiments. The broad EIC coverage provides access to partonic imaging over a substantially enlarged range of longitudinal momentum fractions and resolution scales. Adapted from [3].

kinematic coverage of the EIC. As illustrated in Figure 24, in the case of DVCS the EIC will extend the accessible (x, Q^2) domain well beyond that covered by existing fixed-target and collider measurements, allowing the spatial structure of the nucleon to be studied over a much wider range of parton momentum fractions and resolution scales [3]. Current information is concentrated mainly in the valence-quark region, whereas sea-quark and gluon distributions, particularly their spin and transverse correlations at small x, remain significantly less constrained. The high centre-of-mass energy and luminosity of the EIC will extend multidimensional measurements towards smaller values of x, while polarized beams will provide sensitivity to correlations between the spin of the nucleon and the motion of its constituents. The resulting programme will therefore probe not only where partons are located and how they move, but also how these distributions depend on parton flavour, polarization, longitudinal momentum fraction, and the resolution scale Q^2 [18].

The spatial and momentum-space approaches provide complementary information. Exclusive processes sensitive to GPDs will enable transverse spatial imaging of quarks and gluons as a function of x, while semi-inclusive processes sensitive to TMDs will map their transverse motion and spin–momentum correlations. In addition, GPDs contain information related to the orbital angular momentum of partons and to mechanical properties such as pressure and shear-force distributions inside the nucleon [3]. The combination of these measurements therefore establishes a direct connection between nucleon tomography and some of the fundamental questions introduced above, in particular the origin of nucleon spin and the emergence of its macroscopic properties from QCD dynamics.

1.4.3 Gluon distributions and saturation

A further central goal of the EIC is the exploration of the gluon dominated regime of QCD. The physical origin of the high-density regime is illustrated schematically in Figure 25. As progressively smaller momentum fractions are resolved, the number of low-x gluons and sea quarks increases, eventually motivating the question of whether non-linear gluon interactions can balance further parton multiplication and lead to saturation [18, 3]. Gluons are responsible for a large fraction of the momentum and dynamical structure of hadrons, but their distributions remain considerably less constrained than those of quarks, particularly at small x. The EIC will extend measurements of proton and nuclear structure into this region through several complementary observables, including inclusive structure functions, heavy-flavour production, jet and dihadron correlations, and diffractive processes [3].

As discussed in Sec. 1.3, the gluon density increases rapidly towards low x, eventually raising the question of whether linear QCD evolution remains applicable indefinitely. The possible onset of gluon saturation represents one of the major discovery objectives of the EIC. The importance of nuclear beams follows from the enhancement of the saturation scale with the atomic mass number: heavy nuclei provide a larger transverse gluon density and therefore allow the high-density regime to be approached under less extreme kinematic conditions than in the proton [3, 18]. By varying the nuclear species, the EIC will consequently provide an experimental control parameter for studying how high-density QCD phenomena evolve with the size of the target.

No single observable is expected to provide a complete characterization of this regime. The

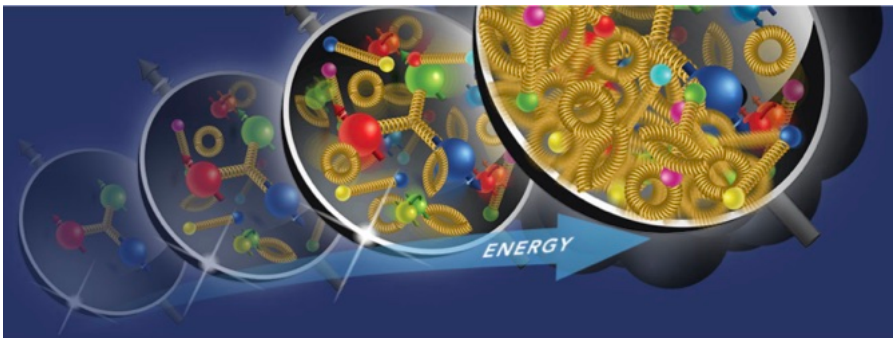


Figure 25: Schematic representation of the proton structure probed at progressively higher energies. Increasing the resolving power exposes an increasingly dense population of low-x gluons and quark-antiquark pairs, providing access to the high-parton-density regime investigated at the EIC. Image from [18].

EIC programme therefore relies on a combination of measurements with different sensitivities to the gluonic structure. Inclusive DIS will constrain the evolution of the gluon density, while charm production offers more direct sensitivity through photon-gluon interactions. Correlations between approximately back-to-back hadrons or jets can probe the transverse momentum structure of low-x gluons and their modification in heavy nuclei. Diffractive reactions are particularly important because of their enhanced sensitivity to the gluon field and to non-linear QCD dynamics [3]. The comparison of these observables in ep and eA collisions, together with their evolution in x, Q^2, and A, will provide a systematic test of whether a universal high-density regime of gluonic matter emerges at sufficiently high energies [18].

1.4.4 Emergence of hadrons and interaction with nuclear matter

Confinement implies that the quarks and gluons participating in a hard interaction are never observed as isolated particles. They must instead evolve into colour-neutral hadrons through the non-perturbative process of hadronization, as can be schematically seen in Figure 26. Although fragmentation functions provide a phenomenological description of this transition, the underlying space-time evolution from an energetic colour charge to an observable hadron is still incompletely understood. Important open questions include the characteristic timescales of colour neutralization and hadron formation, the differences between quark and gluon fragmentation, and the dependence of hadronization on parton flavour and on the surrounding QCD environment [18].

The availability of nuclear beams makes the EIC particularly suited to address these questions. A nucleus can be used as a controlled volume of cold QCD matter through which the struck parton propagates before forming the final-state hadrons. By comparing otherwise similar ep and eA interactions, and by changing the atomic mass number of the target, it becomes possible to vary the amount of nuclear matter encountered by the propagating colour charge. Such measurements are sensitive to partonic energy loss, transverse momentum broadening, modifications of the parton shower, and possible interactions of the forming hadron with the

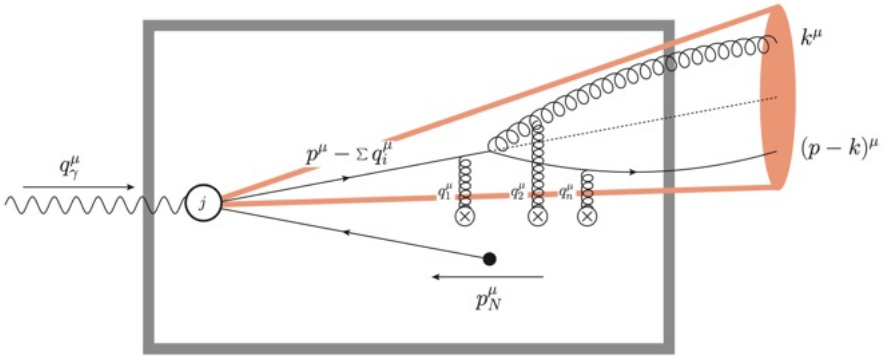


Figure 26: Schematic representation of parton propagation through nuclear matter in semi-inclusive deep-inelastic scattering. After the hard interaction, the struck parton propagates through the nuclear medium and develops a parton shower before the final hadronic state is formed. The nuclear environment can modify the shower evolution and the subsequent hadronization process. Image from [3].

surrounding nuclear environment [3].

The broad kinematic reach of the EIC will allow these phenomena to be studied using both light and heavy flavours and through complementary hadron and jet observables. Identified hadrons retain detailed information on the fragmentation process, while jets provide a closer experimental proxy for the original parton and enable the evolution of the full parton shower to be investigated. Heavy-flavour hadrons add an additional mass scale and are expected to provide distinctive sensitivity to hadron-formation mechanisms. Measurements of light and heavy hadron production in nuclei are identified as an important tool for disentangling different descriptions of in-medium hadronization [18].

These studies also establish a connection between the EIC programme and heavy-ion physics. Unlike the hot and rapidly evolving quark–gluon plasma produced in nucleus–nucleus collisions, the cold nuclear environment of DIS provides a comparatively controlled system in which the initial hard interaction is well defined. Improved knowledge of parton propagation and hadronization in cold nuclear matter will consequently help separate cold-nuclear effects from phenomena associated with a hot QCD medium, providing essential input for the interpretation of measurements in proton–nucleus and nucleus–nucleus collisions [3, 18].

1.5 Experimental requirements to access to EIC physics

The main design requirements of the EIC are [3, 22]:

- Highly polarized electron ($\sim 70\%$) and ion beams;
- Ion beams from deuterons to heavy nuclei such as gold, lead, or uranium;
- Variable $e+p$ center-of-mass energies, with the reference beam configurations spanning approximately $\sqrt{s} \simeq 29 - 141$ GeV;

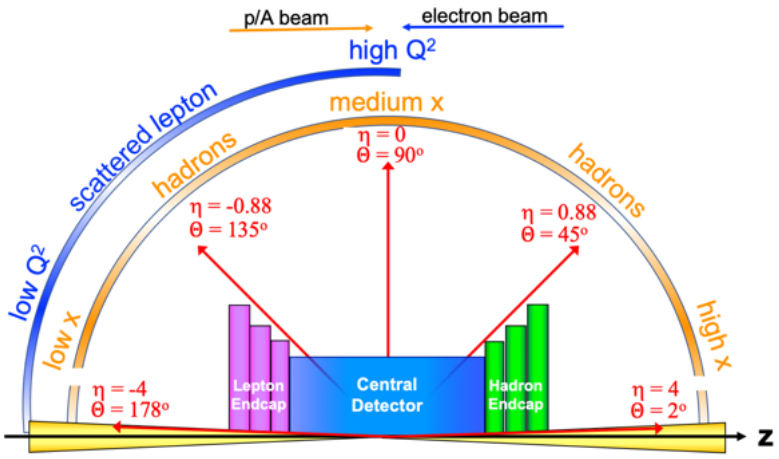


Figure 27: A schematic showing how hadrons and the scattered lepton for different $x - Q^2$ are distributed over the detector rapidity coverage. Image from [22].

- High electron-nucleon luminosity of order $10^{33} - 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$;
- Possibility to have more than one interaction region; for the moment only one detector (ePIC) is being designed, to be placed at the IP6 collision point.

The ePIC experiment has been designed in an almost hermetic layout to provide unprecedented features: high resolution at small transverse momenta to access information at a scale comparable to confinement, and excellent particle identification to access flavour sensitivity [23]. In contrast to symmetric ee and pp colliders, the asymmetric nature of collisions at the EIC leads to unique detector requirements. The hadron endcap, barrel, and electron endcap detector systems see very different particle distributions, in terms of both momentum and particle types. Likewise, the performance requirements on these detector systems vary significantly between the detector regions. This is reflected by the critical detector requirements for the track, vertex, and energy resolution and particle identification separation that are described in the following paragraphs.

In addition to the central region instrumentation, the region most relevant from requirements standpoint for exclusive and diffractive processes is situated beyond the main detector. This region is known as the far-forward area. This unique area allows particles that have extremely small scattering angles to be detected. Further discussion regarding detector requirements is presented, divided into typologies of measurements. The coordinates system taken as a reference for these sections is presented in Figure 27.

1.5.1 Inclusive measurements

Inclusive reactions may be divided into two types of interactions – those that proceed via the exchange of a virtual photon or Z boson, referred to as neutral current (NC) events, and those that proceed via the exchange of a charged W boson, referred to as charged current (CC) events.

For inclusive reactions the kinematic of the interaction are reconstructed by either detecting the scattered electron, or by reconstructing the hadronic recoil. In the case of electron reconstruction the following definitions for x, Q^2 and y are used:

$$Q^2 = 4EE' \cos^2\left(\theta_p^{e'}/2\right) \qquad y = 1 - \frac{E'(1-\cos\theta_p^{e'})}{2E} \qquad x = \frac{Q^2}{sy} \tag{32}$$

where s is the centre-of-mass energy, E and E' are the energy of the incoming and scattered electron and $\theta_p^{e'}$ is the angle of the scattered beam electron measured with respect to the incoming proton axis $(+z)$. If the hadronic recoil is used to reconstruct the event kinematics, then the Jacquet-Blondel [3] definitions are used instead:

$$Q^2_{JB} = \frac{p_T^2}{1-y_{JB}} \qquad y_{JB} = \frac{(E-p^z)}{2E} \qquad x_{JB} = \frac{Q^2_{JB}}{sy_{JB}} \tag{33}$$

where $p_T^2 = \left(\sum_h p_h^x\right)^2 + \left(\sum_h p_h^y\right)^2$ and $(E-p^z) = \sum_h (E_h - p_h^z)$ are summed over the all of the final state hadrons in the event. For CC channels, JB reconstruction is the only option, since the outgoing neutrino cannot be detected, while for NC channels it is possible to use electron, JB or a mixture of electron and hadronic reconstruction techniques. Typically, JB reconstruction is used for events with small y, as the x and y resolutions decrease rapidly in this region for the electron reconstruction variables.

Scattered electron The scattered electron kinematics is very different depending on x, Q^2 so the detector needs excellent electron ID in a wide range of energies and polar angles [22]. As a consequence, an equal rapidity coverage for tracking and EM calorimeter is set. The minimum electron momentum that can be detected is set by the proposed acceptance and the magnetic field of the detector. The vast majority of the inclusive channels require a $Q^2 > 1\ \mathrm{GeV}^2$ and $W^2 > 4\ \mathrm{GeV}^2$ cut for interpretation within a pQCD framework. The requirements exclude nearly the entire available phase space for $\eta < -3.5$. The conclusion is that inclusive reactions with these kinematic cuts do not require detection capabilities beyond $\eta = -3.5$. Furthermore a low material budget is advised to reduce bremsstrahlung.

Hadronic recoil The hadronic recoil in NC events yield and average momentum are highly peaked in the proton beam direction. The accuracy of the JB reconstruction method relies on detecting as much of this hadronic recoil as possible, motivating forward coverage for both electromagnetic and hadronic calorimeters. Charged Current events also rely on the JB reconstruction method so it is critical to investigate the EM and hadronic recoil in CC events, and their distributions also show a highly peaked distribution in the far forward region. However, in contrast to NC events, there is a much larger, higher momentum tail extending into the mid-forward rapidity region. This results were obtained producing momentum distributions (with Pythia6 [3]) of the scattered beam electron, hadronic recoil, for different beam configurations. The present ePIC configuration provides electromagnetic calorimetry approximately over $-3.5 \lesssim \eta \lesssim 3.9$, while the hadronic calorimeter system extends approximately from $\eta \simeq -4.14$ in the electron-going direction to $\eta \simeq 4.0$ in the hadron-going direction [22].

Then, from simulations considering a variety of colliding events, the following constraints were obtained on ECal resolution [22]:

- Good electromagnetic calorimeter resolution in the central and forward detector:

$$\frac{\sigma(E)}{E} \approx \begin{cases} \frac{10\%}{\sqrt{E}} \oplus (2-3)\%, & \text{central region,} \\ \frac{10-12\%}{\sqrt{E}} \oplus (2-3)\%, & \text{forward region,} \end{cases}$$

- Excellent electromagnetic calorimeter resolution at backward rapidities:

$$\frac{\sigma(E)}{E} \approx \frac{2\%}{\sqrt{E}} \oplus (1-2)\%.$$

Meanwhile the HCal resolution requirement is

$$\frac{\sigma(E)}{E} \approx \frac{50\%}{\sqrt{E}} \oplus 10\%$$

in the central and forward regions, while in the backward region it is relaxed to

$$\frac{\sigma(E)}{E} \approx \frac{100\%}{\sqrt{E}} \oplus 10\%.$$

1.5.2 Semi-inclusive measurements

For Semi-inclusive DIS (SIDIS) both the lepton and the outgoing hadrons need to be detected.

For the applicability of perturbative QCD for the hard processes and factorization into non-perturbative distribution and fragmentation functions to be valid, the DIS kinematics require a $Q^2 > 1\ \mathrm{GeV}^2$, together with a selection of $0.01 < y < 0.95$ and $W^2 > 10\ \mathrm{GeV}^2$.

Electron and charged hadron identification In general is necessary to separate electrons from photons and charged hadrons, or to separate charged pions, kaons and protons from each other. In the PID system, the different subdetector will perform the following tasks:

- **Backward RICH**, situated in the electron endcap, with a pion/kaon separation of 3σ up to approximately $7\ \mathrm{GeV/c}$ and a timing reference of approximately $20\ \mathrm{ps}$ [22]. It will be based on the classical technology of a single-volume proximity-focusing aerogel RICH with an extended proximity gap.

- **Barrel DIRC** will provide pion/kaon separation of 3σ at $6\ \mathrm{GeV/c}$, using a high-performance DIRC composed of quartz bar radiators (reusage of BaBar DIRC bars) and performing light detection with MCP-PMTs [22].

- **Forward RICH**, situated in the hadron endcap, will provide pion/kaon separation of 3σ up to approximately $50\,\mathrm{GeV/c}$; it will be constructed with the proven technology of the Dual Radiator RICH, composed of aerogel and C_2F_6 gas, using spherical mirrors for focusing and employing, for the first time in HEP, SiPMs as photosensors with single photon detection efficiency [22, 24].

- **Time of Flight** will perform e, π, K, p separation with single-hit timing resolutions of approximately $35\,\mathrm{ps}$ in the barrel and $25\,\mathrm{ps}$ in the forward region. The BTOF provides 3σ π/K separation up to $p_T \simeq 1.2\,\mathrm{GeV/c}$, while the FTOF extends the separation up to $p \simeq 2.5\,\mathrm{GeV/c}$ [22]. It will be the first employment of AC-LGAD (Low Gain Avalanche Detector) in a collider detector.

In addition to the single hadron requirements, coverage down to low hadron momenta of $p_T > 1\,\mathrm{GeV}$ is required for the partial-wave analysis of di-hadron final states.

1.5.3 Exclusive measurements

In exclusive DIS, $l + p/A \to l' + p'/A' + \gamma/h^{\pm,0}/VM$ the boson emitted by the lepton probes the internal structure of the nucleon while the target remains intact in the final state. It requires the measurement of all particles in the final state, so of the lepton, the hadron and or the photon, in case of DVCS DIS, or the vector meson (in particular, its decay products) for VM DIS. For example, e^+e^- ($\mu^+\mu^-$) pairs from J/ψ decay, or K^+K^- from ϕ decay. It is essential for such reactions to accurately reconstruct all the particles in the event; achieving such reconstruction with sufficient accuracy requires using numerous detector components. Thus exclusive and diffractive reactions demand hermeticity of the setup while also adding the additional requirement of precision far-forward detectors.

1.6 The ePIC experiment

1.6.1 Overall detector concept

The ePIC experiment is the general-purpose detector being developed for the Electron-Ion Collider at Interaction Point 6 (IP6). Its design is driven by the broad EIC physics programme and by the requirement to reconstruct the complete final state over a wide kinematic range. The experimental system is composed of a *Central Detector*, surrounding the interaction point, and a set of *Far Detectors* located along the outgoing electron and hadron beam lines. Together, these systems provide acceptance in regions ranging from the central detector, approximately $-3.5 < \eta < 3.5$, to very small scattering angles that cannot be instrumented within the central apparatus [22].

The central detector, shown schematically in Figure 28, follows a layered architecture. Moving radially outward from the interaction point, charged-particle tracking is followed by dedicated particle identification systems, electromagnetic calorimetry, the superconducting solenoid in the barrel, and hadronic calorimetry. This arrangement minimizes the material traversed by particles before tracking and calorimetric measurements while maintaining nearly hermetic coverage of the final state [22]. The individual detector technologies are not uniform over the full

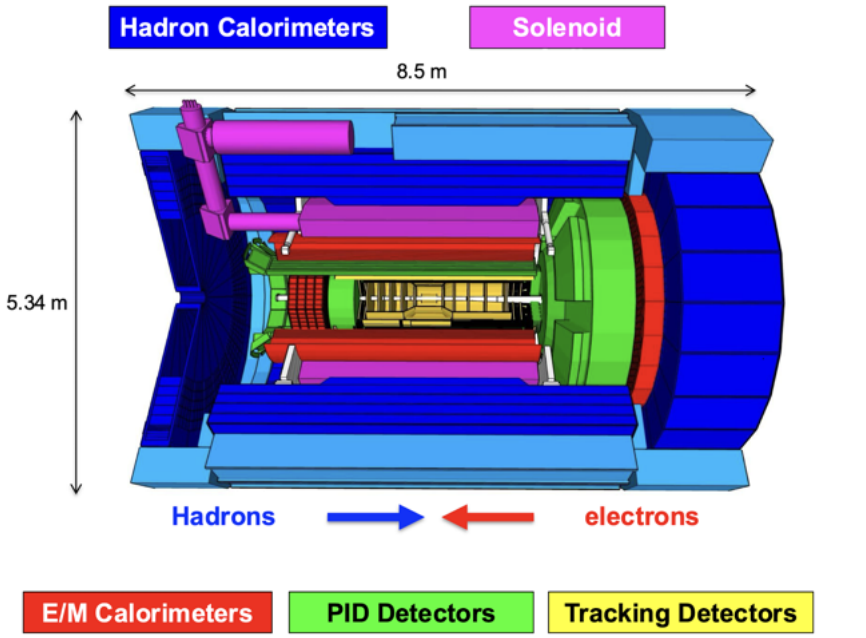


Figure 28: Schematic view of the ePIC Central Detector, showing the principal tracking, particle-identification and calorimeter subsystems surrounding the interaction region. Image from [3].

acceptance, but are instead selected according to the different particle momenta, occupancies and physics requirements of the electron-going, central and hadron-going regions.

At the centre of the detector is the MARCO superconducting solenoid, which provides the magnetic field required for charged-particle momentum determination. The reference operating field for ePIC is $1.7\,\mathrm{T}$, while the magnet is designed to operate up to $2\,\mathrm{T}$. In addition to providing a sufficiently homogeneous field in the central region, its magnetic configuration is optimized in the forward direction to satisfy the requirements of the forward Ring Imaging Cherenkov detector [22]. Detector geometry, magnetic field, tracking and particle-identification systems must therefore be considered as parts of a single integrated design rather than as independent subsystems.

A further defining aspect of ePIC is the adoption of a streaming readout architecture. Instead of using a conventional hardware trigger to decide which collisions are recorded, detector information is continuously read out and processed. This strategy provides flexibility in event reconstruction and selection and is particularly appropriate for the diverse final states foreseen in the EIC physics programme [22].

1.6.2 Asymmetric detector layout

A characteristic feature of ePIC is its intrinsically asymmetric geometry. This configuration reflects both the unequal energies of the colliding electron and hadron beams and the corresponding asymmetry of the final-state particle kinematics. In the convention adopted for ePIC, negative pseudorapidity corresponds to the electron-going, or backward, direction, while pos-

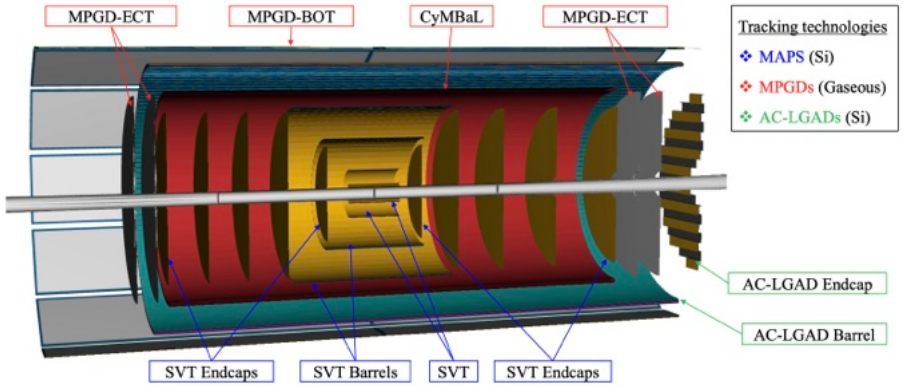


Figure 29: Overview of the ePIC tracking system, showing the silicon vertex tracker, MPGD barrel and endcap layers, and the AC-LGAD timing detectors. Reproduced from [22].

itive pseudorapidity corresponds to the hadron-going, or forward, direction. As illustrated in Figure 27, scattered electrons and produced hadrons populate different regions of the detector depending on the DIS kinematics, making identical instrumentation on the two sides neither necessary nor optimal [22].

The central barrel approximately covers $-1.5 < \eta < 1.5$, while the backward and forward endcaps extend the acceptance toward the electron- and hadron-going directions, respectively. Subsystem-specific acceptance can extend toward $|\eta| \simeq 3.5$–4, with the ultimate limit determined in part by the beam-pipe geometry. These boundaries should therefore not be interpreted as sharp divisions: several detector systems intentionally overlap between neighbouring regions in order to avoid acceptance gaps and to provide redundant measurements [22].

The backward endcap is particularly important for the reconstruction of the scattered electron. Precise measurement of its energy and direction is fundamental for the determination of DIS kinematic variables and consequently requires high-performance tracking and electromagnetic calorimetry. The hadron-going region, in contrast, must accommodate particles with substantially larger longitudinal momenta. This results in demanding requirements on tracking, particle identification and calorimetry over a broad momentum interval.

The asymmetry also extends beyond the Central Detector. Particles scattered at extremely small angles remain close to the outgoing beam trajectories and cannot be measured by the conventional endcap detectors. Dedicated far-forward instrumentation is therefore integrated into the hadron beam line, while luminosity instrumentation and low-Q^2 electron taggers are installed along the electron-going beam line. The full ePIC system consequently extends well beyond the approximately $10\ \mathrm{m}$ longitudinal envelope available for the Central Detector [22].

1.6.3 Tracking and vertexing

Precise reconstruction of charged-particle trajectories is essential to almost every component of the EIC physics programme. Tracking determines particle momenta from their curvature in the magnetic field and provides the direction information required for matching tracks to calorime-

ters and particle-identification detectors. Precise reconstruction of the primary and secondary vertices is also required for the identification of displaced decays, particularly in measurements involving charm and beauty hadrons [22].

The ePIC tracking system combines silicon Monolithic Active Pixel Sensors (MAPS) with Micro-Pattern Gaseous Detectors (MPGDs), as summarized in Figure 29. The innermost part is the Silicon Vertex Tracker (SVT), whose geometry consists of cylindrical barrel layers around the beam pipe together with silicon disks extending the coverage into the forward and backward regions. The three innermost barrel layers are based on the ultra-thin MAPS concept developed for ALICE ITS3, adapted to the larger EIC beam pipe. Their very small material budget is particularly important for reducing multiple scattering close to the interaction point. Additional silicon layers at larger radii provide a longer lever arm for the momentum measurement [22]. Around midrapidity, the target momentum resolution is approximately

$$\frac{\sigma_p}{p} \simeq 0.05\%\, p \oplus 0.5\%,$$

with p expressed in GeV/c, while the transverse impact-parameter resolution is required to approach

$$\sigma_{\mathrm{DCA}_{xy}} \simeq \frac{20\ \mu\mathrm{m}}{p_T} \oplus 5\ \mu\mathrm{m}.$$

In the forward and backward regions the required momentum resolution is approximately $0.05\%\, p \oplus 1\%$. These requirements are closely connected to measurements of SIDIS, exclusive reactions and heavy-flavour production [22].

The silicon tracker is complemented at larger radii by MPGDs. In the barrel these include cylindrical Micromegas and a μRWELL-GEM outer tracking system, while μRWELL–GEM disks extend the measurement in the endcaps. The combination is advantageous because the silicon sensors provide excellent spatial precision and very low material thickness, whereas the gaseous detectors provide additional tracking points and timing information at the level of approximately 10–$20\ \mathrm{ns}$. The outer barrel tracking layers are also important for the hpDIRC, whose Cherenkov-angle resolution depends directly on precise knowledge of the direction of the incident track [22].

Finally, the dedicated AC-LGAD timing layers in the barrel and forward endcap contribute additional high-precision spatial measurements. Tracking, vertexing and timing are therefore strongly interconnected in ePIC: no individual detector technology alone provides all required information, and the complete charged-particle reconstruction results from the combination of the different subsystems.

1.6.4 Calorimetry

The calorimeter system provides measurements of electromagnetic and hadronic energy and is an essential component of event reconstruction in ePIC. Electromagnetic calorimetry is required for precise determination of the scattered-electron energy, photon reconstruction, electron-hadron discrimination and the reconstruction of neutral mesons. Hadronic calorimetry complements the tracking system by measuring neutral hadrons and is required for a complete re-

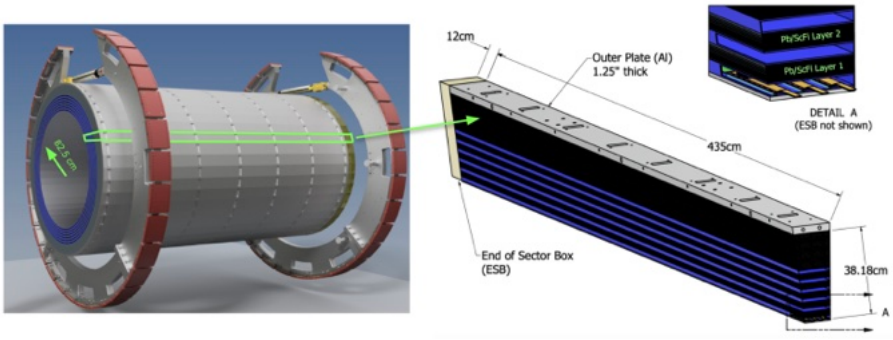


Figure 30: Structure of the ePIC Barrel Imaging Calorimeter. On the left, a 3D rendering of the Barrel Imaging Calorimeter with its 48 sectors. The central drawing shows the structure of a single sector, featuring interleaved Pb/ScFi layers and slots for trays holding AstroPix chips, followed by the Pb/ScFi bulk section. On the right, a zoomed view of the first radially layers is presented. Image from [22].

construction of the hadronic final state and of jets [22]. As for the other ePIC subsystems, the technologies are optimized independently in the backward, barrel and forward regions.

In the electron-going endcap, the Electron-Endcap Electromagnetic Calorimeter (EEEMCal) is a fine-granularity homogeneous calorimeter based on $\mathrm{PbWO_4}$ crystals with SiPM readout. An excellent intrinsic resolution of the crystal technology is required for precise reconstruction of the scattered electron, particularly in the low-x region. The target backward electromagnetic energy resolution is approximately

$$\frac{\sigma_E}{E} \simeq \frac{2\%}{\sqrt{E}} \oplus (1\text{-}2)\%,$$

where E is expressed in GeV [22].

The barrel electromagnetic system is the Barrel Imaging Calorimeter (BIC), whose structure is illustrated in Figure 30. The BIC combines a lead/scintillating-fibre (Pb/SciFi) sampling calorimeter with several imaging layers based on AstroPix MAPS sensors. The Pb/SciFi component provides the energy measurement, while the silicon layers sample the spatial development of the electromagnetic shower. This hybrid architecture improves shower localization, electron/hadron discrimination and the separation of nearby photons, while remaining compatible with the limited radial space available inside the solenoid [22].

In the hadron-going direction, the Forward Electromagnetic Calorimeter (FEMC) uses a compact tungsten-scintillating-fibre technology. Its fine granularity is required to reconstruct photons in the dense forward environment and, in particular, to distinguish isolated photons from the two photons produced in high-energy π^0 decays. The central and forward electromagnetic calorimeters have less stringent energy-resolution requirements than the backward crystal calorimeter, with targets of approximately $10\%/\sqrt{E}\oplus(2\text{–}3)\%$ in the barrel and $(10\text{–}12)\%/\sqrt{E}\oplus(2\text{–}3)\%$ in the forward region [22].

Hadronic calorimetry is likewise divided into three regions. The Barrel Hadronic Calorimeter (BHCal), located outside the solenoid, is based on the outer hadronic calorimeter developed for

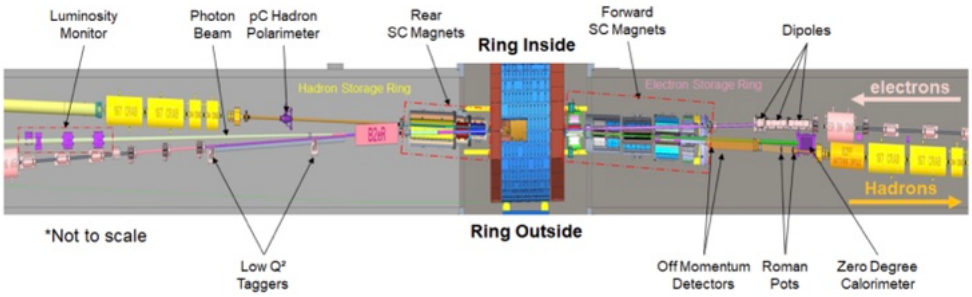


Figure 31: Schematic layout of the ePIC far-detector systems along the outgoing electron and hadron beam lines. Reproduced from [22].

sPHENIX and acts both as a hadronic-energy measurement system and as part of the magnet flux return. In the forward endcap, the Longitudinally-segmented Forward Hadronic Calorimeter (LFHCal) employs steel absorbers and scintillator tiles with SiPM readout. Its longitudinal and transverse segmentation is particularly important for resolving hadronic showers in the high-energy hadron-going region. The backward nHCal is primarily a tail catcher for the relatively low-energy neutral hadronic component and uses a related SiPM-on-tile approach [22]. A common feature of the ePIC calorimeters is the extensive use of SiPMs as photosensors, although the sensor size and pixelization are adapted to each subsystem.

1.6.5 Far-forward and far-backward systems

The acceptance of the Central Detector cannot extend to particles travelling at extremely small angles with respect to the colliding beams. Dedicated instrumentation is therefore integrated into very close to the beam pipes on both sides of IP6. The overall arrangement is shown in Figure 31. These detectors are particularly important for exclusive and diffractive measurements, spectator tagging, luminosity determination and the detection of electrons at very low Q^2 [22].

On the hadron-going side, four far-forward systems extend the $\eta > 4.5$ and are distributed along the outgoing beam line. The B0 system is located inside the first dipole magnet and combines charged-particle tracking with electromagnetic calorimetry. The tracker is based on AC-LGAD technology, while the electromagnetic calorimeter uses $\mathrm{PbWO_4}$ crystals. Farther downstream, Roman Pot (RPOT) and Off-Momentum Detector (OMD) stations measure charged particles travelling very close to the outgoing beam. The RPOT system is optimized for particles retaining a large fraction of the beam momentum, while the OMD provides acceptance for particles with substantially different magnetic rigidity, including spectator protons from the breakup of light nuclei. Together, these systems provide broad acceptance for very-forward charged particles [22].

Neutral particles travelling in the hadron direction are measured by the Zero Degree Calorimeter (ZDC), positioned downstream of the magnetic elements separating charged particles from the beam trajectory. The ZDC provides sensitivity to forward neutrons and photons and is particularly important for nuclear-breakup tagging and exclusive and diffractive measurements. The forward detectors therefore extend the physics acceptance into a region inaccessible to the

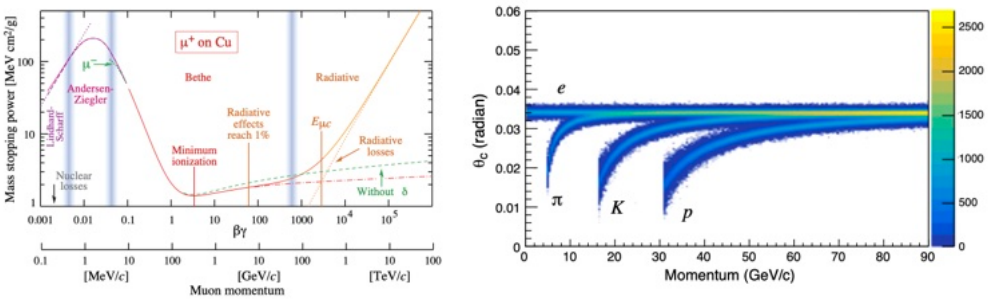


Figure 32: Example of the velocity-dependent detector interactions used for PID devices. The left panel denotes the rate of energy loss of charged particle passing through matter. The right panel denotes the angle of Cherenkov radiation. Image from [3].

central tracking and calorimeter systems.

The far-backward instrumentation is located along the outgoing electron beam line and serves two main purposes: luminosity determination and the detection of electrons scattered at extremely small angles. The luminosity system exploits the precisely calculable bremsstrahlung process. It combines direct detection of bremsstrahlung photons with pair spectrometers that measure the e^+e^- pairs produced by photon conversion. The design goal is an absolute luminosity determination at approximately the 1% level, together with much more precise relative luminosity measurements required for spin-dependent observables [22].

Low-Q^2 taggers are positioned close to the outgoing electron beam and combine tracking stations with electromagnetic calorimetry. They provide access to scattered electrons that remain within a few degrees of the beam direction and therefore fall outside the acceptance of the backward endcap. The far detectors are consequently an integral part of the ePIC physics acceptance rather than auxiliary instrumentation: they recover final-state particles that carry essential information for several of the principal EIC measurements.

1.6.6 General PID role

Particle identification is a central component of the ePIC detector concept because tracking alone determines the momentum and charge of a particle but does not, in general, distinguish pions, kaons and protons. Identified charged hadrons are required for a large fraction of the EIC physics programme, particularly for semi-inclusive measurements in which the species of the observed hadron provides sensitivity to the flavour of the fragmenting parton. PID information also supports the reconstruction of strange and heavy-flavour hadrons and improves the rejection of charged pions from the scattered-electron sample [22]. The physics role of these detectors will be discussed in detail in Chapter 2.

The required momentum range changes strongly with pseudorapidity. The ePIC detector requirements call for $\pi/K/p$ separation up to several GeV/c in the barrel and electron-going region and up to several tens of GeV/c in the hadron-going direction. No single detector principle can span this complete range, so a combination of detectors must be used. Figure 32 shows two example of the velocity-dependent detector interactions used for PID devices. ePIC therefore

combines Cherenkov imaging at intermediate and high momenta with precision time-of-flight measurements at lower momenta.

In the electron-going endcap, identification is provided by the proximity-focusing RICH (pfRICH), which uses an aerogel radiator and High Rate Picosecond PhotoDetectors (HRPPDs). In the barrel, the limited radial space motivates the use of a high-performance DIRC (hpDIRC), in which fused silica acts simultaneously as the Cherenkov radiator and as a light guide. In the forward endcap, the dual-radiator RICH (dRICH) combines aerogel and gas radiators in a common focusing system, extending charged-hadron identification to the large momenta characteristic of the hadron-going region [22].

The Cherenkov systems are complemented by AC-LGAD time-of-flight layers in the barrel and forward endcap. Their excellent time resolution extends hadron separation toward low momentum, where the Cherenkov detectors lose sensitivity, while simultaneously providing additional spatial points for charged-particle tracking. The use of overlapping techniques is therefore essential to maintain continuous PID performance across the broad ePIC acceptance. The detailed physics motivation, quantitative separation requirements and operating principles of these detectors are developed in Chapter 2.

2 Particle Identification at the EIC

The purpose of this Chapter is to discuss the various aspects of Particle Identification (PID) in detail. Sec. 2.2 is dedicated to the role of particle identification in EIC physics and to explain why PID is needed. Sec. 2.3 is devoted to a detailed description of the particle identification requirements, covering momentum and pseudorapidity distributions, required $\pi/K/p$ separation, different requirements in the electron endcap, barrel and hadron endcap and the complementarity of timing and Cherenkov techniques. The subsequent sections are dedicated to the experimental architecture, describing how the ePIC PID systems are structured (Sec. 2.4), with a particular focus on the dual-radiator RICH (Sec. 2.5) and SiPM-based photon detection (Sec. 2.6), as this is the field within which this thesis research falls.

2.1 Particle identification

Particle identification (PID) is a fundamental component of modern particle-physics experiments, complementary to tracking and calorimetry. In its broadest sense, PID consists in assigning a reconstructed particle to a given species, or at least to a particle family, by exploiting differences in the way particles interact with the detector or in observables related to their mass [25]. These two approaches can be broadly distinguished as *identification through interaction properties* and *identification through mass determination*.

The first approach corresponds to the conventional identification strategy employed in general-purpose detectors. Different particle species produce characteristic signatures in the various detector subsystems: charged particles leave tracks in the tracking system, electrons and photons develop electromagnetic showers, hadrons interact predominantly through hadronic processes, while muons can traverse the calorimeters and reach dedicated outer detectors. The combined information from tracking, electromagnetic and hadronic calorimetry, and muon detection therefore allows electrons, photons and muons to be distinguished from charged and neutral hadrons [25]. This approach, however, does not generally provide sufficient discrimination among different charged-hadron species.

The separation of pions, kaons and protons instead relies primarily on their different masses. These particles produce similar signatures in conventional tracking and calorimetric systems, but at the same momentum their velocities are different. Since the momentum p is reconstructed from the particle trajectory, an independent measurement of the velocity $\beta = v/c$ allows the mass to be inferred through

$$m = \frac{p}{c\beta\gamma}, \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}}. \tag{34}$$

Velocity-sensitive observables therefore form the basis of charged-hadron identification. Among the most widely used techniques are ionisation-energy-loss measurements, time-of-flight measurements and Cherenkov detectors [25]. In the following, the discussion focuses primarily on the separation of charged hadrons, and in particular on $\pi/K/p$ identification.

The performance of a PID system is commonly quantified in terms of its separation power.

For two particle hypotheses A and B, it can be expressed as

$$N_\sigma^{A/B} = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \tag{35}$$

where μ_A and μ_B are the mean values of the discriminating observable expected for the two hypotheses, while σ_A and σ_B represent the corresponding resolutions. A larger value of $N_\sigma^{A/B}$ therefore corresponds to a better separation between the two particle species.

No single velocity-sensitive observable provides optimal discrimination over an arbitrarily wide momentum range. Time-of-flight and Cherenkov measurements both determine the particle velocity, but their separation power evolves differently with momentum. This complementarity makes their combination particularly effective when particle identification is required over a broad kinematic range.

In a time-of-flight measurement, the time required by a particle to travel a known path length L is

$$t = \frac{L}{\beta c}. \tag{36}$$

For two different mass hypotheses, the difference in the expected arrival times is therefore

$$\Delta t = \frac{L}{c}\left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\right). \tag{37}$$

At fixed momentum, heavier particles have lower velocities and consequently longer flight times. The separation is strongest at low momentum, where the velocity differences between particle species are comparatively large. As the momentum increases, however,

$$\beta = \frac{pc}{\sqrt{p^2c^2 + m^2c^4}} \longrightarrow 1, \tag{38}$$

and the corresponding time-of-flight differences rapidly decrease.

Cherenkov detectors provide a complementary measurement of the particle velocity. When a charged particle traverses a dielectric medium of refractive index n with a velocity larger than the phase velocity of light in that medium, Cherenkov radiation is emitted. The condition for emission is

$$\beta n > 1, \tag{39}$$

and the photons are emitted at the Cherenkov angle θ_C, defined with respect to the particle direction by

$$\cos\theta_C = \frac{1}{n\beta}. \tag{40}$$

For a given momentum, particles with different masses have different velocities and therefore produce different Cherenkov angles. Measuring θ_C thus provides an independent determination of β and, when combined with the reconstructed momentum, allows the particle species to be identified.

Cherenkov emission occurs only above a mass-dependent threshold. Setting $\beta_{\mathrm{th}} = 1/n$ gives the corresponding threshold momentum,

$$p_{\mathrm{th}} = \frac{mc}{\sqrt{n^2 - 1}}. \tag{41}$$

The refractive index therefore determines the momentum range over which a given radiator is useful for particle identification. Radiators with comparatively large refractive indices have lower Cherenkov thresholds and are suited to lower momenta, whereas low-index radiators have higher thresholds but can retain useful angular separation between particle species at substantially larger momenta. The combination of timing and Cherenkov measurements consequently provides complementary PID information over a broad momentum range.

2.2 Role of particle identification in EIC physics

Particle identification (PID) provides an essential connection between the partonic quantities investigated at the Electron-Ion Collider and the experimentally reconstructed final state. Inclusive deep-inelastic scattering already gives access to fundamental quantities such as structure functions and combinations of parton distribution functions through reconstruction of the scattered lepton. A substantial fraction of the EIC physics programme, however, requires additional information from the hadronic final state. Semi-inclusive measurements, identified hadrons inside jets, heavy-flavour reconstruction and exclusive reactions exploit the species, momentum and angular distributions of final-state particles to obtain information that cannot be reconstructed from the scattered lepton alone [22, 18].

This requirement is ultimately related to confinement. Quarks and gluons produced in a hard interaction do not appear as isolated asymptotic states, but evolve through parton radiation and hadronization into colour-neutral particles. The experimentally reconstructed hadrons therefore retain indirect information about the flavour and momentum of the underlying partonic state and about the non-perturbative fragmentation process. Distinguishing pions, kaons, protons and other hadronic species is consequently required whenever the physics observable depends on the flavour composition or hadronization properties of the final state [22, 18].

The role of PID extends beyond the identification of the struck-parton flavour. It is required for transverse-momentum-dependent measurements, spin-dependent observables, strange and heavy-flavour reconstruction, and the identification of specific exclusive final states. Accordingly, the ePIC physics programme requires charged-hadron identification over a very broad range of momentum and pseudorapidity, making PID an integral component of the physics measurement rather than an auxiliary detector function [22].

2.2.1 Flavour separation in SIDIS

Semi-inclusive deep-inelastic scattering provides one of the clearest examples of the relation between particle identification and partonic flavour. In SIDIS, a hadron is measured in coincidence with the scattered lepton,

$$e + N \to e' + h + X, \tag{42}$$

so that the final-state hadron introduces an additional correlation with the flavour of the struck parton [22].

Within a leading-order collinear description, the production of an identified hadron can be schematically written as

$$\mathrm{d}\sigma^h \propto \sum_q e_q^2 \, f_1^q(x, Q^2) \, D_1^{q\to h}(z, Q^2), \tag{43}$$

where f_1^q represents the parton distribution associated with flavour q and $D_1^{q\to h}$ is the fragmentation function describing the production of hadron h. The variable z denotes the fraction of the fragmenting parton's longitudinal momentum carried by the observed hadron. Different identified hadron species therefore select different combinations of PDFs and fragmentation functions [22].

The flavour sensitivity becomes particularly strong at large z, where the detected hadron is more likely to retain a valence constituent corresponding to the initially struck parton. Positive pions ($\pi^+ : u\bar{d}$), for example, provide enhanced sensitivity to up-quark fragmentation, whereas negative pions ($\pi^- : \bar{u}d$) increase the relative sensitivity to down quarks. Charged kaons ($K^+ : u\bar{s},\ K^- : \bar{u}s$) introduce additional sensitivity to the strange sector through their strange valence content [22]. These differences allow pion and kaon measurements to constrain flavour combinations that cannot be separated using inclusive DIS alone.

The importance of reliable hadron identification is demonstrated directly by full ePIC simulations. Figure 33 shows the kaon identification efficiency and purity in three representative pseudorapidity intervals. The transition between the low-momentum region, in which timing information dominates, and the higher-momentum region covered by the Cherenkov detectors is clearly visible. The figure also illustrates an important experimental point: because pion production is considerably more abundant than kaon production, even a small pion-to-kaon misidentification probability may produce a non-negligible contamination of the kaon sample [22]. Hadron identification is particularly important for the sea-quark structure of the proton. Semi-inclusive pion and kaon measurements provide direct sensitivity to sea-quark polarization and to the strange-quark contribution, which remains significantly less constrained than the dominant light-quark distributions. The ability to identify kaons with high efficiency and controlled contamination is therefore crucial both for unpolarized flavour separation and for spin-dependent measurements [22].

PID consequently acts as a flavour analyser in SIDIS: the measured momentum of the track determines the kinematics, while its identified species determines how strongly the event weights the different partonic flavours.

2.2.2 Fragmentation functions

The interpretation of identified-hadron production necessarily involves fragmentation functions. These non-perturbative quantities describe the transition between a high-energy parton and the observable hadrons produced during fragmentation. In a collinear description, the unpolarized fragmentation function

$$D_1^{q\to h}(z, Q^2) \tag{44}$$

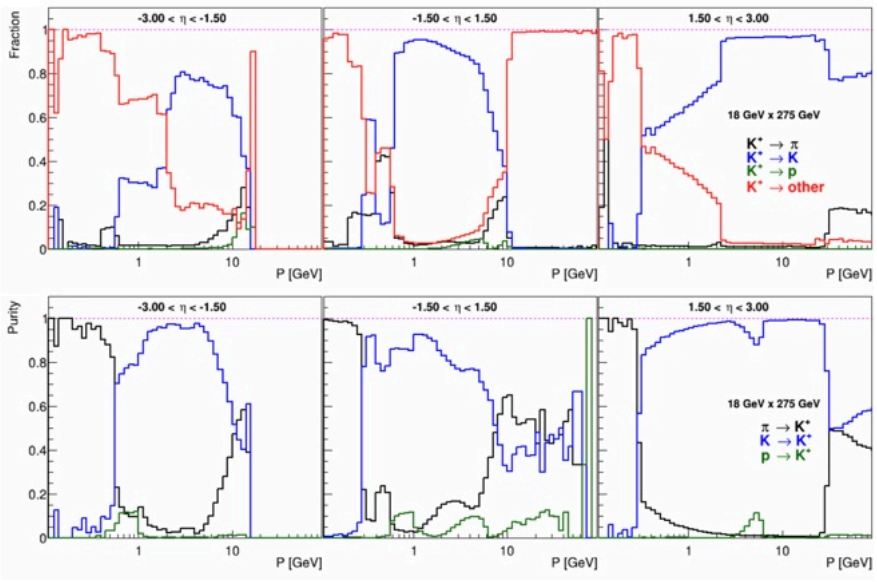


Figure 33: Simulated kaon-identification efficiencies and purities as a function of hadron momentum in three pseudorapidity regions of ePIC. The corresponding pion, proton and other-particle misidentification contributions are also shown. Image from [22].

encodes the distribution of hadrons of species h produced from a parton of flavour q as a function of the longitudinal momentum fraction z. Fragmentation functions should not be interpreted as simple classical probabilities; they are QCD correlation functions entering factorized cross sections, whose phenomenological interpretation nevertheless associates them with parton-to-hadron fragmentation [22].

Different experimental systems provide complementary information on these functions. Inclusive hadron production in e^+e^- annihilation constrains flavour-summed combinations, while hadronic collisions provide important sensitivity to gluon fragmentation. SIDIS is particularly valuable because the flavour dependence of the initial-state PDF is correlated with the species of the measured final-state hadron. This provides sensitivity to individual quark and antiquark fragmentation channels [22].

A useful phenomenological distinction is made between favoured and unfavoured fragmentation. A channel is referred to as favoured when the fragmenting quark is also a valence constituent of the final-state hadron. For example,

$$u \to \pi^+ \tag{45}$$

is favoured because the π^+ has valence structure $u\bar{d}$, whereas

$$u \to \pi^- \tag{46}$$

is unfavoured. Similar relations for charged kaons make kaon identification particularly important for separating strange-quark fragmentation.

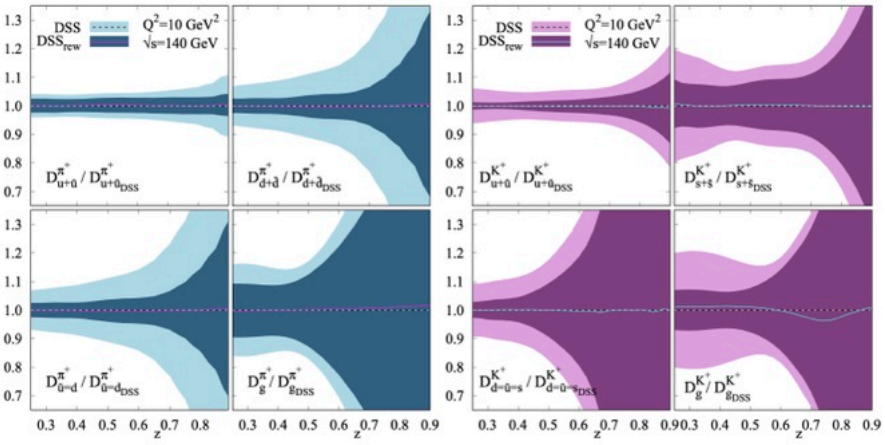


Figure 34: Projected impact of EIC measurements on pion and kaon fragmentation functions. Existing global determinations are compared with fits including simulated EIC SIDIS measurements. Image from [26].

Precise PID is therefore central for fragmentation-function measurements because the experimental pion, kaon and proton samples define the hadron-specific observables entering the extraction. Combining these particles into a single unidentified charged-hadron category would preserve an inclusive hadron yield but would remove much of the flavour-dependent information.

The projected sensitivity of EIC measurements to pion and kaon fragmentation functions can still be illustrated using Figure 34. Fragmentation studies at the EIC extend beyond the collinear single-hadron description. Measurements differential in transverse momentum probe transverse-momentum-dependent fragmentation functions, while dihadron observables retain correlations between pairs of produced hadrons. Identified hadrons reconstructed within jets further allow the perturbative parton shower and the subsequent non-perturbative hadronization stage to be studied with different experimental observables [22, 18].

2.2.3 Transverse-spin and TMD measurements

Transverse-momentum-dependent parton distributions retain information on the intrinsic transverse momentum of partons in addition to their longitudinal momentum fraction. They therefore provide a multidimensional description of partonic motion and, when polarization is included, describe correlations between transverse momentum and the spin of the parton or parent hadron [22, 18].

SIDIS provides experimental access to these distributions because the measured transverse momentum of the final-state hadron receives contributions from both the intrinsic transverse momentum of the initial parton and the transverse momentum generated during fragmentation. Consequently, SIDIS cross sections contain characteristic azimuthal modulations involving different combinations of TMD PDFs and fragmentation functions. Identifying the final-state hadron is important because both the fragmentation function and the flavour weighting de-

pend on the hadron species [22].

A prominent example is the Sivers function,

$$f_{1T}^{\perp q}(x, k_T), \tag{47}$$

a TMD distribution describing the correlation between the intrinsic transverse momentum of an unpolarized quark and the transverse polarization of the parent nucleon [17]. In SIDIS it contributes to a single-spin asymmetry containing a characteristic

$$\sin(\phi_h - \phi_S) \tag{48}$$

modulation, where ϕ_h is the azimuthal angle of the detected hadron momentum and ϕ_S is the azimuthal angle of the transverse nucleon-spin vector, both measured with respect to the lepton scattering plane. Separating positively and negatively charged pions and kaons produces different flavour combinations and is therefore required for a flavour-dependent extraction.

A complementary observable is provided by the Collins mechanism, which describes an azimuthal asymmetry in the fragmentation of a transversely polarized quark into an unpolarized hadron [17]. The transversity distribution h_1^q, which describes the net transverse polarization of quarks inside a transversely polarized nucleon, is chiral odd and cannot appear by itself in the leading-twist inclusive DIS cross section. In single-hadron SIDIS it couples to the chiral-odd Collins fragmentation function, producing an azimuthal modulation conventionally written as

$$\sin(\phi_h + \phi_S). \tag{49}$$

The observable therefore combines information on transverse polarization in the initial state with species-dependent fragmentation in the final state.

The broad EIC range in x, Q^2, z and hadron transverse momentum is expected to provide sensitivity to TMD evolution over scales inaccessible to present fixed-target data. Full ePIC performance studies show that flavour-separated TMD measurements can cover the low-transverse-momentum domain and extend toward the perturbative region [22]. Figure 35 shows representative projected uncertainties for Collins asymmetries of identified positive pions. Dihadron measurements provide additional access to transversity through dihadron fragmentation functions. In all these measurements the precision of the extracted spin and TMD observables depends not only on tracking and angular resolution, but also on maintaining high hadron-identification efficiency and purity throughout the relevant multidimensional phase space [22].

2.2.4 Strange-hadron and heavy-flavour studies

Kaons are particularly important for measurements involving strange quarks. Their strange valence content enhances the sensitivity of identified-kaon SIDIS to the strange component of the nucleon and to strange-quark fragmentation. Combined pion and kaon measurements therefore provide a means of separating strange contributions from the much larger up and down quark components [22].

Strange baryons provide additional observables. Hyperons such as the Λ, which has a mean proper decay length of

$$c\tau_\Lambda \simeq 7.85\ \mathrm{cm}, \tag{50}$$

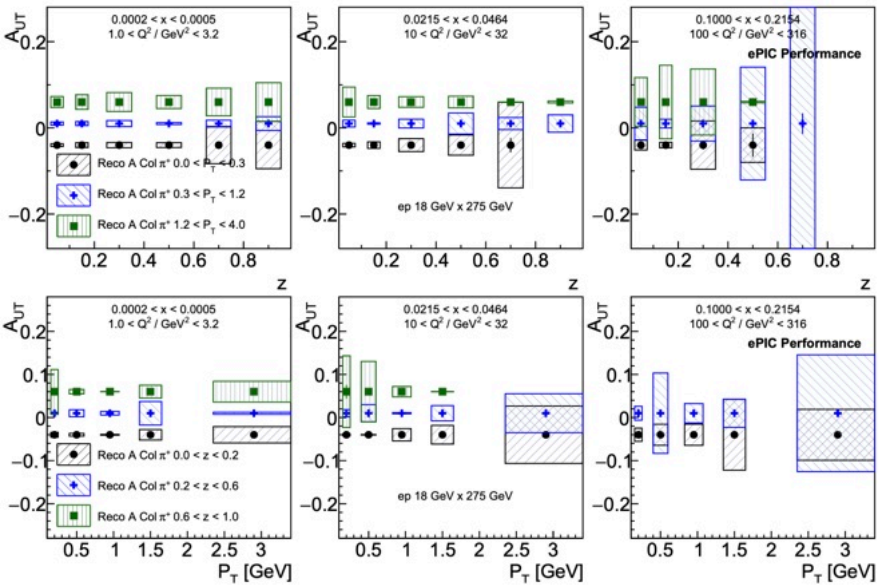


Figure 35: Expected uncertainties on Collins asymmetries for identified π^+ production in representative (x, Q^2) intervals, shown as a function of z and transverse hadron momentum. The projections assume an integrated luminosity of $10\ \mathrm{fb}^{-1}$. Reproduced from [22].

[4] can be reconstructed through weak decays such as

$$\Lambda \to p\pi^-, \qquad \bar{\Lambda} \to \bar{p}\pi^+. \tag{51}$$

Reliable identification of the daughter particles improves invariant-mass reconstruction and reduces combinatorial backgrounds. Such channels also illustrate why the low-momentum reach of the PID system is important: decay daughters can have substantially smaller momenta than the parent hadron.

Heavy flavour extends this flavour-tagging strategy to charm and beauty. Charm production provides an important probe of gluon distributions, while reconstructed heavy-flavour hadrons allow the propagation and hadronization of heavy quarks to be studied in both ep and eA collisions. The characteristic displaced decay vertices provide strong topological discrimination, but particle identification of the decay products remains an crucial part of the reconstruction [22].

A representative channel is

$$D^0 \to K^-\pi^+, \tag{52}$$

for which oppositely charged pion and kaon candidates are combined after particle identification. Vertex displacement, decay length, daughter-track impact parameters and pointing information can then be used to suppress the remaining combinatorial background. The effect is illustrated in Figure 36, where PID is combined with topological selections for reconstructed D^0 candidates in both ep and eA collisions [22]. The heavy-flavour programme also provides direct sensitivity to hadronization in cold nuclear matter. Comparing heavy-flavour production in ep and eA collisions can distinguish different mechanisms of parton energy loss and hadron formation, complementing analogous measurements with light hadrons [22, 18].

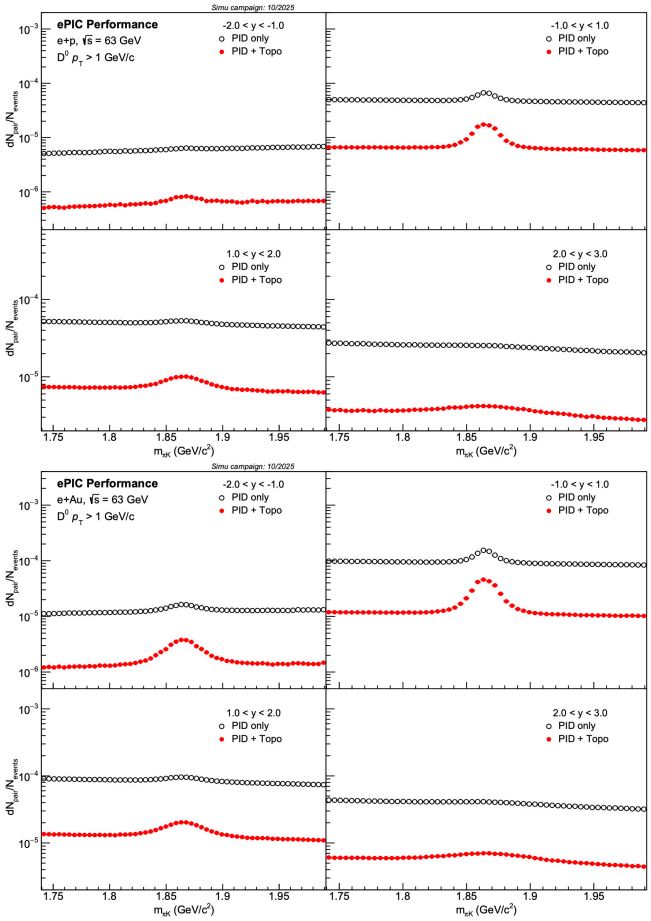


Figure 36: Reconstructed $K\pi$ invariant-mass distributions for D^0 candidates in different rapidity intervals. Results using particle identification alone are compared with those obtained after combining PID with topological selections in simulated ep and eA collisions. Image from [22].

2.2.5 Exclusive final states

In exclusive reactions the function of particle identification differs from its role in SIDIS. Rather than statistically enhancing the correlation between a measured hadron and the flavour of a struck parton, PID contributes directly to the reconstruction of a specific final state. Tracking, calorimetry, PID and far-forward instrumentation must together establish that the measured particles are compatible with the exclusive reaction under study [22].

Exclusive photon and meson production provides access to generalized parton distributions and therefore to correlations between longitudinal parton momentum and transverse spatial position. Different final states probe different combinations of quark and gluon distributions. Vector meson channels are particularly important for gluon imaging, including production of ϕ, J/ψ and Υ mesons [22, 18].

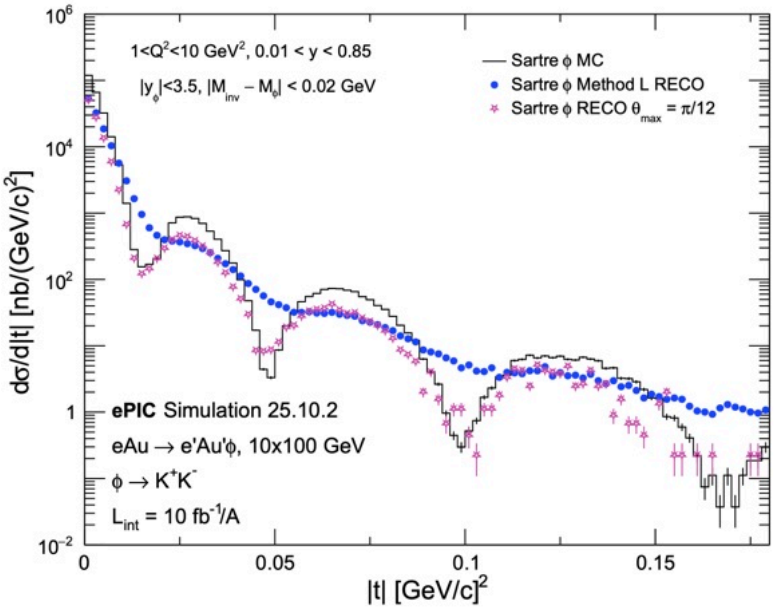


Figure 37: Differential distribution in momentum transfer $|t|$ for coherent $\phi \to K^+K^-$ electroproduction in electron-gold collisions. The reconstructed distribution is obtained from full ePIC simulation. Reproduced from [22].

The role of charged-hadron identification can be illustrated by exclusive ϕ production,

$$e + A \to e' + A' + \phi, \qquad \phi \to K^+K^-. \tag{53}$$

The two kaons must be reconstructed accurately in order to identify the ϕ candidate and determine the meson four-momentum. In nuclear collisions, the coherent channel must additionally be separated from incoherent nuclear breakup. This requires information from the far-forward B0, Roman Pot, Off-Momentum and Zero Degree detectors in addition to the central final-state reconstruction [22].

Figure 37 shows a full ePIC simulation of coherent ϕ electroproduction in electron–gold collisions. The reconstructed momentum-transfer distribution retains the diffractive pattern needed for spatial gluon imaging when the kinematic reconstruction and forward veto information are combined.

Particle identification in exclusive physics must therefore be considered as part of the overall event-completeness requirement. Incorrect particle assignments can transfer events between physically different channels, whereas precise species identification allows specific exclusive reactions to be isolated and interpreted.

2.3 Particle-identification requirements

The physics motivations described in Sec. 2.2 translate into specific detector requirements to fit into an asymmetric collider geometry. Maximum predicted momenta change by more than an

order of magnitude between the electron-going and hadron-going directions, while the available detector volume and magnetic-field configuration also vary with pseudorapidity. A single PID technology is therefore not sufficient to cover the complete ePIC acceptance [22].

The current ePIC baseline combines Cherenkov imaging with precision time-of-flight measurements. The proximity-focusing RICH (pfRICH) covers the electron-going endcap, the high-performance DIRC (hpDIRC) covers the barrel, and the dual-radiator RICH (dRICH) provides high-momentum identification in the hadron-going endcap. AC-LGAD timing layers in the barrel and forward endcap extend the separation toward low momentum [22].

The detailed global PID reconstruction is still under development. Current full-detector physics studies therefore use subsystem-dependent simulated PID efficiencies and misidentification probabilities. These provide a realistic estimate of the physics performance, while the final combination of overlapping detector responses will be developed further [22].

2.3.1 Momentum and pseudorapidity distributions

The kinematic distribution of final-state particles determines the required PID coverage. Pseudorapidity is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{54}$$

where θ is the polar angle of the particle momentum measured with respect to the hadron-beam direction. Positive pseudorapidity is defined to correspond to the hadron-going direction and negative pseudorapidity to the electron-going direction.

Because the beam energies are asymmetric, the centre-of-mass system is longitudinally boosted in the hadron direction. Charged particles produced at positive pseudorapidity can therefore have laboratory momenta extending to several tens of GeV/c, whereas particles in the electron-going region are typically considerably softer. At the same time, large populations of low-momentum hadrons are present throughout the central detector. This combination creates a PID coverage requirement spanning from a few hundred MeV/c to approximately $50\ \mathrm{GeV}/c$ [22].

The current simulated momentum distribution and the corresponding ePIC PID coverage are shown in Figure 38. The underlying histogram represents charged hadrons generated in $18\times275\ \mathrm{GeV}$ ep collisions, while the overlaid contours show the regions in which the individual PID systems provide approximately 3σ separation (definition in Eq. 35) for π/K, K/p and e/π hypotheses [22]. The low-momentum part of this phase space is particularly relevant for decay products of strange and heavy-flavour hadrons and for light-vector-meson reconstruction. It is also important for transverse-momentum-dependent measurements, in which the transverse hadron momentum is one of the principal observables. Conversely, the high-momentum region is critical for forward SIDIS and high-z fragmentation, where the detected hadron retains a large fraction of the fragmenting parton's momentum [22].

