Ultra-Fast Detectors for Particle Physics Applications



E. Giulio Villani

St. Cross College
University of Oxford

Supervised by

Prof. Daniela Bortoletto

Submitted: January 2024

This thesis is submitted to the Department of Physics, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 100,000 words including appendices, bibliography, footnotes, tables, and equations, and has fewer than 155 figures.

E. Giulio Villani
Submitted: January 2024

Dedication

Vorrei dedicare questa tesi a mia madre, Romana, e alle mie figlie, Amaris e Maria.

Abstract

This thesis describes the simulation, fabrication, and characterisation of Low Gain Avalanche Detectors (LGAD) fabricated using a commercial process from Teledyne e2v silicon foundry.

Two runs of device fabrication are described. The first one investigates the performances achievable using the available technology and compares test results with simulations. The second run addresses the specifications of the timing sensor for the High Granularity Timing Detector (HGTD) in the ATLAS detector at the High Luminosity Hadron Collider (HL-LHC).

The characteristics of the LGAD are predicted using TCAD simulations and include full process simulations and electrical performances. Test results include the comparison between the obtained doping profile using Secondary Ion Mass Spectrometry (SIMS) and Spread Resistance Measurement Systems (SRMS) with TCAD simulations and measured and simulated gain when using an optical injection of charge. Effects of gain suppression when a high density of charge is injected are predicted. Besides the intended use as fast timing sensors for High Energy Physics applications (HEP) initial tests on the fabricated LGAD show their potential use as detectors for space radiation and clinical dosimetry.

Issues related to the higher granularity of LGAD sensors, for future HEP applications, are also discussed. A novel approach, which consists of a Graded Epitaxial LGAD (GELGAD) is presented and details of fabrication and TCAD simulations, which show the achievable gain using two different flavours of GELGAD, are presented.

Finally, details of the ongoing projects, which result from the work presented in this thesis, are described along with suggestions for future investigations on aspects of the physics of impact ionisation, upon which LGADs are based.

Acknowledgements

I would like to thank my supervisors Prof. Daniela Bortoletto and Prof. Steve McMahon, for their help and support throughout my time as a PhD student at the University of Oxford and Rutherford Appleton Laboratory.

My deep gratitude to all staff members of the Oxford Physics Microstructure Detector (OPMD) laboratory at the University of Oxford, who provided crucial help and technical support throughout my entire time spent on this project. This includes Richard Plackett for his continuous technical and professional coordination support in all aspects of OPMD activities, Dennis Yborde, Maryam Akhtar, Gale Lockwood, and Alex Knight for their crucial wire bonding work in the clean room. A big thank you to Daniel Hynds, Daniel Weatherill, and Dan Woods for their technical help and Martin Gazi for his crucial technical support. Also a special thank you to Daniel Hynds and Daniel Weatherill for several inspiring and, more importantly, pleasant and friendly discussions on a rather diversified set of different topics.

I would also like to thank staff members of Teledyne e2v foundry for their technical support and contribution to the realization of this project.

Also a big thank you to Prof. Anatoly Rozenfeld, the Director of the Centre of Medical Radiation Physics (CMRP), University of Wollongong, Australia, for several inspiring technical suggestions and his student Jay Archer, for investigating the use of the designed LGAD for use in space and medical dosimetry applications. I would also like to thank Prof. Dave Newbold, the former director of the Particle Physics Department at Rutherford Appleton Laboratory, UK, for his encouragement and sponsorship of my PhD. All my colleagues at RAL PPD, in one way or another, contributed and helped me complete this work.

Finally, I would like to thank Rossana. No doubt I would not have been able to complete this work without her love and support.

This project was financially supported by the Science and Technology Facilities Council UK (STFC) [grant number LGAD ST/T002298/1].

Table of contents

List of figures xv

List of tables xxi

1 Introduction 1

1.1 Detectors for space and time measurement . . . 1
1.2 Timing information at HL-LHC . . . 2
1.3 Thesis Outline . . . 13

2 Silicon Detectors 15

2.1 Introduction to silicon as a semiconductor . . . 16
2.1.1 Crystal Structure . . . 16
2.1.2 Energy Bands . . . 17
2.1.3 Carrier Concentration . . . 19
2.1.4 Generation and recombination . . . 21
2.1.5 Carrier Transport and Basic Semiconductor Equations . . . 22
2.2 The *pn* junction . . . 24
2.3 Principles of detector operation . . . 27
2.3.1 Energy deposition in silicon . . . 27
2.4 Semiconductor Fabrication . . . 32
2.4.1 Crystal growth . . . 33
2.4.2 Oxidation . . . 33
2.4.3 Photolithography . . . 35
2.4.4 Etching . . . 36
2.4.5 Doping . . . 36
2.4.6 Thermal processing . . . 37
2.4.7 Metallization . . . 38

3 Timing detectors **39**
3.1 Timing Detector Introduction . . . 40
3.1.1 Ionisation . . . 41
3.1.2 Distortion . . . 42
3.1.3 Jitter . . . 43
3.1.4 TDC . . . 44
3.2 Sensor with internal gain . . . 44
3.3 Impact Ionisation . . . 45
3.4 The LGAD device . . . 50
3.4.1 LGAD operating principle . . . 51
3.4.2 LGAD implementation . . . 54
3.4.3 LGAD noise . . . 55
3.5 Readout electronics . . . 57
3.5.1 Transimpedance amplifier . . . 57
3.5.2 Charge sensitive amplifiers . . . 59

4 Design and Simulations of Low Gain Avalanche Diode as Timing Detector **61**
4.1 Introduction to Sentaurus TCAD . . . 62
4.2 LGAD Device design . . . 66
4.2.1 Run 1 LGAD design description . . . 66
4.2.2 Run 2 LGAD design description . . . 69
4.3 LGAD simulation . . . 71
4.3.1 Process simulation . . . 72
4.3.2 Process simulation and Secondary Ion Mass Spectrometry (SIMS) 78
4.3.3 Electrical Simulation . . . 81
4.4 Meshing strategy . . . 101
4.5 Validity of the local model . . . 102

5 Fabrication and Performances of Low Gain Avalanche Detectors **105**
5.1 Description of fabricated devices . . . 106
5.1.1 Run 1 devices . . . 106
5.1.2 Run 2 devices . . . 108
5.2 Experimental setups . . . 110
5.2.1 IV and CV test setup . . . 110
5.2.2 Laser source test setup . . . 111
5.2.3 Radioactive source test setup . . . 114
5.3 Readout design and description . . . 116

5.4 Test results: IV and CV characteristics . . . 119
5.4.1 IV plots . . . 120
5.4.2 CV plots . . . 122
5.5 Test results: Time Resolution and Gain . . . 126
5.5.1 Gain . . . 126
5.5.2 Timing . . . 130
5.6 LGAD for dosimetry . . . 133
5.6.1 Gain results . . . 133
5.7 Discussion . . . 135

6 High Granularity LGAD **137**
6.1 LGAD segmentation techniques . . . 138
6.2 Graded Epitaxial Layer LGAD (GELGAD) . . . 141
6.2.1 Graded Epitaxy LGAD structure . . . 141
6.3 GELGAD simulation . . . 143
6.3.1 GELGAD process simulation . . . 143
6.3.2 GELGAD DC electrical simulations . . . 145
6.3.3 GELGAD charge collection simulations . . . 149
6.3.4 GELGAD implementation . . . 157

7 Conclusions and future work **163**
7.1 Conclusions . . . 163
7.2 Future works . . . 165

References **169**

Appendix A Impact Ionisation Models **179**

Appendix B List of Publications and Presentations **183**

List of figures

1.1 Schematic view of the LHC at CERN . . . 2
1.2 Schematic view of the ATLAS detector at CERN . . . 4
1.3 Schematic view of the ATLAS Inner Tracker (ITk) at CERN . . . 5
1.4 Schematic layout of the ITk for the HL-LHC phase of ATLAS detector . . . 5
1.5 Interaction time of pp vertexes at LHC . . . 6
1.6 ATLAS simulation of spatial events at HL-LHC . . . 6
1.7 HL-LHC local pileup vertex densities . . . 7
1.8 ITk resolution of the longitudinal track impact parameter z_0 as a function of η 8
1.9 ATLAS plot of the truth interactions in a single HL-LHC bunch crossing . . . 8
1.10 Position of the HGTD within the ATLAS detector . . . 9
1.11 Global view of the HGTD . . . 10
1.12 Radiation levels of the HGTD . . . 12

2.1 Diamond and zinc blende lattice . . . 16
2.2 Tetrahedron bond and schematic two-dimensional representation . . . 17
2.3 Silicon band diagram . . . 18
2.4 Energy band structure of metals, semiconductors, and insulators . . . 18
2.5 Schematic of recombination processes . . . 22
2.6 Schematic of a *pn* junction in equilibrium . . . 24
2.7 Cross-section of a silicon detector . . . 27
2.8 Mass stopping power for positive muons μ^+ in copper as a function of $\beta\gamma =$ p/Mc . . . 28
2.9 Most probable energy loss in silicon, scaled to the mean loss of a minimum ionising particle (MIP) . . . 29
2.10 Diagram of generation of electrons and holes by absorption of photons . . . 30
2.11 Photon interaction cross-sections in silicon as a function of photon energy . 31
2.12 Example of a silicon process flow for detector fabrication . . . 32
2.13 Shallow Trench Isolation (STI) fabrication process . . . 34

List of figures

3.1 Schematic diagram of a timing detector . . . 40
3.2 Effects of the time walk . . . 41
3.3 Effects of the weighting field for wide and narrow sensor . . . 43
3.4 Effects of jitter . . . 43
3.5 Schematic of the electron impact ionisation process . . . 46
3.6 Ionisation coefficient α_e in silicon . . . 48
3.7 Ionisation coefficient α_h in silicon . . . 48
3.8 PIN diode cross-section . . . 51
3.9 LGAD diode cross-section . . . 51
3.10 PIN cross-section at full depletion . . . 52
3.11 LGAD cross-section at full depletion . . . 53
3.12 An example of the pulse signal in PIN and LGAD . . . 53
3.13 LGAD Signal to Noise ratio (SNR) . . . 56
3.14 Transimpedance Amplifier (TIA) schematic . . . 57
3.15 Combined jitter plot of sensor and amplifier . . . 59

4.1 Production cycles in semiconductor industry . . . 62
4.2 Schematic diagram of a device simulator . . . 63
4.3 Example of Synopsys TCAD simulation flow diagram . . . 64
4.4 An example of TCAD simulation flow in Workbench . . . 65
4.5 Layout of Run 1 fabrication . . . 66
4.6 Layout details of Run 1 fabrication . . . 67
4.7 Run 1 LGAD cell cross-section . . . 67
4.8 Layout of Run 2 fabrication . . . 69
4.9 Run 2 LGAD cell cross-section . . . 70
4.10 Full TCAD simulation flowchart . . . 72
4.11 SPROCESS: Epitaxial growth . . . 73
4.12 SPROCESS: Oxide thermal growth . . . 73
4.13 SPROCESS: Resist deposition . . . 73
4.14 SPROCESS: JTE implantation . . . 74
4.15 SPROCESS: Thermal re-oxidation . . . 75
4.16 SPROCESS: Drive-in for gain layer and p-stop . . . 75
4.17 SPROCESS: LGAD structure after the final thermal process . . . 76
4.18 SPROCESS: Final structure obtained from SPROCESS . . . 77
4.19 SPROCESS: Detailed image of the LGAD structure . . . 77
4.20 SPROCESS: Detailed image of the PIN structure . . . 77
4.21 TCAD and SIMS doping comparison . . . 78

4.22 TCAD and SIMS doping comparison - detailed . . . 79
4.23 TCAD and SIMS net doping comparison . . . 79
4.24 Cross-section of LGAD and PIN from TCAD . . . 81
4.25 JTE optimization for breakdown . . . 85
4.26 TCAD estimated breakdown voltage of LGAD . . . 86
4.27 Cross-section of wafer 2 (WF2) LGAD near breakdown . . . 87
4.28 Cross-section of wafer 21 (WF21) LGAD near breakdown . . . 87
4.29 TCAD Transient simulation setup for charge collection . . . 88
4.30 Meshing for an optical charge injection . . . 90
4.31 Transient charge density from an optical charge injection in LGAD . . . 91
4.32 Transient charge density from an optical charge injection in a PIN diode . . . 92
4.33 Transient electric field and ionisation coefficient from an optical charge injection in LGAD . . . 92
4.34 Transient ionisation coefficient α_e during an optical charge injection in LGAD 93
4.35 Transient current from an optical charge injection in PIN diode and LGAD . 93
4.36 Maximum collected charge for PIN and LGAD . . . 94
4.37 Meshing for MIP charge injection . . . 97
4.38 Transient charge density from a MIP charge injection in LGAD . . . 97
4.39 Transient electric field and ionisation coefficient during a MIP charge injection in LGAD . . . 98
4.40 Transient ionisation coefficient α_e during MIP charge injection in LGAD . . 98
4.41 Maximum ionisation coefficient α_e vs. time during IR and MIP charge injection in LGAD . . . 99
4.42 Simulation results of gain normalized to the percentage of V_{bd} using MIP and IR charge injection for an LGAD . . . 100
4.43 The electric field F, dF/dx, α_e at 95% of V_{bd} and doping for an LGAD . . 103

5.1 Run 1 wafer . . . 106
5.2 Laser cut dice from Run 1 fabrication . . . 107
5.3 Run 1 die on a Compact OPMD LGAD Amplifier (COLA) board . . . 108
5.4 Run 2 wafer . . . 109
5.5 Run 2 array die . . . 109
5.6 OPMD probe station . . . 110
5.7 Laser Setup diagram for gain and timing test . . . 111
5.8 OPMD laser test setup . . . 111
5.9 Example of timing measurement using the SDR method . . . 113
5.10 RAL laser test setup . . . 113

5.11 ^{90}Sr test setup for timing measurements . . . 114
5.12 Example of a signal response from two LGADs detecting the same electron 116
5.13 Schematic diagram of the Compact OPMD LGAD Amplifier (COLA) . . . 117
5.14 A picture of a COLA board . . . 118
5.15 FEMTO DUPVA variable amplifier used as a second stage for LGAD test . 118
5.16 Scope snapshot of COLA rms output noise . . . 119
5.17 Wafer map with field locations . . . 120
5.18 IV plots of WF2, WF8 and WF9 LGAD devices and PIN diodes . . . 121
5.19 A breakdown voltage map of WF19 LGAD array . . . 122
5.20 CV plots of LGAD . . . 124
5.21 CV plots of PIN diode . . . 125
5.22 CV plots of a Run 2 LGAD array . . . 125
5.23 An example of a laser test pulse signal . . . 126
5.24 Example of comparison of PIN and LGAD WF2 pulse signal output for laser charge injection at different bias voltages . . . 127
5.25 Gain plots of WF2 LGAD devices . . . 128
5.26 Gain plots of WF8 LGAD devices . . . 128
5.27 Gain plots of WF9 LGAD devices . . . 128
5.28 Gain and charge plots of WF2, WF8 and WF9 LGAD devices vs. bias voltage normalised to the breakdown voltage V_{bd} . . . 129
5.29 Jitter and slew rate plots of WF2, WF8 and WF9 LGAD devices . . . 130
5.30 Jitter of WF2, WF8 and WF9 LGADs vs. bias voltage normalized to the breakdown voltage V_{bd} . . . 131
5.31 LGAD timing resolution using ^{90}Sr source . . . 132
5.32 Median energy deposition maps for protons incident on LGAD . . . 134
5.33 LGAD gain for protons of different energy . . . 135

6.1 Standard and segmented AC-LGAD . . . 138
6.2 The inverted or double-sided LGAD (iLGAD) . . . 139
6.3 The Trench-Isolated (TI) LGAD . . . 139
6.4 The Deep-Junction (DJ) LGAD . . . 140
6.5 Shallow Graded Epi LGAD (GELGAD) cross-section . . . 141
6.6 Deep GELGAD cross-section . . . 142
6.7 2D section of Te2v Run 2 cell array used for GELGAD TCAD simulation . 142
6.8 The GELGAD fabrication process flow . . . 143
6.9 Shallow GELGAD doping profile obtained by SPROCESS . . . 144
6.10 Deep GELGAD doping profile obtained by SPROCESS . . . 145

6.11 Simulation of current vs. voltage of shallow and deep GELGAD 145
6.12 Deep GELGAD electric field at V_{b95} bias 147
6.13 Field profile of deep GELGAD at V_{b95} bias 147
6.14 Shallow GELGAD electric field at V_{b95} bias 148
6.15 Field profile of shallow GELGAD at V_{b95} bias 148
6.16 Shallow GELGAD transient current for MIP charge injection 149
6.17 Shallow GELGAD gain and collected charge as a function of normalized bias voltage 150
6.18 Deep GELGAD transient current 151
6.19 Deep GELGAD gain and collected charge as a function of normalized bias voltage 151
6.20 Shallow and Deep GELGAD pulse current for central MIP hit 152
6.21 Shallow and Deep GELGAD gain and collected charge as a function of normalized bias voltage 152
6.22 Shallow GELGAD two cells structure 154
6.23 Deep GELGAD two cells structure 154
6.24 Shallow GELGAD two cells MIP transient currents 155
6.25 Deep GELGAD two cells MIP transient currents 155
6.26 GELGAD two cells MIP gain vs. transverse position 156
6.27 Deep GELGAD termination structure 157
6.28 Shallow GELGAD termination structure 158
6.29 Deep GELGAD periphery at V_{b95} bias 158
6.30 Shallow GELGAD periphery at V_{b95} bias 158
6.31 Current vs. voltage of the extended shallow and deep GELGADs 159
6.32 Doping concentration and electric field profile of the high doping deep GELGAD at V_{b95} bias 160
6.33 Doping concentration and electric field profile of the high doping shallow GELGAD at V_{b95} bias 160
6.34 Doping concentration and electric field profile of the low doping deep GELGAD at V_{b95} bias 161
6.35 Doping concentration and electric field profile of the low doping shallow GELGAD at V_{b95} bias 161
6.36 Current vs. voltage of the shallow and deep GELGADs for higher and lower doping 162

List of tables

1.1 Main parameters of HGTD. 11

4.1 Details of Run 1 LGAD cells layout 68
4.2 Dose and energy of the ^{11}B layer (GL) implants in the Run 1 LGAD production 68
4.3 Details of Run 2 LGAD cells layout 70
4.4 Dose and energy of the ^{11}B layer (GL) implants in the Run 2 LGAD production 70
4.5 Results of SIMS and TCAD doping comparison 80
4.6 Breakdown values from TCAD simulation in Volts 86
4.7 Optical charge injection parameters used in the TCAD simulation 89
4.8 Optical charge injection Q_{coll} and Gain from TCAD for the Okuto - Crowell model 94
4.9 Optical charge injection Q_{coll} and Gain from TCAD for the University of Bologna model 95
4.10 MIP charge injection parameters used in TCAD simulations 96
4.11 MIP charge injection Q_{coll} and Gain TCAD results for the Okuto - Crowell model 100
4.12 MIP charge injection Q_{coll} and Gain TCAD results for the University of Bologna model 100

5.1 COLA specifications - T = 20 °C. 117
5.2 FEMTO DUPVA-60 specifications - Gain = 30 dB, T = 20 °C. 119
5.3 IV and CV test results for Run 1 type 3 LGAD and PIN. 123
5.4 Required electrics properties of the single pad of the HGTD sensors at room temperature. 123
5.5 Table of QD laser settings used for gain and timing test. 126
5.6 PIN collected charge. 127
5.7 Laser charge injection average gain $\pm$ 1 σ for WF2, WF8 and WF9. 129
5.8 Laser charge injection average jitter $\pm$ 1 σ [ps] for LGAD devices of WF2. 131

5.9 Laser charge injection average jitter $\pm$ 1 σ [ps] for LGAD devices of WF8. 131
5.10 Laser charge injection average jitter $\pm$ 1 σ [ps] for LGAD devices of WF9. 132

6.1 Shallow GELGAD results - T = 21 °C. 156
6.2 Deep GELGAD results - T = 21 °C. 156

A.1 University of Bologna Impact Ionisation parameters for silicon. 180
A.2 Okuto-Crowell Impact Ionisation parameters for silicon. 181

Chapter 1

Introduction

This chapter describes the use of particle detectors for space and time measurements and presents the topic of this thesis.

Section 1.1 introduces the use of particle detectors for measurements of trajectories and for track arrival times. Section 1.2 describes the pileup issue at the High Luminosity Large Hadron Collider at CERN (HL-LHC) and the expected benefits in track reconstruction obtained from timing information using fast timing detectors. Some description of the High Granularity Timing Detector (HGTD) for ATLAS Phase-II detector Upgrade along with specifications and proposed Low Gain Avalanche Detector (LGAD) technology for the fast timing sensing elements are also given in the same section. Finally, in Section 1.3, an outline of the thesis is shown, describing what is covered in each chapter.

1.1 Detectors for space and time measurement

The use of detectors for the measurement of trajectories of charged particles spans several areas of physics, including space science, the detection of cosmic rays, medical applications, and nuclear and particle physics. Typically such applications require the measurements of particle locations to specified accuracy over a physical area and the capability to deal with incoming rate of the particles.

Developing silicon sensors through the planar fabrication process has allowed accurate particle measurements [1]. Along with the parallel development of very large-scale integration (VLSI) on a silicon chip, it allowed for achieving very high spatial granularity while at the same time covering large areas using arrays and providing the ability to accept data at very high rates [2].

A limiting factor in increasing the rate capability of silicon sensors is the physical process of signal formation. Typically, in standard silicon sensors, the arrival time of a particle is

measured with an accuracy of approximately 0.2 ns which, for a relativistic particle, translates into a flight path uncertainty of around 6 cm. Enhancing the capabilities to measure particle arrival times would provide detectors of simultaneous high spatial granularity and high data rate collection (Ultra Fast Detectors, UFD), a feature that could be very useful not only for the time of flight measurements of a particle but also to aid tracking and vertex reconstruction in high luminosity colliders.

1.2 Timing information at HL-LHC

An example of the importance of timing information in particle colliders can be drawn from the high-luminosity upgrade of the Large Hadron Collider (HL-LHC) at CERN, Figure 1.1:

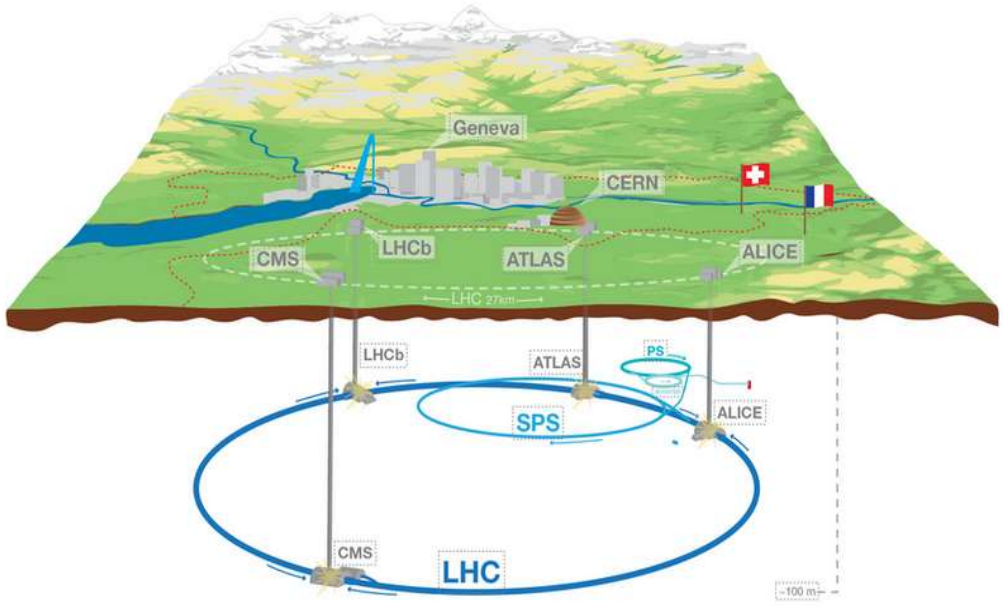


Fig. 1.1 *Schematic view of the LHC at CERN. The 27 km LHC accelerator ring is located approximately 100 m underground. Counter-rotating beams of protons are accelerated to an energy of 7 TeV and made to collide at four points. ALICE, ATLAS, CMS, and LHCb are the four experiments installed in the underground caverns built around the collision points of the LHC beams. Source: CERN.*

At the LHC, the last in a chain of particle accelerators at CERN, protons are accelerated to the highest possible kinetic energy (currently up to 6.8 TeV in each of the two counter-rotating beams) and then made to collide at four different interaction points, where dedicated

detectors are located, like the ATLAS detector shown in Figure 1.2. The analysis of the result of these collisions, which rely on sophisticated algorithms of track reconstruction to identify the particles through their trajectories, allows us to investigate the structure of the subatomic world and test the predictions of various theories of particle physics, including the Standard Model (SM) of particle physics, a theoretical model that describes all known elementary particles [3]. The SM describes the electromagnetic, the weak, and the strong interactions in the same theoretical framework, which does not include the gravitational interaction. Besides confirmation and test of additional components of the SM one of the main objectives of the LHC was to discover, or to exclude, the existence of the Higgs boson, a neutral scalar boson that is the quantum of a field that is required to describe massive particles in the SM and in which it plays a fundamental role. On July 4$^{\text{th}}$, 2012, the ATLAS and CMS collaborations jointly announced the discovery of a new particle with a mass of 125 GeV [4, 5], which would become known as the Higgs boson.[1].

Each of the two counter-rotating colliding beams at the LHC consists of many bunches of protons, each bunch having a length of approximately 7 cm, transverse dimensions of about 10 μm, and containing a few 10^{11} protons. The proton bunches are separated in time by 25 ns, corresponding to a spatial separation of approximately 750 cm. There are nominally about 2700 filled bunches in each beam and this number cannot be substantially changed, due to constraints in bunch injection into the LHC. The bunch crossing occurs at a rate of 40 MHz. At the nominal luminosity[2] of the LHC the average number of proton-proton (pp) interactions in a single bunch crossing (pileup) is approximately 23 and is limited to a maximum of 60. However, the planned luminosity upgrade to HL-LHC will increase this number to about 200 [6]. Most of these interactions are "soft", i.e. glancing collisions, that do not contribute

[1]Despite the excellent description of most observed phenomena and outstanding prediction capabilities, the SM does not contain particles that are consistent with being the constituents of dark matter, the existence of which has been inferred by astronomical observations from its gravitational effects. Also, the SM does not explain why our universe shows an excess of matter although the same amount of matter and anti-matter should have been generated in the Big Bang. Another important question is whether a unifying theory exists that provides a common description of all forces (strong, electromagnetic, weak, and gravitational). All this suggests the existence of new physics (BSM or beyond the Standard Model physics) around the TeV energy scale that is not contained in the SM and the discovery of which is a key objective of the LHC. For direct tests of the SM, searches for dark matter and beyond the Standard Model physics (BSM) a deeper understanding of the interactions and self-interactions of the Higgs boson are crucial, with many of these processes characterised by extremely low cross-section. To be able to draw robust conclusions a larger amount of data is therefore needed, which requires an increase in luminosity and related integrated luminosity from the LHC. The High Luminosity LHC (HL-LHC) upgrades are currently underway and focused on delivering ten times more integrated luminosity than the current LHC program.

[2]The quantity that provides a measurement of the number of collisions that can be produced in a detector per cm^2 and second is called luminosity and is the proportionality factor $\mathscr{L}$ between the number of events per second $\frac{dR}{dt}$ and the cross-section of the event σ_p: $\frac{dR}{dt} = \mathscr{L}\sigma_p$. The bigger the value of $\mathscr{L}$, the bigger the number of collisions. At LHC, this value is $\approx 10^{34}$ cm^{-2}·s^{-1}.

to the search for new physics at the TeV energy scale. A relatively small fraction of all collisions are "hard" collisions that contain high transverse momentum particles that may come from new high-mass objects. Nevertheless, the presence of some tracks and energy from the (on average) 200 minimum bias collisions[3] can confuse or degrade the triggers and the offline track reconstruction of the hard collisions. Since the number of bunches cannot be increased, the luminosity increase at the LHC results in a higher pileup, producing many more hits in the tracking detectors and leading to a higher fraction of mismeasured or misidentified tracks. The spatial pileup line density, i.e. the number of collisions per unit length along the beam axis during one bunch crossing, represents therefore an important parameter to evaluate the performance of the Inner Tracker (ITk) of ATLAS detector in the HL-LHC [7–9], shown in Figure 1.3. Because of the pileup, track reconstruction becomes particularly challenging in the end-cap region at high η[4], Figure 1.4.