The strong pseudorapidity dependence of the momentum spectrum is therefore one of the main reasons for the asymmetric PID architecture. High-momentum Cherenkov identification is indispensable in the hadron-going endcap, whereas lower-momentum techniques play a proportionally larger role in the barrel and electron-going regions.

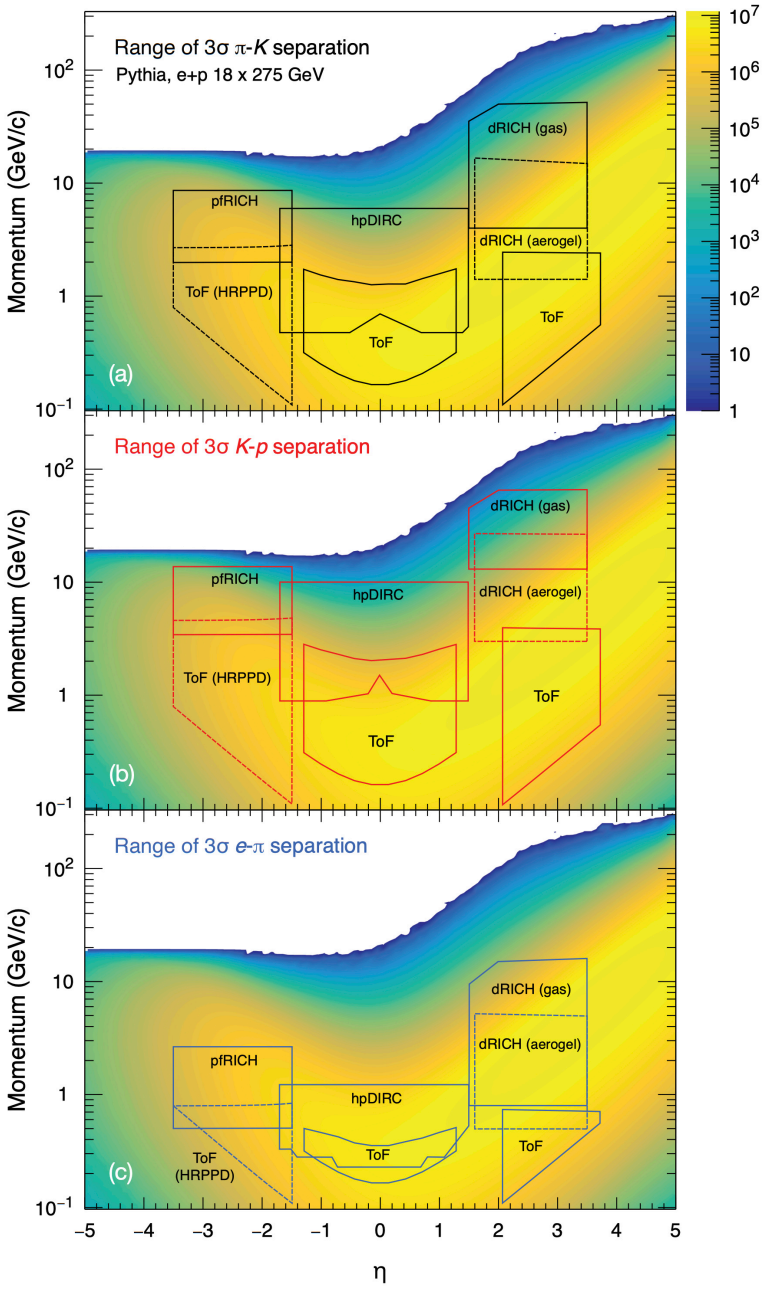


Figure 38: Charged-hadron momentum as a function of pseudorapidity for simulated 18×275 GeV ep collisions. The contours indicate the approximate 3σ separation regions provided by the pfRICH, hpDIRC, dRICH and time-of-flight systems for π/K, K/p and e/π identification. Reproduced from Figure 3.90 of [22].

Table 1: Representative charged-hadron PID coverage of the current ePIC baseline [22]. The momentum values correspond to the π/K performance benchmarks of the individual subsystems.

Subsystem	Acceptance	Momemntum range
pfRICH	$-3.5 < \eta < -1.5$	$p < 7\ \mathrm{GeV}/c$
BTOF	$-1.2 < \eta < 1.6$	$p_T < 1.2\ \mathrm{GeV}/c$
hpDIRC	$\lvert\eta\rvert < 1.57$	$p < 6\ \mathrm{GeV}/c$
FTOF	$1.9 < \eta < 3.6$	$p < 2.5\ \mathrm{GeV}/c$
dRICH	$1.5 < \eta < 3.5$	hadron ID $3 < p < 50\ \mathrm{GeV}/c$

2.3.2 Required $\pi/K/p$ separation

The ultimate goal of charged-hadron PID is the discrimination between pions, kaons and protons. These particles are sufficiently long lived to reach the dedicated PID detectors and constitute the identified-hadron samples used in most SIDIS, fragmentation and heavy-flavour analyses. A separation of approximately 3σ (defined as Eq.35) is adopted as the main performance benchmark for identified charged hadrons [22].

The required momentum reach is strongly dependent on pseudorapidity. For the present ePIC baseline, the π/K separation requirements and the complementary timing ranges can be summarized as follows:

The π/K separation is particularly demanding because pions are substantially more abundant than kaons over much of the EIC phase space. Consequently, even a percent-level probability for a pion to be misidentified as a kaon can produce a non-negligible contamination of the kaon sample, as observed in the full-detector SIDIS simulations at $18 \times 275\ \mathrm{GeV}$ [22]. Indeed, the relative pion contamination of an identified-kaon sample scales approximately as

$$\frac{N_{\pi\to K}}{N_{K\to K}} \simeq \frac{N_\pi}{N_K}\frac{\varepsilon_{\pi\to K}}{\varepsilon_K}, \tag{55}$$

so that the large N_π/N_K ratio amplifies even a small pion misidentification probability.

This effect is visible in the full-simulation PID studies shown in Figure 33. Consequently, efficiency alone does not characterize PID performance: sample purity and misidentification probability must also be controlled [22].

The upper momentum limit should not be interpreted as the point at which tracking ceases. Tracks can be reconstructed well beyond the PID separation boundary, but their species can no longer be assigned with sufficient confidence for flavour-sensitive measurements. The useful identified-hadron phase space is therefore the intersection between tracking acceptance and PID acceptance.

At low momentum a complementary limitation arises from magnetic bending and detector geometry: sufficiently low transverse-momentum particles may curl before reaching an outer Cherenkov detector. The inner timing layers are therefore particularly important because they identify particles that either do not reach the outer Cherenkov system or have momenta below its Cherenkov threshold [22].

2.3.3 Different requirements in the electron endcap, barrel and hadron endcap

The asymmetric collision kinematics lead to three qualitatively different PID regions. The electron-going endcap is where the scattered DIS electron is detected together with comparatively low-momentum hadrons; the barrel is where a large population of intermediate-momentum particles is reconstructed within a geometrically constrained volume; and the hadron-going endcap is where the most energetic hadrons produced at the collider are observed.

In the *electron-going endcap*, approximately

$$-3.5 < \eta < -1.5,$$

the identification of the scattered electron is a major requirement, but charged-hadron identification remains important for SIDIS and exclusive reactions. The pfRICH uses an aerogel radiator and HRPPD photosensors [27] to provide approximately 3σ π/K separation up to $7\ \mathrm{GeV}/c$. Its fast HRPPD response also provides time-of-flight information at low momentum and a timing reference of order $20\ \mathrm{ps}$ for the dedicated barrel and forward timing systems [22].

The pfRICH also contributes directly to electron-pion separation in the kinematic region in which calorimetry alone is insufficient. Its design is therefore required to combine hadron identification with low material budget, since the backward electromagnetic calorimeter is located immediately downstream.

The *barrel* occupies the central rapidity region and is subject to particularly restrictive radial detector integration constraints. High-momentum identification is provided by the hpDIRC, which uses synthetic fused-silica bars simultaneously as Cherenkov radiators and photon light guides. The current requirement is at least 3σ π/K separation for $p < 6\ \mathrm{GeV}/c$ over its acceptance [22].

The hpDIRC is complemented at smaller radius by the Barrel Time-of-Flight detector (BTOF). The BTOF uses AC-LGAD strip sensors [28] with a single-hit timing resolution of approximately $35\ \mathrm{ps}$. It extends the π/K separation to low momentum, approximately $p_T < 1.2\ \mathrm{GeV}/c$ over $-1.2 < \eta < 1.6$, and provides an additional high-precision space point for charged-particle tracking [22].

In the *hadron-going endcap*, the longitudinal boost produces much larger particle momenta. The dRICH consequently has the most demanding requirement and has to provide charged-hadron identification in the momentum range $3 < p < 50\ \mathrm{GeV}/c$ over $1.5 \leq \eta \leq 3.5$. The highest-momentum requirement occurs in the most forward part of the acceptance, where the gas radiator must retain approximately 3σ π/K separation up to $50\ \mathrm{GeV}/c$ [22].

Low-momentum forward identification is provided by the Forward Time-of-Flight detector (FTOF), based on pixelated AC-LGAD sensors. Its single-hit timing resolution is approximately $25\ \mathrm{ps}$ and provides 3σ π/K separation for $p < 2.5\ \mathrm{GeV}/c$ for $1.9 < \eta < 3.6$ [22].

2.3.4 Complementarity of timing and dRICH

The broad momentum range required for ePIC cannot be covered efficiently by a single particle-identification technique. As discussed in Sec. 2.1, both time-of-flight and Cherenkov detectors determine the particle velocity and, when combined with the momentum measured by the tracking system, allow different mass hypotheses to be distinguished. Their separation power,

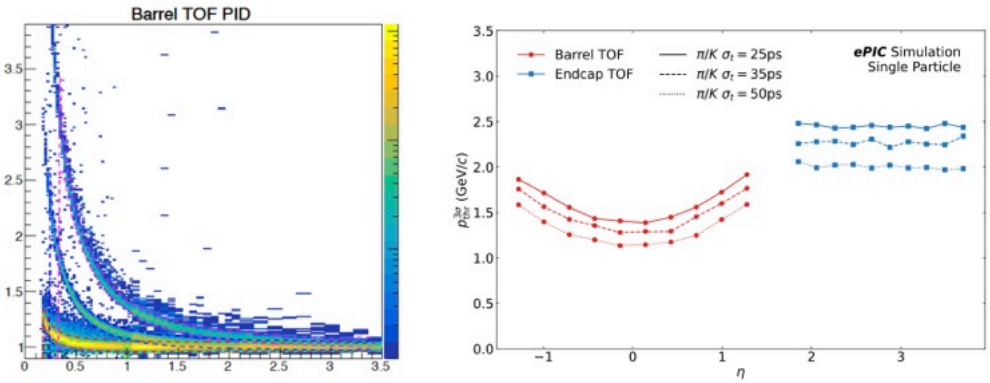


Figure 39: Simulated performance of the ePIC time-of-flight system. Left: BTOF $1/\beta$ as a function of momentum (p) in the simulation with PYTHIA DIS events. Right: the approximate upper momentum limits for 3σ particle separation as a function of pseudorapidity for the barrel and forward AC-LGAD timing layers. Image from [22].

however, evolves differently with momentum. Time-of-flight measurements are most effective at low momentum, where the velocity differences between pions, kaons and protons are comparatively large, while their discriminating power decreases as the particles become increasingly relativistic. The ePIC AC-LGAD timing layers are designed to cover this low-momentum region: the barrel and forward TOF systems achieve single-hit timing resolutions of approximately 35 ps and 25 ps, respectively, with the corresponding π/K momentum coverage shown in Figure 39 [22].

At higher momenta, Cherenkov detectors provide the required continuation of the PID coverage. As described in Sec. 2.1, the momentum range over which a Cherenkov detector is effective depends strongly on the refractive index of its radiator. Aerogel, with its comparatively high refractive index, typically $n \sim 1.02$–1.04, provides sensitivity from relatively low momentum, whereas low-index gaseous radiators, typically characterized by $n \sim 1.001$, have higher Cherenkov thresholds but preserve useful angular separation between particle species up to much larger momenta. The combination of timing and Cherenkov techniques therefore provides continuous and partially overlapping hadron identification over the wide momentum range required by the ePIC physics programme.

The forward dRICH exploits this complementarity within a single detector. Its aerogel has a reference refractive index of approximately $n = 1.026$, while the baseline gas radiator is C_2F_6 with $n \simeq 1.00086$ at standard conditions [22]. The aerogel extends identification from the end of the FTOF range at approximately 2.5 GeV$/c$ and overlaps the gas response above its kaon threshold, approximately 12 GeV$/c$. The gas radiator then maintains separation up to the highest required momenta.

This interplay is illustrated in Figure 40, where reconstructed mass distributions obtained independently from the aerogel and gas radiators demonstrate the overlap between the two regimes.

The resulting PID strategy therefore follows a continuous sequence in momentum. Precision

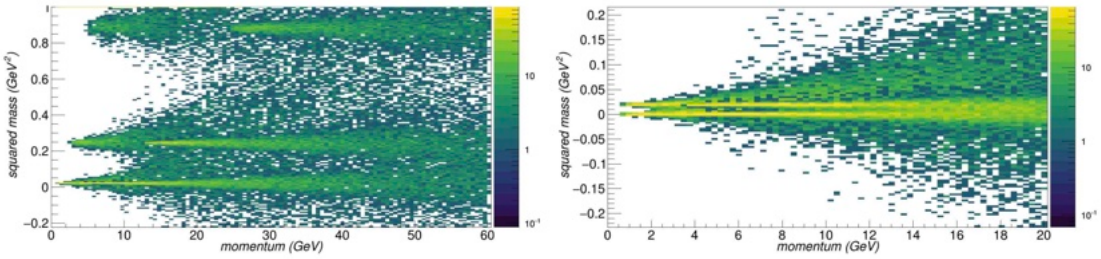


Figure 40: Reconstructed particle mass as a function of momentum in the forward dRICH for pions, kaons and protons (left) and for pions and electrons (right). Results obtained from the aerogel and gas radiators illustrate their complementary momentum coverage. The simulation corresponds to $3.2 < \eta < 3.5$. Reproduced from [22].

timing provides strong separation while the species have measurably different velocities; aerogel Cherenkov detectors extend identification beyond the useful ToF range; and low-index gas Cherenkov radiation preserves separation when the particles become highly relativistic. Overlapping regions between the different techniques provide continuity of the physics acceptance and offer valuable control of particle-identification efficiencies and misidentification probabilities.

The detailed implementation of the pfRICH, hpDIRC, AC-LGAD timing layers and dRICH is described in the following section.

2.4 ePIC PID systems

The broad momentum and pseudorapidity coverage discussed in Sec. 2.3 is obtained in ePIC through a combination of Cherenkov-imaging and precision time-of-flight detectors. The choice of technology changes across the detector because the momentum spectrum, available integration volume and magnetic-field conditions differ substantially between the electron-going, barrel and hadron-going regions. The resulting architecture consists of a proximity-focusing RICH in the backward endcap, a high-performance DIRC in the barrel, AC-LGAD time-of-flight layers in the barrel and forward endcap, and a dual-radiator RICH in the forward region [22].

These detectors have deliberately overlapping acceptances. At low momentum, timing provides the strongest velocity discrimination, whereas at larger momentum the Cherenkov angle becomes the principal observable. The overlap between neighbouring systems prevents large gaps in particle-identification capability and provides redundant information that can be used to control efficiencies and misidentification probabilities. The following subsections summarize the operating principle and present baseline design of each subsystem.

2.4.1 Backward proximity-focusing RICH

Particle identification in the electron-going endcap is provided by the *proximity-focusing Ring Imaging Cherenkov detector* (pfRICH). The detector covers approximately

$$-3.5 < \eta < -1.5$$

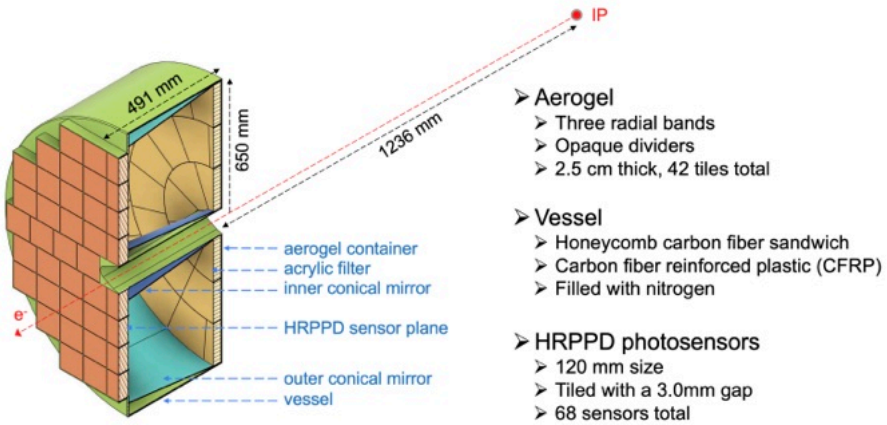


Figure 41: Baseline layout of the ePIC proximity-focusing RICH in the electron-going endcap. The aerogel radiator, expansion volume, mirror elements and HRPPD sensor plane are indicated. Reproduced from [22].

and is designed to provide at least 3σ pion-kaon separation up to about $7\ \mathrm{GeV}/c$ across this acceptance [22]. In addition to hadron identification, the pfRICH contributes to the rejection of the large low-momentum pion background accompanying the scattered DIS electron.

The detector employs a $2.5\ \mathrm{cm}$-thick silica-aerogel radiator with a nominal refractive index of $n \simeq 1.040$, followed by an extended proximity gap of about $40\ \mathrm{cm}$, as determined by the separation between the aerogel wall and the HRPPD photon-detection plane [22]. When a charged particle crosses the aerogel above the corresponding Cherenkov threshold, the emitted photons propagate through the expansion region before reaching the photon-detection plane. In a proximity-focusing geometry no conventional focusing lens is required to form the ring: the transverse position of the photons at the sensor plane is determined primarily by their Cherenkov emission angle and by the length of the expansion gap. Increasing this gap improves the geometrical separation between different Cherenkov angles, although it must be balanced against the finite dimensions of the detector [22].

The baseline geometry is shown in Figure 41. The vessel has a diameter of approximately $1.3\ \mathrm{m}$ and a longitudinal length of about $49\ \mathrm{cm}$. The radiator plane contains 42 trapezoidal aerogel tiles, approximately $2.5\ \mathrm{cm}$ thick, arranged around the beam-pipe opening. A thin acrylic filter downstream of the aerogel suppresses the short-wavelength part of the spectrum, for which chromatic dispersion and Rayleigh scattering would otherwise degrade the single-photon angular resolution. The vessel is continuously flushed with dry purified nitrogen [22].

The photon plane is instrumented with 68 High Rate Picosecond Photon Detectors (HRPPDs), large-area microchannel-plate photomultipliers specifically developed for high-rate single-photon detection. Each device has an active area of about $104 \times 104\ \mathrm{mm}^2$ segmented into a 32×32 pixel matrix. Their fast response is important not only for Cherenkov imaging but also for timing: the pfRICH can provide additional time-of-flight information at low momentum and a reference time of order $20\ \mathrm{ps}$ for the dedicated barrel and forward ToF detectors [22].

The cylindrical geometry also incorporates inner and outer mirrors. These are not used to

focus the complete Cherenkov image in the same sense as in a focusing RICH, but recover photons that would otherwise escape through the lateral boundaries of the proximity volume. In this way the effective acceptance can be extended toward both edges of the backward pseudorapidity interval.

A further constraint is imposed by the Electron-Endcap Electromagnetic Calorimeter, which is located immediately downstream. The pfRICH must therefore achieve the required photon yield and angular resolution while maintaining a low material budget, limiting the degradation of the scattered-electron energy measurement [22].

2.4.2 Barrel DIRC

The barrel region is instrumented with the *high-performance Detection of Internally Reflected Cherenkov light* (hpDIRC), which provides charged-hadron identification over approximately $|\eta| \leq 1.57$. The main performance requirement is at least 3σ π/K separation up to approximately $6\ \mathrm{GeV}/c$ [22].

The choice of a DIRC architecture is particularly well suited to the barrel because it provides Cherenkov imaging with an active radiator thickness of only about $17\ \mathrm{mm}$ in the radial direction, fitting within the compact ePIC central detector. The DIRC principle uses long bars of synthetic fused silica simultaneously (refractive index $n \sim 1.47$) as the Cherenkov radiator and as photon light guides, with the Cherenkov photons transported along the bars by total internal reflection [22].

Cherenkov photons produced by a charged particle are trapped inside a rectangular bar by total internal reflection. Provided that the optical surfaces have sufficiently high quality, this propagation preserves the angular information carried by the photons. The Cherenkov angle can therefore be reconstructed after the photons have travelled to the end of the bar, allowing most of the expansion volume and readout equipment to be displaced away from the central tracking acceptance [22].

The hpDIRC baseline geometry, shown in Figure 42, consists of twelve optically isolated sectors arranged around the beam line. Each sector contains a bar box and a corresponding readout box, forming an approximately twelve-sided polygonal barrel with an inner radius of about $760\ \mathrm{mm}$. Each bar box contains ten long fused-silica radiator bars, giving a total of 120 bars. The radiator length is approximately $4.56\ \mathrm{m}$ [22]. A mirror at one end of each radiator bar reflects the photons toward the readout side. At the opposite end, a three-layer spherical lens focuses the light onto a synthetic-fused-silica prism that acts as the expansion volume. The focusing optics are an important distinction between the hpDIRC and the original DIRC concept [29]: they reduce optical aberrations and improve the reconstruction of the Cherenkov angle while preserving the radial compactness of the detector [22].

The rear surface of each prism is instrumented with pixelated single-photon detectors. The present baseline uses lifetime-enhanced MCP-PMTs[30], with an effective pixel size of approximately $3.2\ \mathrm{mm}$. Both photon position and arrival time are used in the reconstruction. In particular, time-based imaging compares the measured photon arrival times with the distributions expected under different particle hypotheses, complementing the geometrical information from the photon impact positions [22].

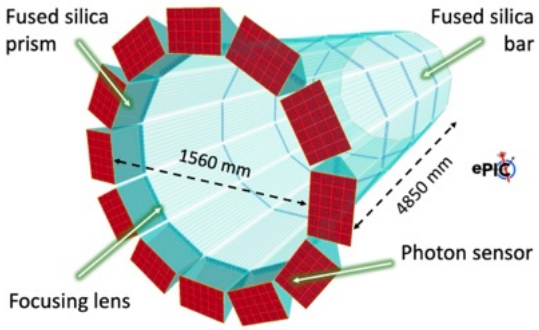


Figure 42: Baseline geometry of the ePIC high-performance DIRC. Cherenkov photons generated in the fused-silica bars are transported by total internal reflection toward the focusing optics, expansion prisms and photon-sensor arrays. Image from [22].

The final PID performance depends not only on the intrinsic optical resolution of the hpDIRC but also on the accuracy with which the charged-particle direction is known when the track enters the radiator. For π/K separation near the upper momentum limit, a track angular resolution of order $0.5\ \mathrm{mrad}$ is required. The outer barrel tracking layers immediately upstream of the hpDIRC therefore form an important part of the complete PID measurement [22].

2.4.3 Barrel and forward TOF

The Cherenkov systems are complemented at low momentum by two dedicated time-of-flight detectors: the Barrel Time-of-Flight detector (BTOF) and the Forward Time-of-Flight detector (FTOF). Both employ *AC-coupled Low-Gain Avalanche Diodes* (AC-LGADs), thin silicon sensors that combine internal charge multiplication with fast signal formation and fine spatial segmentation [22].

The AC-LGAD structure contains an internal gain layer that amplifies the charge generated by a traversing particle. The induced signal is shared through a resistive layer and capacitively coupled to segmented metal electrodes on the sensor surface. This architecture provides precise time measurements while preserving accurate position information. The ToF layers consequently have a dual role within ePIC: they extend particle identification below the useful momentum range of the Cherenkov systems and simultaneously provide additional high-precision points for charged-particle tracking [22].

The BTOF and FTOF geometries are illustrated in Figure 43. Different segmentation is adopted in the two regions because of their different occupancies, integration constraints and tracking requirements. The BTOF uses strip-like readout elements with a characteristic segmentation of approximately $0.5 \times 10\ \mathrm{mm}^2$, whereas the FTOF employs $0.5 \times 0.5\ \mathrm{mm}^2$ pixels [22].

The BTOF consists of 144 longitudinal staves positioned at a radius of approximately $65\ \mathrm{cm}$ and extending over a total length of about $2.6\ \mathrm{m}$. The strip geometry reduces the required number of electronic channels while maintaining fine resolution in the azimuthal coordinate. The nominal single-hit time resolution is approximately $35\ \mathrm{ps}$, providing about $3\sigma\ \pi/K$ discrimination up to $p_T \simeq 1.2\ \mathrm{GeV}/c$ over approximately $-1.2 < \eta < 1.6$ [22].

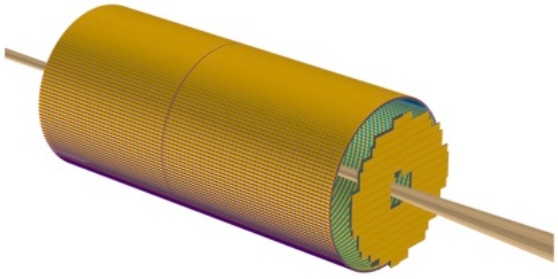


Figure 43: Geometry of the ePIC AC-LGAD time-of-flight systems, with the cylindrical barrel timing layer and the pixelated forward endcap detector. Reproduced from [22].

The FTOF is installed as a disk in the hadron-going endcap, approximately $1.9\ \mathrm{m}$ from the interaction point. Its pixelated geometry is better suited to the higher track density and forward incidence angles characteristic of this region. The nominal single-hit time resolution is approximately $25\ \mathrm{ps}$ and extends $3\sigma\ \pi/K$ discrimination to momenta of about $2.5\ \mathrm{GeV}/c$ over approximately $1.9 < \eta < 3.6$ [22].

The different implementations therefore serve the same physical purpose but are optimized for their local detector environments. In the barrel, BTOF identifies low-momentum particles that may not reach the larger-radius hpDIRC or that remain below its useful Cherenkov range. In the forward endcap, FTOF provides the low-momentum continuation of the dRICH acceptance. Together they remove the principal low-momentum gaps that would remain in a Cherenkov-only PID system.

2.4.4 Forward dRICH

The highest-momentum charged hadrons are concentrated in the hadron-going direction, where particle identification is provided by the *dual-radiator Ring Imaging Cherenkov detector* (dRICH). It covers approximately

$$1.5 \leq \eta \leq 3.5$$

and is designed to extend hadron identification from the upper reach of the FTOF to approximately $50\ \mathrm{GeV}/c$ [22]. This large dynamic range cannot be achieved efficiently with a single Cherenkov radiator.

The detector therefore combines a solid aerogel radiator, optimized for the intermediate-momentum region, with a low-refractive-index gas radiator that preserves pion-kaon angular separation to the highest momenta. Cherenkov photons from both radiators are transported by a common focusing mirror system toward photon-detection surfaces located outside the direct charged-particle acceptance. This arrangement provides the required momentum reach while remaining compatible with the limited longitudinal space of the forward endcap [22].

The forward environment introduces additional constraints absent, or less severe, in the other PID systems. The photon detectors operate in a non-negligible and non-uniform magnetic field, making conventional photomultiplier solutions impractical in this region. Silicon photomultipliers are therefore employed for single-photon detection. Their compact dimen-

sions also allow the focal surface to be instrumented with the non-planar geometry required by the optical system [22].

Only this general role within the global PID architecture is introduced here. The choice and optimization of the two radiators, the common focusing optics, the SiPM photon-detection plane, the front-end electronics, readout strategy, radiation environment and geometrical constraints are discussed in detail in Sec. 2.5.

2.5 Dual-radiator RICH detector

2.5.1 Detector principle

In the forward (hadron-going) end-cap of ePIC, particle identification is provided by a dual-radiator ring-imaging Cherenkov detector (dRICH), complemented by the forward time-of-flight (FTOF) system that covers the low-momentum region up to approximately $2.5\,\mathrm{GeV/c}$. To extend the particle-identification capability up to momenta of about $50\,\mathrm{GeV/c}$, the dRICH employs two radiators simultaneously, namely aerogel and gas, as shown in Figure 44. Given the practically achievable refractive indices, minimum radiator thicknesses of $4\,\mathrm{cm}$ for the aerogel and $1\,\mathrm{m}$ for the gas are required to ensure an adequate photon yield [31]. A focusing mirror system is therefore needed to preserve accurate ring imaging despite the uncertainty associated with the photon emission point. To maintain a compact and cost-effective detector layout, a common imaging system is shared by both radiators.

The dRICH must provide continuous acceptance over the forward ePIC pseudorapidity range $1.5 \leq \eta \leq 3.5$ [3], extending from the beam pipe to the boundaries of the detector barrel. The mirror system focuses the Cherenkov photons onto six active sensor areas positioned within the shadow of the barrel detector and its support ring, close to the ePIC solenoid coils. In this region, the magnetic field is non-uniform and reaches values of up to $0.6\,\mathrm{T}$, favouring the use of silicon photomultipliers (SiPMs), whose operation is essentially unaffected by magnetic fields. Their implementation in the dRICH for single-photon detection will represent the first use of SiPMs for Cherenkov imaging in a collider experiment. To satisfy the demanding requirements of ePIC, the dRICH incorporates several innovative technological solutions. Among the most relevant developments discussed here are aerogel tiles with a refractive index of $n = 1.026$, a C_2F_6 gas radiator, SiPM sensors equipped with in-situ annealing, lightweight carbon-fiber mirrors with an optimized core structure, a curved photon-detection surface designed to match the mirror focal region, dedicated ALCOR electronics for precision timing and digitization, and a streaming readout system supported by machine-learning-assisted online data filtering.

2.5.1.1 Aerogel and gas radiators

Aerogel radiator Initial simulations of the dRICH performance indicated an optimal aerogel refractive index of $n = 1.020$ [1]. During the $R\&D$ phase, several $2\,\mathrm{cm}$-thick samples produced by Aerogel Factory (Japan), with refractive indices ranging from $n = 1.015$ to $n = 1.030$, were characterized through laboratory measurements [31]. The samples were subsequently tested using detector prototypes at the CERN test-beam facility, also in collaboration with ALICE3 [32].

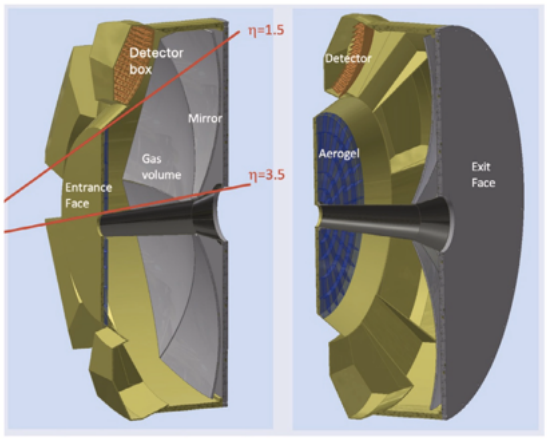


Figure 44: dRICH detector model with major components highlighted [31].

The measurements showed that the aerogel transparency is typically greater than 70% at a wavelength of $400\,\mathrm{nm}$. In addition, the optical quality was found to improve with increasing refractive index, reaching its highest value at $n = 1.030$, consistently with trends observed in previous developments. Following an optimization study based on the comparison between experimental data and simulations, a baseline refractive index of $n = 1.026$ was therefore selected [31].

Compared with the initially proposed value, this choice improves the overlap with the FTOF system by reducing the kaon momentum threshold to approximately $2.1\,\mathrm{GeV/c}$. At the same time, it maintains a momentum reach well above the kaon threshold of the gas radiator, which is approximately $12\,\mathrm{GeV/c}$. This is made possible by the larger photon yield, which compensates for the reduced separation between the Cherenkov angles of different particle species. During the $R\&D$ phase, the chromatic dispersion, $dn/d\lambda$, was measured to be $6 \times 10^{-6}, \mathrm{nm}^{-1}$ at a wavelength of $400\,\mathrm{nm}$ [31].

An engineering study is currently being conducted to determine the maximum dimensions and optimal geometry of the aerogel tiles. The aim is to cover the dRICH entrance window, corresponding to a disk with a radius of $90\,\mathrm{cm}$ surrounding the central beam-pipe envelope, while minimizing the inactive area. Tiles with a reference maximum size of $18\,\mathrm{cm}$, previously achieved for the BELLE II experiment, have been successfully produced and cut using a water-jet technique in both square and trapezoidal geometries, as shown in the top-left panel of Figure 45. A feasibility study is ongoing to extend the tile dimensions to $20\,\mathrm{cm}$ or more, thereby reducing dead regions and maximizing the detectable photon yield.

Gas radiator The reference gas selected for the dRICH radiator is hexafluoroethane, C_2F_6, whose optical properties satisfy the detector requirements. At standard temperature and pressure (STP), it is characterized by a refractive index of $n = 1.00086$ and a low chromatic dispersion of $dn/d\lambda = 0.2 \times 10^{-6}, \mathrm{nm}^{-1}$ at a wavelength of $\lambda = 350\,\mathrm{nm}$. The performance of C_2F_6 in terms of photon yield and Cherenkov-angle resolution was validated using the dRICH prototype at the CERN test-beam facility [33]. With an optical arm comparable to that foreseen for ePIC, a

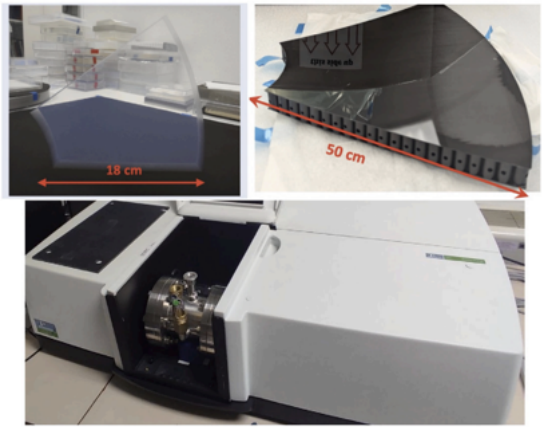


Figure 45: (Top left) Example of aerogel tile produced with the dRICH baseline refractive index $n = 1.026$ and the reference dimension of BELLE-II. (Top-right) Mid-size CFRP mirror demonstrator with optimized core structure. (Bottom) Photo-spectrometer with compact high-pressure chamber for gas transparency measurements. [31].

single-photoelectron (SPE) angular resolution close to $1\,\mathrm{mrad}$ was achieved, consistently with expectations.

The dRICH will operate at atmospheric pressure, using standard techniques to minimize pressure gradients across the detector. Particular attention must therefore be devoted to the gas-handling system, which must allow efficient purging and separation of the radiator gas while limiting losses. This requirement is especially important because C_2F_6 is a greenhouse gas with a high global-warming potential. Its physical properties present both advantages and technical challenges. In particular, its boiling point of $-78.1\,^{\circ}\mathrm{C}$ ensures that it remains in a single gaseous phase under all operating conditions, while making recovery through liquefaction more difficult [31].

One option currently under investigation is the use of CO_2 as a stand-by gas during maintenance operations. Since CO_2 can be liquefied before C_2F_6 at manageable pressures of approximately $10\,\mathrm{bar}$, the two gases could potentially be separated through controlled liquefaction. A more promising solution involves the use of selective membranes that exploit the low permeability of the relatively large fluorocarbon molecule. This approach follows a pioneering study performed by LHCb on the separation of CF_4 and CO_2 using C0–C10 membranes manufactured by UBE Industries. Separation tests involving C_2F_6 are currently being carried out at CERN with the support of gas-system experts.

The transparency of C_2F_6 was measured at the COMPASS monochromator facility after approximately four years of storage in gas bottles. Values greater than 98% were obtained in the near-ultraviolet wavelength range from $170\,\mathrm{nm}$ to $220\,\mathrm{nm}$, which is particularly sensitive to the presence of contaminants. A dedicated system has also been developed to measure the gas transparency over the broader wavelength range from $200\,\mathrm{nm}$ to $900\,\mathrm{nm}$, which more closely matches the spectral response of the SiPM sensors. The setup is based on a commercial photospectrometer coupled to a compact high-pressure chamber, operating at pressures of up to

$10\,\mathrm{bar}$ and equipped with transparent quartz windows, as shown in the bottom panel of Figure 45. This apparatus is expected to become part of the continuous monitoring instrumentation installed in the high-pressure section of the ePIC dRICH gas system.

Measurements are also ongoing to quantify the scintillation yield of C_2F_6. Preliminary results indicate an emission peak at wavelengths around $300\,\mathrm{nm}$. The observed yield is approximately twice that of C_4F_{10}, while remaining about twenty times lower than that of CF_4, whose comparatively high scintillation emission can represent a significant source of background [31].

2.5.2 Common focusing system

The dRICH mirror array is designed to cover a spherical surface with a radius of $2200\,\mathrm{mm}$ and a total area of approximately $10\,\mathrm{m}^2$. Carbon-fiber-reinforced polymer (CFRP) technology has been selected as the baseline solution for the mirror substrate, as it provides the required mechanical stiffness while minimizing the detector material budget.

A medium-scale demonstrator with a diagonal dimension of $60\,\mathrm{cm}$ was manufactured by Composite Mirror Applications (Arizona, USA) according to the dRICH specifications, as shown in the top-right panel of Figure 45. The internal CFRP structure was optimized to preserve the accuracy of the mirror surface while limiting the overall mass to approximately $5\,\mathrm{kg/m}^2$. The design combines a lightweight C-shaped cellular structure, previously adopted by LHCb, in the central region with a more robust cylindrical-cell structure, derived from the CLAS12 design, along the outer edges.

As visible in Figure 45, the optical surface of the CFRP substrate already exhibits partial reflectivity even before the deposition of the Al+SiO_2 reflective coating. Prior to coating, the mirror was characterized using a point-like-source imaging test. The resulting D_0 parameter characterizes how tightly the reflected image of a point-like source is concentrated, and therefore provides an overall measure of the mirror optical quality: smaller values correspond to a sharper image and a better surface quality. The measured value, $D_0 = 1.8\,\mathrm{mm}$, was better than the required maximum of $2.5\,\mathrm{mm}$. The same measurement yielded a curvature radius of $(2254 \pm 5)\,\mathrm{mm}$, slightly above the target value but still close to the requirement that deviations remain within 1% of the nominal radius [31].

2.5.3 Photon detection plane

The dRICH photon-detection surface has a spherical geometry with a radius of approximately $110\,\mathrm{cm}$, closely matching the three-dimensional focal surface of the mirror system, as shown in Figure 44. Silicon photomultiplier (SiPM) technology has been selected for photon detection because it provides excellent single-photon counting performance within the ePIC magnetic field, while its compact dimensions allow the efficient tessellation of the curved active surface. Each SiPM sensor has an active area of $3 \times 3, \mathrm{mm}^2$, ensuring the required spatial resolution, together with an intrinsic time resolution better than $150\,\mathrm{ps}$.

Sec. 2.6 discusses in detail the motivation for selecting SiPMs for this application. It also presents the studies performed on their dark count rate, operating temperature, and response to radiation damage, together with the possible mitigation strategies (called *annealing*) considered to reduce radiation-induced damages.

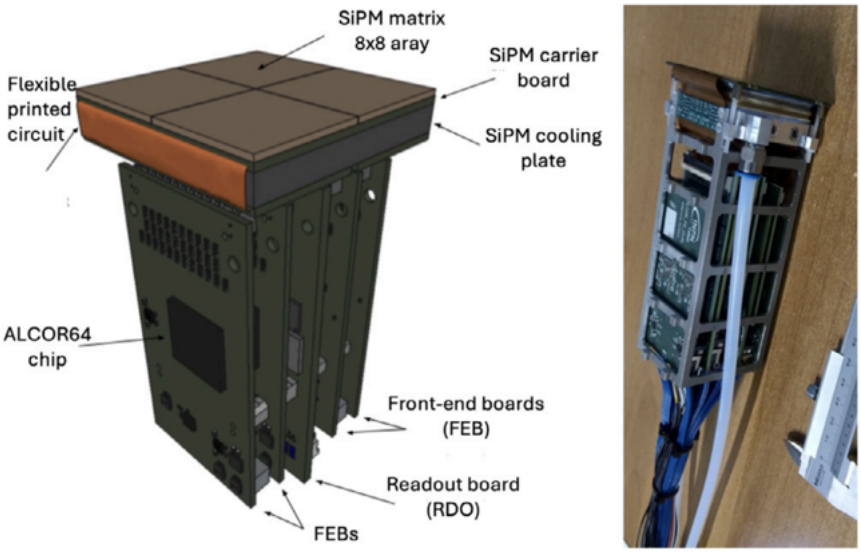


Figure 46: The PDU concept (left) and preliminary realization (right). [31].

ALCOR readout The selected front-end application-specific integrated circuit (ASIC) is ALCOR [34], a custom 32-channel chip whose input coupling and rate capability are optimized for SiPM operation. The chip adopts a threshold-based timestamp architecture with a time binning finer than $50\,\mathrm{ps}$. It also provides a measurement of the time-over-threshold (ToT) or of the signal slew rate, which can be used to correct time-walk effects. Tests have demonstrated that ALCOR can sustain rates exceeding $500\,\mathrm{kHz}$ per channel, corresponding to almost twice the reference limit required by ePIC [31]. The timing performance of the complete SiPM–ALCOR readout chain was evaluated using single-photon laser events. The measured resolution was fully compatible with the ePIC requirements, remaining comfortably below $150\,\mathrm{ps}$ RMS. Moreover, the timing resolution was found to improve for larger SPAD sizes, owing to the correspondingly larger signal amplitudes and faster slew rates. To minimize the volume occupied within the dRICH envelope while maximizing the active-area packing factor, the photon-detection system is organized into compact photodetector units (PDUs). Each PDU has approximate dimensions of $52 \times 52 \times 140\ mm^3$ and integrates the SiPM sensors, the front-end readout electronics, a cooling system for sub-zero-temperature operation, and the capability to perform in-situ high-temperature annealing, as shown in the left panel of Figure 46.

Each PDU comprises 256 SiPM channels read out by four front-end boards (FEBs), equipped with ALCOR TDCs, and a single readout board (RDO), which provides the interface to the ePIC data-acquisition (DAQ) and detector-control systems. A cooling block is placed in thermal contact with the rear side of the printed-circuit board hosting the SiPMs, referred to as the carrier board. The carrier board is connected to the readout electronics through flexible Kapton circuits. A low-viscosity silicone fluid circulates through the cooling block in a closed loop connected to an external thermostat. The fluid temperature can be regulated either to maintain the SiPMs at low operating temperature or to preheat the sensors during forward-bias annealing, thereby

reducing the current required for the procedure.

The RDO integrates the main control and data-handling functions, including configuration, data transfer, and scrubbing, within a board approximately the size of a credit card. A preliminary 32-channel version of ALCOR has been used for several years in SiPM characterization activities together with the ePIC-oriented readout system. A prototype PDU equipped with this version of the ASIC was successfully constructed and operated with dRICH prototypes at the CERN test-beam facility, as shown in the right panel of Figure 46. The results demonstrated ring-imaging and particle-identification capabilities consistent with the requirements of ePIC [35]. A new 64-channel version of the ALCOR ASIC, foreseen for integration in a ball-grid-array (BGA) package, has been under development together with an upgraded front-end board. The previous multi-project-wafer (MPW) production did not yield usable devices because of issues related to the under-bump metallization (UBM) and packaging process. The ASIC is therefore not yet available for characterization, and a new MPW submission is currently being prepared before the upgraded FEB can enter the full evaluation phase.

Streaming readout The ePIC data-acquisition system operates in streaming mode and employs the FELIX-155 card [36] as a data-aggregation module (DAM). The dRICH represents one of the main contributors to the overall ePIC data throughput, since the significant dark count rate (DCR) of the SiPMs generates a substantial continuous data stream whose magnitude increases with the accumulated radiation dose and integrated luminosity.

As discussed above, low-temperature operation and periodic annealing are foreseen to mitigate the radiation-induced increase of the SiPM dark rate. A value of approximately $300\,\mathrm{kHz}$ per channel is adopted in the present studies as a conservative operating reference point, rather than as a strict limit guaranteed by the annealing procedure. At this reference DCR, the unreduced data stream would amount to several terabits per second. The quoted maximum throughput of approximately $1.3\,\mathrm{Tbps}$ for the complete dRICH already assumes a reduction of the DCR-dominated data stream by a factor of order $\mathcal{O}(5)$ [31, 22].

Several complementary strategies are being considered to achieve this reduction. A shutter signal implemented in the front-end ASIC could restrict the readout to time windows associated with the EIC bunch crossings, which occur every $10\,\mathrm{ns}$. Additional timing or event-selection tags may also be derived from signals in other detector systems, for example from scintillators or calorimetric energy deposits. In the streaming, triggerless architecture of ePIC, these tags do not constitute a conventional hardware trigger that globally accepts or rejects an event; rather, they provide auxiliary information that can be used to selectively retain or discard buffered dRICH data.

In parallel, online data-reduction algorithms implemented in the dRICH DAMs are being investigated. These approaches exploit information from the dRICH itself, or from multiple subdetectors, to identify data dominated by background and with no relevant signal content. Machine-learning implementations of this selective readout strategy are currently under study, with preliminary results showing promising performance [37].

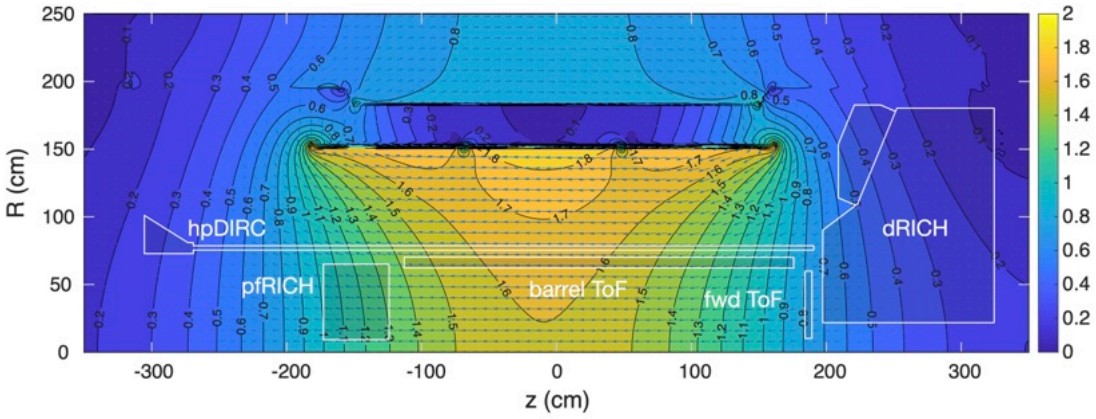


Figure 47: Magnetic-field map of the ePIC detector for the $1.7\,\mathrm{T}$ solenoid configuration, with the envelopes of the PID detectors superimposed. In the dRICH region the field is non-uniform and reaches values of approximately $0.3\,\mathrm{T}$–$0.6\,\mathrm{T}$. Image from [22].

2.5.4 Magnetic field and geometrical constraints

The integration of the dRICH in the hadron-going endcap is strongly constrained by both the available detector volume and the magnetic-field configuration. The detector must preserve an open acceptance over approximately $1.5 \leq \eta \leq 3.5$, while fitting inside the forward endcap and remaining compatible with the accelerator beam pipe and the far-forward detector acceptance. The resulting dRICH has a ring-shaped geometry, with a length of approximately $1.27\,\mathrm{m}$ and an outer diameter of about $3.6\,\mathrm{m}$, and is divided into six independent sectors around the beam line [22, 31].

An important geometrical constraint arises from the crossing configuration of the collider beams. The electron and hadron beams cross at an angle of $25\,\mathrm{mrad}$; moreover, the outgoing beam pipe must preserve acceptance for particles reaching the far-forward detectors. Consequently, its cross section is displaced from the detector axis at the dRICH location and increases with the distance from the interaction point. The innermost detector elements therefore cannot be arranged with a simple cylindrically symmetric geometry, and the radiator, mirror and support structures must accommodate an intrinsically asymmetric central region [22].

The limited longitudinal space also determines the configuration of the optical system. Cherenkov photons are redirected away from the direct charged-particle acceptance and focused onto six detection regions located behind the geometrical shadow of the barrel detector and its support ring. This allows the photon sensors and their associated services to be placed outside the main particle trajectories while retaining the required forward acceptance. The resulting optical configuration is therefore intrinsically off-axis, and the curved photon-detection surfaces described in the previous section are designed to follow as closely as possible the three-dimensional focal region generated by the mirror system [22, 31].

The location of these detection regions places the photosensors close to the coils of the MARCO solenoid. Although the central solenoid operates at a reference field of $1.7\,\mathrm{T}$, the field in the dRICH photon-detection region is smaller but highly non-uniform, ranging approximately

from $0.3\,\mathrm{T}$ to $0.6\,\mathrm{T}$, and the field lines are tangent to the detection plane, as shown in Figure 47 [22]. Such conditions strongly constrain the choice of the photon-detection technology: conventional vacuum photomultipliers are unsuitable in this environment, whereas SiPMs maintain their photon-detection performance in high magnetic fields. Their compact dimensions additionally allow the sensor plane to conform to the non-planar focal surface required by the optical design [31, 38]. The magnetic field also affects the trajectory of charged particles while they traverse the radiator volume. Because Cherenkov photons are emitted continuously along the particle path, significant bending inside the radiator introduces an additional spread in the photon emission directions and therefore degrades the achievable Cherenkov-angle resolution. This loss cannot be fully recovered simply by modelling the curved trajectory accurately.

For this reason, the position of the dRICH and the configuration of the MARCO coils have been optimized to minimize the charged-particle bending within the approximately metre-long gas radiator. The detector simulation nevertheless includes the residual curvature of the particle trajectories, together with tracking resolution and multiple scattering [22, 31].

This requirement is particularly stringent at the highest momenta, where the difference between the Cherenkov angles of pions and kaons becomes very small. Most contributions to the single-photon angular resolution are consequently required to remain at or below approximately $0.5\,\mathrm{mrad}$. Unlike statistically independent single-photon uncertainties, the effects associated with the reconstructed track direction and its magnetic bending are correlated among all photons belonging to the same charged particle and therefore cannot simply be reduced by increasing the number of detected photons. Despite the focusing system and curved detection surface, the uncertainty associated with the photon emission point along the extended gas radiator remains one of the dominant geometrical contributions, at the level of approximately $1\,\mathrm{mrad}$ [22, 31].

The curved readout geometry provides an additional means of adapting the detector to these spatial constraints. The photon-detection plane is assembled from compact, four-side-buttable PDUs, allowing the active area to approximate the curved focal surface while minimizing inactive regions between adjacent modules [38]. This modular approach is particularly important because the photosensors, front-end electronics and thermal-management components must all be accommodated within the restricted volume available behind the barrel support structure.

The complete mechanical and optical configuration has also been addressed through a real-scale engineering prototype reproducing the geometry of one dRICH sector. Such a prototype allows the mechanical stability, gas tightness, thermal behaviour, assembly of the optical components and the extended photon-detection plane to be studied under a realistic off-axis configuration [31]. The magnetic and geometrical constraints are therefore not independent engineering considerations, but directly determine the optical layout, the placement and technology of the photon sensors, and ultimately the achievable particle-identification performance of the dRICH.

2.6 dRICH SiPM-based photon detection

The photon-detection system is one of the most distinctive technological aspects of the ePIC dRICH. The detector must reconstruct Cherenkov patterns formed by a relatively small number

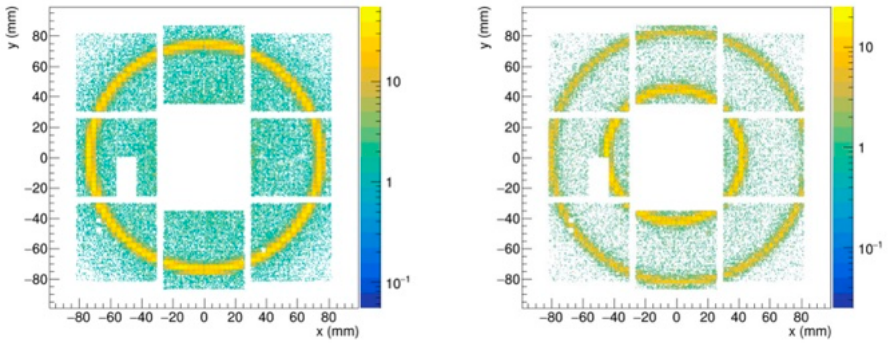


Figure 48: Cherenkov-ring images reconstructed with the SiPM-based dRICH prototype. The left panel shows the response obtained with the aerogel radiator, while the right panel shows the simultaneous aerogel and C_2F_6 ring patterns. Adapted from [38].

of photons while operating in a restricted volume, in a non-uniform magnetic field and under a moderate but non-negligible radiation load. These conditions simultaneously require high single-photon sensitivity, fine spatial segmentation, precise timing and stable long-term operation [22, 31].

Silicon Photomultipliers (SiPMs)[39] provide a compact solid-state solution satisfying these requirements. Their implementation in the dRICH, however, also introduces challenges associated mainly with intrinsic dark-count background and radiation-induced degradation [40]. The present detector concept therefore treats the SiPM sensor, front-end electronics, cooling system and annealing capability as components of a single integrated photo-detection system [22, 38].

The discussion in this section is restricted to the requirements arising from the dRICH application. The physical operating principle of SiPMs, their characteristic parameters, noise processes and radiation-damage mechanisms are introduced systematically in Chapter 3.

2.6.1 Need for single-photon sensitivity

Ring-imaging Cherenkov detectors reconstruct the velocity of a charged particle from the directions of the individual photons emitted while it traverses the radiator. Unlike a calorimetric measurement, where the relevant observable is the total collected light, the dRICH relies on the spatial pattern formed by individual photon hits. The photon detector must therefore register isolated photons with high efficiency and sufficiently low noise that their positions can be associated with a Cherenkov ring [24, 31].

The requirement is particularly stringent because the number of detected photons per track is intrinsically limited. For particles well above Cherenkov threshold, simulations predict on average approximately 12 recorded photons per track from the aerogel radiator and about 18 from the gas radiator, with moderate variations across pseudorapidity [22, 31]. Each individual detected photon consequently carries a significant fraction of the information used to reconstruct the Cherenkov angle.

The feasibility of this approach has been demonstrated with dRICH prototypes. In recent beam tests, more than 23 photoelectrons per track were measured for $11.5\,\mathrm{GeV/c}$ pions and

muons in the tested configuration, and Cherenkov rings from the aerogel and gas radiators were reconstructed using SiPM readout [38]. A representative result is shown in Figure 48. Efficient photon detection is also important because the track-level angular resolution is obtained by combining the information from the photons assigned to the same charged particle. Increasing the number of correctly reconstructed photons improves the statistical component of the Cherenkov-angle measurement, while lost photons or accidental noise hits reduce the effectiveness of the ring reconstruction. High photon-detection efficiency must therefore be achieved together with controlled accidental rates rather than considered as an isolated sensor property.

For this reason, the dRICH sensor requirements are specified directly at the single-photon level. The baseline photosensors must provide efficient photon detection over a broad spectral range from approximately 300 to $900\ \mathrm{nm}$ and achieve a peak photon-detection efficiency greater than 40%, with the maximum sensitivity lying within the 400–$450\ \mathrm{nm}$ wavelength range [22]. The microscopic processes producing this response will be discussed in Sec. 3.1.2.

2.6.2 Spatial and timing requirements

The position of each detected photon enters directly into the reconstruction of its Cherenkov angle. Spatial segmentation must consequently be fine enough that the finite sensor size does not become a dominant contribution to the single-photon angular resolution. For the dRICH, the active area of one readout channel is $3 \times 3\ \mathrm{mm}^2$

a value that optimise the required spatial precision [22, 31].

The complete dRICH photon-detection surface covers approximately $3\ \mathrm{m}^2$ and contains more than 3×10^5 SiPM channels [24, 38]. The sensors are distributed over six curved detection surfaces following the focal geometry of the mirror system. This large number of independent channels makes compact integration particularly important and motivates the organization of the sensors into modular Photo-Detector Units.

The importance of controlling the spatial contribution is illustrated in Figure 49, which shows the different terms entering the single-photon angular resolution for the two radiators. Most individual contributions are designed to remain at or below approximately $0.5\ \mathrm{mrad}$. This scale is driven by the small difference between the pion and kaon Cherenkov angles at the upper end of the gas radiator momentum range [22]. Precise timing provides a second and complementary handle. Individual SiPM channels are required to have an intrinsic time resolution better than approximately $150\ \mathrm{ps}$. The baseline ALCOR front-end uses a TDC architecture with a time binning finer than $50\ \mathrm{ps}$, and the complete SiPM-ALCOR chain is expected to provide a time resolution better than approximately $200\ \mathrm{ps}$ RMS [22, 31].

Timing is not used to determine the Cherenkov angle itself, but is crucial for assigning photon hits to the correct track and rejecting random dark counts. This becomes increasingly important as radiation exposure raises the SiPM dark-count rate during detector operation. The spatial and temporal specifications are therefore closely connected: the former preserves the angular information contained in the Cherenkov image, while the latter preserves the purity of that image.

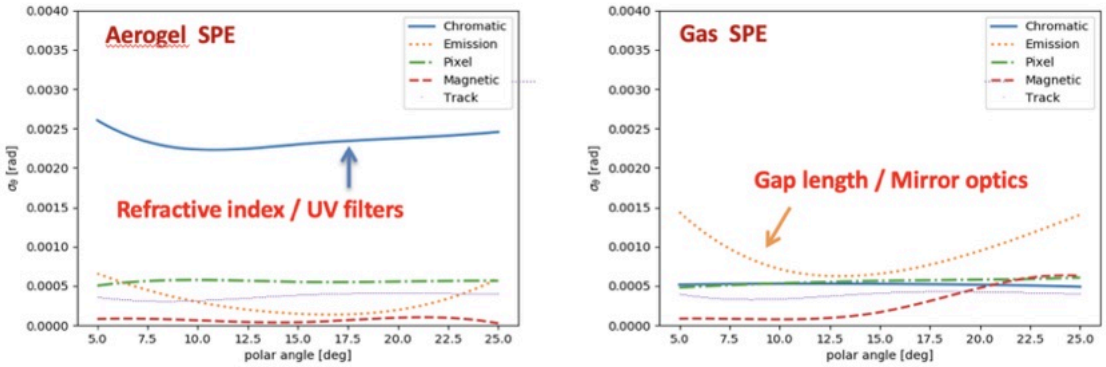


Figure 49: Contributions to the dRICH single-photon angular resolution for the aerogel radiator (left) and gas radiator (right). The finite sensor pixel size is one of the contributions controlled through the choice of a $3 \times 3\ \mathrm{mm}^2$ SiPM channel. Reproduced from [22].

2.6.3 Choice of SiPMs

The dRICH photosensor must satisfy several requirements simultaneously. It must detect single photons efficiently over a broad visible and near-UV wavelength interval, provide millimetre-scale segmentation and sub-nanosecond timing, tolerate the magnetic environment, and allow the construction of a large curved active surface at a manageable cost. SiPM technology provides this combination in a commercially available solid-state device [24].

The baseline sensor specifications require a $3 \times 3\ \mathrm{mm}^2$ active area per channel and sensitivity over approximately

$$300\ \mathrm{nm} < \lambda < 900\ \mathrm{nm}.$$

The peak photon-detection efficiency should exceed 40%, with the maximum sensitivity lying approximately in the range

$$400\ \mathrm{nm} < \lambda_{\mathrm{peak}} < 450\ \mathrm{nm}.$$

This wavelength range provides good overlap with the Cherenkov spectrum transmitted by the radiators and optical elements [22].

Sensor selection also requires a compromise between photon-detection efficiency, timing response and noise. Larger microcells generally provide larger signals and can improve the single-photon timing response, but sensor geometry and operating conditions also affect the dark-count and correlated-noise behaviour. Several commercial and prototype SiPMs have therefore been characterized before and after irradiation as part of the dRICH $R\&D$ programme [31, 38].

The current $R\&D$ baseline reported for the dRICH is the Hamamatsu S13361-3075, although the investigation of faster devices and sensors with enhanced UV sensitivity is continuing [31]. The final detector requirements are therefore expressed primarily in terms of performance specifications rather than relying solely on a particular commercial device.

A large-area SiPM implementation has already been validated with detector prototypes. Figure 50 shows the 2024 readout plane used in beam tests, where arrays of individual sensors

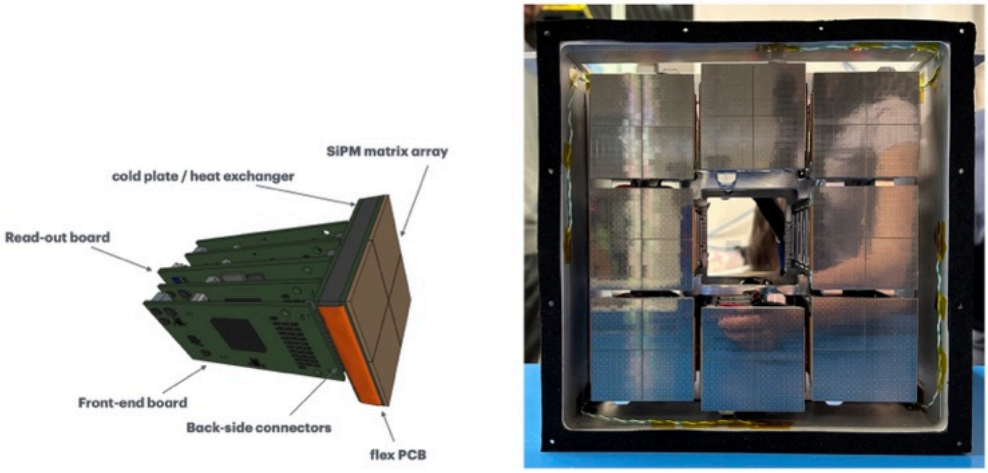


Figure 50: SiPM-based dRICH readout prototype used during the 2024 beam tests. The modular structure allows large photon-detection surfaces to be assembled from compact SiPM arrays. Image taken from [38].

were assembled into modular units and operated together with the dRICH readout electronics. The internal structure, breakdown behaviour, gain, PDE and dependence on overvoltage that ultimately determine these sensor properties will be explained in detail in Chapter 3.

2.6.4 Magnetic-field insensitivity

The magnetic environment represents one of the strongest arguments for the use of SiPMs in the dRICH. As discussed in Sec. 2.5.4, the photon-detection surfaces are located close to the MARCO solenoid coils in a region where the magnetic field is non-uniform and reaches approximately

$$B \simeq 0.3 - 0.6\,\mathrm{T}.$$

The orientation of the field is also unfavourable for conventional vacuum photodetectors [22, 24].

Classical photomultiplier tubes rely on the transport of electrons through vacuum between a photocathode and a sequence of multiplication stages. Their trajectories are therefore strongly affected by an external magnetic field. Even microchannel-plate photomultipliers, which are considerably more tolerant than conventional PMTs, become difficult to operate when the field direction is poorly aligned with the multiplication channels [24].

SiPMs, in contrast, perform photon conversion and charge multiplication entirely within a solid-state silicon structure. Their single-photon response is consequently essentially unaffected by magnetic fields of the magnitude expected in the dRICH region [24, 22]. This property removes the need for bulky magnetic shielding and allows the active surface to be positioned close to the solenoid coils.

The location of the dRICH relative to the central magnet is illustrated in Figure 51, where the forward RICH volume is highlighted within the full ePIC detector. This geometrical integration,

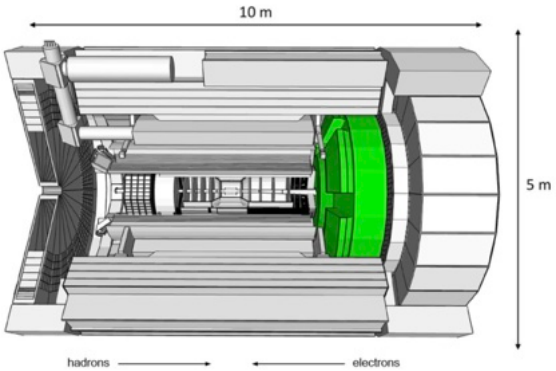


Figure 51: Schematic view of the ePIC detector with the forward dRICH volume highlighted. Its integration close to the central solenoidal magnet motivates the use of a magnetic-field-insensitive photon-detection technology. Adapted from [24].

combined with the non-uniform field described previously, strongly favours a compact solid-state photodetector technology.

Magnetic-field insensitivity is therefore not merely a general advantage of SiPM technology: it is one of the characteristics that enables the chosen optical layout of the dRICH itself. A more detailed description of SiPM avalanche multiplication is postponed to Chapter 3.

2.6.5 Dark-count background

The principal disadvantage of SiPMs for Cherenkov imaging is their intrinsic dark-count rate (DCR). A thermally or electrically generated carrier can initiate an avalanche that is indistinguishable at the readout level from an avalanche triggered by a real photon. In a detector in which the physical signal consists of isolated single photons, such pulses directly produce accidental hits in the reconstructed Cherenkov image [24].

For new SiPMs, the dRICH TDR quotes a room-temperature dark count rate of the order of

$$50\,\mathrm{kHz/mm^2},$$

although the exact value depends on the sensor type and operating conditions [22]. For a $3 \times 3\,\mathrm{mm}^2$ channel, this would correspond to several hundred kilohertz at room temperature, illustrating the importance of operating the sensors at substantially lower temperature.

The radiation-damage and annealing studies adopted for the present dRICH design are instead based on Hamamatsu S13360-3050 sensors operated at $T = -30^{\circ}\mathrm{C}$ and $V_{\mathrm{over}} = 4\,\mathrm{V}$. Under these conditions, a DCR of

$$300\,\mathrm{kHz/channel}$$

is used as a conservative operational reference for assessing the evolution of the dark rate and the need for annealing. This value should not be interpreted as a strict detector limit; operation at lower temperature and overvoltage, for example $T = -40^{\circ}\mathrm{C}$ and $V_{\mathrm{over}} = 3\,\mathrm{V}$, would further reduce the DCR [22].

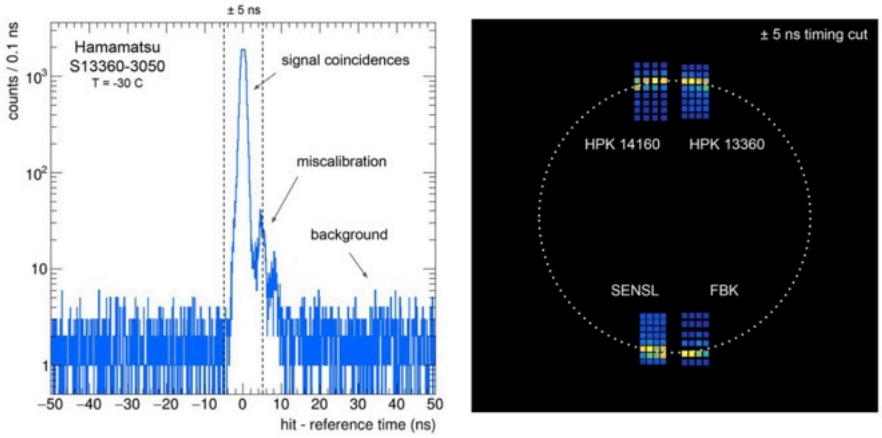


Figure 52: Timing suppression of the SiPM dark-count background in a dRICH prototype. Left: hit-time distribution showing the correlated signal and approximately uniform DCR background. Right: reconstructed partial Cherenkov ring after application of a $\pm 5\,\mathrm{ns}$ timing selection. Image from [24].

Simulations indicate that the particle-identification performance is only marginally affected at this level, provided that timing information and the appropriate reconstruction strategy are used [22, 31].

The usefulness of timing is demonstrated experimentally in Figure 52. The distribution measured with a SiPM prototype contains a narrow time-correlated component associated with the beam and an approximately uniform background originating from random dark counts. Selecting photon hits inside a restricted time interval strongly suppresses this accidental contribution and allows the Cherenkov ring to emerge clearly [24].

The final ALCOR electronics exploit this principle more aggressively. Single-photon time-tagging with a system resolution better than approximately $200\,\mathrm{ps}$ enables out-of-time DCR to be rejected in both firmware and software. A programmable hardware shutter is included specifically to reduce the data bandwidth once radiation-induced DCR becomes significant [22]. At the conservative $300\,\mathrm{kHz/channel}$ reference level, the total dRICH throughput would otherwise reach approximately $1.3\,\mathrm{Tbps}$ before the planned additional data reduction [22, 31].

The microscopic origins of primary and correlated SiPM noise and their dependence on temperature and overvoltage are discussed separately in Sec. 3.3.

2.6.6 Expected radiation damage

Radiation exposure represents the second major challenge associated with the use of SiPMs in the dRICH. The photon sensors and front-end electronics are intentionally positioned away from the beam pipe, where the radiation environment is substantially less severe than in the innermost forward region. Nevertheless, an accumulated non-ionizing radiation fluence is expected throughout the ePIC lifetime [22].

At the location of the dRICH photodetectors, simulations for ep collisions at the highest EIC

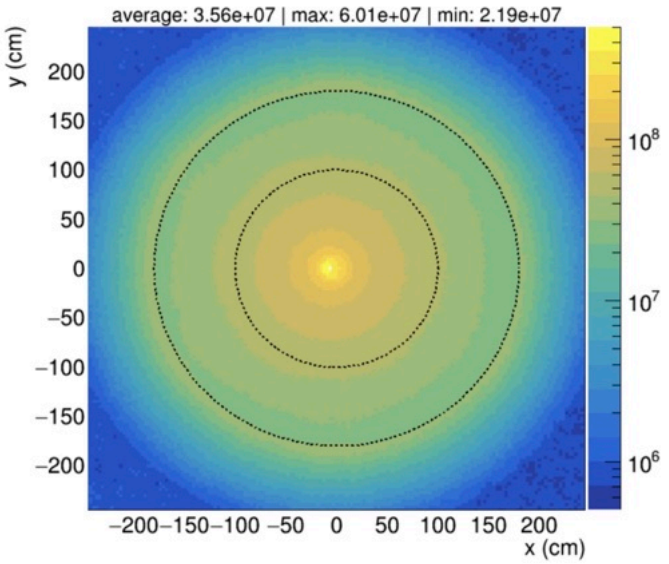


Figure 53: Transverse map of the expected 1-MeV-neutron-equivalent fluence per $1\,\mathrm{fb}^{-1}$ of integrated luminosity at the longitudinal position of the dRICH photon detector for highest-energy ep collisions. The dashed boundaries indicate the radial region occupied by the photon-detection system. Reproduced from [22].

centre-of-mass energy predict a maximum 1-MeV-neutron-equivalent fluence of approximately

$$\Phi_{\mathrm{eq}}^{\mathrm{max}} \simeq 6.0 \times 10^{7}\ \mathrm{cm}^{-2}/\mathrm{fb}^{-1},$$

with an average of approximately

$$\Phi_{\mathrm{eq}}^{\mathrm{avg}} \simeq 3.6 \times 10^{7}\ \mathrm{cm}^{-2}/\mathrm{fb}^{-1}.$$

When the estimated contribution from beam-induced background is included, these values become approximately

$$6.4 \times 10^{7} \quad \text{and} \quad 3.7 \times 10^{7}\ \mathrm{cm}^{-2}/\mathrm{fb}^{-1},$$

respectively [22].

The non-uniformity of the exposure is visible in Figure 53. Sensors closer to the beam line experience the largest fluence, while those at larger radii receive substantially smaller doses. The complete photo-detection system is expected to remain below a maximum integrated fluence of order $10^{11}\ n_{\mathrm{eq}}/\mathrm{cm}^2$ [22].

Within the radiation range relevant to the dRICH, the principal observed macroscopic consequence is an increase of the dark-count rate. Other operating parameters, including PDE, gain, quenching resistance, pixel capacitance and breakdown voltage, are not expected to undergo comparably large changes at these moderate fluences [22]. The preservation of single-photon sensitivity is therefore determined primarily by the ability to control the radiation-induced increase in DCR.

Dedicated irradiation campaigns have consequently exposed candidate sensors to fluences extending to and beyond those expected during dRICH operation [31, 38, 41]. The microscopic mechanisms by which displacement damage introduces electrically active defects and increases generation current will be discussed in detail in Sec. 3.4.

2.6.7 Cooling and annealing as mitigation strategies

The dRICH operating strategy combines two complementary approaches to control the SiPM dark-count rate: continuous operation at reduced temperature and periodic high-temperature annealing. Cooling suppresses the thermally generated component of the DCR during data taking, whereas annealing partially recovers the radiation-induced degradation accumulated over time [22, 24].

For new SiPMs, the DCR typically decreases by approximately a factor two for every 7–$10^{\circ}\mathrm{C}$ reduction in temperature [22]. The dRICH sensors are therefore designed to operate at

$$T \simeq -30^{\circ}\mathrm{C}$$

or below. Operation near $-40^{\circ}\mathrm{C}$ remains a possible option for further DCR reduction. The prototype PDU thermal system has already demonstrated sensor operation down to approximately $-40^{\circ}\mathrm{C}$ [38].

In the detector design, the SiPM carrier board is placed in thermal contact with a cooling block connected to a closed circulating-fluid system. The same thermal infrastructure is designed to participate in the annealing procedure by raising the sensor assembly toward high temperature during periods without physics data [22, 31].

Annealing can also be produced directly within the SiPM. Under forward bias the microcells behave as forward-polarized diodes, and the current flowing through the device generates heat through the Joule effect. Tests performed during the dRICH R&D programme have shown that this procedure can recover approximately 97% of radiation-induced degradation in Dark Current under appropriate conditions [22, 31]. The circulating thermal system can pre-heat the sensors to about $100\,^{\circ}\mathrm{C}$ or above, reducing the electrical power required for the subsequent forward-bias annealing. Annealing temperatures up to approximately $150\,^{\circ}\mathrm{C}$ are currently considered.

Temperatures significantly above this value are disfavoured for the present devices. Tests at $175\,^{\circ}\mathrm{C}$, for example, showed discoloration of the resin covering the SiPMs, with the potential to reduce their photon-detection efficiency [38].

The projected evolution of the DCR and the effect of repeated recovery cycles are shown in Figure 54. Without annealing, the most exposed sensors would reach the conservative $300\,\mathrm{kHz/channel}$ reference level after only about $9.5\,\mathrm{fb}^{-1}$. The repeated recovery strategy substantially extends this interval [22].

A possible operational scenario consists of relatively frequent, softer recovery cycles, approximately every $3\,\mathrm{fb}^{-1}$, interleaved with more complete annealing cycles roughly every $30\,\mathrm{fb}^{-1}$. Because annealing cannot completely remove the accumulated damage, a residual DCR is expected to build up progressively. For the most exposed detector region, maintaining the DCR below the present $300\,\mathrm{kHz/channel}$ reference becomes increasingly demanding beyond an integrated luminosity of approximately $200\,\mathrm{fb}^{-1}$ [22]. These numbers should be regarded as an

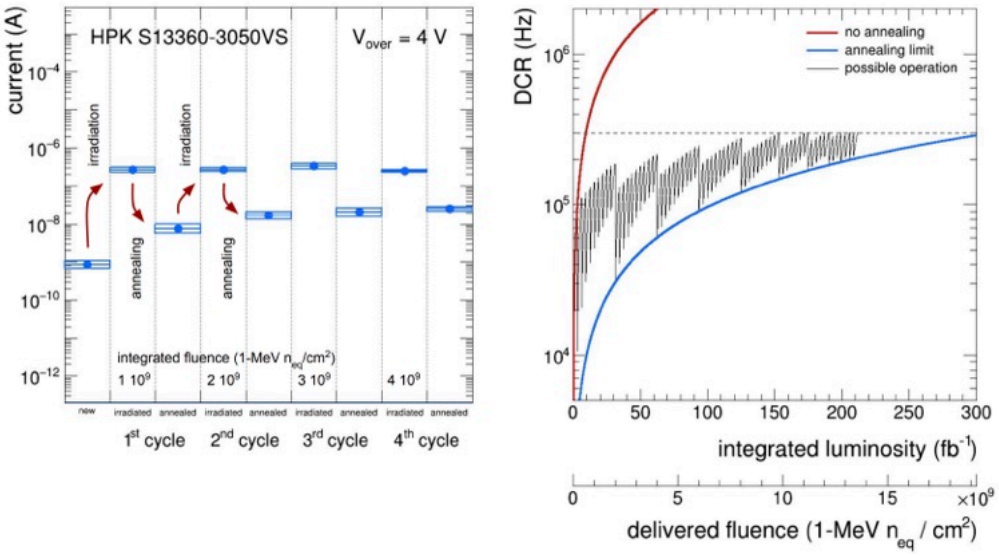


Figure 54: SiPM radiation-damage and annealing studies for the dRICH. Left: dark-current evolution during repeated irradiation and annealing cycles. Right: projected DCR evolution for the most exposed PDU as a function of delivered luminosity, showing the cases without annealing, the residual annealing limit and a representative operating scenario. Reproduced from [22].

operational projection rather than a fundamental detector limit, since they depend on sensor choice, temperature, overvoltage, reconstruction performance and the achievable annealing efficiency.

Cooling and annealing are therefore integral elements of the dRICH photon-detection concept rather than auxiliary services. Their purpose is to preserve the single-photon sensitivity required for Cherenkov imaging throughout detector operation. The physical origin of the temperature dependence, radiation-induced defects and their thermal recovery will form the subject of the following chapter and provide the direct motivation for the experimental studies presented later in this thesis.

3 Silicon Photomultipliers

The silicon photomultiplier (SiPM) is a solid-state photodetector consisting of an array of integrated single-photon avalanche diodes (SPADs), commonly referred to as microcells or pixels, connected in parallel and typically read out through a single channel [39]. Each SPAD has a square geometry, with a characteristic size ranging from approximately 10 to 100 μm. By operating the individual microcells above their break-down voltage, SiPMs provide a large electrical signal even for the detection of a single photon, together with high photon-detection efficiency and fast response. Commercial devices typically have active areas ranging from about 1 mm^2 up to a few mm^2 [24]. Their main advantages include low operating voltage, compactness, robustness, insensitivity to magnetic fields, relatively low cost, commercial availability, and a high degree of customizability.

In the context of the dRICH detector, SiPMs are operated in a regime in which the detection of individual photons is required. Under these conditions, one of the main limitations is represented by the dark count rate (DCR), since thermally generated carriers and electrically or radiation-induced defects can trigger avalanches that are indistinguishable from those initiated by photons. The DCR is therefore closely related to the SiPM operating conditions, increasing with temperature and overvoltage, and can be significantly enhanced by radiation damage. Understanding the origin and evolution of these noise contributions is essential for assessing the performance of SiPMs in a radiation environment. Their impact can be mitigated through strategies such as cooling, the use of narrow timing windows, and annealing. In addition, emission microscopy (EMMI) provides a complementary tool for spatially resolving the avalanche activity of the device and identifying localized regions associated with electrically active defects.

Following this logical progression, the first sections of this Chapter introduce the SiPM working principle, starting from its basic components (Sec. 3.1), and describe its main operating parameters and characteristics (Sec. 3.2). The different sources of primary and secondary noise are then discussed in Sec. 3.3, with particular attention to their microscopic origin and their macroscopic effects on detector performance. Sec. 3.4 describes the main effects of radiation on SiPMs, emphasizing the distinction between damage induced by ionizing and non-ionizing radiation. The mitigation strategies that can be adopted to reduce the impact of radiation damage, including cooling, narrow timing windows, and annealing, are presented in Sec. 3.5. Finally, Sec. 3.6 reviews previous studies devoted to the characterization of defects in SiPMs, with particular emphasis on spatially resolved investigations of avalanche activity, and discusses their main results in relation to the approach adopted in this thesis.

3.1 Silicon Photomultipliers operating principle

3.1.1 Diode working principle: the photoconversion

Semiconductors are materials whose electronic properties lie between those of conductors and insulators. They include elemental semiconductors, such as silicon and germanium, as well as compound semiconductors formed by elements belonging to different groups of the periodic table. In a crystalline semiconductor, the interaction between neighbouring atoms causes the

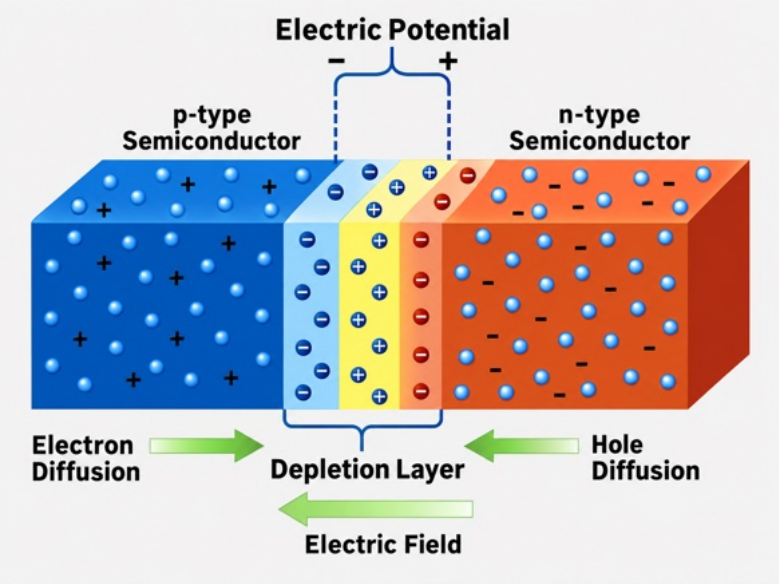


Figure 55: Representation of diffusion processes within a p-n junction with different doping concentrations.

discrete atomic energy levels to broaden into energy bands. The highest occupied band is referred to as the *valence band*, while the higher-energy band in which electrons can contribute to electrical conduction is the *conduction band*. The two are separated by an energy interval known as the *band gap*.

At finite temperature, the occupation of these energy states is governed by the Fermi–Dirac distribution. Although the characteristic thermal energy $k_{\mathrm{B}}T$ at room temperature is considerably smaller than the semiconductor band-gap energy, thermal excitation produces a small but finite probability for electrons to be promoted from the valence band to the conduction band. Consequently, an intrinsic semiconductor contains a finite concentration of thermally generated electron–hole pairs. Electrons in the conduction band and the corresponding holes left in the valence band can both contribute to electrical conduction.

The concentration and type of charge carriers can be modified through *doping*. In silicon, the introduction of Group III atoms creates acceptor states and leads to *p-type* material, in which holes are the majority carriers, whereas Group V atoms act as donors and produce *n-type* material, in which electrons are the majority carriers. When p-type and n-type materials are brought into contact to form a *p–n junction*, the carrier concentration gradient causes electrons and holes to diffuse across the junction, as schematically illustrated in Figure 55.

As electrons and holes diffuse across the junction, they recombine, leaving behind ionized donor and acceptor atoms that are fixed within the crystal lattice. A region depleted of mobile charge carriers is therefore established around the junction, known as the *depletion region*. The fixed charges generate an internal electric field and a corresponding potential barrier that oppose further carrier diffusion.

The width of the depletion region and the strength of its electric field can be modified by

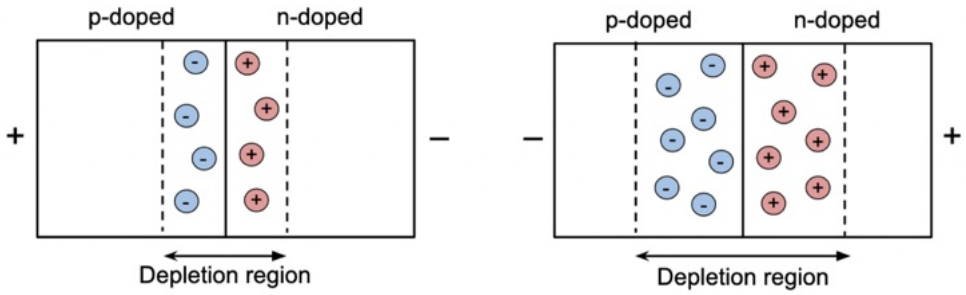


Figure 56: A p-n junction under two different polarisation conditions. On the left, forward bias; on the right, reverse bias.

applying an external voltage across the junction, as illustrated in Figure 56. In particular, reverse bias increases the depletion-region width and the electric field within it. When an ionizing particle traverses this region, or when a photon with sufficient energy is absorbed in or near the active region, electron–hole pairs are generated. The electric field separates the charge carriers and drives electrons and holes towards opposite electrodes, producing a measurable electrical signal. This conversion of deposited or absorbed energy into an electrical signal constitutes the basic detection mechanism of a semiconductor photodiode.

3.1.2 SPAD working principle: the multiplication

SPADs are photodiodes specifically designed to operate under reverse-bias conditions (p-type doping connected to the negative voltage terminal, n-type doping to the positive terminal) beyond the break-down voltage.

Depending on the strength of the electric field applied to the diode, there are three operating modes (corresponding to three different devices):

1. If the field is low, the e-h pairs produced by interaction with the particle are collected but not multiplied, and the device operates as a unity-gain photodiode [24].

2. When the applied electric field is increased, charge carriers can gain enough kinetic energy between collisions to generate additional electron-hole pairs through **impact ionisation**. This regime corresponds to the operation of an Avalanche Photodiode (APD). Below the break-down voltage, the electric field is sufficiently high to produce carrier multiplication, but not high enough to satisfy the condition for a self-sustained avalanche. As a consequence, the multiplication process remains finite and the avalanche naturally terminates once the initially generated carriers and their secondary products leave the high-field region. No external quenching circuit is therefore required in this regime. In silicon, the impact-ionisation probability is higher for electrons than for holes, so electron-initiated multiplication generally dominates the avalanche development.

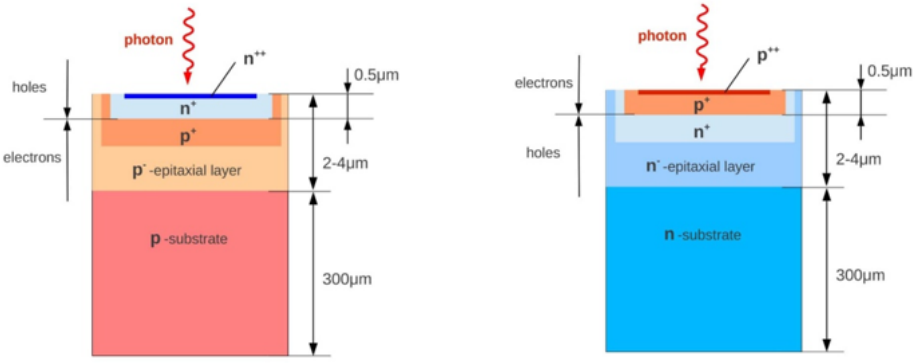


Figure 57: (a) The n-on-p SPAD structure, more sensitive to the red light and (b) the p-on-n structure, more sensitive in the blue and near UV. Pictures from [39].

3. If the applied electric field is even stronger, exceeding what is known as the break-down voltage, the holes will also have sufficient kinetic energy to create secondary e-h pairs. This is the operating regime of the SPADs used in SiPMs (electric field of the order of $10^5\ V \cdot cm^{-2}$), also known as **Geiger mode** [39]. Under these conditions, a single incident photon can trigger an avalanche process, very rapidly generating a macroscopic current, which allows the passage of the particle to be detected with high temporal resolution. Once the current begins to flow, the process is self-sustaining unless the voltage is reduced below the break-down voltage; this is the role of the quenching circuit. Indeed, in a passively quenched microcell, the avalanche current flows through the quenching resistor R_q, producing a voltage drop across it. As a consequence, the voltage across the junction decreases and eventually falls below the break-down voltage, interrupting the self-sustained avalanche. Once the avalanche has been quenched, the microcell does not return instantaneously to its operating voltage. Instead, it progressively recharges through the quenching resistor toward the applied bias voltage, with a finite recovery time determined by the effective resistance and capacitance of the cell. During this recharge period, the microcell has a reduced sensitivity to subsequent photons and fully recovers only after its voltage has returned sufficiently close to the nominal operating value. It is also worth noting that, in this operating regime, the electrical signal produced by a single microcell is not proportional to the energy deposited by the incident particle or photon.

3.1.3 SPAD fabrication

There are two ways of constructing a SPAD, either n-doped on p-doped or p-doped on n-doped, which can be seen in Figure 57.

SPADs are usually composed of a shallow very thin ($\sim\ 100nm$) n^+/p^+ layer characterized by a high number of donors/acceptors, a p^+/n^+ implant that creates the junction with the n^+/p^+ layer, a low doping p/n epitaxial layer, and a final p/n substrate with low resistivity that acts as contact for the cathode. The choice between n-on-p and p-on-n structures depends

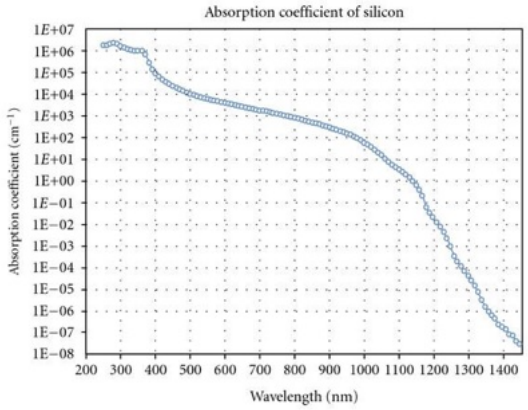


Figure 58: The absorption coefficient of silicon in cm^{-1} as a function of the wavelength. The data is graphed on a y log scale. Image from [42].

on the wavelength range for which the detector is intended to achieve the highest sensitivity. Photon absorption in silicon is strongly wavelength dependent, as shown in Figure 58. Blue and near-ultraviolet photons are absorbed within a shallow region close to the device surface, whereas photons at longer wavelengths, such as red light, penetrate deeper into the silicon before being absorbed.

The resulting spectral response depends not only on the absorption depth, but also on the electric-field profile, the position of the high-field multiplication region, and on which type of charge carrier reaches this region and initiates the avalanche. In silicon, electrons generally have a higher impact-ionisation probability than holes. Therefore, structures in which short-wavelength photons predominantly generate electrons that are efficiently driven into the multiplication region tend to exhibit a higher sensitivity in the blue and near-ultraviolet range. Conversely, when photons are absorbed deeper in the device, the relative contribution of electrons and holes to avalanche triggering changes according to the junction geometry and electric-field configuration.

As a result, p-on-n SPAD structures are generally optimized for enhanced sensitivity at shorter wavelengths, whereas n-on-p structures are typically more favourable for detection at red and longer wavelengths [39].

Figure 59 shows a cross-sectional diagram of two different blue sensitive SPADs. The sensor cell is connected to the quenching resistor—shown in grey in the figure—via a layer of SiO_2. It can also be seen that the detection area is covered by a layer of anti-reflective material, to prevent the loss of part of the incident signal. In configurations where SPADs are arranged in a matrix, as in Silicon Photomultipliers, isolating each individual cell is crucial to reduce noise effects such as charge diffusion and optical crosstalk (see Sec. 3.3). This isolation can be achieved by introducing physical trenches along the edges of each SPAD cell. The structure described above is fundamental to understand the possible damages caused by radiation, presented in Sec. 3.4.

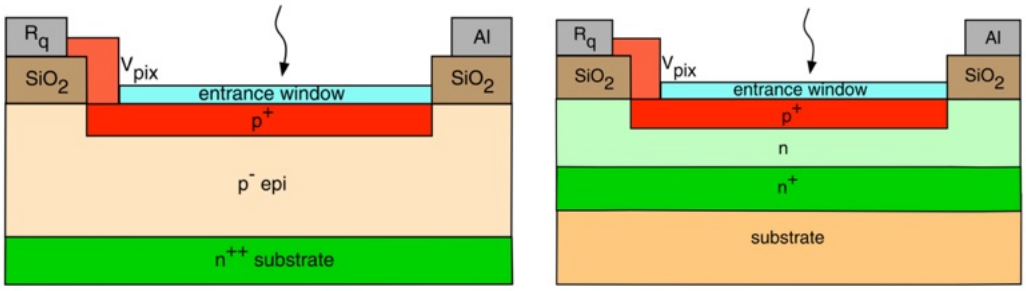


Figure 59: Schematic cross-section of two possible single pixels of blue sensitive SiPMs. The poly-silicon quenching resistor R_q and the Al-contact line are isolated from the silicon via a SiO_2 layer. Image from [40].

3.2 Main SiPM parameters

3.2.1 SPAD electrical model and Gain

To better understand the behaviour of the device, it is useful to introduce an electrical model of the SPAD. Several models can be found in the literature, differing mainly in the way the avalanche process is represented (for a comprehensive review, see [24]). In this work, the electrical model shown in Figure 60a is considered, where the photon-triggered avalanche is represented by a switch that closes when a break-down occurs, connecting the junction to a voltage source corresponding to the break-down voltage (V_{BD}).

The SPAD and its quenching resistor (R_Q) are represented while the device is operated under a reverse bias voltage (V_{bias}). The intrinsic capacitance and resistance of the diode space-charge region are modelled by C_d and R_D, respectively, while C_q represents the parasitic capacitance associated with the quenching resistor R_q. The depleted junction can be approximated as a parallel-plate capacitor with area A, so that its capacitance C_D can be expressed as

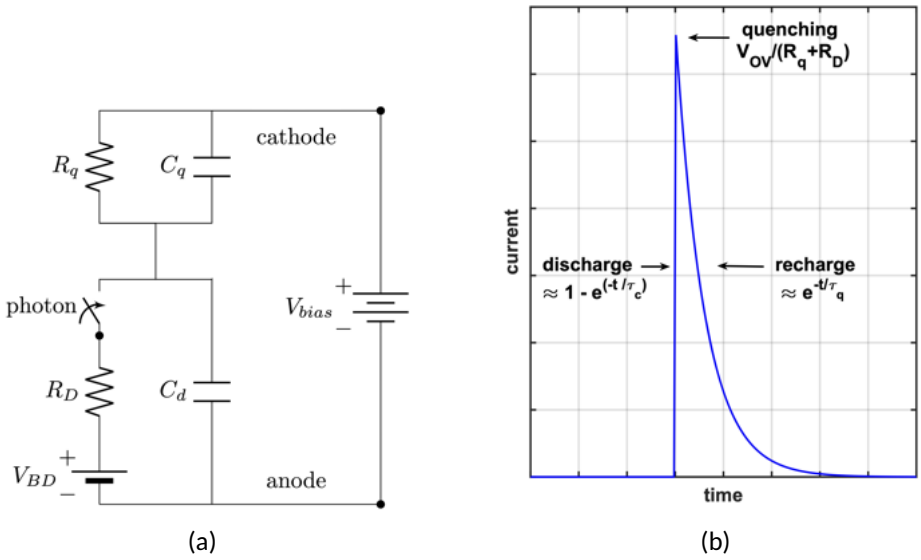
$$C_D = \frac{\epsilon_0 \epsilon_{\text{Si}} A}{W_D} \tag{56}$$

where ϵ_0 and ϵ_{Si} are the dielectric constants of vacuum and silicon, respectively, and W_D is the width of the depletion region. The latter can be expressed as [24]

$$W_D = \sqrt{\frac{2\epsilon_0 \epsilon_{\text{Si}}}{q} \left(\frac{N_A + N_D}{N_A N_D} \right) \left(V_{\text{bi}} + V_{\text{bias}} - \frac{2kT}{q} \right)} \tag{57}$$

where q is the elementary charge, N_A and N_D are the acceptor and donor doping concentrations, respectively, and V_{bi} is the built-in voltage, or contact potential, of the junction.

For an asymmetrically doped junction in which the donor concentration is much larger than the acceptor concentration, $N_D \gg N_A$, and therefore $N = N_A$, Eq. (57) becomes



(a) (b)

Figure 60: (a) Electrical model of a SPAD with its quenching resistor (R_q). (b) Current pulse produced by the SPAD following photon detection. Pictures from [24].

$$W_D = \sqrt{\frac{2\epsilon_0\epsilon_{\mathrm{Si}}}{q}\left(\frac{1}{N}\right)\left(V_{\mathrm{bi}} + V_{\mathrm{bias}} - \frac{2kT}{q}\right)} \tag{58}$$

Substituting Eq. (58) into Eq. (56), the dependence of C_D on the bias voltage can be written as

$$C_D = A\sqrt{\frac{Nq\epsilon_0\epsilon_{\mathrm{Si}}}{2}}\left(V_{\mathrm{bi}} + V_{\mathrm{bias}} - \frac{2kT}{q}\right)^{-1/2} \tag{59}$$

The junction capacitance depends only weakly on temperature and is proportional to the active area of the SPAD. As the reverse bias voltage increases, the depletion region becomes wider, although its width tends to saturate before the break-down voltage is reached. Consequently, above V_{BD}, the capacitance of the depletion region remains approximately constant.

The operation of the circuit shown in Figure 60a can then be described as follows. Before photon detection, the switch is open and C_D is charged to the voltage V_{bias} applied between the anode and cathode. When a photon triggers an avalanche, the switch closes and a current begins to flow, discharging C_D through the resistive elements of the circuit. Recalling that $v_{\mathrm{over}} = V_{\mathrm{bias}} - V_{\mathrm{BD}}$, the initial discharge current can be approximated as

$$I_D = \frac{V_{over}}{R_D + R_Q} \tag{60}$$

The voltage drop across C_D also induces a corresponding voltage variation across C_q, producing an additional and faster current component. Since C_q is typically smaller than C_D, this

contribution gives rise to a faster current spike, as in Figure 60b. The total charge released during the avalanche is proportional to the sum of the two capacitances multiplied by V_{over}. The gain (G), defined as the total avalanche charge divided by the elementary charge, is therefore given by

$$G = \frac{V_{over}\,(C_D + C_q)}{q} \tag{61}$$

In practice, the gain of a SPAD depends on the microcell capacitance and on the applied overvoltage. For relatively large microcells, with capacitances of the order of a few hundred femtofarads, corresponding approximately to cell sizes of order $100\ \mu\mathrm{m}$ depending on the device technology, gain values approaching 10^7 can be reached at overvoltages of several volts.

As the avalanche develops, the voltage across the junction decreases until the current falls below the threshold required to sustain the multiplication process. The capacitances then recharge through the quenching resistor, restoring the initial operating conditions. The resulting current pulse therefore reflects the discharge, quenching, and recharge phases of the SPAD response, as shown in Figure 60b.

The characteristic discharge and recharge times, which determine the speed of the SPAD response, are described by the time constants

$$\tau_c \approx (C_D + C_q)(R_D \parallel R_q)$$

and

$$\tau_q \approx R_q(C_q + C_D),$$

respectively. For typical values of $R_D \sim 1\ \mathrm{k\Omega}$ and $R_q \sim 100\ \mathrm{k\Omega}$, the rise time is extremely short, of the order of a few hundred picoseconds, whereas the decay time is considerably longer, typically of the order of $100\ \mathrm{ns}$. Both time constants depend on the junction capacitance and therefore on the size of the SPAD. Consequently, at the same V_{over}, smaller SPADs generally provide a faster response but a lower gain than devices with a larger active area.

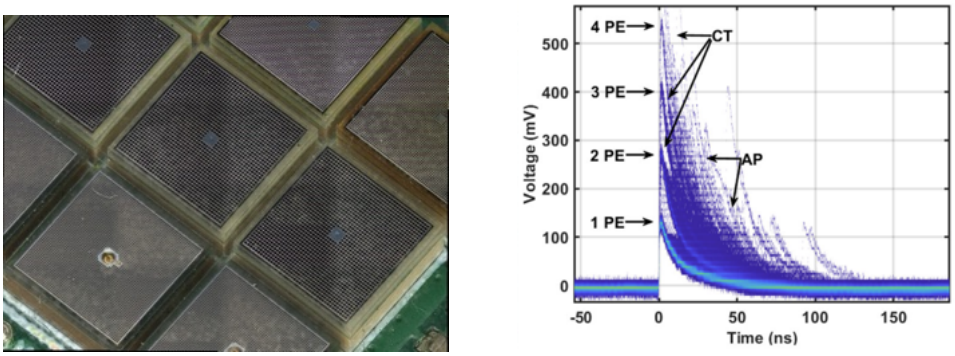
A trade-off between response speed and gain must therefore be considered depending on the requirements of the application. In particular, in low-light measurements where a high gain is especially advantageous, larger SPADs may be preferred despite their longer response time.

3.2.2 Photodetection probability

The ability of a SPAD to detect a photon is defined by the photodetection probability (PDP):

$$PDP(\lambda, V_{over}) = QE(\lambda) \times P_t(\lambda, V_{over}) \tag{62}$$

where $QE(\lambda)$ is the quantum efficiency and $P_t(\lambda, V_{over})$ is the avalanche triggering probability [24]. The quantum efficiency is the probability that a photon is absorbed in the device and it generates a carrier, in the active region, that reaches the high field region. The $QE(\lambda)$ is therefore obtained combining the probability for a photon to transmit through the anti-reflective coating (ARC) at the detector surface with the probability for a photon to create a carrier that can



(a) (b)

Figure 61: (a) Picture of Hamamatsu S13360 SiPMs. (b) Multiple waveform in persistence of a SiPM. It is visible the response of the device when multiple (4) SPADs are firing. The cross-talk (CT) and afterpulse (AP) contributions are highlighted (see Sec. 3.3). Image from [24].

reach the high-field region. The wavelength dependency of the QE arises from the fact that the photon absorption is wavelength dependent, as explained in the previous section. $P_t(\lambda, V_{over})$ is the probability that an electron–hole pair, photogenerated by a photon with wavelength λ, triggers an avalanche in a detector operated at a voltage V. The avalanche triggering probability depends on the electric field, thus on the overvoltage, and on the position where the carriers are generated.

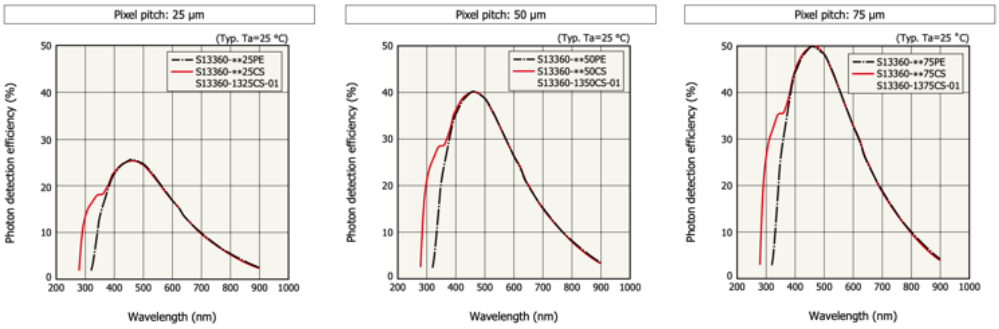
3.2.3 SiPM structure

As said above, SPADs can register individual photons at room temperature but, being operated in Geiger mode, they do not have a linear response to the number of photons in a light pulse. The SiPM solves this limitation by incorporating hundreds of SPADs connected within a single device. In this way, it is possible to detect the amount of incident light with single-photon precision; the only exception is when two photons strike the same cell within a time interval shorter than the quenching time; in that case, only one is detected. To reduce the probability of multiple photons impinging on the same SPAD, the dimension of the pixel is kept small, usually between 10×10 μm^2 and 100×100 μm^2. Figure 61a is a picture of a board of SiPMs.

Depending on the SPAD connection, SiPMs are divided into:

- Analog SiPMs: SPADs are connected in parallel, and it is possible to detect the amount of incident light, with single-photon precision, by summing the currents from the individual SPADs, as it can be seen in Figure 61b;
- Digital SiPMs: each SPAD is connected to its own readout electronics, including front-end and digitization of the signal's time stamp via a time to digital converter (TDC).

In the following paragraphs, greater attention will be paid to analog SiPMs, given their wider use in high-energy physics [39].



(a) (b) (c)

Figure 62: Photon Detection Efficiency (%) vs wavelength of Hamamatsu S13360 SiPM series with pixel pitch of (a) $25\mu m$ (b) $50\mu m$ and (c) $75\mu m$. Plots from [43].

3.2.4 Photodetection efficiency

The probability that a SiPM produces an output signal in response to an incident photon is called *Photon Detection Efficiency* (PDE) and is defined as [24]:

$$PDE = \frac{N(photons_detected)}{N(photons_impinging)} \tag{63}$$

It can also be expressed in term of the PDP defined in Eq. 62:

$$PDE(\lambda, V_{over}) = FF \times PDP(\lambda, V_{over}) \tag{64}$$

where FF denotes the microcell Fill Factor, defined as the ratio between the photosensitive area and the total SiPM surface, considering that the insulation trenches between SPADs are not sensitive to light. As a consequence, the $PDE(\lambda, V_{over})$ depends on the dimension of the SPADs, as it can be seen in Figure 62.

Through the PDP, the PDE also shows a dependency on the applied overvoltage, as it can be seen in Figure 63, since the avalanche triggering probability depends on the electric field.

3.2.5 Break-down voltage and overvoltage

The break-down voltage (V_{BD}) represents a fundamental parameter of a SPAD and defines the minimum reverse-bias voltage above which the device operates in Geiger mode. At this voltage, the electric field in the multiplication region becomes sufficiently high for the impact-ionization process to become self-sustaining. Its value is determined by the internal structure of the junction, in particular by the doping profile and by the width of the depleted region [44, 45, 39].

In practical SiPM operation, the absolute bias voltage is generally less informative than the voltage applied above break-down, referred to as the *overvoltage*:

$$V_{over} = V_{\mathrm{bias}} - V_{\mathrm{BD}}. \tag{65}$$

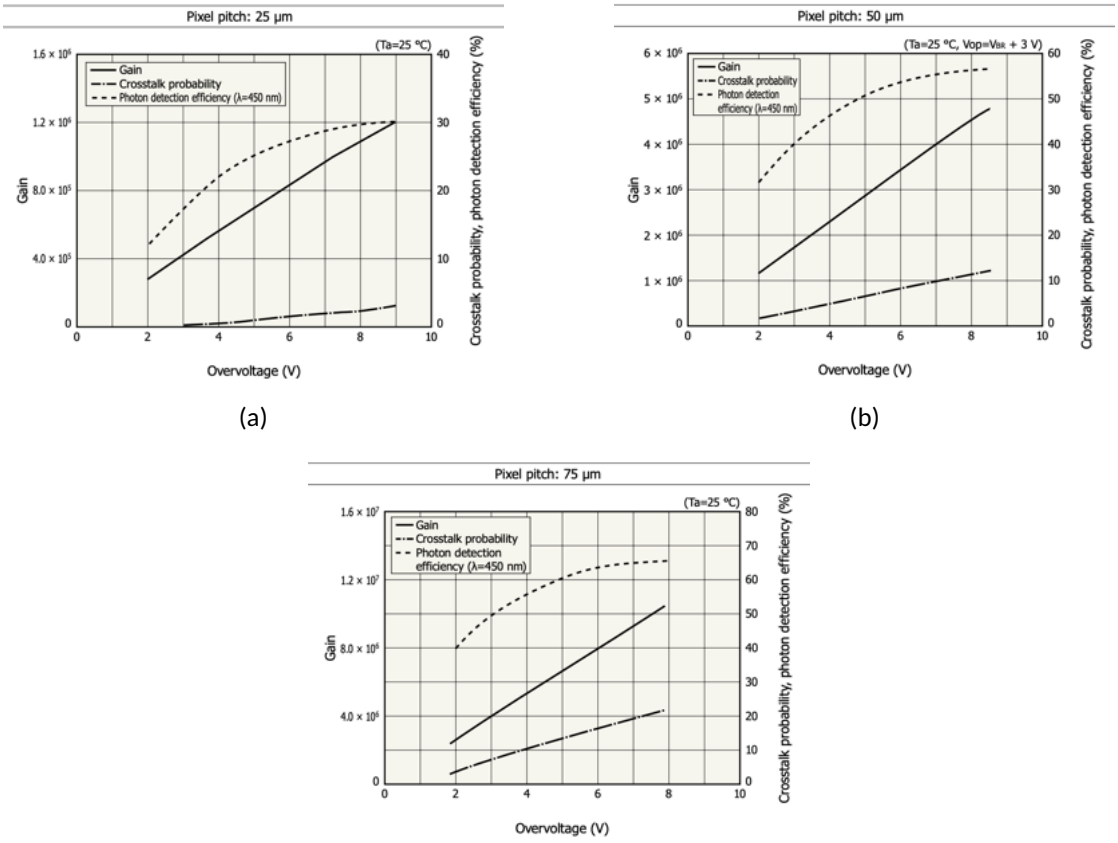


Figure 63: SiPMs' characteristics varying with the operating voltage for pixel pitch of (a) $25\ \mu m$ (b) $50\ \mu m$ and (c) $75\ \mu m$. Although increasing the operating voltage improves the photon detection efficiency and time resolution, it also increases the dark count and crosstalk at the same time (see Sec. 3.3), so an optimum operating voltage must be selected to match the application. Plots from [43].

As already introduced in the description of the SPAD electrical model, V_{over} determines the voltage excursion experienced by a microcell during an avalanche and therefore directly affects several SiPM parameters. For an approximately constant cell capacitance, the avalanche charge and consequently the gain increase approximately linearly with V_{over}, as expressed in Eq. (61). At the same time, increasing the overvoltage increases the electric field in the multiplication region and therefore the probability that a photo-generated carrier triggers a self-sustaining avalanche. This explains the increase of the PDE with V_{over} discussed in the previous subsection [44, 39].

The choice of the operating overvoltage therefore results from a compromise between different detector properties. A larger V_{over} generally provides a higher gain and photon-detection probability, but it also affects the probability of noise avalanches and correlated effects. In par-

ticular, dark-count rate, optical crosstalk and afterpulsing become increasingly important at high electric fields. These effects will be discussed in detail in Sec. 3.3. For this reason, SiPM characteristics are commonly reported as a function of V_{over}, rather than of the absolute bias voltage [45, 39].

An additional important property of V_{BD} is its temperature dependence. As the temperature increases, interactions of accelerated charge carriers with the silicon lattice become more frequent. Consequently, a larger electric field, and therefore a larger applied voltage, is required to reach the condition for self-sustaining avalanche multiplication. To first order, this dependence can be approximated as

$$V_{\mathrm{BD}}(T) = V_{\mathrm{BD}}(T_0) + \alpha_{\mathrm{BD}}(T - T_0), \tag{66}$$

where α_{BD} is the break-down-voltage temperature coefficient and T_0 is a reference temperature. Its value depends on the device structure and is typically of the order of a few tens of $\mathrm{mV/K}$ [24, 39]. Therefore, if V_{bias} is kept constant while the temperature changes, the actual operating overvoltage also changes. This effect must be taken into account whenever SiPM properties are compared at different temperatures, either by correcting the applied bias or by explicitly evaluating V_{over} at each temperature.

Experimentally, V_{BD} can be determined using different methods. A common procedure consists of measuring the gain as a function of the applied bias and extrapolating the approximately linear dependence to zero gain. Alternatively, the onset of the Geiger regime can be identified directly from the current-voltage characteristic, for example by analysing its derivative. Small differences between the break-down voltages obtained with different methods can occur because they probe different aspects of the transition between the avalanche-photodiode and Geiger regimes [45, 39].

3.2.6 Dynamic range

An important consequence of the multi-cell structure of a SiPM is its capability to provide information on the number of detected photons. Since the individual SPADs are operated in Geiger mode, the signal produced by a single microcell is essentially independent of the number of photons that initiate the avalanche. Photon-number information is instead recovered by combining the responses of the many independent microcells composing the SiPM. When several cells fire simultaneously, their signals are summed at the common output and the total collected charge is approximately proportional to the number of fired cells, as visible in Figure 61b [45, 39].

This proportionality is valid as long as the number of detected photons is sufficiently small compared with the number of available microcells. If two or more photons reach the same microcell within a time interval shorter than its recovery time, they cannot produce independent full-amplitude signals. As the incident photon flux increases, the probability of this occurrence becomes progressively larger and the SiPM response deviates from linearity until it eventually approaches saturation.

For an instantaneous and spatially uniform light pulse, neglecting correlated noise and assuming that each cell can fire only once, the mean number of fired cells can be approximated

by [45, 39]

$$N_{\text{fired}} = N_{\text{cell}} \left[1 - \exp\left(-\frac{N_\gamma \, \text{PDE}}{N_{\text{cell}}} \right) \right], \tag{67}$$

where N_γ is the number of photons incident on the SiPM and N_{cell} is the total number of microcells. In the low-occupancy limit,

$$N_\gamma \, \text{PDE} \ll N_{\text{cell}}, \tag{68}$$

Eq. (67) reduces to

$$N_{\text{fired}} \simeq N_\gamma \, \text{PDE}, \tag{69}$$

and the detector response is therefore approximately linear. Conversely, when the number of detected photons becomes comparable with or larger than the number of cells, an increasing fraction of photons reaches microcells that have already fired and the response progressively saturates.

The dynamic range of a SiPM is consequently strongly related to the total number of microcells. For a fixed sensor area, reducing the cell pitch increases the number of available cells and extends the range over which the detector response remains approximately linear. Smaller cells also have a lower capacitance and consequently a shorter recovery time, further increasing the capability of the detector to respond to high photon rates. These advantages are accompanied by other design trade-offs: reducing the cell dimensions generally decreases the gain of the individual cell and may reduce the geometrical fill factor because a larger fraction of the sensor area is occupied by isolation structures, quenching elements and other inactive regions [39, 45].

The simple expression in Eq. (67) represents an idealized case. The actual response also depends on the spatial and temporal distribution of the incident light. In particular, if the duration of the light pulse is comparable with or longer than the cell recovery time, the same microcell may partially or completely recharge and fire more than once during a single optical pulse. The effective dynamic range can therefore extend beyond that predicted by the instantaneous-pulse approximation. Conversely, non-uniform illumination can lead to an earlier local saturation of the most strongly illuminated regions. Correlated avalanches, such as optical crosstalk and afterpulsing, introduce additional deviations from the ideal response and must be included when a precise reconstruction of the incident photon number is required [45, 39].

3.2.7 Timing resolution

In addition to single-photon sensitivity, one of the most relevant characteristics of SiPMs is their capability to determine the arrival time of individual photons with high precision. The intrinsic timing performance is commonly characterized through the *Single Photon Time Resolution* (SPTR), defined as the statistical uncertainty in the reconstructed arrival time of a single detected photon [44, 45, 39]. Experimentally, the SPTR is generally measured by illuminating the detector with short laser pulses at sufficiently low intensity that the probability of detecting more than one photon per pulse is negligible, and studying the distribution of the delay between the reference laser pulse and the SiPM signal. Some examples can be seen in Figure 64.

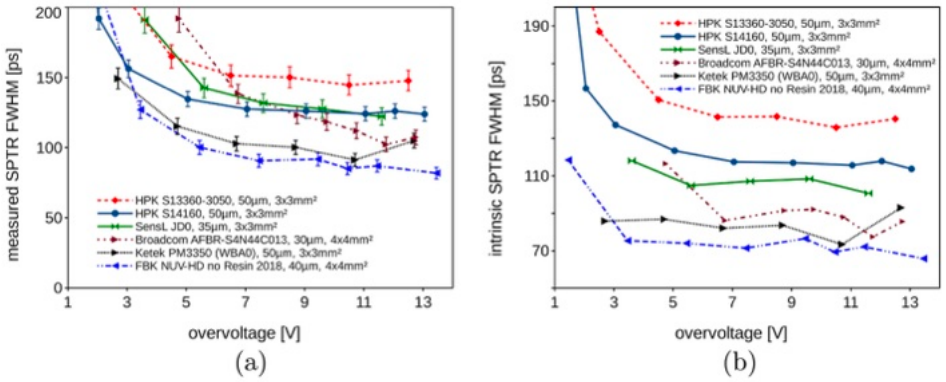


Figure 64: Single-photon time resolution (SPTR) as a function of the SiPM overvoltage for devices from different manufacturers. (a) Measured SPTR, including the contributions from the laser pulse width and electronic noise. (b) Intrinsic SPTR obtained after quadratically subtracting the electronic-noise contribution and the 42 ps FWHM laser pulse width. Figure from [39].

The finite width of this distribution originates from several physical mechanisms. The avalanche multiplication itself is a stochastic process, and therefore the time required for the initial carrier to develop into a macroscopic avalanche fluctuates from event to event. Additional fluctuations can arise from non-uniformities of the electric field across the active region and from differences in the position at which the photon is absorbed. Photo-generated carriers must travel from the absorption point to the high-field multiplication region, resulting in a *collection time spread* (CTS). The magnitude of this contribution depends on the device structure and on the photon wavelength, since the photon absorption depth in silicon is wavelength dependent [44].

When considering the complete SiPM rather than an individual SPAD, further contributions appear. Avalanches generated in different microcells can experience different propagation times towards the common readout electrode, giving rise to a *transit time spread* (TTS). The magnitude of this effect depends on the sensor geometry, on the microcell interconnection scheme and on the position of the firing cell within the device. Electric-field non-uniformities near the borders of individual microcells can also worsen the timing resolution because avalanches generated in these regions develop differently from those initiated in the central high-field region [44, 45].

The measured arrival-time distribution is therefore generally characterized by a relatively narrow prompt component accompanied by a tail towards longer times. The latter is mainly associated with photons absorbed in undepleted regions of the silicon. In this case, the generated carriers can reach the active region by diffusion rather than by the much faster drift process, producing delayed avalanches. The importance of this contribution depends on the thickness and carrier lifetime of the undepleted regions and becomes particularly relevant for structures optimized for longer wavelengths [44].

The measured timing resolution also includes an important contribution from the front-end electronics. For a leading-edge time measurement, the electronic contribution to the timing jitter can be approximately expressed as

$$\sigma_{t,\mathrm{el}} \simeq \frac{\sigma_V}{\left.\dfrac{dV}{dt}\right|_{\mathrm{thr}}}, \tag{70}$$

where σ_V is the RMS voltage noise at the amplifier output and $dV/dt|_{\mathrm{thr}}$ is the signal slew rate at the discrimination threshold [44, 39]. A low-noise and sufficiently high-bandwidth readout is therefore required in order to exploit the intrinsic timing capability of the SiPM. In particular, the total SiPM capacitance becomes increasingly important for large-area devices, since it can reduce the amplitude and the rising-edge slope of the fast single-cell signal. The fast component transmitted through the quenching capacitance C_q is consequently especially valuable in timing-oriented readout schemes [45, 39].

Modern SiPMs can achieve single-photon time resolutions of the order of a few tens of picoseconds, although the exact value depends strongly on the device dimensions, internal structure, and readout electronics. For $1 \times 1\ \mathrm{mm}^2$ devices, single-photon time resolutions of approximately 40–$80\ \mathrm{ps}$, expressed as the full width at half maximum (FWHM) of the single-photon timing distribution, have been reported [24]. The operating overvoltage also plays an important role, since it affects the electric field, the avalanche-triggering probability, and the amplitude of the single-cell signal. The operating point must nevertheless be selected taking into account the noise performance of the device, since increasing V_{over} also enhances dark counts and correlated avalanches. These contributions and their dependence on the operating conditions are discussed in the following section.

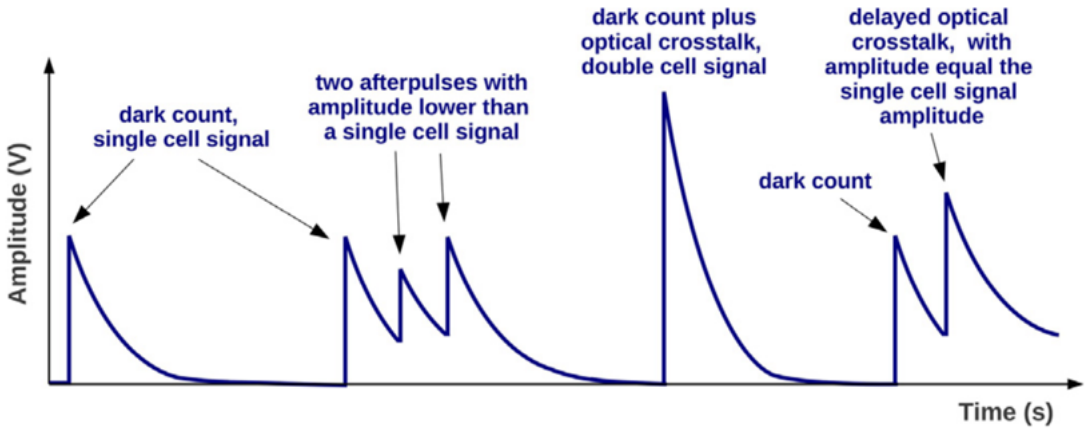


Figure 65: Representation of the SiPM output signal of the different kinds of noise observables: primary events, prompt crosstalk, afterpulsing and delayed crosstalk events. Image from [45].

.

3.3 Noise and dark-count rate

In SiPMs, noise sources are generally classified into primary contributions, such as DCR, and secondary (or correlated) noises, which identifies all the avalanche pulses following a primary event, such as afterpulse and crosstalk. Figure 65 shows a schematic representation of the amplitudes of the different SiPM noise contributions as a function of time.

3.3.1 Dark Count Rate

The main source of noise in a SiPM is the dark count rate (DCR) [24], which is mainly caused by free charge carriers generated within the depletion region and subsequently reaching the high-field region, where they can trigger a Geiger discharge. Since the break-down of a micro-cell is independent of whether it is initiated by a photon-generated carrier or by a thermally generated one, the resulting signals are indistinguishable. Two main processes contribute to the generation of free charge carriers.

The first mechanism is trap-assisted generation, in which an electron is promoted from the valence band to the conduction band through an intermediate energy state, or trap, located within the band gap. These traps originate from imperfections in the silicon lattice and may be introduced during device fabrication or generated by radiation damage. This contribution can be described using Shockley–Read–Hall (SRH) statistics [39].

The rate of trap-assisted generation depends strongly on temperature, as well as on the density and energy distribution of the available trap states, which may arise from impurities and structural defects. As an empirical rule of thumb, the dark current of SiPMs is often observed to decrease by approximately a factor of two for every $10^{\circ}\mathrm{C}$ reduction in temperature [45]. However, the exact temperature dependence is device-specific and can vary with the operating conditions and the temperature range considered.

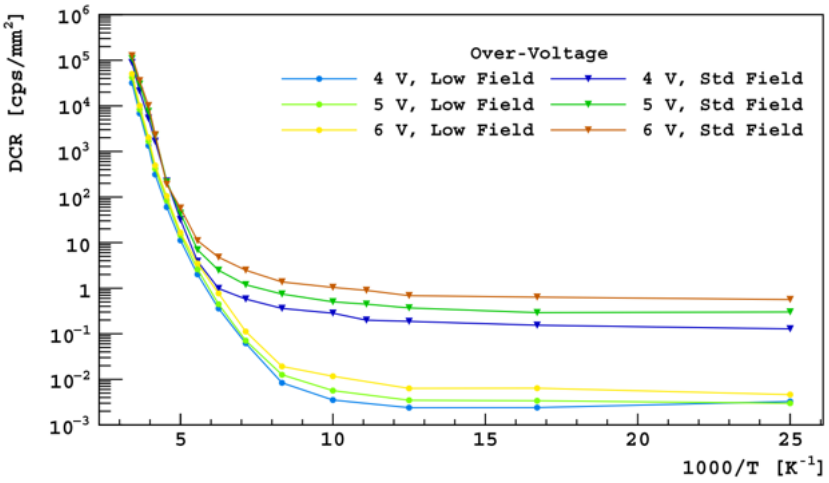


Figure 66: DCR for NUV-HD FBK-SF (triangular markers) and NUV-HD-LF (circular markers) SiPMs as a function of function of 1/T and overvoltage. Image taken from [46].

The second contribution is given by direct and trap-assisted band-to-band tunnelling. In this case, the generation rate is mainly determined by the strength of the electric field and shows only a weak dependence on temperature. For this reason, an appropriate design of the electric field is particularly important for reducing the DCR at low temperatures. At room temperature the first mechanism dominates, while at cryogenic temperatures ($T < 100–150K$), the dark rate is basically a band-to-band tunnel effect. The different temperature dependency of the two contributions can be seen clearly in Figure 66, where at high temperatures the trap-assisted generation is dominant and there is a strong temperature dependency, meanwhile at low temperatures we enter in the tunnelling regime and the DCR saturates.

The DCR increases with increasing bias voltage and linearly with the active area.

3.3.2 Afterpulse

The afterpulse is caused by carriers trapped within the cells during a primary avalanche and subsequently released, or by secondary photons; it therefore usually occurs in the same SPAD where the primary signal was detected. The afterpulse is caused by the capture of charge carriers by impurities or defects in the silicon, resulting in a delayed release, or by secondary photons produced during the avalanche and re-absorbed in the same SPAD. The afterpulse is problematic because it can be mistakenly interpreted as the arrival of a subsequent photon, generating false signals and compromising the accuracy of measurements. Furthermore, the afterpulse can affect the time resolution of the SiPM, introducing delays or distortions in the arrival times of photons [39]. The afterpulse probability depends on the number of traps, release time constants and the recharge time of the SPAD. Thus the main approaches to reduce afterpulse are:

1. Reduce the number of traps using pure Silicon wafers;

2. Use specific materials or geometries, known as low-lifetime substrates, to reduce the release time constant [47];

3. Tune the recharge time constant in order that most of the trapped charge carrier release happens when the cell is not yet (fully) recharged [39].

3.3.3 Prompt, delayed and external crosstalk

Each avalanche in a SPAD produces optical photons, that can travel to a neighbouring SPAD and trigger an avalanche; this process is called crosstalk. Depending on the travel path of the optical photon, one can have prompt, delayed or external crosstalk:

- **Prompt CT**: Prompt crosstalk, called also direct crosstalk, happens when the optical photon is absorbed in the active volume of the neighbouring SPAD (see Figure 67, on the left). The direct crosstalk probability has its peak value when the primary avalanche has reached roughly its maximum of produced e-h pairs, this happens some hundreds of picoseconds after the initial event [39]. Due to the small distance the light has to travel, prompt crosstalk is timed very close to the primary event, starting from almost no time delay to a delay of, at maximum, several hundreds of picoseconds. The result is two SPADs firing at the same time, giving 2 PE signals. A way to reduce prompt crosstalk is to introduce optical trenches in-between the cells.

- **Delayed CT**: Delayed crosstalk happens when the optical photon is absorbed in the non depleted volume of the second SPAD (see Figure 67, in the middle). The generated carriers must diffuse to the multiplication region, taking a time on the order of nanoseconds, resulting in a delayed correlated signal. However, part of the delayed crosstalk events will occur too shortly following the first avalanche in which case this type of noise is not distinguishable from prompt crosstalk. It should be mentioned that delayed afterpulsing might lead to the same type of signal at the SiPM terminal as delayed crosstalk, this especially if the SPAD of the afterpulse event was almost or already fully charged [39]. Since the thickness of the superficial un-depleted region of the entrance window is normally very small, it can be assumed that the most important contribution to the delayed crosstalk comes from the region below the active area [44]. Delayed crosstalk can be minimized by reducing the un-depleted region thickness, using a substrate with a low minority carrier lifetime [47], a buried junction below the active area to prevent diffusion or by extending the trenches deep into the substrate to confine the optical photons within the same SPAD.

- **External CT**: The external crosstalk happens when the photon is emitted towards the optical window, is reflected and re-enter the SiPM and triggers another avalanche (see Figure 67, on the right). The secondary avalanches are no longer limited to neighbouring SPADs any more and the external crosstalk probability might exceed the other sources of crosstalk [39]. Crosstalk in general and in particular external crosstalk can be minimized by reducing the SPAD gain, which will reduce the produced e-h pairs and, hence, the generated optical photons.

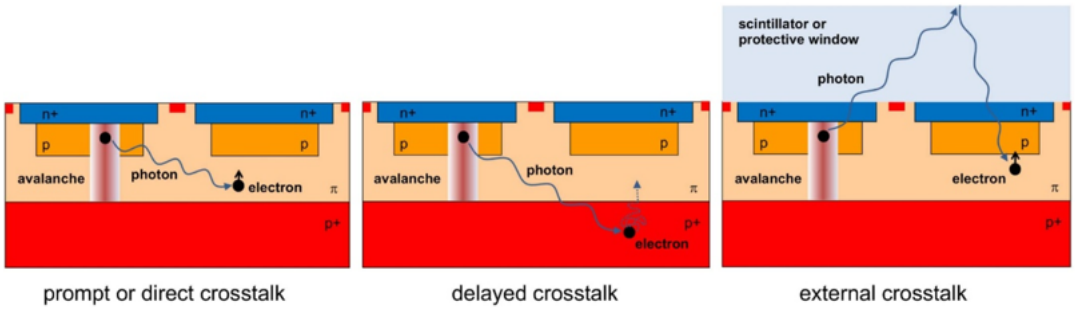


Figure 67: Different types of crosstalk observable in analog SiPMs. Reprint from [44].

3.3.4 Dark current and leakage current

Under reverse bias and in the absence of illumination, an operational distinction can be made between the current measured below and above the breakdown voltage. Below breakdown, the measured current is generally referred to as the reverse leakage current, I_{leak}, and is associated with thermally generated and leakage carriers without Geiger-mode multiplication. Above breakdown, the device operates in Geiger mode and the current measured in the dark, I_{dark}, also includes the contribution of self-sustained avalanches triggered by thermally or field-generated carriers [24]. This distinction is therefore useful from an experimental point of view, rather than representing a fundamental separation between two independent current components. The effect of radiation damage is different on these two currents.

The I_{leak} shows little proportionality to the bias voltage; it is composed of the charges generated in the detector when no outside radiation in entering the detector and, considering that the bias voltage is below the break-down value, there is no Geiger multiplication. The main mechanisms contributing to the leakage current are the diffusion of minority charge carriers from quasi neutral regions into the depleted region and the generation of electron-hole pairs due to defects in the depletion region [40]. Therefore the leakage current shows a strong dependence on the temperature, divided into the two contributions:

$$I_{diff} \propto T^3 e^{-\frac{E_g}{kT}}, \tag{71}$$

$$I_{gen} \propto T^2 e^{-\frac{E_a}{kT}}, \tag{72}$$

where E_g is the gap energy and E_a is the activation energy.

Also for the dark current the main mechanisms are the diffusion of minority carriers (Eq. 71) and the generation of electron-hole pairs, but Eq. 72 has to be modified to take into account also the tunnelling effect, obtaining:

$$I_{gen+tun} \propto (1+\Gamma) T^2 e^{-\frac{E_a}{kT}}, \tag{73}$$

where Γ is the electric-field-dependent enhancement factor associated with trap-assisted tunnelling. It accounts for the increase in the carrier-generation rate with respect to purely ther-

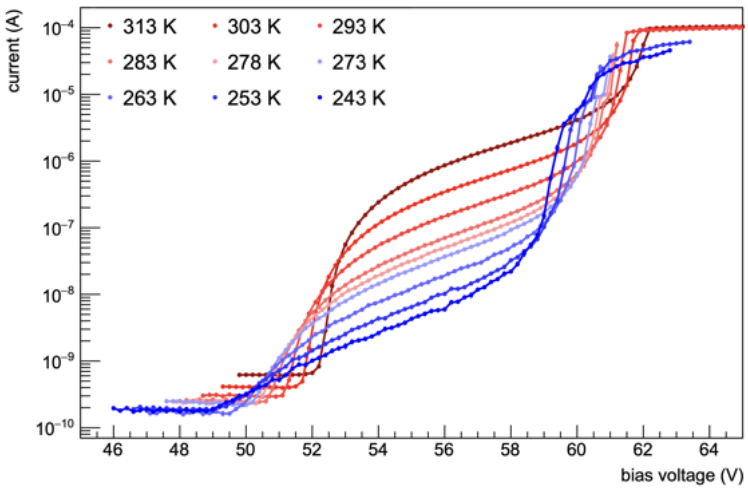


Figure 68: Dark current as a function of the bias voltage of a reverse biased Hamamatsu S13360-3050VS SiPM. Image taken from [24].

mal generation and becomes more relevant as the electric field increases. For a more detailed discussion on the topic, refer to [40].

The I_{dark} can be expressed with the following formula:

$$I_{dark} = DCR(T, V_{over}) \cdot G(V_{over}) \cdot q \cdot ECF(V_{over}) \tag{74}$$

where DCR is the primary dark count rate, G is the gain and ECF is the excessive charge factor, which quantifies the amount of excess charge produced per primary avalanche by correlated noise. So the dark current has a strong dependency both on the temperature and the bias voltage, as it can be seen from Figure 68.

3.4 Radiation damage of SiPM

The use of SiPMs in high-energy and nuclear-physics experiments naturally raises the question of their stability in radiation environments. This aspect is particularly relevant for detectors located relatively close to an interaction point, where the integrated particle fluence can become significant over the lifetime of the experiment. At distances of the order of one metre, fluences up to approximately 10^{14} cm^{-2} are expected in the most demanding HL-LHC environments, while values of the order of 10^{11}–10^{12} cm^{-2} are representative of the radiation levels considered for the future Electron Ion Collider [24]. Since SiPMs rely on the detection of individual charge carriers, even relatively small modifications of the silicon crystal can strongly affect their noise performance.

Radiation damage in silicon is conventionally separated into two main classes according to the mechanism through which the incident radiation transfers energy to the material: Ionizing Energy Loss (IEL), associated mainly with the generation of charge in dielectric layers and at interfaces, and Non-Ionizing Energy Loss (NIEL), which produces displacement damage in the silicon crystal [40, 24]. The two mechanisms affect different regions of the device and consequently

produce different macroscopic signatures. Their microscopic origin and their consequences for SiPM operation are discussed in the following subsections.

3.4.1 Ionizing and Non-Ionizing Energy Loss

Non-Ionizing Energy Loss corresponds to the fraction of energy transferred by an incident particle to the silicon lattice through atomic collisions. When the transferred energy is sufficiently large, a silicon atom can be displaced from its equilibrium lattice position. The resulting structural modifications are referred to as *bulk damage*, since they occur throughout the sensitive silicon volume. Neutrons and energetic charged hadrons are particularly effective in producing this type of damage, although sufficiently energetic electrons and photons can also contribute [40].