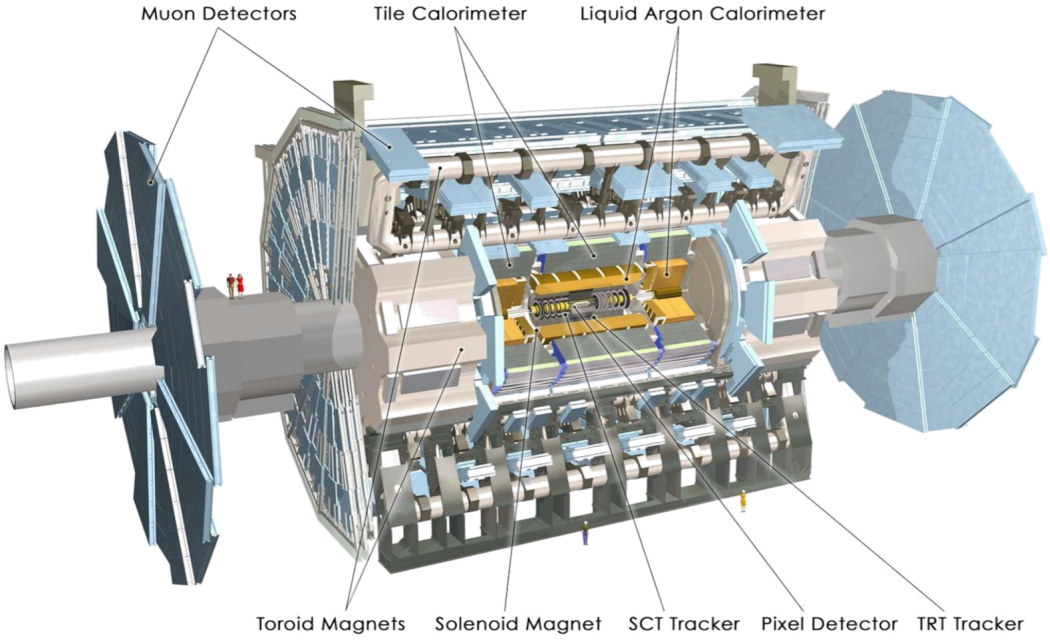


Fig. 1.2 *Schematic view of the ATLAS detector at CERN with the various sub-systems. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is about 7000 tons. Source: CERN.*

[3]The minimum bias collisions in this context is defined as the "spectator collisions" to the hard collision of interest.

[4]The ATLAS reference system is a right-handed Cartesian coordinate system, with the nominal collision point at the origin. The positive x-axis is defined as pointing from the collision point to the centre of the LHC ring and the positive y-axis points upwards, the z-axis is defined along the beam pipe. The azimuthal angle, ϕ is measured around the beam axis, and the polar angle, θ is measured with respect to the z-axis. The pseudorapidity is defined as η = -ln tan(θ/2).

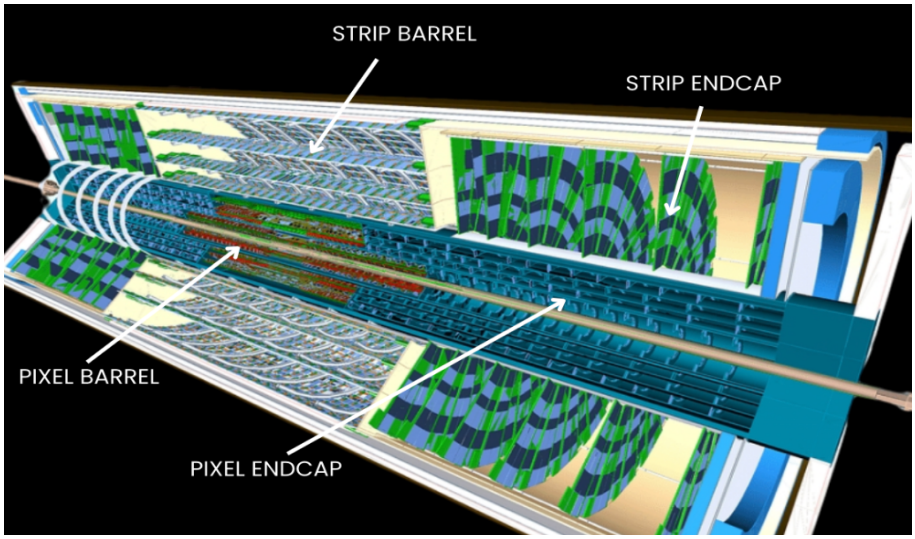


Fig. 1.3 *Schematic view of the upgraded central tracking system (Inner Tracker, ITk) for the HL-LHC phase of ATLAS detector at CERN. It is an all-silicon system that consists of a Pixel Detector at a small radius and a large area Strip Detector surrounding it. Source: CERN.*

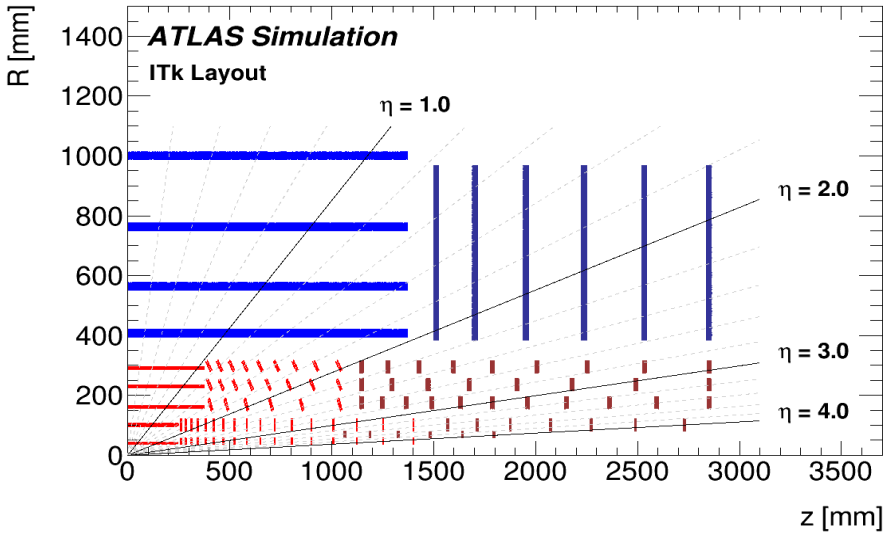


Fig. 1.4 *Schematic layout of one quadrant of the ITk for the HL-LHC phase of ATLAS detector. Only the active detector elements are shown. The detector elements of the barrel and end-cap Strip detector are shown in blue, for the ITk Pixel are shown in red. Tracks at high* η *traverse more material hence the accuracy in their reconstruction is degraded by multiple scattering. Reprinted from [10].*

Figure 1.5 illustrates an LHC example of track density with a relatively high pileup for one bunch crossing with ≈ 50 collisions. In this case, the density of events occurring in different locations is such that spatial tracking information is sufficient to reconstruct each vertex correctly.

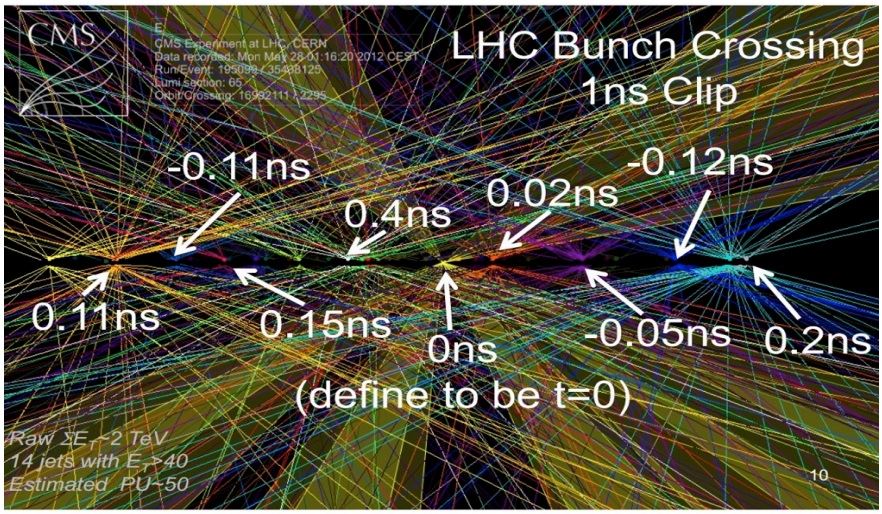


Fig. 1.5 *Interaction time of many pp vertexes happening in the same bunch crossing at LHC in the case of ≈ 50 overlapping events. The vertexes are spaced 10's of picoseconds apart. Source: CERN.*

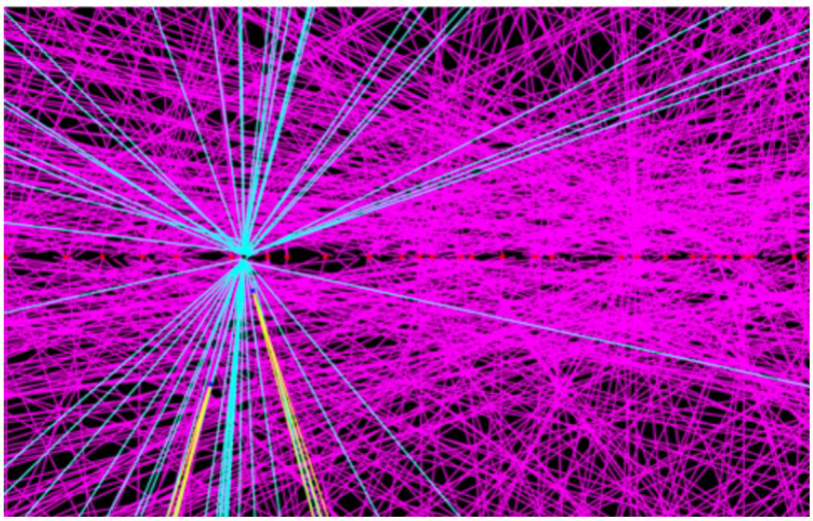


Fig. 1.6 *ATLAS simulation of spatial events at HL-LHC with $< \mu >= 140$ pileup. Source: CERN.*

Figure 1.6 shows the results of an ATLAS simulation at HL-LHC. In such a scenario the number of events per bunch crossing is higher, rendering the process of track reconstruction

much more difficult. Figure 1.7 shows that at the HL-LHC, with an average pileup of 200, the vertex line density reaches an average (most probable) value of 1.8 vertices/mm (1.44 vertices/mm) [10].

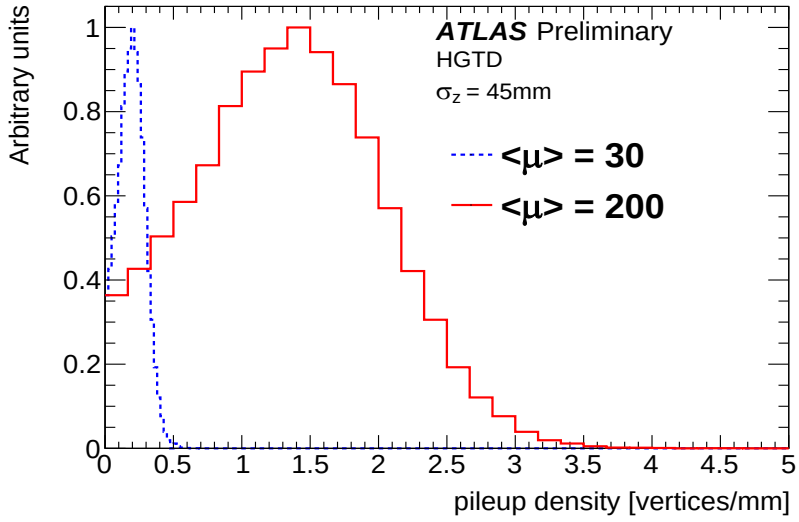


Fig. 1.7 *Simulation of HL-LHC local pileup vertex densities for two values of* $< \mu >= 30$ *and* $< \mu >= 200$. *Reprinted from [10].*

When the ITk impact parameter resolution in z, i.e. the direction along the beamline, is better than the typical distance between two vertices, the vertices can be distinguished. This is the case in the central region of the detector, as shown in Figure 1.8. When the z_0 resolution degrades and becomes larger than the distance between two vertices, precision timing can help separate the vertices, by reducing the density of vertices which are considered for a given track. Figure 1.9 shows an ATLAS simulation of the events dispersion in time for a given z_0 in a single HL-LHC bunch crossing, assuming an average pileup of $< \mu >= 200$. The timing distribution of the events in Figure 1.9 has an rms of $\approx$ 150 ps. Therefore, a timing resolution of 30 ps would allow dividing such events into 5 groups, each with several concurrent events equal to one-fifth of the total, which would provide an almost complete pileup suppression, eliminating the overlapping events by precise time tagging.

Using this concept, a High-Granularity Timing Detector (HGTD), a novel detector introduced to enhance the performances of the ITk in the forward region by adding the capability to measure charged-particle trajectories in time as well as space, has been designed [10]. The HGTD will measure the arrival times of charged particles with an average time resolution of approximately 30 ps per track at the beginning of the operation of HL-LHC, increasing to 50 ps due to radiation damage at the end of its operation. Such high-precision time measure-

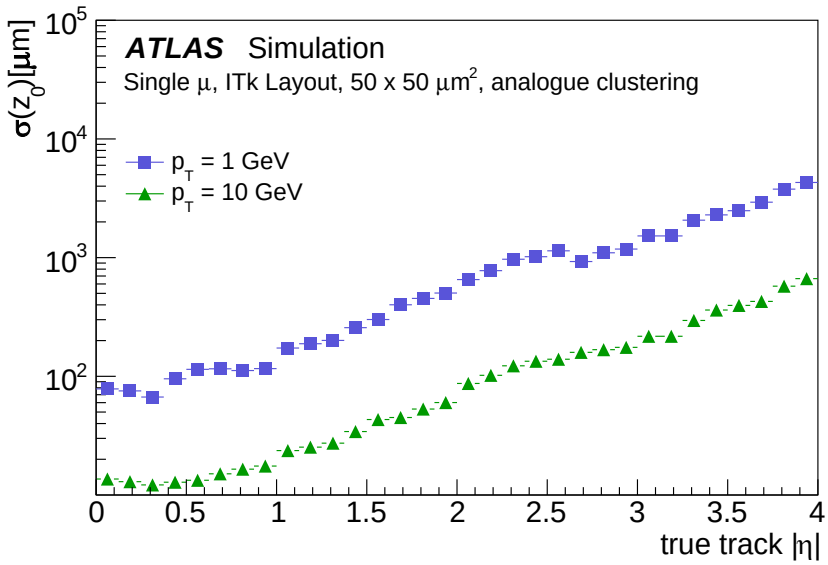


Fig. 1.8 *Resolution of the longitudinal track impact parameter, z_0, as a function of η for muons of transverse momentum p_T = 1 GeV and p_T = 10 GeV using ITk alone. Reprinted from [10].*

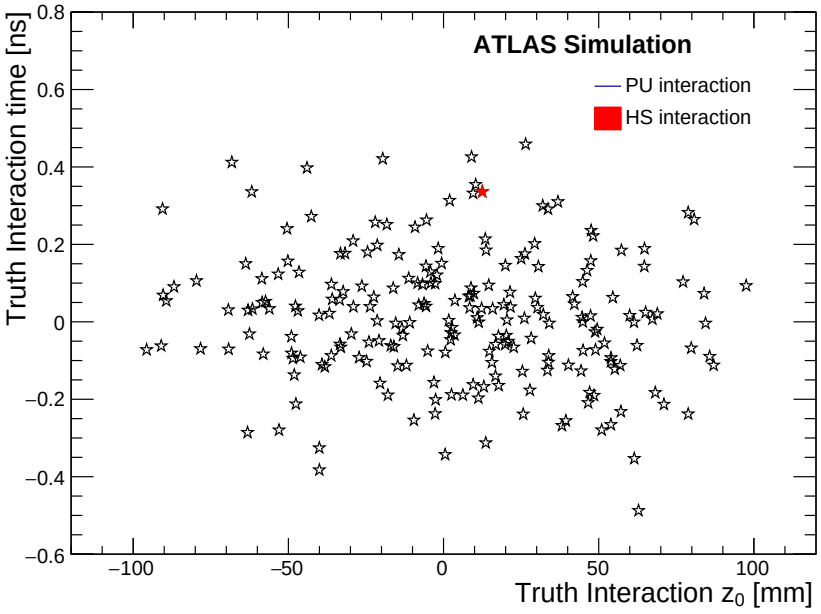


Fig. 1.9 *ATLAS plot of the truth interactions in a single bunch HL-LHC crossing in the z_0–t plane, showing the simulated Hard Scatter (HS) tt event interaction (red) with superimposed pileup interactions (black) for $< \mu >= 200$. Reprinted from [10].*

ments for charged particles will enhance the performance of physics object reconstruction, complementing the ITk in the high-η forward region and improving the physics potential of ATLAS at the HL-LHC.

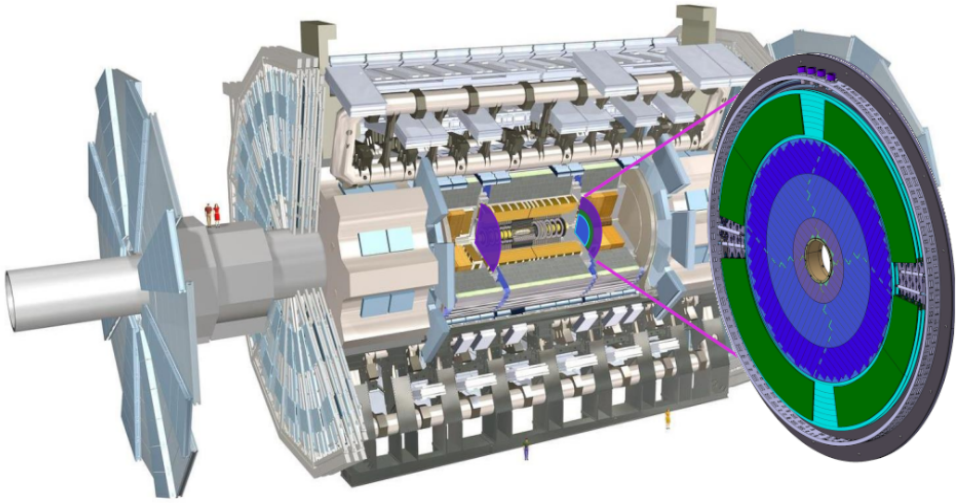


Fig. 1.10 *Position of the HGTD within the ATLAS Detector. The HGTD acceptance is defined as the surface covered by the HGTD between a radius of 120 mm and 640 mm at a position of $z = \pm$ 3.5 m along the beamline, on both sides of the detector. Reprinted from [10].*

Figure 1.10 shows the arrangement of the HGTD within the ATLAS detector. It will be located in the gap region between the barrel and the end-cap calorimeters, at a distance in z of approximately $\pm$ 3.5 m from the nominal interaction point. The HGTD has been designed for operation with $< \mu >= 200$ and a total integrated luminosity of 4000 fb^{-1}[5]. Each end-cap of the HGTD consists of one hermetic vessel, two instrumented double-sided layers (mounted on two cooling/support disks), and two moderators placed inside and outside the hermetic vessel, to reduce the back-scattered neutrons created by the end-cap/forward calorimeters, protecting both the ITk and the HGTD. Each cooling/support disk is physically separated into two half circles. To maximize the hit efficiency the layers are rotated in opposite directions with respect to one another by 15 to 20°.

A global view of the various components of the detector and its main parameters is shown in Figure 1.12.

[5]The integral of the delivered luminosity over time is called integrated luminosity and it is a measurement of the collected data size. It is usually expressed in the inverse of cross-section, like 1 fb^{-1}, where 1 fb = 10^{-15} barn, where a barn = 10^{-24} cm^2. Note that 1 fb^{-1} $\approx 7 \times 10^{12}$ pp collisions.

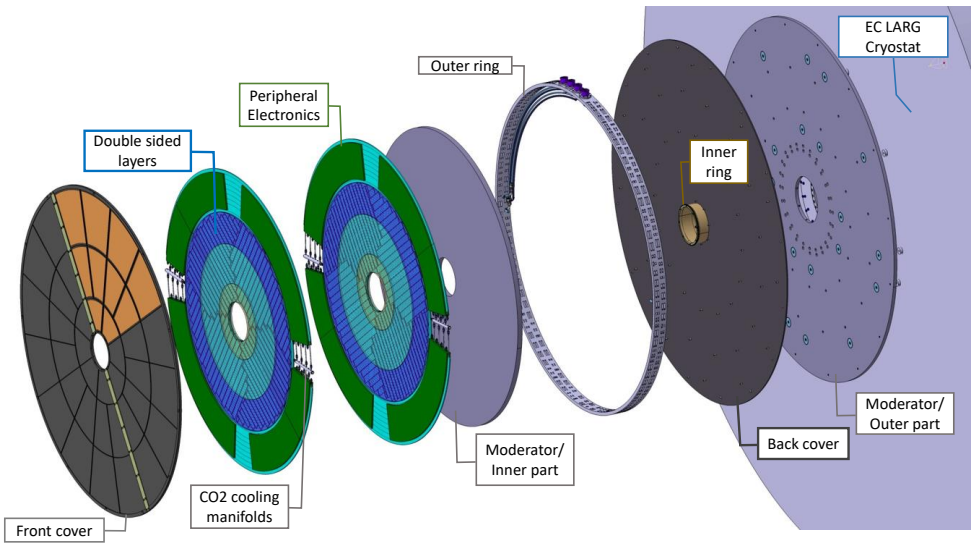


Fig. 1.11 *Global view of the HGTD to be installed on each of two end-cap calorimeters. The various components are shown: hermetic vessel (front and rear covers, inner and outer rings), two instrumented double-sided layers (mounted in two cooling disks with sensors on the front and back of each cooling disk), two moderator pieces placed inside and outside the hermetic vessel. Reprinted from [10].*

Besides suppressing pileup by its timing capabilities, the HGTD can provide luminosity measurement at HL-LHC. Its occupancy which is low due to its high granularity and large distance from the interaction area, is correlated with the interaction rate (i.e. the luminosity), providing good linearity between the average number of hits and the average number of simultaneous pp interactions over the full range of HL-LHC luminosity. Also, with detector signal durations in the few ns range, the charged-particle multiplicities within the acceptance can be determined accurately for each bunch crossing separately. To reduce the bandwidth required from the readout ASIC, the current plan is to perform the luminosity measurement in a reduced η range and to read out the ASICs for sensors located in the region 430 mm $<$ r $<$ 640 mm (equivalent to η range $2.4 < \eta < 2.8$) [10]. For the HGTD readout, a custom ASIC (ALTIROC [11]), which will be bump-bonded to the sensors, is being developed to meet the requirements of time resolution and radiation hardness. The ASIC will also provide the functionality to count the number of hits registered in the sensor and transmit it at 40 MHz to allow unbiased, bunch-by-bunch measurements of the luminosity and the implementation of a minimum-bias trigger. The main parameters of the HGTD are reported in Table 1.1.

One of the crucial parameters of the HGTD will be the radiation hardness of the sensors and electronics. Since the HGTD will be installed with a pseudo-rapidity coverage of η range $2.4 < \eta < 4.0$, it is essential that the detector can withstand the radiation levels throughout

the HL-LHC operations. The neutron-equivalent fluence (1 MeV n_{eq} cm^{-2}) at a radius of 120 mm, is expected to reach 5.6×10^{15} n_{eq} cm^{-2} and the Total Ionising Dose (TID) about 3.3 MGy as shown in Figure 1.12. After applying safety factors, to account for uncertainties in the simulation, the detector would need to withstand 8.3×10^{15} n_{eq} cm^{-2} and 7.5 MGy.

Table 1.1 *Main parameters of HGTD.*

Mechanical parameter	value
Pseudo-rapidity coverage	$2.4 < \|\eta\| < 4.0$
Thickness in z	75 mm (+50 mm moderator)
Position of active layers in z	$\pm$ 3.5 m
Weight per end-cap	350 kg
Radial extension:	
Total	110 mm $<$ r $<$ 1000 mm
Active area	120 mm $<$ r $<$ 640 mm
Sensor parameter	**value**
Pad size	1.3 mm $\times$ 1.3 mm
Active sensor thickness	50 μm
Number of channels	3.6 M
Active area	6.4 m^2
Module size	30 $\times$ 15 pads (4 cm $\times$ 2 cm)
Modules	8032
Collected charge per hit	$>$ 4.0 fC
Average number of hits per track	
$2.4 < \|\eta\| < 2.7$ (640 mm $>$ r $>$ 470 mm)	$\approx$ 2.0
$2.7 < \|\eta\| < 3.5$ (470 mm $>$ r $>$ 230 mm)	$\approx$ 2.4
$3.5 < \|\eta\| < 4.0$ (230 mm $>$ r $>$ 120 mm)	$\approx$ 2.6
Average time resolution per hit (start and end of operational lifetime)	
$2.4 < \|\eta\| < 4.0$	$\approx$ 35 ps (start), $\approx$ 70 ps (end)
Average time resolution per track (start and end of operational lifetime)	$\approx$ 30 ps (start), $\approx$ 50 ps (end)

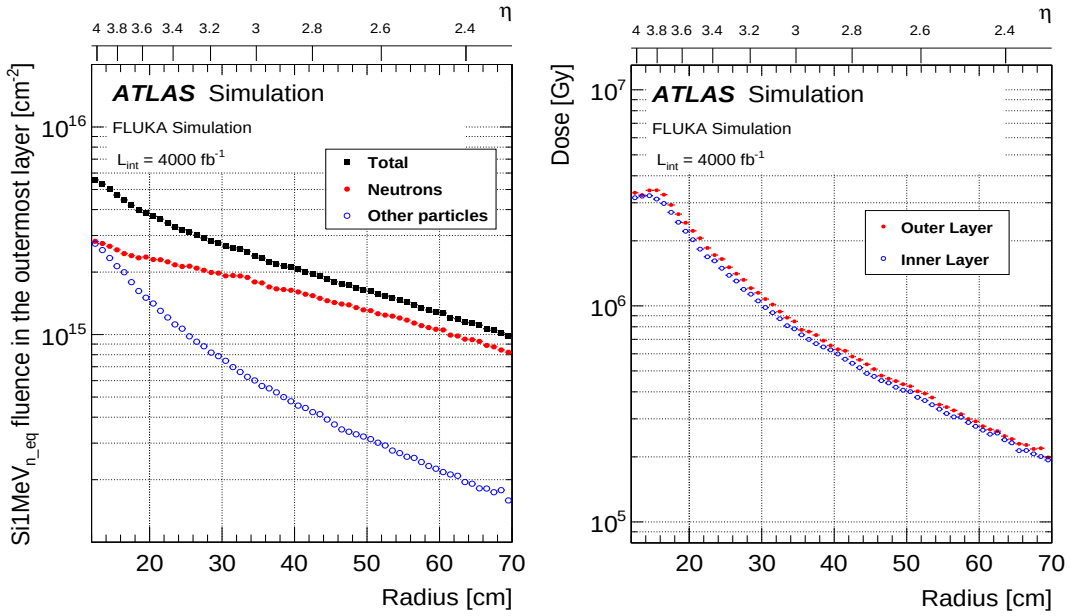


Fig. 1.12 *Expected nominal Si 1 MeV n_{eq} fluence and ionising dose as functions of the radius in the outermost sensor layer of the HGTD for 4000 fb^{-1}, i.e. before including safety factors. The contribution from charged hadrons is included in 'Others'. These estimations used Fluka simulations using ATLAS Fluka geometry 3.1Q7 (from December 2019) [10].*

To achieve sufficient performance of the sensors and ASICs given the expected radiation levels, the detector layout has been designed considering a replacement scenario during the HL-LHC program, with the sensors and electronics within the lowest radius (r $<$ 230 mm) that will be replaced after each 1000 fb^{-1} and the sensors and ASICs within 230 mm $<$ r $<$ 470 mm that will be replaced at half of the data-taking (2000 fb^{-1}).