Since the displacement efficiency depends on both particle type and energy, fluences produced by different radiation fields cannot be compared directly. The standard convention is to express the displacement damage in terms of the equivalent fluence of $1\,\mathrm{MeV}$ neutrons,

$$\Phi_{\mathrm{eq}} = \kappa\Phi, \tag{75}$$

where Φ is the actual particle fluence and κ is the corresponding hardness factor. The hardness factor is obtained from the non-ionizing energy deposited by the radiation relative to that produced by $1\,\mathrm{MeV}$ neutrons in silicon. This NIEL scaling provides a practical method to compare irradiation campaigns performed with particles of different types and energies, although it remains an approximation because it does not explicitly account for differences in the microscopic defect populations or for their subsequent evolution [48, 40]. It is also useful in experimental simulations where there are many particles with different energies, to predict the NIEL damage received

Ionizing Energy Loss instead corresponds to energy transferred to the electronic system of the material. In SiPMs its most relevant effects occur in the dielectric layers and at the Si–SiO_2 interfaces. In SiO_2, approximately one electron–hole pair is generated for every $18\,\mathrm{eV}$ of deposited energy. The electrons have high mobility and can generally leave the oxide rapidly, whereas a fraction of the holes becomes trapped in pre-existing or radiation-induced states. The accumulation of positive oxide charge, together with the formation of interface traps, modifies the electric-field distribution close to the silicon surface and generates additional surface current [40, 24].

Ionizing damage is usually quantified through the *Total Ionizing Dose* (TID), expressed in Gray(Gy), whereas displacement damage is described through Φ_{eq}. These quantities therefore characterize two physically different processes and should not be directly interchanged. Moreover, the classification does not imply that a given radiation species produces exclusively one type of damage. Energetic protons, electrons and photons may deposit energy through both ionizing and non-ionizing processes; which contribution dominates depends on the particle energy and on the detector structure [40].

The distinction is experimentally visible in the electrical response of irradiated SiPMs. Ionizing irradiation primarily modifies surface-related currents and can significantly increase the current measured below break-down. Displacement damage, instead, introduces generation-

active defects in the silicon bulk and produces a particularly strong increase of the current above break-down and of the DCR.

3.4.2 Point and cluster defects

The microscopic structure of displacement damage depends on the amount of energy transferred to the recoiling silicon atom. Approximately $25\,\mathrm{eV}$ are required to remove an atom from its lattice position. The displaced atom, referred to as a *Primary Knock-on Atom* (PKA), leaves behind a vacancy and occupies an interstitial position elsewhere in the lattice. The vacancy–interstitial pair represents the simplest radiation-induced point damage [40, 24].

Vacancies and interstitials are not necessarily stable in their initial configuration. They become mobile at temperatures above approximately $150\,\mathrm{K}$ and can migrate through the lattice. A vacancy and an interstitial may recombine, restoring the original lattice configuration, or the defects may migrate towards surfaces and interfaces. Alternatively, they can interact with other lattice imperfections or impurities and form more stable complexes. Examples include vacancy–phosphorus (VP), vacancy–oxygen (VO), divacancy (V_2) and trivacancy (V_3) configurations [40]. This evolution is important because the electrically active defects observed after irradiation do not necessarily correspond to the primary defects produced during the initial collision.

When the PKA receives a sufficiently large kinetic energy, the process is no longer limited to a single displacement. For PKA energies above approximately $1\,\mathrm{keV}$, the recoiling atom can displace additional silicon atoms and initiate a collision cascade, producing a spatially concentrated region of damage. These structures are referred to as *cluster defects*. At still higher recoil energies, above approximately $12\,\mathrm{keV}$, several clusters may be generated by the same primary interaction [40]. The distinction explains why $1\,\mathrm{MeV}$ electrons predominantly create isolated point defects, whereas protons and neutrons can efficiently produce both point and cluster damage.

Clusters are of particular interest for SiPMs because their effect can be strongly localized. Instead of producing a uniform increase of the generation rate across the device, a single particle interaction may create a highly active region within one microcell. Measurements performed with digital SiPMs, where the DCR can be monitored independently for each SPAD, provide a direct example of this behaviour. As shown in Figure 69, individual cells can exhibit sudden step-like increases in DCR as the proton fluence is accumulated, with the magnitude of the increase varying considerably from one cell to another. Such observations are consistent with the formation of localized cluster-like defects [49, 40].

The formation and subsequent evolution of point and cluster defects provide the microscopic basis for the macroscopic degradation observed after irradiation. However, the presence of a geometrical lattice distortion alone is not sufficient to determine its impact on SiPM performance. What ultimately matters is whether the resulting defect introduces an electrically active state in the silicon band gap.

3.4.3 Radiation-induced energy levels in the band gap

In an ideal silicon crystal, the valence and conduction bands are separated by a band gap of approximately $E_g \simeq 1.12\,\mathrm{eV}$ at room temperature. Structural defects break the periodicity

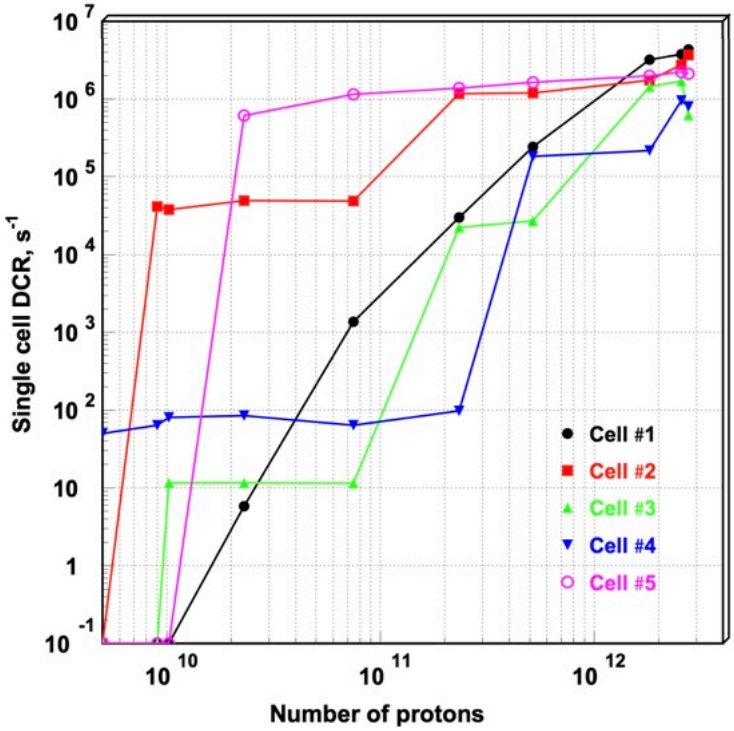


Figure 69: Dark-count rate of individual cells of a digital SiPM as a function of the accumulated number of incident protons. The step-like variations show the localized character of radiation-induced damage and are compatible with the creation of strongly generation-active cluster defects. Figure from [49].

of the lattice potential and can introduce allowed electronic states inside this forbidden region. The electrical effect of a radiation-induced defect is therefore determined not only by its physical structure, but also by the energy position of the corresponding level and by its carrier-capture cross sections [40].

The principal effects are schematically represented in Figure 70. Deep levels located close to the middle of the band gap are particularly efficient generation-recombination centres. They allow electrons and holes to be generated through intermediate states, strongly increasing the probability of thermal carrier generation. Since the middle of the silicon band gap lies at approximately $0.56\,\mathrm{eV}$ from either band edge, defects introducing levels in this region have a particularly important contribution to the generation current.

Other defect states can act predominantly as trapping centres. A charge carrier generated by a particle or photon can be captured and subsequently released after a characteristic time that depends on the position of the trap and on its capture and emission probabilities. Deep states have longer release times and may therefore prevent the trapped carrier from contributing to the original signal. The same physical concept is related to the trapping mechanisms already introduced in the discussion of afterpulsing in Sec. 3.3, although here the emphasis is on defects

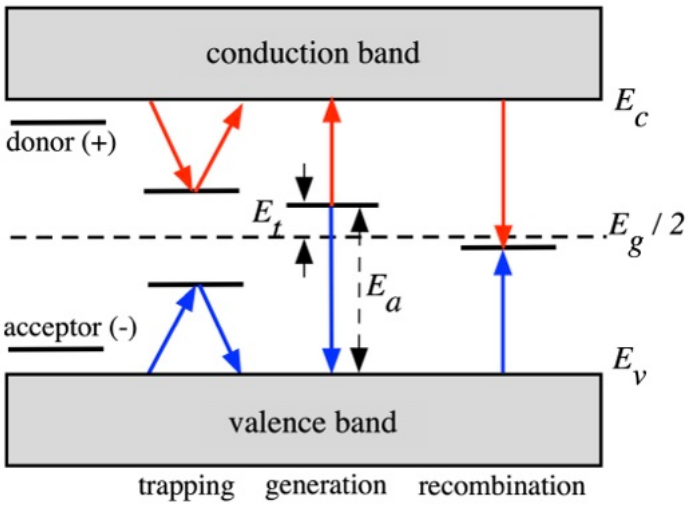


Figure 70: Schematic representation of defect levels inside the silicon band gap and of their principal macroscopic effects: generation-recombination, carrier trapping and modification of the effective doping concentration. Figure from [40].

introduced or modified by radiation.

Finally, electrically charged defect levels can alter the effective space-charge density of the material. Their occupation changes the effective donor or acceptor concentration and consequently modifies the electric-field profile inside the depleted region. In parallel, nuclear interactions may deactivate electrically active dopants. At sufficiently high radiation levels these effects can modify the depletion conditions and produce a measurable variation of the breakdown voltage [40, 24]. Such variations are usually much smaller than the increase observed in dark noise at moderate fluences, but become relevant when severe displacement damage is accumulated. The importance of these levels also explains why a simple count of displaced atoms is not sufficient to predict detector degradation. Two irradiation conditions producing a similar amount of displacement damage may generate different populations of electrically active states. In particular, deep levels located near mid-gap are of primary relevance for SiPM operation because the carriers generated through them can subsequently trigger Geiger avalanches. The following subsection considers this connection at the macroscopic level.

3.4.4 Increase of generation current and DCR

The physical mechanisms responsible for dark-carrier generation and their dependence on temperature and electric field have already been introduced in Sec. 3.3. Radiation does not introduce a new type of SiPM signal; instead, it increases the concentration of electrically active centres from which carriers can be generated in the absence of light. The consequences are therefore observed mainly as an increase of the generation current and, when the generated carriers reach the high-field region, of the DCR [40].

This amplification of the microscopic effect is particularly important in a Geiger-mode detector. A carrier released by a radiation-induced defect and a carrier generated by the absorption

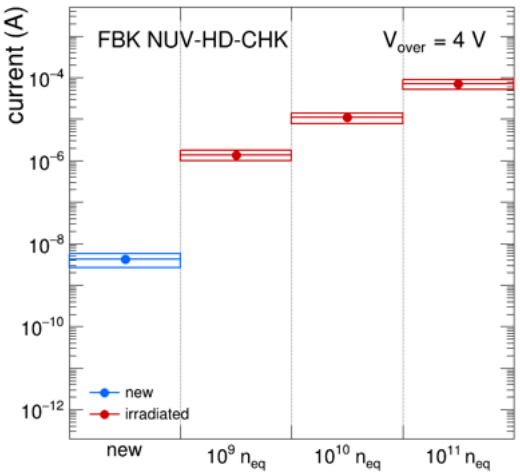


Figure 71: Dark current measured at -30°C for FBK NUV-HD-CHK SiPMs exposed to different neutron-equivalent fluences. The points represent the average measured over several sensors, while the rectangles indicate the RMS dispersion of the sample. Figure from [41].

of a photon produce the same avalanche once they enter the multiplication region. Therefore, a relatively small increase in the probability of spontaneous carrier generation may result in a large increase in the number of full-amplitude dark pulses.

Bulk and surface radiation damage contribute differently to this process. Bulk defects located in, or sufficiently close to, the depleted region can generate carriers that directly reach the high-field region and trigger an avalanche. Surface generation centres instead primarily increase the non-multiplied current. They contribute to the DCR only when the device electric-field configuration allows the generated carriers to migrate towards the multiplication region [40, 24]. This distinction explains why the variation of the current below and above break-down can be very different after irradiation.

For hadron irradiation, the most prominent experimental signature at moderate fluence is a strong increase of dark current and DCR, while several other SiPM parameters remain substantially unchanged. Measurements performed on different devices commonly show an approximately proportional increase of the radiation-induced current with neutron-equivalent fluence. This behaviour is visible in Figure 71, obtained in the irradiation programme developed for the EIC dRICH photosensors. Measurements performed at -30°C show a systematic increase of the dark current with fluence for FBK devices [41].

At moderate displacement-damage levels the correlation between DCR and dark current introduced in Sec. 3.3 remains a useful description, since gain and correlated-noise parameters are generally found to vary much less than the primary dark-event rate. At very high fluence this simple interpretation becomes progressively less accurate: the large DCR increases the average microcell occupancy, modifies the effective operating conditions of individual cells and makes a direct pulse-counting determination of the DCR increasingly difficult [40]. One of the first

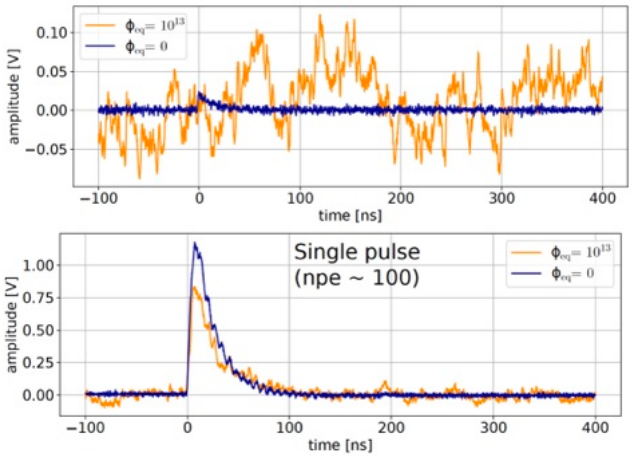


Figure 72: Single-photoelectron charge signals measured at $V - V_{\mathrm{BD}} = 3\ \mathrm{V}$ before irradiation and after neutron irradiation to $\Phi_{\mathrm{eq}} = 10^{9}\ \mathrm{cm}^{-2}$. The resolved photoelectron structure visible before irradiation is lost after irradiation at room temperature. Figure from [50].

experimental consequences of this high-noise regime is the disappearance of the single-photon structure of the SiPM signal.

3.4.5 Loss of single-photon resolution

For a non-irradiated SiPM operated under suitable conditions, the charge spectrum obtained with low light intensity exhibits distinct peaks corresponding to different numbers of fired microcells. This separation is possible because individual avalanches are sufficiently well isolated and the fluctuations of the baseline remain small compared with the single-cell signal.

As the DCR increases after irradiation, dark avalanches occur at progressively shorter time intervals. Their signals start to overlap with one another and with the signal induced by incident photons. The resulting pile-up increases the baseline fluctuations and broadens the charge distribution until neighbouring photoelectron peaks can no longer be separated. The detector therefore loses its *single-photon counting resolution*, even though the individual microcells remain capable of undergoing Geiger discharges.

An example is shown in Figure 72, where the single-photoelectron structure is clearly visible before irradiation and disappears after neutron irradiation to $\Phi_{\mathrm{eq}} = 10^{9}\ \mathrm{cm}^{-2}$ at room temperature. The exact fluence at which this transition occurs depends on the sensor design, operating temperature, overvoltage and irradiation conditions. As a general scale, the single-photon resolution of room-temperature SiPMs is commonly lost in the range $\Phi_{\mathrm{eq}} \sim 10^{9}$–$10^{10}\ \mathrm{cm}^{-2}$ [50, 40].

The loss of peak separation should not be interpreted as a complete loss of photosensitivity. A sufficiently intense optical signal can still be distinguished from the noise background after the single-photoelectron peaks have disappeared. This distinction is important for applications in which many photons are detected simultaneously, such as in calorimetry, but it represents a severe limitation for detectors that rely on individual-photon identification.

At still higher DCR, the effect extends beyond the broadening of the charge spectrum. A microcell requires a finite time to recover after each avalanche. If a new dark event occurs before the cell is completely recharged, the avalanche develops at a lower effective overvoltage and produces a reduced signal. When this condition occurs in a significant fraction of cells, the number of fully available microcells decreases and the effective detector response can also be reduced [40].

This aspect is particularly important for Cherenkov detectors such as the ePIC dRICH, where individual photons carry the information required to reconstruct the Cherenkov ring. Radiation-induced DCR must therefore be maintained below a level at which single-photon discrimination becomes inefficient. Cooling, timing selection and recovery through annealing are consequently integral parts of the SiPM operating strategy in this application [24, 41].

3.4.6 Fluence and temperature dependence

The degradation produced by radiation is determined by the combined effect of the accumulated displacement damage and the operating conditions of the SiPM. Fluence controls primarily the number and distribution of radiation-induced defects, while temperature controls the probability that electrically active defects generate carriers during detector operation. It is therefore useful to keep these two dependences conceptually separated.

For hadron irradiation, the increase of dark current and DCR is the dominant effect over a broad range of moderate fluences. Up to approximately $\Phi_{\mathrm{eq}} \sim 10^{12}\ \mathrm{cm}^{-2}$, measurements on several SiPM technologies generally show no large variation of parameters such as pixel capacitance, quenching resistance, gain or PDE, within the experimental sensitivity, while dark current and DCR increase considerably [40]. The precise limit is device dependent and should therefore not be regarded as universal.

At larger fluence, the very high dark-event rate itself begins to influence the SiPM response. A useful quantity to describe this regime is the average dark-count occupancy, defined as the fraction of time for which a microcell is unavailable, or only partially recovered, because of dark avalanches. For a SiPM with N_{pix} microcells and recovery time τ, and assuming that all cells have, on average, the same DCR:

$$\eta_{\mathrm{DC}} = \frac{\tau\, ECF}{N_{\mathrm{pix}}}\, DCR. \tag{76}$$

The occupancy therefore increases with the DCR and consequently with radiation fluence. The effect is illustrated in Figure 73. At sufficiently high fluence a non-negligible fraction of the cells is not fully recovered when a photon arrives. This can reduce the effective PDE and signal amplitude even if the intrinsic photon-conversion mechanism of an individual available cell has not changed [40].

The effect of temperature on the dark-carrier generation mechanisms has already been discussed in Sec. 3.3 and illustrated in Figure 66. The same principles apply after irradiation: lowering the operating temperature suppresses thermally activated generation through defect levels and can therefore strongly reduce the radiation-induced DCR. The underlying lattice damage, however, is still present. Cooling reduces the probability that an electrically active defect generates a detectable carrier; it does not restore the original crystal structure.

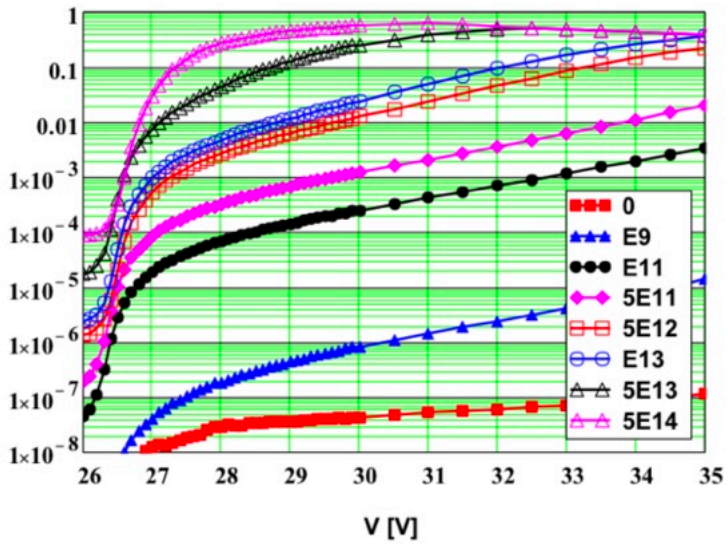


Figure 73: Average dark-count occupancy as a function of bias voltage for SiPMs exposed to different neutron-equivalent fluences and operated at -30°C. The increasing occupancy at high fluence reduces the fraction of fully recovered cells available for photon detection. Figure from [40].

This distinction is especially important when comparing the temperature at which the irradiation is performed with the temperature at which the damaged device is subsequently operated. No strong difference has been observed in the permanent hadron-induced damage when otherwise comparable SiPMs are irradiated at room or cryogenic temperatures, whereas the DCR measured after irradiation changes considerably with the operating temperature [24]. At sufficiently low temperature, thermally activated generation is progressively suppressed and field-assisted tunnelling becomes relatively more important, as already discussed in Sec. 3.3.

The combined dependence on fluence and temperature determines the practical radiation tolerance of a SiPM. A fluence at which single-photon detection becomes impossible at room temperature may still be compatible with operation at reduced temperature. Dedicated studies for the EIC dRICH have therefore characterized irradiated sensors at -30°C and investigated fluences up to $10^{11}\ n_{\mathrm{eq}}/\mathrm{cm}^2$, above the expected integrated radiation level of the detector. A progressive increase of both dark current and DCR is observed with fluence, while the low-temperature operation substantially reduces their absolute values [41, 24].

Cooling acts only on the manifestation of the radiation damage during operation. A complementary strategy is to modify the defect population itself by promoting defect migration and recombination through thermal treatment. This process, known as *annealing*, is discussed in Sec. 3.5.

3.5 Radiation-damage mitigation

The results presented in Sec. 3.4 show that the main effect caused by radiation damage in SiPMs is the progressive increase of the dark count rate, which can compromises single-photon detection at high integrated fluences. For applications in which individual photons have to be identified, such as the ePIC dRICH, radiation tolerance must therefore be addressed through a combination of complementary strategies. Some of them act only on the operating conditions of the detector, reducing the contribution of radiation-induced dark counts without modifying the underlying defect population. Others, most notably thermal annealing, can partially recover the sensor by promoting the reduction of the radiation-induced dark count rates [24, 40, 41].

The principal mitigation strategies considered for SiPM-based photon detection are discussed in the following. Particular attention is given to thermal annealing, since the study of the evolution of radiation-induced defects during successive annealing treatments represents the central subjects of this thesis.

3.5.1 Operation at reduced temperature

Operating the SiPM at reduced temperature is one of the most immediate methods to limit the consequences of radiation damage. As discussed in Sec. 3.3, lowering the temperature strongly suppresses the thermally activated component of the DCR. Radiation-induced defects are not removed by this procedure; instead, their probability of generating carriers able to trigger an avalanche is reduced. Cooling must therefore be regarded as an *operational mitigation* rather than as a recovery of the damaged material.

This distinction becomes particularly important after irradiation, when the concentration of generation-active centres is much larger than in a new sensor. Measurements performed on irradiated SiPMs show that the temperature dependence remains strong also after substantial neutron fluence. In the temperature interval between room temperature and approximately -30°C, a reduction of the DCR by roughly a factor of 2–3 every 10°C has been observed. Consequently, cooling can extend by several orders of magnitude the radiation level at which useful signals remain distinguishable from the dark background [51].

A direct example is shown in Figure 74. At room temperature the optical signals of increasingly irradiated devices rapidly disappear into the dark background, whereas at -30°C the same laser pulses remain visible up to considerably larger fluences. At still lower temperatures, thermal generation can be suppressed even further. After high-temperature annealing, single-photon detection has been demonstrated at $77\,\mathrm{K}$ for devices exposed to fluences as high as $10^{14}\ n_{\mathrm{eq}}/\mathrm{cm}^2$, with DCR values below the kHz level [51].

For the ePIC dRICH, low-temperature operation is part of the detector design rather than an auxiliary laboratory procedure. The SiPMs are foreseen to operate at sub-zero temperature, around -30°C, using an actively controlled cooling system integrated into the photodetector unit [31]. Cooling alone, however, cannot prevent the progressive accumulation of radiation damage and is therefore combined with the additional strategies described below.

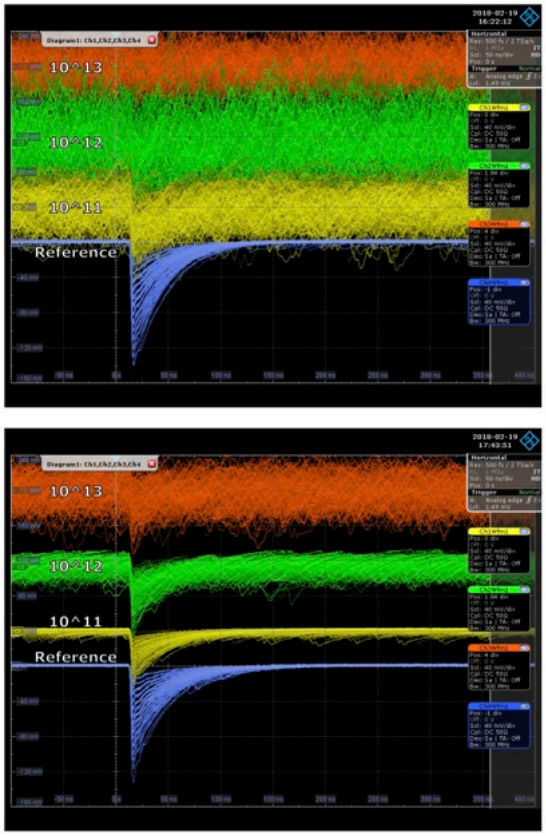


Figure 74: Response to $400\ \mathrm{nm}$ laser pulses of Hamamatsu S13360-1350CS SiPMs before irradiation and after neutron-equivalent fluences of 10^{11}, 10^{12} and $10^{13}\ n_{\mathrm{eq}}/\mathrm{cm}^2$. Measurements are shown at room temperature (top) and at $-30°$C (bottom), illustrating the reduction of the radiation-induced background obtained by cooling the sensors. Figure from [51].

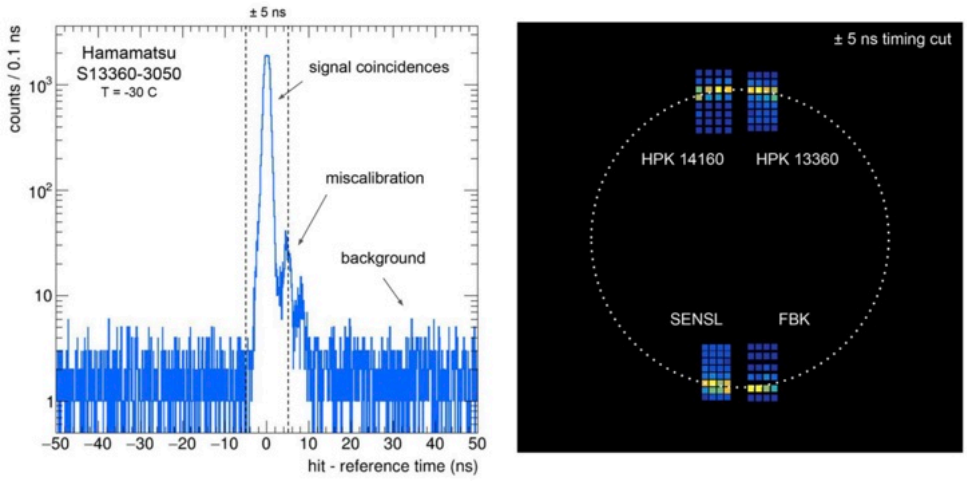


Figure 75: Use of timing information for DCR suppression in the dRICH SiPM readout. Left: time distribution of the detected hits, showing the signal coincidence peak, residual calibration structures and the random DCR background; the ± 5 ns selection window is indicated. Right: accumulated Cherenkov-ring hits after application of the timing selection. Figure from [24].

3.5.2 Narrow timing windows

A second mitigation strategy exploits the fact that dark counts and Cherenkov photons have very different temporal characteristics. Dark events occur randomly in time, whereas signal photons are correlated with the passage of the charged particle and therefore with the collision time. If sufficiently precise time information is available, only hits recorded within a narrow interval around the expected signal arrival time need to be accepted.

For a random dark-count rate DCR and a time window of duration Δt, the mean number of accidental dark hits accepted per channel is approximately

$$N_{\mathrm{dark}} \simeq DCR\,\Delta t, \tag{77}$$

provided that $DCR\,\Delta t \ll 1$. Reducing the time window therefore decreases the accepted dark background approximately linearly, without requiring any modification of the sensor itself. The effectiveness of this approach ultimately depends on the timing performance of the complete SiPM-electronics chain and on an accurate channel to channel time calibration.

This principle has been demonstrated with the dRICH prototype using ALCOR-based readout electronics. Figure 75 shows the time distribution measured during a beam test. The Cherenkov signal forms a well-defined coincidence peak, whereas random dark events generate an approximately flat background. Even a relatively broad ± 5 ns coincidence interval substantially suppresses the DCR contribution and allows the Cherenkov-ring pattern to emerge clearly after the timing selection [24].

This method is particularly attractive because its effectiveness does not decrease simply because the sensor has accumulated defects: as long as the total hit rate remains within the capabilities of the electronics, most temporally uncorrelated dark events can be rejected. High time

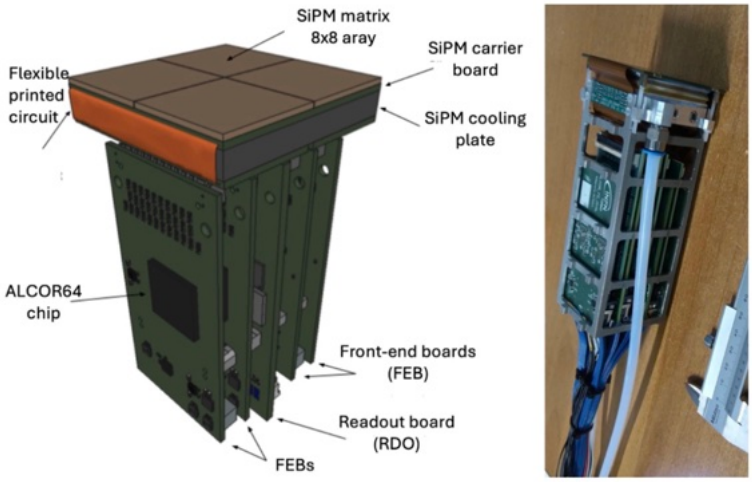


Figure 76: Concept of the ePIC dRICH photodetector unit and preliminary realization. The compact module integrates the SiPM sensor plane, front-end readout electronics and the thermal infrastructure required for low-temperature operation and in-situ annealing. Figure from [31].

resolution and programmable gating are consequently important components of the radiation-tolerance strategy for large SiPM-based single-photon systems.

3.5.3 Sensor and electronics optimization

Radiation tolerance also depends on the characteristics of the SiPM and on the readout architecture. Sensor selection should not be based on radiation hardness alone, but on the combined requirements on PDE, DCR, timing performance and operating voltage. Measurements performed on different commercial and prototype SiPMs show that devices characterized by a low initial DCR generally preserve an advantage after irradiation and annealing, making the non-irradiated dark-rate performance an important parameter when selecting the sensor technology [41, 24].

The operating point represents an additional degree of optimization. As discussed in Sec. 3.2, increasing the overvoltage improves gain and photon-detection probability but simultaneously increases the probability that a dark carrier triggers an avalanche. In a radiation environment, the selected overvoltage must therefore provide sufficient single-photon efficiency without unnecessarily amplifying the radiation-induced background. At the device-design level, reducing the volume of the high-field multiplication region has also been proposed as a possible way of decreasing the number of bulk defects capable of producing detectable avalanches while preserving the photosensitive area [24].

The front-end electronics must in turn remain efficient when the DCR progressively increases. Low electronic noise, high rate capability, precise time stamping and the possibility of implementing a programmable time gate are all required. In the ePIC dRICH these functions are integrated within compact photodetector units (PDUs), which combine SiPM sensors, ALCOR TDC electronics and the thermal system required both for sub-zero operation and for high-

temperature annealing. The present ALCOR architecture uses 64 channels ASIC that integrates amplification, discrimination and TDC electronics with time-over-threshold, with time binning below $50\ \mathrm{ps}$ [31]. Sensor, operating-point and electronics optimization therefore do not remove radiation-induced defects, but increase the range over which the detector remains usable before a recovery treatment becomes necessary.

3.5.4 Thermal annealing

Unlike cooling and timing selection, thermal annealing acts on the damaged material itself. The term *annealing* refers to the partial recovery of radiation-induced degradation when the sensor is maintained at an elevated temperature for a given period of time. The microscopic evolution is complex and depends on the type of defect, but the general interpretation is that increasing the temperature enhances the mobility and reconfiguration probability of radiation-induced defects. Vacancies and interstitials can recombine, while more complex defect configurations may transform into states with a lower electrical activity [48, 40].

The recovery is both time- and temperature-dependent. Different defect populations are characterized by different activation energies and therefore by different characteristic annealing times. As a consequence, the decrease of radiation-induced current is generally not described by a single universal time constant. At a fixed temperature a rapid initial recovery is commonly followed by a progressively slower evolution, while increasing the annealing temperature accelerates the process and can activate the recovery of defect populations that remain essentially stable at lower temperature [40, 51].

Annealing can already occur spontaneously at room temperature. Measurements on neutron-irradiated SiPMs show that the radiation-induced current progressively decreases even during storage at room temperature, while increasing the annealing temperature accelerates this recovery, leading to a faster reduction of the current [52].

Annealing is generally incomplete: after a given treatment a residual radiation-induced contribution may remain, indicating that some defects are stable under the applied conditions or require different temperature–time scales. Moreover, a precise microscopic identification of all the defect transformations responsible for the macroscopic recovery of SiPM performance has not yet been established. This is one of the motivations for studying the evolution of individual radiation-induced hotspots during controlled annealing treatments.

3.5.5 Conventional oven annealing

The most direct implementation of high-temperature annealing consists of placing the irradiated devices in a temperature-controlled oven. This approach provides accurate and homogeneous control of the annealing temperature and is therefore particularly suitable for systematic studies of the recovery as a function of time and temperature.

A long-duration study performed on neutron-irradiated SiPMs used a sequence of annealing temperatures progressively increasing from $50^{\circ}\mathrm{C}$ to $175^{\circ}\mathrm{C}$ over approximately 600 hours. For a device irradiated to $10^{11}\ n_{\mathrm{eq}}/\mathrm{cm}^2$, the DCR measured at $-40^{\circ}\mathrm{C}$ decreased from more than $700\ \mathrm{kHz}/\mathrm{mm}^2$ to approximately $40\ \mathrm{kHz}/\mathrm{mm}^2$. For a device irradiated to $10^{12}\ n_{\mathrm{eq}}/\mathrm{cm}^2$, the

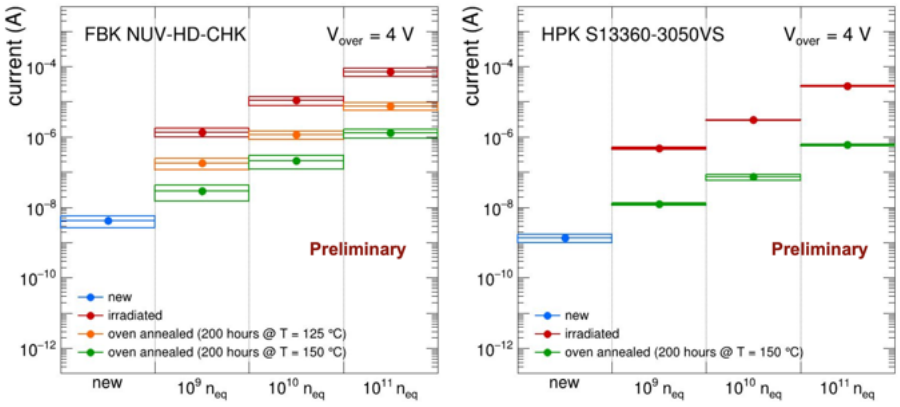


Figure 77: Dark-current measurements for FBK (left) and Hamamatsu (right) SiPMs exposed to different proton fluences and measured at different stages before and after high-temperature oven annealing. Points represent the average over several sensors, while rectangles indicate the RMS dispersion of the sample. Figure from [41].

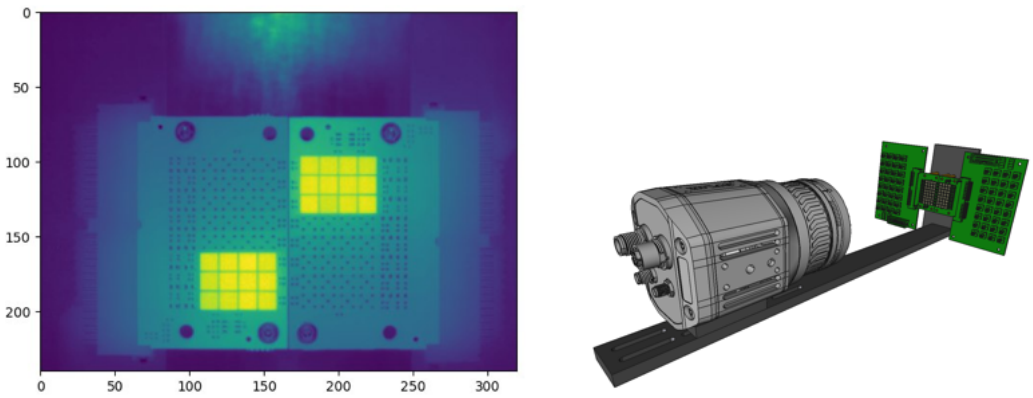
corresponding reduction was from about $7\ \mathrm{MHz/mm^2}$ to $400\ \mathrm{kHz/mm^2}$ [51]. The largest improvement at each temperature was observed during the first part of the treatment, followed by a tendency towards a plateau.

High-temperature oven annealing has also been extensively investigated for the ePIC dRICH $R\&D$. For proton-irradiated FBK SiPMs, the dark current was reduced by approximately one order of magnitude after 200 hours at $125°$C and by approximately two orders of magnitude after a subsequent treatment at $150°$C. A similar behaviour was observed for Hamamatsu devices [41]. The evolution is shown in Figure 77. The main limitation of oven annealing for a large detector is practical rather than physical. Sensors installed inside an experimental apparatus cannot in general be repeatedly removed, heated and reinstalled without significant intervention. Moreover, the maximum usable temperature and duration are constrained by the thermal compatibility of the sensor package, optical coating, resin, soldering and nearby electronics. For example, prolonged treatments above $100°$C may alter some encapsulating materials and must therefore be validated for the specific SiPM technology [51]. These limitations motivate annealing techniques that can be performed directly on the installed detector.

3.5.6 In-situ annealing

An alternative approach is to heat the SiPM electrically while it remains installed in the detector. When a SiPM is forward biased, its microcells behave as forward-polarized diodes and a substantial current can flow through the device and its quenching resistors. The corresponding Joule power raises the sensor temperature, allowing high-temperature annealing to be performed without an external oven. This method is generally referred to as *in-situ*, *online*, or *forward-bias* annealing.

A dedicated experimental system was developed within the ePIC dRICH $R\&D$ programme



(a) (b)

Figure 78: Left: sensor temperature monitored with an infrared camera during forward-bias heating. Right: schematic of the experimental setup developed for the in-situ annealing of SiPMs. The sensors are heated through forward-bias current, while their temperature is monitored with an infrared thermal camera and controlled through an Arduino-based feedback system. Figures from [35].

to investigate the possibility of performing annealing directly on the SiPMs exploiting the Joule heating. The objective is to bring each sensor to a controlled temperature and maintain it for a predefined time, while avoiding the need to remove the photodetectors from the experimental apparatus. In the adopted configuration, the SiPM is operated in forward bias so that the current flowing through the device produces Joule heating. The sensor temperature is calibrated as a function of the applied forward current and is continuously monitored with a FLIR infrared camera. An Arduino-based control system uses the temperature information to regulate the delivered power and maintain the desired annealing condition. For the Hamamatsu S13360-3050, S13360-3075, and S14160-3050 devices, corresponding to pixel pitches of $50\ \mu\mathrm{m}$, $75\ \mu\mathrm{m}$, and $50\ \mu\mathrm{m}$, respectively, a dissipated power of approximately $1\ \mathrm{W}$ per sensor was sufficient to raise the device temperature above 175°C [35].

The experimental arrangement developed for these measurements is shown in Figure 78.

The procedure was systematically tested at progressively increasing annealing temperatures, from 100°C to 175°C, and for different treatment durations. The evolution of the residual radiation damage is shown in Figure 79. Repeated measurements on several SiPMs showed a consistent recovery behaviour, with higher temperatures producing a faster reduction of the radiation-induced dark current. In particular, annealing at 150°C for approximately 100 hours was sufficient to recover about 97% of dark current induced by radiation damage, reaching a recovery level comparable with that obtained through conventional oven annealing. Increasing the temperature beyond 150°C did not result in a significant additional improvement of the dark-current recovery after prolonged annealing times [35]. These results demonstrate that forward-bias self-heating can provide an effective in-situ alternative to oven annealing while preserving the possibility of treating the sensors after their installation in the detector.

The concept has subsequently been incorporated into the dRICH photodetector architecture

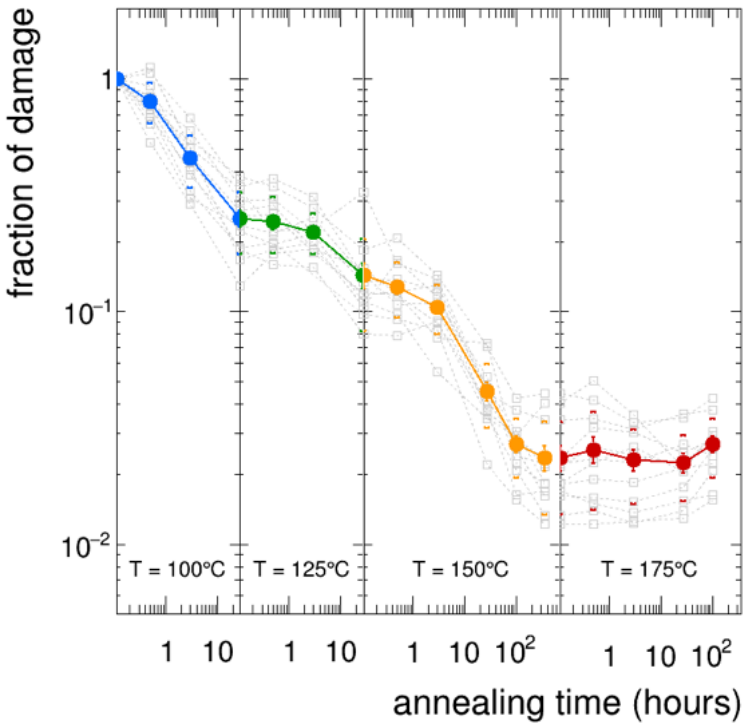


Figure 79: Fraction of residual radiation-induced damage as a function of the integrated annealing time for successive temperature steps at $100°$C, $125°$C, $150°$C and $175°$C. Measurements from individual SiPMs are shown together with the averaged evolution for each annealing temperature. Figure from [35].

[31]. The same thermal circuit used for sub-zero SiPM operation can be operated at elevated temperature to pre-heat the sensor carrier, reducing the electrical power that must be supplied through forward bias. This makes it possible to envisage repeated annealing cycles on selected portions of the detector during periods without beam, while more extensive treatments can be scheduled during longer shutdowns.

The combination of low-temperature operation, precise time selection, optimized sensor and readout characteristics, and periodic annealing therefore provides a complete strategy for maintaining the single-photon performance of SiPMs throughout operation in a radiation environment. Among these approaches, annealing is unique in directly modifying the population of radiation-induced defects. Understanding how individual defects respond to successive thermal treatments is consequently essential for optimizing the recovery procedure and motivates the defect-characterization studies discussed in Sec. 3.6.

3.6 Defect characterization

The discussion of radiation damage developed in the previous sections has so far been based mainly on macroscopic electrical observables, such as dark current and dark-count rate. These quantities provide an effective measure of the overall degradation of the SiPM, but they integrate the response of the complete device and do not retain information on the spatial distribution of the regions responsible for the increased avalanche activity. A complementary approach is therefore to exploit the optical emission produced during avalanche multiplication and image it with a sufficiently sensitive optical system.

This approach, generally referred to as *emission microscopy* or EMMI, provides spatial information that is not accessible from a conventional current measurement. Localized regions producing an enhanced optical signal can be identified, their positions can be followed under different operating conditions, and their evolution can be compared before and after irradiation or thermal treatments. The interpretation of these measurements nevertheless requires a clear distinction between a microscopic defect in the silicon and the optical signal ultimately recorded by the camera. This distinction will be central throughout the experimental analysis presented in the following chapters.

3.6.1 Light emission during avalanche processes

As discussed in Sec. 3.1, a Geiger discharge involves a large number of carriers accelerated by the intense electric field in the multiplication region. In addition to producing the electrical signal used for photon detection, this avalanche process is accompanied by the emission of secondary photons. The phenomenon has been investigated both through spectroscopic measurements and through spatially resolved imaging of biased SiPMs [53, 54].

Measurements performed on different SiPM technologies show that the secondary emission extends over a broad spectral interval from the visible region towards the near infrared. The underlying emission is associated with energetic carriers involved in avalanche multiplication, and different radiative processes can contribute in different spectral regions. Proposed mechanisms include indirect and direct interband transitions and intraband Bremsstrahlung emission [53]. An example of the measured emission spectrum is shown in Figure 80, where the broad spectral distribution of the secondary photons can be observed directly.

The measurements represented in Figure 80 also show that the amount of emitted light depends on the SiPM operating voltage. Increasing the overvoltage enhances the secondary-photon emission, indicating that the optical output of the avalanche cannot be considered independent of the electrical operating point [53]. Measurements in which avalanches were externally stimulated lead to the same conclusion: after normalizing the optical signal to the number of avalanches, the photon yield per avalanche was still observed to increase with overvoltage [54], as a consequence of gain increase with the overvoltage.

The temperature dependence provides a complementary result. Over the investigated temperature interval, no significant variation of the secondary-photon yield per avalanche was observed within the experimental uncertainties [54], as a consequence of the fact that the gain does not change with temperature at a fixed overvoltage. This distinction is particularly important for the interpretation of emission-microscopy measurements. If the optical intensity accu-

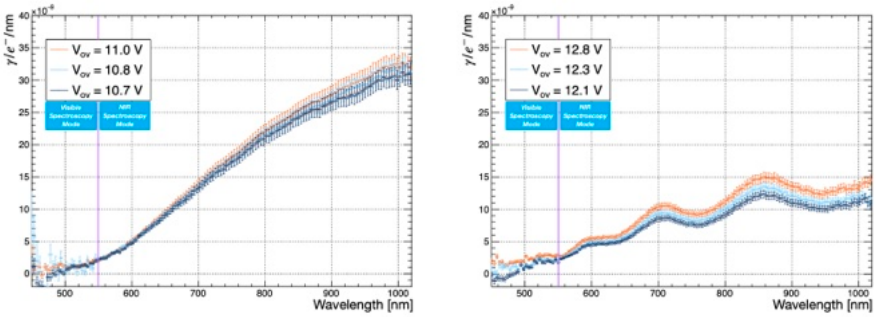


Figure 80: Example of the secondary-photon emission measured from a SiPM operated in Geiger mode. The emission extends over a broad wavelength range and depends on the operating overvoltage. Figure from [53].

mulated during a fixed exposure increases with temperature, the increase does not necessarily imply that each avalanche emits more photons. Instead, it predominantly reflect an increase in the number of avalanches occurring during the acquisition time. This point will be relevant in the quantitative analysis of the hotspot emission presented in the following chapters.

3.6.2 Emission hotspots and electrically active defects

If the dark-avalanche activity were completely uniform over the active area of a SiPM, the corresponding accumulated optical emission would also be expected to be approximately uniform once geometrical effects are taken into account. Experimentally, however, emission images show spatially localized regions with intensities substantially larger than their surroundings. These regions are conventionally referred to as *hotspots*.

A representative example is shown in Figure 81, where the optical emission recorded from a SiPM operated above break-down is clearly non-uniform. Localized bright regions can be distinguished within individual microcells, indicating that the avalanche activity is concentrated preferentially at specific positions of the device [55].

The spatial structures visible in Figure 81 are directly related to non-uniformities in the local dark-avalanche activity. Spatially resolved measurements showed that a relatively small fraction of the microcells containing high-intensity emission regions can contribute disproportionately to the total DCR. In particular, less than 5% of the investigated microcells were found to account for as much as approximately 56% of the total dark-count rate [55]. The temperature dependence of the emission associated with these regions was also found to be compatible with carrier generation dominated by Shockley–Read–Hall processes. These results provide the physical basis for using emission microscopy as a diagnostic of electrically active non-uniformities in a SiPM. A defect or defect complex capable of increasing the local carrier-generation probability can increase the frequency with which a particular region initiates Geiger discharges. Since every avalanche produces secondary photons, the enhanced local avalanche activity produces a correspondingly enhanced accumulated optical signal.

The relation is nevertheless indirect. The photons recorded by the camera are predominantly associated with the avalanche process, rather than constituting a direct optical image

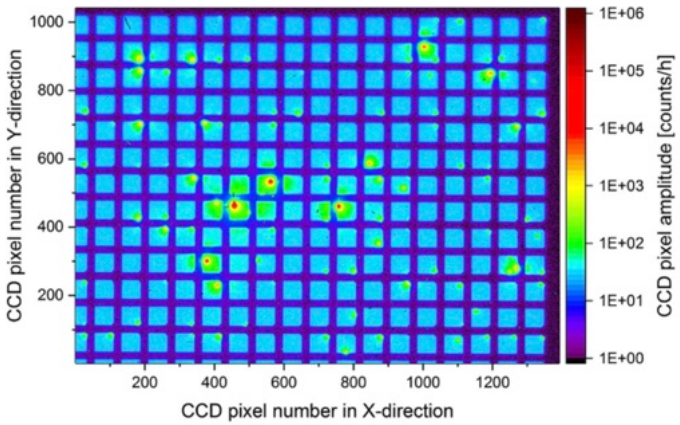


Figure 81: Spatially resolved optical emission from a SiPM operated above break-down. Localized regions with enhanced emission intensity are visible within a subset of the microcells and are identified as emission hotspots. Figure from [55].

of the atomic-scale defect itself. A hotspot should therefore be understood as a spatial region characterized by enhanced avalanche-associated emission. The microscopic defect, or group of defects, responsible for the enhanced carrier-generation probability is located at a much smaller scale than the optical resolution of the measurement. Furthermore increasing the optical magnification does not indefinitely improve the spatial resolution, since the achievable resolution is ultimately limited by diffraction. In the visible range, structures significantly smaller than the wavelength of the detected light cannot be spatially resolved, regardless of the magnification employed.

An additional limitation is introduced by the protective resin covering the SiPM surface. The emitted light propagates through materials with different refractive indices before reaching the imaging system, and the refractive-index discontinuity at the resin-air interface can affect the light propagation through refraction and related optical effects.

Consequently, a one-to-one correspondence between one detected hotspot and one physical lattice defect cannot in general be established.

3.6.3 Emission microscopy for SiPM characterization

Emission microscopy has consequently become a useful complementary technique for SiPM characterization. Spectroscopic and imaging studies have shown that different SiPM technologies exhibit different distributions of secondary emission and different populations of localized hotspots [53]. The technique therefore provides information not only on the total avalanche-related optical activity of the device, but also on its spatial distribution.

This capability is particularly relevant in the context of radiation damage. As discussed in Sec. 3.4, displacement damage introduces additional electrically active centres in the silicon bulk. If these centres enhance the local probability of carrier generation, they can increase the dark-avalanche activity of specific regions of the sensor. Emission microscopy can therefore

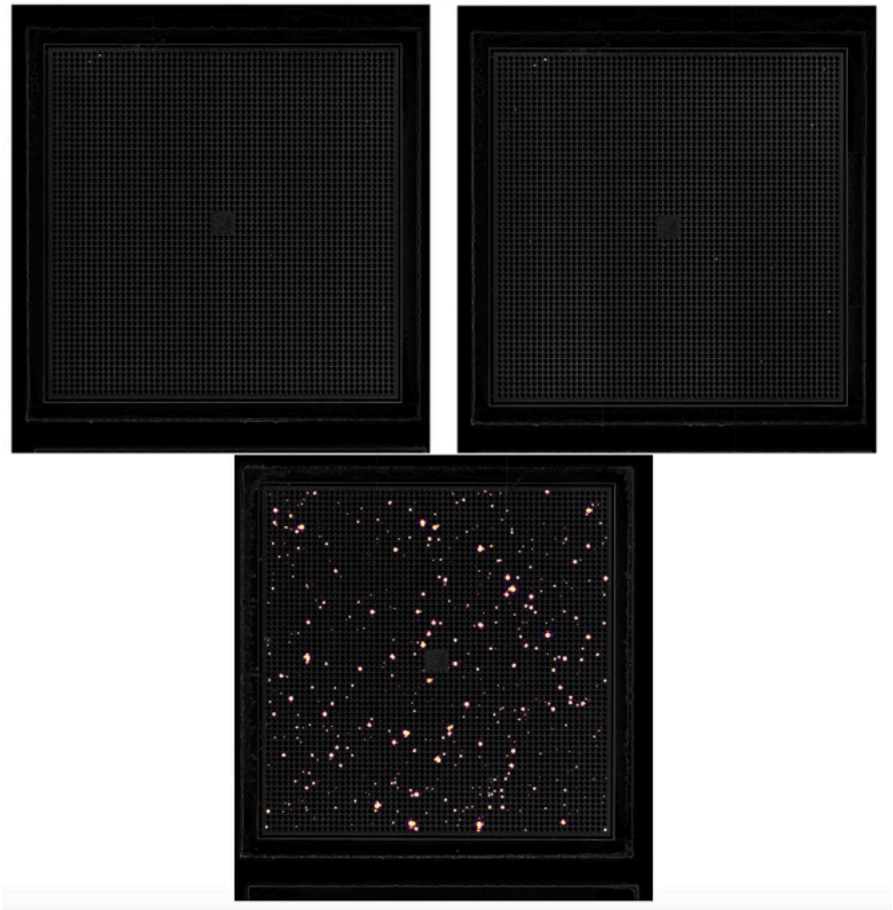


Figure 82: Emission-microscopy images of a Hamamatsu SiPM before and after irradiation. Upper left: intrinsic hotspots observed on a new SiPM at 5 V overvoltage. Upper right: intrinsic hotspots observed on a new SiPM at 7 V overvoltage. Bottom: irradiated sensor with radiation-induced hotspots. Figure from [56].

reveal radiation-induced modifications through the appearance of new hotspots, changes in the intensity of pre-existing sites, or both.

A dedicated application of this approach to the ePIC dRICH SiPM R&D programme has been developed at INFN Bologna [56]. High-resolution emission images were acquired under controlled temperature and bias conditions, first on non-irradiated Hamamatsu SiPMs and subsequently after proton irradiation to a neutron-equivalent fluence of approximately $10^9\ n_{\mathrm{eq}}/\mathrm{cm}^2$. The comparison obtained from these measurements is shown in Figure 82.

As visible in Figure 82, the non-irradiated device contains only a limited number of detectable emission sites, whereas after irradiation a much denser population of hotspots appears across the sensor surface. The upper panels also show that increasing the overvoltage from 5 to $7\ \mathrm{V}$ makes additional weak emission sites detectable even before irradiation. The number of experimentally observed hotspots therefore depends both on the underlying distribution of

electrically active regions and on the operating conditions of the SiPM [56]. After irradiation, the large increase in the number of light-emitting sites is accompanied by an increase in the device dark current, further supporting the connection between the optical pattern and the radiation-induced increase of avalanche activity.

This comparison illustrates the complementary information provided by emission microscopy and conventional electrical characterization. A current measurement establishes the overall increase of the dark activity of the sensor, whereas EMMI additionally provides information on *where* this activity is localized and on how many distinct emitting regions become experimentally observable. This spatial information is the basis for the hotspot-by-hotspot characterization and the study of their evolution during irradiation and annealing developed in the following chapters.

3.6.4 Observables and limitations of EMMI

The interpretation of emission images requires defining precisely which quantities are measured. In the experimental configuration used in this thesis, the camera integrates the optical emission produced during a finite exposure time. The measured intensity associated with a hotspot can therefore be represented schematically as

$$L_{\mathrm{meas}} \propto R_{\mathrm{av}}\, \Delta t\, Y_{\gamma}\, \varepsilon_{\mathrm{opt}}, \tag{78}$$

where R_{av} is the local avalanche rate, Δt is the exposure time, Y_{γ} is the average detected secondary-photon yield associated with an avalanche, and $\varepsilon_{\mathrm{opt}}$ accounts for the collection efficiency of the microscope and for the spectral response of the camera. The quantity referred to as *hotspot luminosity* throughout this thesis is therefore the integrated optical intensity measured by the experimental apparatus and is expressed in arbitrary units.

This definition specifies both the capabilities and the limitations of the technique. EMMI can identify the coordinates of localized regions of enhanced avalanche activity, compare their optical intensities, determine the spatial distribution of the visible hotspot population and follow the same regions under different operating or treatment conditions. By repeating the measurements at different temperatures and overvoltages, it is also possible to investigate how the accumulated emission associated with individual hotspots responds to changes in the SiPM operating point.

EMMI does not, however, directly resolve an atomic defect, determine its microscopic structure or identify its corresponding energy level in the silicon band gap. The number of detected hotspots is therefore not a measurement of the absolute number of lattice defects. Several microscopic generation centres may contribute to the activity observed within a single optical hotspot, while other defects may remain invisible because their contribution is below the detection threshold.

Likewise, the measured hotspot luminosity is not an intrinsic “luminosity of the defect”. Changes in the measured intensity can originate from changes in avalanche frequency, avalanche charge, correlated avalanches or secondary-photon yield. These contributions cannot be completely separated from a single integrated emission image. The present setup also does not perform spectroscopy: only photons transmitted by the optical system and falling within the

sensitivity range of the camera contribute to the measured signal. A hotspot that is no longer detected after a treatment must therefore not automatically be interpreted as evidence that the corresponding microscopic defect has physically disappeared. Its electrical activity may have decreased below the detection threshold, or its observable optical contribution may have changed.

These limitations do not reduce the usefulness of EMMI; rather, they define the physical meaning of its observables. In the present work, the terms *hotspot* and *light-emitting site* will consequently be used for the experimentally detected optical structures, while the term *defect* will be reserved, where appropriate, for the underlying microscopic modification of the silicon responsible for enhanced carrier generation.

3.6.5 Motivation for studying irradiation and annealing

The combination of radiation damage and annealing provides a particularly suitable context in which to exploit this spatial information. As discussed in Sec. 3.4, irradiation strongly increases the dark current and DCR through the introduction of additional electrically active centres. Conventional electrical measurements quantify the resulting degradation but cannot determine whether the increased current originates from a large number of moderately active regions or from a smaller population of particularly active sites.

Emission microscopy makes it possible to address this question experimentally by comparing the hotspot population before and after irradiation. Furthermore, once the same emitting regions can be identified across successive measurements, the technique can be extended from a static characterization to a study of their evolution. This is especially relevant for annealing, where the reduction of the total dark current indicates a partial recovery of the device but does not reveal how that recovery is distributed among individual spatial regions.

The first EMMI measurements performed within the Bologna programme established the appearance of a substantially larger light-emitting population after irradiation and proposed the systematic monitoring of these sites during subsequent thermal treatments [56]. The work presented in this thesis develops this programme further by quantitatively characterizing the detected hotspots and following their evolution throughout a controlled annealing sequence.

The following chapters therefore move from the general SiPM physics discussed so far to the experimental investigation developed for this purpose. The next chapter describes the EMMI apparatus, the acquisition and image-processing procedures, the identification and spatial matching of individual hotspots, and the method used to determine their optical intensity and associated uncertainties. The subsequent results are then organized around three complementary questions: how the measured hotspot emission depends on temperature and overvoltage, how irradiation modifies the number and intensity distribution of the light-emitting sites, and how this population evolves during successive annealing treatments. In this way, emission microscopy is used as a spatially resolved complement to conventional electrical characterization, with the objective of obtaining a more detailed picture of the formation and recovery of radiation-induced activity in SiPMs.

4 Emission Microscopy measurements

The aim of this work is to directly characterize the evolution of radiation damage in SiPMs through a series of high-precision emission microscopy measurements performed with a dedicated experimental setup. In particular, the study focuses on monitoring the evolution of the emission hotspots throughout different annealing steps and, for two of the investigated sensors, on characterizing the dependence of luminosity on temperature and overvoltage.

In Section 4.1 a description of the sensors under study is given, along with the workflow utilised for the whole analysis. Section 4.2 contains the description of the instrumentation necessary to perform the emission images measurements. In Section 4.3 a brief description of the codes for the data acquisition is presented, with a focus on the precaution taken in order to obtain a statistically meaningful measurement. Section 4.4 is devoted to the measures that have been implemented to reduce noise and align the images, so that they are consistent throughout the entire analysis. Section 4.5 contains a detailed description of how the hotspots were characterised, setting out the procedure for determining the coordinates and calculating the luminosity and its statistical uncertainty. Lastly, Section 4.6 describes how the systematic uncertainties have been evaluated.

4.1 Sensors under study

Four Hamamatsu SiPMs were investigated in this work: two sensors of type S13360-3050 and two sensors of type S13360-3075. According to the manufacturer data sheet [43], both types belong to the S13360 MPPC (Multi-Pixel Photon Counter) series and have an effective photosensitive area of $3.0 \times 3.0\ \mathrm{mm}^2$. Throughout this thesis, the two S13360-3050 sensors are referred to as A1 and A2, while the two S13360-3075 sensors are referred to as B1 and B2. The main characteristics of the investigated sensors are summarized in Table 2.

ID	Sensor type	Pixel pitch	No. SPADs	Fill factor	PDE	Gain	V_{break}
A1	S13360-3050	$50\ \mu\mathrm{m}$	3600	74%	40%	1.7×10^6	$53 \pm 5\ \mathrm{V}$
A2	S13360-3050	$50\ \mu\mathrm{m}$	3600	74%	40%	1.7×10^6	$53 \pm 5\ \mathrm{V}$
B1	S13360-3075	$75\ \mu\mathrm{m}$	1600	82%	50%	4.0×10^6	$53 \pm 5\ \mathrm{V}$
B2	S13360-3075	$75\ \mu\mathrm{m}$	1600	82%	50%	4.0×10^6	$53 \pm 5\ \mathrm{V}$

Table 2: Main characteristics of the SiPMs investigated in this work. The PDE values refer to a peak sensitivity at $450\ \mathrm{nm}$, with a spectral response range of 270–$900\ \mathrm{nm}$. The gain values are measured at $V_{\mathrm{over}} = 3\ \mathrm{V}$, while the breakdown voltage is specified at $T = 25^{\circ}\mathrm{C}$.

The sensors were irradiated in the experimental area operated by TIFPA at the Trento Proton Therapy Centre [57], using a proton beam of $\sim (140\ \mathrm{MeV}$. The irradiation level was expressed as a 1-MeV neutron-equivalent (n_{eq}) fluence of $10^9\ n_{\mathrm{eq}}/\mathrm{cm}^2$. The main goal of this work is to study hotspots' evolution after irradiation and after annealing cycles. Figure 83 represents the workflow followed to perform the study. Baseline studies were done on new SiPMs.

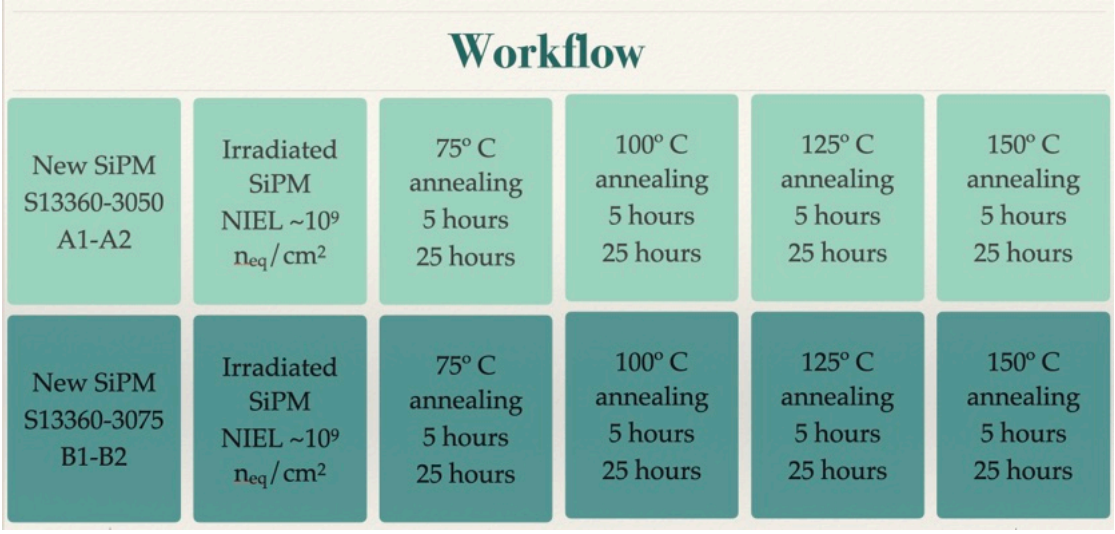


Figure 83: Diagram illustrating the workflow followed for SiPM studies. First stage on new sensors, second stage after irradiation, followed by eight successive annealing stages with progressively increasing temperature and duration.

4.2 EMMI Experimental Set-Up

The experimental set-up used to perform the measurements is presented in Figure 84. A single board of SiPMs is hosted in the apparatus during the data-taking, as visible in Figure 85. The temperature of the SiPM devices is controlled and stabilized via a Peltier-based system, which uses a temperature sensor mounted onto the SiPM carrier PCB board to provide feedback and temperature monitoring. Data are taken with EMMI (EMission MIcroscope), which is a system composed of a Thorlabs CERNA optical microscope, a 4x Nikon objective and a Thorlabs Quantalux Quantalux 2.1 MP sCMOS camera. The EMMI system is mounted on a motor-driven 3-axis moving system, in order to acquire images at different positions, and is placed inside an aluminum dark box, in order to ensure shielding from external light. A cooling unit is installed on the box wall, to prevent the air inside the box to overheat because of Peltier operation. The box is internally coated with neoprene and equipped with six removable patch panels, which provide access for the power-supply cables, data cables, and auxiliary services, including the fan connections, Arduino cables, the camera USB cable, and the cooling unit. A picoammeter, power supplies for the cooling system and the sensor, a fiber optic illuminator and the EMMI's PC complete the overall setup. A more thorough description of the components is reported in the following paragraphs.

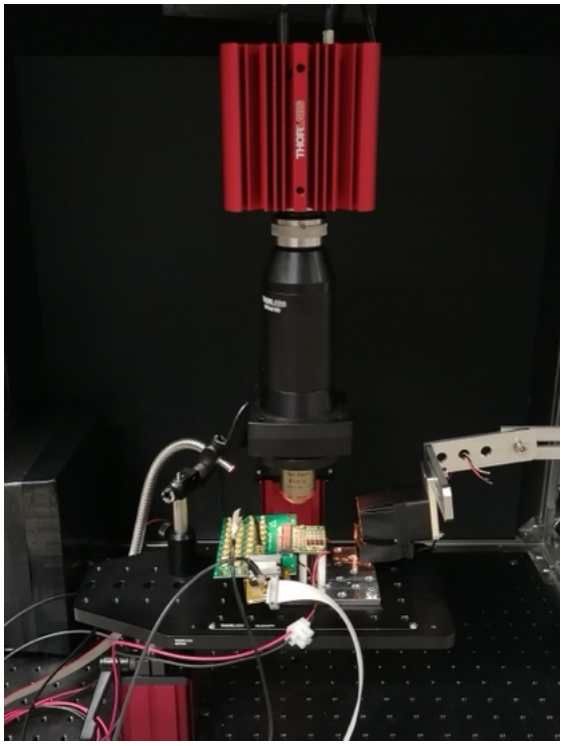


Figure 84: Picture of the EMMI (EMission MIcroscopy) experimental set-up, taken at INFN's Silicon Laboratory in Bologna.

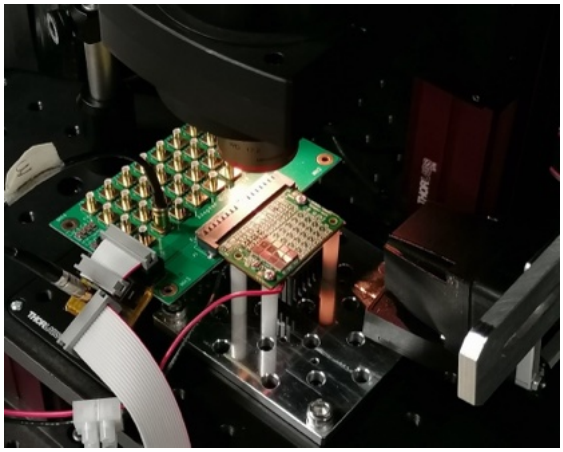
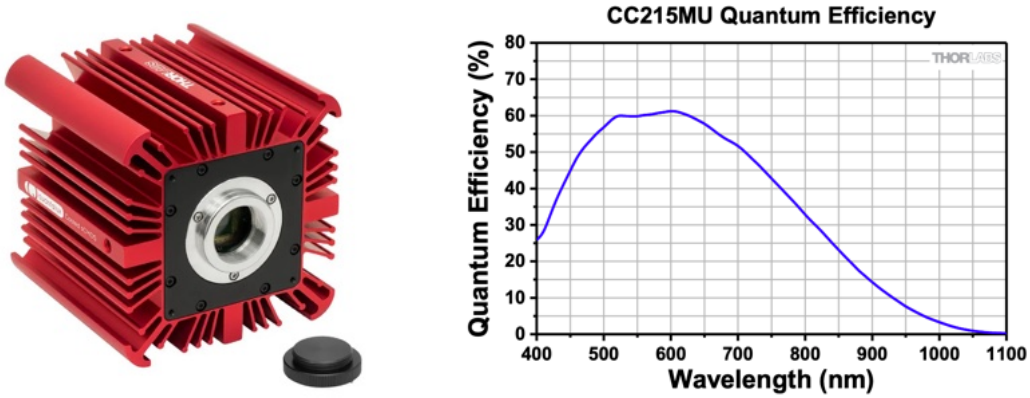


Figure 85: Board containing the four studied SiPMs: two sensors of type S13360-3050 and two sensors of type S13360-3075.



(a) (b)

Figure 86: (a) Picture of the Quantalux 2.1 MP sCMOS camera, model CC215MU. (b) CC215MU quantum efficiency.

4.2.1 Peltier cooler and temperature sensor

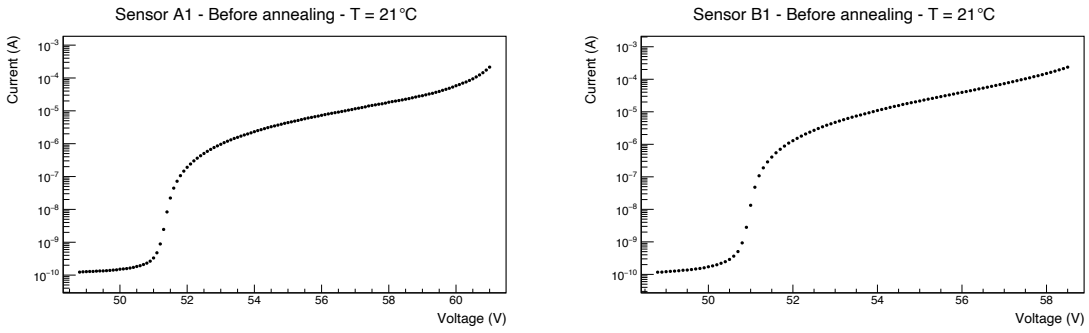
The SiPM board is mounted on a copper plate, beneath which is a Peltier module with a heat sink attached to its hot side; a fan blows air to effectively remove heat from the heat sink. The temperature is monitored by the LM73 Texas Instruments temperature sensor, mounted onto the SiPM carrier board, which sends feedback to the Peltier power supply to actively regulate the temperature of the SiPMs. The system changes the temperature by setting a new target value for the Peltier plate. The software then adjusts the Peltier voltage progressively until the sensor approaches the requested temperature. After each temperature change, the system waits for a fixed holding period of $300\ s$ before proceeding. This sequence is repeated for each target temperature, ensuring that measurements are taken only after a controlled stabilization interval.

4.2.2 EMMI acquisition system

Data are acquired using the EMMI system, whose main optical components are a Thorlabs Cerna Mini Microscope with Epi-Illuminator, a $4\times$ Nikon Plan Fluorite objective, and a Thorlabs Quantalux 2.1 MP sCMOS camera. The main components of the optical system, together with their corresponding product codes, are summarized in Table 3.

Component	Product code
Thorlabs Cerna Mini Microscope with Epi-Illuminator	SFM
Nikon Plan Fluorite $4\times$ Imaging Objective	N4X-PF
Thorlabs Quantalux 2.1 MP sCMOS Camera	CC215MU

Table 3: Main optical components of the EMMI system and their corresponding product codes.



(a) (b)

Figure 87: Two exemplary IV scans performed on the SiPMs under test, before annealing, at $T = 21°C$.

The CC215MU is a monochrome, actively cooled scientific camera specifically designed for low-light imaging applications. A picture of the camera used in the setup is shown in Figure 86a. The sensor consists of 1920×1080 active pixels, with a pixel size of $5.04 \times 5.04\ \mu\mathrm{m}^2$, corresponding to an active area of approximately $9.68 \times 5.44\ \mathrm{mm}^2$. The camera is optically coupled to the microscope through a C-mount interface, and the optical system provides a magnification factor of $4\times$ through the Nikon objective. Consequently, each camera pixel corresponds to an effective size of ($1.26 \times 1.26\ \mu\mathrm{m}^2$) on the SiPM plane. The detector is characterized by a low readout noise, below $1\ e^-$ median RMS, which is particularly advantageous for detecting the weak optical emission associated with SiPM avalanches. Its spectral response lies in the visible spectrum, with a peak quantum efficiency of approximately 61% at $600\ nm$. The wavelength dependence of the quantum efficiency is reported in Figure 86b.

4.2.3 SiPM biasing and I-V characterization

The Keithley 2450 SourceMeter is employed to supply the reverse bias voltage to the SiPM sensors and to measure the current. The electrical characterization of the SiPMs is performed by measuring the current as a function of the applied bias voltage, thus obtaining the corresponding I–V curves. The breakdown voltage is then extracted from these curves by identifying the onset of the Geiger multiplication regime. Since the breakdown voltage of a SiPM depends on temperature, its value cannot be considered constant during measurements performed at different temperatures. A linear temperature dependence is therefore assumed, with a reference breakdown voltage at $20°\mathrm{C}$ and a temperature coefficient describing its variation with temperature. In the present setup, the reference values have been measured to be 51.3 V for type-A sensors and 50.9 V for type-B sensors, with a temperature coefficient of about 55 mV/°C. For each measurement temperature, the corresponding breakdown voltage is calculated and used to determine the required bias voltage. The applied voltage is therefore adjusted according to $V_{\mathrm{bias}} = V_{\mathrm{break}}(T) + V_{\mathrm{over}}$, so that the overvoltage remains under control at each change in temperature. This ensures that measurements acquired at different temperatures are performed under comparable and controlled operating conditions.

4.2.4 Oven for annealing

The annealing treatments of the SiPMs were performed using a Memmert UF55 universal laboratory oven. The oven provides controlled and uniform heating, allowing the sensors to be kept at a well-defined temperature for the required annealing time. Its operating range extends up to $300^{\circ}\mathrm{C}$, with precise temperature regulation suitable for reproducible thermal treatments. The progressive annealing steps are set out in the Tab.4.

Annealing step	Temperature (°C)	Time (hours)	Total annealing time (hours)
Step 1	75	5	5
Step 2	75	25	30
Step 3	100	5	35
Step 4	100	25	60
Step 5	125	5	65
Step 6	125	25	90
Step 7	150	5	95
Step 8	150	25	120

Table 4: The progressive annealing steps adopted in this study.

4.3 Image acquisition

The imaging system is mounted on a three-axis motorized positioning stage, which controls the relative position between the microscope and the SiPM along the two directions parallel to the sensor surface and along the vertical z axis. Before each acquisition, the correct transverse position and vertical position are determined automatically.

The positioning in the x-y plane is obtained from the optical image of the SiPM by identifying a reference Through-Silicon Via (TSV), which provides a well-defined and reproducible geometrical feature of the device. A dedicated region of the image is searched for the TSV, whose centre is then used as the reference point for adjusting the transverse position of the motorized stage. An example of this procedure is shown in Figure 88, where the identified reference position is indicated within the search region. This procedure allows the same region of the SiPM to be located independently at the beginning of each measurement.

Once the transverse position has been determined, the optimal position along the z axis is found through an autofocus procedure. A sequence of images is acquired while progressively varying the distance between the objective and the SiPM. For each position, the image sharpness is quantified through the variance of the Laplacian: this quantity increases as the image becomes sharper and reaches a maximum close to the optimal focal plane. The dependence of the focus metric on the z position is shown in Figure 89. Around the maximum, the measured values are described by a quadratic function, and the position of its maximum is taken as the best-focus position.

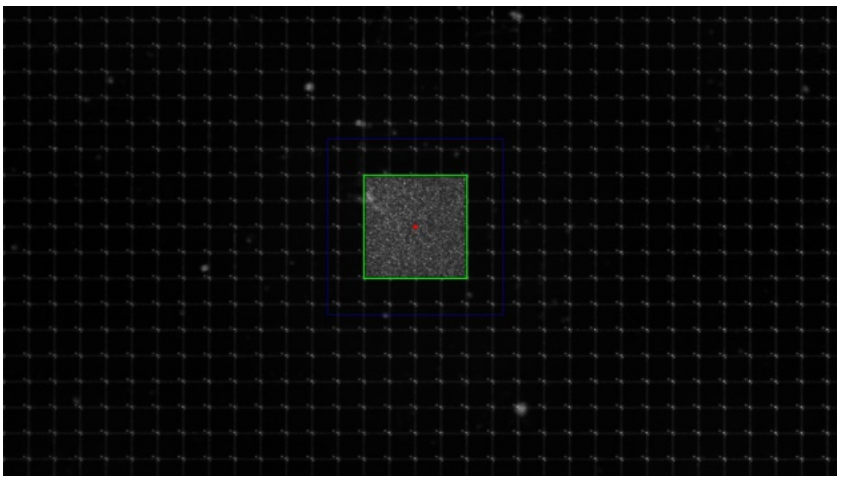


Figure 88: Example of the automatic transverse positioning procedure. A reference TSV is identified in the optical image and its centre is used to determine the position of the SiPM in the plane parallel to its surface.

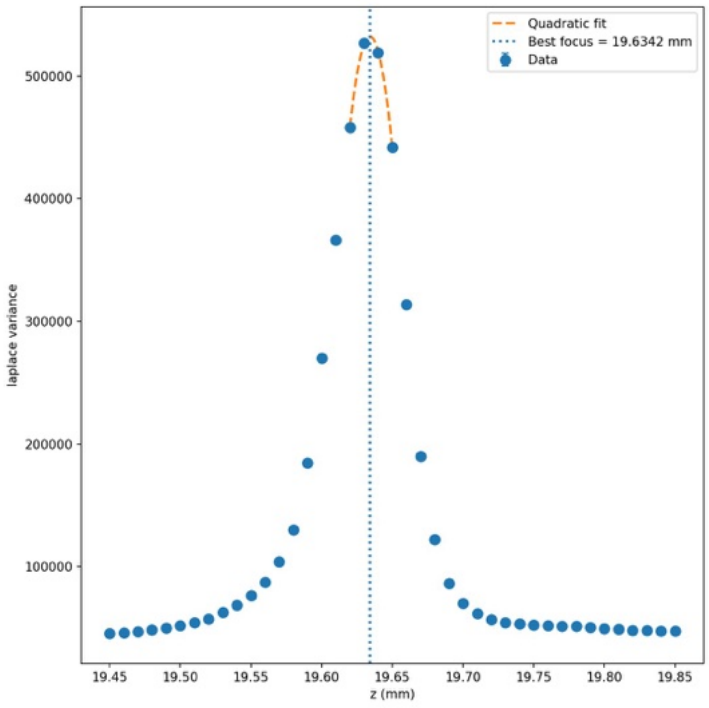


Figure 89: Example of the autofocus procedure. The image sharpness, quantified through the variance of the Laplacian, is measured as a function of the vertical position z. A quadratic fit is performed around the maximum, whose position defines the optimal focal distance.

The transverse centring and autofocus procedures are performed before each measurement. The motorized system is then moved to the newly determined x, y, and z positions, and the image acquisition is performed under the required experimental conditions. In this way, the positioning of the SiPM and the focal condition are independently re-established for every acquisition, reducing the sensitivity of the measurement to small mechanical changes between different runs. For full-device measurements, the system automatically moves over a predefined two-dimensional grid of positions, keeping the focal coordinate fixed, so that several partially overlapping images can be acquired to cover the entire SiPM.

Once each measurement condition is fixed, meaning the camera position with respect to the SiPM, temperature and bias voltage of the SiPM, the acquisition procedure starts. The basic principle of the measurement is the subtraction of the background contribution from the signal recorded while the SiPM is biased. In particular, the signal image is acquired with the SiPM switched on and therefore contains both the emission produced by the device and the background contribution, whereas the background image is acquired with the SiPM switched off. An optical image of the SiPM, illuminated using the fiber-optic illuminator, is also acquired.

A single acquisition performed by the camera is referred to as a *frame*, i.e. an image recorded with a specified exposure time. Since the maximum exposure time available for an individual camera acquisition is limited, measurements requiring a longer effective integration time are obtained by combining several frames. For each measurement condition, M signal frames and N background frames are therefore acquired and averaged separately. In the schematic representation shown in Figure 90, these two sets are indicated as *hot* and *dark*, respectively.

The signal and background contributions are first averaged independently pixel by pixel. The mean signal image is defined as

$$\langle I_{\mathrm{signal}}(x,y)\rangle = \frac{1}{M}\sum_{j=1}^{M} I_j^{\mathrm{signal}}(x,y), \tag{79}$$

while the mean background image is

$$\langle I_{\mathrm{background}}(x,y)\rangle = \frac{1}{N}\sum_{j=1}^{N} I_j^{\mathrm{background}}(x,y). \tag{80}$$

The background-subtracted EMMI image is then obtained as

$$I_{\mathrm{EMMI}}(x,y) = \langle I_{\mathrm{signal}}(x,y)\rangle - \langle I_{\mathrm{background}}(x,y)\rangle\,. \tag{81}$$

This subtraction isolates the optical contribution associated with the biased SiPM from the background signal. This intermediate image is indicated as *diff* in Figure 90. Averaging several frames before the subtraction increases the effective signal integration time and reduces frame-to-frame statistical fluctuations and isolated noise contributions.

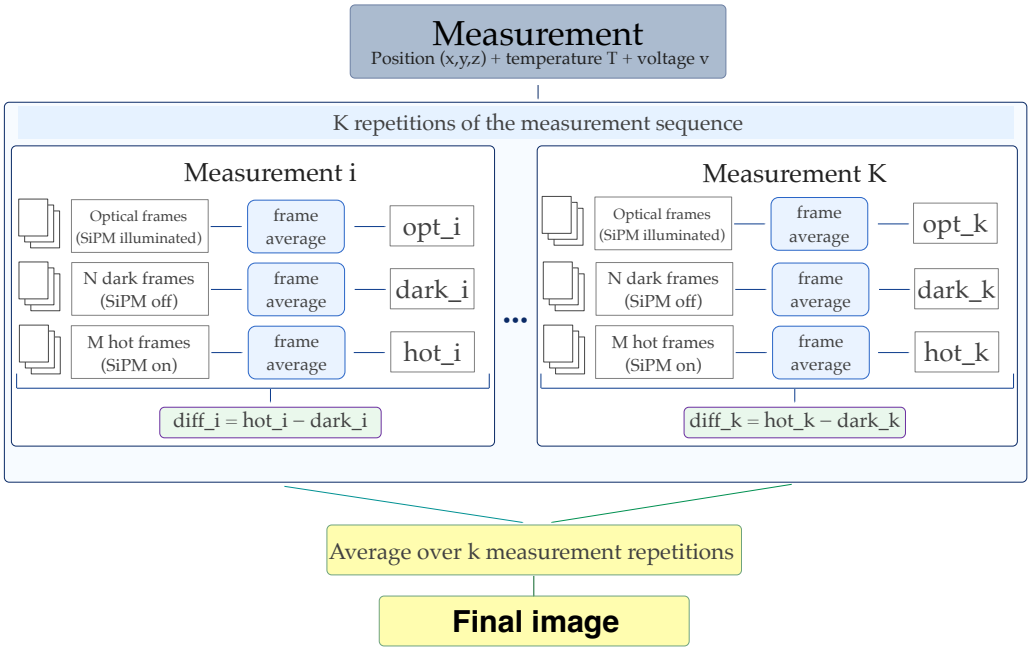


Figure 90: Schematic representing the data acquisition steps to obtain the final image used for the data analysis.

To further improve the statistical precision of the measurement, the complete procedure described above is repeated K times under the same experimental conditions. This results in K independent background-subtracted EMMI images of the same physical configuration. The final emission image is obtained by averaging the K measurements pixel by pixel. For each pixel, the mean intensity is therefore defined as

$$\langle I_{\mathrm{EMMI}}(x,y)\rangle = \frac{1}{K}\sum_{k=1}^{K} I_{\mathrm{EMMI},k}(x,y), \tag{82}$$

where $I_{\mathrm{EMMI},k}(x,y)$ is the intensity measured at pixel (x,y) in the k-th background-subtracted image.

The statistical uncertainty associated with the mean is obtained from the dispersion of the K independent measurements. The sample standard deviation is

$$s_I(x,y) = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left[I_{\mathrm{EMMI},k}(x,y) - \langle I_{\mathrm{EMMI}}(x,y)\rangle\right]^2}, \tag{83}$$

and the statistical uncertainty on the mean intensity is consequently

$$\sigma_{\langle I_{\mathrm{EMMI}}\rangle}(x,y) = \frac{s_I(x,y)}{\sqrt{K}} = \sqrt{\frac{1}{K(K-1)}\sum_{k=1}^{K}\left[I_{\mathrm{EMMI},k}(x,y) - \langle I_{\mathrm{EMMI}}(x,y)\rangle\right]^2}. \tag{84}$$

Applying this procedure independently to every pixel produces both the final averaged EMMI image and the corresponding statistical-uncertainty image. The values of N, M, and K were not kept identical throughout the entire measurement campaign, since the acquisition parameters were progressively optimized according to the requirements of the different measurements. Tab. 5 reports some exemplary values of N, M, and K. Nevertheless, since both the frame-level combination and the final combination of repeated measurements are are always referred to an exposure time of $5000\ \mathrm{ms}$, the measured luminosities can be consistently compared.

Measurement	N **frames background**	M **frames signal**	K **repetitions**
New device – whole-sensor measurement	20	20	20
New device – temperature scan	20	20	50
New device – overvoltage scan	20	20	50
After irradiation – whole-sensor measurement	20	20	20
After irradiation – temperature scan	20	20	50
After irradiation – overvoltage scan	20	20	50
Annealing – whole-sensor measurements	20	20	20
Annealing – central-region measurements	20	20	25

Table 5: Number of background frames (N), signal frames (M), and repeated background-subtracted measurements (K) used for the different EMMI acquisition configurations. The same acquisition parameters were maintained throughout the different annealing stages.

This two-level averaging procedure substantially increases the total integrated signal exposure while preserving a consistent normalization between measurements. As a consequence, the statistical significance of the weak optical emission is improved and the associated statistical uncertainties are reduced. At the same time, averaging independent acquisitions limits the influence of random fluctuations and occasional transient events, such as cosmic-ray interactions, providing a more stable and reproducible image for the subsequent quantitative EMMI analysis.

4.4 Image processing

Before performing the quantitative analysis of the emission images, obtained with the procedure described in Sec. 4.3, a preliminary processing is applied to reduce instrumental artifacts, introduced by the camera sensor and its readout electronics, and to eliminate any shifts or rotations caused by the repositioning the board between annealing stages. The following paragraphs describe these processing.

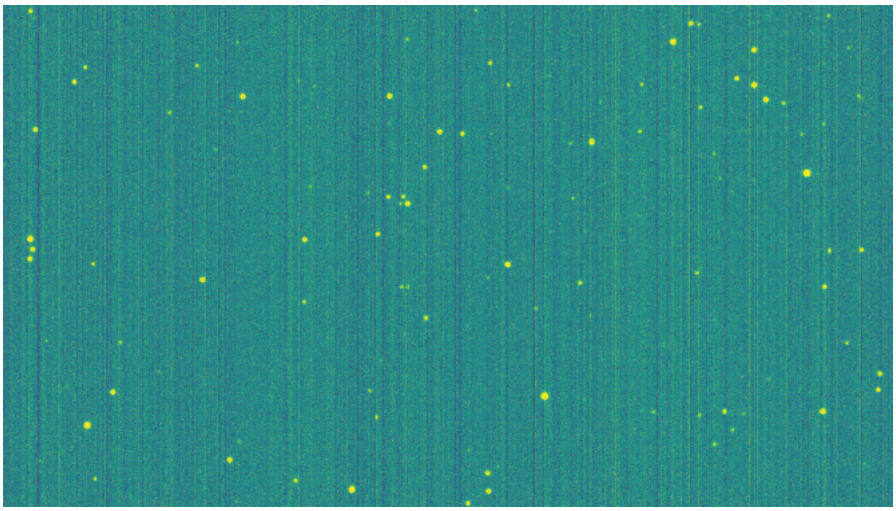


Figure 91: Image before clean-up.

4.4.1 Image clean-up

The clean-up procedure addresses two main effects: column-dependent offsets and isolated anomalous pixels. First, a column-bias correction is applied to compensate for differences in the baseline level associated with individual camera columns. Since this contribution affects the pixels belonging to the same column in a correlated way, it can generate artificial vertical structures extending across the image, as can be seen in Figure 91. The correction removes these column-dependent variations, producing a more uniform baseline while preserving localized structures associated with the optical emission from the SiPM, as can be seen in Figure 92.

A second correction targets isolated pixels whose intensity is anomalously different from that of their local surroundings. Pixels with an abnormally high response are identified as *hot pixels*, whereas pixels with an abnormally low response are classified as *cold pixels*. Their identification relies on their localized character: in contrast to a physical emission hotspot, which extends over several neighbouring pixels according to the optical response of the imaging system, a defective camera pixel appears as an isolated intensity anomaly. Once identified, these pixels are corrected using the information provided by the surrounding image region, preventing a single detector pixel from producing an artificial source or altering the measured intensity of a genuine emission hotspot. The final result is visible in Figure 93.

The preprocessing therefore acts specifically on structures that can be attributed to the camera rather than to the SiPM. This step is particularly important before the automatic source-identification procedure, since residual column structures or isolated anomalous pixels could otherwise be interpreted as physical emission features. The resulting image retains the spatial emission pattern of the device while providing a cleaner and more uniform input for the subsequent hotspot detection, matching, and luminosity analysis.

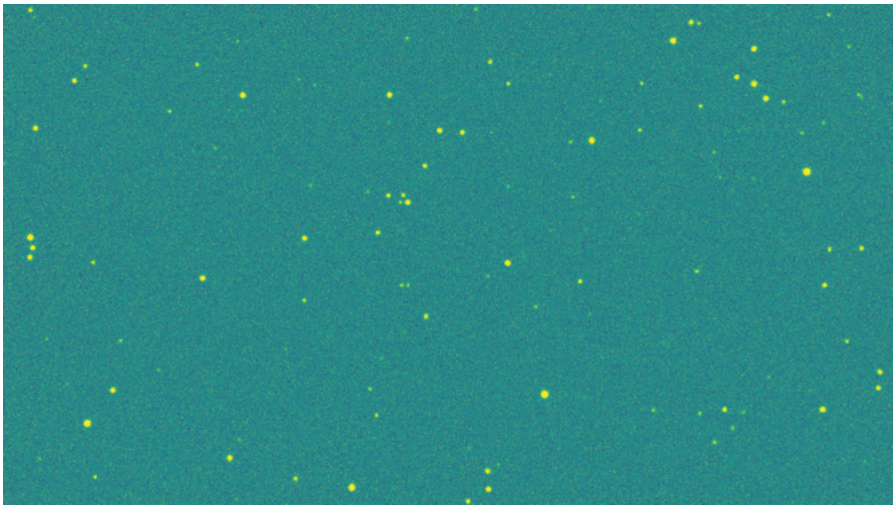


Figure 92: Image after removal of column dependent bias.

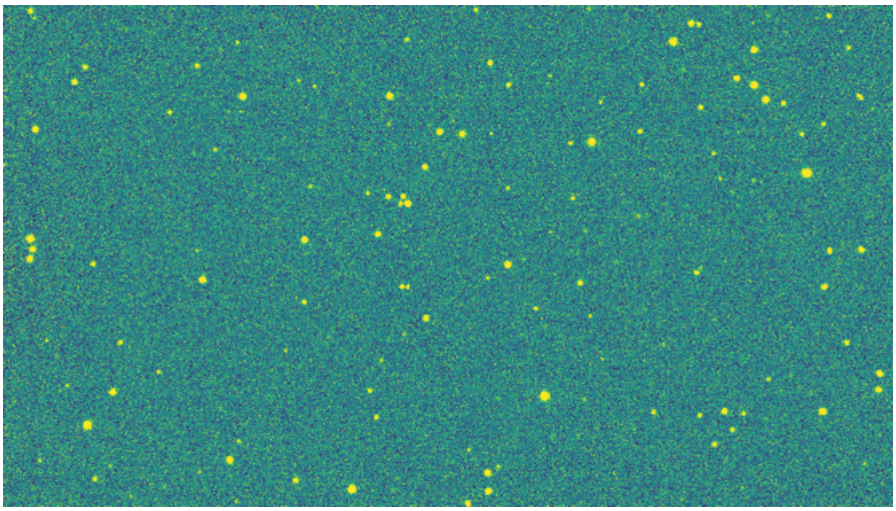


Figure 93: Image after removal of column dependent bias and hot/cold pixels.

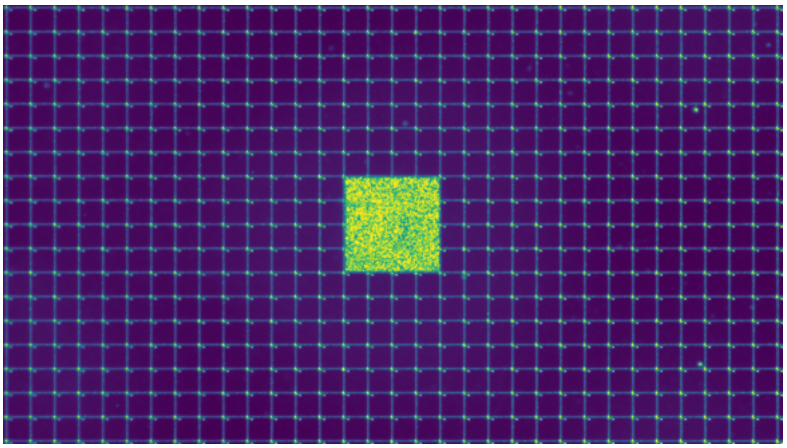
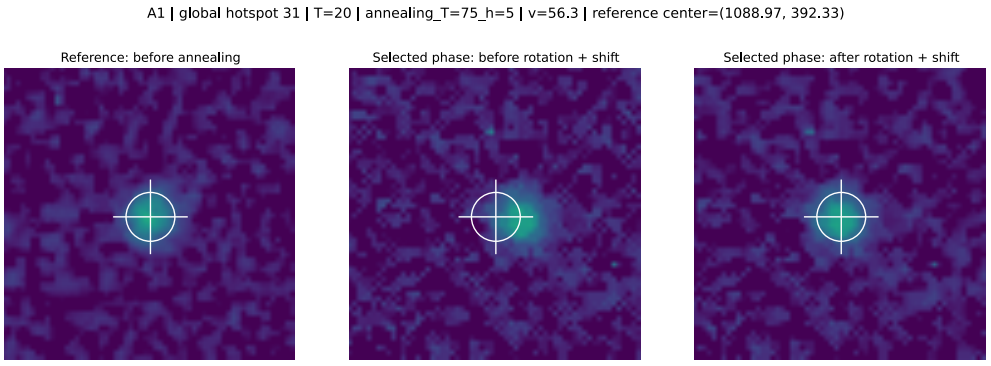


Figure 94: Optical image of central region of sensor A1, obtained with an external light illuminating the SiPM. Geometrical structures of the device, such as SPADs and the Through-Silicon Vias (TSVs) are visible.

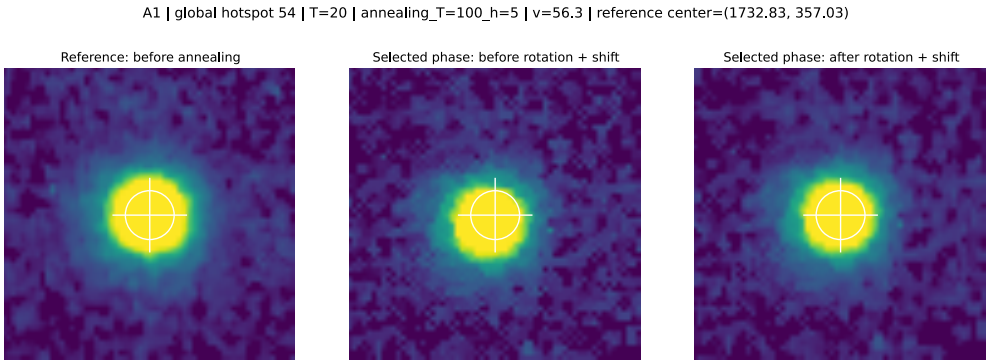
4.4.2 Image rotation and shift

A geometrical alignment procedure is applied to the images acquired at the different annealing stages in order to ensure that the same physical position on the SiPM corresponds to the almost the same image coordinates throughout the analysis. This correction is particularly important because, during the annealing procedure, the SiPM board must be removed from the EMMI apparatus and subsequently mounted again for the following measurement campaign. Repositioning the board with micrometric reproducibility is not experimentally achievable; therefore, small differences in both the angular orientation and the lateral position of the device are inevitably introduced between consecutive annealing phases. Even displacements of only a few pixels are significant for the present analysis, since, with the $4\times$ magnification, one camera pixel corresponds to approximately $1.26\ \mu\mathrm{m}$ on the SiPM plane. Thus, a displacement of 3–4 pixels corresponds to about 4–$5\ \mu\mathrm{m}$, which is non-negligible when tracking individual localized hotspots and compared with the nominal integration radius of 20 pixels, corresponding to approximately $25\ \mu\mathrm{m}$.

The alignment is performed using the optical image, obtained with an external light illuminating the SiPM. An exemplary picture is represented in Figure 94. The optical image contains permanent geometrical structures of the device, in particular the Through-Silicon Vias (TSVs) and the SPAD grid structure, whose positions are fixed with respect to the SiPM active area and are therefore independent of the annealing treatment and of the electrical operating conditions. These structures provide stable spatial references that can be identified in all measurement phases. In contrast, the emission pattern itself cannot be used as a reliable reference because the number and intensity of luminous hotspots can change as a consequence of irradiation and annealing. Furthermore, it would not be methodologically adequate to use the emission image to correct the image itself. The first correction accounts for differences in the angular orientation of the device. For each measurement phase, the rotation angle is determined



(a)



(b)

Figure 95: Example images of the rotation+shifting correction procedure. In the left images, the reference images from which the coordinates are calculated. In the middle images, evident geometrical displacement caused by the removal and repositioning of the SiPM board. In the right images, correction made by the rotation+shifting procedure.

from the corresponding optical image using the dedicated rotation-measurement procedure. The measured angle is then applied to all the images belonging to that phase, including the optical, background-subtracted emission, and the matrices containing the statistical uncertainty of the measured images. Since image rotation may introduce regions outside the original field of view, the procedure also keeps track of the pixels generated by the transformation and of the corresponding image boundaries.

After the rotational correction, the optical image acquired before annealing is defined as the common geometrical reference for the entire analysis. For every subsequent annealing stage, the already rotated optical image is compared with this reference to determine the residual translation along the horizontal and vertical image directions. The displacement is expressed through two components, corresponding to the shifts along the x and y axes. Once determined, exactly the same translation is applied to all the rotated images acquired during that annealing phase.

The combination of rotation and translation therefore brings the images from all annealing

stages into a common reference coordinate system defined by the pre-annealing measurement. The procedure corrects the geometrical displacement caused by repeated removal and repositioning of the SiPM board without relying on the position or intensity of the emission hotspots themselves. Consequently, differences observed in hotspot coordinates after alignment are not dominated by macroscopic changes in the position of the board inside the experimental apparatus. The effectiveness of this procedure can be seen in the image in Figure 95, using two example hotspots. Shown are the reference image from which the coordinates are calculated, the original image from the device following the annealing stage, and the image corrected using the rotation and shift procedure This spatial registration is essential for the subsequent hotspot-matching procedure. Once all images are expressed in the same reference system, hotspots detected at different annealing stages can be compared according to their spatial coordinates and associated with the same physical region of the SiPM. This considerably reduces the probability of assigning different global identifiers to the same defect because of a displacement of the device, or conversely of incorrectly associating physically distinct hotspots. The alignment therefore provides the geometrical basis required to study reliably the appearance, disappearance, and luminosity evolution of individual emission centres throughout the annealing sequence.

4.5 Hotspot characterization

4.5.1 Identification and localisation

The first step for the data analysis is to find the hotspots location inside the SiPM. In order to do so, an algorithm based on Photutils was developed, which is a Python library that provides commonly-used tools and key functionality for detecting and performing photometry of astronomical sources [58]. Considering the analogy with the dark background and luminous spots, it was possible to employ the library in this context.

The analysis is organized into a sequence of processing stages.

Background estimation and subtraction The slowly varying background was evaluated dividing the image into boxes of (50×50) pixels, and the background level within each box is evaluated using a median estimator. A (3×3) filtering operation is subsequently applied to smooth the background model and reduce discontinuities between adjacent regions. The estimated background is then subtracted from the original image: $I_{\text{processed}} = I_{\text{original}} - I_{\text{background}}$. The resulting image enhances localized luminous structures while suppressing large-scale variations in the illumination level.

Detection threshold The source-detection threshold is defined locally as three times the root-mean-square fluctuation of the estimated background: $T(x, y) = 3\sigma_{\text{background}}(x, y)$. Consequently, a region is considered potentially associated with a luminous source only when its intensity is sufficiently higher than the local background fluctuations. Since the background noise may vary across the image, the threshold is represented by a two-dimensional map rather than by a single constant value.

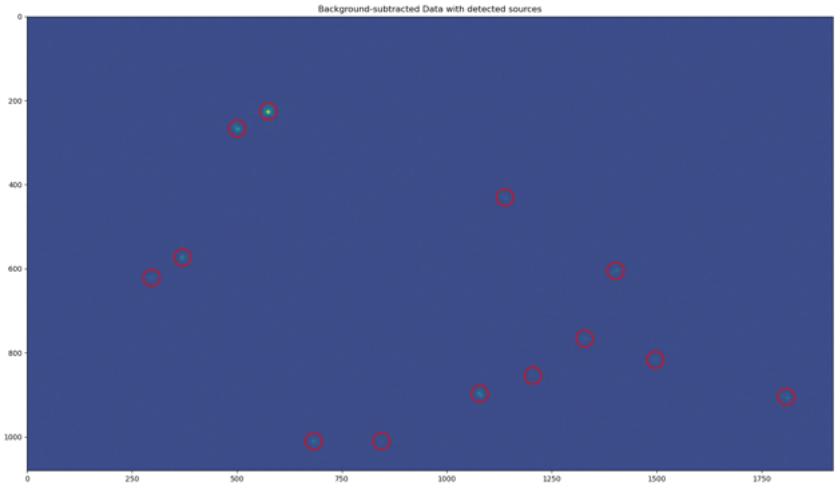


Figure 96: Example of application of the described code for individuating the defects coordinates on the detector; in the picture is presented a new SiPM (before irradiation) in the best visibility conditions (maximum temperature and overvoltage).

Source detection and deblending The identification of luminous sources is based on a segmentation procedure that searches for spatially connected regions whose intensity exceeds the detection threshold. To suppress isolated fluctuations and very small structures that are unlikely to correspond to genuine emission hotspots, a detected source is required to contain at least 30 connected pixels above threshold.

The systematic uncertainty associated with this requirement was evaluated by varying the minimum source size by $\pm 20\%$, corresponding to thresholds of 24 and 36 connected pixels. The complete hotspot-counting procedure was repeated for both alternative values, and the maximum variation in the number of detected hotspots with respect to the nominal 30-pixel threshold was assigned as the corresponding systematic uncertainty. A more detailed description of this methodology is given in Sec. 4.6.

Particular care is taken in regions where two or more luminous sources are close to one another or partially overlap. In these cases, the intensity distribution within the detected region is examined at several intermediate levels in order to determine whether distinct local maxima are present. This deblending procedure allows nearby emission centres to be separated and treated as independent sources when their relative intensity contribution is sufficiently significant. In the present analysis, 32 intermediate intensity levels are used, together with a low contrast requirement so that relatively faint hotspots can still be distinguished in the vicinity of brighter ones.

The outcome of this procedure is a segmentation map in which each identified hotspot is associated with a distinct region of pixels, while pixels not assigned to any source are classified as background. This map therefore provides the basis for determining the position and spatial extent of each detected emission centre in the subsequent analysis.

Figure 97: Example of inspection of an individual hotspot, from sensor A1 before annealing at $T = 21°C$ and $v_{over} = 5\ V$. The red circle represents the integration radius of 20 pixels.

Measurement of source properties Once the luminous regions have been identified, their main geometrical properties are determined from the corresponding segmented pixels. For each source, the centroid is calculated as the intensity-weighted centre of the segmented region according to

$$x_c = \frac{\sum_i x_i I_i}{\sum_i I_i}, \qquad y_c = \frac{\sum_i y_i I_i}{\sum_i I_i}, \tag{85}$$

where I_i is the intensity of the i-th pixel belonging to the source and (x_i, y_i) are its coordinates. The resulting centroid (x_c, y_c) is used to define the hotspot position in the image. The spatial extent of each detected source is instead estimated from the area of its segmented region, obtained from the number of pixels assigned to that source. In order to express this area through a single characteristic length, an equivalent circular radius is defined as

$$r_{\mathrm{eq}} = \sqrt{\frac{A_{\mathrm{segment}}}{\pi}} \tag{86}$$

where A_{segment} is the area of the segmented source in pixels squared. The quantity r_{eq} corresponds to the radius of a circle having the same area as the detected region. It should therefore not be interpreted as implying a circular shape of the source, but rather as a convenient scalar measure of its characteristic spatial size. Accordingly, more extended segmented regions are associated with larger equivalent radii, while more compact sources are assigned smaller values.

Inspection of an individual source It is also possible to manually inspect an individual source. Focusing on a hotspot is a useful way to examine the properties of the light source, and it allows us to observe how it changes during the annealing stages. An example of single hotspot inspection is presented in Figure 97.

4.5.2 Luminosity calculation

After finding the hotspot's coordinates, the other main characteristic of the hotspot is its brightness, or luminosity. For each selected hotspot, the measured luminosity L_{hotspot} is defined as the integrated pixel content inside a circular region of fixed radius centred on the source centroid. The choice of the radius value, together with the computation of its systematic uncertainty, is discussed in Sec. 4.6.

Luminosity calculation For each hotspot, the corresponding luminosity, expressed in arbitrary camera units, is obtained by integrating the background-subtracted EMMI intensity over a circular region centred on the hotspot coordinates. The integration region is defined by

$$(x - x_0)^2 + (y - y_0)^2 \leq R^2, \tag{87}$$

where (x_0, y_0) are the hotspot coordinates and R is the integration radius.

Using the background-subtracted intensity map $\langle I_{\mathrm{EMMI}}(x, y)\rangle$ defined above, the hotspot luminosity is therefore calculated as

$$L = \sum_{\substack{x,y \\ (x-x_0)^2+(y-y_0)^2 \leq R^2}} \langle I_{\mathrm{EMMI}}(x, y)\rangle \, . \tag{88}$$

The sum includes all pixels whose centres lie within the circular integration region.

Statistical uncertainty calculation The statistical uncertainty associated to a luminosity measurement is calculated using the errors assigned to the individual histogram bins $\sigma_{\langle I_{\mathrm{EMMI}}\rangle}(x, y)$. We assume the pixels to be statistically independent, therefore the uncertainty is calculated by summing the bin errors in quadrature $S = \sqrt{\sum_{i=1}^{N} \sigma_{\langle I_{\mathrm{EMMI}}\rangle}(x, y)^2}$.

4.5.3 Hotspot identification across annealing phases

Since hotspot identification is performed independently for each annealing phase, the list of detected emission centres obtained from one measurement cannot be directly compared with that obtained from another. This is particularly important in the present study because the hotspot population itself is expected to evolve: some emission centres may disappear during annealing, others may appear only after a given treatment, and some may temporarily fall below the detection threshold and become visible again at a later stage. A common catalogue is therefore required in order to follow the same physical emission centres throughout the complete annealing sequence.

The purpose of this catalogue is to contain every spatially distinct hotspot observed at least once during the measurements and to preserve its identity across the different annealing phases. Each hotspot included in the catalogue is assigned a unique *global ID*, which is used throughout the rest of this thesis to identify the same emission centre independently of the phase in which it is detected. The coordinates obtained independently from the source-identification procedure at each annealing stage are referred to as *phase-specific coordinates*, whereas the corresponding reference position in the common catalogue is referred to as the *global coordinate*.

The post-irradiation measurement performed before annealing is used to initialise the catalogue. Each hotspot detected in this image is assigned a global ID and its position, expressed in the common coordinate system obtained after the geometrical alignment described in Sec. 4.4, defines its initial global coordinate. The subsequent annealing phases are then considered in chronological order. Whenever a hotspot is detected close to the position of a hotspot already present in the catalogue, it is interpreted as the same physical emission centre and retains the same global ID. In the nominal analysis, two detections are considered compatible when the distance between their centroids is smaller than 10 pixels.

This correspondence is required to be unique within each annealing phase, so that two different detected hotspots cannot be associated with the same global ID. Conversely, when a detected hotspot cannot be associated with any previously catalogued emission centre, it is considered a newly observed hotspot and a new global ID is introduced. In this way, the catalogue progressively contains all distinct hotspots observed during the complete sequence, without incorrectly counting the repeated observation of the same hotspot as the appearance of a new one.

The reliability of the spatial association was checked by varying the matching radius between 5 and 25 pixels. The global-ID assignment remained unchanged throughout this range, while multiple possible associations appeared only for radii larger than approximately 20 pixels, where distinct nearby hotspots can fall within the same matching region. Moreover, even with the 10-pixel radius requirements, the distance between a phase-specific coordinate and the corresponding global coordinate never exceeded 4 pixels.

Once the global catalogue has been constructed, the detection history of each hotspot can be followed throughout the full annealing sequence. For every phase, each global ID can therefore be classified as detected or absent, allowing emission centres that remain visible, appear, disappear, or reappear to be distinguished. The same correspondence is also used to follow the luminosity of individual hotspots as annealing progresses. The global-ID catalogue thus provides the common reference needed to study the evolution of the hotspot population and of individual emission centres independently of the source-identification performed in each single measurement.

4.6 Evaluation of systematic uncertainties

In this section, the procedure adopted to evaluate the systematic uncertainties associated with the analysis is described. In particular, the choice of the integration radius used to calculate the hotspot luminosity can affect the measured value. Therefore, a dedicated study was performed to quantify the dependence of the measured luminosity on this analysis parameter and to assign the corresponding systematic uncertainty.

4.6.1 Study on isolation cut

In order to perform a more reliable quantitative study of the luminosity of individual emission hotspots, a subset of *isolated* hotspots was selected. The purpose of this requirement is to reduce the possible contamination of the measured luminosity of one hotspot by the optical emission associated with neighbouring hotspots. Considering integration radii between 5 and 25 pixels, an emission site is classified as isolated when the distance between its centroid and that of any other detected hotspot is at least 60 pixels.

This condition is particularly relevant for the luminosity measurement, which is obtained by integrating the image intensity within a finite region around each hotspot. As long as the separation between two hotspots is sufficiently large compared with the integration radius, the corresponding integration regions do not significantly overlap and the measured luminosity can be associated more reliably with a single emission site. Conversely, when two hotspots are too close, part of the light originating from a neighbouring source may fall within the integration

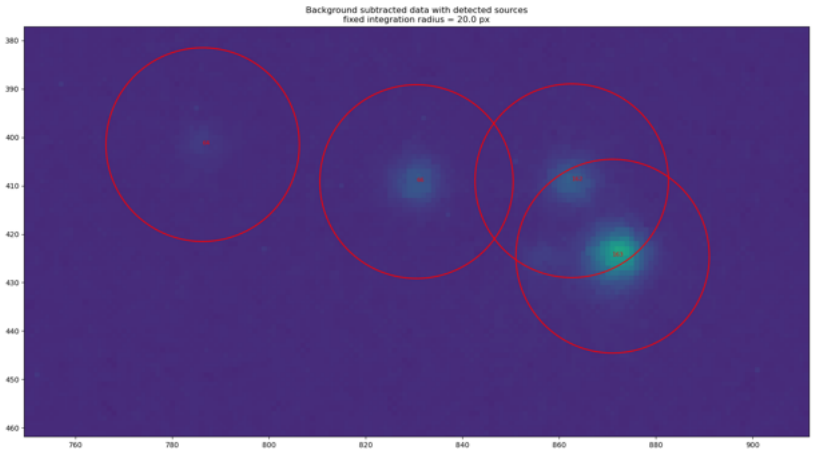


Figure 98: Possible critical situation of close hotspots. Increasing the integration radius of all the defects could lead to an overestimation of the luminosity, due to integration over an area containing a different hotspot. Hotspot like 66-162-163 (in the local ID coordinates) are excluded by the isolation cut.

region and lead to an overestimation of the luminosity attributed to the hotspot under study. An example of such a critical configuration is shown in Figure 98.

Since the isolation requirement selects only a subset of the detected hotspots, it is necessary to verify that this subset does not correspond to a particular class of emission sites and therefore remains representative of the full hotspot population. For this purpose, the luminosity distribution of all detected hotspots was compared with that of the subset satisfying the isolation requirement. The two distributions are shown before normalization in Figure 99 and after normalization in Figure 100.

The normalized luminosity distributions show no significant difference between the complete hotspot population and the isolated subset. In particular, the ratio of the two normalized distributions is compatible with unity within the uncertainties. This indicates that requiring spatial isolation does not preferentially select hotspots with a particular luminosity range. The isolated hotspots can therefore be used for the more detailed studies presented in the following sections while reducing the probability of luminosity contamination from neighbouring emission sites.

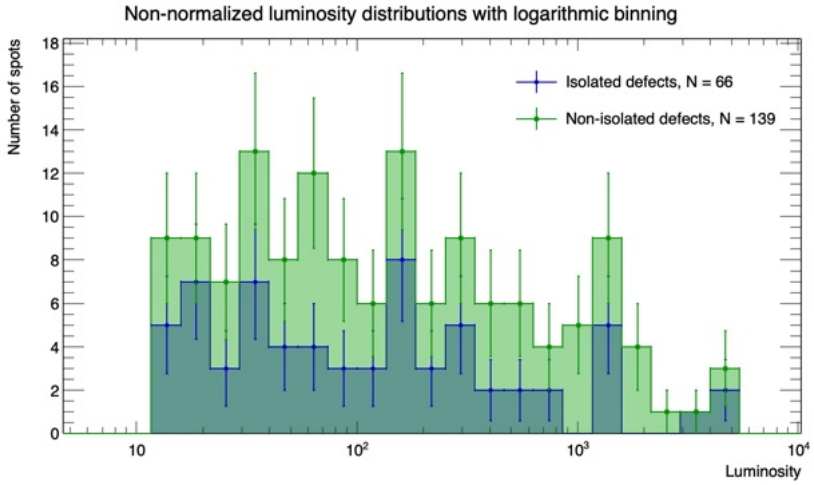


Figure 99: Plot of the luminosity distribution (non normalized) with logarithmic binning. In green the whole sample, in blue the isolated defects.

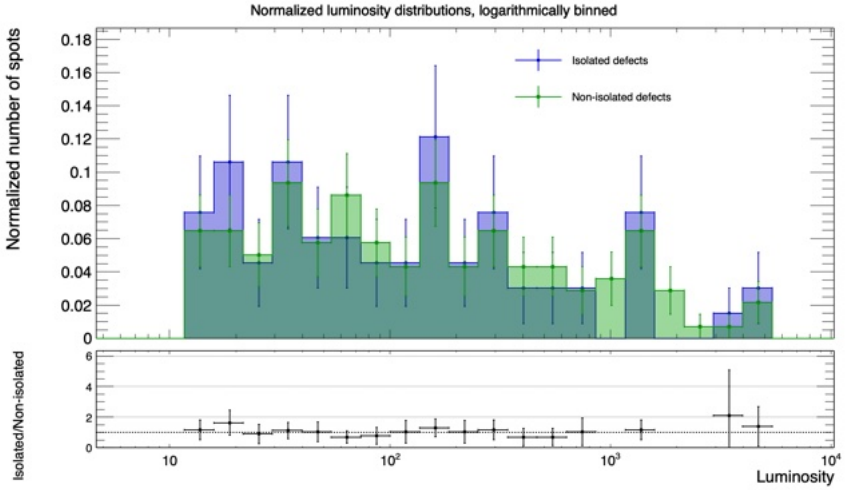


Figure 100: Plot of the normalized luminosity distribution with logarithmic binning. In green the whole sample, in blue the isolated defects. In the lower pad, the ratio between isolated and non-isolated defects.

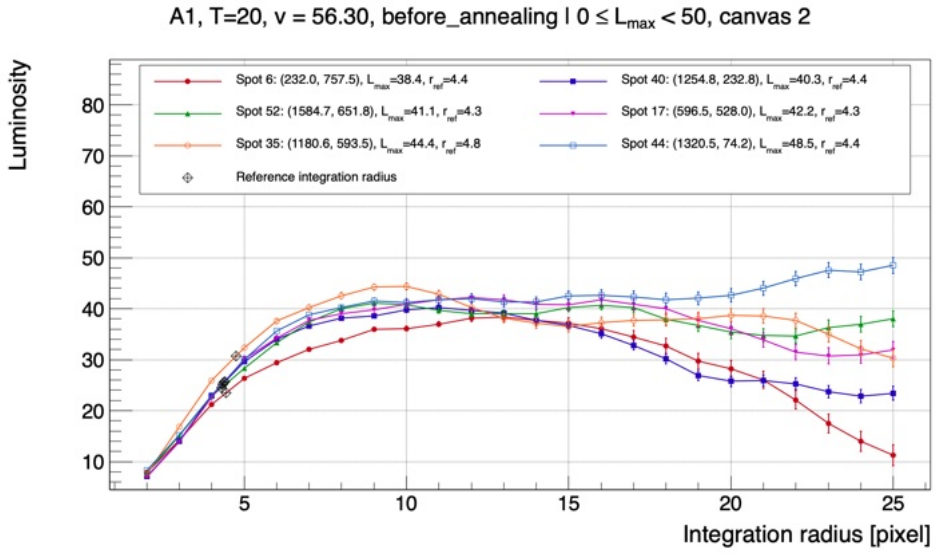


Figure 101: Luminosity of each hotspot recalculated for different values of the integration radius, while keeping its position fixed. Hotspot with maximum luminosity between 0 and 50 a.u.

4.6.2 Study on radius

We used isolated defects to perform a detailed study and evaluate the influence of the choice of the integration radius. The luminosity of each hotspot was recalculated for different values of the integration radius while keeping its position fixed. Also the reference radius value, computed with the algorithm described in Sec. 4.5.1, was evaluated in this study. As shown in Figures 101-103, the measured luminosity generally increases as the integration radius is enlarged, since a progressively larger fraction of the emitted signal is included. For sufficiently large radii the luminosity tends to reach a plateau, indicating that most of the signal associated with the hotspot is contained within a circular integration region with radius approximately $15 - 20$ pixels, as can be seen in Figures 101-103. Some faint hotspots with luminosity lower than 50 a.u. and a few cases of hotspots with luminosity lower than 100 a.u. do not appear to follow this trend. In fact, their luminosity appears to decrease as the integration radius increases. One possible explanation is that, as these hotspots are not very bright, they are more affected by fluctuations in the background, which can even take on negative values. This explanation is supported by the increase in statistical uncertainties for high values of the radius.

As it can be seen from figures Figures 101-103, the so-called "reference integration radius" value, defined in Eq. 4.5.1 discussed in Sec. 4.5, always sits on the rising slope of the luminosity versus radius lines, well before the plateau. Using this radius would therefore lead to an underestimation of the luminosity. With the considerations described above, a sufficiently large integration radius of $R = 20$ pixels was chosen as the nominal integration radius for the analysis. The next section will nonetheless discuss the associated systematic checks and uncertainties associated with this choice.

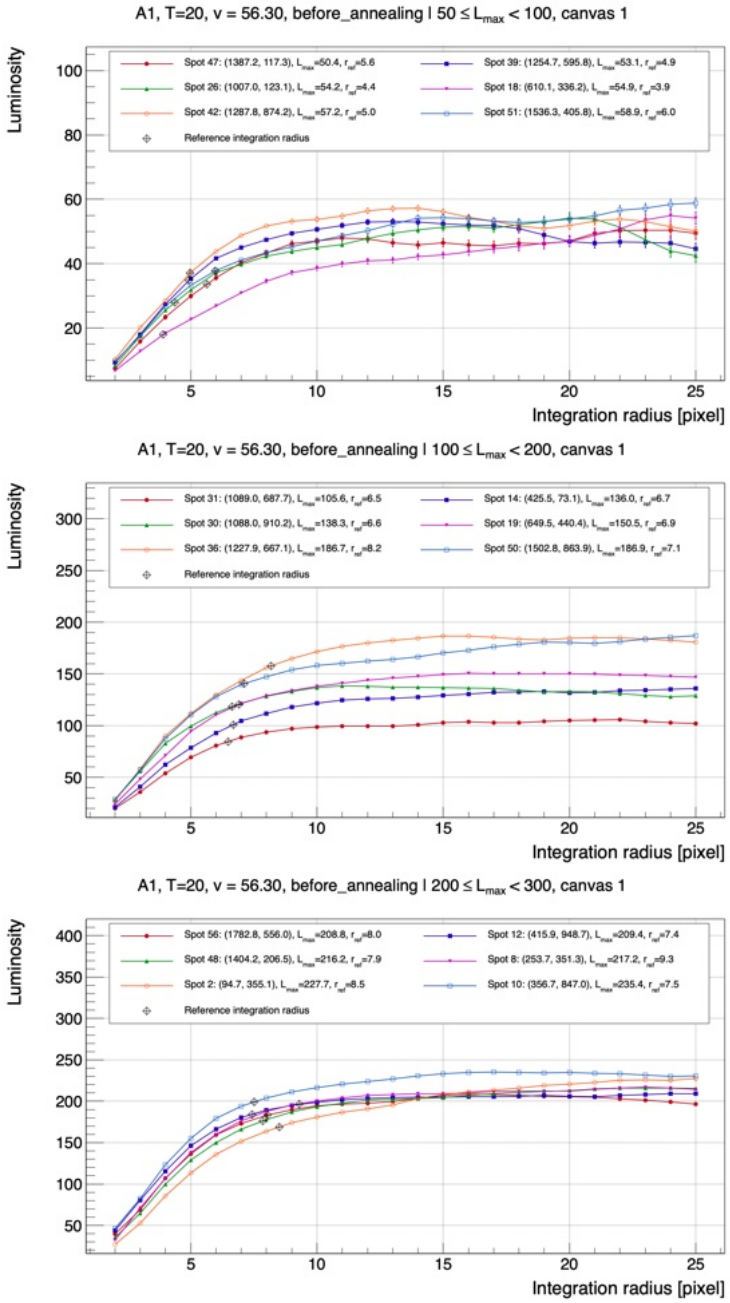


Figure 102: Luminosity of each hotspot recalculated for different values of the integration radius, while keeping its position fixed. (Top) Hotspot with maximum luminosity between 50 and 100 a.u., (Middle) Hotspot with maximum luminosity between 100 and 200 a.u., (Bottom) Hotspot with maximum luminosity between 200 and 300 a.u.

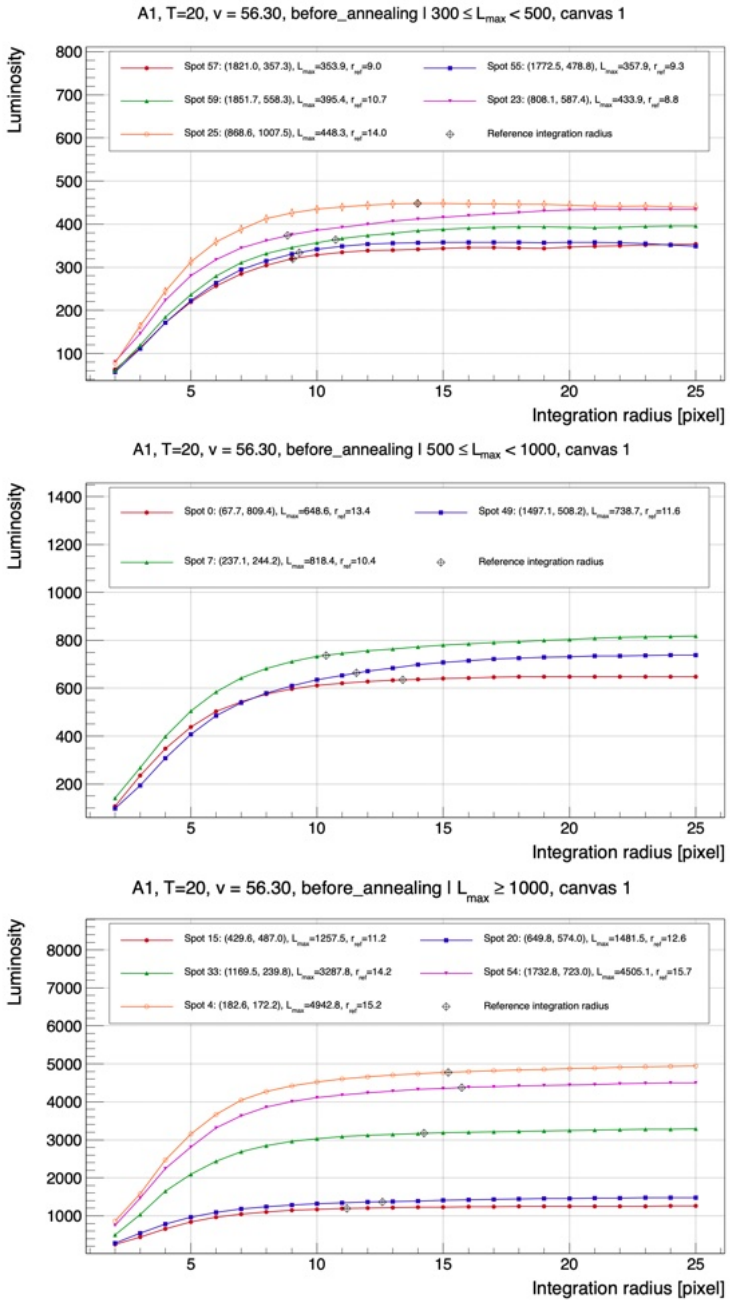
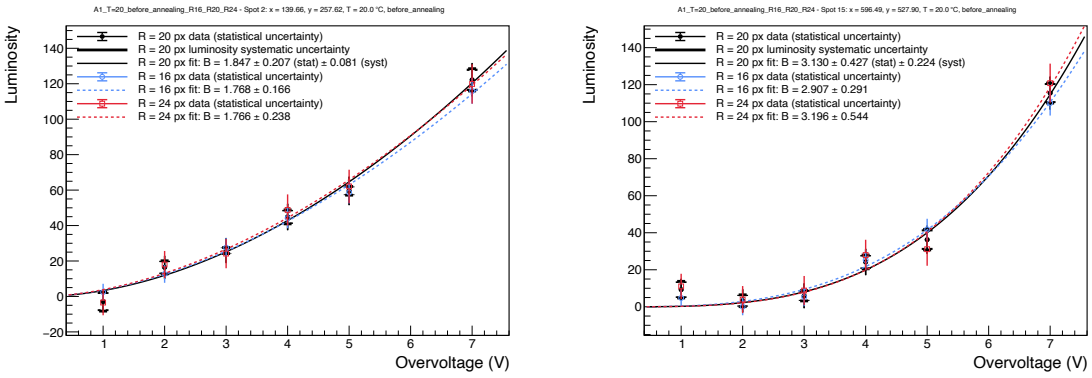


Figure 103: Luminosity of each hotspot recalculated for different values of the integration radius, while keeping its position fixed. (Top) Hotspot with maximum luminosity between 300 and 500 a.u., (Middle) Hotspot with maximum luminosity between 500 and 1000 a.u., (Bottom) Hotspot with maximum luminosity greater than 1000 a.u.



(a) (b)

Figure 104: Plots showing the luminosities at different overvoltages, for two examples hotspots, measured with the three integration radii as L_{16}, L_{20}, and L_{24}. The systematic uncertainty associated to L_{20} is showed.

4.6.3 Calculation of systematic uncertainties

The systematic uncertainty contribution is evaluated to quantify the dependence of the hotspot luminosity measurements on the choice of the integration radius. A radius that is too small may exclude part of the emitted signal, whereas a larger radius may include an increased contribution from the surrounding background. The sensitivity of the measured luminosity to this choice is therefore used to estimate the corresponding systematic uncertainty. The results obtained with an integration radius of $R = 20$ pixels is taken as the nominal measurement, while two additional analyses are performed considering a 20% change on the fixed radius, practically $R = 16$ and $R = 24$ pixels. For each hotspot, annealing phase, temperature, and bias voltage, the luminosity obtained in the three analyses is compared. The correspondence between the measurements is established using the global hotspot identifier together with the annealing phase and the operating conditions, ensuring that the luminosities being compared refer to the same physical hotspot under exactly the same experimental conditions. Denoting the luminosities measured with the three integration radii as L_{16}, L_{20}, and L_{24}, two absolute variations with respect to the nominal result are calculated:

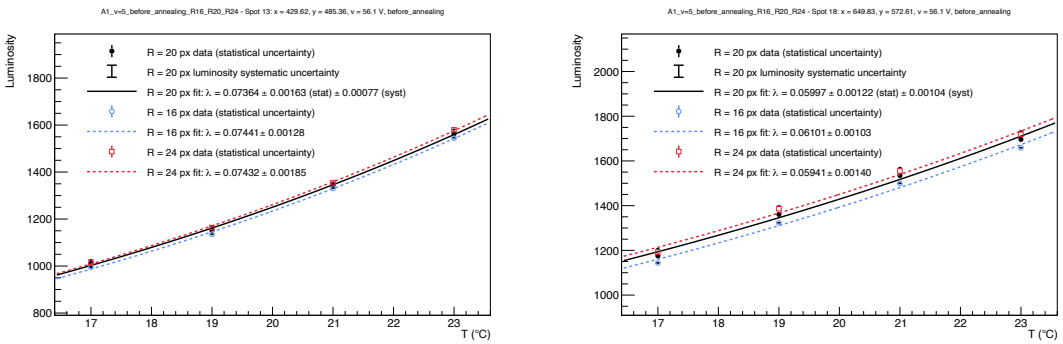
$$\Delta L_{+} = |L_{24} - L_{20}| \,,$$

and

$$\Delta L_{-} = |L_{16} - L_{20}| \,.$$

The systematic uncertainty assigned to the luminosity measurement is then defined as

$$\sigma(L_{\text{syst}}) = \max\left(\Delta L_{+}, \Delta L_{-}\right) .$$



(a) (b)

Figure 105: Plots showing the luminosities at different temperatures, for two examples hotspots, measured with the three integration radii as L_{16}, L_{20}, and L_{24}. The systematic uncertainty associated to L_{20} is showed.

Thus, the largest variation observed when increasing or decreasing the integration radius with respect to its nominal value is taken as the systematic uncertainty of the nominal $R = 20$ luminosity. Each hotspot measurement is therefore characterized by both a statistical uncertainty, originating from the image and luminosity-extraction procedure, and a systematic uncertainty, $\sigma(L_{\text{syst}})$, describing its sensitivity to the choice of integration region. On average, this systematic contribution is approximately 10%. Examples of its representation are shown in Figures 104 and 105.

The effect of the integration-radius choice is also propagated through the subsequent stages of the analysis discussed in Chapter 5. To do so, the complete analysis is repeated using $R = 16$ and $R = 24$ pixels in addition to the nominal $R = 20$ pixels, so that each derived quantity is independently recalculated for the three integration radii.

If a generic quantity extracted from the analysis is denoted by P, the corresponding values are P_{16}, P_{20}, and P_{24}, with P_{20} representing the nominal result. The deviations are defined as

$$\Delta P_+ = |P_{24} - P_{20}|\,, \qquad \Delta P_- = |P_{16} - P_{20}|\,,$$

and the associated systematic uncertainty is assigned as

$$\sigma(P_{\text{syst}}) = \max\left(\Delta P_+, \Delta P_-\right).$$

This procedure evaluates the effect of the integration-radius choice directly on the final quantity of interest and provides a consistent estimate of the corresponding systematic uncertainty.

5 Results and Discussion

This chapter presents the experimental results obtained from the EMMI characterization of the investigated SiPMs, with particular emphasis on the study of damages induced by irradiation and on their subsequent evolution during thermal annealing. Section 5.1 introduces the reference measurements performed on non-irradiated device, including the initial hotspot population and the distribution of the measured optical intensities. The dependence of the hotspot emission intensity on temperature and overvoltage is then investigated in Secs. 5.2 and 5.3, respectively. Section 5.4 presents the direct comparison before and after irradiation, considering the number and spatial distribution of the emission hotspots and their intensity distribution. The evolution of the irradiated devices throughout the annealing sequence is discussed in Sec. 5.5, combining whole-device hotspot counting, individual hotspot tracking in the central region, the evolution of the luminosity distributions and a description of the individual hotspots behaviours. Finally, Sec. 5.6 summarizes the main observations and discusses their physical interpretation within a unified framework.