Due to space limitations and the need for the HGTD sensor to be thin and configurable in arrays, a low gain avalanche detector (LGAD [12]), a silicon-based timing detector technology, has been chosen. An extensive R&D program is still ongoing within the framework of RD50 [13]. However, baseline sensors that can provide the required timing resolution in the harsh radiation environments of HL-LHC (the sensors will have to be operated at -30 °C to mitigate the effect of the accumulated radiation dose) have been identified and produced by different vendors.

In this thesis the design, simulation, and testing of a silicon LGAD, targeting the HGTD specifications, is presented. All aspects of the design process, including layout, TCAD simulation of the fabrication process, and electrical characteristics, will be described. Characterization of the fabricated devices, including charge collection and timing performances, will be presented and discussed. TCAD simulation of a novel LGAD structure, with details

of the proposed fabrication process which aims at increasing the fill-factor, i.e. the percentage of the sensor surface that has gain, of conventional LGAD, will also be shown.

1.3 Thesis Outline

This introductory chapter describes the use of detectors for space and time measurements and highlights the potential benefits of fast-timing detectors (or Ultra-Fast Detectors, UFD) in Particle Physics experiments at the HL-LHC to reduce the pileup effect. A brief description of the HGTD at HL-LHC, a timing detector designed to mitigate pileup effect and based upon an LGAD silicon technology, is provided.

Chapter 2 gives an overview of the fundamental aspects of semiconductor physics relevant to particle detection. Some general aspects of the process steps involved in device manufacturing, with a focus on the applications for detectors in physics experiments, are also provided.

Chapter 3 introduces timing detectors and their implementation using Low Gain Avalanche Diode (LGAD). Their operating principle, based on impact ionisation, is discussed, with a brief review of the impact ionisation process and its modelling. Typical circuitry for the readout of LGAD devices is also shown in this chapter.

Chapter 4 describes the design and simulation of LGAD devices. Details of the sensor's single cell and array, designed to meet the HGTD specifications, and the simulation of the fabrication process along with electrical and charge collection simulation are presented. Some aspects of TCAD simulation, including meshing strategy and validity of a local model for impact ionisation, are discussed.

Chapter 5 describes the fabricated devices and the experimental techniques and setups used for their characterisation. Obtained test results are shown and discussed. Some initial test results on LGAD devices for radiation dosimetry applications are also described in this chapter.

Chapter 6, after an introductory summary of current solutions to increase the fill-factor, proposes a novel technological solution based on a deep gain layer obtained via graded epitaxial growth (GELGAD), for highly segmented LGADs. Details of TCAD simulations, which include fabrication and initial electrical simulations, are presented, which suggests that the proposed solution might provide better uniformity in terms of gain with respect to conventional LGAD. Additional simulations investigating the edge effects of the structure for their implementation are also provided.

Finally, Chapter 7 summarises the presented work and discusses future developments.

Chapter 2

Silicon Detectors

This chapter presents a summary of semiconductor physics relevant to particle detection, along with some general details on the process steps involved in device manufacturing.

In Section 2.1 silicon as a semiconductor is introduced, and its crystal structure and electrical characteristics are described, including aspects of charge transport. The *pn* junction, arguably the most important electronic structure, is presented in Section 2.2 along with a description of its operating principle. Section 2.3 describes the operating principle of particle and radiation detectors, including the signal formation. Some common steps used in the fabrication process of semiconductor devices, focusing mostly on those used in the fabrication of the devices presented in this thesis and used during the design and simulation phase, are discussed in Section 2.4.

2.1 Introduction to silicon as a semiconductor

2.1.1 Crystal Structure

Most commonly used semiconductors are single crystals with diamond (Si and Ge, belonging to the IVth group of the periodic table) or zinc blende lattice (GaAs and other III–V compound semiconductors), as shown in Figure 2.1.

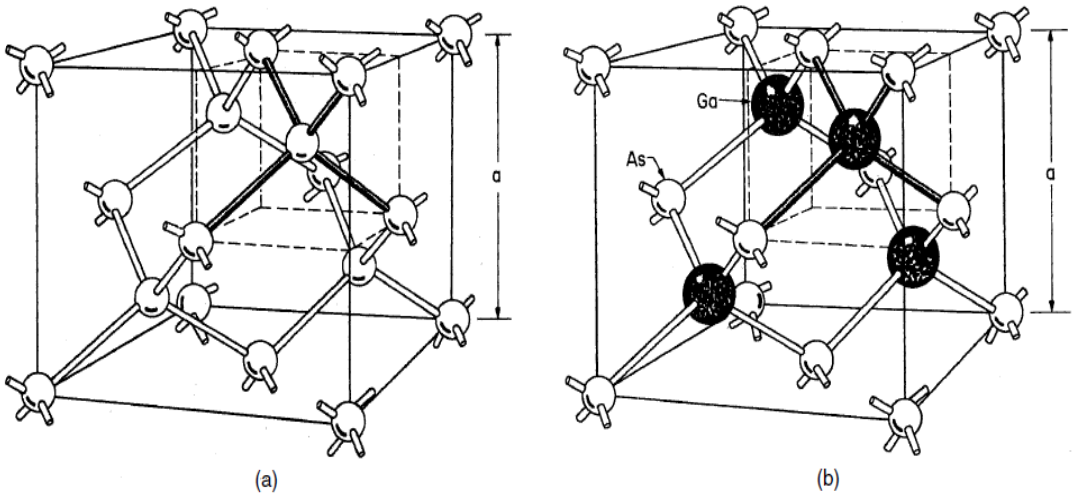


Fig. 2.1 *Diamond (a) and zinc blende (b) lattice. Reprinted from Sze [14].*

Both lattices can be viewed as being composed of two interpenetrating face-centred cubic (fcc) sublattices that are displaced by one-quarter of the distance along the diagonal of the cube. All atoms in the diamond lattice are identical whereas the two fcc sublattices consist of different atoms in the case of III–V compounds such as GaAs. As shown in Figure 2.2a each atom is surrounded by four close atoms belonging to the other fcc sublattice and they are arranged in a tetrahedron. Covalent bonds are formed with each of the neighbouring atoms by sharing the outer four valence electrons, as shown in Figure 2.2b.

The lattice orientation of a silicon wafer is described using the Miller indices [15], also shown in Figure 2.2a. Typically, three wafer orientations (<100>, <110>, and <111>) are used to produce silicon devices. The crystalline difference between the three orientations significantly affects the electrical properties of the devices. In high energy physics the <100> orientation is normally used since it is more radiation tolerant compared with the <111>. This stems from the lower density of oxide charge and dangling bonds of <100> compared to <111>, which affect the surface current and the interpad capacitance through the Si–SiO_2 interface trapped charge (around 10^{10} cm^{-2} for <100> and 10^{11} cm^{-2} for <111>), especially after exposure to ionising radiation [16].

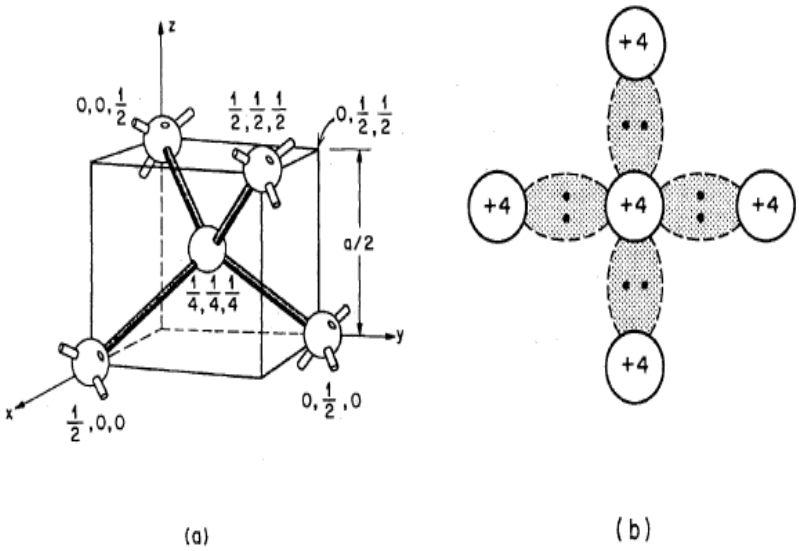


Fig. 2.2 *Tetrahedron bond (a) and schematic two-dimensional representation (b). Reprinted from [14].*

2.1.2 Energy Bands

The formation of energy bands in a semiconductor can be understood using two different approaches. The first approach starts from a single, isolated atom, which shows discrete energy levels. When two atoms come close to each other, their energy levels split into an upper and lower level, forming the bonding and antibonding states, as described in covalent bond theory, and delocalizing across the two atoms. By adding more atoms the number of levels increases and the groups of levels form bands. Typically many bands are present in most semiconductors and, after band formation, an energy gap is found between occupied and unoccupied bands. The second approach starts with free electron waves. Upon the introduction of the electrostatic potential associated with the atoms of the bulk material, the waves undergo Bragg reflection and cannot penetrate the bulk, which results in opening the energy band gap in the free electron energy-momentum $E(k)$ dispersion relations [17]. Depending on the crystallographic direction, different dispersion relations are obtained, for which approximate analytical formulas exist.

Examples of the band structure of silicon are shown in Figure 2.3. As the minimum of the conduction band, E_C, lies at different k values of the maximum of the valence band, E_V, silicon is an indirect bandgap semiconductor. Other materials, like GaAs, where the extrema are found at the same k value are direct bandgap semiconductors.

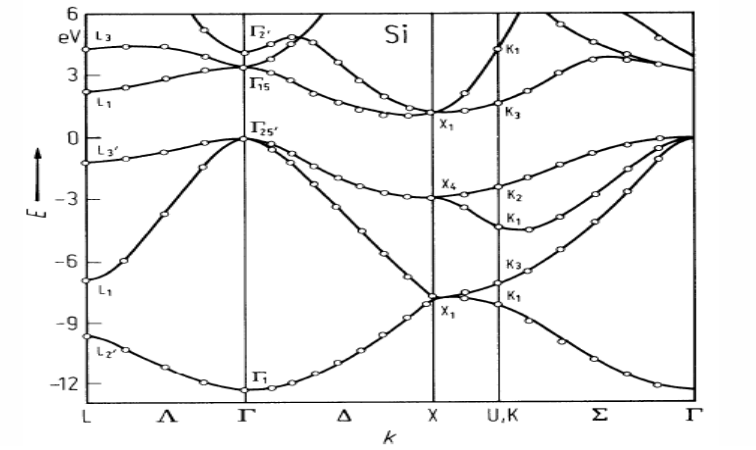


Fig. 2.3 *Silicon band diagram along the <100> (Δ), <110> (Σ), and <111> (Λ) direction. The valence band consists of a light, heavy, and split-off band, which have a maximum at the* Γ *point ($k = 0$). The energy band gap at this point is around 3.2 eV. Reprinted from [14].*

At low temperatures, the electrons fill the low energy band (valence band), and all states in the next higher energy band (conduction band) will be empty. By increasing the temperature, the thermal energy might become high enough to promote a few electrons to the conduction band, while simultaneously creating a vacancy (hole) in the valence band. The free electrons and holes created increase the conductivity of the semiconductor. The structure of an insulator, like SiO_2, is similar except that the band gap is much larger so that the occupation probability of states in the conduction band is nearly zero. Conductors may have overlapping valence and conduction bands or a partially filled conduction band, as shown in Figure 2.4.

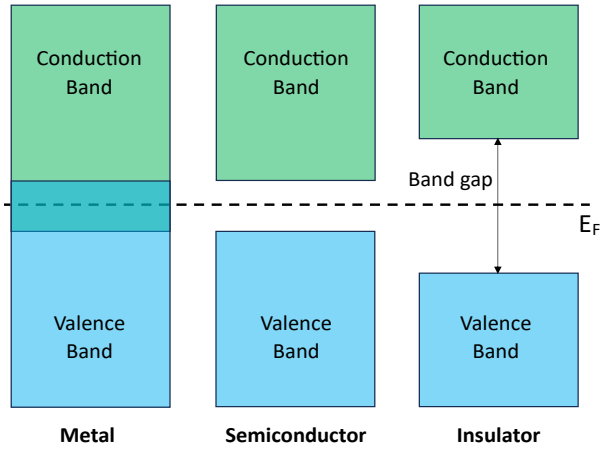


Fig. 2.4 *Energy band structure of metals, semiconductors, and insulators. E_F is the Fermi level, defined as the level at which the occupation probability is one-half.*

Usually, it is possible to treat electrons in the conduction band and holes in the valence band similarly to free particles, but with an effective mass m_n^* or m_p^*, which account for their motion inside a crystal and is different from the free electron mass. Such an effective mass also depends on other parameters, like the direction of movement with respect to the crystal axis. Knowledge of the band structure allows for a full energy description of the carriers in it, i.e. the kinetic energy of electrons is obtained from its upward distance from E_C, the lower edge of the conduction band, and that of the holes downward from E_V, the upper edge of the valence band.

2.1.3 Carrier Concentration

Intrinsic Semiconductors

A semiconductor is considered intrinsic when it does not contain impurities, or the impurities are negligible compared to the number of thermally generated electrons and holes. Under equilibrium conditions the electrons (n) and holes (p) carrier concentration can be calculated by integrating over all the valence and conduction bands, using the density of available states in each band $g_c(E)$ and the probability of states occupation, described by the Fermi-Dirac distribution $f(E)$:

$$n = \int_{E_c}^{E_{top}} g_c(E) f(E) dE \tag{2.1}$$

$$p = \int_{E_{bot}}^{E_v} g_c(E)(1 - f(E)) dE \tag{2.2}$$

For non-degenerate semiconductors, the Fermi level E_F, i.e. the energy level where the probability of finding an electron is 50%, is more than $3k_BT$ away from both the valence E_V and conduction E_C band edges, and equations (2.1), (2.2) simplify to:

$$n = N_C e^{\frac{E_F - E_C}{k_B T}} \tag{2.3}$$

$$p = N_V e^{\frac{E_V - E_F}{k_B T}} \tag{2.4}$$

where N_C and N_V are the effective densities of states in the conduction and valence bands respectively [18] and k_B is the Boltzmann's constant. For an intrinsic semiconductor, the concentration n of electrons in the conduction band is equal to the concentration p of holes in

the valence band, hence $n = p = n_i$, where n_i is defined as the intrinsic carrier concentration. Using the equations above, the product np gives:

$$np = n_i^2 = N_C N_V e^{\frac{E_V - E_C}{k_B T}} = N_C N_V e^{\frac{-E_G}{k_B T}} \tag{2.5}$$

i.e. the product np of electron and hole concentration, known as the mass-action law, depends on the band gap E_G only and it is independent of the Fermi level E_F. In silicon at room temperature, $n_i = 1.45 \times 10^{10}$ cm^{-3}. Introducing the intrinsic carrier density n_i and the intrinsic energy level E_i, i.e. the Fermi level close to the middle of the band gap for an intrinsic semiconductor, one can reformulate (2.3) and (2.4) as:

$$n = n_i e^{\frac{E_F - E_i}{k_B T}} \tag{2.6}$$

$$p = n_i e^{\frac{E_i - E_F}{k_B T}} \tag{2.7}$$

Extrinsic Semiconductors

The process of doping consists of adding a controlled amount of specific impurity atoms in the crystal to alter its conductivity. This procedure can be performed either during crystal growth or later in selected regions of the crystal, by diffusion or ion implantation, methods discussed in later sections of this chapter. Depending on the type of added material, one obtains n-type semiconductors with an excess of electrons in the conduction band or p-type with an excess of holes in the valence band. For example, in the case of doping with phosphorus, which has five valence electrons, only four electrons are used for the covalent bonding with the neighbouring silicon atoms, while the fifth is only weakly bound and becomes readily available for conduction throughout the bulk of the material. In contrast, in the case of doping with boron, which has only three valence electrons, an additional electron is used to form the covalent bonds.

In the energy band model, the replacement of an atom of the lattice by a different impurity atom leads to the creation of localized energy levels in the band gap. These energy levels may be of donor E_D or acceptor E_A type. In the case of phosphorus $E_C - E_D = 0.045$ eV, hence these states will be almost completely ionised at room temperature, where $k_B T \approx 0.026$ eV and the electrons will be promoted to the conduction band. Similar considerations apply in the case of doping with boron, where holes are created in the valence band when these acceptor states are filled by electrons promoted from the valence band. The term shallow doping is used when the energy gaps between these states and the edges of the valence or

conduction band are small. Conversely, deep defects create states near the middle of the band gap. These are, for example, produced by ionising radiation or specific types of dopants, e.g. Au and Cu in silicon [19], and are usually detrimental to the device.

2.1.4 Generation and recombination

The generation of carriers in a semiconductor may be obtained by different processes, including electrical, optical, or ionising means. For instance, electron-holes (e-h) pairs are created in a semiconductor when photons with energies exceeding the band gap energy of the semiconductor are absorbed and promote electrons from the valence to the conduction band. The inverse process to the generation of carriers is carrier annihilation, or recombination, which takes place via different mechanisms: non-radiative, band-to-band radiative, and Auger recombination.

The non-radiative, or Shockley-Read-Hall (SRH) recombination process is the predominant recombination process for indirect band gap semiconductors such as silicon and germanium. In this process, recombination is accomplished via a deep-level recombination centre in the forbidden gap and the energy of the excess carriers is released via phonon emission. Band-to-band radiative recombination is usually the predominant process occurring in direct band gap semiconductors such as GaAs and InP and is negligible in silicon. In this case, the band-to-band recombination of electron-hole pairs is accompanied by the emission of a photon. The Auger recombination is the dominant recombination process in degenerate semiconductors or under high-injection conditions. Unlike the non-radiative and radiative recombination, which are two-particle processes, Auger band-to-band recombination is a three-particle process that involves two electrons and one hole for n-type semiconductors or one electron and two holes for p-type semiconductors. Figure 2.5 illustrates schematically the three different recombination mechanisms.

The general expression for the net recombination rate under steady-state conditions in the SRH process, assuming a trap state located at E_t with density N_t in the band gap is given by:

$$U = U_n = U_p = \frac{(np - n_i^2)}{\tau_{p0}(n + n_1) + \tau_{n0}(p + p_1)} \tag{2.8}$$

where $\tau_{p0} = \frac{1}{C_p N_t}$ and $\tau_{n0} = \frac{1}{C_n N_t}$ are the minority hole lifetime for an n-type semiconductor and minority electron lifetime for a p-type semiconductor: C_p and C_n are the hole and electron capture coefficients and n_1 and p_1 denote the electron and hole densities respectively when

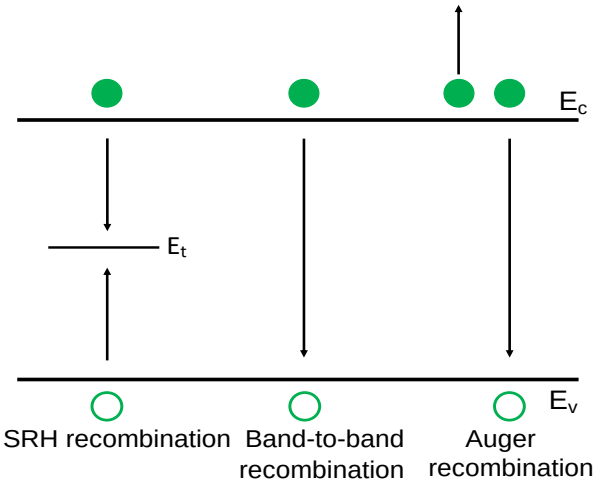


Fig. 2.5 *Schematic of recombination processes.*

the Fermi level coincides with the trap level E_t, as described in [20].
The rate of Auger recombination under steady-state conditions is described by:

$$U_A = C_n(n^2 p - n_0^2 p_0) + C_p(p^2 n - p_0^2 n_0) \tag{2.9}$$

The electron and hole capture coefficients C_n and C_p can be calculated from the process of impact ionisation, which is the inverse process of Auger recombination [20].

2.1.5 Carrier Transport and Basic Semiconductor Equations

The spatial and time-dependent processes of the excess carrier in a semiconductor under non-equilibrium conditions can be analysed using the semiconductor equations. An often-used scheme, which includes drift and diffusion components along with recombination and generation processes is the following:

$$\frac{\partial n}{\partial t} = \frac{1}{e}(\nabla J_n + G_n - R_n) \tag{2.10}$$

$$\frac{\partial p}{\partial t} = \frac{1}{e}(-\nabla J_p + G_p - R_p) \tag{2.11}$$

$$J_n = e\mu_n n F + eD_n \nabla n \tag{2.12}$$

$$J_p = e\mu_p p F - eD_p \nabla p \tag{2.13}$$

$$\nabla F = \frac{\rho}{\varepsilon_s} = \frac{e}{\varepsilon_s}(N_D^+ - N_A^- + p - n) \tag{2.14}$$

The five equations above describe the so-called Drift-Diffusion (DD) model [21], which can be formally derived from the zeroth and first moment of Boltzmann's Transport Equation (BTE), expressing mass and momentum conservation [22]. Equations (2.10) and (2.11) express the continuity equations for electrons (holes) and include the total generation rates G_n (G_p) that may be due to external sources, incoming radiation, or impact ionisation process, and the total recombination rates R_n (R_p). Equations (2.12) and (2.13) describe the electrons (holes) current densities J_n (J_p) which consist of a drift and diffusion component with F being the electric field, μ_n (μ_p) the electron (hole) mobility and D_n (D_p) the diffusivity. The drift component describes the motion of charge due to an electric field whereas the second component describes the diffusion motion arising from differences in carriers' concentration. The last equation of the model, (2.14), is Poisson's equation, which relates the divergence of the electric field F to the charge density, with N_D^+ (N_A^-) the density of ionised donor (acceptor) impurities.

The five equations above in the five unknown quantities (V, n, p, J_n, J_p), with V the electric potential, constitute a system of coupled non-linear equations that must be solved self-consistently. In general, it is not possible to obtain a solution directly in one step, thus a non-linear iteration method is usually required. The two most popular methods used to solve the coupled set of equations are Newton's [23] and Gummel's [24] discretization schemes. It is however possible to reformulate the current equation (2.12) in terms of quasi-Fermi level as:

$$\begin{aligned} J_n &= e\mu_n V_T \nabla n - en\mu_n \nabla V \\ &= en\mu_n (V_T (\frac{n_i}{n}) \nabla (\frac{n}{n_i}) - \nabla V) \\ &= en\mu_n \nabla (V_T ln(\frac{n}{n_i}) - V) \\ &= -n\mu_n \nabla E_{F,n} \end{aligned} \tag{2.15}$$

making use of Einstein's relations $D_n = (\frac{k_B T}{e})\mu_n$ ($D_p = (\frac{k_B T}{e})\mu_p$) with the term in parenthesis being the thermal voltage V_T and where the definition of quasi-Fermi level has been used to write $E_{F,n} - E_{i,0} = -e\phi_n = -e(V_T ln(\frac{n}{n_i}) - V)$. Thus, the current equations (2.12), (2.13) can be written in a linear form (2.15), as the gradient of the quasi-Fermi level, merging the drift and diffusion contribution into a single variable. This simplifies solving the system of equations, as Gummel's iteration scheme has a faster rate of convergence when the coupling between the equations is reduced [25].

As mentioned, the Drift-Diffusion (DD) model (2.10) - (2.14) expresses charge and momentum conservation only. More sophisticated models, like the Hydrodynamic model [26], are obtained from higher moments of BTE and include energy conservation and effects of rapidly changing fields within semiconductor volumes, a typical scenario in very small devices used in advanced processes [27]. In most cases of practical importance in the field of particle and radiation detectors, the internal field does not change abruptly either in time or space, making the DD model sufficiently accurate for the description of the devices.

2.2 The *pn* junction

One of the most important electronic structures is the *pn* junction. Given its importance in the field of electronics, its operating principle is described in practically all books on semiconductor devices, for example [28, 29]. This section describes its basic behaviour from a point of view of its use as a detector, more details can be found in [30].

A *pn* junction, shown in Figure 2.6, is obtained by joining together extrinsic semiconductors of opposite doping, that can be initially thought of as being separated, electrically neutral, and in thermal equilibrium, with electrons and holes homogeneously distributed in their respective volumes.

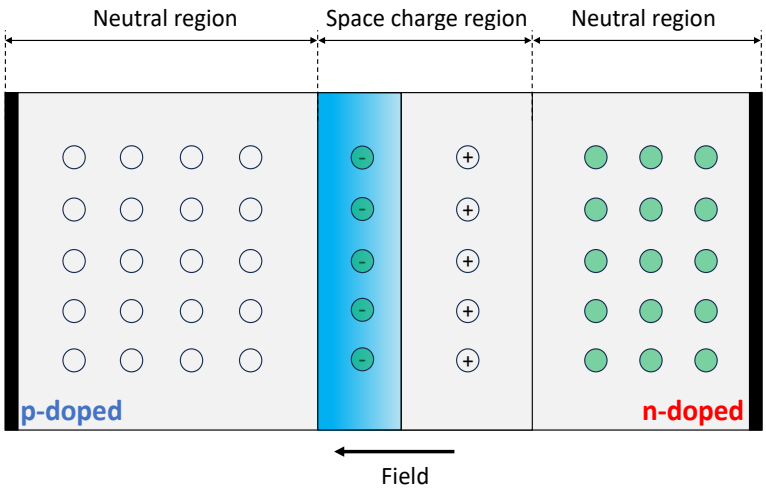


Fig. 2.6 *Schematic of a pn junction in equilibrium. Ionised dopants in the n region and p region show positive and negative charges respectively. The created field prevents further carrier diffusion.*

Once the two parts are brought into contact, electrons from the *n* region will diffuse into the *p* region and holes from the *p* region into the *n* region. Such a process eventually creates

a surplus of fixed negative electric charge in the p region and of positive charge in the n region. The resulting electric field counteracts the diffusion process, eventually stopping it and creating a space-charge (SC) region around the interface boundary, depleted of any mobile charge carriers, in which the ionised dopants sustain the electric field. The transition region of carriers' concentration around the depleted zone is, for most practical cases, very thin compared with the SC region. Hence, assuming an abrupt change between the neutral and the completely depleted SC region, the built-in potential V_{bi} developed across it is given by:

$$V_{bi} = \frac{k_B T}{e} ln(\frac{N_A^- N_D^+}{n_i^2}) \tag{2.16}$$

with N_A^-, N_D^+ the acceptor and donor concentrations, assumed fully ionised and thus equal to the majority carrier concentrations n and p in the doped regions, and n_i the intrinsic carrier concentration. Under the assumption of an abrupt junction, the solution to Poisson's equation provides the width w of the SC region and the value of the maximum electric field F_{max} at the junction as:

$$w = \sqrt{\frac{2\varepsilon_s(N_A + N_D)}{eN_A N_D}(V_{bi} - V)} \tag{2.17}$$

$$F_{max} = \sqrt{\frac{2eN_A N_D}{\varepsilon_s(N_A + N_D)}(V_{bi} - V)} \tag{2.18}$$

with ε_s the permittivity of silicon and V the voltage applied across the junction conventionally assumed positive if applied to the p-side. From (2.17) it is seen that, decreasing the voltage V applied across the junction, i.e. by "biasing" it in the reverse mode of operation, the extension w of the depletion region increases. For a device of thickness d, full depletion is achieved with an applied voltage V_{fd} of:

$$V_{fd} = \frac{eN_A N_D}{2\varepsilon_s(N_A + N_D)} d^2 - V_{bi} \tag{2.19}$$

In the reverse mode of operation, the increased strength of the electric field further prevents the diffusion flow of charge across the junction. The thermally generated carriers

inside the depletion region are the main contributors to the leakage current, which can be approximated by:

$$J_g \approx \frac{en_i w}{\tau} \tag{2.20}$$

with τ being the effective carrier lifetime, which characterizes the generation-recombination process in the SC region. Conversely, applying a positive bias voltage V, i.e. in the forward bias mode, the strength of the field across the junction decreases, which enhances the diffusion flow of carriers. Neglecting the recombination in the depletion region, the current density is now given by:

$$J = J_0(e^{\frac{eV}{k_B T}} - 1) \tag{2.21}$$

with k_B the Boltzmann's constant, T the absolute temperature. The saturation current J_0 is given by:

$$J_0 = en_i^2(\frac{D_p}{N_D L_p} + \frac{D_n}{N_A L_n}) \tag{2.22}$$

where $D_{p/n}$ ($L_{p/n}$) are the diffusion constants (diffusion lengths) for holes and electrons respectively. In the forward region, the current shows an exponential dependence on the applied voltage V, whereas it depends on $\approx \sqrt{V}$ in the reverse region. The junction shows a rectifying property, as it allows the flow of current mostly in one direction.