In order to evaluate these results, it is important to bear in mind that the counting measurements were taken across the entire sensor ($3 \times 3\ mm^2$), whilst the measurements of luminosity as a function of temperature and overvoltage were taken only in the central region of the sensor ($2.4 \times 1.3\ mm^2$), visible in Figure 106, due to the long data acquisition times. The convention used to refer to the sensors is:

- A1 and A2 for the two sensors S13360-3050VS;
- B1 and B2 for the two sensors S13360-3075VS.

5.1 Reference measurements on non-irradiated SiPMs

Initial measurements were performed on the new devices before irradiation, in order to count and evaluate damages already present in new sensors. The discussion will cover the counting and spatial distribution of hotspots due defects already present in the material, their emission intensity (also referred to as *luminosity*, and their dependence on temperature and overvoltage.

5.1.1 Hotspots counting and distribution

The initial measurements were carried out by considering the entire sensor ($3 \times 3\ mm^2$) and counting the detected hotspots with the algorithm discussed in Sec. 4.5. The images were taken with an exposure time of $5000\ ms$ per 20 frames, 20 repetitions, keeping the overvoltage fixed at $5\ \mathrm{V}$ and the temperature fixed at $20\ ^\circ C$. Figure 107 left show a picture of sensor A1 with the superposition of the emission image and the optical image; Figure 107 right shows the emission image with the detected hotspots circled in red.

The hotspots seams to be distributed uniformly across the sensor. Tab. 6 reports the quantitative measurements obtained from the analysis. The mean number of hotspots per mm^2 are computed excluding the insensitive area of the sensor (meaning the TSV, $area = 0.2\ mm \times 0.2\ mm = 0.04\ mm^2$, from the whole SiPM area ($3\ mm \times 3\ mm = 9\ mm^2$).

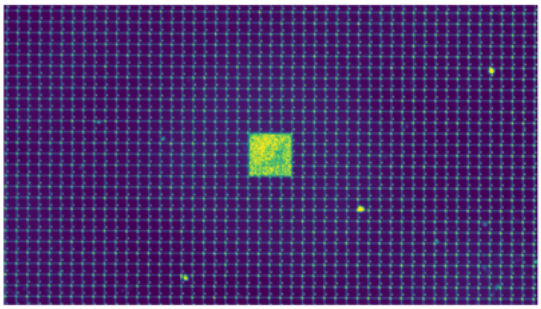


Figure 106: Example picture of the central region (area = $2.4 \times 1.3\ mm^2$) of the investigated detectors.

Table 6: Number and spatial density of hotspots measured in the new device.

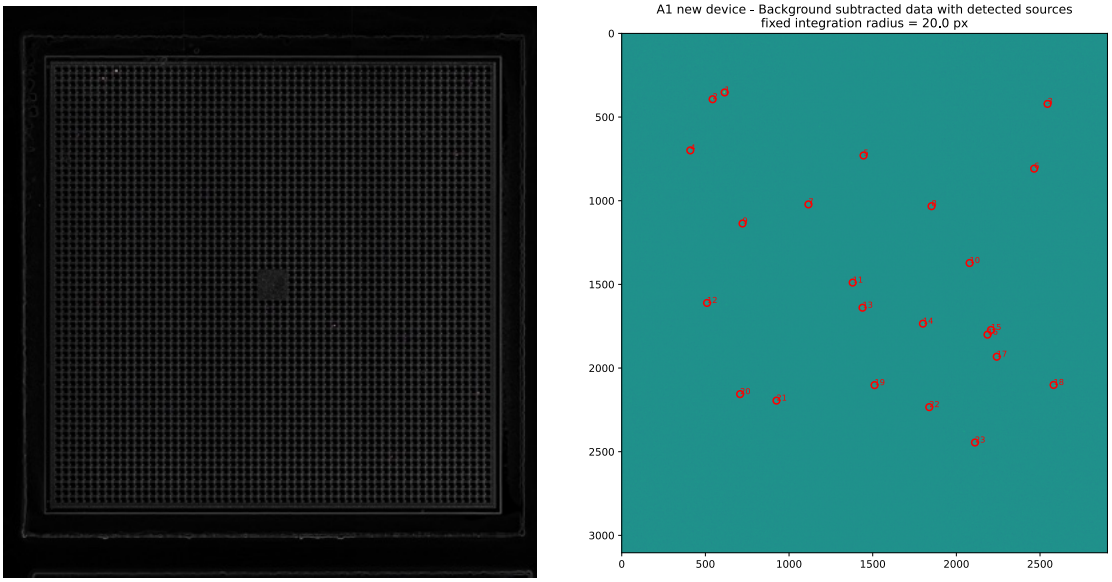
Region	Area	N. hotspots	Mean number of hotspots/mm^2
Whole sensor	$3.0 \times 3.0\ \mathrm{mm}^2$	23 ± 2	2.56 ± 0.22
Central region	$2.4 \times 1.3\ \mathrm{mm}^2$	11 ± 1	3.57 ± 0.32

5.1.2 Hotspots intensity

The detected emission hotspots do not all contribute the same optical intensity to the EMMI image. In order to define a reference distribution that can later be compared directly with the irradiated and annealed devices, only measurements acquired in the central region of the detector ($2.4 \times 1.3\ \mathrm{mm}^2$), at $T = 20°$C and $V_{over} = 7\ V$, were considered. The resulting intensity distribution is shown in Figure 108.

For the selected sample, 11 emission hotspots are detected. As the data collection covers a smaller area, this analysis is based on a smaller number of hotspots; hence the discrepancy in numbers compared with the previous paragraph. Their measured optical intensities span more than one order of magnitude, from approximately 22.6 to 478.6 a.u., indicating that the light-emitting sites already present in the non-irradiated device are characterized by markedly different levels of avalanche-associated optical emission. The distribution is strongly asymmetric: most hotspots populate the low-intensity region, while only a small number of sites extend toward substantially larger values. The median intensity is approximately 56.4, whereas the mean is about 108.1, the higher value of the latter being driven by the bright tail of the distribution. In particular, 9 of the 11 detected hotspots have an intensity below approximately 120, while two brighter sites reach values of about 204 and 479.

This broad and positively skewed distribution shows that even before irradiation the detected light-emitting sites form a heterogeneous population in terms of measured optical intensity. As discussed in Secs. 5.2 and 5.3, this quantity should not be interpreted as an intrinsic luminosity of a physical defect. The camera records the optical emission accumulated from the avalanches occurring at each location during the exposure, so differences in intensity can re-



(a) (b)

Figure 107: Counting measurements carried out on sensor A1 (type S13360-3050VS), considering the whole sensor area $3 \times 3 mm^2$. (a) Superposition of the emission image and the optical image. (b) Emission image with the detected hotspots (23) circled in red.

flect differences in the local avalanche rate, in the secondary photon yield associated with the avalanches, or in both. The distribution shown in Figure 108 therefore provides a reference baseline for evaluating how irradiation modifies both the population and the relative intensity of the detected emission hotspots.

5.1.3 Dependence on temperature

A second set of measurements, only on sensor A1, was performed keeping the overvoltage fixed, once with $V_{over} = 7\ V$ and once with $V_{over} = 9\ V$, whilst changing the temperature. The images were taken with an exposure time of $5000\, ms$ per 20 frames, 50 repetitions. Considering the duration of the data taking, this measurements were performed only on the central part of the sensor.

For both investigated operating conditions, $V_{\mathrm{over}} = 7$ and 9 V, the measured optical intensity (*luminosity*) increases with temperature, as shown in Figure 109. For each hotspot, the dependence was parametrized empirically as $L_{\mathrm{spot}}(T) = A\, e^{\lambda T}$.

The extracted λ values were subsequently compared for each operating condition and fitted with a constant, as shown in Figure 110. The constant fits were performed using the statistical uncertainties of the individual λ values. At $V_{\mathrm{over}} = 7$ V, the fit gives

$$\lambda = (0.061 \pm 0.003)\ ^{\circ}\mathrm{C}^{-1},$$

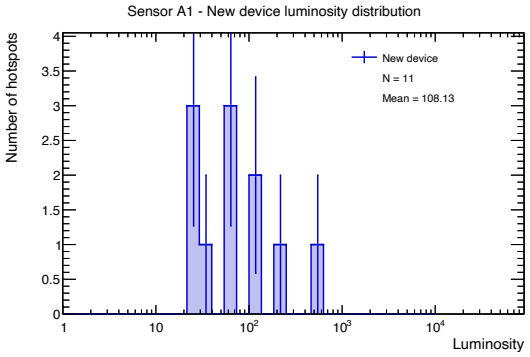
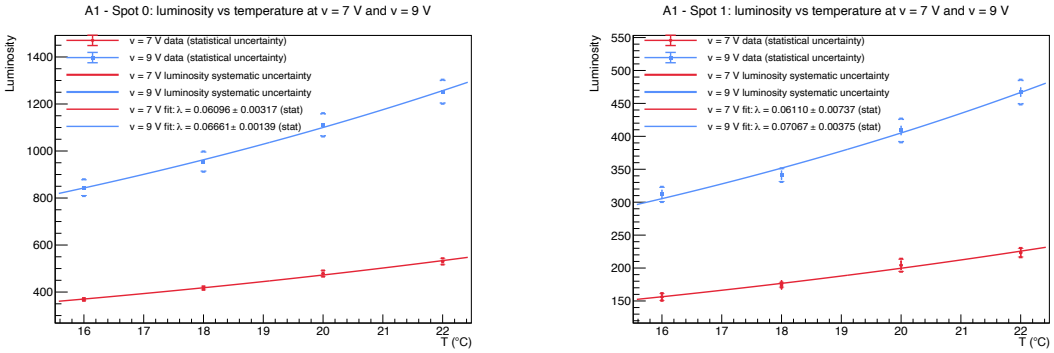


Figure 108: Distribution of the measured optical intensity of the emission hotspots detected in the non-irradiated A1 sensor at $T = 20°$C. The distribution is dominated by relatively low-intensity sites and presents a tail toward larger values.



(a) (b)

Figure 109: The luminosity as a function of temperature keeping the SiPM overvoltage fixed. Two example hotspots in the images. Both working conditions are represented: $V_{over} = 7\,V$ in red and $V_{over} = 9\,V$ in blue.

with $\chi^2/\mathrm{ndf} = 7.17/10 = 0.72$. At $V_{\mathrm{over}} = 9$ V, the fitted value is

$$\lambda = (0.065 \pm 0.001)\ {}^{\circ}\mathrm{C}^{-1},$$

with $\chi^2/\mathrm{ndf} = 15.92/11 = 1.45$. The goodness of the two fits therefore indicates that, within the observed hotspot-to-hotspot fluctuations, the temperature coefficients are compatible with a common characteristic value at each overvoltage.

From the fitted coefficients, the temperature interval corresponding to a factor-two variation of the measured emission,

$$\Delta T_2 = \frac{\ln 2}{\lambda},$$

is found to be

$$\Delta T_2 = (11.40 \pm 0.51)^{\circ}\mathrm{C}$$

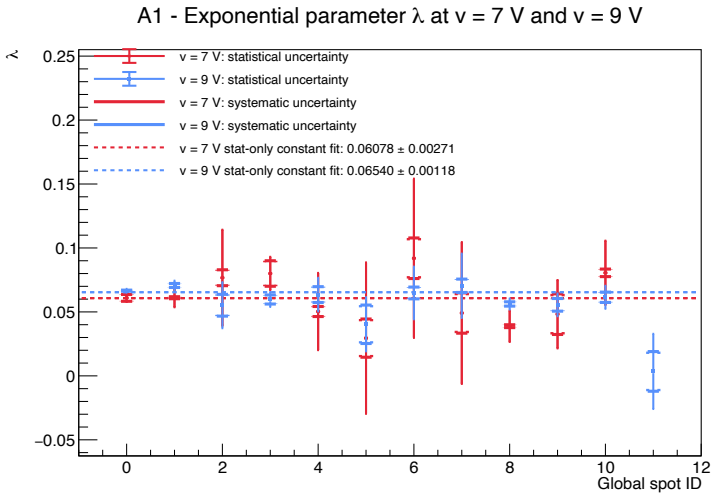


Figure 110: The extracted λ values, for both overvoltages, fitted with a constant for each operating condition. $V_{over} = 7\ V$ in red and $V_{over} = 9\ V$ in blue.

at $V_{\mathrm{over}} = 7\ \mathrm{V}$, and

$$\Delta T_2 = (10.60 \pm 0.19)^{\circ}\mathrm{C}$$

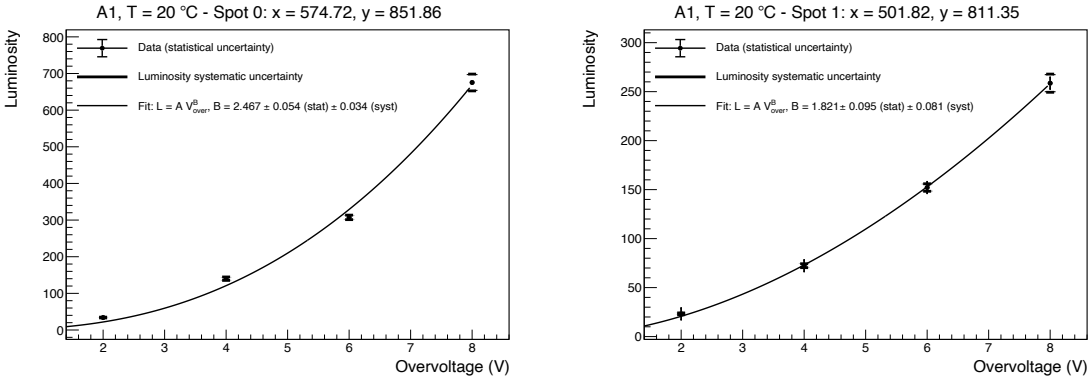
at $V_{\mathrm{over}} = 9\ \mathrm{V}$.

This temperature dependence is interpreted as being primarily associated with the temperature dependence of the dark-count rate (DCR), and hence with the number of thermally generated avalanches occurring during the camera exposure. Indeed, both the DCR and the dark current of SiPMs exhibit a strong temperature dependence, decreasing approximately exponentially as the temperature is lowered, with characteristic variations corresponding to roughly a factor of two every $10^{\circ}\mathrm{C}$. The similar temperature scale observed for the optical emission therefore suggests that its increase with temperature is mainly driven by the increasing avalanche rate, rather than by a larger amount of light emitted by each individual avalanche. This interpretation is consistent with Raymond et al. [54], who found no significant temperature dependence of the secondary-photon emission per avalanche.

5.1.4 Dependence on overvoltage

A third set of measurements, again only for sensor A1, were taken keeping the temperature fixed at $20\ ^{\circ}C$, whilst changing the overvoltage. The images were taken with an exposure time of $5000\ ms$ per 20 frames, 50 repetitions. As in the previous section, for reasons relating to the duration of data collection, the measurements are performed only on the central part of the sensor.

The dependence of luminosity on overvoltage was fitted with the empirical power law $L_{\mathrm{spot}} = A\, V_{\mathrm{over}}^{B}$, as can be seen in Figure 111, that is found to fit well for all hotspots. As illustrated in Figure 112 most of the analysed hotspots had a B parameter compatible with ≈ 2. This behaviour is consistent with the results of McLaughlin et al. [53] and Raymond et al. [54], which show that



(a) (b)

Figure 111: The luminosity as a function of overvoltage keeping the SiPM temperature fixed. Two example hotspots in the images.

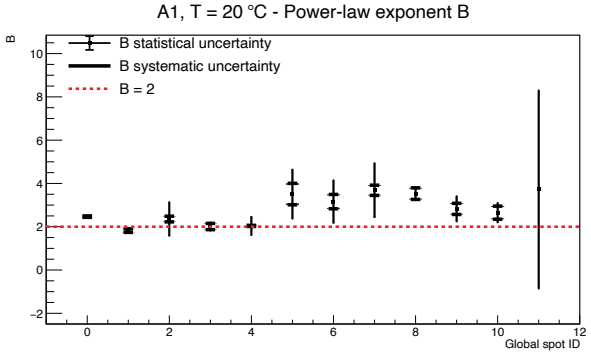


Figure 112: The extracted B values from the empirical power law $L_{\mathrm{spot}} = A\,V_{\mathrm{over}}^{B}$. A red dashed line for B=2 is plotted for reference.

SiPM photon emission depends on the operating voltage and on the secondary-photon yield per avalanche.

This approximately quadratic behaviour is consistent with the expected increase of the SiPM response with overvoltage. In particular, both the dark-count rate, and therefore the number of avalanches occurring during the camera exposure, and the avalanche gain increase with V_{over}. Since the measured EMMI signal results from the light emitted by the ensemble of avalanches integrated over the exposure time, the combined increase of avalanche rate and avalanche charge naturally leads to a stronger-than-linear dependence of the measured luminosity on overvoltage. Additional contributions from triggering probability, correlated avalanches, and photon production per avalanche may also play a role, so the observed approximately quadratic trend cannot yet be attributed to a single microscopic mechanism.

5.2 Temperature dependence of the emission intensity

After characterizing the emission properties of the new devices, the same analysis was extended to the irradiated SiPMs in order to investigate whether the observed temperature dependence is preserved after radiation damage. The temperature dependence of the optical emission provides a first quantitative characterization of the light-emitting sites identified in the EMMI images. Its interpretation requires, however, a clear distinction between the different physical quantities involved in the measurement. Physical defects in the silicon can modify the local carrier-generation probability and consequently the probability of initiating a Geiger avalanche. The EMMI analysis does not directly image these microscopic defects, but identifies spatially localized *light-emitting sites*, hereafter referred to as *emission hotspots*, or just *hotspots*. The quantity measured for each hotspot is the optical intensity accumulated by the camera during the complete exposure time.

The camera therefore does not measure an intrinsic "luminosity of a defect". Instead, the recorded intensity results from the secondary photons produced by all avalanches occurring in the corresponding region during the exposure. Schematically, the measured optical intensity can be expressed as

$$L_{\mathrm{meas}} \propto N_{\mathrm{av}}\, Y_{\gamma}\, \varepsilon_{\mathrm{opt}}, \tag{89}$$

where N_{av} is the number of Geiger avalanches occurring during the exposure, Y_{γ} is the average secondary-photon yield contributing to the detected signal per avalanche, and $\varepsilon_{\mathrm{opt}}$ represents the optical collection and detection efficiency of the experimental apparatus. For a fixed camera exposure time Δt,

$$N_{\mathrm{av}}(T) = R_{\mathrm{av}}(T)\, \Delta t, \tag{90}$$

where R_{av} represents the local avalanche rate.

This distinction is particularly important when studying the temperature dependence. The thermally generated contribution to the dark-count rate increases rapidly with temperature, since a larger number of charge carriers become available to initiate Geiger discharges. Defect-related generation centres can further enhance this process locally. A strong increase of the measured emission intensity with temperature can therefore arise primarily from an increase of the local avalanche rate, corresponding to a larger number of avalanches during the same camera exposure, rather than from an increase of the optical emission produced by each individual avalanche.

The measurements reported by Raymond et al. [54] provide an important reference for this interpretation. In that work, the secondary photon emission was measured under controlled avalanche stimulation and normalized to the number of avalanches. No significant temperature dependence of the photon yield per avalanche was observed within the experimental uncertainties over the investigated temperature range. The devices and experimental conditions are different from those considered in the present work; nevertheless, this result indicates that a pronounced temperature dependence of the optical signal accumulated by the camera does not require a correspondingly strong temperature dependence of the secondary-photon production mechanism itself.

The temperature evolution observed with EMMI is therefore interpreted primarily as a consequence of the temperature dependence of the underlying avalanche activity. This suggest that the measured intensity is correlated to the primary dark-count rate. Other processes, including afterpulsing and optical cross-talk, can also contribute generating additional avalanches and can themselves depend on the operating conditions.

5.2.1 Parametrization of the temperature dependence

For all the device conditions considered in this work, the optical intensity of the individual emission hotspots increases with temperature. Within the investigated temperature interval, this dependence is parametrized empirically by an exponential function,

$$L_i(T) = A_i e^{\lambda T}, \tag{91}$$

where i identifies the emission hotspot, A_i is a normalization parameter and λ quantifies the temperature dependence. The parameter A_i determines the absolute normalization of the optical signal of the i-th hotspot. When the temperature is expressed in degrees Celsius, it formally represents the extrapolated intensity at $T = 0\ ^\circ$C. Its numerical value depends on the absolute avalanche activity of the site, on the acquisition conditions, on the optical collection efficiency and on the adopted intensity-integration procedure. It should therefore not be interpreted as an intrinsic optical property of a physical defect.

The parameter λ, on the other hand, describes the relative temperature dependence of the measured intensity. In fact,

$$\frac{1}{L_i}\frac{dL_i}{dT} = \lambda, \tag{92}$$

so that λ can be regarded as an effective fractional temperature coefficient of the accumulated optical emission. A useful quantity for expressing this dependence in a more intuitive form is the temperature interval required for the measured intensity to increase by a factor of two. From Eq. 91,

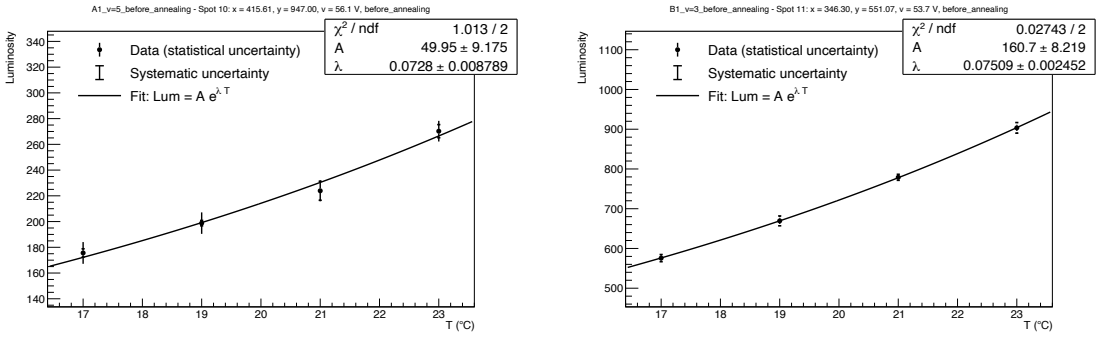
$$\frac{L(T + \Delta T_2)}{L(T)} = e^{\lambda \Delta T_2} = 2, \tag{93}$$

which gives

$$\Delta T_2 = \frac{\ln 2}{\lambda}. \tag{94}$$

5.2.2 Fit quality and hotspot-to-hotspot variability

The exponential dependence introduced in Eq. 91 is fitted independently to each emission hotspot. This approach makes it possible both to verify whether the chosen parametrization provides an adequate description of the temperature dependence of the individual sites and to quantify the variation of the temperature coefficient between different hotspots.



(a) Sensor A1.

(b) Sensor B1.

Figure 113: Representative examples of the optical intensity measured for individual emission hotspots as a function of temperature. The curves represent fits with the function $L_i(T) = A_i e^{\lambda_i T}$. The two examples illustrate the approximately exponential increase observed in both irradiated sensors.

The representative examples shown in Figure 113 illustrate the typical behaviour observed in both irradiated sensors. For the post irradiation (pre-annealing) measurements, the selected samples contain 50 hotspots for A1 and 52 for B1. The distributions of the reduced chi-square values are concentrated around values of order unity, with median values of approximately 0.67 for A1 and 0.70 for B1. The exponential function therefore provides an adequate empirical description of the temperature evolution for the majority of the selected individual hotspots.

The values of the fitted coefficient λ_i, however, are not identical for all emission sites. This is already evident in the irradiated sensors before annealing. For A1, the selected hotspot population has an average

$$\langle \lambda \rangle = (0.086 \pm 0.008)\,^{\circ}\mathrm{C}^{-1}, \tag{95}$$

with an RMS of

$$\mathrm{RMS}(\lambda) = 0.053\,^{\circ}\mathrm{C}^{-1}. \tag{96}$$

For B1, the corresponding values are

$$\langle \lambda \rangle = (0.069 \pm 0.009)\,^{\circ}\mathrm{C}^{-1}, \tag{97}$$

and

$$\mathrm{RMS}(\lambda) = 0.0413\,^{\circ}\mathrm{C}^{-1}. \tag{98}$$

This values shows that the temperature dependence is characterized by a significant hotspot-to-hotspot dispersion. This effect can be seen in Figures 114-115

For this reason, the average coefficient should not be interpreted as a universal temperature constant shared by all emission hotspots. Instead, it provides a compact characterization

of the hotspot population at a given annealing stage, while the RMS contains complementary information on the degree of heterogeneity among the individually fitted sites.

For a population containing N selected hotspots, the quantities used in the following comparison are calculated as

$$\langle \lambda \rangle = \frac{1}{N} \sum_{i=1}^{N} \lambda_i, \tag{99}$$

with statistical uncertainty

$$\sigma_{\langle \lambda \rangle}^{\mathrm{stat}} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\lambda_i - \langle \lambda \rangle\right)^2} \tag{100}$$

and population RMS

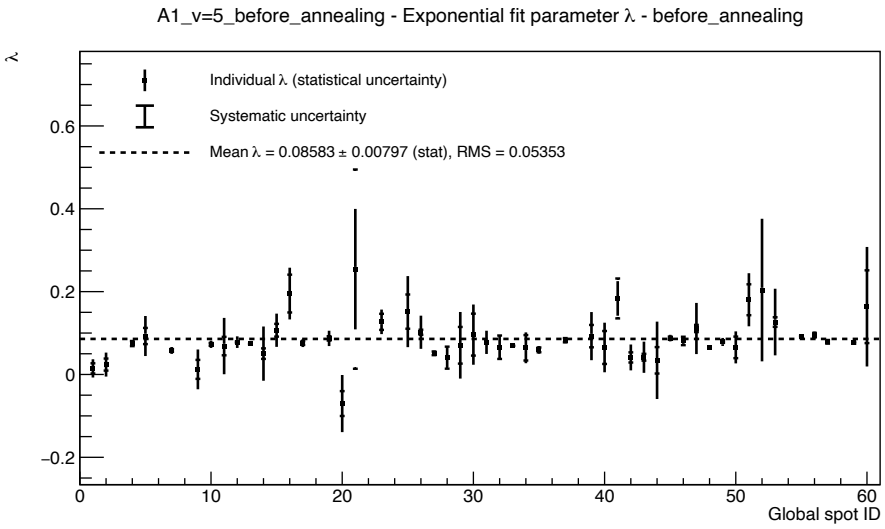
$$\mathrm{RMS}(\lambda) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\lambda_i - \langle \lambda \rangle\right)^2}. \tag{101}$$

The statistical uncertainty in Eq. 100 quantifies the uncertainty propagated from the individual fitted coefficients, whereas the RMS in Eq. 101 describes the actual spread of the hotspot population. The two quantities therefore have different meanings and should not be directly compared as alternative estimates of the same uncertainty.

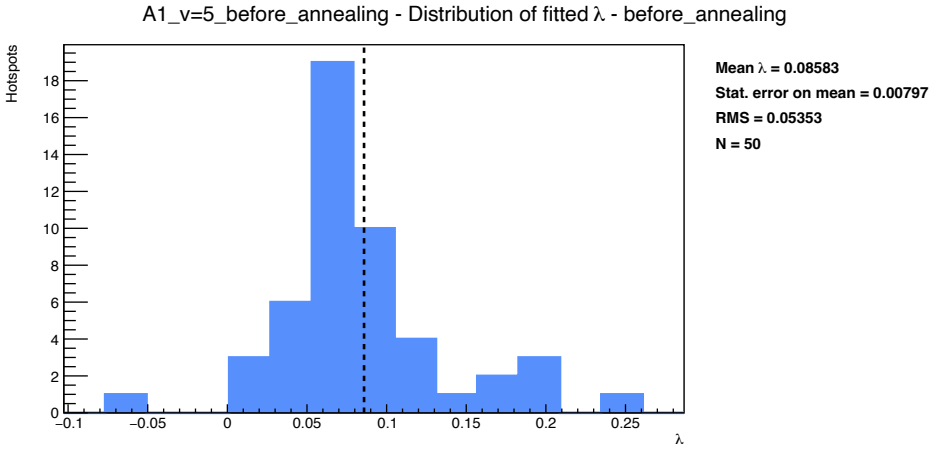
5.2.3 Comparison between new and irradiated sensor

A direct comparison with the measurements performed before irradiation provides an indication of whether radiation damage modifies the temperature dependence of the optical emission. For the new sensor, at $V_{\mathrm{over}} = 7$ V, the temperature coefficient from the constant fit was found to be $\lambda = (0.061 \pm 0.003)\ ^{\circ}\mathrm{C}^{-1}$. After irradiation, the corresponding A1 measurement before annealing gives a mean value of $\langle \lambda \rangle = 0.086 \pm 0.008\ ^{\circ}\mathrm{C}^{-1}$, with an RMS of $0.054\ ^{\circ}\mathrm{C}^{-1}$ across the hotspot population.

Considering these results, the temperature coefficient therefore increases significantly after irradiation. This indicates that irradiation modifies the characteristic temperature dependence of the optical emission, leading to a stronger increase of the emitted luminosity with temperature.

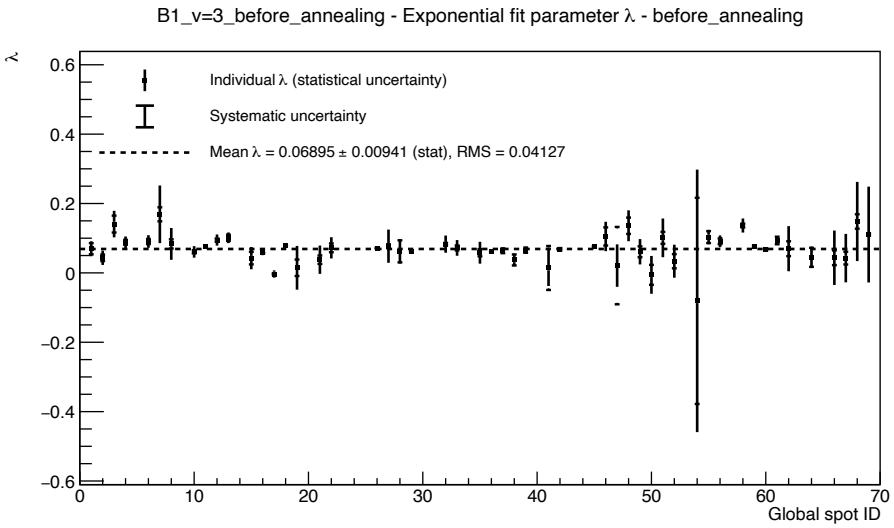


(a)

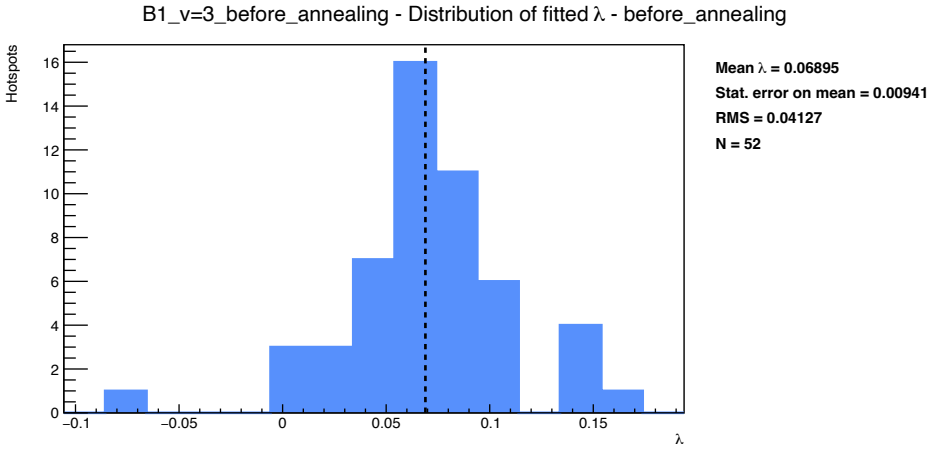


(b)

Figure 114: Values of the temperature coefficient λ_i obtained from the independent exponential fits for sensor A1 before annealing. Up: λ_i values for each emission hotspots; the mean is represented as a dashed line. Down: Distributions of the λ_i values; it shows the significant hotspot-to-hotspot dispersion around the mean value.



(a)



(b)

Figure 115: Values of the temperature coefficient λ_i obtained from the independent exponential fits for sensor B1 before annealing. Up: λ_i values for each emission hotspots; the mean is represented as a dashed line. Down: Distributions of the λ_i values; it shows the significant hotspot-to-hotspot dispersion around the mean value.

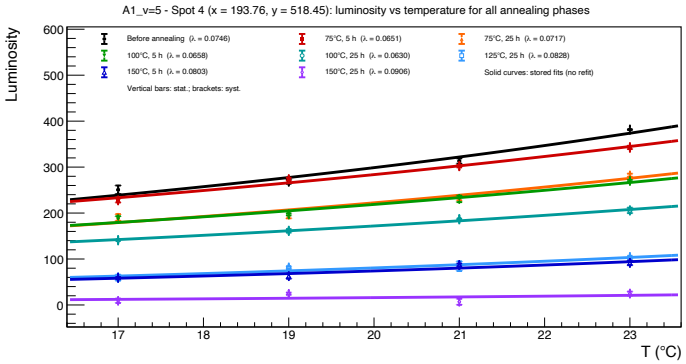


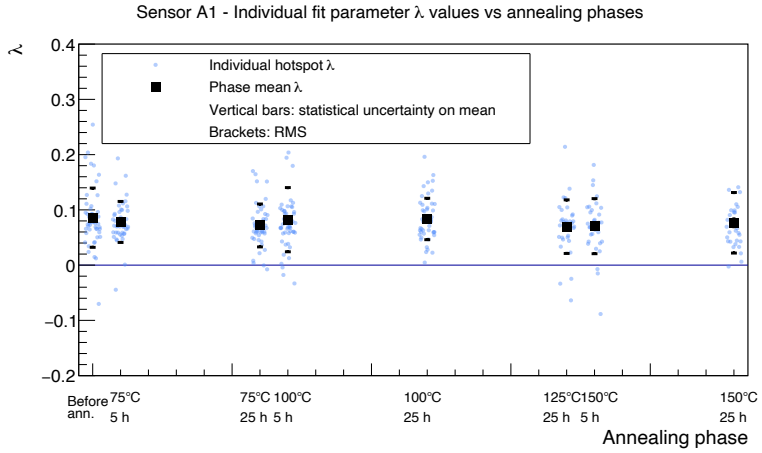
Figure 116: Luminosity versus temperature for the same hotspot (global ID 4), the different annealing phases are represented with different colours.

5.2.4 Evolution of the temperature coefficient with annealing

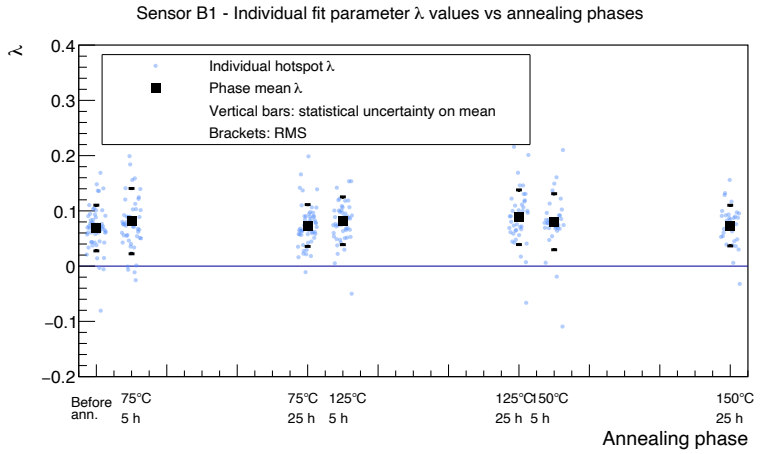
The individual hotspots exponential fit and then the averaging across al detected hotspots is repeated independently for each annealing phase. The average temperature coefficient, its propagated statistical uncertainty, and the RMS of the corresponding hotspot distribution are reported in Tab. 7. The results show that the characteristic temperature coefficient remains of the same order of magnitude. An exemplary hotspot can be seen in Figure 116.

Table 7: Average temperature coefficient obtained from the individual hotspot fits at the different annealing stages and corresponding temperature interval ΔT_2. Missing phases are due to weather-related issues that occurred during data collection, which rendered the data-taking problematic.

Sensor	Phase	N_{spots}	$\langle\lambda\rangle$ [°C^{-1}]	RMS(λ) [°C^{-1}]	ΔT_2 [°C]
A1	Before annealing	50	0.086 ± 0.008	0.054	8.1 ± 0.7
A1	75°C, 5 h	46	0.078 ± 0.006	0.037	8.9 ± 0.6
A1	75°C, 25 h	47	0.072 ± 0.006	0.039	9.7 ± 0.8
A1	100°C, 5 h	52	0.082 ± 0.008	0.058	8.4 ± 0.8
A1	100°C, 25 h	52	0.084 ± 0.005	0.037	8.3 ± 0.5
A1	125°C, 25 h	34	0.070 ± 0.008	0.048	10.0 ± 1.2
A1	150°C, 5 h	33	0.071 ± 0.009	0.050	9.8 ± 1.2
A1	150°C, 25 h	32	0.077 ± 0.010	0.055	9.0 ± 1.2
B1	Before annealing	52	0.069 ± 0.009	0.041	10.1 ± 0.8
B1	75°C, 5 h	50	0.081 ± 0.008	0.059	8.5 ± 0.9
B1	75°C, 25 h	52	0.073 ± 0.005	0.038	9.4 ± 0.7
B1	125°C, 5 h	48	0.082 ± 0.006	0.043	8.4 ± 0.6
B1	125°C, 25 h	41	0.088 ± 0.008	0.049	7.8 ± 0.7
B1	150°C, 5 h	36	0.080 ± 0.009	0.051	8.6 ± 0.9
B1	150°C, 25 h	32	0.073 ± 0.007	0.037	9.5 ± 0.8



(a) Sensor A1.

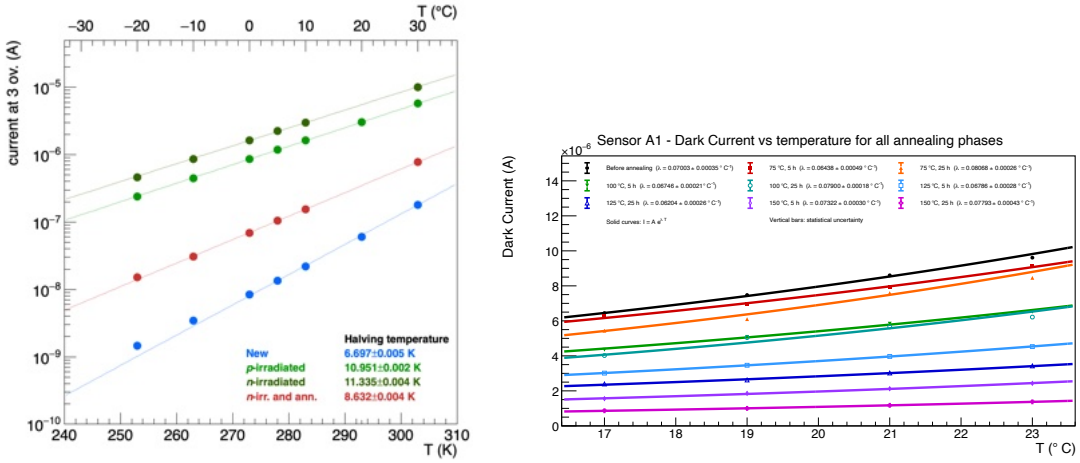


(b) Sensor B1.

Figure 117: Average $\langle\lambda\rangle$ coefficient obtained from the individual hotspot fits as a function of the annealing phase for sensors A1 and B1, plotted alongside the individual fit values for each hotspot (in blue). Missing phases are due to weather-related issues that occurred during data collection, which rendered the data inaccessible.

For A1 the average coefficient varies between 0.070 and $0.086\ ^{\circ}\mathrm{C}^{-1}$. Sensor B1 shows a similar range, with average values between 0.069 and $0.088\ ^{\circ}\mathrm{C}^{-1}$ throughout the complete annealing sequence. These values correspond to a characteristic intensity-doubling interval of 8–10°C, according to Eq. 94.

Thus, despite the substantial evolution of the radiation-induced hotspot population during annealing, the typical temperature dependence of the detectable emission sites remains approximately of the same scale over most of the thermal treatment. No clear monotonic evolution of $\langle\lambda\rangle$ with annealing temperature or duration is observed. For example, in A1 the mean



(a) (b)

Figure 118: Behaviour of the dark current at different temperatures. Left: Halving temperature for S13360-3050VS in the new, proton-irradiated, neutron-irradiated, and annealed ($150°C$ for 150 hours) states; the image from [59] is reported here for reference across a broader temperature range. Right: dark current for sensor A1 at each annealing stages.

coefficient decreases during the $75°$C treatment, increases again during the $100°$C treatment, and subsequently returns to values around $0.07\ °\mathrm{C}^{-1}$. B1 shows similar moderate fluctuations. This behaviour is illustrated in Figure 117, where the population averages are shown as a function of the annealing phase.

The RMS values provide an important additional element for interpreting this figure. The RMS lies between 0.04 and $0.06\ °\mathrm{C}^{-1}$, demonstrating that the spread between individual hotspots remains significant throughout the annealing sequence. Consequently, relatively small variations of the population mean from one phase to the next should not be interpreted as evidence for a coherent modification of the temperature response of every individual hotspot.

As a final consideration on the temperature dependence of the optical emission, it is useful to compare its evolution with that observed for the dark current. The approximately constant value of the mean temperature coefficient λ throughout the annealing sequence differs from the behaviour of the dark current: as shown in the Figure 118, ΔT increases after irradiation and subsequently decreases during annealing, which would correspond to an increase of the associated λ_{DC}. This different evolution indicates that the temperature dependences of the optical emission and of the dark current are not governed by exactly the same underlying processes, suggesting that additional mechanisms contribute to the measured luminosity; further dedicated studies will be required to clarify this behaviour.

5.3 Overvoltage dependence of the emission intensity

The dependence of the measured optical emission on the SiPM overvoltage provides complementary information to the temperature dependence discussed in Sec. 5.2. As in the previous analysis, it is important to distinguish between the microscopic defects present in the silicon, the light-emitting sites identified in the EMMI images, the optical intensity measured from these sites, and the underlying avalanche activity.

The optical intensity accumulated by the camera is produced by all Geiger avalanches occurring in the region associated with a given hotspot during the exposure. Increasing the overvoltage simultaneously affects the avalanche charge, triggering probability, dark-count rate, correlated avalanches and secondary-photon production. The measured EMMI signal therefore represents the combined response of several overvoltage-dependent processes rather than a direct measurement of a single microscopic quantity.

Schematically, the measured intensity can be written as

$$L_{\mathrm{meas}}(V_{\mathrm{over}}) \propto R_{\mathrm{av}}(V_{\mathrm{over}})\Delta t\, Y_{\gamma}(V_{\mathrm{over}})\varepsilon_{\mathrm{opt}}, \tag{102}$$

where R_{av} represents the total local avalanche rate, Δt is the camera exposure time, Y_{γ} is the effective photon yield associated with the avalanches, and $\varepsilon_{\mathrm{opt}}$ accounts for the optical collection and detection efficiency.

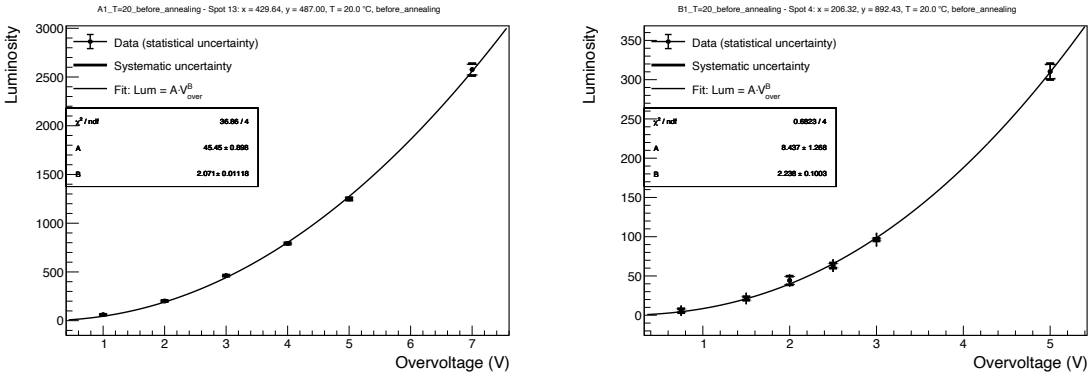
One of the most direct dependences arises from the avalanche charge. To first order,

$$Q_{\mathrm{av}} = C_{\mathrm{cell}} V_{\mathrm{over}}, \tag{103}$$

where Q_{av} is the charge released during a single Geiger avalanche, C_{cell} is the effective capacitance of the SiPM microcell, and V_{over} is the overvoltage, defined as the difference between the applied bias voltage and the breakdown voltage, $V_{\mathrm{over}} = V_{\mathrm{bias}} - V_{\mathrm{bd}}$. This relation shows that, to first order, the avalanche charge and therefore the SiPM gain increase linearly with overvoltage.

The accumulated optical intensity, however, is not expected to follow this dependence alone. Spatially resolved studies of SiPM dark noise have shown that the emitted intensity increases with excess bias as a consequence of both increasing gain and increasing dark-count rate [55]. In addition, measurements of the secondary emission have observed an increase of the photon yield with overvoltage even after normalisation to the charge passing through the device [53]. Measurements of the photon yield per avalanche lead to the same conclusion [54], indicating that the emission process itself also depends on the operating voltage.

Correlated avalanches provide a further contribution. Both optical cross-talk and afterpulsing become more probable as the overvoltage increases, increasing the total number of avalanches contributing to the optical signal. Consequently, an approximately quadratic dependence of the measured EMMI intensity on overvoltage should not be interpreted as an intrinsic quadratic law associated with an individual physical defect, but rather as an effective scaling of the complete measured avalanche-related emission.



(a) Sensor A1.

(b) Sensor B1.

Figure 119: Representative examples of the optical intensity measured from individual emission hotspots as a function of overvoltage at $T = 20°$C before annealing. The curves represent the individual fits with $L_i = A_i V_{\mathrm{over}}^{B_i}$.

5.3.1 Parametrization of the overvoltage dependence

The measurements were performed at a fixed temperature of $T = 20°$C while varying the SiPM overvoltage. For each hotspot, six overvoltage values were included in the fit, spanning the intervals

$$V_{\mathrm{over}} \in [1, 2, 3, 4, 5, 7]\ \mathrm{V} \qquad \text{for A1,} \tag{104}$$

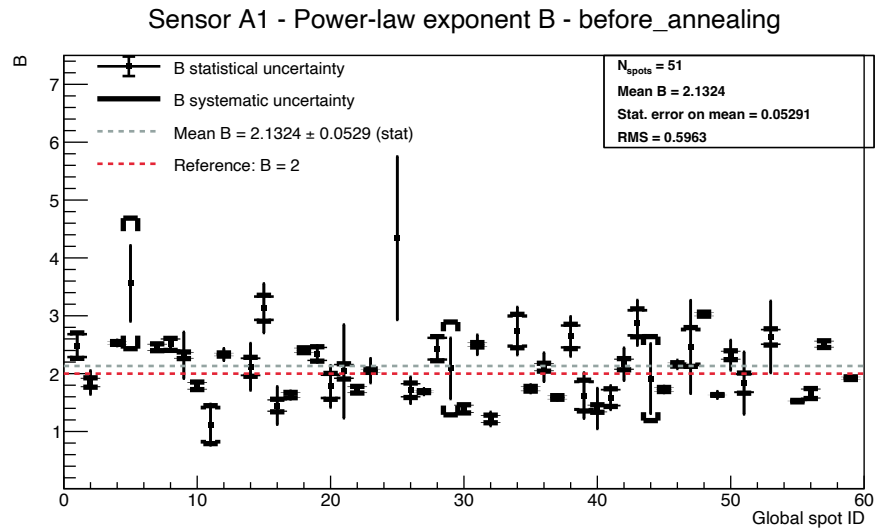
and

$$V_{\mathrm{over}} \in [0.75, 1.5, 2, 2.5, 3, 5]\ \mathrm{V} \qquad \text{for B1.} \tag{105}$$

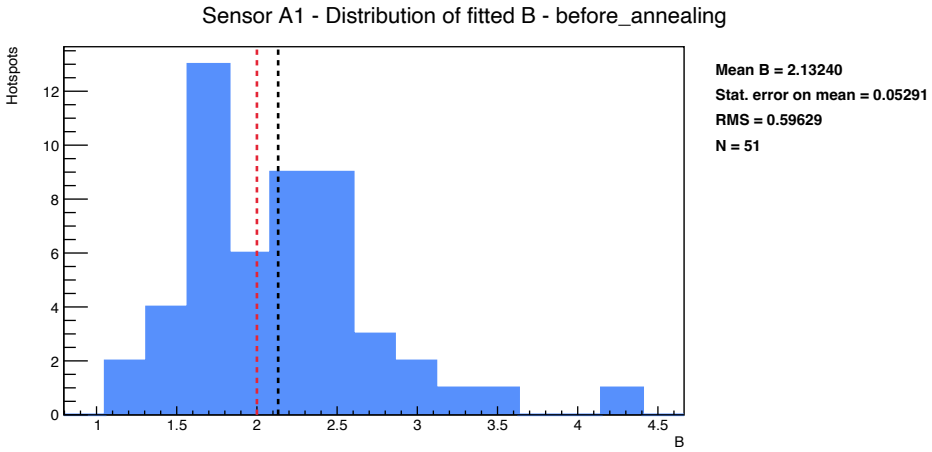
The dependence is fitted independently for each hotspot using the empirical power law

$$L_i(V_{\mathrm{over}}) = A_i V_{\mathrm{over}}^{B_i}, \tag{106}$$

where i identifies the hotspot, A_i determines its intensity scale and B_i describes the power of its overvoltage dependence; both A_i and B_i are therefore determined independently for each emission site. The power law provides a compact empirical parametrization over the experimentally investigated voltage range and should not be interpreted as a fundamental microscopic law for defect-related carrier generation. Representative individual fits are shown in Figure 119. Both examples illustrate the increase of the optical signal with overvoltage and the ability of the empirical power law to reproduce the measured dependence over the selected voltage interval.

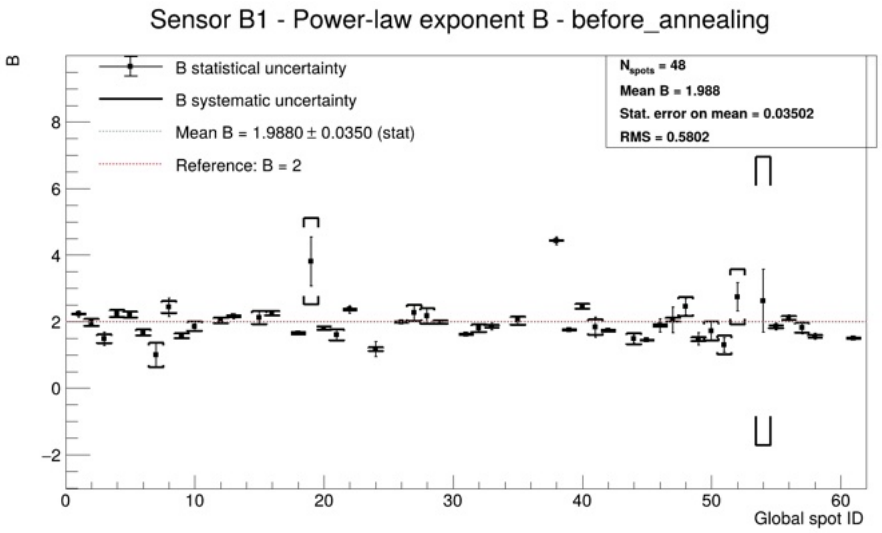


(a)

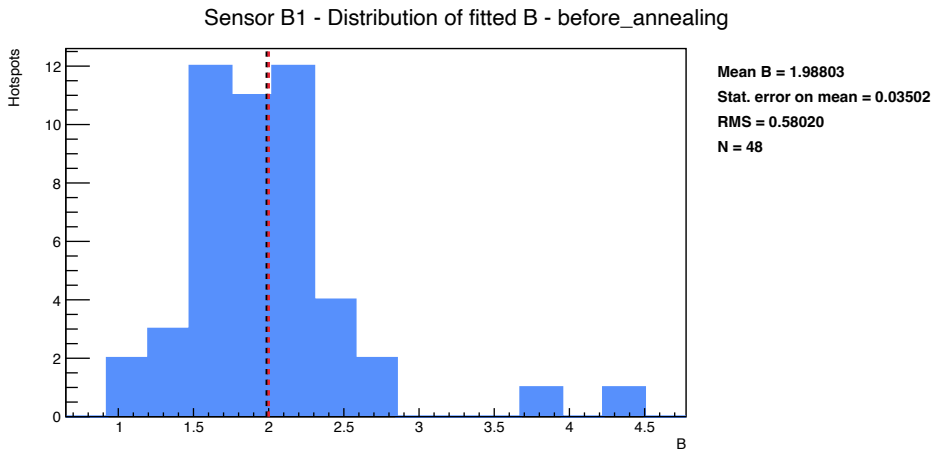


(b)

Figure 120: Values of the B coefficient obtained from the independent exponential fits for sensor A1 before annealing. Left: B values for each emission hotspots; the mean is represented as a dashed line. Right: Distributions of the λ_i values; it shows the significant hotspot-to-hotspot dispersion around the mean value.



(a)



(b)

Figure 121: Values of the coefficient B obtained from the independent exponential fits for sensor B1 before annealing. Left: B values for each emission hotspots; the mean is represented as a dashed line. Right: Distributions of the B values; it shows the significant hotspot-to-hotspot dispersion around the mean value.

5.3.2 Variation among individual emission hotspots

The independent determination of B_i makes it possible to investigate directly whether the different emission hotspots exhibit the same response to overvoltage. The individual fits show that this is not strictly the case. Although the extracted exponents are generally distributed around values of order two, a sizeable hotspot-to-hotspot dispersion is observed for both sensors.

For this reason, the characteristic overvoltage dependence of a given sensor and annealing phase is described by the arithmetic mean of the exponents obtained from the individual hotspot fits,

$$\langle B \rangle = \frac{1}{N} \sum_{i=1}^{N} B_i, \tag{107}$$

where N is the number of individually fitted hotspots. The statistical uncertainty associated with the mean is obtained by propagating the statistical uncertainties of the individual fit parameters,

$$\sigma_{\langle B \rangle}^{\mathrm{stat}} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} \left(B_i - \langle B \rangle\right)^2}. \tag{108}$$

The hotspot-to-hotspot variability is instead quantified through the RMS of the B_i distribution,

$$\mathrm{RMS}(B) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(B_i - \langle B \rangle\right)^2}. \tag{109}$$

The statistical uncertainty on $\langle B \rangle$ and the RMS therefore describe two conceptually different quantities. The former quantifies the uncertainty associated with the determination of the population mean, whereas the latter characterizes the actual dispersion of the individual hotspot responses. In the following, both quantities are considered when assessing how representative the mean exponent is of the complete hotspot population.

5.3.3 Quality of the mean exponent description

The individual results obtained before annealing provide a first indication of how accurately a single mean exponent represents the hotspot population. Figures 120–121 show the distribution of the B values for each device. Both sensors exhibit distributions centred around values of order two.

For sensor A1, the individual fits before annealing give

$$\langle B \rangle_{\mathrm{A1}} = 2.13 \pm 0.05, \tag{110}$$

with an RMS of 0.60. Sensor B1 gives

$$\langle B \rangle_{\mathrm{B1}} = 1.99 \pm 0.035, \tag{111}$$

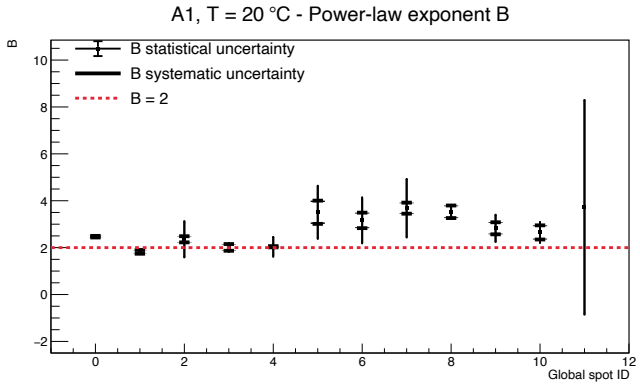


Figure 122: Exponents B_i extracted from the individual overvoltage fits for the non-irradiated sensor A1 at $T = 20°$C. The horizontal reference at $B = 2$ indicates an exactly quadratic dependence.

with a smaller RMS of 0.58. In both cases, the population average is therefore compatible with a quadratic dependence of the measured optical intensity on overvoltage.

However, the individual blue points in Figure 123 demonstrate that this characteristic value should not be interpreted as a universal exponent shared by all emission hotspots. This is particularly evident for A1, whose distribution before annealing is considerably broader than that observed for B1. Individual emission centres can therefore exhibit substantially different sensitivities to overvoltage even though their population average remains close to $B = 2$.

The simultaneous representation of the mean and of all individual results in Figure 123 is particularly useful in this respect, since it makes it possible to distinguish variations of the population average from variations of its internal dispersion.

5.3.4 Comparison between new and irradiated sensors

The measurements performed on the non-irradiated sensor provide the reference to confront to the overvoltage response after irradiation. As discussed in Sec. 5.1, the individual hotspots observed in the new sensor were characterized by exponents predominantly distributed around $B \simeq 2$, as shown again in Figure 122. The measured emission therefore already exhibited an approximately quadratic dependence on overvoltage before irradiation.

After irradiation, the characteristic exponents remain of the same order. Before the beginning of the annealing treatment, the mean values obtained from the individual fits are $\langle B \rangle = 2.13 \pm 0.05$ for A1 and $\langle B \rangle = 1.99 \pm 0.04$ for B1. The first annealing phase shown for each sensor in Figure 123 therefore remains centred close to the quadratic behaviour observed in the new device.

The comparison between Figure 122 and Figure 123 indicates that irradiation does not produce a transition towards a qualitatively different dependence of the optical intensity on overvoltage. This is noteworthy because irradiation produces a very large modification of the EMMI hotspot population, as discussed in Sec. 5.4, while the characteristic exponent describing the overvoltage response remains of order two.

5.3.5 Evolution during annealing

The same individual-fit procedure was repeated at every stage of the annealing sequence for both sensors. The evolution is summarized in Figure 123, where the mean exponent at each phase is shown together with the complete population of individual B_i values. The statistical uncertainty associated with the mean describes its determination from the individual fits, while the RMS brackets quantify the spread of the corresponding hotspot population. The quantitative results are also reported in Tab. 8

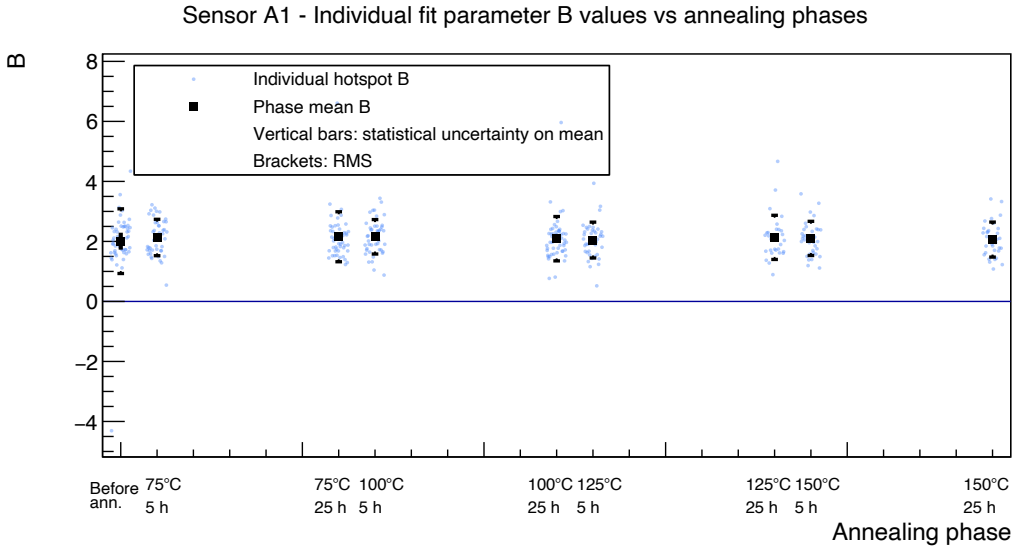
“‘latex “‘

Table 8: Mean values of the exponent B obtained from the individual hotspot fits at $T = 20°$C for sensors A1 and B1 throughout the annealing sequence. The uncertainty associated with $\langle B \rangle$ is statistical, while $\mathrm{RMS}(B)$ quantifies the hotspot-to-hotspot dispersion of the individual fitted exponents.

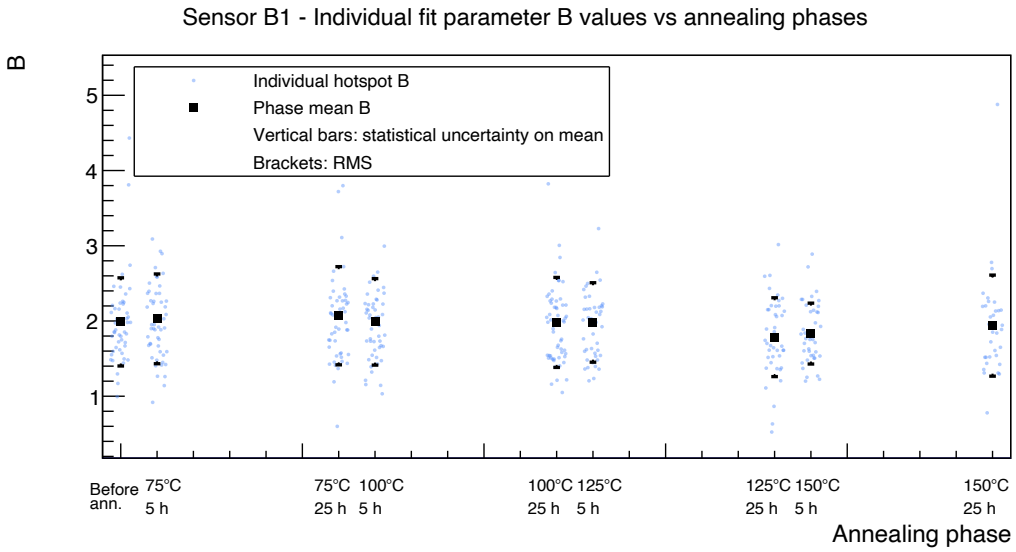
Sensor	Phase	N_{spots}	$\langle B \rangle$	$\mathrm{RMS}(B)$
A1	Before annealing	51	2.13 ± 0.05	0.60
A1	$75°$C, 5 h	44	2.13 ± 0.06	0.59
A1	$75°$C, 25 h	46	2.16 ± 0.05	0.82
A1	$100°$C, 5 h	46	2.15 ± 0.06	0.56
A1	$100°$C, 25 h	52	2.09 ± 0.05	0.72
A1	$125°$C, 5 h	41	2.05 ± 0.05	0.58
A1	$125°$C, 25 h	30	2.14 ± 0.07	0.72
A1	$150°$C, 5 h	35	2.10 ± 0.07	0.56
A1	$150°$C, 25 h	32	2.06 ± 0.07	0.57
B1	Before annealing	48	1.99 ± 0.03	0.58
B1	$75°$C, 5 h	51	2.03 ± 0.04	0.59
B1	$75°$C, 25 h	50	2.07 ± 0.07	0.65
B1	$100°$C, 5 h	50	1.99 ± 0.03	0.57
B1	$100°$C, 25 h	57	1.98 ± 0.03	0.59
B1	$125°$C, 5 h	48	1.98 ± 0.03	0.52
B1	$125°$C, 25 h	41	1.78 ± 0.03	0.52
B1	$150°$C, 5 h	40	1.83 ± 0.04	0.40
B1	$150°$C, 25 h	34	1.94 ± 0.04	0.67

The behaviour of A1 is shown in Figure 123a. The mean exponent remains close to two throughout the complete annealing sequence, with only moderate variations between consecutive treatments. No systematic increase or decrease of $\langle B \rangle$ is observed with increasing annealing temperature or duration. The characteristic overvoltage scaling of the A1 hotspot population therefore remains approximately quadratic even though the population itself evolves substantially during the thermal treatment.

The individual blue points in Figure 123a provide additional information that is not contained in the mean alone. A sizeable dispersion is present at all phases, showing that the stability of $\langle B \rangle$ does not imply that all hotspots respond identically to overvoltage. Changes in the composition of the detectable population can occur while leaving its average exponent unchanged.



(a) Sensor A1.



(b) Sensor B1.

Figure 123: Evolution of the exponent B during the annealing sequence for sensors A1 and B1. The highlighted points represent the mean exponent $\langle B \rangle$ obtained from the individual hotspot fits at each phase, while the blue points show the corresponding individual B_i values. The brackets represent the RMS of each hotspot population.

Sensor B1, shown in Figure 123b, exhibits a somewhat different evolution. During the first part of the annealing sequence, up to the $125°$C, 5 h treatment, the mean exponent remains close to its initial value and therefore close to an approximately quadratic dependence. At the later stages, a more pronounced variation of the mean becomes visible: the population shifts towards lower B values after the $125°$C, 25 h and $150°$C, 5 h treatments, followed by a partial recovery after the final $150°$C, 25 h treatment.

This variation is nevertheless accompanied by a broad distribution of individual B_i values, clearly visible in Figure 123b. The distributions corresponding to successive annealing stages substantially overlap, so that the evolution cannot be described simply as a uniform displacement of all hotspot exponents. Rather, the position and width of the population evolve together.

The different behaviours observed for A1 and B1 also indicate that no universal monotonic dependence of B on the annealing temperature or accumulated annealing time can be identified. In A1 the population mean is comparatively stable, whereas B1 develops a more evident variation during the high-temperature stages. This difference further supports the interpretation of $\langle B \rangle$ as an effective property of the experimentally observed hotspot population rather than as a direct parameter characterizing a unique microscopic defect mechanism.

An additional aspect to consider is that the population of detectable hotspots itself changes during annealing. As discussed in Sec. 5.5, emission centres can decrease in intensity, disappear below the detection threshold, or become detectable at different stages of the treatment. Consequently, the set of hotspots contributing to the distribution of B_i is not necessarily identical at every annealing phase. An evolution of $\langle B \rangle$ can therefore result both from changes in the overvoltage response of individual emission centres and from changes in the composition of the observable population.

Overall, the overvoltage measurements consistently yield characteristic exponents of order two for the new sensor, the irradiated devices and most stages of the annealing treatment. The individual-fit approach, however, shows that this approximately quadratic scaling represents an average property of the hotspot population and not a universal response of individual emission sites. The evolution of $\langle B \rangle$ and of the corresponding hotspot-to-hotspot dispersion should therefore be interpreted together with the changes in hotspot number and luminosity discussed in Secs. 5.5 and 5.6.

5.4 Effects of irradiation

After establishing the reference behaviour of the non-irradiated devices in Sec. 5.1, the effect of irradiation can be investigated by directly comparing the EMMI images acquired before and after irradiation. The comparison is performed in terms of the number detected emission hotspots and the distribution of their measured luminosity.

These observables describe different aspects of the radiation-induced modification of the SiPM. In particular, an increase in the number of detected hotspots indicates larger generated radiation damage, while changes in the intensity distribution characterize the level of avalanche activity.

5.4.1 Number of the emission hotspots

The most immediate effect observed after irradiation is the strong increase in the number of light-emitting sites detected across the SiPM surface. As discussed in Sec. 5.1, the non-irradiated devices contain only a limited population of detectable hotspots under the reference operating conditions. After irradiation, in contrast, the EMMI images are characterized by several hundred spatially resolved emission sites.

The hotspot-detection procedure described in Sec. 4.5 was applied independently to the four irradiated sensors. The number of detected hotspots obtained before the beginning of the annealing sequence is reported in Tab. 9, expressing the mean number of hotspots per mm^2 with respect to the active area $= 8.96\, mm^2$ for sensors A1 and A2 and $= 8.91\, mm^2$ for sensors B1 and B2.

Table 9: Number of emission hotspots detected in the four irradiated sensors after irradiation with a proton beam of approximately $140\,\mathrm{MeV}$, with a neutron-equivalent radiation level of $10^9\ 1\ MeV\ n_{eq}/cm^2$, before the annealing sequence.

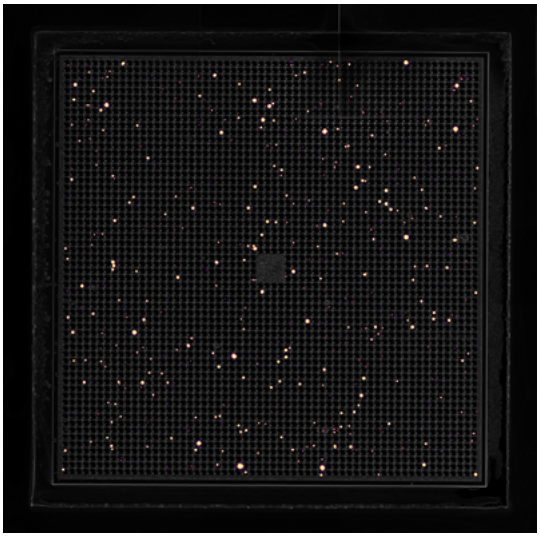
Sensor	Sensor type	After irradiation	Mean n. $hotspots/mm^2$
A1	S13360-3050VS	562 ± 29	62.7 ± 3.2
A2	S13360-3050VS	596 ± 34	66.5 ± 3.8
B1	S13360-3075VS	535 ± 15	60.0 ± 1.7
B2	S13360-3075VS	527 ± 13	59.1 ± 1.5

For all four sensors, the post-irradiation population is of the order of $500 - 600$ detected hotspots. The effect of therefore involves a substantial increase in the number of spatial locations from which avalanche-induced optical emission can be detected.

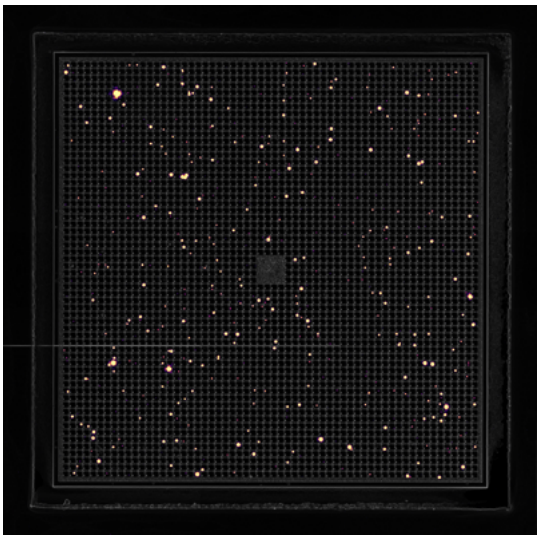
Figure 124 shows superimposition of the EMMI images and the optical images of the four sensors after irradiation (before annealing). The same detection criteria are used for the four devices, allowing their spatial patterns to be compared directly.

The images in Figure 124 show that the large increase in the number of detected hotspots is not confined to a single macroscopic region of the devices. Instead, emission sites are observed throughout the imaged sensor surface, which is consistent with a uniform radiation damage cause by the uniform irradiation field. This observation is qualitatively consistent with the microscopic picture of displacement damage discussed in Sec. 3.4. Radiation-induced modifications of the silicon can introduce generation centres within the depleted regions of many different microcells. When such centres enhance the probability of carrier generation, the corresponding SPAD can undergo dark Geiger avalanches more frequently. Since the EMMI camera integrates the secondary optical emission produced by these avalanches, spatially localized increases in avalanche activity appear as additional emission hotspots.

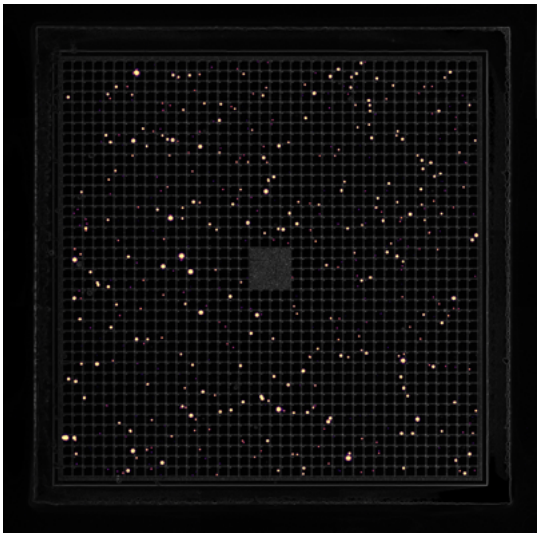
It is important, however, not to establish a one-to-one correspondence between each detected optical hotspot and a single microscopic defect. The spatial resolution of the optical measurement is much larger than the atomic scale of the radiation-induced defects, and several microscopic generation centres may contribute to the avalanche activity associated with the same detected emission site. Conversely, whether a damaged region becomes visible as a



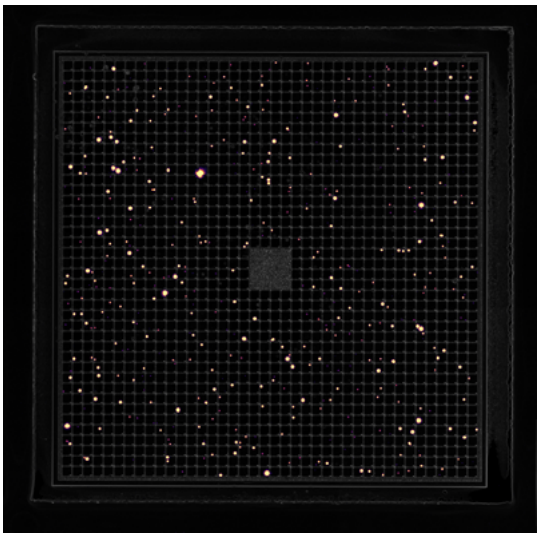
(a) Sensor A1.



(b) Sensor A2.

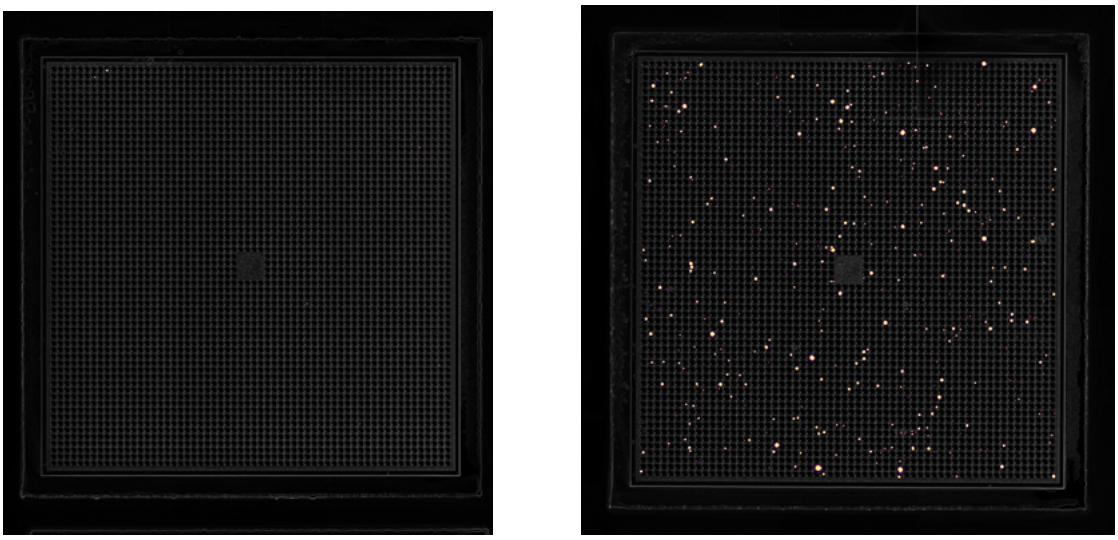


(c) Sensor B1.



(d) Sensor B2.

Figure 124: Spatial distribution of the emission hotspots detected after irradiation and before the annealing sequence for sensors A1, A2, B1 and B2. The detected light-emitting sites are visible in the images, obtained superimposing the EMMI image and the optical image. Irradiation produces a dense population of emission hotspots distributed over the imaged sensor area.



(a) Before irradiation.

(b) After irradiation, before annealing.

Figure 125: Direct comparison of the spatial distribution of the detected emission hotspots before and after irradiation for sensor A1. The sensor is placed under the same conditions of $T = 20°C$ and $V_{over} = 5\ V$. The comparison illustrates the strong increase in the number of spatially localized emission sites following irradiation.

hotspot depends on its avalanche rate, the operating conditions and the detection threshold of the EMMI analysis. The number of detected hotspots should therefore be regarded as an experimental observable characterizing the spatial distribution of enhanced avalanche activity, rather than as a direct count of radiation-induced lattice defects.

Figure 125 is intended to show the same sensors (A1) before and after irradiation using comparable visualization conditions.

The principal effect established by the counting measurement is therefore a large increase in the number of detectable emission hotspots following irradiation, accompanied by a spatially widespread distribution over the sensor surface. This result provides the first indication that irradiation creates, or activates above the EMMI detection threshold, a large population of additional regions characterized by enhanced avalanche activity. The next section investigates whether irradiation also modifies the distribution of the measured optical intensities.

5.4.2 Distribution of the emission intensities

In addition to the strong increase in the number of detected emission hotspots discussed in Sec. 5.4.1, the emission hotspots generated through irradiation exhibit a different distribution of optical intensities compared to the hotspots already present before irradiation.

To enable a direct comparison with the reference measurements presented in Sec. 5.1, the analysis was restricted to the central region of the sensor, corresponding to an active area of $3.08\ \mathrm{mm}^2$. Only measurements acquired at $T = 20°\mathrm{C}$ and $V_{\mathrm{over}} = 7\ \mathrm{V}$ were considered, and

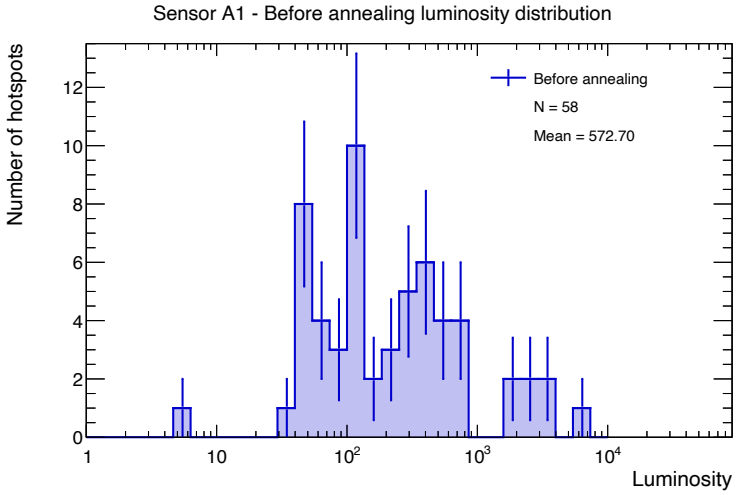


Figure 126: Distribution of the optical intensity measured for the emission hotspots of sensor A1 after irradiation at $T = 20°$C and $V_{\text{over}} = 7$ V. Logarithmic binning is used because of the broad range of measured intensities.

the same isolated-hotspot selection adopted throughout the luminosity analyses was applied.

The intensity distribution measured for sensor A1 after irradiation is shown in Figure 126. Since the measured intensities extend over several orders of magnitude, logarithmic binning is used in order to represent simultaneously the large population of relatively weak emission sites and the much brighter tail of the distribution.

Under these selection criteria, 60 emission hotspots are detected. Their measured optical intensities span more than three orders of magnitude, from approximately 5.8 up to

$$L_{\text{max}} \simeq 1.2 \times 10^4. \tag{112}$$

The distribution is strongly asymmetric, with most hotspots concentrated in the lower-intensity region and a comparatively small population extending towards very large values. The median intensity is approximately

$$L_{\text{median}} \simeq 210, \tag{113}$$

whereas the arithmetic mean is larger,

$$L_{\text{mean}} \simeq 572. \tag{114}$$

The large difference between the mean and median reflects the pronounced high-intensity tail and shows that the population cannot be represented adequately by a single average value.

Comparison with the non-irradiated sensor. The effect of irradiation on intensity distribution can be seen more directly by comparing the measurements obtained before irradiation and after irradiation, before annealing. The two distributions are shown together in Figure 127. Since

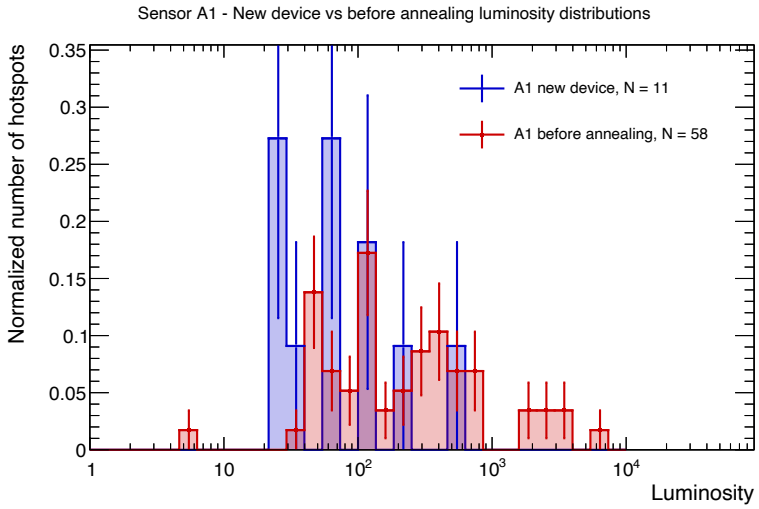


Figure 127: Normalized distributions of the optical intensity of the emission hotspots measured in sensor A1 before irradiation and after irradiation, before annealing, at $T = 20°$C and $V_{\text{over}} = 7$ V. Only sites detected at the selected operating condition are considered. Both distributions are normalized and represented using logarithmic binning, allowing their shapes to be compared independently of the different number of hotspots.

the two data sets contain different numbers of detected hotspots, both histograms are normalized. The comparison therefore emphasizes differences in the relative intensity distributions independently of the increase in the absolute number of detected sites.

Before irradiation, the reference sample contains 11 detected hotspots. The measured intensities lies within the range [23,479]a.u., with a median of about 56 a.u. and a mean of 108 a.u.. The non-irradiated distribution is therefore concentrated predominantly in the low-intensity region and covers a comparatively limited range.

After irradiation, a clear change in the shape of the distribution is observed. The upper limit of the measured intensities range increases significantly; the new range is in fact [40,2600]a.u. The median increases from 56 a.u. to 210 a.u., almost a factor of four, while the distribution becomes substantially broader and develops a pronounced tail towards high intensities. 9 hotspots, corresponding to $\sim 15\%$ of the population, reach intensities above 10^3 a.u., well beyond the range covered by the non-irradiated sample.

The normalized comparison therefore shows that the effect of irradiation cannot be described simply as the appearance of a larger number of hotspots drawn from the same underlying intensity distribution. Even after removing the difference in population size through normalization, the post-irradiation distribution is shifted towards larger intensities, spans a much broader range and develops a high-intensity tail that is absent from the reference measurement.

The effect of irradiation is consequently visible through two distinct but complementary observables. First, the number of detected light-emitting sites increases strongly, as discussed in Sec. 5.4.1. Second, the sites that are detected after irradiation exhibit a substantially different distribution of accumulated optical emission. Irradiation therefore modifies both the population

of spatial regions contributing to the EMMI signal and the range of avalanche-associated optical intensities observed across that population.

The limited statistics of the non-irradiated sample should be kept in mind when comparing the detailed shapes of the two normalized distributions. Nevertheless, the extension of the post-irradiation population far beyond the complete pre-irradiation intensity range provides a clear indication that irradiation produces a qualitatively broader distribution of light-emitting activity.

5.5 Evolution with annealing

5.5.1 Hotspot counting – whole device

The first observable used to characterize the evolution of the irradiated SiPMs during annealing is the number of detected emission hotspots over the complete active area of each device. In contrast with the analyses involving the optical intensity of individual hotspots, the counting measurement does not require the selected emission sites to be spatially isolated. The isolation criterion introduced for the quantitative luminosity analysis is therefore not applied in this case.

This distinction is important because the objective of the present analysis is exclusively to determine how many spatially resolved light-emitting sites are detected at each annealing phase. No integration of the optical signal around the individual centres is required. The source-identification algorithm can separate nearby emission sites and is consequently suitable for counting the complete detected hotspot population, including regions where several hotspots are located relatively close to one another. The whole-device measurement therefore provides the most inclusive characterization of the evolution of the hotspot population during annealing.

The number of detected hotspots for the four irradiated devices throughout the complete annealing sequence is reported in Tab. 10.

Table 10: Number of emission hotspots detected over the complete device area at the different annealing phases for sensors A1, A2, B1 and B2.

Annealing phase	A1	A2	B1	B2
Before annealing	562 ± 29	596 ± 34	535 ± 15	527 ± 13
75°C, 5 h	554 ± 31	564 ± 30	518 ± 15	509 ± 12
75°C, 25 h	526 ± 29	545 ± 33	505 ± 15	483 ± 16
100°C, 5 h	486 ± 21	498 ± 29	472 ± 14	466 ± 16
100°C, 25 h	433 ± 23	414 ± 19	420 ± 8	423 ± 17
125°C, 5 h	408 ± 20	387 ± 20	378 ± 16	371 ± 13
125°C, 25 h	345 ± 19	360 ± 19	341 ± 18	320 ± 13
150°C, 5 h	318 ± 12	316 ± 22	284 ± 13	256 ± 14
150°C, 25 h	251 ± 17	240 ± 21	219 ± 17	231 ± 12

From Figure 128 a common trend is immediately visible for all four devices: the number of detected emission hotspots decreases monotonically throughout the annealing sequence, with steps corresponding to the temperature increase. The reduction is already measurable after

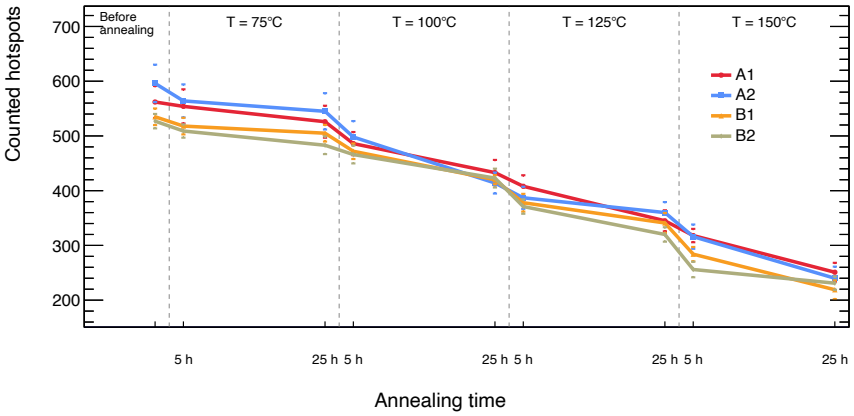


Figure 128: Graphical representation of Tab. 10: number of hotspots at different annealing stages for the four sensors.

the first 75°C treatment, although at this stage it remains relatively small. After 5 h at 75°C, the number of hotspots decreases from 562 to 554 for A1, from 596 to 564 for A2, from 535 to 518 for B1, and from 527 to 509 for B2. These variations correspond to reductions of only a few percent, indicating that the first low-temperature annealing step produces a comparatively modest modification of the detected population.

A more pronounced decrease becomes evident during the subsequent 100°C and 125°C treatments. After the complete 100°C, 25 h phase, the populations have fallen to 433 hotspots for A1, 414 for A2, 420 for B1 and 423 for B2. After annealing at 125°C for 25 h, the corresponding values are further reduced to 345, 360, 341 and 320. Thus, by this stage of the annealing sequence, approximately one third to almost one half of the hotspots detected immediately after irradiation are no longer identified under the same measurement and detection conditions.

The largest overall reduction is reached after the final 150°C treatments. After 25 h at 150°C, only 251 hotspots remain detected in A1, 240 in A2, 219 in B1 and 231 in B2. Relative to the post-irradiation condition before annealing, this corresponds to reductions of approximately

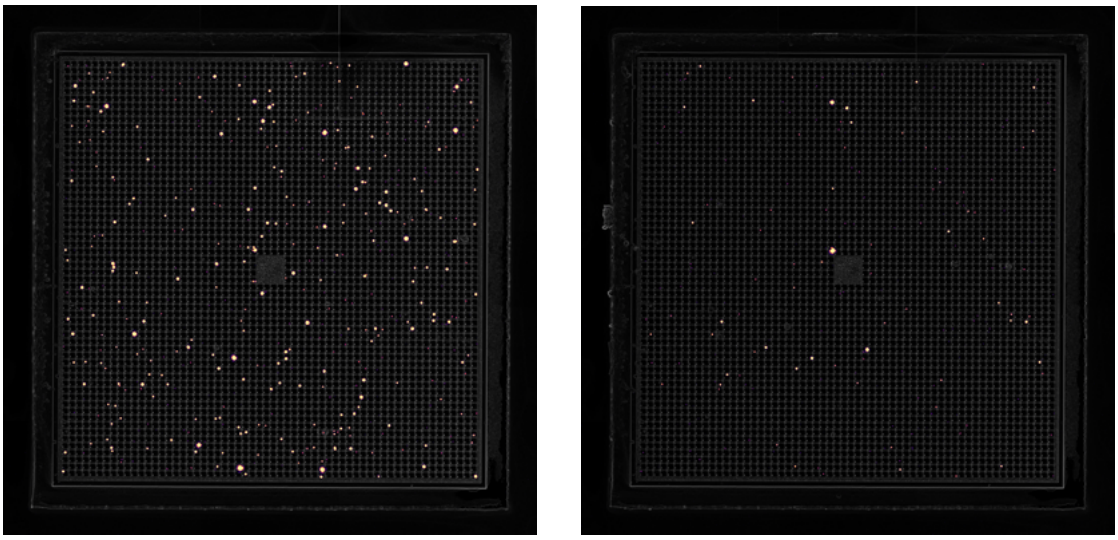
$$55.3\% \quad (\mathrm{A1}), \qquad 59.7\% \quad (\mathrm{A2}), \tag{115}$$

and

$$59.1\% \quad (\mathrm{B1}), \qquad 56.2\% \quad (\mathrm{B2}). \tag{116}$$

Equivalently, the population remaining detectable after the complete annealing sequence corresponds to only about 40–45% of the hotspot population initially observed after irradiation. No device remains approximately unchanged, nor does any sensor show a qualitatively different trend from the others at the level of whole-device hotspot counting.

The evolution can be appreciated visually by comparing the complete EMMI images acquired before annealing with those obtained after the final 150°C, 25 h treatment. These comparisons are shown separately for the four sensors in Figures 129– 132.



(a) Before annealing: 562 hotspots.

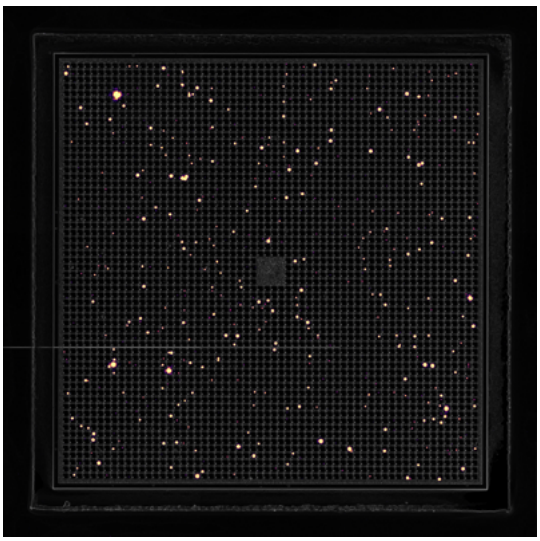
(b) 150°C, 25 h: 251 hotspots.

Figure 129: Whole-device EMMI images of sensor A1 before annealing and after the final treatment at 150°C for 25 h. The reduction in the number of detected emission hotspots is visible over the complete sensor area.

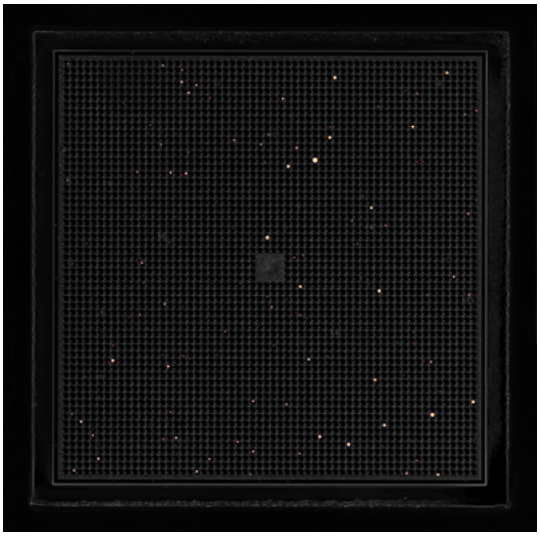
The image comparisons confirm that the decrease reported in Tab. 10 is a global feature of the sensor surface rather than the consequence of the disappearance of activity from a single restricted region. Emission hotspots remain distributed across the device after annealing, but their overall spatial density is progressively reduced.

It is important to stress that the disappearance of a hotspot from the EMMI image does not necessarily imply the physical removal of the microscopic defect or defects responsible for carrier generation at that position. The observable considered here is the number of *detected light-emitting sites* in the visible spectrum. Annealing can modify the microscopic configuration of radiation-induced defects and reduce their contribution to carrier generation and dark avalanche activity. As a consequence, the accumulated optical emission associated with a given location can fall below the detection threshold even if residual electrically active defects remain present. Alternatively, it is possible that, following the annealing stages, the photons are emitted at wavelengths outside the visible spectrum and are therefore no longer detectable by our apparatus.

The whole-device counting measurement therefore demonstrates that annealing progressively reduces the number of spatial regions producing detectable avalanche-associated optical emission. Because the annealing treatments are performed sequentially, the evolution shown here reflects the cumulative thermal history of each device: the effects of annealing time and temperature cannot be completely separated from the counting data alone. Nevertheless, two robust features are common to all four sensors. First, every successive phase produces a decrease in the detected hotspot population. Second, the final reduction is large and comparable among the four devices, leaving only approximately 40–45% of the original post-irradiation

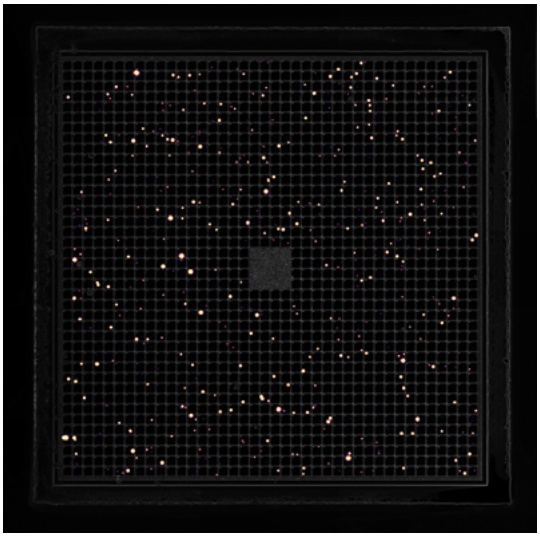


(a) Before annealing: 596 hotspots.

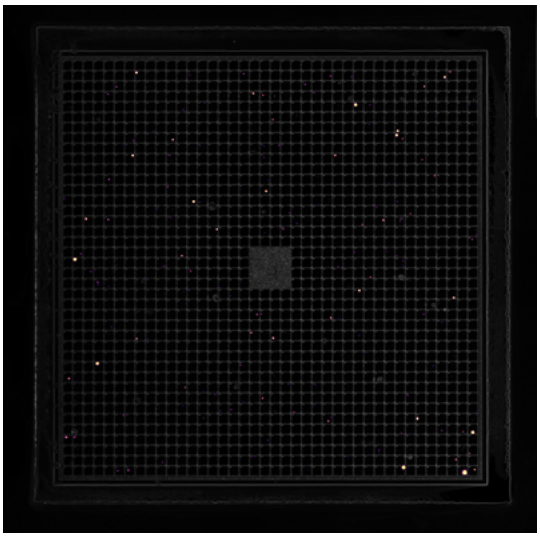


(b) $150°$C, 25 h: 240 hotspots.

Figure 130: Whole-device EMMI images of sensor A2 before annealing and after the final treatment at $150°$C for 25 h. A substantial reduction of the detected hotspot population is observed after the complete annealing sequence.

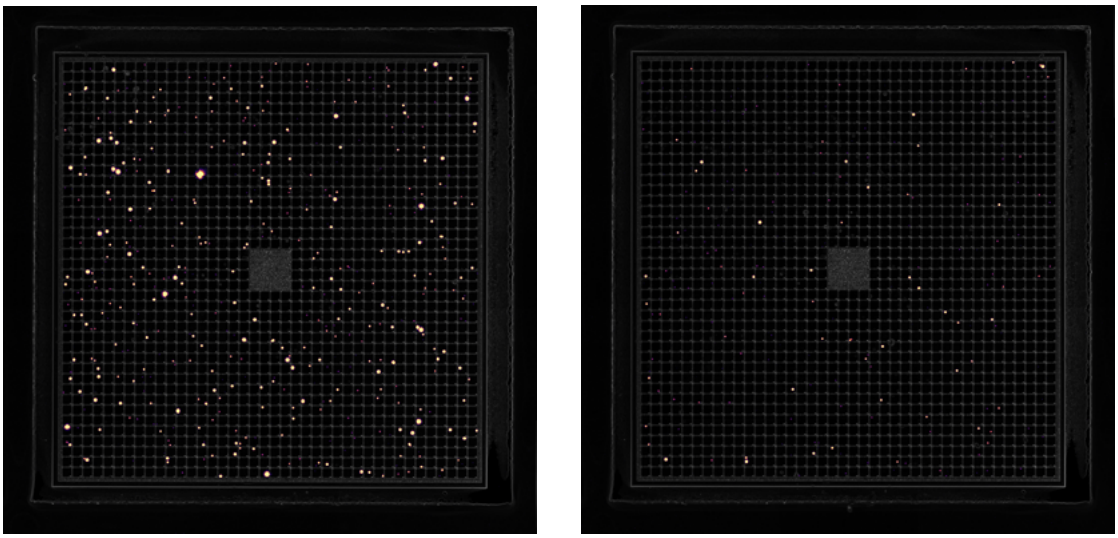


(a) Before annealing: 535 hotspots.



(b) $150°$C, 25 h: 219 hotspots.

Figure 131: Whole-device EMMI images of sensor B1 before annealing and after the final treatment at $150°$C for 25 h. The final image contains considerably fewer detected emission hotspots than the irradiated sensor before annealing.



(a) Before annealing: 527 hotspots.

(b) $150°$C, 25 h: 231 hotspots.

Figure 132: Whole-device EMMI images of sensor B2 before annealing and after the final treatment at $150°$C for 25 h. As for the other devices, the annealing sequence produces a marked reduction in the number of detected emission hotspots.

hotspot population detectable after the complete annealing sequence.

The whole-device result establishes the macroscopic evolution of the radiation-induced emission pattern. In the following subsection, the same evolution is studied within the central sensor region used for the more detailed hotspot-by-hotspot analysis, allowing the counting results to be connected directly with the subsequent measurements of optical intensity.

5.5.2 Hotspot counting – central region

In order to connect the evolution of the hotspot population with the subsequent hotspot-by-hotspot analysis, the counting procedure was repeated within the central region of sensors A1 and B1. This is the same region used for the measurements of the temperature and overvoltage dependences and for the quantitative study of the optical intensity.

As in the whole-device analysis of Sec. 5.5.1, the counting is performed on the complete hotspot population within the selected region, without requiring the emission sites to satisfy the isolation criterion used for the luminosity integration. The purpose of this analysis is again only to determine the presence or absence of each detected light-emitting site.

The main advantage of restricting the analysis to the central region is that the spatial matching procedure described in Sec. 4.5 allows the same hotspot to be followed through the successive annealing phases by means of its global identifier. The evolution can therefore be classified in more detail. At each phase, a hotspot can be: *detected*, if it is present in the corresponding image; *appeared*, if it is detected for the first time during the annealing sequence; *disappeared*, if it was detected in the preceding evolution but is no longer observed; or *reappeared*, if it

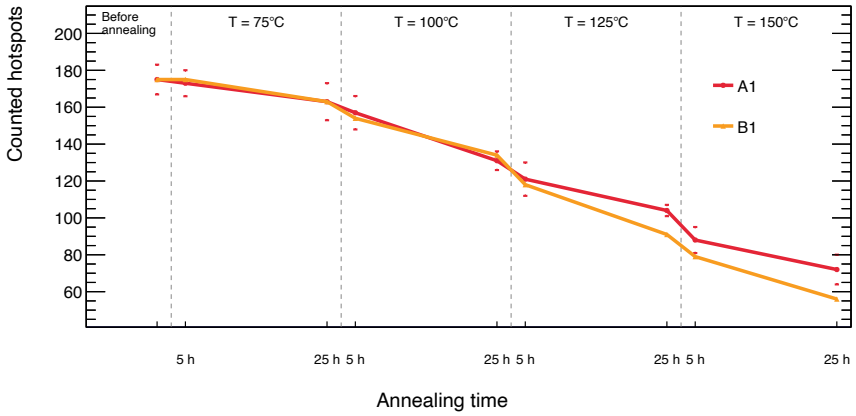


Figure 133: Graphical representation of Tabs. 11- 12: number of hotspots in the cental region at different annealing stages for the two sensors A1 and B1.

becomes detectable again after a previous disappearance.

The resulting evolution for sensor A1, measured at $T = 20°$C and $V_{\text{over}} = 5$ V, is reported in Tab. 11. For reappearing hotspots, the table also indicates the most recent annealing phase in which the hotspot had been classified as disappeared.

The corresponding results for sensor B1, measured at $T = 20°$C and $V_{\text{over}} = 3$ V, are shown in Tab. 12.

The overall evolution is consistent with the whole-device counting, as can be seen in Figure 133. In both central regions, the number of detected hotspots decreases progressively as the annealing sequence proceeds. For A1, the population decreases from 175 hotspots before annealing to 72 after the final $150°$C, 25 h treatment, corresponding to a reduction of approximately 59%. For B1, the population decreases from 175 to 56 hotspots, corresponding to a reduction of approximately 68%.

The global decrease, however, does not imply that the evolution consists only of a progressive disappearance of emission sites. The global-ID tracking reveals a more dynamic behaviour: at several stages, a small number of hotspots is detected for the first time, while other sites that had previously disappeared become visible again. Reappearance can occur after a single annealing step or after several subsequent phases in which the hotspot remains undetected. For example, among the nine A1 hotspots classified as reappeared after $150°$C for 5 h, some had disappeared as early as the $100°$C treatments, whereas others had disappeared only during the preceding $125°$C phase.

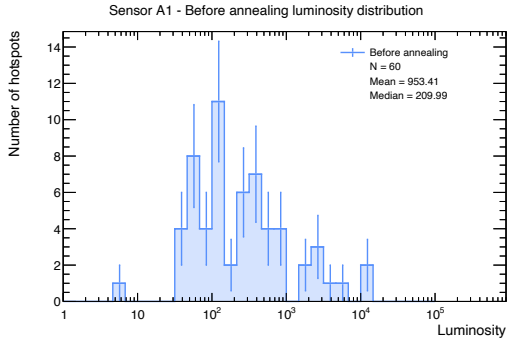
These classifications refer exclusively to the experimental detectability of the light-emitting sites. In particular, “appearance”, “disappearance” and “reappearance” do not necessarily correspond to the creation, removal or reconstruction of an individual microscopic defect. A hotspot can move above or below the EMMI detection threshold as its local avalanche activity changes. The tracking analysis nevertheless provides an important additional observable: instead of describing annealing only through the total number of hotspots, it makes it possible to follow the

Table 11: Evolution of the hotspot population in the central region of sensor A1 at $T = 20°$C and $V_{\mathrm{over}} = 5$ V. The "reappeared" column specifies the most recent phase in which the corresponding hotspots had disappeared.

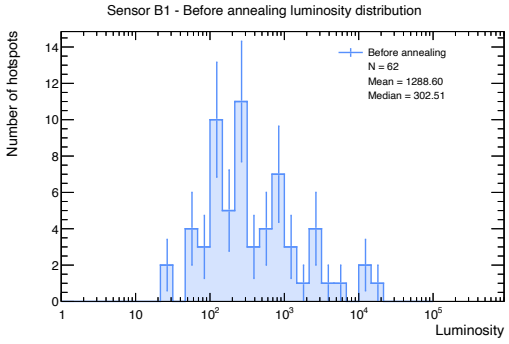
Annealing phase	Detected	Appeared	Reappeared	Disappeared
Before annealing	175 ± 8	0 ± 0	0 ± 0	0 ± 0
$75°$C, 5 h	173 ± 7	6 ± 4	0 ± 0	8 ± 3
$75°$C, 25 h	163 ± 10	3 ± 6	2 ± 2 total from $75°$C, 5 h: 2	15 ± 1
$100°$C, 5 h	157 ± 9	6 ± 1	5 ± 2 total from $75°$C, 5 h: 1 from $75°$C, 25 h: 4	17 ± 3
$100°$C, 25 h	131 ± 5	1 ± 1	3 ± 1 total from $75°$C, 25 h: 2 from $100°$C, 5 h: 1	30 ± 4
$125°$C, 5 h	121 ± 9	3 ± 1	6 ± 0 total from $100°$C, 5 h: 1 from $100°$C, 25 h: 5	19 ± 3
$125°$C, 25 h	104 ± 3	3 ± 0	10 ± 2 total from $100°$C, 5 h: 2 from $100°$C, 25 h: 3 from $125°$C, 5 h: 5	30 ± 5
$150°$C, 5 h	88 ± 7	3 ± 0	9 ± 1 total from $100°$C, 5 h: 1 from $100°$C, 25 h: 4 from $125°$C, 5 h: 1 from $125°$C, 25 h: 3	28 ± 5
$150°$C, 25 h	72 ± 8	0 ± 1	7 ± 0 total from $125°$C, 25 h: 1 from $150°$C, 5 h: 6	23 ± 5

Table 12: Evolution of the hotspot population in the central region of sensor B1 at $T = 20°C$ and $V_{over} = 3$ V. The "reappeared" column specifies the most recent phase in which the corresponding hotspots had disappeared.

Annealing phase	Detected	Appeared	Reappeared	Disappeared
Before annealing	175 ± 10	0 ± 0	0 ± 0	0 ± 0
$75°C$, 5 h	175 ± 5	13 ± 2	0 ± 0	13 ± 7
$75°C$, 25 h	163 ± 6	6 ± 2	3 ± 2 total from $75°C$, 5 h: 3	21 ± 2
$100°C$, 5 h	154 ± 4	4 ± 1	7 ± 1 total from $75°C$, 5 h: 3 from $75°C$, 25 h: 4	20 ± 4
$100°C$, 25 h	134 ± 5	5 ± 1	4 ± 1 total from $75°C$, 5 h: 1 from $100°C$, 5 h: 3	29 ± 1
$125°C$, 5 h	118 ± 8	3 ± 0	11 ± 2 total from $100°C$, 5 h: 2 from $100°C$, 25 h: 9	30 ± 2
$125°C$, 25 h	91 ± 8	1 ± 3	5 ± 3 total from $125°C$, 5 h: 5	33 ± 4
$150°C$, 5 h	79 ± 3	3 ± 1	8 ± 2 total from $100°C$, 25 h: 1 from $125°C$, 5 h: 2 from $125°C$, 25 h: 5	23 ± 4
$150°C$, 25 h	56 ± 3	1 ± 1	5 ± 2 total from $125°C$, 25 h: 1 from $150°C$, 5 h: 4	29 ± 1



(a) Sensor A1, $T = 20°\text{C}$ and $V_{\text{over}} = 7$ V.



(b) Sensor B1, $T = 20°\text{C}$ and $V_{\text{over}} = 5$ V.

Figure 134: Luminosity distributions of the emission hotspots detected before annealing for sensors A1 and B1. The distributions are evaluated at fixed operating conditions for each sensor and shown using logarithmic binning.

history of individual emission sites throughout the thermal treatment.

This information will be used in the following subsection to relate the evolution of the hotspot population to the evolution of their measured optical intensity.

5.5.3 Luminosity distributions

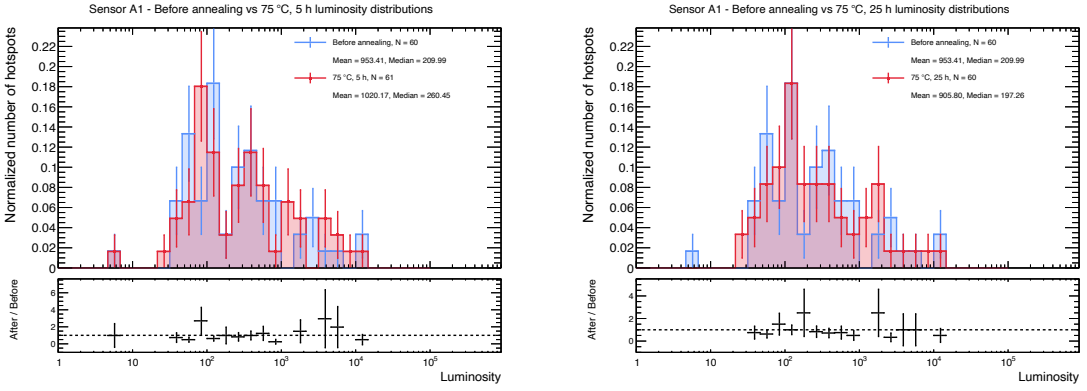
The hotspot-counting analyses presented in Secs. 5.5.1 and 5.5.2 show that annealing progressively reduces the number of detectable light-emitting sites. A complementary question is whether the luminosity distribution of the hotspots that remain detectable is also modified during the thermal treatment.

For this purpose, the luminosity distributions are compared at fixed operating conditions for each sensor. Sensor A1 is evaluated at $T = 20°\text{C}$ and $V_{\text{over}} = 7$ V, while sensor B1 is evaluated at $T = 20°\text{C}$ and $V_{\text{over}} = 5$ V. Only hotspots detected under the selected operating condition are included in the distribution. Furthermore, as with all measurements relating to brightness, only isolated hotspots have been taken into account in these distributions; for this reason, the number of hotspots differs from that obtained during the count in Sec. 5.5.2.

Since the number of selected hotspots changes throughout the annealing sequence, the distributions are normalized. This removes the effect of the changing population size and allows the comparison to focus on the *shape* of the luminosity distribution. Logarithmic binning is used because the measured intensities span several orders of magnitude.

Before annealing, both irradiated sensors exhibit broad and asymmetric luminosity distributions, with a large population of relatively weak emission sites and a long tail extending towards high intensities, as can be seen from Figure 134. As already discussed in Sec. 5.4.2, this broad distribution is one of the characteristic effects observed after irradiation.

Sensor A1 For A1, the pre-annealing distribution contains 60 selected hotspots and has a median luminosity of approximately



(a) (b)

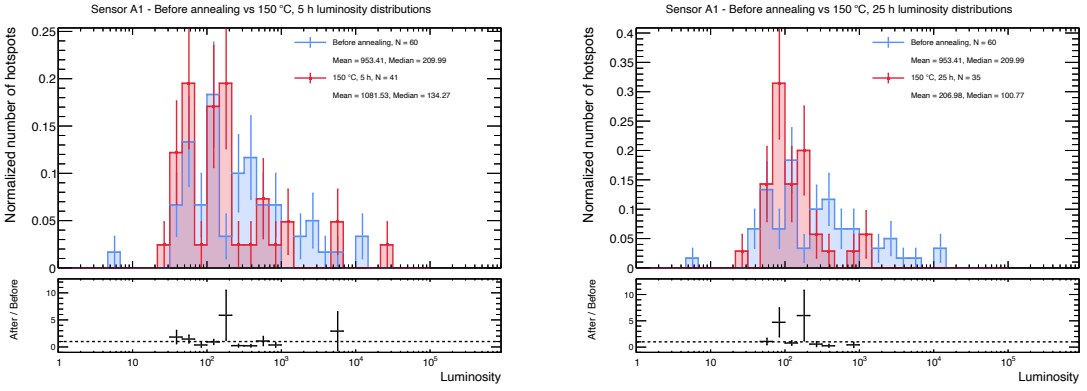
Figure 135: Normalized luminosity distributions of the emission hotspots detected at the different annealing stages for sensors A1. The distributions are evaluated at fixed operating conditions for each sensor and shown using logarithmic binning. Normalization allows the evolution of the distribution shape to be compared independently of the reduction in the number of detected hotspots.

$$L^{\mathrm{A1}}_{\mathrm{median}} \simeq 210. \tag{117}$$

Meanwhile the high-intensity tail extends beyond 10^4. During the first annealing stages, the shape of the distribution does not evolve monotonically. After 75°C for 5 h, for example, the median increases to approximately 260, whereas after 25 h at the same temperature it returns to about 197, as can be seen in Figure 135. Similarly, the median after 100°C for 5 h is approximately 226, before decreasing to about 141 after 25 h.

The substantial overlap among these early distributions indicates that the initial annealing treatments do not simply rescale the optical intensity of the complete hotspot population by a common factor. Some bright sites remain present, while other hotspots become weaker or disappear from the selected sample. This is consistent with the individual-site evolution discussed in Sec. 5.5.2, where appearance, disappearance and reappearance were found to occur simultaneously with the overall reduction in hotspot number.

A clearer modification becomes visible at the highest annealing temperatures. After the 150°C, 5 h treatment, the median luminosity of A1 decreases to approximately 134. A very bright isolated site is still present in this phase, so that the arithmetic mean remains strongly affected by the high-intensity tail and is not representative of the central part of the distribution (See Figure 136a). After extending the 150°C treatment to 25 h, the distribution changes much more markedly: the median decreases to approximately 101. The high-intensity tail is also strongly suppressed, with the maximum measured luminosity reduced to approximately 1.4×10^3 (See Figure 136b).



(a) (b)

Figure 136: Comparison between normalized luminosity distributions of the emission hotspots detected before annealing and after annealing at $150°C$ for 5 and 25 hours.

Sensor B1 The evolution is similarly visible in B1. Before annealing, the selected population contains 62 hotspots with a median luminosity of approximately

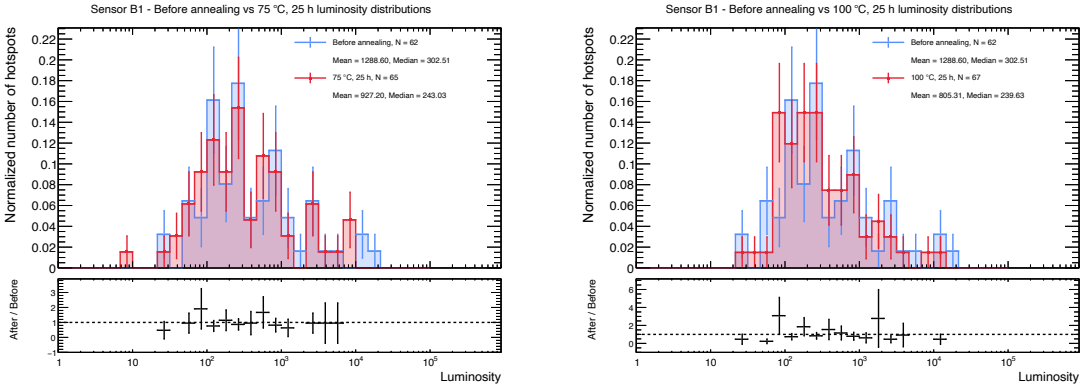
$$L^{\mathrm{B1}}_{\mathrm{median}} \simeq 303, \tag{118}$$

The distribution extends to luminosities above 10^4. During the $75°$C and $100°$C treatments, the distributions remain broad and continue to overlap substantially with the initial one. The medians remain in the approximate range 240–300, showing that no large uniform reduction of the intensity of the surviving population occurs during the earliest stages, as can be seen in Figure 137.

At higher annealing temperatures, however, a progressive displacement towards lower intensities becomes more evident. The B1 median decreases to approximately 223 after $125°$C for 5 h and to about 181 after 25 h. After the final $150°$C treatment, the change is particularly clear: the median is approximately 198 after 5 h and reaches approximately 126 after 25 h. At the same time, the high-luminosity tail is strongly reduced. After the final annealing phase, the brightest selected B1 hotspot has a luminosity of only approximately 703, compared with values extending above 1.6×10^4 before annealing. This results are reported in Figure 138.

Considerations The suppression of the high-intensity tail is also apparent from the relative population of bright hotspots. Before annealing, approximately 15% of the selected A1 hotspots and 21% of the B1 hotspots have luminosities above 10^3. After the final $150°$C, 25 h treatment, this fraction is reduced to approximately 6% for A1, while no B1 hotspot in the selected sample remains above 10^3.

The comparison therefore reveals two related effects of the annealing procedure. First, as established by the counting measurements, the number of detectable emission hotspots decreases. Second, the normalized luminosity distributions show that the population that remains detectable also evolves: at the later annealing stages it becomes increasingly dominated



(a) (b)

Figure 137: Normalized luminosity distributions of the emission hotspots detected at the different annealing stages for sensors B1, compared to the luminosity distribution before annealing

by lower-intensity sites, while the bright tail characteristic of the post-irradiation condition is progressively suppressed.

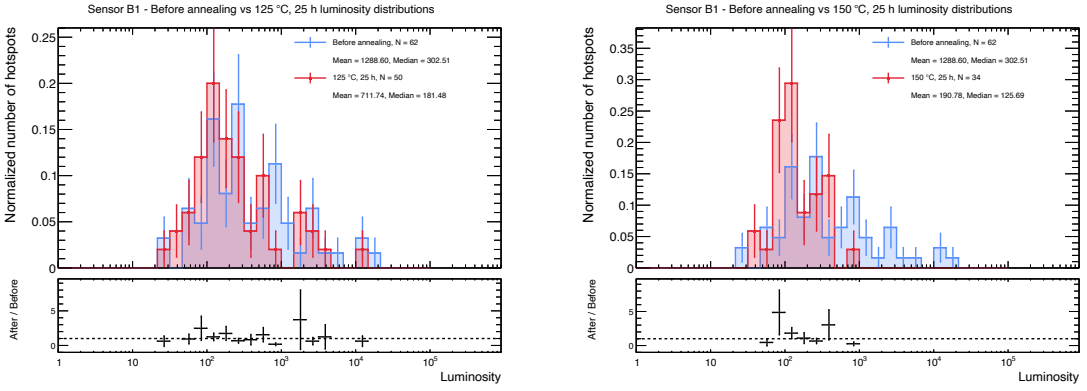
The evolution is nevertheless not strictly monotonic from one annealing phase to the next. Individual bright hotspots can persist through several thermal treatments, and the set of sites entering the luminosity distribution changes as hotspots disappear, appear or reappear. The luminosity histograms at different phases therefore do not represent repeated measurements of an exactly identical population. They describe the distribution of the *detectable and measurable* hotspot population at each stage of the annealing sequence.

This distinction is important for the physical interpretation. A shift of the normalized distribution towards lower luminosity can result both from a decrease in the avalanche-associated optical emission of individual hotspots and from a change in the composition of the detected population. For example, preferential disappearance of highly active sites would modify the normalized distribution even if the intensity of the surviving sites were unchanged. Conversely, a decrease in the luminosity of hotspots that remain spatially identifiable would represent a direct evolution of their local avalanche activity. The population distributions alone cannot fully separate these two effects.

For this reason, the normalized luminosity distributions provide a population-level description of the annealing process and should be considered together with the individual hotspot tracking introduced in Sec. 5.5.2 and discussed in Sec.5.5.4.

5.5.4 Individual hotspots behaviour

The population-level analyses presented in the previous sections show that annealing produces both a progressive reduction in the number of detectable hotspots and a modification of their luminosity distribution. However, these observables alone do not determine how the optical emission associated with a *single* spatial location evolves during the thermal treatment. In particular, the disappearance of a hotspot from the source-identification procedure does not imply



(a) (b)

Figure 138: Normalized luminosity distributions of the emission hotspots detected at the higher temperature annealing stages for sensors B1, compared to the luminosity distribution before annealing

that its optical emission instantaneously vanishes. Its intensity may instead progressively decrease until it falls below the detection threshold. Conversely, a hotspot detected for the first time at an intermediate annealing stage may already have produced a weaker optical signal during the preceding measurements.

To investigate these effects, the luminosity analysis was extended through a *"forced measurement"* procedure. Once a spatial position has been associated with a global hotspot identifier, its luminosity is evaluated at the same position at every annealing phase, independently of whether the hotspot is classified as detected by the source-identification algorithm in that particular phase. The measurement is therefore not restricted to the phases in which the emission site is sufficiently intense to be independently identified.

This approach makes it possible to follow the optical activity of a given spatial region both before its first detection and after its disappearance.

The global-ID population considered in this analysis is consequently larger than the hotspot population detected in any individual annealing phase. Whenever a previously unidentified emission site appears during the annealing sequence, it is assigned a new global identifier and its position is subsequently included in the forced-measurement procedure. For the dataset considered here, this results in a total of 98 global hotspots for sensor A1 and 118 for sensor B1. This definition is particularly useful for newly appearing hotspots, since their assigned coordinates can also be used to inspect the images acquired *before* their first detection. The data presented below refer to the dataset at $T = 20°C$, with the overvoltage $7\ \mathrm{V}$ for A1 and $5\ \mathrm{V}$ for B1.

The evolution of individual hotspots reveals a substantially more heterogeneous picture than that suggested by the population averages.

For a qualitative description of the observed trajectories, the individual hotspot evolutions are grouped into the following categories:

- **Decreasing and disappearing**: hotspots whose luminosity progressively decreases, either

approximately monotonically or with superimposed fluctuations, until the emission falls below the detection threshold;

- **Decreasing**: hotspots whose luminosity decreases throughout the annealing sequence but remains above the detection threshold;
- **Constant**: hotspots whose luminosity fluctuates between successive measurements without exhibiting a clear overall increasing or decreasing trend;
- **Increasing**: hotspots whose luminosity at the end of the investigated annealing sequence is higher than its initial value;
- **Peaks**: hotspots characterized by large transient variations in luminosity during one or more annealing phases, without an otherwise well-defined monotonic trend;
- **Appeared**: hotspots that are initially below the detection threshold and become detectable only at an intermediate stage of the annealing sequence, in some cases subsequently becoming undetectable again.

The different behaviour are discussed individually in the following paragraphs.

Decreasing and disappearing The most frequently observed behaviour is a progressive reduction in luminosity followed by disappearance below the detection threshold. For sensor A1, around 20 hotspots show an approximately monotonic decrease before disappearing, while $\sim$30 exhibit a decreasing evolution accompanied by appreciable fluctuations. Also sensor B1 contains around 30 hotspots each.

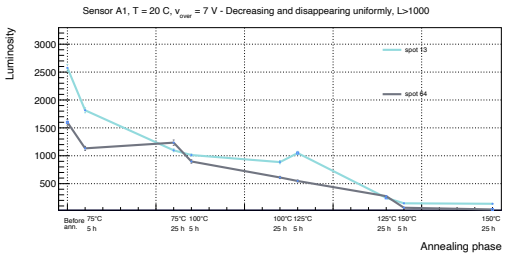
This behaviour is observed over a broad range of initial luminosities, although it is particularly common among relatively weak hotspots. In both sensors, around 40 of the hotspots belonging to this category have luminosities below approximately 400.

Representative high-luminosity examples are shown in Figures 139a and 139b for approximately monotonic trajectories, and in Figures 139c and 139d for trajectories characterized by larger fluctuations before disappearance. Examples from the lower-luminosity population are reported in Figures 139e and 139f.

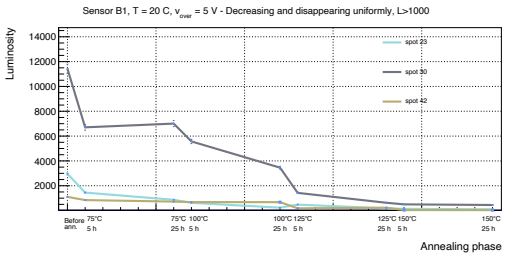
Decreasing A second population consists of hotspots whose luminosity decreases during annealing but remains detectable throughout the investigated sequence. Unlike the disappearing population, this behaviour is observed with comparable frequency over different luminosity ranges.

For sensor A1, 11 hotspots exhibit an approximately monotonic decrease, while 20 show an overall decreasing trend accompanied by fluctuations. For sensor B1, 14 hotspots decrease approximately monotonically and 21 show a fluctuating decreasing behaviour.

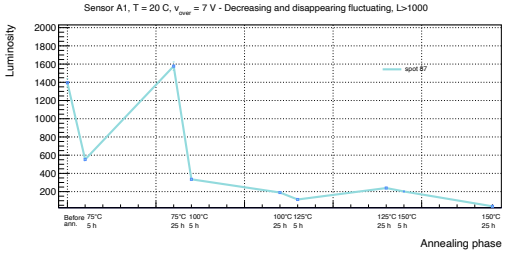
Representative examples of these two types of evolution are shown in Figure 140a, for an approximately monotonic decrease in B1, and in Figure 140b, for a fluctuating decreasing trajectory in A1.



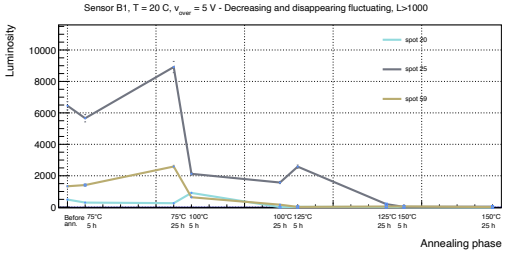
(a) A1, high luminosity: approximately monotonic decrease.



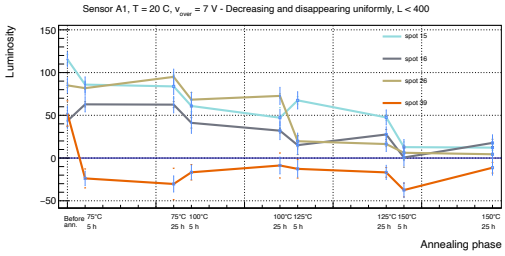
(b) B1, high luminosity: approximately monotonic decrease.



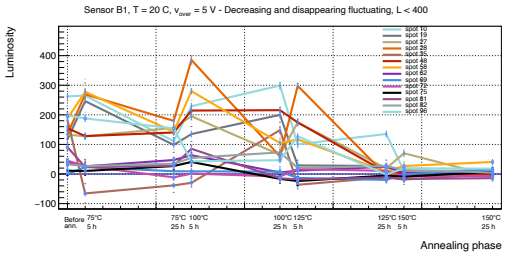
(c) A1, high luminosity: fluctuating decrease.



(d) B1, high luminosity: fluctuating decrease.

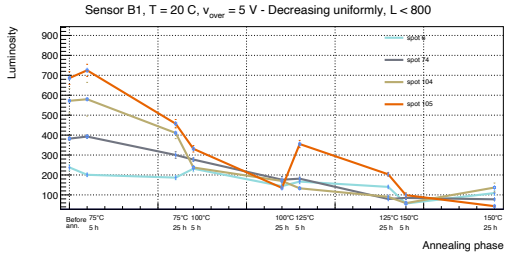


(e) A1, low luminosity: approximately monotonic decrease.

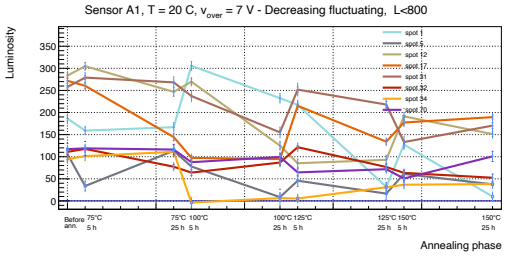


(f) B1, low luminosity: fluctuating decrease.

Figure 139: Representative hotspots whose luminosity decreases during annealing until the emission falls below the detection threshold. Both approximately monotonic and fluctuating trajectories are observed over different luminosity ranges.

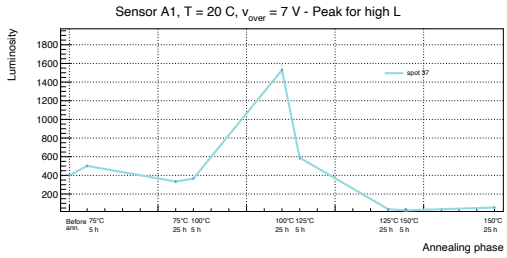


(a) Sensor B1: decreasing evolution with approximately monotonic decrease.

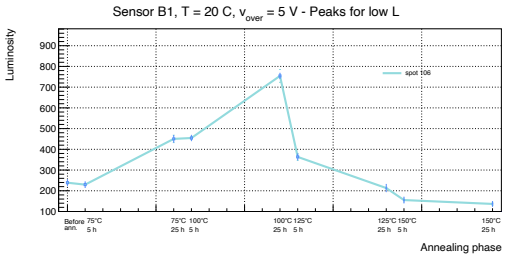


(b) Sensor A1: decreasing evolution with fluctuations.

Figure 140: Representative hotspots showing an overall decrease in luminosity while remaining detectable throughout the investigated annealing sequence.



(a) Example of a high-luminosity hotspot exhibiting a pronounced transient peak.



(b) Example of a lower-luminosity hotspot exhibiting a pronounced transient peak.

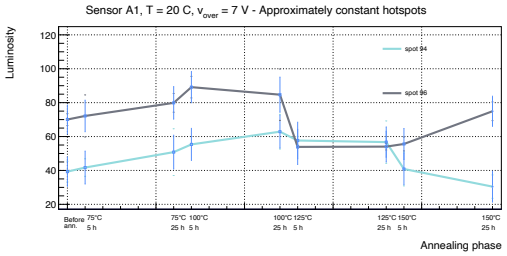
Figure 141: Examples of hotspots characterized by pronounced transient variations in luminosity during the annealing sequence. Such behaviour is observed only for a limited fraction of the studied hotspot population.

Peaks A smaller group of hotspots exhibits pronounced variations in luminosity during individual annealing phases without showing a clear monotonic evolution over the complete sequence. These variations appear as localized peaks superimposed on an otherwise less regular trajectory.

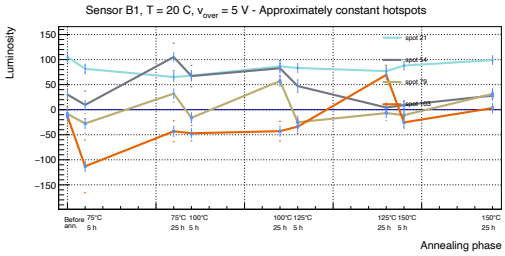
Five hotspots with this behaviour are identified in sensor A1 and nine in sensor B1. This behaviour is not restricted to a specific luminosity range: an example involving a relatively bright hotspot is shown in Figure 141a, while Figure 141b shows a corresponding example at lower luminosity.

Constant A small fraction of the hotspot population remains approximately stable throughout the annealing sequence. These hotspots exhibit fluctuations between successive measurements, but no clear overall increasing or decreasing trend can be identified.

Only five hotspots of this type are observed for each sensor. Representative examples for A1 and B1 are shown in Figures 142a and 142b, respectively.



(a) Representative approximately constant hotspot in sensor A1.



(b) Representative approximately constant hotspot in sensor B1.

Figure 142: Examples of hotspots whose luminosity remains approximately constant, apart from fluctuations, throughout the investigated annealing sequence.

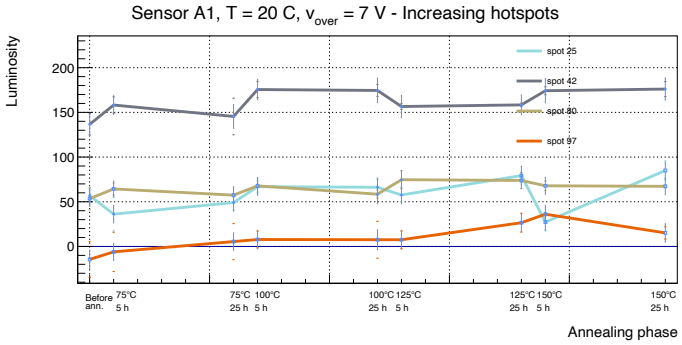
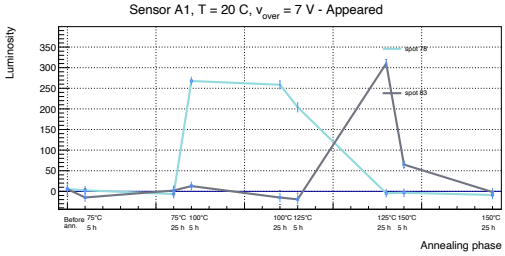


Figure 143: Representative hotspot in sensor A1 whose luminosity at the end of the investigated annealing sequence is higher than its initial value.

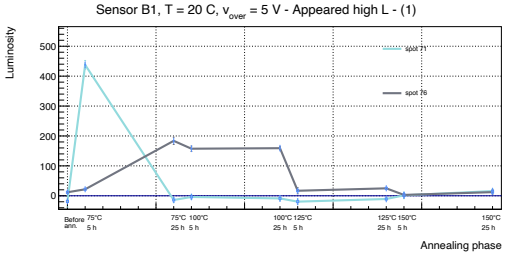
Increasing An increase in luminosity over the complete annealing sequence is considerably less common. Four hotspots with this behaviour are identified in sensor A1 and only one in sensor B1.

A representative A1 example is shown in Figure 143. This behaviour should not necessarily be interpreted as evidence for a distinct class of radiation-induced emission centres. Temporary increases in luminosity are also observed in hotspots that subsequently decrease during later annealing phases. The increasing trajectories observed here may therefore represent transient stages of the evolution rather than a fundamentally different response to annealing. Longer or more intense thermal treatments would be required to determine whether their luminosity would eventually decrease as well.