Approximating the junction to a parallel plate capacitor of width w, i.e neglecting edge effects, the capacitance per unit area is given by:

$$C_j = \frac{\varepsilon_s}{w} = \sqrt{\frac{\varepsilon_s e N_A N_D}{2(N_A + N_D)(V_{bi} - V)}} \tag{2.23}$$

Equation (2.23) shows that the inverse of capacitance square, $\frac{1}{C_j^2}$, varies linearly with the applied voltage V. Thus, a plot of $\frac{1}{C_j^2}$ vs. V yields a straight line, and the intercept with the voltage axis is the built-in potential V_{bi}. In the case of a one-sided junction, if for example $\mathrm{N_A} >> \mathrm{N_D}$, the slope of $\frac{1}{C_j^2}$ yields the doping density N_D.

2.3 Principles of detector operation

A schematic cross-section of a silicon sensor crossed by an ionising particle is shown in Figure 2.7. The external bias voltage V_{bias} polarizes the *pn* junction in the reverse mode, thus creating a large depleted volume with a strong electric field. The absence of charge carriers in this SC region allows the detection of even a small amount of charge generated by the incoming ionising particle. When the particle crosses the sensor it deposits energy in it, creating e-h pairs along its path. The number of pairs created depends on the mechanism of energy deposition, the type of particle, its energy, and the sensor's thickness. Due to the strong electric field in the depleted region, the electrons and holes drift towards opposite electrodes (electrons to the n^{++} and holes to the p^{++} in Figure 2.7), inducing a current signal on both electrodes [31], which starts when the carriers begin to move and ends when the last carrier is collected. The signal current from the sensor is then amplified and filtered by an external front-end (FE) amplifier and fed into the readout electronics (RO).

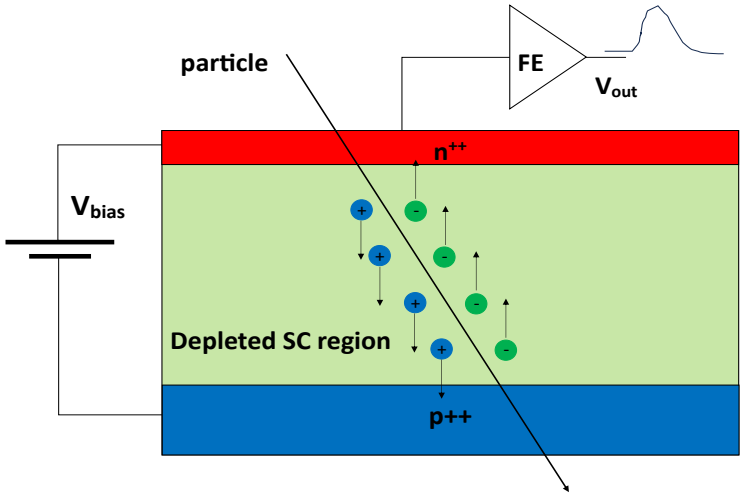


Fig. 2.7 *Cross-section of a silicon detector.*

2.3.1 Energy deposition in silicon

Energy deposition by Charged Particles

When a charged particle crosses a silicon bulk it interacts electromagnetically with the electrons of the atoms, gradually losing its energy along its path. Electrons of the host atom can be raised to a higher energy atomic orbitals, causing atomic excitation, or acquire

sufficient energy to leave the atom, producing an e-h pair. The process of average energy loss by a particle, or stopping power, is described by the Bethe-Bloch formula [32]:

$$\frac{-dE}{dx} = 2\pi N_A r_e^2 m_e c^2 \rho \frac{Z}{A}\frac{z^2}{\beta^2}[ln(\frac{2m_e\gamma^2 v^2 W_{max}}{I^2}) - 2\beta^2 - \delta - 2\frac{C}{Z}] \tag{2.24}$$

In equation (2.24), $2\pi N_A r_e^2 m_e c^2 = 0.1535$ MeV cm^{2}g^{-1}, with $r_e = 2.817 \times 10^{-13}$ cm the classical electron radius and N_A the Avogadro's number, I is the effective ionisation potential, Z the atomic number of the medium, A its atomic weight, ρ the density of the medium, z the charge of the traversing particle, $\beta = \frac{v}{c}$ the normalized velocity of the traversing particle, δ is a high-energy density correction, W_{max} is the maximum energy transfer in a single collision and C is a shell correction.
An example of the energy loss rate as a function of particle momentum is shown in Figure 2.8.

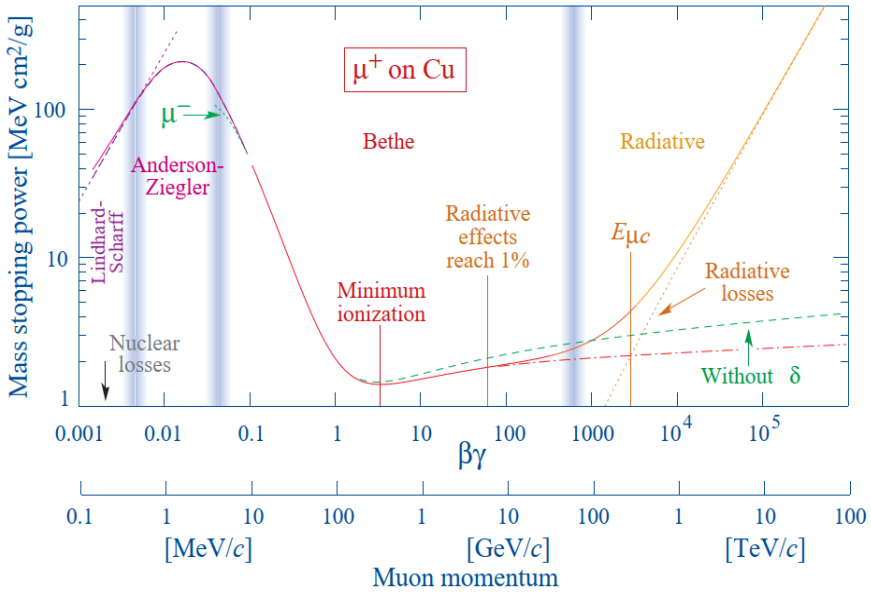


Fig. 2.8 *Mass stopping power for positive muons μ^+ in copper as a function of $\beta\gamma = p/Mc$ over nine orders of magnitude in momentum (12 orders of magnitude in kinetic energy). The solid curves indicate the total stopping power [32].*

At low (non-relativistic) energies, the energy loss rate is inversely proportional to the energy (or velocity squared). With rising energy a minimum in the ionisation loss is reached, at $\beta\gamma \approx 3$ - 3.5. Particles at this energy are called Minimum Ionising Particles (MIP). Higher energies show a relativistic rise of logarithmic characteristics and finally, at the highest energies, the radiative losses due to bremsstrahlung dominate over ionisation.

The statistical fluctuations of energy losses around the average are described by the Landau distribution [32], an asymmetric distribution skewed towards higher energies. The high energy tail of the distribution is caused by rare collisions with small impact parameters, where electrons with energies, sufficient to ionise other atoms, are produced. Because of this asymmetry, the average value of the energy loss does not match the most probable value (MPV), which is lower. The MPV and width of the Landau distribution depend on the sensor thickness, resulting in smaller signals for path length, as shown in Figure 2.9.

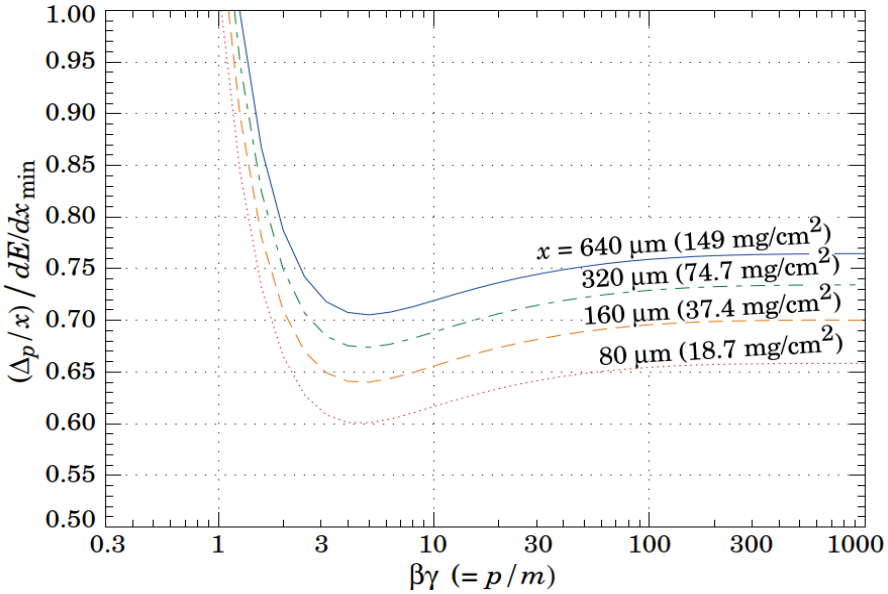


Fig. 2.9 *Most probable energy loss in silicon, scaled to the mean loss of a minimum ionising particle, 388 eV/µm (1.66 MeV g^{-1} cm^2) [33].*

In all semiconductors, only a fraction of the energy loss is used for the creation of e-h pairs. As electronic and lattice excitations are involved in the process of ionisation, the variance σ_N^2 in the number of charge carriers $N = \frac{E}{E_i}$ produced by absorbed energy E where E_i is the mean energy required for e-h pairs creation is reduced by the Fano factor F [18], about 0.1 in silicon and germanium. Thus, $\sigma_N = \sqrt{FN}$, which corresponds to an energy resolution of $\frac{\sigma_E}{E} = \sqrt{\frac{FE_i}{E}}$.

Energy deposition by Electromagnetic Radiation

The schematic diagram of the basic process of energy deposition by electromagnetic radiation, upon which photodetectors are based, is shown in Figure 2.10. The energy of the absorbed photon is used to promote the electron from the valence into the conduction band. If the photon energy is higher than the band gap energy E_G, the electron will be promoted into one of the empty states of the conduction band, leaving a hole behind in the valence band. For indirect band gap semiconductors, like silicon, additional energy is needed to supply momentum transfer to the electron, or, for a direct transition without momentum change, a larger energy than E_G is required. The generated electron and hole will subsequently move towards the band-gap edges, emitting energy in the form of lattice vibrations (phonons) or lower energy photons. Absorption of photons with energies below E_G is possible if there are local states in the band gap due to lattice imperfections.

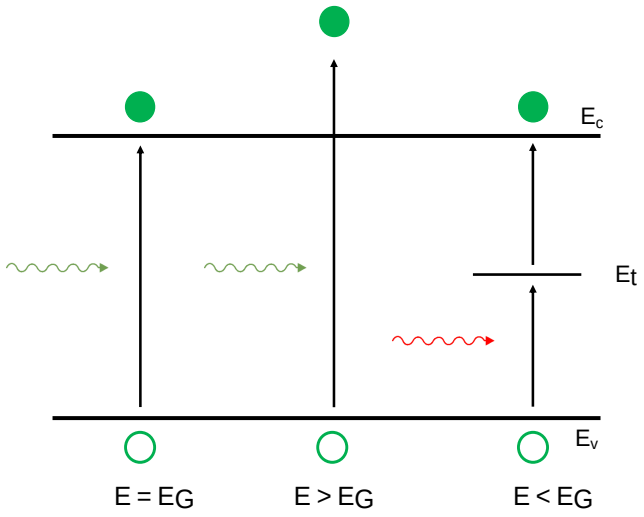


Fig. 2.10 *Diagram of generation of electrons and holes by absorption of photons of energies $E = E_G$, $E > E_G$, and $E < E_G$. E_c is the bottom of the conduction band, E_v the top of the valence band.*

Figure 2.11 shows the contributions to the photon cross-section in silicon. At low energies, the photoelectric effect dominates the interaction cross-section, although Compton and Rayleigh's scattering also contributes. The photoelectric cross-section is characterized by discontinuities (absorption edges) as thresholds for photoionisation of various atomic levels are reached. Compton scattering, i.e. the elastic scattering of a photon on a quasi-free electron, contributes at higher energies, up to a few MeV's. Finally, the electron-positron

pair production by a photon in the field of a nucleus dominates at the highest energies, with a threshold energy of 2 m_e, i.e. 1022 keV.

Tests on silicon photodiodes [34] show that for photon energies below 4 eV one e-h pair is created per incident photon, whereas above $\approx$ 1.5 keV a mean energy of 3.67 eV is needed for pair creation. This is approximately three times larger than the band gap of 1.12 eV because momentum conservation requires phonon excitations.

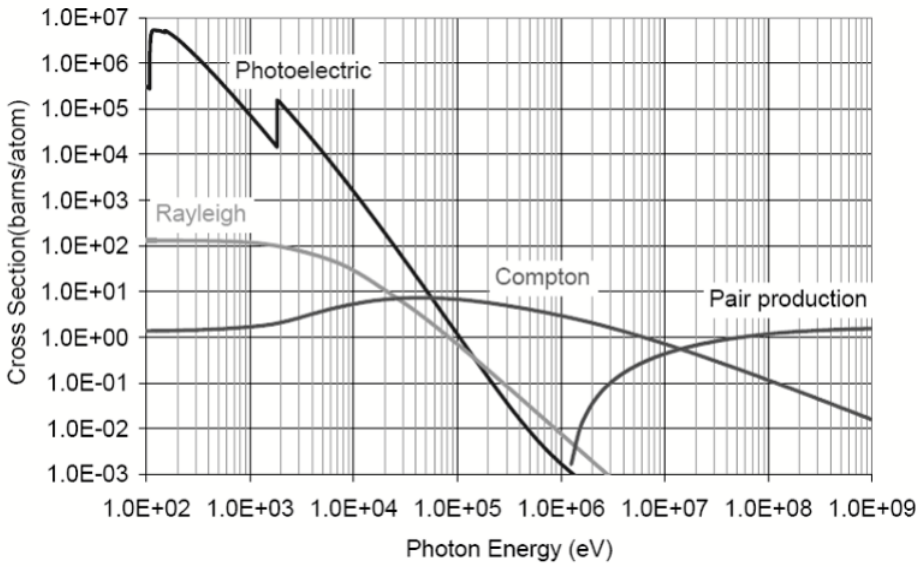


Fig. 2.11 *Photon interaction cross-sections in silicon as a function of photon energy [35].*

2.4 Semiconductor Fabrication

Semiconductor detector technology is to a large extent based on planar technology, which was originally developed for microelectronics and later adapted to detector-grade silicon processing [36]. The requirement of maintaining high-quality semiconductor properties for detector fabrication initially prevented planar processing, with its high-temperature processing steps, from being used in detector technology. Nowadays the complexity and integration of sensor technology into the modern microelectronics processes implies that many of the methods used in microelectronics have been introduced into detector processing and fabrication. Figure 2.12 shows an example of a process flow used in the fabrication of silicon devices, where many processing steps, sometimes requiring precision at the atomic level, are used.

This section describes the basic steps of device fabrication used in detector and microelectronics processing, focusing on those implemented in the design and simulation of this project, which are described in Chapter 4.

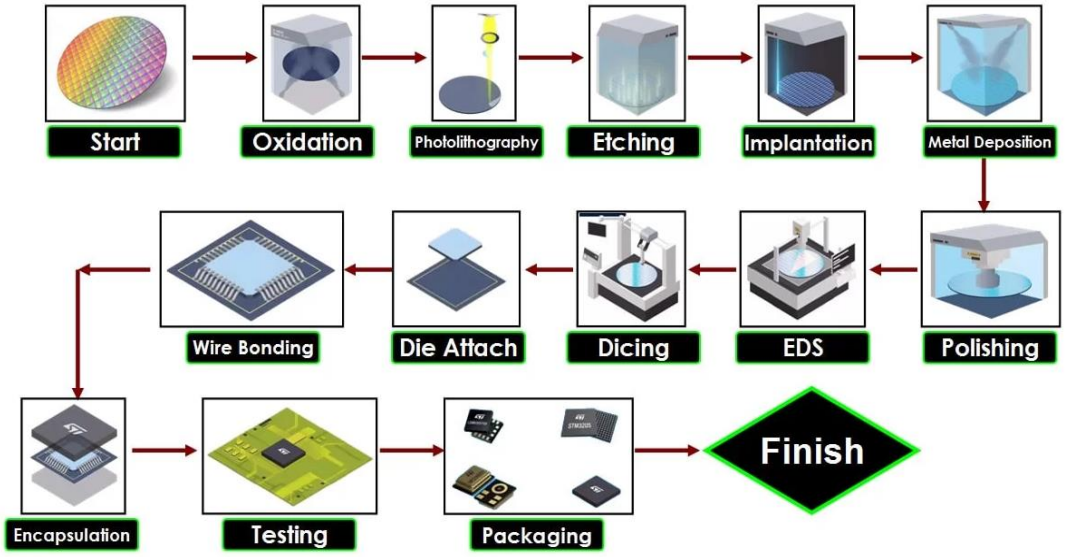


Fig. 2.12 *Example of a silicon process flow for detector fabrication. Source: http://www.electronicsandyou.com.*

2.4.1 Crystal growth

The techniques of growing silicon wafers have been developed and perfected in the microelectronics industry, where a purity of 99.9999999% or higher is usually required [37]. The need to keep up with consumer demand has had the effect of increasing the wafer size [38] to reduce the relative cost per device. The starting point to grow a single-element semiconductor is a melted high-purity polycrystalline material and a seed crystal of defined orientation, which determines the orientation of the full-grown ingot.

For microelectronics applications, the Czochralski (CZ) process is usually employed [39], which consists of keeping the melt in a quartz crucible and slowly pulling up and rotating the seed crystal. Vertical speed and rotation rate are important parameters in defining the characteristics of the final ingot. This growing method is not usually suited for detector-grade material as many impurities, originating mostly from the crucible, are left in the crystal. This reduces the minority carrier lifetime, in turn decreasing the diffusion length and affecting the charge collection efficiency[1]. Detector-grade silicon requires an almost defect-free crystal structure and as few as possible foreign atoms, except for dopants that might be added to change the intrinsic material properties. The most common silicon detector material is produced using the float zone (FZ) process [40], which results in a very low impurity concentration in the crystal, as the melt does not come into contact with the quartz crucible. In the FZ process a high-purity polysilicon rod, vertically suspended in a vacuum or an inert gas, is melted in a narrow horizontal zone using high-frequency induction. This zone is slowly moved upwards from the seeding crystal on the bottom and the material at the bottom of the molten zone solidifies into a single crystal. During this process many of the foreign elements, having a small segregation coefficient (solubility) in silicon, are pushed upwards towards the top of the rod. Repeating this zone-melting procedure several times results in very pure single crystals. One limitation of the FZ technique is the difficulty in growing large-diameter crystals, due to the need to grow the ingot suspended to avoid contamination.

The obtained single crystal ingots are subsequently cut into thin wafers and their surface is lapped and polished, to make them ready for subsequent processing steps.

2.4.2 Oxidation

The purpose of oxidation is twofold:

[1]The minority carrier lifetime τ is the average time needed for a decrease of 1/e of the minority carrier density. For very high purity silicon is of the order of 1 ms, reducing to less than 1 ns for heavily doped or irradiated silicon. It is related to the diffusion length L_d by $L_d = \sqrt{\tau D}$, where D is the diffusion coefficient. A short lifetime reduces the distance over which the carriers can travel before recombination, reducing the charge signal collected by the electrode.

- to provide electrical isolation among devices;
- to act as a barrier, either as passivation of the silicon surface or as a mask against dopant diffusion and implantation.

Due to the conductivity of a silicon wafer, two devices located close to each other might be electrically connected via the substrate. The oxidation process is used to prevent separate devices from interacting with each other through the substrate. The most common approach to device isolation is through Shallow Trench Isolation (STI), depicted in Figure 2.13, which uses a small silicon area of the wafer and allows high device density.

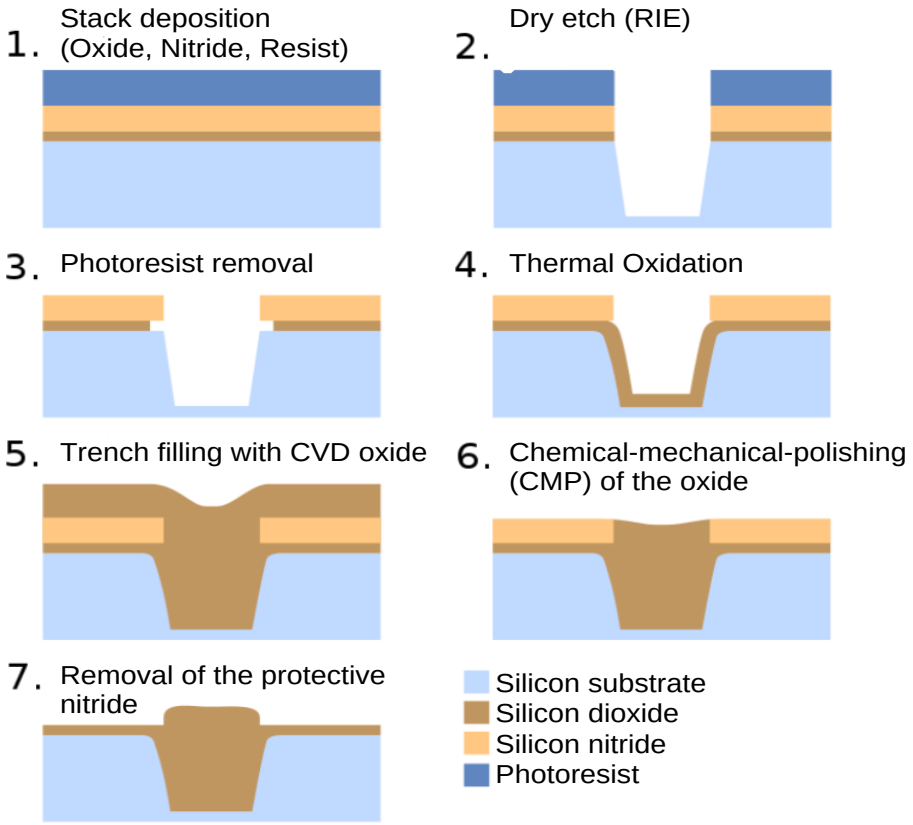


Fig. 2.13 *STI fabrication process of modern integrated circuits in cross-sections (from Wikipedia).*

Passivation of detectors and other silicon semiconductor devices consists of covering a large fraction of the outer surface with SiO_2, which saturates most of the open bonds of silicon in the Si–SiO_2 transition region, leaving only a small amount (approximately 10^{10} - 10^{11} cm^{-2}, to be compared with the surface density of atoms at the crystal boundaries, of the order of 5×10^{15} cm^{-2}) of dangling bonds. The oxide is also used as a mask against dopant diffusion, as most dopants used in the semiconductor industry have a lower diffusion rate in

SiO_2 than in silicon, and block ion implantation in areas covered by the oxide. The use of a relatively thin ($\approx$ 10 - 100 nm) oxide, called screen or sacrificial oxide, grown onto the silicon surface is also employed to increase the scattering of ions of dopants during the process of implantation, to minimise the effect of channeling [41]. Oxidation is the highest-temperature process and is performed at the beginning of device fabrication, immediately after a surface cleaning process, by heating silicon in furnaces at temperatures of $\approx$ 1000 °C. The high temperature renders many of the impurities and crystal defects mobile, which implies that great care has to be taken to avoid introducing impurities that could affect the properties of the semiconductor. The 'dry' oxidation is performed in an O_2 atmosphere and produces the best oxide quality, with the least concentration of defects, at the expense of a slow growth rate. The 'wet' oxidation is performed in an H_2O atmosphere and produces a faster-growing oxide of lower quality compared to the dry process. In both cases, oxygen diffuses through the layer of oxide already grown and reacts with silicon at the Si–SiO_2 interface. The process slows down with increasing thickness of the oxide layer and is strongly temperature-dependent.

The dry and wet oxidation processes require the availability of silicon for the reaction to occur, thus they cannot be used to fabricate interlayer dielectrics (ILD) between metal layers in modern microelectronics devices. In such cases, chemical vapour deposition (CVD) [42] can be used to deposit SiO_2.

2.4.3 Photolithography

Photolithography is used to transfer the pattern from a mask (usually chrome on glass) onto the wafer surface. The semiconductor wafer is covered with a thin (1 μm or less) layer of photosensitive resin, obtained by dropping a small amount of liquid resin on a fast-spinning wafer. After drying, the resin is illuminated, with UV light through the mask, which is either pressed against the wafer (contact illumination) or held at a close (10 - 20 μm) distance (proximity illumination). Contact illumination allows higher resolution and smaller feature sizes compared to proximity illumination but has the disadvantage of defect introduction by particles. The resin can also get stuck to the mask producing faults in consecutive illuminations. In advanced microelectronics processing, projective illumination, in which an optical system transfers the structures from the mask onto the wafer, is routinely used. After illumination, the resist is developed. For commonly used positive photoresists the illuminated resist is dissolved using a chemical solvent, a process called stripping, so that the wafer surface becomes exposed in these regions and ready for the subsequent processing steps of etching or doping. The same process of photolithography is used to produce masks, illuminating the photoresist on top of chrome-covered glass through a collimator, which is moved step by step over the glass to produce the desired mask topology. Using a focussed

electron beam to directly write into the photoresist allows even higher quality masks and the production of finer structures.

For the production of double-sided structures, the front and backside photolithography must be aligned with respect to each other. This is accomplished by using infrared light to look through the wafer or by the storage of mask images against which the other side can be aligned once the direct view is blocked by the wafer put in front of the mask.

2.4.4 Etching

After photolithography, the process of etching follows. A pattern of oxide from the wafer is removed, to create exposed regions of silicon for the subsequent processing steps. The etchant is applied to the wafer as a whole and acts on all regions not protected by the resist. The etchant may be liquid (wet etching) or gaseous (ion etching).

Wet etching is commonly used when requirements on structure size are not too stringent and etching isotropy, i.e. equal etching along the vertical and horizontal direction with respect to the surface of the wafer, which is inherent in the process of wet etching, can be tolerated. This leads to under-etching below the resist (typically comparable to the photoresist thickness of 1 - 2 μm).

Dry etching uses ions rather than chemical treatment, where the ions typically used are reactive gases such as oxygen, boron, and fluorocarbons [43]. It allows high anisotropy in etching, i.e. higher vertical etch rate compared to horizontal etch rate, but it might lead to some oxide radiation damage, which has detrimental effects in terms of increased leakage and charge trapping.

2.4.5 Doping

Dopants typically used for silicon are boron (p-type), phosphorus, and arsenic (n-type). To change the doping in selected regions of the wafer surface, exposed by etching, a process of diffusion or implantation, or a combination of both, can be used.

Diffusion consists of having the dopant atoms move from the open semiconductor surface into the semiconductor volume, which is held at a high temperature during the process. The dopant concentration is either controlled at the surface during the process or a defined amount of dopant is deposited at the surface at the beginning of the procedure. The process of implantation is performed by bombarding the wafer with a (defocussed) ion beam of defined energy. Ions will be stopped in the semiconductor in the regions that have been opened in the preceding photolithographic and etching steps. In the other regions, the ions will be stopped in the insulating layer(s) above the semiconductor, unless their energy is high enough to

penetrate the semiconductor. Oxide and photoresist may also be used as a stopping layer for the implantation. The implantation of the dopants in the crystal is not sufficient to obtain the desired electrical performance of the doped regions: the dopant atoms may, for instance, not be in a regular lattice position, which is needed for being electrically active, but in an interstitial site in the crystal structure. Additionally, the process of ion implantation creates some damage to the crystal structure of silicon. Subsequent heat treatment of the crystal is always necessary to bring the dopant atoms into regular lattice positions and to regrow the lattice.

Doping via implantation allows the production of buried layers of dopants by choosing high implantation energies and in general the doping concentration and implantation depth can be controlled with an accuracy of a few % [44]. A relatively small implantation depth (typically less than 1 μm) can be reached with the standard energies used in semiconductor manufacturing, i.e. a few 100s keV. Therefore a process of drive-in diffusion may be needed to allow the dopants to reach a deeper depth in the silicon. Doping via diffusion introduces a smaller amount of defects than implantation but it is less flexible in terms of achievable doping profiles. It also poses safety concerns, as it usually requires the use of highly toxic gases.

For certain applications, including imaging detectors and LGAD devices as discussed in this thesis, a semiconductor wafer with an epitaxially grown layer of different doping is used as a starting material, onto which the process of doping is subsequently performed. The process of epitaxial growing, i.e. growing of the crystal in the correct lattice structure on top of a single-crystal wafer, is performed at high temperatures in a gaseous atmosphere.

2.4.6 Thermal processing

Besides the thermal treatments inherent in some of the processing steps described before, which happen at high temperatures, additional treatments are usually needed in the manufacturing process of silicon devices.

After ion implantation, the dopant atoms will generally not be located on a regular lattice location but at an interstitial position while distorting or destroying the crystal lattice in its proximity. Damage to the lattice will also be done by silicon recoil atoms produced in collisions with the dopant atoms or ions. Thus, a thermal treatment is needed to place the dopant atoms into regular lattice sites and repair the damage that has occurred during implantation. Due to the incomplete chemical matching on the Si–SiO_2 interface, dangling bonds are left at the interface, causing positive charges and corresponding flat-band voltage

shifts in Metal Oxide Semiconductor (MOS) transistors[2]. These dangling bonds can partially be bound with hydrogen. This is accomplished by heating the device in a partial hydrogen atmosphere (hydrogen annealing), which enables hydrogen to diffuse through the oxide towards the interface, thus leading to a saturation of the dangling bonds.

Thermal treatment is usually needed after the deposition of the conductor to ensure good electrical contact between deposited metal (e.g., aluminium) and silicon. In the case of aluminium, this is done by heating the device to a temperature somewhat below the melting point of the metal, i.e. to about 400 °C.

2.4.7 Metallization

Deposited metal layers are used for low-resistivity interconnection and for supplying connections, like bond pads, to the outside. A commonly used metal is aluminium. However, copper is often used in modern microelectronic processes due to its lower resistance.

Aluminium is deposited either by evaporation or by sputtering, performed in a vacuum. Evaporation is done by heating in a tungsten boat. Sputtering is performed by bombarding an aluminium target with high energy ions, usually Ar, which sputters off aluminium atoms from the target which then deposit on the wafer substrate to form a thin film. The use of copper for metallization requires extra care, as it introduces deep defects in the band gap of silicon. More sophisticated metallization processes, like the Dual-Damascene, are used for this purpose [45].

Aluminium has the advantage of excellent electrical conductivity, but it has the drawback of a low melting temperature (660 °C), which creates problems when more than one metallization layer is needed. In such cases, an insulating layer formed at low temperatures between the two metallization layers or a metal layer of tungsten which has a higher melting point can be used.

[2]The flat-band condition is defined as when the charge carrier concentration in the homogeneously doped bulk semiconductor is uniform up to the boundary with the oxide and the electric field strength throughout the semiconductor is zero. It depends on the bias voltage applied and on the density and sign of trapped charge at the Si–SiO_2 interface.

Chapter 3

Timing detectors

As discussed in a previous chapter, Ultra-Fast Detectors (UFD) able to achieve a time resolution of 30 ps with a spatial resolution of the order of 10 μm [46] can bring many advantages to HEP applications. Most of these devices rely, for their operation, on the avalanche or impact ionisation process. In this chapter, timing detectors and their implementation using Low Gain Avalanche Diode (LGAD) are introduced. Their operating principle, based on impact ionisation, is discussed, and techniques for their read out are presented.

Section 3.1 introduces timing detectors and the characteristics of a time tagging detector along with parameters that describe and define their performances are discussed. Sensors with internal gain, based on charge multiplication, are described in Section 3.2 along with a description of how these sensors can be used for timing applications. The impact ionisation process, upon which the mechanism of charge multiplication relies, is briefly reviewed in Section 3.3. Section 3.4 introduces the LGAD device. The operating principle of LGAD and noise sources that affect the overall performances are discussed, along with some technical aspects related to their implementation. The front-end readout (RO) circuitry for LGAD, based on transimpedance amplifiers (TIA) that can be used to operate in fast TIA or charge sensitive amplifier (CSA) mode, is described in Section 3.5.

3.1 Timing Detector Introduction

Figure 3.1 shows the main components of a timing detector. The sensor, modelled as a capacitor C_d with a current source I_d in parallel, is read out by a preamplifier that amplifies and shapes the signal. The preamplifier output is then fed into a voltage comparator of threshold V_{th} to determine the time of the hit. The comparator output is digitized by a time-to-digital converter (TDC).

To accurately determine the time of passage of a particle both the readout electronics and the sensor need to minimise possible causes of uncertainty. In particular, the sensor should show a current signal that scales with the amount of energy released by the particle without distortion, and the readout should minimise the input and output noise affecting the determination of the time of the hit.

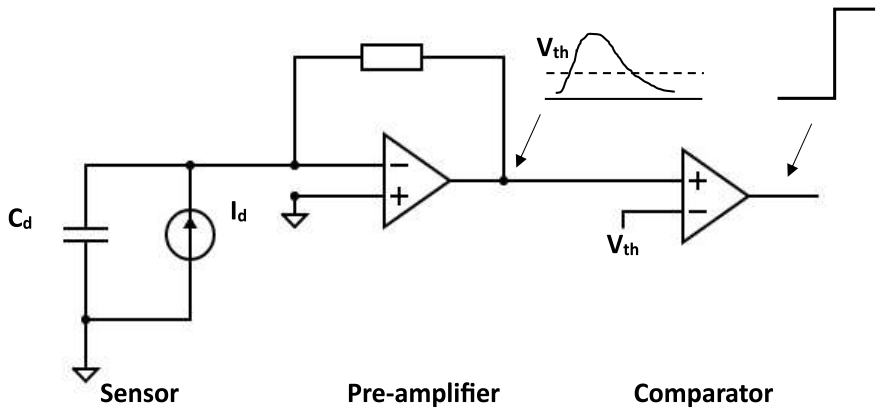


Fig. 3.1 *Schematic diagram of a timing detector.*

The most relevant factors contributing to the temporal resolution σ_t are:

$$\sigma_t^2 = \sigma_{ionis}^2 + \sigma_{dist}^2 + \sigma_{jit}^2 + \sigma_{TDC}^2 \tag{3.1}$$

where:

- σ_{ionis} is due to the irregular shape in the sensor signal, arising from non-uniform energy deposition by the impinging particle;
- σ_{dist} is due to the distorted shape of the electric and weighting field inside the sensor, which affects the drift velocity of the carriers and the signal formation;
- σ_{jit} is due to the jitter caused by either the sensor or the electronics;

- σ_{TDC} is due to the quantization of the time bin used by the TDC.

These factors will be described in the following subsections.

3.1.1 Ionisation

A charged particle crossing a silicon sensor generates along its path e-h pairs, the density of which varies on an event-by-event basis. This has the effect of creating a non-uniform current signal (Landau noise effect) and changing the signal amplitude (time walk effect). The Landau noise, due to the statistical nature of the ionisation process, represents the ultimate limit to achievable signal uniformity, as the irregularities in the signal shape degrade the achievable time resolution. To mitigate the effect of Landau noise thin sensors can be used, as their signal is less sensitive to fluctuations. The intrinsic limit of Landau noise is about 25 ps in a 50 μm-thick UFD, while it is about 60 ps in a 300 μm-thick one [47]. The time walk term arises since larger signals cross a fixed threshold earlier than smaller ones, as shown in Figure 3.2.

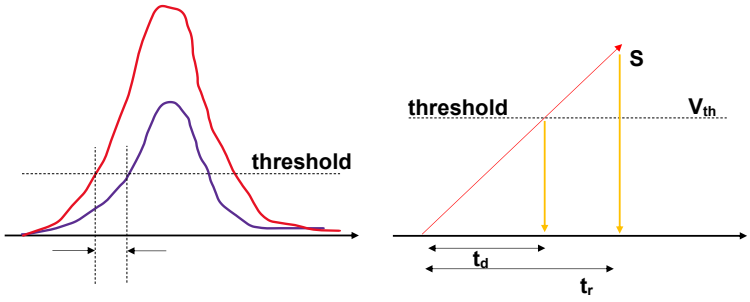


Fig. 3.2 *Left side: Signals of different amplitude cross a fixed threshold at different times, generating a delay t_d on the firing of the discriminator that depends upon the signal amplitude. Right side: a linear signal, with amplitude S and rise time t_r crosses the threshold V_{th} with a delay t_d.*

We can use the linear approximation to evaluate the time t_d for a signal with amplitude S and rise time t_r to cross the threshold V_{th}. As $t_d/t_r = V_{th}/S$, the time when the signal crosses the threshold is $t_d = t_r V_{th}/S$. The time walk contribution, σ_{tw}, to the temporal resolution is the RMS of t_d:

$$\sigma_{tw} = \left[\frac{V_{th}}{\frac{S}{t_r}}\right]_{rms} \propto \left[\frac{N}{\frac{dV}{dt}}\right]_{rms} \tag{3.2}$$

where the slew rate (SR) is defined as $S/t_r = dV/dt$, once we write the threshold V_{th} as a multiple of the system noise N. The time walk is therefore minimised by systems with low noise and high slew rate. Time walk mitigation techniques can also be implemented in the readout electronic, using the constant fraction discriminator technique [48].

3.1.2 Distortion

In a silicon sensor, the shape of the current signal can be calculated using Ramo-Shockley's theorem [49]:

$$i(t) \propto e v_d F_w \tag{3.3}$$

where e is the unit charge, v_d is the drift velocity of carriers, and F_w is the weighting field. Equation (3.3) shows that for accurate timing, v_d and F_w need to be as constant as possible to generate very uniform current signals $i(t)$ across the sensor. The drift velocity v_d of the charge carriers must be uniform everywhere in the active volume of the sensor as any non-uniformity induces variations in the signal shape that depend upon the particle interaction position, affecting the overall time resolution. This can be obtained by an electric field high enough to saturate the charge carriers' drift velocity. In silicon, a field of ≈ 30 kV$\cdot$ cm^{-1} or higher saturates the electron's drift velocity [50, 51]. The weighting field F_w expresses the coupling between a charge positioned in a given location and the readout electrode. If this coupling changes along the electrode pitch then the signal shape depends on the hit position of the particle, which spoils the time resolution. A wide strip geometry, where the strip width is similar to the strip pitch, has a weighting field along the pitch almost constant, whereas, in the case of thin strips, where the strip width is much smaller than the strip pitch, the weighting field is concentrated solely underneath the strip implant, as shown in Figure 3.3. Therefore, the time resolution can be affected to an extent that depends on the effective value of (3.3). An additional factor that can contribute to worsening the time resolution is the effect of the electrode. In an UFD the signal propagates from the impact point to the readout electronics. If the propagation of the signal happens in the electrode's well, a non-zero delay occurs due to the well's resistivity, which might affect the time resolution. Therefore, UFDs are often fully metalized since propagation in a metal does not cause significant delay.

In summary, to minimise the distortion effects, the sensor is required to have a geometry as close as possible to a parallel plate capacitor, with an implant pitch larger than the sensor thickness and implant width very similar to the pitch, to guarantee uniform electric and

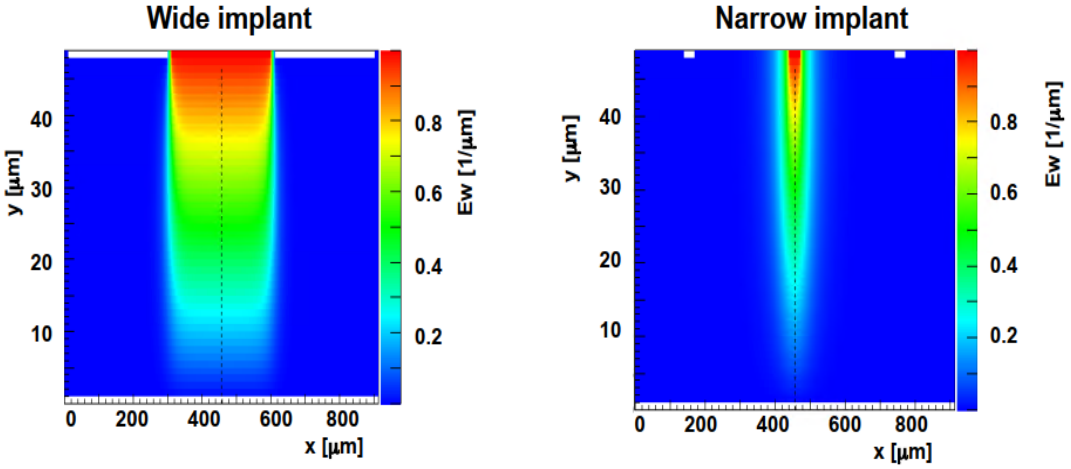


Fig. 3.3 *The weighting field for two electrode configurations: (left) wide implants, and (right) narrow implants. From [46].*

weighting fields. Also having a metalized surface is beneficial as it avoids position-dependent delays.

3.1.3 Jitter

The presence of noise on the signal, coming from the sensor, the preamplifier electronics, or on the discriminator threshold V_{th}, affects the firing time of the comparator, spoiling the time resolution. This effect is directly proportional to the noise N and inversely proportional to the slope of the signal around V_{th}, as sketched in Figure 3.4:

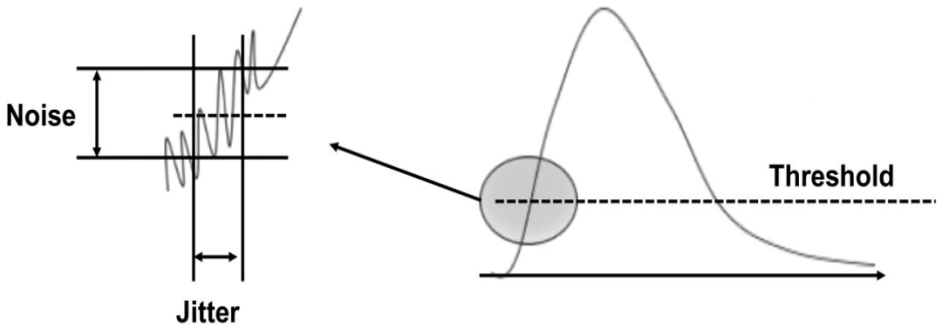


Fig. 3.4 *The jitter is the uncertainty in time tagging due to the noise superimposed onto the signal, which causes the early or late firing of the comparator.*

Assuming a constant slope of the signal output, the slew rate dV/dt can be approximated by the ratio of signal amplitude S over the rise time t_r:

$$\sigma_{jit} = \frac{N}{\frac{dV}{dt}} \propto \frac{t_r}{\frac{S}{N}} \tag{3.4}$$

The minimisation of the jitter term requires a high signal-to-noise ratio S/N and high slew rate SR, which can be achieved by optimizing the sensor and the readout. High signal S and slew rate reduce the amplifier requirements, contributing to reducing the noise. On the other hand, a high slew rate sensor requires a high bandwidth amplifier, which increases the noise. For example, in the HGTD, the jitter term is expected to be around 25 ps at 10 fC of input charge signal [10].

3.1.4 TDC

The TDC records the time of the discriminator firing in a time bin of finite width ΔT, given by the TDC least significant bit. This process contributes a time uncertainty equal to $\Delta T/\sqrt{12}$. Due to the fine binning of the TDCs commonly used in high-energy physics experiments, for example in the HGTD where the bin width of the internal TDC is 20 ps [10], the σ_{TDC} term can often be neglected.

3.2 Sensor with internal gain

To understand the potential advantages offered by a sensor with internal gain in terms of timing resolution, one can look first at the mechanisms of signal formation in a standard PIN diode[1], using Ramo-Shockley's theorem. Assuming a pixel width W much larger than the sensor thickness d, the weighting field is almost constant $F_w \propto 1/d$. If the electric field is high enough to saturate the drift velocity to v_{sat}, the initial current following charge injection can be written as:

$$I_{max} \propto Ne\frac{1}{d}v_{sat} = n_{e-h}de\frac{1}{d}v_{sat} = n_{e-h}ev_{sat} \tag{3.5}$$

[1] A PIN diode consists of an undoped intrinsic semiconductor region sandwiched between the p-type and an n-type semiconductor of a *pn* junction. In contrast to a normal *pn* junction the PIN diode has inferior rectifying characteristics but it operates better as a fast switch and photodetector. Standard silicon sensors are often identified as PIN diodes, even if the bulk consists of lightly doped silicon.

where $N = n_{e-h}d$ is the number of e-h pairs, assuming a uniform charge creation of e-h pairs per unit length. This result shows that in PIN silicon sensors the peak current does not depend on the sensor thickness. Typical values for a Minimum ionising Particle (MIP) crossing a silicon PIN diode result in $I_{max} \approx 1.5\ \mu$A. The rise time of the current signal can be estimated by differentiating (3.5):

$$dI_{max} \propto e\frac{1}{d}v_{sat}dN \approx e\frac{1}{d}v_{sat}n_{e-h}v_{sat}dt \propto n_{e-h}ev_{sat}^2\frac{1}{d}dt \tag{3.6}$$

yielding:

$$\frac{dI_{max}}{dt} \propto v_{sat}^2\frac{1}{d} \tag{3.7}$$

Recalling that, from section 3.1, a high slew rate is crucial to obtain excellent time resolution, it can be seen that the time precision of sensors with no internal gain essentially is limited by the saturation velocity and the maximum value of the weighting field F_w. On the contrary, in a sensor with an internal charge gain G, the number of electron-hole pairs is GN, and therefore I_{max}, dI_{max}, the peak signal amplitude, and the slew rate all increase. As shown in section 3.1 this improves the time resolution. The impact ionisation process, which leads to charge multiplication and hence internal gain, is described in the next section.

3.3 Impact Ionisation

Impact ionisation is a three-particle generation process, as shown in Figure 3.5. In such a process, free carriers are accelerated by a strong electric field until they gain enough energy to promote, after a collision, electrons from the valence to the conduction band creating additional e-h pairs. The scattering process requires energy and momentum conservation, therefore the minimum threshold energy of the ionising particle must be at least as large as the energy band gap. However, momentum conservation usually implies an energy larger than the band gap [52]. Such threshold energy depends on the band structure of the semiconductor and it usually corresponds to electric fields of the order of 10^4 V$\cdot$ cm^{-1} in low band gap materials up to values over 10^5 V$\cdot$ cm^{-1} at room temperature.

In silicon, using the approximation of two parabolic conduction and valence bands and assuming the same effective mass for electrons and holes, a well-known approximation for the impact ionisation threshold energy E_i is the 3/2 bandgap rule, i.e. $E_i = \frac{3}{2}E_G$, with $E_G = 1.12$ eV at T = 300 K [53]. If the secondary e-h pair also acquires high enough energy

to create additional pairs an avalanche effect is triggered and the carrier density increases exponentially, usually leading to an avalanche breakdown.

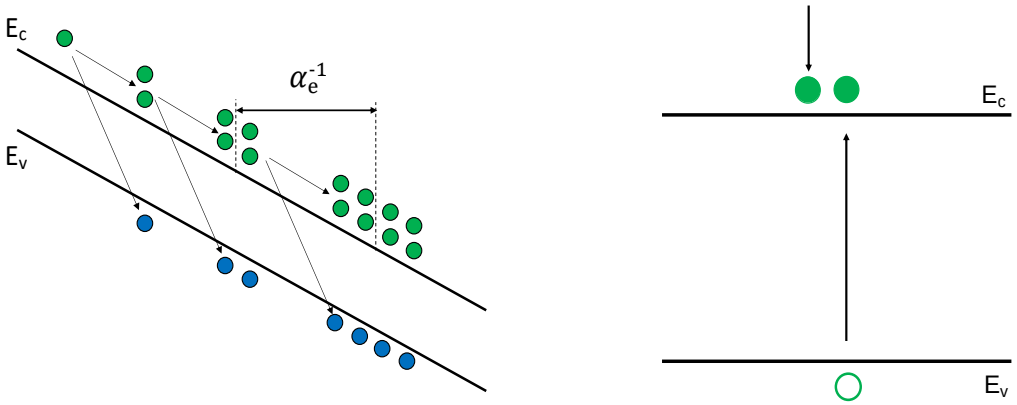


Fig. 3.5 *Schematic of the electron impact ionisation process. Left: The electron, accelerated along an average distance α_e^{-1}, undergoes a collision and its excess energy produces a new e-h pair. Consecutive collisions can ignite an avalanche. Right: The band diagram of the impact ionisation process, essentially the inverse of the Auger recombination, as shown in Figure 2.5.*

Impact ionisation theories give the ionisation coefficients α_e and α_h, which describe the number of e-h pairs generated per unit distance travelled by a carrier between two collisions, in analytic forms and/or plots versus parameters, like the mean free path and the threshold energy. In most cases, such theories assume:

- fields spatially uniform and time-independent;
- long avalanche regions;
- carriers rapidly reach a steady spatial state under the competing influences of the field and phonon collisions.

Under such conditions, the ionisation coefficient depends only on the strength of the electric field, i.e. it is a local function of the field F, independent of the carrier position, and provides the basis for simulations based on local variables, like drift-diffusion or the hydrodynamic scheme. The ionisation coefficient is the reciprocal of the average distance $\bar{x}$ along the direction of the field between ionising collisions, with the assumption that:

$$\bar{x} \gg \frac{E_i}{eF} \tag{3.8}$$

where the RHS of (3.8) represents the minimum distance required by a carrier to create an e-h pair in a uniform field. As the field F increases, this distance becomes smaller than the distance between scattering events, making the ionisation coefficient position-dependent, i.e. a non-local variable [54]. Assuming that the above locality conditions are satisfied, a typical relation for the ionisation coefficient on the local value of electric field F, stemming from theory and validated experimentally, takes the form [55] :

$$\alpha_\nu(F) = \alpha_\nu^\infty e^{\frac{-F_{th}^\nu}{F^\delta}} \tag{3.9}$$

where ν represents the carrier (electrons or holes) , F_{th}^ν is the critical field for ionisation, α_ν^∞ is the asymptotic limit of the ionisation rate and δ is the exponent of the field F, ranging between and 1 and 2. The parameter values in the formula have well-defined physical meanings but are often determined from fits to the experimental data. The temperature dependence, such as the increase in ionisation due to the reduced phonon scattering at lower temperatures when the average distance between ionising collisions is smaller, is accounted for in the parameters used in (3.9).

The formula (3.9) essentially consists of the product of the limit α_ν^∞ and the probability for a carrier to gain energy equal to the ionisation threshold which shows an exponential dependence on the field strength. Such probability is proportional to the energy distribution function of the carriers and can be obtained formally by solving the Boltzmann transport equation for the momentum distribution function [56]. From (3.9) it is seen that, as the field F increases, the probability of reaching the ionisation threshold tends to 1 hence the $(\alpha_\nu^\infty)^{-1}$ must tend to $\frac{E_i}{eF_{ava}}$, the minimum distance required to create an e-h pair. Initial theoretical works suggested the value of coefficient δ in (3.9) to be 1 in [56] or 2 in [53], corresponding to the two cases of energy gained by carriers between phonon collisions being less or more than the energy lost per collision. Further first-principle modelling from the Boltzmann transport equation (BTE) in [57] recognized the values of 1 and 2 for δ as limiting cases for high and low fields. No closed-form solution to this approach exists, but the related ionisation coefficients are usually expressed in a polynomial form like:

$$\alpha_\nu(F) = \frac{e^{C_o(r)+C_1(r)x(F)+C_2x(F)^2}}{\lambda} \tag{3.10}$$

where $r = \frac{E_r}{E_i}$ is the ratio between the average energy loss per collision and the ionisation energy and $x = \frac{E_i}{e\lambda F}$, with λ the mean free path between collisions with high energetic

phonons. The full expression of the polynomials C(r) is reported in [58]. Figures 3.6 and 3.7 show examples of α_e and α_h in silicon using different ionisation models.

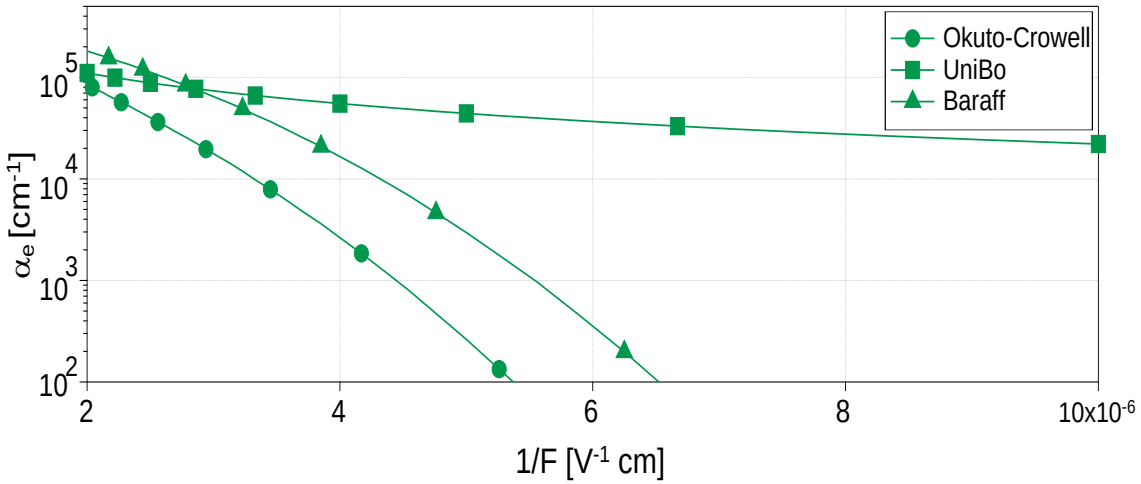


Fig. 3.6 *Electron ionisation coefficient α_e in silicon at at T = 21 °C. The Okuto-Crowell and University of Bologna model uses default Synopsys TCAD SDEVICE parameters [59] while Baraff's model uses a value of r = 0.02. Full parameter values are reported in Appendix A.*

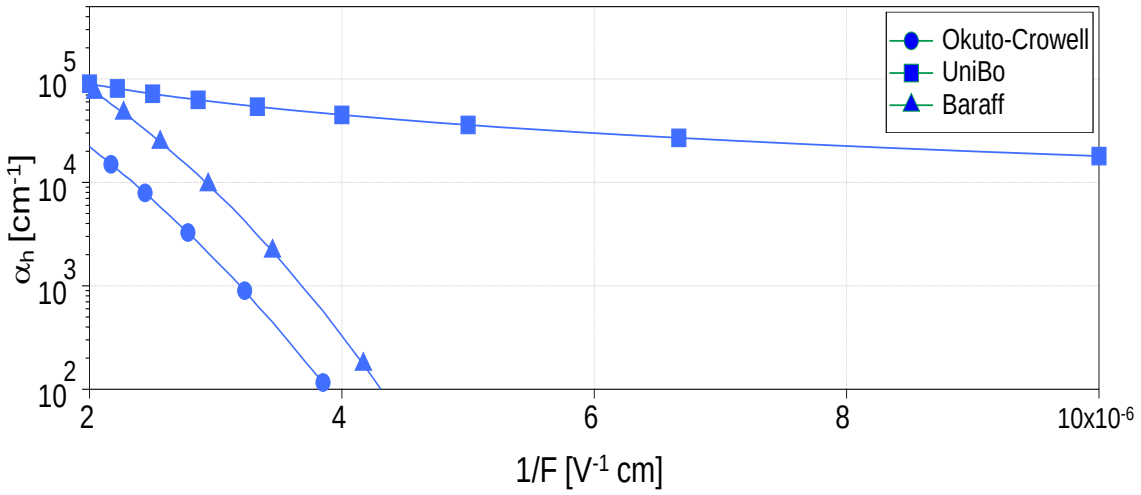


Fig. 3.7 *Hole ionisation coefficient α_h in silicon at at T = 21 °C. The Okuto-Crowell and University of Bologna model uses default Synopsys TCAD SDEVICE parameters while Baraff's model uses a value of r = 0.02. Full parameter values are reported in Appendix A.*

From an experimental and modelling perspective, the impact ionisation rate is usually expressed using the ionisation coefficients α_e and α_h. Considering for example the one-dimensional case where only electrons are generated by electrons, one can write:

$$\alpha_e(F) = \frac{1}{n}\frac{dn}{d(tv_n)} = \frac{1}{nv_n}\frac{dn}{dt} \tag{3.11}$$

where n is the electron concentration, v_n is the electron velocity and $\frac{dn}{d(tv_n)}$ therefore describes the generated carriers per unit distance, with the electron generation rate $\frac{dn}{dt}$. Including the impact ionisation generated holes, which can also contribute to the impact ionisation multiplication, the total generation rate can be written as:

$$G_{II} = \frac{dn}{dt} + \frac{dp}{dt} = \alpha_e n v_n + \alpha_h p v_p = \alpha_e \frac{J_e}{e} + \alpha_h \frac{J_h}{e} \tag{3.12}$$

Let us assume a depletion region of width W and boundary conditions with an initial electron current density at $x = 0$, $J_{e(0)} = J_{e0}$. As the impact ionisation starts, the density $J_e(x)$ increases throughout the depletion region W due to the electrons generated by impact ionisation. At the same time, the generated holes are accelerated in the opposite direction, resulting in a hole current density J_h. In steady-state, the total current density $J = J_e + J_h$ is constant along the structure. At the end of the depletion region W the electron current J_e reaches its maximum described by the current multiplication factor M_e :

$$J_e(W) = J_{e0} M_e = J \tag{3.13}$$

Using (3.12) the change of J_e through the depletion regions can be written as:

$$dJ_e = J_e \alpha_e dx + J_h \alpha_h dx \tag{3.14}$$

Since $J = J_e + J_h$:

$$\frac{dJ_e}{dx} - (\alpha_e - \alpha_h) J_e = \alpha_h J \tag{3.15}$$

which, once solved with the stated boundary conditions, yields a multiplication factor M_e:

$$M_e = \frac{1}{1 - \int_0^W \alpha_e e^{-(\int_0^x \alpha_e - \alpha_h) d\xi} dx} \tag{3.16}$$

Assuming only impact ionisation of electrons, i.e. neglecting α_h in (3.14), the expression of M_e simplifies to:

$$M_e = e^{\int_0^W \alpha_e dx} \tag{3.17}$$

If the velocity of the carrier v_n has reached saturation, a condition usually met at the high fields when impact ionisation occurs [60], the current multiplication factor M_e coincides with the electron multiplication factor.