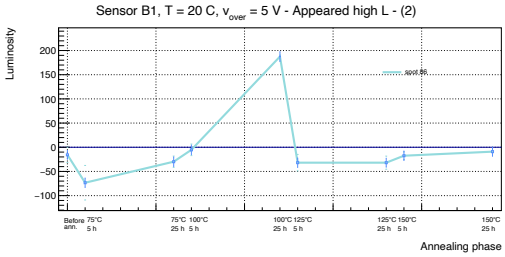
Appeared The most unusual behaviour is observed for a small population of hotspots that are initially below the detection threshold and become detectable only during the annealing sequence. Two such cases are identified for sensor A1 and ten for sensor B1. Some remain detectable after their first appearance, whereas others subsequently fall below the detection threshold again.



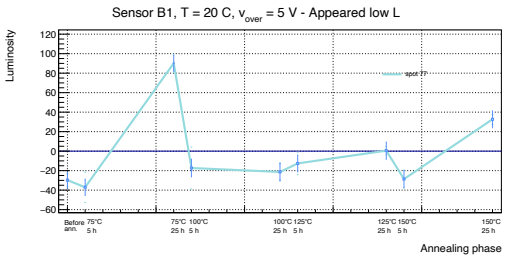
(a) Representative hotspot detected for the first time during annealing.



(b) Representative hotspot detected for the first time during annealing.



(c) Example of a hotspot appearing during annealing and subsequently evolving in luminosity.



(d) Example of a hotspot becoming detectable and subsequently falling below the detection threshold again.

Figure 144: Representative hotspots that become detectable only after one or more annealing stages. The forced-measurement procedure allows the optical signal at the same spatial coordinates to be followed also before the first formal detection.

Representative examples of these different trajectories are shown in Figs. 144a, 144b, 144c and 144d. The forced-measurement procedure is particularly informative in these cases, since it allows the luminosity at the corresponding spatial position to be evaluated also during the phases preceding the first formal detection.

The appearance of these hotspots should not be interpreted as direct evidence for the microscopic formation of new defects during annealing. The EMMI measurement is sensitive only to the spectral range accessible to the camera, while photons emitted outside this range, in particular in the infrared, are not included in the present measurement. Moreover, the classification is based on whether the visible optical signal exceeds the detection threshold. An "appearing" hotspot therefore indicates that the visible emission associated with a given position becomes sufficiently intense to be detected, rather than demonstrating that the underlying microscopic defect itself was created at that annealing stage.

5.6 Discussion

The measurements presented in this chapter provide a characterization of the optical emission of the investigated SiPMs before irradiation, after irradiation and throughout the subsequent annealing sequence. The reference measurements on the non-irradiated device showed a rel-

atively small population of emission hotspots, with a broad distribution of optical intensities. The luminosity of individual hotspots was found to increase approximately exponentially with temperature and to exhibit an approximately quadratic dependence on overvoltage. After irradiation, the number of detectable emission sites increased considerably and the luminosity distribution became broader, extending towards substantially larger values. During annealing, both the number of detected hotspots and the luminosity of the observable population progressively decreased, although the evolution of individual sites was found to be highly heterogeneous, including decreasing, disappearing, reappearing, approximately constant and, more rarely, increasing trajectories. At the same time, the characteristic temperature coefficient λ remained of the same order throughout most of the annealing sequence, while the overvoltage exponent B remained close to two for most investigated conditions.

A final comparison can be made between three observables that independently characterize the evolution of the irradiated sensors during annealing: the dark current, the mean luminosity of the detected hotspots, and the number of detected hotspots. Their evolution throughout the annealing sequence is shown in Figures 145, 146 and 147, respectively. The purpose of this comparison is exclusively to determine whether the three measured quantities exhibit comparable relative variations; no physical inference concerning the origin of the observed changes is made here.

For this purpose, the evolution of each observable X can be expressed through the ratio

$$R_X = \frac{X(150^\circ\mathrm{C},\ 25\ \mathrm{h})}{X(\mathrm{before\ annealing})}. \tag{119}$$

For the dark-current measurements, which were performed at several temperatures, the ratio is evaluated independently at each temperature. For sensor A1, the values obtained between 17°C and 23°C are approximately

$$R_{\mathrm{DC}}^{\mathrm{A1}} = 0.134,\ 0.134,\ 0.136,\ 0.143, \tag{120}$$

respectively. The relative decrease is therefore very similar over the investigated temperature interval, with an average ratio of approximately 0.137. After the complete annealing sequence, the dark current is thus about 14% of its pre-annealing value.

The corresponding ratio for the mean hotspot luminosity of A1 is

$$R_{\mathrm{L}}^{\mathrm{A1}} = \frac{37.41}{396.60} \simeq 0.094, \tag{121}$$

while the number of detected hotspots in the central region changes from 175 to 72, giving

$$R_{\mathrm{N}}^{\mathrm{A1}} = \frac{72}{175} \simeq 0.411. \tag{122}$$

The three observables therefore exhibit the same overall direction of evolution, but with different relative amplitudes. The mean luminosity is reduced to approximately 9% of its initial value and the dark current to approximately 14%, whereas approximately 41% of the initially detected hotspots remain detectable. In other words, the relative reduction of the mean optical

intensity is considerably larger than the reduction in the number of detected sites. The dark-current decrease is also substantially larger than the decrease in hotspot number and, in terms of the final-to-initial ratio, is closer to the evolution of the mean luminosity.

A qualitatively similar comparison can be performed for sensor B1. The mean hotspot luminosity decreases from approximately 493.93 before annealing to 26.88 after the final treatment, corresponding to

$$R_{\mathrm{L}}^{\mathrm{B1}} = \frac{26.88}{493.93} \simeq 0.054. \tag{123}$$

The number of detected hotspots in the central region decreases from 175 to 56, giving

$$R_{\mathrm{N}}^{\mathrm{B1}} = \frac{56}{175} = 0.320. \tag{124}$$

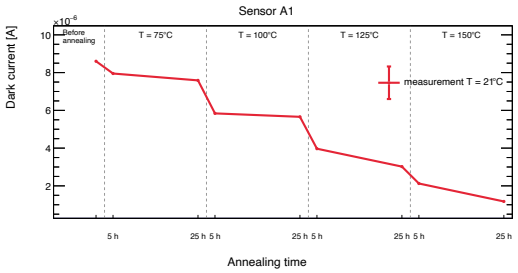
The dark-current gives instead

$$R_{\mathrm{DC}}^{\mathrm{B1}} \simeq 0.138. \tag{125}$$

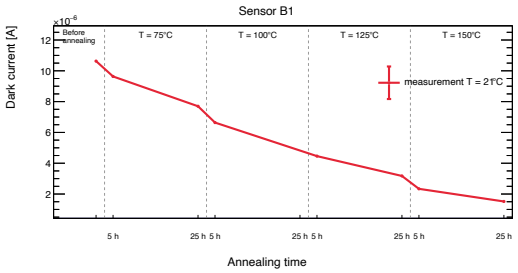
For B1, the available results again show that the three observables decrease substantially during annealing, but not by the same relative amount. After the final treatment, the mean hotspot luminosity is only approximately 5% of its initial value, while the detectable hotspot population is approximately 32% of the initial population. The dark-current value corresponds to approximately 14% of the initial current. As for A1, the decrease in the number of detected hotspots is therefore less pronounced than the decrease observed for the luminosity and dark current.

The evolution at the intermediate annealing stages also shows that the correspondence between the three quantities is not point-by-point. In particular, the mean luminosity can temporarily increase between consecutive measurements, as observed after the first 75°C, 5 h treatment for both sensors, while the dark current and the hotspot population do not exhibit the same increase. The three observables should therefore not be regarded as quantities related by a simple proportionality throughout the annealing sequence.

Nevertheless, when only the initial and final conditions are compared, a common experimental decreasing trend emerges clearly.

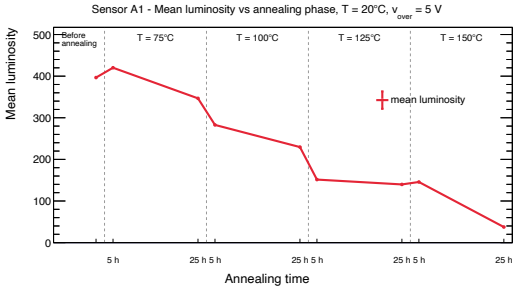


(a) Sensor A1.

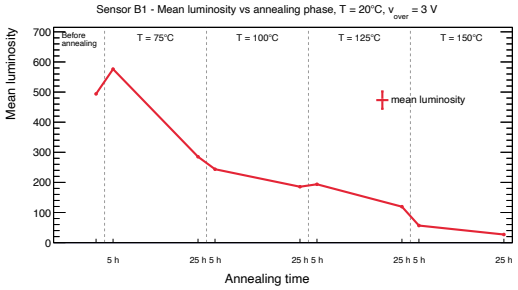


(b) Sensor B1.

Figure 145: Evolution of the dark current throughout the annealing sequence for sensors A1 and B1. The different curves correspond to the investigated operating temperatures. For both sensors, the dark current progressively decreases as the annealing sequence proceeds. The final 150°C, 25 h measurement for sensor B1 is not yet available.



(a) Sensor A1.



(b) Sensor B1.

Figure 146: Evolution of the mean measured luminosity of the emission hotspots throughout the annealing sequence for sensors A1 and B1. The mean is evaluated over the hotspot population included in the luminosity analysis at each annealing stage. Despite some non-monotonic variations between consecutive phases, both sensors show a substantial overall decrease between the pre-annealing condition and the final 150°C, 25 h treatment.

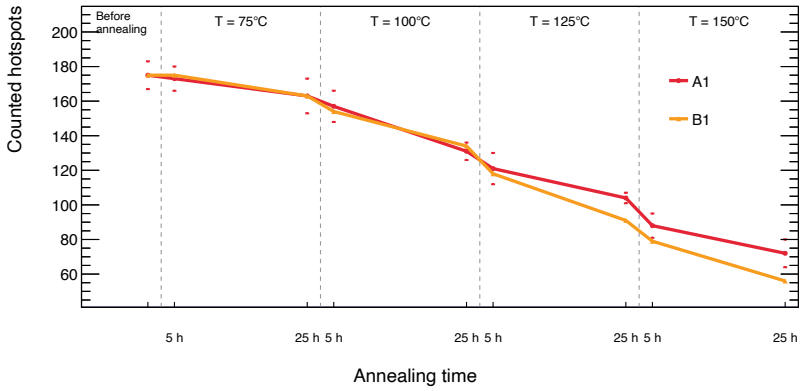


Figure 147: Evolution of the number of detected emission hotspots in the central regions of sensors A1 and B1 throughout the annealing sequence. The hotspot population decreases progressively for both devices, from 175 hotspots before annealing to 72 for A1 and 56 for B1 after the final 150°C, 25 h treatment.

Conclusions

This thesis investigated radiation damage and its evolution during thermal annealing in Silicon Photomultipliers using Emission Microscopy (EMMI), an optical microscope optimized for low light detection. The study was motivated by the possible use of SiPMs as single-photon detectors in the dRICH of the ePIC experiment, where radiation-induced dark noise represents an important issue for long-term detector operation. More generally, the work also addresses the characterization of radiation damage and recovery processes in SiPMs.

EMMI provides a spatially resolved observable that is complementary to conventional electrical measurements, which integrate the response of the whole device. The secondary photons emitted during Geiger avalanches allows localized regions of enhanced avalanche activity, referred to as emission hotspots, to be identified and followed. These hotspots should not be interpreted as direct images of individual microscopic defects: their visibility depends on the local avalanche activity, the operating conditions and the sensitivity of the experimental apparatus, and several microscopic defects may contribute to the signal observed at a single spatial location.

A clear modification of the optical emission pattern was observed after proton irradiation with $1\ MeV$ neutron-equivalent (n_{eq}) fluence of $10^9\ \mathrm{cm}^{-2}$. In the sensor characterized both before and after irradiation, the number of hotspots detected over the complete device increased from 23 ± 2 to 562 ± 29. Similar post-irradiation populations, of the order of 500–600 hotspots, were observed in all four investigated sensors. Irradiation gives rise to a new population of detectable hotspots characterized by a luminosity distribution that differs significantly from that observed before irradiation. Under the same operating conditions, the emission observed before irradiation was confined within the range [20, 480] a.u, whereas after irradiation the distribution broadened across the range [40, 2600] a.u. and developed a pronounced high-luminosity tail. Irradiation therefore affects both the number of detectable hotspots and the intensity distribution of the observed hotspot population.

The dependence of the hotspot luminosity on the SiPM operating conditions was also characterized. Over the investigated temperature range, the measured luminosity follows an exponential dependence on temperature. Irradiation changes the corresponding temperature coefficient, while during the subsequent annealing sequence its characteristic scale remains similar for most of the investigated stages. The dependence on overvoltage is instead well described by a power law with an exponent close to two for the new sensor, the irradiated devices and most annealing conditions. In both analyses, a significant hotspot-to-hotspot dispersion was observed. The fitted parameters should therefore be interpreted as characteristic properties of the measured hotspot populations rather than as universal parameters shared by all emission sites.

The evolution during thermal annealing represents the central result of this work. A progressive reduction in the number of detected emission hotspots was observed in all four sensors. After the complete annealing sequence, the whole-device hotspot population was reduced by 55–60%, independently of the differences between the devices and their intermediate evolution. At the same time, the luminosity distributions shifted towards lower values and their

high-luminosity tails were strongly suppressed. The reduction in luminosity was considerably stronger than the reduction in the number of detectable hotspots. In the central regions, the final mean luminosity corresponded to 9% and 5% of its initial post-irradiation value for the two sensors investigated in detail, while 41% and 32% of the initial hotspot populations remained detectable. The dark current exhibited the same overall decreasing trend, with a reduction stronger than that observed in the hotspot count.

The spatial tracking of individual hotspots provides additional information on this evolution. Although the population-level trend is dominated by a decrease in hotspot number and luminosity, individual emission sites do not all follow the same trajectory. Many hotspots progressively decrease in luminosity and eventually fall below the detection threshold, while others remain detectable, temporarily increase, disappear and later reappear, or become detectable only at an intermediate annealing stage. Appearance and disappearance must therefore be understood in terms of experimental detectability and do not necessarily correspond to the creation or removal of a microscopic defect.

The main added value of this work is the use of EMMI to follow radiation damage and annealing with spatial information that is not available from whole-device electrical measurements alone. Dark current and dark-count rate quantify the integrated electrical degradation of a SiPM, whereas EMMI makes it possible to determine where avalanche-associated emission occurs, how the population of active regions evolves, and how the optical activity of individual locations changes throughout annealing. The combined picture obtained in this thesis therefore shows that irradiation increases both the population of detectable hotspots and the range of their luminosities, while thermal annealing reduces both quantities, with a substantially stronger relative effect on hotspot luminosity than on hotspot number.

Further measurements using in-situ annealing would allow the recovery process to be investigated under thermal and electrical conditions closer to those expected during actual SiPM operation at the EIC. Extending the study to a larger number of sensors and performing a more systematic quantitative correlation between EMMI observables, dark current and dark-count rate would further strengthen the connection between spatially resolved optical measurements and the electrical recovery of the devices. These developments could contribute to the optimization of annealing strategies for the long-term operation of SiPMs in the radiation environment expected for the ePIC dRICH detector.

List of Figures

1 Schematic representing the planned Electron Ion Collider [1]. 4

2 The Standard Model of Particle Physics displays the fundamental particles of matter and the force that govern their interactions [4]. 6

3 On the left, a schematic draw of the behaviour of α_s. Summary of determinations of α_s as a function of the energy scale Q compared to the running of the coupling computed at five loops taking as an input the current PDG average, $\alpha_s(m_z^2) = 0.11800.0009$. [4]. 8

4 Schematic trend of the Cornell potential [5] 9

5 The conventional string breaking: the QCD string spanned between the static quark Q and the static antiquark $\bar{Q}$ breaks due to light $\bar{q}q$ pair creation [7] . . . 9

6 Illustrative representation of the different contributions to the proton mass, emphasizing that only a small fraction originates from the quark masses, while the dominant contribution emerges from QCD dynamics. 10

7 Schematic representation of the different contributions to the nucleon spin, including quark and gluon helicities and orbital angular momentum. The precise separation of these terms depends on the adopted gauge-invariant decomposition. 12

8 Example of lepton-hadron inelastic scattering with an electron and a proton [5]. 14

9 Early structure function measurements from fixed-target electron-proton inelastic scattering at SLAC. Left: measurements of $F_2(x, Q^2)$ showing Bjorken scaling. Right:measurements of $2xF_1/F_2$ showing the Callan-Gross relation. Image from [5]. Adapted from [11] and [12]. 16

10 Proton parton distribution functions obtained from the combined H1 and ZEUS deep-inelastic scattering data in the HERAPDF2.0 NLO analysis at $Q^2 = 10\,\mathrm{GeV}^2$. The distributions xu_v, xd_v, xS and xg are shown as functions of Bjorken-x; the sea and gluon distributions are scaled down by a factor of 20 for visualization. Adapted from [15]. 18

11 Measurements of $F_2^{ep}(x, Q^2)$ at HERA. Results from both the H1 and ZEUS experiments are shown. Image from [5]. 19

12 Schematic representing the collinear approximation, used to obtain 1 dimensional PDFs [17]. 20

13 Schematic of the Inclusive Deep Inelastic Scattering process, for (a) Neutral Current and (b) Charged Current categories. 21

14 Schematic representing the key variables for Wigner distributions and TMDs: $x, \vec{k}_T, \vec{b}_T$ [17]. 22

15 Quantities characterizing the multi-dimensional parton structure of hadrons and the relations between them. **WD**: Wigner Distribution; **TMD**: Trasnverse Momentum Distribution; **IPD**: Impact Parameter Distribution; **GPD**: Generalized Parton Distribution; **GTMD**: Generalized TMD; **GPD**: Generalized Parton Distribution [17]. The variable ξ, known as the *skewness*, quantifies the longitudinal momentum transferred to the nucleon. In a commonly used symmetric parametrization, the active parton enters the hard scattering carrying a longitudinal momentum fraction $x + \xi$ and leaves it with a fraction $x - \xi$. 23

16 Schematic representing the Semi-Inclusive Deep Inelastic Scattering process [3]. $V(q)$ denotes the virtual gauge boson exchanged between the incoming lepton and the nucleon, carrying four-momentum $q = k - k'$. In electromagnetic SIDIS it corresponds to a virtual photon, γ^*, with virtuality $Q^2 = -q^2$, which probes the internal partonic structure of the proton. 24

17 Ratio of the nucleon structure functions measured in iron and deuterium as a function of Bjorken-x. The departure of the ratio from unity, and in particular its decrease in the region $0.3 \lesssim x \lesssim 0.7$, provided the original experimental evidence for what became known as the EMC effect. Adapted from [19]. 27

18 Nuclear modification of parton distributions. The upper panel shows the ratio between the virtual-photon cross section on a nucleus and that on a proton, normalized by the nuclear mass number, highlighting the shadowing, antishadowing, and EMC regions. The lower panel illustrates the expected reduction of the uncertainty on the gluon distribution in a gold nucleus after the inclusion of projected EIC inclusive and charm measurements. Adapted from [18]. 28

19 Illustration of the gluon saturation regime and of one of its possible experimental signatures. (a) shows the expected transition between the dilute and saturated regions in the (x, Q^2) plane. (b) shows the predicted dihadron azimuthal correlation for electron-proton and electron-+-gold collisions, where the suppression of the away-side peak at $\Delta\phi \simeq \pi$ reflects the larger saturation scale of the nuclear target. Adapted from [18]. 30

20 Nuclear modifications in diffractive scattering. The left panel shows ratios of nuclear to proton diffractive structure functions for different nuclei within saturation-based dipole models, while the right panels show corresponding nuclear modifications of diffractive sea-quark and gluon distributions in a leading-twist shadowing description. The different predictions illustrate the strong sensitivity of diffraction to the underlying small-x dynamics. Adapted from [3]. 31

21 Predicted nuclear modification of hadron production in eA relative to ep scattering as a function of the fragmentation variable z. Light hadrons and heavy-flavour mesons exhibit different nuclear modifications, providing complementary sensitivity to the mechanisms governing parton propagation and hadron formation in nuclear matter. Adapted from [18]. 33

22 Examples of the expected EIC contribution to the understanding of the nucleon spin and mass. (a) Gluon helicity distribution obtained from a global analysis: the broader band represents the uncertainty from existing measurements, while the narrower band illustrates the expected improvement after the inclusion of EIC data. Image from [21] (b) Projected sensitivity to the trace-anomaly contribution to the proton mass through near-threshold quarkonium production. Adapted from [3]. 36
23 Schematic representation of the multidimensional partonic structure of a polarized proton for different values of the longitudinal momentum fraction x. The images on the left represent transverse spatial distributions, related to GPDs, while those on the right represent transverse-momentum distributions, related to TMDs. Image from [18]. 37
24 Kinematic coverage in the (x, Q^2) plane expected for deeply virtual Compton scattering at the EIC, compared with existing and planned DVCS measurements at fixed-target and collider experiments. The broad EIC coverage provides access to partonic imaging over a substantially enlarged range of longitudinal momentum fractions and resolution scales. Adapted from [3]. 37
25 Schematic representation of the proton structure probed at progressively higher energies. Increasing the resolving power exposes an increasingly dense population of low-x gluons and quark-antiquark pairs, providing access to the high-parton-density regime investigated at the EIC. Image from [18]. 39
26 Schematic representation of parton propagation through nuclear matter in semi-inclusive deep-inelastic scattering. After the hard interaction, the struck parton propagates through the nuclear medium and develops a parton shower before the final hadronic state is formed. The nuclear environment can modify the shower evolution and the subsequent hadronization process. Image from [3]. . 40
27 A schematic showing how hadrons and the scattered lepton for different $x - Q^2$ are distributed over the detector rapidity coverage. Image from [22]. 41
28 Schematic view of the ePIC Central Detector, showing the principal tracking, particle-identification and calorimeter subsystems surrounding the interaction region. Image from [3]. 45
29 Overview of the ePIC tracking system, showing the silicon vertex tracker, MPGD barrel and endcap layers, and the AC-LGAD timing detectors. Reproduced from [22]. 46
30 Structure of the ePIC Barrel Imaging Calorimeter. On the left, a 3D rendering of the Barrel Imaging Calorimeter with its 48 sectors. The central drawing shows the structure of a single sector, featuring interleaved Pb/ScFi layers and slots for trays holding AstroPix chips, followed by the Pb/ScFi bulk section. On the right, a zoomed view of the first radially layers is presented. Image from [22]. 48
31 Schematic layout of the ePIC far-detector systems along the outgoing electron and hadron beam lines. Reproduced from [22]. 49

32 Example of the velocity-dependent detector interactions used for PID devices. The left panel denotes the rate of energy loss of charged particle passing through matter. The right panel denotes the angle of Cherenkov radiation. Image from [3]. 50
33 Simulated kaon-identification efficiencies and purities as a function of hadron momentum in three pseudorapidity regions of ePIC. The corresponding pion, proton and other-particle misidentification contributions are also shown. Image from [22]. 56
34 Projected impact of EIC measurements on pion and kaon fragmentation functions. Existing global determinations are compared with fits including simulated EIC SIDIS measurements. Image from [26]. 57
35 Expected uncertainties on Collins asymmetries for identified π^+ production in representative (x, Q^2) intervals, shown as a function of z and transverse hadron momentum. The projections assume an integrated luminosity of $10\ \mathrm{fb}^{-1}$. Reproduced from [22]. 59
36 Reconstructed $K\pi$ invariant-mass distributions for D^0 candidates in different rapidity intervals. Results using particle identification alone are compared with those obtained after combining PID with topological selections in simulated ep and eA collisions. Image from [22]. 60
37 Differential distribution in momentum transfer $|t|$ for coherent $\phi \to K^+K^-$ electroproduction in electron-gold collisions. The reconstructed distribution is obtained from full ePIC simulation. Reproduced from [22]. 61
38 Charged-hadron momentum as a function of pseudorapidity for simulated $18 \times 275\ \mathrm{GeV}$ ep collisions. The contours indicate the approximate 3σ separation regions provided by the pfRICH, hpDIRC, dRICH and time-of-flight systems for π/K, K/p and e/π identification. Reproduced from Figure 3.90 of [22]. 63
39 Simulated performance of the ePIC time-of-flight system. Left: BTOF $1/\beta$ as a function of momentum (p) in the simulation with PYTHIA DIS events. Right: the approximate upper momentum limits for 3σ particle separation as a function of pseudorapidity for the barrel and forward AC-LGAD timing layers. Image from [22]. 66
40 Reconstructed particle mass as a function of momentum in the forward dRICH for pions, kaons and protons (left) and for pions and electrons (right). Results obtained from the aerogel and gas radiators illustrate their complementary momentum coverage. The simulation corresponds to $3.2 < \eta < 3.5$. Reproduced from [22]. 67
41 Baseline layout of the ePIC proximity-focusing RICH in the electron-going endcap. The aerogel radiator, expansion volume, mirror elements and HRPPD sensor plane are indicated. Reproduced from [22]. 68
42 Baseline geometry of the ePIC high-performance DIRC. Cherenkov photons generated in the fused-silica bars are transported by total internal reflection toward the focusing optics, expansion prisms and photon-sensor arrays. Image from [22]. 70
43 Geometry of the ePIC AC-LGAD time-of-flight systems, with the cylindrical barrel timing layer and the pixelated forward endcap detector. Reproduced from [22]. 71

44 dRICH detector model with major components highlighted [31]. 73
45 (Top left) Example of aerogel tile produced with the dRICH baseline refractive index $n = 1.026$ and the reference dimension of BELLE-II. (Top-right) Mid-size CFRP mirror demonstrator with optimized core structure. (Bottom) Photo-spectrometer with compact high-pressure chamber for gas transparency measurements. [31]. 74
46 The PDU concept (left) and preliminary realization (right). [31]. 76
47 Magnetic-field map of the ePIC detector for the $1.7\,\mathrm{T}$ solenoid configuration, with the envelopes of the PID detectors superimposed. In the dRICH region the field is non-uniform and reaches values of approximately $0.3\,\mathrm{T}$–$0.6\,\mathrm{T}$. Image from [22]. 78
48 Cherenkov-ring images reconstructed with the SiPM-based dRICH prototype. The left panel shows the response obtained with the aerogel radiator, while the right panel shows the simultaneous aerogel and C_2F_6 ring patterns. Adapted from [38]. 80
49 Contributions to the dRICH single-photon angular resolution for the aerogel radiator (left) and gas radiator (right). The finite sensor pixel size is one of the contributions controlled through the choice of a $3 \times 3\,\mathrm{mm}^2$ SiPM channel. Reproduced from [22]. 82
50 SiPM-based dRICH readout prototype used during the 2024 beam tests. The modular structure allows large photon-detection surfaces to be assembled from compact SiPM arrays. Image taken from [38]. 83
51 Schematic view of the ePIC detector with the forward dRICH volume highlighted. Its integration close to the central solenoidal magnet motivates the use of a magnetic-field-insensitive photon-detection technology. Adapted from [24]. . . 84
52 Timing suppression of the SiPM dark-count background in a dRICH prototype. Left: hit-time distribution showing the correlated signal and approximately uniform DCR background. Right: reconstructed partial Cherenkov ring after application of a $\pm 5\,\mathrm{ns}$ timing selection. Image from [24]. 85
53 Transverse map of the expected 1-MeV-neutron-equivalent fluence per $1\,\mathrm{fb}^{-1}$ of integrated luminosity at the longitudinal position of the dRICH photon detector for highest-energy ep collisions. The dashed boundaries indicate the radial region occupied by the photon-detection system. Reproduced from [22]. . . . 86
54 SiPM radiation-damage and annealing studies for the dRICH. Left: dark-current evolution during repeated irradiation and annealing cycles. Right: projected DCR evolution for the most exposed PDU as a function of delivered luminosity, showing the cases without annealing, the residual annealing limit and a representative operating scenario. Reproduced from [22]. 88
55 Representation of diffusion processes within a p-n junction with different doping concentrations. 90
56 A p-n junction under two different polarisation conditions. On the left, forward bias; on the right, reverse bias. 91

57 (a) The n-on-p SPAD structure, more sensitive to the red light and (b) the p-on-n structure, more sensitive in the blue and near UV. Pictures from [39]. 92
58 The absorption coefficient of silicon in cm^{-1} as a function of the wavelength. The data is graphed on a y log scale. Image from [42]. 93
59 Schematic cross-section of two possible single pixels of blue sensitive SiPMs. The poly-silicon quenching resistor R_q and the Al-contact line are isolated from the silicon via a SiO_2 layer. Image from [40]. 94
60 (a) Electrical model of a SPAD with its quenching resistor (R_q). (b) Current pulse produced by the SPAD following photon detection. Pictures from [24]. 95
61 (a) Picture of Hamamatsu S13360 SiPMs. (b) Multiple waveform in persistence of a SiPM. It is visible the response of the device when multiple (4) SPADs are firing. The cross-talk (CT) and afterpulse (AP) contributions are highlighted (see Sec. 3.3). Image from [24]. 97
62 Photon Detection Efficiency ($\%$) vs wavelength of Hamamatsu S13360 SiPM series with pixel pitch of (a) $25\mu m$ (b) $50\mu m$ and (c) $75\mu m$. Plots from [43]. . . . 98
63 SiPMs' characteristics varying with the operating voltage for pixel pitch of (a) $25\ \mu m$ (b) $50\ \mu m$ and (c) $75\ \mu m$. Although increasing the operating voltage improves the photon detection efficiency and time resolution, it also increases the dark count and crosstalk at the same time (see Sec. 3.3), so an optimum operating voltage must be selected to match the application. Plots from [43]. . 99
64 Single-photon time resolution (SPTR) as a function of the SiPM overvoltage for devices from different manufacturers. (a) Measured SPTR, including the contributions from the laser pulse width and electronic noise. (b) Intrinsic SPTR obtained after quadratically subtracting the electronic-noise contribution and the 42 ps FWHM laser pulse width. Figure from [39]. 102
65 Representation of the SiPM output signal of the different kinds of noise observables: primary events, prompt crosstalk, afterpulsing and delayed crosstalk events. Image from [45]. 104
66 DCR for NUV-HD FBK-SF (triangular markers) and NUV-HD-LF (circular markers) SiPMs as a function of function of 1/T and overvoltage. Image taken from [46]. 105
67 Different types of crosstalk observable in analog SiPMs. Reprint from [44]. . . . 107
68 Dark current as a function of the bias voltage of a reverse biased Hamamatsu S13360-3050VS SiPM. Image taken from [24]. 108
69 Dark-count rate of individual cells of a digital SiPM as a function of the accumulated number of incident protons. The step-like variations show the localized character of radiation-induced damage and are compatible with the creation of strongly generation-active cluster defects. Figure from [49]. 111
70 Schematic representation of defect levels inside the silicon band gap and of their principal macroscopic effects: generation-recombination, carrier trapping and modification of the effective doping concentration. Figure from [40]. 112

71 Dark current measured at -30°C for FBK NUV-HD-CHK SiPMs exposed to different neutron-equivalent fluences. The points represent the average measured over several sensors, while the rectangles indicate the RMS dispersion of the sample. Figure from [41]. 113
72 Single-photoelectron charge signals measured at $V - V_{\mathrm{BD}} = 3\ \mathrm{V}$ before irradiation and after neutron irradiation to $\Phi_{\mathrm{eq}} = 10^{9}\ \mathrm{cm}^{-2}$. The resolved photoelectron structure visible before irradiation is lost after irradiation at room temperature. Figure from [50]. 114
73 Average dark-count occupancy as a function of bias voltage for SiPMs exposed to different neutron-equivalent fluences and operated at -30°C. The increasing occupancy at high fluence reduces the fraction of fully recovered cells available for photon detection. Figure from [40]. 116
74 Response to $400\ \mathrm{nm}$ laser pulses of Hamamatsu S13360-1350CS SiPMs before irradiation and after neutron-equivalent fluences of 10^{11}, 10^{12} and $10^{13}\ n_{\mathrm{eq}}/\mathrm{cm}^{2}$. Measurements are shown at room temperature (top) and at -30°C (bottom), illustrating the reduction of the radiation-induced background obtained by cooling the sensors. Figure from [51]. 118
75 Use of timing information for DCR suppression in the dRICH SiPM readout. Left: time distribution of the detected hits, showing the signal coincidence peak, residual calibration structures and the random DCR background; the $\pm 5\ \mathrm{ns}$ selection window is indicated. Right: accumulated Cherenkov-ring hits after application of the timing selection. Figure from [24]. 119
76 Concept of the ePIC dRICH photodetector unit and preliminary realization. The compact module integrates the SiPM sensor plane, front-end readout electronics and the thermal infrastructure required for low-temperature operation and in-situ annealing. Figure from [31]. 120
77 Dark-current measurements for FBK (left) and Hamamatsu (right) SiPMs exposed to different proton fluences and measured at different stages before and after high-temperature oven annealing. Points represent the average over several sensors, while rectangles indicate the RMS dispersion of the sample. Figure from [41]. 122
78 Left: sensor temperature monitored with an infrared camera during forward-bias heating. Right: schematic of the experimental setup developed for the in-situ annealing of SiPMs. The sensors are heated through forward-bias current, while their temperature is monitored with an infrared thermal camera and controlled through an Arduino-based feedback system. Figures from [35]. 123
79 Fraction of residual radiation-induced damage as a function of the integrated annealing time for successive temperature steps at 100°C, 125°C, 150°C and 175°C. Measurements from individual SiPMs are shown together with the averaged evolution for each annealing temperature. Figure from [35]. 124
80 Example of the secondary-photon emission measured from a SiPM operated in Geiger mode. The emission extends over a broad wavelength range and depends on the operating overvoltage. Figure from [53]. 126

81 Spatially resolved optical emission from a SiPM operated above break-down. Localized regions with enhanced emission intensity are visible within a subset of the microcells and are identified as emission hotspots. Figure from [55]. . . 127
82 Emission-microscopy images of a Hamamatsu SiPM before and after irradiation. Upper left: intrinsic hotspots observed on a new SiPM at 5 V overvoltage. Upper right: intrinsic hotspots observed on a new SiPM at 7 V overvoltage. Bottom: irradiated sensor with radiation-induced hotspots. Figure from [56]. 128
83 Diagram illustrating the workflow followed for SiPM studies. First stage on new sensors, second stage after irradiation, followed by eight successive annealing stages with progressively increasing temperature and duration. 132
84 Picture of the EMMI (EMission MIcroscopy) experimental set-up, taken at INFN's Silicon Laboratory in Bologna. 133
85 Board containing the four studied SiPMs: two sensors of type S13360-3050 and two sensors of type S13360-3075. 133
86 (a) Picture of the Quantalux 2.1 MP sCMOS camera, model CC215MU. (b) CC215MU quantum efficiency. 134
87 Two exemplary IV scans performed on the SiPMs under test, before annealing, at $T = 21°C$. 135
88 Example of the automatic transverse positioning procedure. A reference TSV is identified in the optical image and its centre is used to determine the position of the SiPM in the plane parallel to its surface. 137
89 Example of the autofocus procedure. The image sharpness, quantified through the variance of the Laplacian, is measured as a function of the vertical position z. A quadratic fit is performed around the maximum, whose position defines the optimal focal distance. 137
90 Schematic representing the data acquisition steps to obtain the final image used for the data analysis. 139
91 Image before clean-up. 141
92 Image after removal of column dependent bias. 142
93 Image after removal of column dependent bias and hot/cold pixels. 142
94 Optical image of central region of sensor A1, obtained with an external light illuminating the SiPM. Geometrical structures of the device, such as SPADs and the Through-Silicon Vias (TSVs) are visible. 143
95 Example images of the rotation+shifting correction procedure. In the left images, the reference images from which the coordinates are calculated. In the middle images, evident geometrical displacement caused by the removal and repositioning of the SiPM board. In the right images, correction made by the rotation+shifting procedure. 144
96 Example of application of the described code for individuating the defects coordinates on the detector; in the picture is presented a new SiPM (before irradiation) in the best visibility conditions (maximum temperature and overvoltage). 146

97 Example of inspection of an individual hotspot, from sensor A1 before annealing at $T = 21°C$ and $v_{over} = 5\,V$. The red circle represents the integration radius of 20 pixels. 147
98 Possible critical situation of close hotspots. Increasing the integration radius of all the defects could lead to an overestimation of the luminosity, due to integration over an area containing a different hotspot. Hotspot like 66-162-163 (in the local ID coordinates) are excluded by the isolation cut. 150
99 Plot of the luminosity distribution (non normalized) with logarithmic binning. In green the whole sample, in blue the isolated defects. 151
100 Plot of the normalized luminosity distribution with logarithmic binning. In green the whole sample, in blue the isolated defects. In the lower pad, the ratio between isolated and non-isolated defects. 151
101 Luminosity of each hotspot recalculated for different values of the integration radius, while keeping its position fixed. Hotspot with maximum luminosity between 0 and 50 a.u. 152
102 Luminosity of each hotspot recalculated for different values of the integration radius, while keeping its position fixed. (Top) Hotspot with maximum luminosity between 50 and 100 a.u., (Middle) Hotspot with maximum luminosity between 100 and 200 a.u., (Bottom) Hotspot with maximum luminosity between 200 and 300 a.u. 153
103 Luminosity of each hotspot recalculated for different values of the integration radius, while keeping its position fixed. (Top) Hotspot with maximum luminosity between 300 and 500 a.u., (Middle) Hotspot with maximum luminosity between 500 and 1000 a.u., (Bottom) Hotspot with maximum luminosity greater than 1000 a.u. 154
104 Plots showing the luminosities at different overvoltages, for two examples hotspots, measured with the three integration radii as L_{16}, L_{20}, and L_{24}. The systematic uncertainty associated to L_{20} is showed. 155
105 Plots showing the luminosities at different temperatures, for two examples hotspots, measured with the three integration radii as L_{16}, L_{20}, and L_{24}. The systematic uncertainty associated to L_{20} is showed. 156
106 Example picture of the central region (area = $2.4 \times 1.3\,mm^2$) of the investigated detectors. 158
107 Counting measurements carried out on sensor A1 (type S13360-3050VS), considering the whole sensor area $3 \times 3mm^2$. (a) Superposition of the emission image and the optical image. (b) Emission image with the detected hotspots (23) circled in red. 159
108 Distribution of the measured optical intensity of the emission hotspots detected in the non-irradiated A1 sensor at $T = 20°$C. The distribution is dominated by relatively low-intensity sites and presents a tail toward larger values. 160
109 The luminosity as a function of temperature keeping the SiPM overvoltage fixed. Two example hotspots in the images. Both working conditions are represented: $V_{over} = 7\,V$ in red and $V_{over} = 9\,V$ in blue. 160

110 The extracted λ values, for both overvoltages, fitted with a constant for each operating condition. $V_{over} = 7\,V$ in red and $V_{over} = 9\,V$ in blue. 161
111 The luminosity as a function of overvoltage keeping the SiPM temperature fixed. Two example hotspots in the images. 162
112 The extracted B values from the empirical power law $L_{\text{spot}} = A\,V_{\text{over}}^{B}$. A red dashed line for B=2 is plotted for reference. 162
113 Representative examples of the optical intensity measured for individual emission hotspots as a function of temperature. The curves represent fits with the function $L_i(T) = A_i e^{\lambda_i T}$. The two examples illustrate the approximately exponential increase observed in both irradiated sensors. 165
114 Values of the temperature coefficient λ_i obtained from the independent exponential fits for sensor A1 before annealing. Up: λ_i values for each emission hotspots; the mean is represented as a dashed line. Down: Distributions of the λ_i values; it shows the significant hotspot-to-hotspot dispersion around the mean value. 167
115 Values of the temperature coefficient λ_i obtained from the independent exponential fits for sensor B1 before annealing. Up: λ_i values for each emission hotspots; the mean is represented as a dashed line. Down: Distributions of the λ_i values; it shows the significant hotspot-to-hotspot dispersion around the mean value. 168
116 Luminosity versus temperature for the same hotspot (global ID 4), the different annealing phases are represented with different colours. 169
117 Average $\langle\lambda\rangle$ coefficient obtained from the individual hotspot fits as a function of the annealing phase for sensors A1 and B1, plotted alongside the individual fit values for each hotspot (in blue). Missing phases are due to weather-related issues that occurred during data collection, which rendered the data inaccessible. 170
118 Behaviour of the dark current at different temperatures. Left: Halving temperature for S13360-3050VS in the new, proton-irradiated, neutron-irradiated, and annealed ($150°C$ for 150 hours) states; the image from [59] is reported here for reference across a broader temperature range. Right: dark current for sensor A1 at each annealing stages. 171
119 Representative examples of the optical intensity measured from individual emission hotspots as a function of overvoltage at $T = 20°$C before annealing. The curves represent the individual fits with $L_i = A_i V_{\text{over}}^{B_i}$. 173
120 Values of the B coefficient obtained from the independent exponential fits for sensor A1 before annealing. Left: B values for each emission hotspots; the mean is represented as a dashed line. Right: Distributions of the λ_i values; it shows the significant hotspot-to-hotspot dispersion around the mean value. 174
121 Values of the coefficient B obtained from the independent exponential fits for sensor B1 before annealing. Left: B values for each emission hotspots; the mean is represented as a dashed line. Right: Distributions of the B values; it shows the significant hotspot-to-hotspot dispersion around the mean value. 175

122 Exponents B_i extracted from the individual overvoltage fits for the non-irradiated sensor A1 at $T = 20°$C. The horizontal reference at $B = 2$ indicates an exactly quadratic dependence. 177
123 Evolution of the exponent B during the annealing sequence for sensors A1 and B1. The highlighted points represent the mean exponent $\langle B \rangle$ obtained from the individual hotspot fits at each phase, while the blue points show the corresponding individual B_i values. The brackets represent the RMS of each hotspot population. 179
124 Spatial distribution of the emission hotspots detected after irradiation and before the annealing sequence for sensors A1, A2, B1 and B2. The detected light-emitting sites are visible in the images, obtained superimposing the EMMI image and the optical image. Irradiation produces a dense population of emission hotspots distributed over the imaged sensor area. 182
125 Direct comparison of the spatial distribution of the detected emission hotspots before and after irradiation for sensor A1. The sensor is placed under the same conditions of $T = 20°C$ and $V_{over} = 5\,V$. The comparison illustrates the strong increase in the number of spatially localized emission sites following irradiation. 183
126 Distribution of the optical intensity measured for the emission hotspots of sensor A1 after irradiation at $T = 20°$C and $V_{\mathrm{over}} = 7$ V. Logarithmic binning is used because of the broad range of measured intensities. 184
127 Normalized distributions of the optical intensity of the emission hotspots measured in sensor A1 before irradiation and after irradiation, before annealing, at $T = 20°$C and $V_{\mathrm{over}} = 7$ V. Only sites detected at the selected operating condition are considered. Both distributions are normalized and represented using logarithmic binning, allowing their shapes to be compared independently of the different number of hotspots. 185
128 Graphical representation of Tab. 10: number of hotspots at different annealing stages for the four sensors. 187
129 Whole-device EMMI images of sensor A1 before annealing and after the final treatment at $150°$C for 25 h. The reduction in the number of detected emission hotspots is visible over the complete sensor area. 188
130 Whole-device EMMI images of sensor A2 before annealing and after the final treatment at $150°$C for 25 h. A substantial reduction of the detected hotspot population is observed after the complete annealing sequence. 189
131 Whole-device EMMI images of sensor B1 before annealing and after the final treatment at $150°$C for 25 h. The final image contains considerably fewer detected emission hotspots than the irradiated sensor before annealing. 189
132 Whole-device EMMI images of sensor B2 before annealing and after the final treatment at $150°$C for 25 h. As for the other devices, the annealing sequence produces a marked reduction in the number of detected emission hotspots. . 190
133 Graphical representation of Tabs. 11- 12: number of hotspots in the cental region at different annealing stages for the two sensors A1 and B1. 191

134 Luminosity distributions of the emission hotspots detected before annealing for sensors A1 and B1. The distributions are evaluated at fixed operating conditions for each sensor and shown using logarithmic binning. 194
135 Normalized luminosity distributions of the emission hotspots detected at the different annealing stages for sensors A1. The distributions are evaluated at fixed operating conditions for each sensor and shown using logarithmic binning. Normalization allows the evolution of the distribution shape to be compared independently of the reduction in the number of detected hotspots. 195
136 Comparison between normalized luminosity distributions of the emission hotspots detected before annealing and after annealing at $150°C$ for 5 and 25 hours. . 196
137 Normalized luminosity distributions of the emission hotspots detected at the different annealing stages for sensors B1, compared to the luminosity distribution before annealing 197
138 Normalized luminosity distributions of the emission hotspots detected at the higher temperature annealing stages for sensors B1, compared to the luminosity distribution before annealing 198
139 Representative hotspots whose luminosity decreases during annealing until the emission falls below the detection threshold. Both approximately monotonic and fluctuating trajectories are observed over different luminosity ranges. . . . 200
140 Representative hotspots showing an overall decrease in luminosity while remaining detectable throughout the investigated annealing sequence. 201
141 Examples of hotspots characterized by pronounced transient variations in luminosity during the annealing sequence. Such behaviour is observed only for a limited fraction of the studied hotspot population. 201
142 Examples of hotspots whose luminosity remains approximately constant, apart from fluctuations, throughout the investigated annealing sequence. 202
143 Representative hotspot in sensor A1 whose luminosity at the end of the investigated annealing sequence is higher than its initial value. 202
144 Representative hotspots that become detectable only after one or more annealing stages. The forced-measurement procedure allows the optical signal at the same spatial coordinates to be followed also before the first formal detection. . 203
145 Evolution of the dark current throughout the annealing sequence for sensors A1 and B1. The different curves correspond to the investigated operating temperatures. For both sensors, the dark current progressively decreases as the annealing sequence proceeds. The final $150°$C, 25 h measurement for sensor B1 is not yet available. 206
146 Evolution of the mean measured luminosity of the emission hotspots throughout the annealing sequence for sensors A1 and B1. The mean is evaluated over the hotspot population included in the luminosity analysis at each annealing stage. Despite some non-monotonic variations between consecutive phases, both sensors show a substantial overall decrease between the pre-annealing condition and the final $150°$C, 25 h treatment. 206

147 Evolution of the number of detected emission hotspots in the central regions of sensors A1 and B1 throughout the annealing sequence. The hotspot population decreases progressively for both devices, from 175 hotspots before annealing to 72 for A1 and 56 for B1 after the final 150°C, 25 h treatment. 207

List of Tables

1 Representative charged-hadron PID coverage of the current ePIC baseline [22]. The momentum values correspond to the π/K performance benchmarks of the individual subsystems. 64

2 Main characteristics of the SiPMs investigated in this work. The PDE values refer to a peak sensitivity at $450\,\mathrm{nm}$, with a spectral response range of 270–$900\,\mathrm{nm}$. The gain values are measured at $V_{\mathrm{over}} = 3\,\mathrm{V}$, while the breakdown voltage is specified at $T = 25^{\circ}\mathrm{C}$. 131

3 Main optical components of the EMMI system and their corresponding product codes. 134

4 The progressive annealing steps adopted in this study. 136

5 Number of background frames (N), signal frames (M), and repeated background-subtracted measurements (K) used for the different EMMI acquisition configurations. The same acquisition parameters were maintained throughout the different annealing stages. 140

6 Number and spatial density of hotspots measured in the new device. 158

7 Average temperature coefficient obtained from the individual hotspot fits at the different annealing stages and corresponding temperature interval ΔT_2. Missing phases are due to weather-related issues that occurred during data collection, which rendered the data-taking problematic. 169

8 Mean values of the exponent B obtained from the individual hotspot fits at $T = 20^{\circ}$C for sensors A1 and B1 throughout the annealing sequence. The uncertainty associated with $\langle B \rangle$ is statistical, while $\mathrm{RMS}(B)$ quantifies the hotspot-to-hotspot dispersion of the individual fitted exponents. 178

9 Number of emission hotspots detected in the four irradiated sensors after irradiation with a proton beam of approximately $140\,\mathrm{MeV}$, with a neutron-equivalent radiation level of $10^9\ 1\ MeV\ n_{eq}/cm^2$, before the annealing sequence. 181

10 Number of emission hotspots detected over the complete device area at the different annealing phases for sensors A1, A2, B1 and B2. 186

11 Evolution of the hotspot population in the central region of sensor A1 at $T = 20^{\circ}$C and $V_{\mathrm{over}} = 5$ V. The “reappeared” column specifies the most recent phase in which the corresponding hotspots had disappeared. 192

12 Evolution of the hotspot population in the central region of sensor B1 at $T = 20^{\circ}$C and $V_{\mathrm{over}} = 3$ V. The “reappeared” column specifies the most recent phase in which the corresponding hotspots had disappeared. 193

References

[1] A. Accardi, J. L. Albacete, M. Anselmino, N. Armesto, E. C. Aschenauer, A. Bacchetta, D. Boer, W. K. Brooks, T. Burton, N. B. Chang, W. T. Deng, A. Deshpande, M. Diehl, A. Dumitru, R. Dupré, R. Ent, S. Fazio, H. Gao, V. Guzey, H. Hakobyan, Y. Hao, D. Hasch, R. Holt, T. Horn, M. Huang, A. Hutton, C. Hyde, J. Jalilian-Marian, S. Klein, B. Kopeliovich, Y. Kovchegov, K. Kumar, K. Kumerički, M. A. C. Lamont, T. Lappi, J. H. Lee, Y. Lee, E. M. Levin, F. L. Lin, V. Litvinenko, T. W. Ludlam, C. Marquet, Z. E. Meziani, R. McKeown, A. Metz, R. Milner, V. S. Morozov, A. H. Mueller, B. Müller, D. Müller, P. Nadel-Turonski, H. Paukkunen, A. Prokudin, V. Ptitsyn, X. Qian, J. W. Qiu, M. Ramsey-Musolf, T. Roser, F. Sabatié, R. Sassot, G. Schnell, P. Schweitzer, E. Sichtermann, M. Stratmann, M. Strikman, M. Sullivan, S. Taneja, T. Toll, D. Trbojevic, T. Ullrich, R. Venugopalan, S. Vigdor, W. Vogelsang, C. Weiss, B. W. Xiao, F. Yuan, Y. H. Zhang, and L. Zheng. Electron ion collider: The next qcd frontier - understanding the glue that binds us all, 2014.

[2] S. Ozaki and Thomas Roser. Relativistic heavy ion collider, its construction and upgrade. *Progress of Theoretical and Experimental Physics*, 2015:3A102–0, 01 2015.

[3] Abdul Khalek et al. Science requirements and detector concepts for the electron-ion collider. *Nuclear Physics A*, 1026:122447, October 2022.

[4] F. Takahashi et al. Review of Particle Physics. *Int. J. Mod. Phys. A*, 41(22):2630011, 2026.

[5] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[6] Bikash Sinha Sourav Sarkar, Helmut Satz. *The Physics of the Quark Gluon Plasma*. Springer, 2010.

[7] M. N. Chernodub. Qcd string breaking in strong magnetic field. *Modern Physics Letters A*, 29(30):1450162, 09 2014.

[8] Xiangdong Ji. Gauge-invariant decomposition of nucleon spin. *Physical Review Letters*, 78(4):610–613, January 1997.

[9] Alessandro Bacchetta, Markus Diehl, Klaus Goeke, Andreas Metz, Piet J. Mulders, and Marc Schlegel. Semi-inclusive deep inelastic scattering at small transverse momentum. *Journal of High Energy Physics*, 2007(02):093, feb 2007.

[10] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

[11] Jerome I. Friedman and Henry W. Kendall. Deep inelastic electron scattering. *Ann. Rev. Nucl. Part. Sci.*, 22:203–254, 1972.

[12] A. Bodek, M. Breidenbach, D. L. Dubin, J. E. Elias, J. I. Friedman, H. W. Kendall, J. S. Poucher, E. M. Riordan, M. R. Sogard, D. H. Coward, and D. J. Sherden. Experimental studies of the neutron and proton electromagnetic structure functions. , 20(7):1471–1552, October 1979.

[13] Chris Llewellyn Smith. From concrete quarks to qcd: a personal perspective. *The European Physical Journal H*, 48(1):13, 2023.

[14] L.W. Whitlow E.M. Riordan S. Dasu S. Rock A. Bodek. Precise measurements of the proton and deuteron structure functions from a global analysis of the slac deep inelastic electron scattering cross sections. *Physics Letters B Volume 282, Issues 3–4*, 1992.

[15] H. Abramowicz et al. Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and qcd analysis of hera data. *European Physical Journal C*, 75:580, 2015.

[16] J. Beringer, J. F. Arguin, R. M. Barnett, K. Copic, O. Dahl, D. E. Groom, C. J. Lin, J. Lys, H. Murayama, C. G. Wohl, W. M. Yao, P. A. Zyla, C. Amsler, M. Antonelli, D. M. Asner, H. Baer, H. R. Band, T. Basaglia, C. W. Bauer, J. J. Beatty, V. I. Belousov, E. Bergren, G. Bernardi, W. Bertl, S. Bethke, H. Bichsel, O. Biebel, E. Blucher, S. Blusk, G. Brooijmans, O. Buchmueller, R. N. Cahn, M. Carena, A. Ceccucci, D. Chakraborty, M. C. Chen, R. S. Chivukula, G. Cowan, G. D'Ambrosio, T. Damour, D. de Florian, A. de Gouvêa, T. DeGrand, P. de Jong, G. Dissertori, B. Dobrescu, M. Doser, M. Drees, D. A. Edwards, S. Eidelman, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, B. Foster, T. K. Gaisser, L. Garren, H. J. Gerber, G. Gerbier, T. Gherghetta, S. Golwala, M. Goodman, C. Grab, A. V. Gritsan, J. F. Grivaz, M. Grünewald, A. Gurtu, T. Gutsche, H. E. Haber, K. Hagiwara, C. Hagmann, C. Hanhart, S. Hashimoto, K. G. Hayes, M. Heffner, B. Heltsley, J. J. Hernández-Rey, K. Hikasa, A. Höcker, J. Holder, A. Holtkamp, J. Huston, J. D. Jackson, K. F. Johnson, T. Junk, D. Karlen, D. Kirkby, S. R. Klein, E. Klempt, R. V. Kowalewski, F. Krauss, M. Kreps, B. Krusche, Yu. V. Kuyanov, Y. Kwon, O. Lahav, J. Laiho, P. Langacker, A. Liddle, Z. Ligeti, T. M. Liss, L. Littenberg, K. S. Lugovsky, S. B. Lugovsky, T. Mannel, A. V. Manohar, W. J. Marciano, A. D. Martin, A. Masoni, J. Matthews, D. Milstead, R. Miquel, K. Mönig, F. Moortgat, K. Nakamura, M. Narain, P. Nason, S. Navas, M. Neubert, P. Nevski, Y. Nir, K. A. Olive, L. Pape, J. Parsons, C. Patrignani, J. A. Peacock, S. T. Petcov, A. Piepke, A. Pomarol, G. Punzi, A. Quadt, S. Raby, G. Raffelt, B. N. Ratcliff, P. Richardson, S. Roesler, S. Rolli, A. Romaniouk, L. J. Rosenberg, J. L. Rosner, C. T. Sachrajda, Y. Sakai, G. P. Salam, S. Sarkar, F. Sauli, O. Schneider, K. Scholberg, D. Scott, W. G. Seligman, M. H. Shaevitz, S. R. Sharpe, M. Silari, T. Sjöstrand, P. Skands, J. G. Smith, G. F. Smoot, S. Spanier, H. Spieler, A. Stahl, T. Stanev, S. L. Stone, T. Sumiyoshi, M. J. Syphers, F. Takahashi, M. Tanabashi, J. Terning, M. Titov, N. P. Tkachenko, N. A. Törnqvist, D. Tovey, G. Valencia, K. van Bibber, G. Venanzoni, M. G. Vincter, P. Vogel, A. Vogt, W. Walkowiak, C. W. Walter, D. R. Ward, T. Watari, G. Weiglein, E. J. Weinberg, L. R. Wiencke, L. Wolfenstein, J. Womersley, C. L. Woody, R. L. Workman, A. Yamamoto, G. P. Zeller, O. V. Zenin, J. Zhang, R. Y. Zhu, G. Harper, V. S. Lugovsky, and P. Schaffner. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012.

[17] Renaud Boussarie, Matthias Burkardt, Martha Constantinou, William Detmold, Markus Ebert, Michael Engelhardt, Sean Fleming, Leonard Gamberg, Xiangdong Ji, Zhong-Bo Kang, Christopher Lee, Keh-Fei Liu, Simonetta Liuti, Thomas Mehen, Andreas Metz, John Negele, Daniel Pitonyak, Alexei Prokudin, Jian-Wei Qiu, Abha Rajan, Marc Schlegel, Phiala Shana-

han, Peter Schweitzer, Iain W. Stewart, Andrey Tarasov, Raju Venugopalan, Ivan Vitev, Feng Yuan, and Yong Zhao. Tmd handbook, 2023.

[18] US Department of Energy (USDOE). A new era of discovery: The 2023 long range plan for nuclear science. Technical report, US Department of Energy (USDOE), Washington, DC (United States). Office of Science, 10 2023.

[19] J. J. Aubert et al. The ratio of the nucleon structure functions f_2^n for iron and deuterium. *Physics Letters B*, 123:275–278, 1983.

[20] Francois Gelis, Edmond Iancu, Jamal Jalilian-Marian, and Raju Venugopalan. The color glass condensate. *Annual Review of Nuclear and Particle Science*, 60:463–489, 2010.

[21] Ignacio Borsa, Gonzalo Lucero, Rodolfo Sassot, Elke C. Aschenauer, and Ana S. Nunes. Revisiting helicity parton distributions at a future electron-ion collider. *Phys. Rev. D*, 102:094018, Nov 2020.

[22] Silvia Dalla Torre and John Haggerty. *The ePIC Detector Preliminary Technical Design Report - Version3.1.* Zenodo, January 2026.

[23] Chandradoy Chatterjee. Particle Identification with the ePIC detector at the EIC. *PoS*, DIS2024:266, 2025.

[24] Luigi Pio Rignanese, Pietro Antonioli, Roberto Preghenella, and Eugenio Scapparone. Sipms and examples of applications for low light detection in particle and astroparticle physics. *La Rivista del Nuovo Cimento*, 2024.

[25] Christian Lippmann. Particle identification. *Nucl. Instrum. Meth. A*, 666:148–172, 2012.

[26] E. C. Aschenauer I. Borsa R. Sassot C. Van Hulse. Semi-inclusive deep-inelastic scattering, parton distributions and fragmentation functions at a future electronion collider. *Phys. Rev. D 99 (9) 094004*, 2019.

[27] A.V. Lyashenko, J. Agarwala, A. Asaturyan, M. Aviles, B. Azmoun, C. Chatterjee, S.M. Clarke, S. Cwik, C.J. Hamel, Y. Jin, A. Kiselev, M.J. Minot, B. Page, M.A. Popecki, M. Purschke, S. Stoll, and C. Woody. Hrppd photosensors for rich detectors with a high resolution timing capability. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1082:170964, February 2026.

[28] M. Davis, G. Stage, A. Borjigin, S. Beringer, N. Lynch, S. Nakarmi, C. Galmes Altafulla, A. Drumm, A. Molnar, A. Tiernan, S. M. Mazza, H. Sadrozinski, B. A. Schumm, A. Seiden, F. MartinezMckinney, and T. Shin. Characterization of the first full-size production for epic tof layers. 2025.

[29] Jochen Schwiening. Dirc, a new type of particle identification system for babar. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 408(1):211–220, May 1998.

[30] Chandradoy Chatterjee. R&d efforts in cherenkov imaging technologies for particle identification in future experiments. *Nuclear Instruments and Methods in Physics Research Section A*, 1088:171516, 2026.

[31] M. Contalbrigo et al. The ePIC dual-radiator RICH detector. *Nucl. Instrum. Meth. A*, 1088:171537, 2026.

[32] Rocco Liotino, Eugenio Nappi, Nicola Nicassio, and Giacomo Volpe. *Aerogel characterization for RICH applications*. Nuclear Instruments and Methods in Physics Research Section A, 2026.

[33] S. Vallarino et al. Prototype of a dual-radiator rich detector for the electron–ion collider. *Nuclear Instruments and Methods in Physics Research Section A*, 2024.

[34] F. Cossio et al. A mixed-signal asic for the drich detector of the epic experiment at the eic. *Nuclear Instruments and Methods in Physics Research Section A*, 2024.

[35] B. Rajesh Achari, N. Agrawal, M. Alexeev, C. Alice, P. Antonioli, C. Baldanza, L. Barion, A. Calivà, M. Capua, M. Chiosso, et al. The sipm photodetector of the epic dual-radiator rich at the eic: overview and beam test results. *Journal of Instrumentation*, 20(08):C08025, 2025.

[36] Alexander Paramonov. FELIX: the Detector Interface for the ATLAS Experiment at CERN. *EPJ Web Conf.*, 251:04006, 2021.

[37] C. Rossi et al. Online ai-based distributed data reduction for the dual-radiator rich detector in the epic experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1087:171416, 2026.

[38] A. Paladino et al. The sipm readout plane for the epic-drich photo-detector at the eic: overview and beam test results. *JINST 21 C01031*, 2026.

[39] Stefan Gundacker and Arjan Heering. The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector. *Physics in Medicine and Biology*, 2020.

[40] E. Garutti and Yu. Musienko. Radiation damage of sipms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926:69–84, May 2019.

[41] Roberto Preghenella, Maxim Alexeev, Pietro Antonioli, Casimiro Baldanza, Luca Barion, Michela Chiosso, Marco Contalbrigo, Fabio Cossio, Silvia Dalla Torre, Manuel Dioniso Da Rocha Rolo, Giulio Dellacasa, Davide Falchieri, Roberto Malaguti, Marco Mignone, Luigi Pio Rignanese, Nicola Rubini, and Simone Vallarino. Study of radiation effects on sipm for an optical readout system for the eic dual-radiator rich. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1056:168578, 2023.

[42] N. M. Miskovsky, P. H. Cutler, A. Mayer, B. L. Weiss, Brian Willis, T. E. Sullivan, and P. B. Lerner. Nanoscale devices for rectification of high frequency radiation from the infrared through the visible: A new approach. *Journal of Nanotechnology*, 2012(1):512379, 2012.

[43] Hamamatsu Photonics K.K. *MPPC (Multi-Pixel Photon Counter) S13360 Series*, 2024. Datasheet.

[44] Claudio Piemonte and Alberto Gola. Overview on the main parameters and technology of modern silicon photomultipliers. *Nuclear Instruments and Methods in Physics Research Section A*, 926, 2019.

[45] Fabio Acerbi and Stefan Gundacker. Understanding and simulating sipms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926:16–35, 2019. Silicon Photomultipliers: Technology, Characterisation and Applications.

[46] F. Acerbi et al. Cryogenic characterization of fbk hd near-uv sensitive sipms. *IEEE Transactions Electron Devices*, 64 (2), 2016.

[47] Acerbi Fabio, Ferri Alessandro, Zappala Gaetano, Paternoster Giovanni, Picciotto Antonino, Gola Alberto, Zorzi Nicola, and Piemonte Claudio. Nuv silicon photomultipliers with high detection efficiency and reduced delayed correlated-noise. *IEEE Transactions on Nuclear Science*, 62(3), 2015.

[48] G. Lindstrom. Radiation damage in silicon detectors. *Nuclear Instruments and Methods in Physics Research*, 2003.

[49] M. Yu. Barnyakov, T. Frach, S. A. Kononov, I. Kuyanov, and V. Prisekin. Radiation hardness test of the philips digital photon counter with proton beam. *Nuclear Instruments and Methods in Physics Research Section A*, 824:83–84, 2016.

[50] T. Tsang, T. Rao, S. Stoll, and C. Woody. Neutron radiation damage and recovery studies of sipms. *Journal of Instrumentation*, 11(12):P12002, 2016.

[51] M. Calvi, P. Carniti, C. Gotti, C. Matteuzzi, and G. Pessina. Single photon detection with sipms irradiated up to 1014 cm2 1-mev-equivalent neutron fluence. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 922:243–249, April 2019.

[52] Yi Qiang, Carl Zorn, Fernando Barbosa, and Elton Smith. Radiation hardness tests of sipms for the jlab hall d barrel calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 698:234–241, 2013.

[53] J.B. McLaughlin et al. Characterisation of SiPM Photon Emission in the Dark. *Sensors*, 21(17):5947, 2021.

[54] K. Raymond et al. Stimulated secondary emission of single-photon avalanche diodes. *IEEE Transactions on Electron Devices 71 6871–6879*, 2024.

[55] E. Engelmann, E. Popova, and S. Vinogradov. Spatially Resolved Dark Count Rate of SiPMs. *Eur. Phys. J. C*, 78(11):971, 2018.

[56] Paladino et al. Probing radiation damage in sipms with emission microscopy. *JINST 21 C05019*, 2026.

[57] Francesco Tommasino et al. A new facility for proton radiobiology at the trento proton therapy centre: Design and implementation. *Physica Medica*, 58:99–106, 2019.

[58] Larry Bradley, Brigitta M. Sipőcz, T. P. Robitaille, E. J. Tollerud, Zé Vinícius, Christoph Deil, Kyle Barbary, Tom J. Wilson, Ivo Busko, Axel Donath, Hans Moritz Günther, Mihai Cara, P. L. Lim, Sebastian Meßlinger, Simon Conseil, Michael Droettboom, K. Azalee Bostroem, E. M. Bray, Lars Andersen Bratholm, Zach Burnett, William Jamieson, Adam Ginsburg, Dan Taranu, Geert Barentsen, Matthew W. Craig, Brett M. Morris, Marshall Perrin, and Shivangee Rathi. Photutils, April 2026.

[59] Bologna ePIC group. Radiation damage studies and annealing of sipm for the epic-drich detector at the eic. *In preparation*.

Acknowledgement

Ringrazio il mio relatore, il Professor Pietro Antonioli, per essersi reso disponibile a seguirmi per questo lavoro di tesi magistrale.

Ringrazio il Dott. Roberto Preghenella per avermi offerto la possibilità di lavorare su un progetto così nuovo ed entusiasmante, e per avermi seguita in questi mesi di lavoro. Sono cresciuta moltissimo, come persona e come ”fisico”, sia per le abilità tecniche che ho acquisito per affrontare le sfide del progetto, sia per le opportunità di presentare a conferenze e di partecipare al *test-beam* al CERN.

Ringrazio il Dott. Antonio Paladino per il supporto costante durante questo percorso, per la disponibilità nel chiarire gli aspetti tecnici del lavoro e per il prezioso contributo nella preparazione delle diverse scadenze.

Un grazie va anche al Dott. Luigi Pio Rignanese, per avermi dato le basi per fare questo lavoro durante la tesi triennale e per avermi aiutato a prepararmi per gli orali di dottorato.

Un grazie speciale va anche al Professor Andrea Alici, che purtroppo non è più con noi oggi, per essere stato il mio primo mentore; porterò sempre con me la sua disponibilità e la sua gentilezza nei modi di fare, per ricordarmi che si può essere persone di grande successo rimanendo profondamente buoni umanamente.