In (3.16), assuming $\alpha_e = \alpha_h = \alpha$ one can see that the multiplication factor M_e tends to infinity, i.e. avalanche breakdown, when the *ionisation integral* reads:

$$1 = \int_0^W \alpha dx \tag{3.18}$$

This shows that, in principle, the current density is not needed for avalanche breakdown modelling, i.e. the solution to the Poisson equation alone would provide an approximate indication of the breakdown condition.

3.4 The LGAD device

As shown in the previous section, the process of impact ionisation leads to charge multiplication, an effect that can be exploited to obtain sensors with internal gain. The same effect is exploited to increase the timing resolution in timing detectors. The gain effect due to impact ionisation has been used to produce silicon photo-sensors such as avalanche photodiodes (APD [61]) with gain in the 100's, silicon photon multipliers (SiPM [62]) with a gain of the order 10^6 and Low Gain Avalanche Detectors (LGAD) with a gain of 10 - 20.

In a standard PIN silicon sensor, like in the n^{++} - π^+ - p^{++} of Figure 3.8, with π^+ indicating a high resistivity p-type epitaxial region, a value of the electric field of the order of ~ 300 kV$\cdot$ cm^{-1}, which is sufficient to ignite impact ionisation, can be obtained by applying a high external bias voltage to the cathode. Such a value of the electric field might lead to charge multiplication but can cause a device breakdown since it induces a very high field on the device periphery.

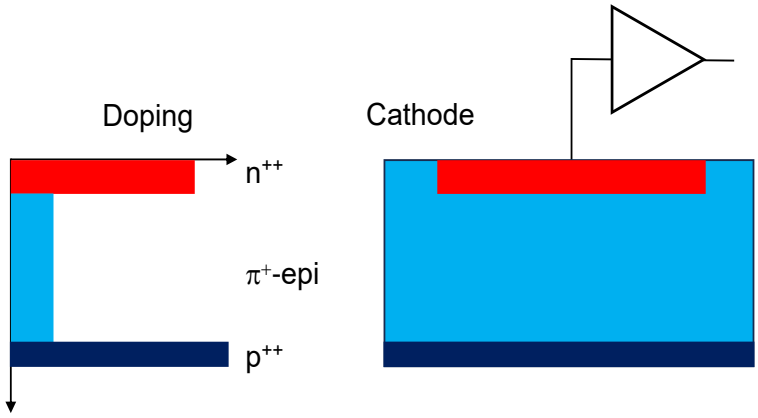


Fig. 3.8 *PIN diode cross-section.*

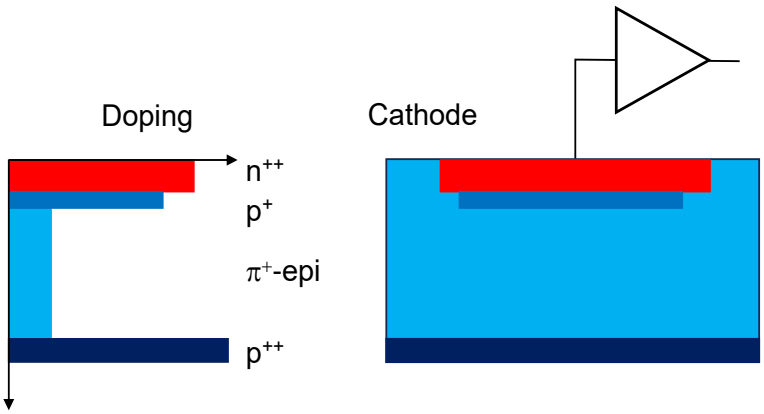


Fig. 3.9 *LGAD diode cross-section.*

In the LGAD device [63], an additional p^+ doped layer (usually by boron, with a density of around 10^{16} cm^{-3} and thickness 0.5 - 1 μm) is implanted near the *pn* junction, as shown in Figure 3.9. This layer usually called the gain layer (GL), creates an electric field high enough to start impact ionisation when depleted. In a typical design, the GL is placed just below the cathode n^{++} region and does not extend to the periphery of the sensor.

3.4.1 LGAD operating principle

Figure 3.10 shows a cross-section of a PIN device and the internal electric field at full depletion, obtained from a TCAD simulation.

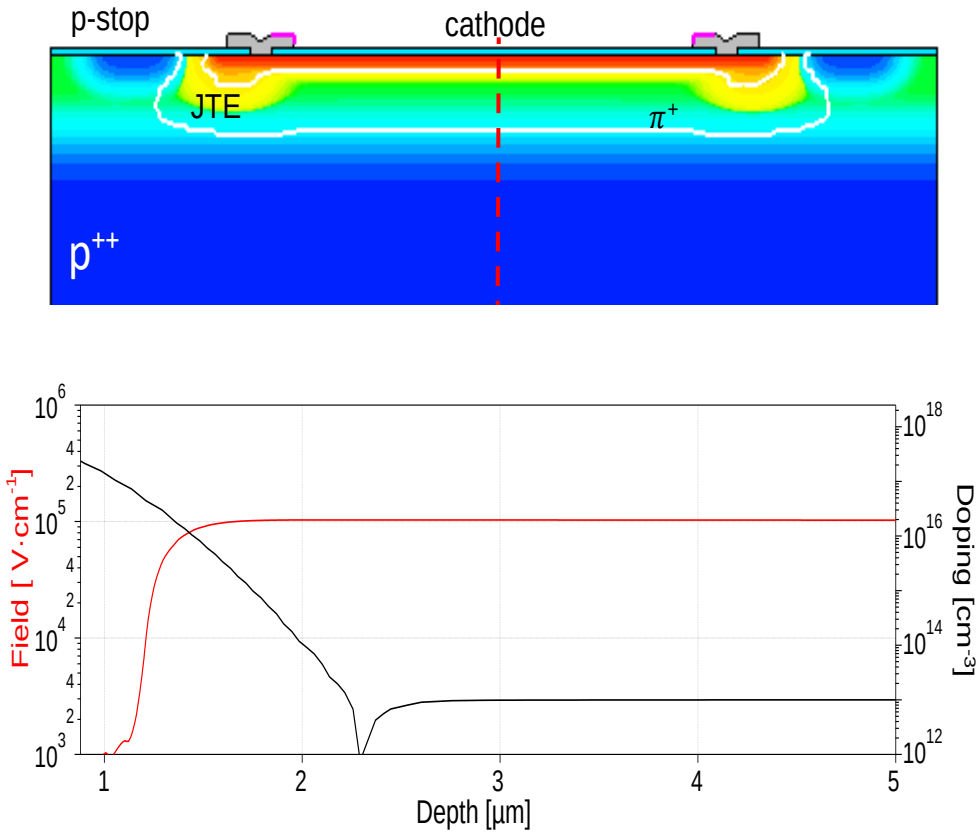


Fig. 3.10 *Top: A PIN cross-section, with the doping concentration shown in different colours. Bottom: the doping and electric field profile at full depletion along the middle cut line.*

In the PIN device, there is a region of constant field, or drift region, which extends from the collecting well through the high resistivity epitaxial π^+ region to the low resistivity substrate p^{++}. The carriers in the drift region, generated by an incoming ionising particle, travel at saturated speed until collected by the electrode. Figure 3.11 shows the cross-section and electric field at full depletion obtained with a TCAD simulation of an LGAD device with the same layout as the PIN diode in Figure 3.10. Two field regions can be distinguished: a drift and a gain region. In the gain region, due to the strong electric field of ≈300 kV·cm^{-1} created by the presence of the GL, the carriers acquire a kinetic energy sufficient to ignite the impact ionisation process, leading to charge multiplication.

The different internal field configurations in PIN and LGAD devices cause very different signal pulses. Figure 3.12 shows a TCAD simulation of the signal current in a PIN and an LGAD following charge injection from a laser pulse, with both devices biased to 60 and 90% of breakdown voltage V_{bd}.

For the PIN diode, after the initial pulse due to Ramo's effect [49], the current decreases steadily, due to the combined collection of electrons and holes. The current decreases until

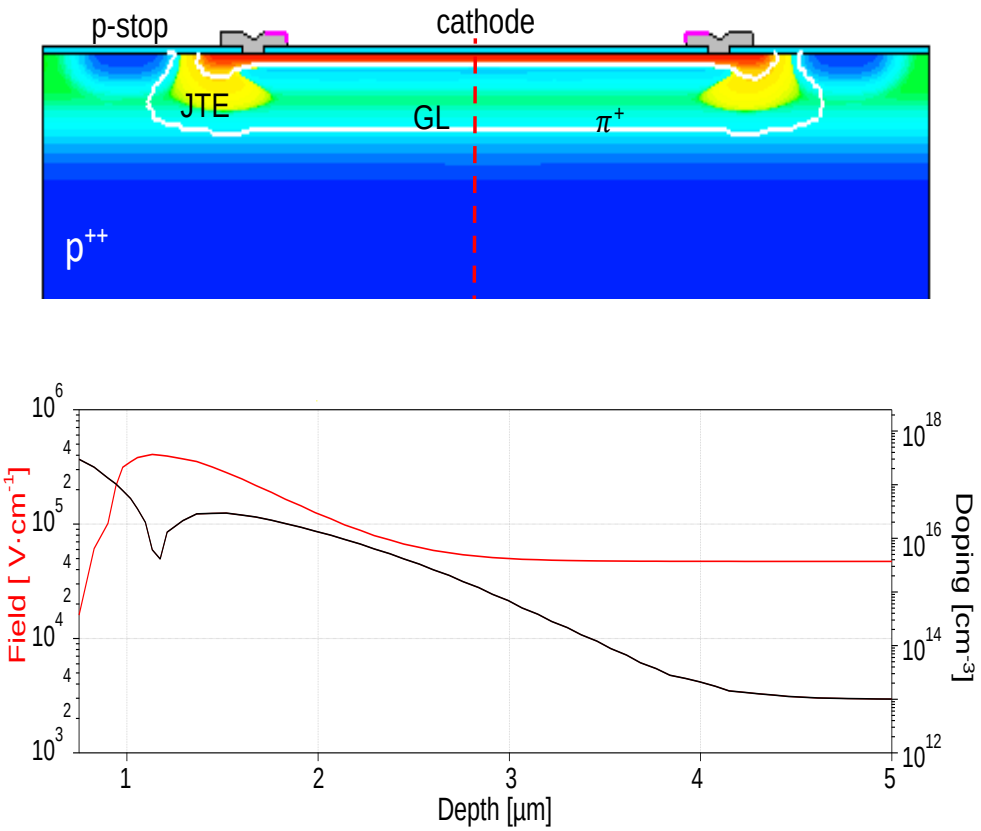


Fig. 3.11 *Top: An LGAD cross-section, with the doping concentration shown in different colours. Bottom: the doping and electric field profile at full depletion along the middle cut line.*

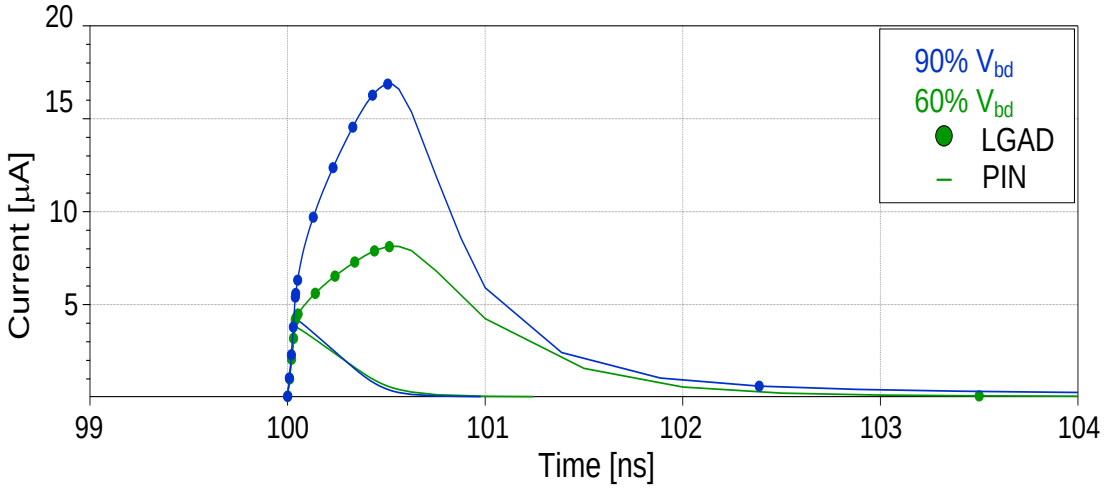


Fig. 3.12 *TCAD simulation of the first 4 ns, following an optical charge injection at t = 100 ns, of a pulse signal in PIN and LGAD biased to 60 and 90% of breakdown voltage V_{bd}.*

all electrons have been collected and the only remaining current is due to the more slowly drifting holes. The overall time length of the pulse signal is dictated by the hole's drift time.

For the LGAD device, the primary electrons and holes generated by the incoming radiation drift toward the n^{++} and p^{++} electrodes respectively. The primary electrons enter the GL, acquire energy high enough to start the multiplication mechanism and produce secondary e-h pairs. The slope in this region, i.e. the charge gain, is dictated by the strength of the electric field in the GL, controlled by the bias voltage applied. Since the multiplication happens very near the cathode, the secondary electrons are quickly collected, whereas the secondary holes drift almost the full bulk thickness before being collected by the anode. The signal in the LGAD thus increases up to the collection of the last primary electron, it stays almost constant until the first secondary hole reaches the anode and then decreases to zero in a period dictated by the hole's drift velocity. The overall time length of the LGAD signal is thus longer than in a PIN diode because it is given by the sum of electrons and holes drift times.

3.4.2 LGAD implementation

In silicon, the average ionisation energy for electrons, i.e. the energy required to create an additional e-h pair following the scattering with an atom, is lower than for holes. This, along with the higher mobility of electrons, makes it preferable to design devices exploiting impact ionisation to be based on electrons rather than holes multiplication. Hence, an n^{++} - π^{+} - p^{++} design, like in Figure 3.9, is almost invariably used. The possibility offered by modern silicon technology to fine-tune the GL doping makes it also possible to design an LGAD capable of operating at different bias voltages with similar performances.

In both PIN and conventional LGAD designs, shown in Figures 3.10 and 3.11, the collecting n^{++} well is surrounded by a deep n^{++} implant, called Junction Termination Extension (JTE), equipped with a metal field plate. Its purpose is twofold:

- to decrease the strength of the electric field at the edge of the n^{++} well which could lead to premature breakdown. In a multipad design, the JTE is located around each pad;
- in an LGAD device, to ensure that the e-h pairs generated by particles impinging in the region between the pads do not reach the gain layer. To start the impact ionisation process, electrons generated in this region would have first to drift to the gain implant and this would delay the onset of charge multiplication by 10 or 100s ps, causing a wrong assignment of the particles' time of arrival.

The n^{++} implants are isolated from each other by an extra p^{++} implant, the so-called p-stop. The p-stop has a doping concentration in the range 10^{15} - 10^{17}cm^{-3}. Without p-stops, the inversion layer created by the positive charges at the Si–SiO_2 interface would short together all n^{++} implants [64].

At the physical edge of the sensor, guard rings are also used. These structures help gradually decrease the voltage from the sensor edge, increasing the breakdown voltage and reducing the leakage current [65]. Multi-guard ring structures are often used to enhance their effectiveness [66] but this increases the size of the sensor cell and also reduces the active area.

3.4.3 LGAD noise

In a PIN diode the shot noise, which is always present in silicon sensors and is due to the random crossing of charge through a *pn* junction, [67, 68], does not usually represent a problem unless the sensor is heavily irradiated. In this case, the leakage current is very large and its spectral power density is given by:

$$\sigma_{sh}^2 = 2e(I_{surf} + I_{bulk})\Delta f \tag{3.19}$$

where I_{surf} and I_{bulk} are the surface and bulk leakage currents respectively. Shot noise is frequency-independent over an extremely wide frequency region.

In devices with internal gain, like LGADs, the shot noise might become significant even before irradiation, since the current is multiplied by the gain factor G. Additionally, the internal gain mechanism, a random process, contributes to the total noise, increasing the overall shot noise by a factor of F [69]:

$$F \approx G^x = Gk + (2 - \frac{1}{G})(1-k) \tag{3.20}$$

where x is the excess noise index and $k = \frac{\alpha_h}{\alpha_e}$ is the ratio between the holes and electrons ionisation coefficients, introduced in (3.9). Expression (3.20) shows that to minimise the noise, the gain G should not be too high and it is preferable to operate the device with a low value of k, i.e. with the holes ionisation coefficient smaller than the electron's one. In silicon, α_h is always smaller than α_e, unless at a very high field, as shown in Figures 3.6 and 3.7, suggesting that electron multiplication, i.e. n^{++} - π^+ - p^{++} structures, are preferable. Including (3.20) into (3.19), the total shot noise power density becomes:

$$\sigma_{sh}^2 = 2e(I_{surf} + I_{bulk}G^2F)\Delta f \tag{3.21}$$

where the G^2 factor arises from the current noise I_{bulk} being multiplied along with its fluctuations. The overall expression of the signal-to-noise ratio (SNR) can then be written as:

$$SNR = \frac{IG}{\sqrt{2e(I_{surf} + I_{bulk}G^2F)\Delta f}} \propto \frac{I}{\sqrt{F}} \tag{3.22}$$

which would suggest that a device with internal gain G would have a worse SNR if this was the only noise source. In practice, the readout electronics contribute to the overall noise and usually dominate it up to the point where the shot noise of the sensor becomes predominant, as shown qualitatively in Figure 3.13.

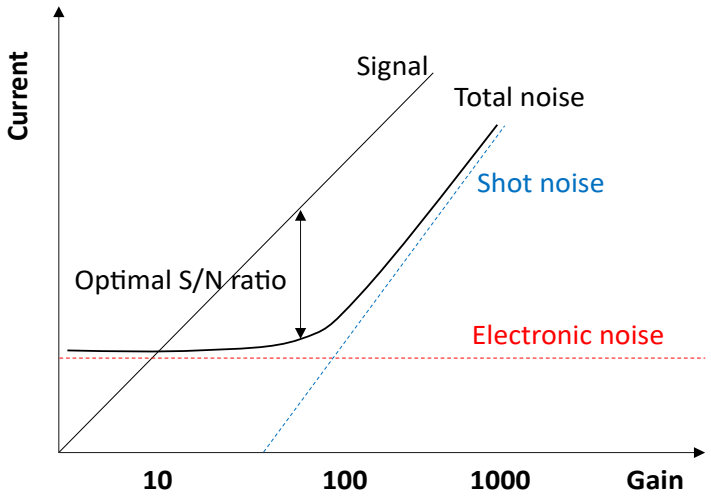


Fig. 3.13 *LGAD Signal to Noise Ratio (SNR).*

The overall SNR increases with gain G up to the point where the shot noise starts to increase. A gain region that maximizes the SNR depends on the specific design, of both the sensor and the readout, but it usually corresponds to moderate gain G of the order of 10 - 20 [47].

3.5 Readout electronics

The typical circuit topology used in LGAD sensor readout preamplifier consists of a Trans Impedance Amplifier (TIA) and it is shown in Figure 3.14.

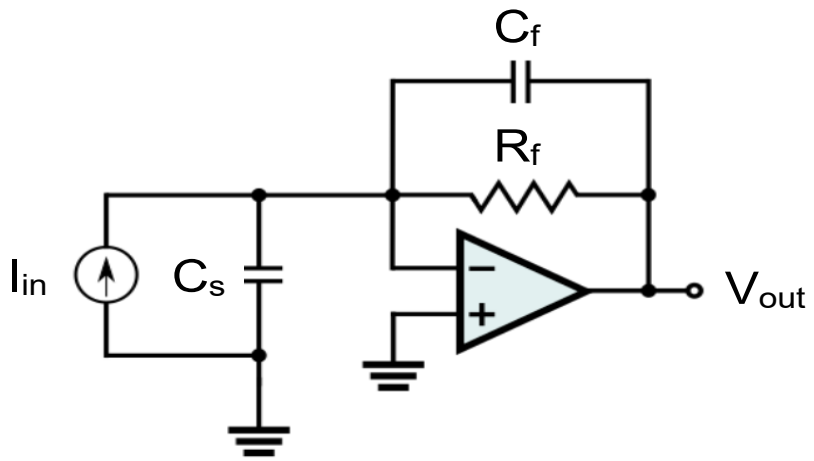


Fig. 3.14 *Transimpedance Amplifier (TIA) schematic.*

In it, a fraction of the output voltage V_{out} is converted to a current by the feedback path and subtracted from the input, lowering the total input impedance by a factor due to the Miller effect [70]. Neglecting the effect of the input impedance Z_{inA} of the amplifier, the input impedance Z_{in} seen by the sensor is given by:

$$Z_{in} = \frac{1}{sC_s} // \frac{R_f}{A(1+s\tau_f)} = \frac{R_f}{A(1+s\tau_f)+sR_fC_s} \tag{3.23}$$

where $\tau_f = R_fC_f$, A is the open-loop gain of the amplifier, C_s is the capacitance of the sensor and $s = i2\pi f$ the complex variable. The choice of the feedback components values R_f and C_f, which define the characteristic time constant of the circuit, leads to a distinction between a Fast Current Amplifier, or fast TIA, and a Charge Sensitive Amplifier (CSA).

3.5.1 Transimpedance amplifier

If, in (3.23), $\tau_f << 1$ one gets:

$$Z_{in} \approx \frac{R_f}{A + sR_fC_s} \tag{3.24}$$

In such case, the output voltage V_{out} from a current I_{in} is given by:

$$V_{out} = -AV_{in} = -AI_{in}Z_{in} \approx -AI_{in}\frac{R_f}{A + sR_fC_s} \tag{3.25}$$

which, for frequencies $f << \frac{1}{2\pi R_f C_s}$ gives:

$$V_{out} \approx -I_{in}R_f \tag{3.26}$$

In such configuration, the TIA converts the fast input current signal I_{in} from the sensor into a voltage output V_{out} with gain R_f, essentially preserving the pulse signal shape. This allows full exploitation of the fast slew rate SR of the LGAD sensor for timing. The wide bandwidth required for the amplifier might, however, increase the overall noise and the jitter, which, from 3.4 and assuming the sensor current is a short pulse of t_{dr} characteristic time, can be expressed as [10]:

$$\sigma_{jit} \propto \frac{t_r}{\frac{S}{N}} = \frac{e_n C_s}{Q_{inj}} \sqrt{\frac{t_{rt}^2 + t_{dr}^2}{2t_{rt}}} \tag{3.27}$$

where e_n is the voltage noise spectral density of the amplifier, C_s is the sensor capacitance, Q_{inj} is the input signal charge, i.e. the integral of the pulse current over t_{dr}, and t_{rt} is the rise time of the amplifier, inversely proportional to its bandwidth, combined in quadrature with t_{dr} to estimate the total rise time t_r. From (3.27) it is clear that to minimise the jitter, a low C_s, a large Q_{inj} and small t_{dr} are required, in combination with optimised characteristics of the amplifier for the specific sensor. Any added capacitance in parallel to C_s, for example, due to Z_{inA}, decreases the overall bandwidth and affects the jitter making it important to choose an amplifier with a small input impedance and to operate the sensor at full depletion, to reduce its capacitance. Figure 3.15 shows qualitatively the behaviour of (3.27), with the minimum of jitter found at $t_{rt} = t_{dr}$. An additional aspect to consider for the bandwidth optimisation of the amplifier is the requirement for a potentially high bias current, which increases the power consumption, an important factor in multi-channel systems.

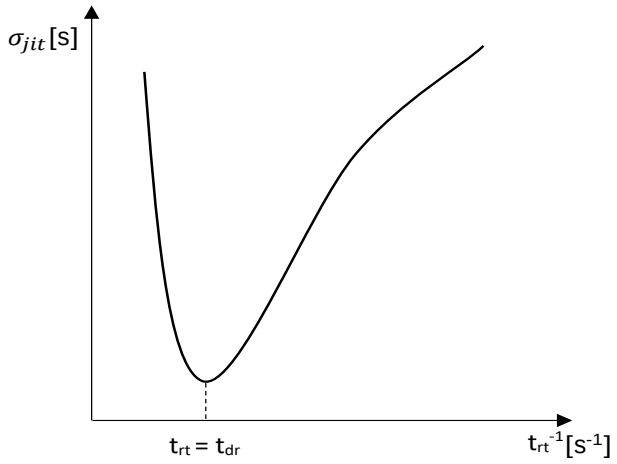


Fig. 3.15 *An example plot of total jitter σ_{jit} from (3.27). An excessive bandwidth (low t_{rt}) of the amplifier increases the noise, whilst a low bandwidth (high t_{rt}) decreases the slew rate, in both cases increasing the jitter. An optimum value is found for $t_{rt} = t_{dr}$.*

3.5.2 Charge sensitive amplifiers

If, on the other hand, in 3.23, $\tau_f >> 1$ and A $C_f >> C_s$, one gets:

$$Z_{in} \approx \frac{R_f}{As\tau_f + sR_fC_s} \approx \frac{R_f}{As\tau_f} = \frac{1}{AsC_f} \tag{3.28}$$

In this case, an input current I_{in} is converted into:

$$V_{out} = -AV_{in} = -AI_{in}Z_{in} \approx -AI_{in}\frac{1}{AsC_f} = -I_{in}\frac{1}{sC_f} \tag{3.29}$$

which, in the time domain, corresponds to the integral of I_{in}, i.e. Q_{in} multiplied by the factor $\frac{1}{C_f}$. In such configuration, the TIA converts the input charge Q_{in} collected from the sensor into a voltage output V_{out} with gain $\frac{1}{C_f}$, effectively a CSA. The integration of the input signal reduces the noise and therefore the jitter, but does not preserve the input signal shape. Additional elements normally included in the CSA configuration to improve the SNR include differentiating (CR) and integrating (RC) stages, to implement a shaper [71].

In conclusion, the choice of the amplifier configuration depends on the application: for timing measurements, the fast TIA configuration is preferable, as it preserves fully the input signal shape, whereas a CSA provides a more accurate measurement of the collected charge, which improves the gain measurement.

Chapter 4

Design and Simulations of Low Gain Avalanche Diode as Timing Detector

The Technology Computer-Aided Design (TCAD) package used for this research project was Synopsys Sentaurus [59]. Such a package includes tools for the complete simulations of the fabrication process and electrical, optical, and thermal characteristics of semiconductor devices. The simulation of typical manufacturing processes like epitaxial growth, oxidation, ion implantation, etching, thermal diffusion and metal deposition was used in the design phase of this project. The correct implementation of such simulation steps allows accurate predictions of the structures resulting from the fabrication process.

In this chapter, Section 4.1 provides an introduction to the Synopsys TCAD simulation environment, followed by the description of the layout of the devices designed for this project in Section 4.2. In Section 4.3 details of the simulation processes of the devices including manufacturing and electrical simulations are provided, along with comparisons of simulations with test results. The meshing strategies used in the simulations are illustrated in Section 4.4 and the validity of the local model for impact ionisation for the LGAD devices designed in this project is discussed in Section 4.5.

4.1 Introduction to Sentaurus TCAD

Technology Computer-Aided Design (TCAD) refers to the use of computer simulations to develop and optimize semiconductor processing technologies and devices. Its widespread use is related to the high cost of semiconductor manufacturing, which has become increasingly complex over the last few decades, at the same time attaining to continuously reduce the unit price of semiconductor chips. Owing to the continuously increasing demand for consumer electronics, the market requires cycles of design to production of 18 months or less. As typically 2-3 months are required for wafer tape-out, a short time is left for development, as shown in Figure 4.1 [72]. The traditional methodology of trial and error is no longer feasible due to cost and time constraints. Thus, computer-aided design (CAD) is routinely accepted for the design and development of ICs.

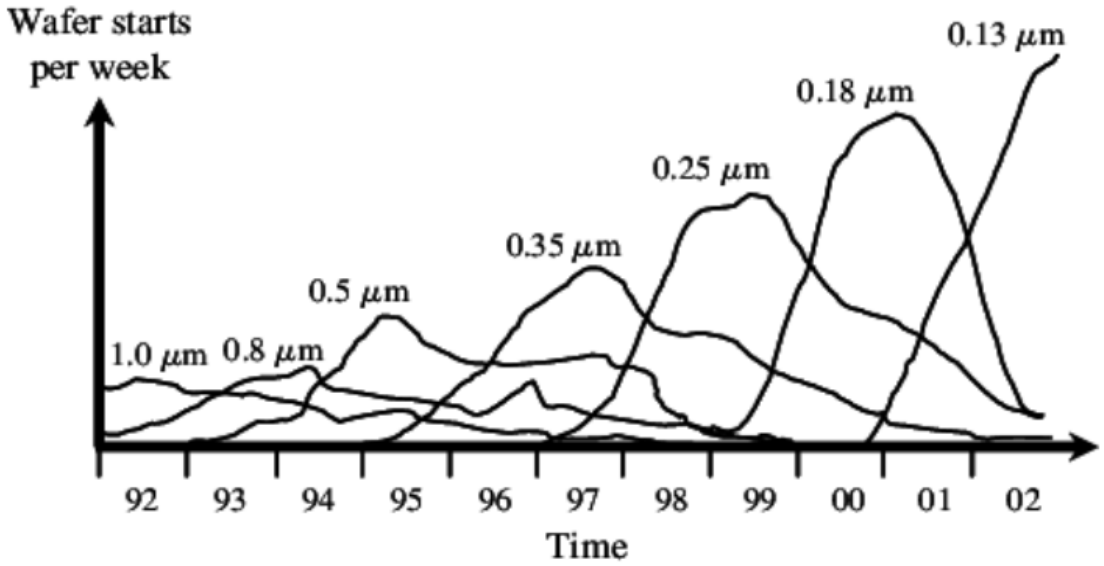


Fig. 4.1 *Shrinking product life cycles in the semiconductor industry over time (years) - from [72].*

Synopsys Sentaurus TCAD provides a collection of simulation tools that includes industry-leading processes and device simulation packages, and a powerful GUI-driven environment for managing tasks and analysing the results. Process and device simulation of a variety of devices, including Complementary Metal-Oxide-Semiconductor Field-Effect Transistor (CMOSFET), Fin-shaped field-effect transistor (FinFET), power devices, memory devices, image sensors, solar cells, and analogue/RF devices are possible. Tools are available for interconnect modelling and parameter extraction, including SPICE parameters for macro models.

For this specific project, simulations of LGAD devices and PIN diodes operating in the impact ionisation region were performed. The main components of a generic semiconductor device simulator are shown in Figure 4.2.

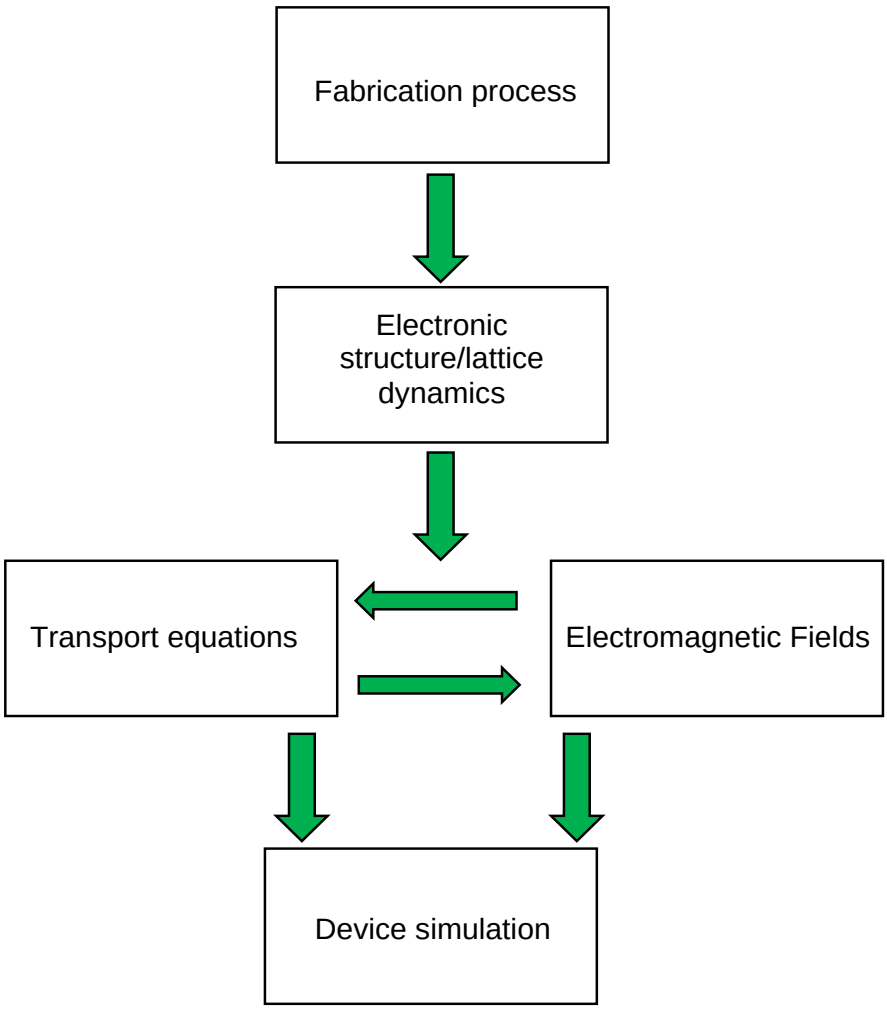


Fig. 4.2 *Schematic diagram of a typical semiconductor device simulator.*

The fabrication process tool simulates the manufacturing steps involved in the making of a semiconductor device. In modern processes, tens of steps are usually required. The obtained structure includes spatial information on doping concentration, defects in the crystal, stress and in general all the relevant details needed for the description of an electronic device. For the step of the electrical simulation, semiconductor transport equations and fields that drive the charge are coupled and need to be solved self-consistently. The results of the simulations can then be presented graphically or saved to files for additional processing.
In Sentaurus TCAD an example of a simulation flow chart, implementing processes of Figure

4.2, is shown in Figure 4.3. The Sentaurus Structure Editor allows the definition of the semiconductor device in 1, 2 or 3 dimensions and includes the geometry along with doping spatial information. The resulting structure, once suitably discretized for finite element method (FEM) electrical analysis, is sent as input to tools for electrical simulation, Sentaurus Device. The results might include current vs. voltage, transient response of the device to charge injection, capacitance vs. voltage and other variables depending on the analysis. Results can then be plotted within the Sentaurus framework or saved to files for additional analysis.

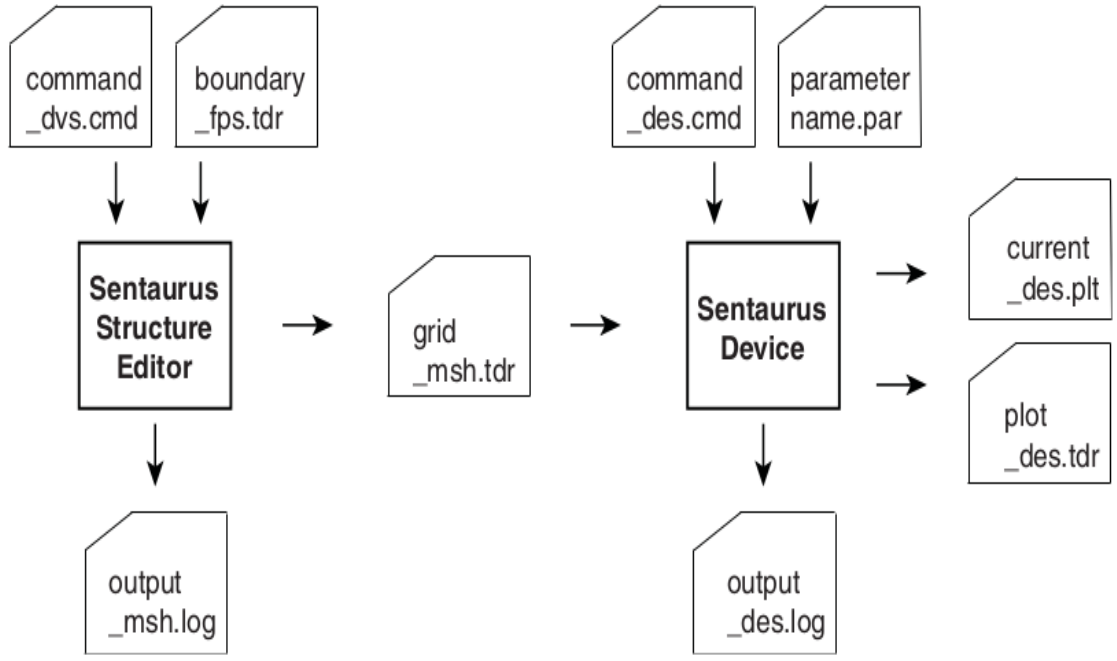
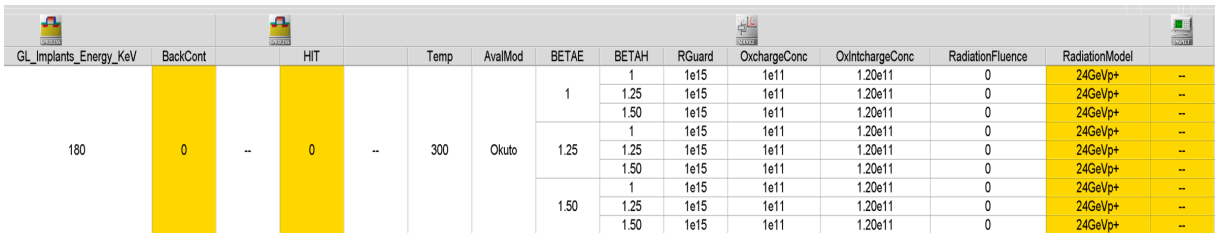


Fig. 4.3 *Typical simulation flow implemented in Sentaurus TCAD.*

The input files to the tools, describing the desired device structure or the type of analysis are scripts compatible with tool command language (TCL) [73]. Each TCL script is related to a specific tool in the graphical environment (or Sentaurus Workbench) of Synopsys TCAD. An example GUI implementation of the flowchart used in this project is shown in the Sentaurus Workbench environment in Figure 4.4.



GL_Implants_Energy_KeV	BackCont		HIT		Temp	AvalMod	BETAE	BETAH	RGuard	OxchargeConc	OxIntchargeConc	RadiationFluence	RadiationModel	
180	0	--	0	--	300	Okuto	1	1	1e15	1e11	1.20e11	0	24GeVp+	--
								1.25	1e15	1e11	1.20e11	0	24GeVp+	--
								1.50	1e15	1e11	1.20e11	0	24GeVp+	--
							1.25	1	1e15	1e11	1.20e11	0	24GeVp+	--
								1.25	1e15	1e11	1.20e11	0	24GeVp+	--
								1.50	1e15	1e11	1.20e11	0	24GeVp+	--
							1.50	1	1e15	1e11	1.20e11	0	24GeVp+	--
								1.25	1e15	1e11	1.20e11	0	24GeVp+	--
								1.50	1e15	1e11	1.20e11	0	24GeVp+	--

Fig. 4.4 *An example of simulation flow in TCAD Sentaurus Workbench as used in this project. The simulation flow proceeds from left to right. The input parameters to each of the tools used can be modified and correspond to a specific simulation.*

The graphical implementation of Sentaurus Workbench (SWB) allows for organizing the technical process flow. The architecture supports user-defined databases, which can input the custom processes and results of semiconductor fabrication process technology or electrical property tests. It is also possible, using SWB, to parametrize input files to run simulation groups automatically. In Figure 4.4 the first two tools (SPROCESS) simulate the fabrication process of the device. Such simulations include all the steps used by the semiconductor foundry during device manufacturing, i.e. epitaxial growth, diffusion, ion implantation, etching and metalization for contact creation. In the shown example the first tool is used to perform the actual fabrication of the devices, and the second tool is used to re-mesh the structure to make it suitable for electrical simulation. The third tool (SDEVICE) performs an electrical simulation of the device, in this case sweeping the voltage applied to the cathode of the device up to the breakdown point to obtain current versus voltage curves (IV plots). The fourth tool (INSPECT) is used to plot and extract current and voltage properties of the simulated devices. An additional visualization tool, SVISUAL, can be used and also allows internal probing of the simulated structures, to obtain information such as the local strength and direction of the electric field, carrier mobility and recombination.

In this thesis, the simulation tools SPROCESS, SDEVICE, and SVISUAL were used. Details and descriptions of the tools used are discussed in this chapter.

4.2 LGAD Device design

In this project, two fabrication runs were made, in collaboration with the semiconductor foundry Teledyne e2v (Te2v [74]). The first fabrication run aimed at investigating the performances achievable with the available process, comparing the simulation predictions with test results and selecting the best-performing process combinations to address the specifications of the High Granularity Timing Detector (HGTD) for the ATLAS detector [10]. The design of the second run included the full array of 15 $\times$ 15 LGAD cells as per HGTD layout specifications.

4.2.1 Run 1 LGAD design description

A total of eight 6" wafers, each of 50 μm thick high resistivity epitaxial layer, were processed in the first run. Each wafer had a different LGAD gain layer doping profile, to investigate the achievable gain and timing performances.

The layout of the Run 1 design is shown in Figure 4.5. LGAD cells with implanted gain layers and various cathode sizes are shown on the left side of the layout. For reference comparison, PIN diodes with the same layout but lacking the gain layer implant were implemented in the same design and are shown on the right.

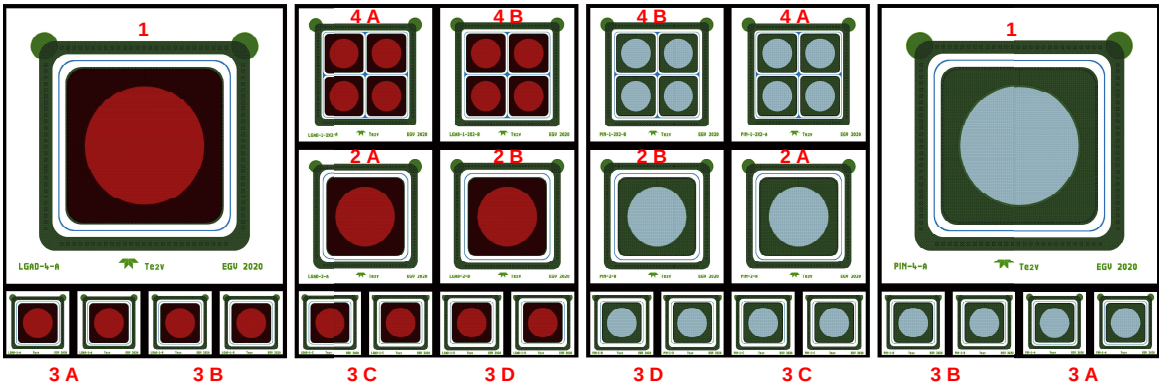


Fig. 4.5 *Run 1 layout, including LGAD on the left and PIN diodes of the same layout on the right.*

Individual cells of cathode sizes 1, 2 and 4 mm, along with an array of 2 $\times$ 2 cells of 1 mm cathodes were implemented in the first design, as shown in Figure 4.6. Different layouts with varying distances of the guard ring to the cathode were included to investigate the effects of this distance on the breakdown voltage. The central hole in the top metal layer was included in all layouts to allow laser injection for charge collection studies. An example of the cross-section of an LGAD device is shown in Figure 4.7. Details of cell dimensions are

reported in Table 4.1. Details of the (normalized) dose and energy of gain layer ^{11}B implant as used in Run 1 fabrication are reported in Table 4.2.

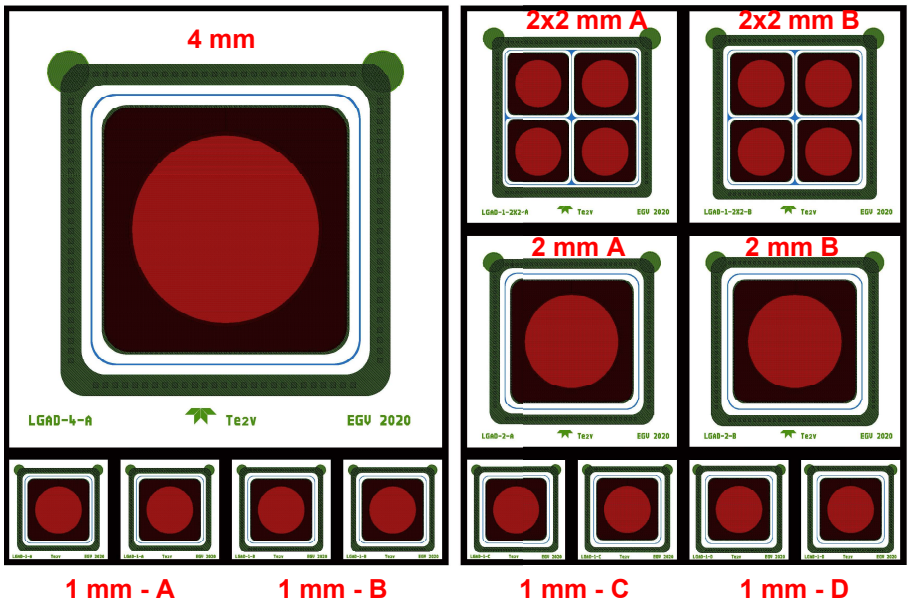


Fig. 4.6 *The different cell layouts implemented in Run 1. The cathode size is shown in the figure while the different cell flavours (A, B, C, D) are reported in Table 4.1.*

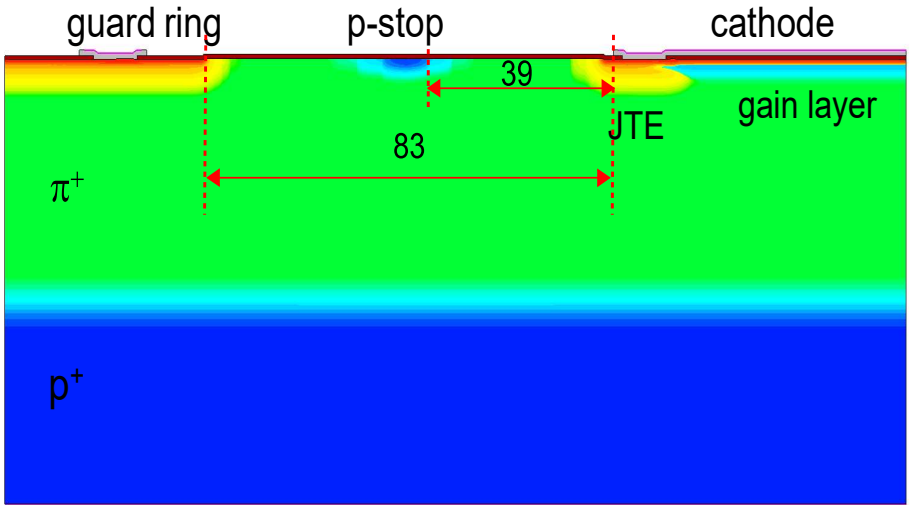


Fig. 4.7 *Example of the cross-section of a 1 mm LGAD device (first flavour of layout 3 in Table 4.1) from Run 1. Dimensions in µm.*

Table 4.1 *Details of Run 1 LGAD cells layout. For each layout, the different flavours A, B, C, and D refer to the p-stop to Guard Ring distance. Values in µm.*

Cell layout	Cathode diameter	Laser hole diameter	Cathode to p-stop	p-stop width	p-stop to Guard Ring	Guard Ring width
1	4000	3020	156	6	152	332
2	2000	1510	78	6	76/96	166
3	1000	755	39	6	38/48/58/68	83
4	1000	755	39	6	38/68	166

Table 4.2 *Dose and energy of the ^{11}B layer (GL) implants in the Run 1 LGAD production. At the request of Te2v, the values are normalized to the dose and energy adopted for wafers 19, 20 and 21.*

Wafer	GL dose	GL energy
19,20,21	1.00	1.00
17,18	1.07	1.00
15,16	0.92	1.05
12,13,14	1.00	1.05
9,10,11	1.07	1.05
7,8	1.15	1.05
4,5,6	1.00	1.11
2,3,24	1.07	1.11

4.2.2 Run 2 LGAD design description

For Run 2 fabrication a total of four 6" wafers were processed, featuring the same 50 μm thick high resistivity epitaxial layer as in the first run and with each wafer of different gain layer doping. The gain layer doping in Run 2 was selected based on the extrapolation of the performance observed in the measurements of the Run 1 devices. Three of the Run 2 wafers have the same gain layer as wafers 2, 4 and 7 in Run 1.

The layout of Run 2 is shown in Figure 4.8, along with the picture of an individual LGAD cell of the array. The full array consists of 15 $\times$ 15 LGAD cells (PIN diodes were not included in the design), each cell having a size of 1.3 $\times$ 1.3 mm^2, with a cell layout similar to that of Run 1. The total size of the array is 21 $\times$ 21 mm^2. Each cell includes a 90 μm diameter passivation apertures for bump-bonding, a 300 μm diameter central hole with a rectangular slit for laser charge injection across the full length of the cell and an additional 200 $\times$ 150 μm^2 aperture for probe testing.

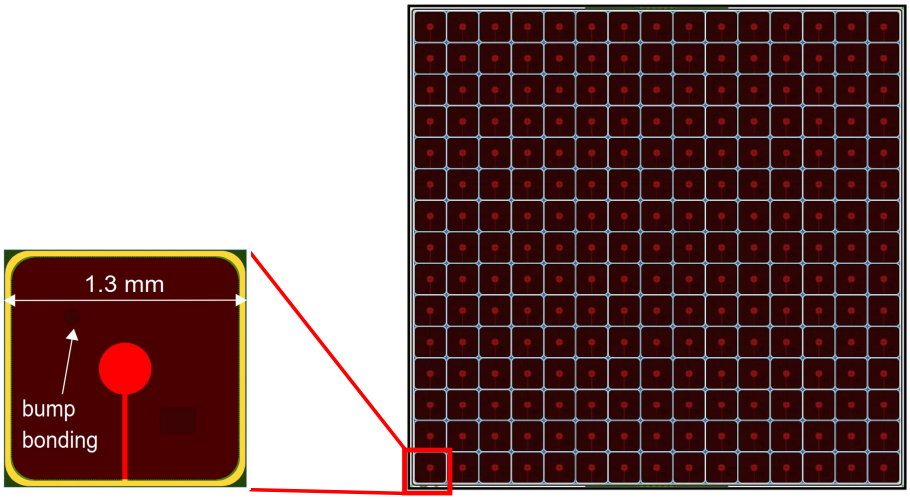


Fig. 4.8 *Run 2 15 $\times$ 15 cells array LGAD layout (Right) and a single LGAD cell (Left). On each cell the central hole and rectangular slit on the metal cathode allow laser charge injection to investigate the gain uniformity.*

An example of the cross-section of an LGAD cell of the array is shown in Figure 4.9 and the layout details are summarized in Table 4.3. The details of the (normalized) dose and energy of the gain layer ^{11}B implant used in Run 2 are reported in Table 4.4.

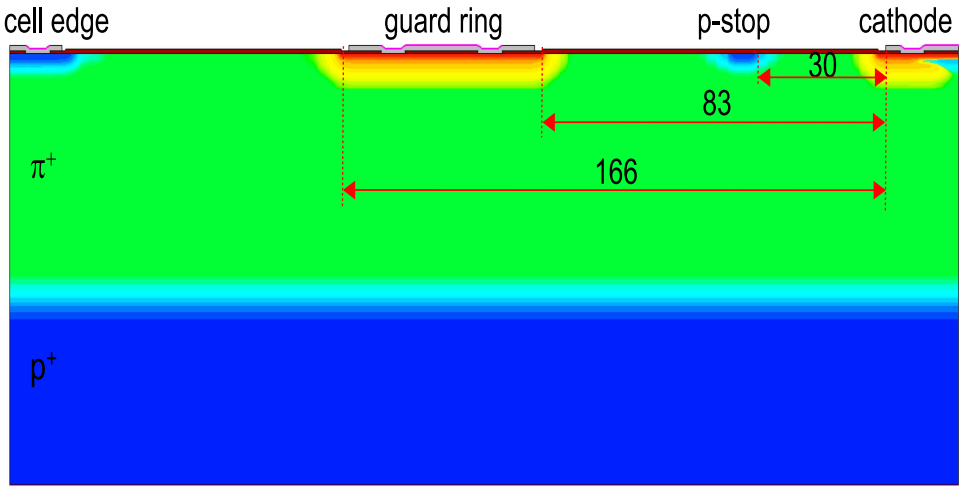


Fig. 4.9 *Cross-section of a cell array from Run 2. Dimensions in µm.*

Table 4.3 *Details of Run 2 LGAD array cells layout. Values in µm.*

Cathode diameter	Cathode to GR	Cathode to p-stop	p-stop width	Cathode to Cathode	GL to GL
1300	83	30	6	66	86

Table 4.4 *Dose and energy of the ^{11}B layer (GL) implants in the Run 2 LGAD production. The values are normalized to the dose and energy adopted for wafers 19, 20 and 21 of Run 1.*

Wafer	GL dose	GL energy
1,5,7	1.07	1.11
8,12,6	1.15	1.05
13,23,24	1.15	1.11
19,20	1.00	1.11

4.3 LGAD simulation

The main TCAD tools used in this project were Sentaurus SPROCESS [75] for the fabrication process simulations and Sentaurus SDEVICE [76] for the electrical and optical simulations. Figure 4.10 shows the full simulation flowchart, including process fabrication and electrical characteristics. The process parameters for the doping profile of the gain layer and the Junction termination extension (JTE) were determined via an iterative approach since no previously fabricated devices were available at the start of this project. The design approach followed consisted of first optimizing the accuracy of the fabrication process simulation via comparison with Secondary Ion Mass Spectrometry (SIMS [77]) results and then calibrating the process simulator parameters to reduce errors from inaccurate doping profiles from simulations. Both methods of analytical and Monte Carlo implantation were used [75] in the process simulation[1].

A PIN diode was simulated to predict more accurately its breakdown voltage, as explained in 4.3.3, to optimize its layout to reach high operating voltages (of the order of 500 V or above) and to estimate its charge collection performance from transient simulations. The next step was the design of an LGAD device, which included an investigation of the breakdown and charge collection, using different impact ionisation models and different GL doping. Process parameters were then optimized and the entire simulation process was repeated until a gain over 10 and a predicted breakdown voltage, around half that expected for the PIN diode, was obtained. Due to the expected uncertainties in modelling the impact ionisation, as explained in section 4.3.3, a range of different doping profiles were chosen for the device fabrication. Two-dimensional simulations were performed to investigate the effects of the gain layer and the device edge on the device performance. Besides the parameters of the implants, i.e. dose and energy of the implanted ions, changes in the temperature during the process (the thermally activated diffusion and dopants activation steps in the fabrication process) were investigated to optimize the predicted characteristics of the devices.

A summary of the various fabrication and electrical simulation steps is provided in the next subsections.

[1] In Sentaurus SPROCESS either analytic functions or Monte Carlo (MC) methods can be used to compute the profile distributions of implanted ions and the implantation damage. In the analytical approach various functions that reproduce these profiles, based on the theory of ions collisions or empirical, are employed and their parameters are optimized to fit the obtained SIMS data. The MC method uses a statistical approach to calculate the penetration of implanted ions into the target and the accumulation of crystal damage by computing a large number of individual ion trajectories in the target. Generally, the MC approach guarantees more accurate results at the cost of increased computational time.

4.3.1 Process simulation

The steps of lithography, photoresist deposition, oxidation, implantation, thermal diffusion, etching and metallization were all implemented in the simulation of the fabrication process. The diagram of the Sentaurus SPROCESS flowchart is shown in Figure 4.10 and follows all the fabrication steps used at the silicon foundry. Implantation and dopant diffusion models were selected based on test results provided by the foundry with additional optimization of some of the parameters to improve the simulation accuracy. Details of the process steps, along with intermediate SPROCESS outputs, are provided below. The simulated structure consisted of a two-dimensional cell half the width of the real one to reduce the required computational resources. The region of the p-type implant under the cathode (the gain layer GL), the p-stop and the guard ring were included to investigate their effects on the gain and breakdown voltage.

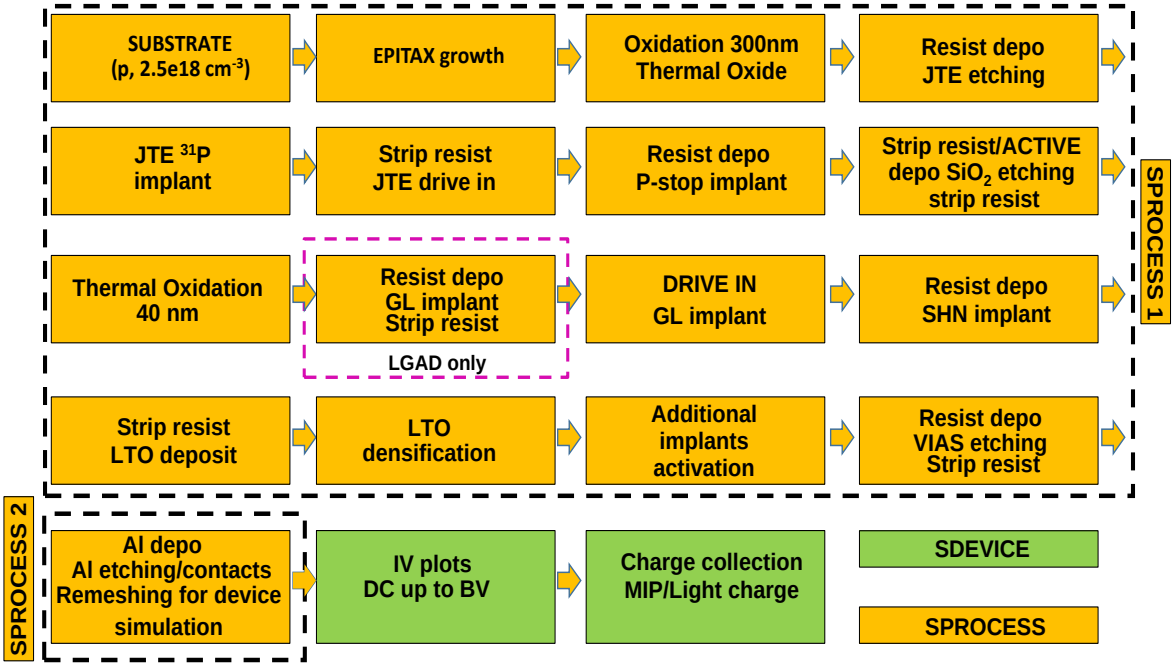


Fig. 4.10 *TCAD simulation flowchart. The SPROCESS steps (yellow boxes in the flowchart) refer to the manufacturing process steps. The output structure from SPROCESS is sent to SDEVICE (green boxes in the flowchart) for the simulation of the electrical characteristics.*

Epitaxial growth

The first step consisted of the growing of high resistivity ($\sim 1\ k\Omega$) silicon epitaxial layer on a low resistivity Czochralski (CZ) substrate, Figure 4.11. The isotropic growth was performed at a temperature of 1000 °C.

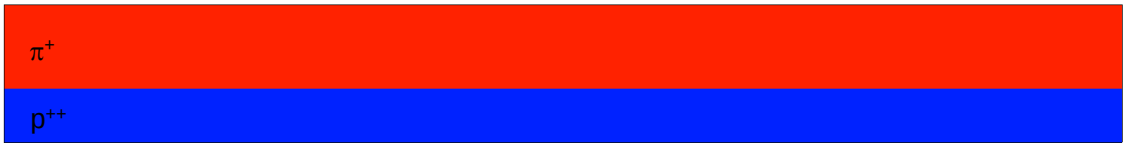


Fig. 4.11 *Silicon epitaxial layer (π^+) growth on CZ silicon p^{++} substrate.*

Thermal oxidation

Next, a layer of 300 nm oxide was thermally grown in a dry atmosphere, Figure 4.12.

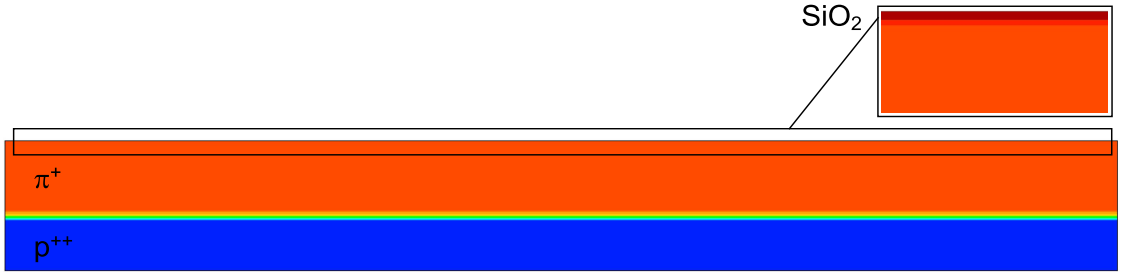


Fig. 4.12 *Thermal growth of 300 nm thick SiO_2.*

Resist deposition and etching

After the deposition of JTE and guardring photomask an anisotropic oxide etching was performed, Figure 4.13.

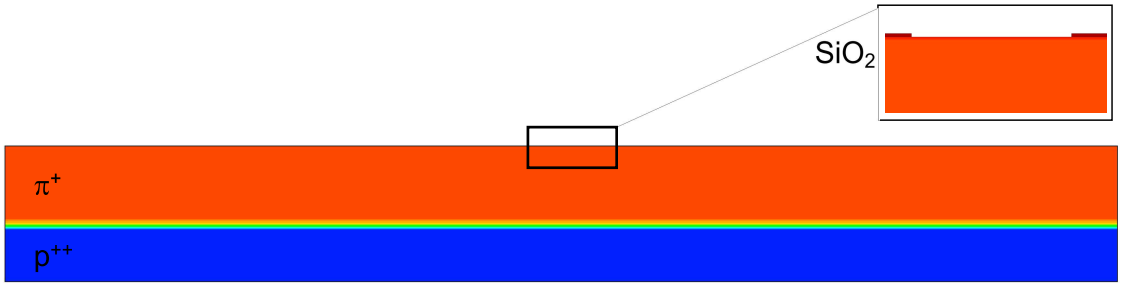


Fig. 4.13 *Resist deposition and anisotropic SiO_2 etching.*

JTE implant

The first implantation process was performed, using an analytical implantation model for ^{31}P ions with an energy of 25 keV and a dose of 10^{13} cm^{-2}. To avoid the channelling effect[2], a tilting of 7 degrees was applied to the beam and the entire implantation process was performed through several successive implantations rotated by around 15 degrees each.

JTE annealing

After stripping off the residual resist, the drive-in, or thermal diffusion of introduced impurities, for the JTE implant activation was performed, Figure 4.14. This step was slightly modified in the second fabrication run to guarantee an increased breakdown voltage.

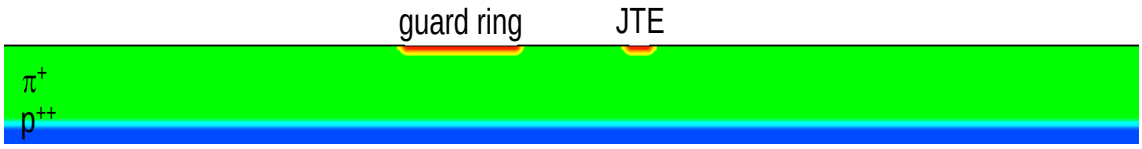


Fig. 4.14 *JTE implantation and drive-in.*

P-stop and p-channel implantation

After the photomask resist deposition of the p-stop mask, the ^{11}B implantation was performed through the 300 nm thick oxide. A Monte Carlo (MC) implant simulation was performed to guarantee better accuracy in the doping profiles.

Active area etching

After the photomask resist deposition of the active area mask, which corresponds to where the actual device is located, etching of the oxide was performed with an over-etching of 100 nm.

Re-oxidation

A thin layer of 40 nm of sacrificial oxide was thermally grown onto the silicon, as sketched in Figure 4.15. This has the purpose of reducing the channelling effect during the next step of implantation via the scattering of the ions through the oxide.

[2]The influence of the crystal lattice on the trajectories of ions penetrating the crystal is known as channelling. The atomic rows and planes of the crystal act as guides that steer energetic ions along preferred directions, which increase the ion penetration depth, affecting the final doping profile.

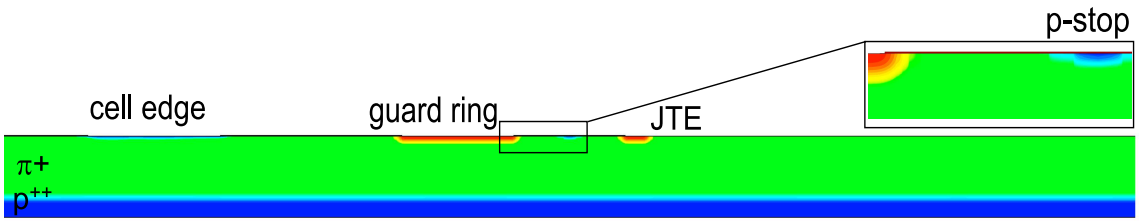


Fig. 4.15 *Thermal re-oxidation. The inset shows the p-stop region and the 40 nm thermally grown oxide, with the small consumption of silicon at the interface following the oxide growth.*

Gain layer implant

After the resist deposition of the GL mask, the process of ^{11}B implantation was performed, using the MC engine to guarantee better accuracy in the doping profile. This step was skipped in the case of PIN diodes.

Thermal drive-in of the GL implant

The drive-in for the GL and the p-stop activation was then performed, Figure 4.16.

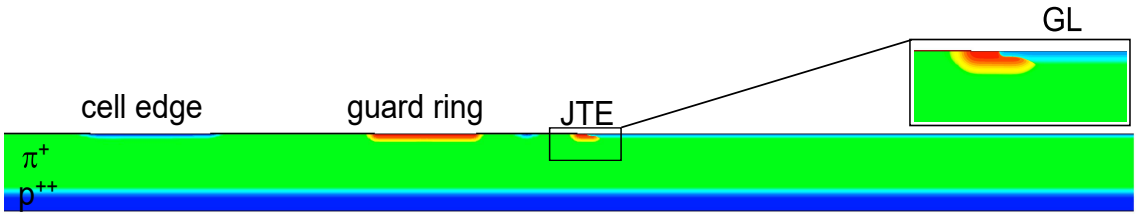


Fig. 4.16 *Drive-in for GL and p-stop. The inset shows details of the GL and JTE.*

Shallow N implant

After the resist deposition of the shallow N layer mask, the ^{31}P implantation was performed, using an analytical profile distribution as done for the JTE doping. The photoresist was then stripped off.

LTO deposition and densification

The Low Temperature Oxide (LTO) was deposited, via chemical vapour deposition (CVD). This step was simulated by an isotropic oxide deposition in an inert atmosphere. A thermal step to densify the LTO was then performed. In manufacturing this step is usually done

because LTO has poor mechanical properties, as it is often filled with cracks and voids. Thermal annealing after deposition densifies the oxide film and makes it more mechanically and chemically stable.

Implant activation

A final implant activation and diffusion were then performed, at a temperature of 1000 °C.

Contact VIAs etching

After the resist deposition of the vertical electrical connections (VIAs) layer mask, the oxide was etched anisotropically and the photoresist was stripped off, Figure 4.17.

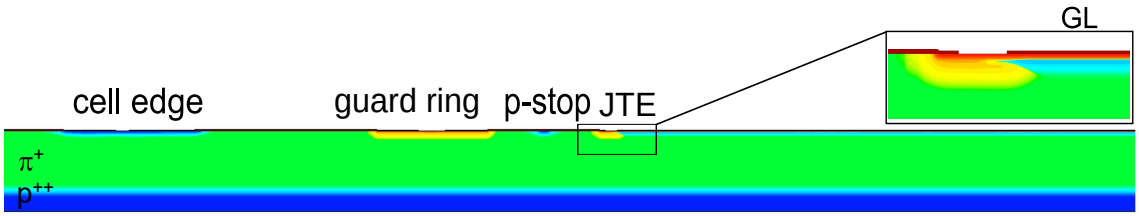


Fig. 4.17 *LGAD structure after the final thermal budget. The inset shows the etched oxide before the metal deposition for contact formation.*

Aluminium deposition

The final step of the simulation process was the isotropic deposition of aluminium over the entire structure and the filling of the VIAs. The photoresist mask of the contact layer was then applied and the metal anisotropically etched, with the remaining photoresist stripped off. Contact electrodes were defined and assigned to the metal regions, for the subsequent step of electrical simulation.

Figure 4.18 shows an example of a final Run 1 full structure obtained from SPROCESS, with detailed images of LGADs and PINs in Figure 4.19 and Figure 4.20.

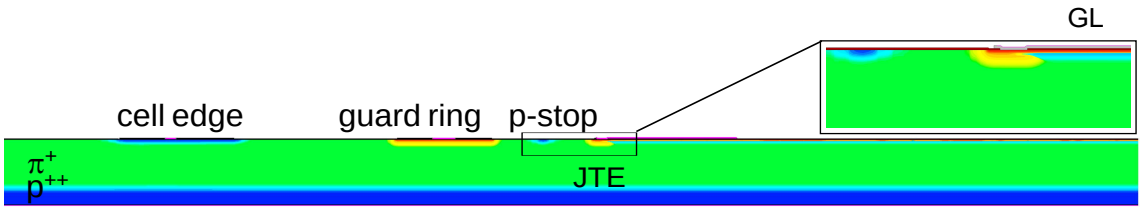


Fig. 4.18 *Final structure obtained from SPROCESS. The inset shows details of the aluminium cathode contact and the adjacent p-stop.*

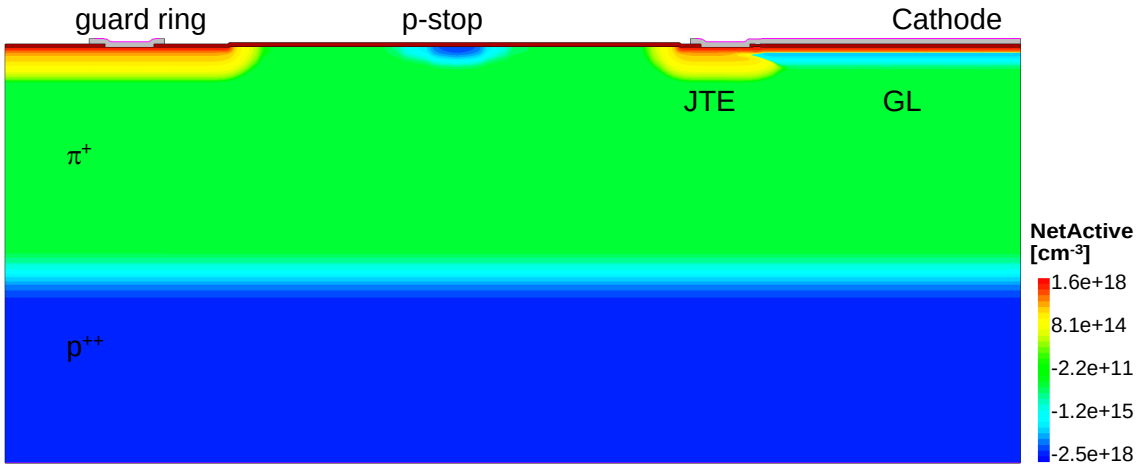


Fig. 4.19 *Detailed image of the final Run 1 LGAD structure, showing the doping and the GL below the n-well.*

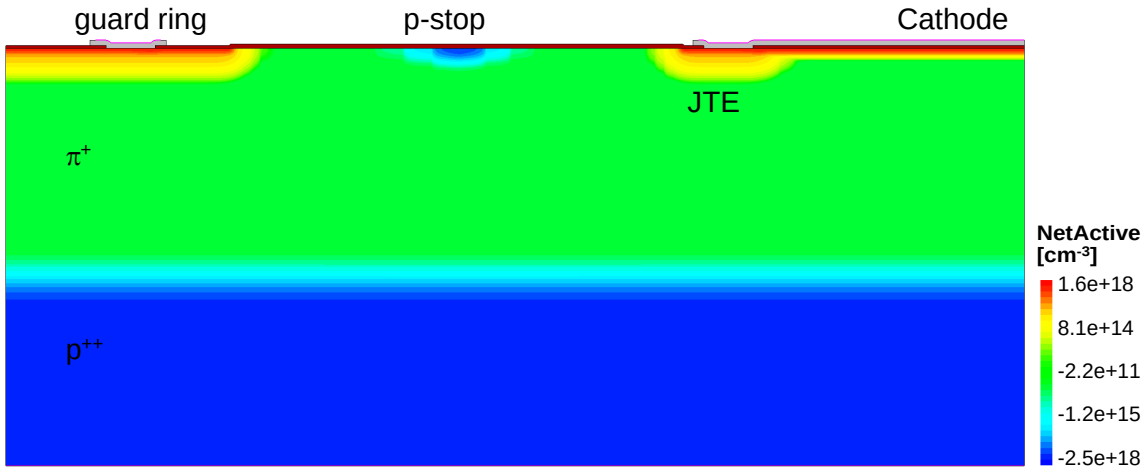


Fig. 4.20 *Detailed image of the final PIN structure, with doping.*

4.3.2 Process simulation and Secondary Ion Mass Spectrometry (SIMS)

The obtained structures from SPROCESS simulations have been compared with fabricated devices using Secondary Ion Mass Spectrometry (SIMS [77]) and Spread Resistance Measurement System (SRMS [78]) techniques, to verify the accuracy of the process modelling. Figure 4.21 shows the superimposed TCAD and SIMS/SRMS test results to compare with the predicted LGAD doping profiles for wafer 2 (WF2) of Run 1. Details of the profile in the GL region are shown in Figures 4.22 and 4.23.

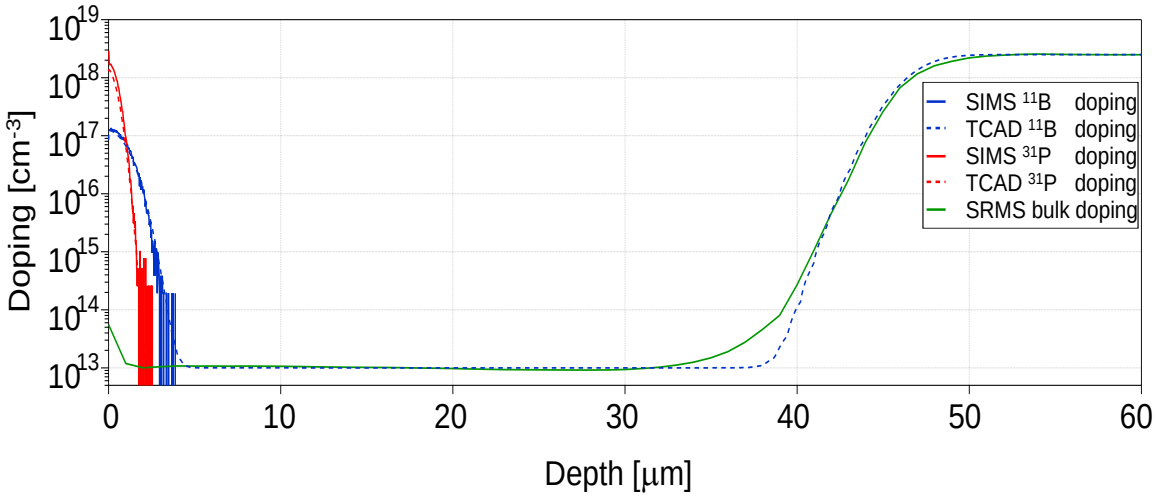


Fig. 4.21 *Full doping profile TCAD (dotted line) and SIMS/SRMS (thick line) comparison for LGAD from wafer 2 of Run 1.*

As shown in Figure 4.21 the measured doping profile agrees with the TCAD predictions. The epitaxial and substrate doping show the same predicted and measured value in both cases, with discrepancies observed near the transition boundary between the substrate and the epitaxial layer, starting at a depth of around 32 μm and around 42 μm. This deep region is not affected by impact ionisation or edge breakdown phenomena so such discrepancy is not relevant for predicting the gain performances of the device. An expanded view of the GL region with the ^{11}B and ^{31}P doping is shown in Figure 4.22. The discrepancy in total ^{11}B dose, obtained by integrating the profile from the top surface to the epitaxial region, nearly at the depth limit for the SIMS technique, is estimated to be around 4%, which is of the same order of accuracy of the SIMS technique [79]. In the same figure, the comparison of the ^{31}P doping shows a discrepancy which increases towards the device top surface but practically disappears, nearly at the depth limits of SIMS technique, at a depth of 1.3 μm, inside the region of the GL. The effect of the peak value of ^{31}P close to the surface is not crucial as this region is essentially undepleted, due to its high doping concentration.

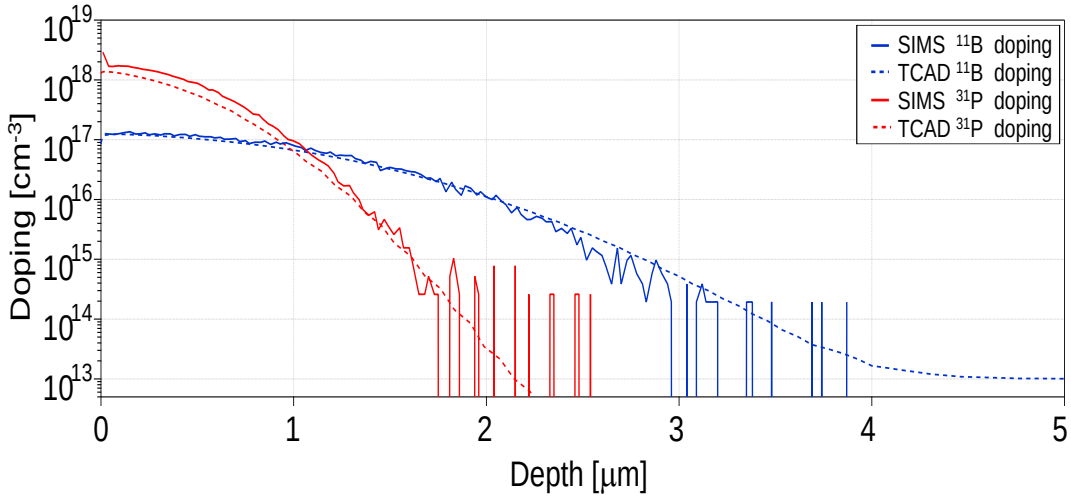


Fig. 4.22 *An expanded view of GL doping from Figure 4.21. SIMS (thick line) and TCAD (dotted line) results are shown for comparison.*

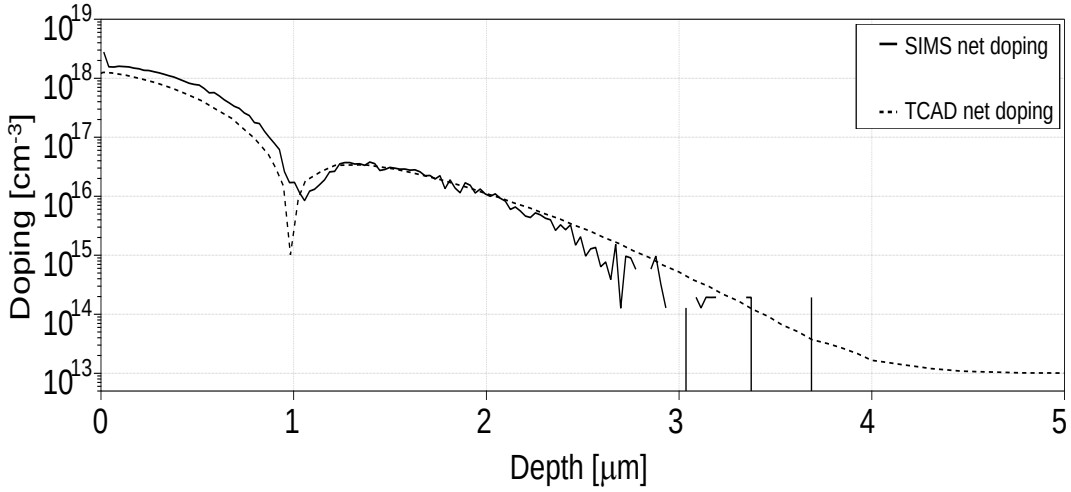


Fig. 4.23 *Absolute net doping profile comparison in the first 5 µm from the top surface from Figure 4.21. SIMS (thick line) and TCAD (dotted line) results are shown for comparison.*

The peak value of ^{11}B is found near the surface and the discrepancy observed between SIMS and TCAD results might also be due to different segregation coefficients for boron and SiO_2 used by default in SPROCESS. This value is not relevant for the operation of the device, as in this region the ^{11}B doping is fully compensated by the ^{31}P doping of the n-well. The absolute net doping profile comparison is shown in Figure 4.23. The discrepancy in peak value and location of the peak of ^{11}B in the gain layer, around 10%, might also be due to the granularity in SIMS data near the peak, noticeable in the same figure. Additional SIMS data from more samples would be required for a better statistical estimate.

Table 4.5 summarizes the TCAD and SIMS results for WF2 of Run 1.

Table 4.5 *Results of SIMS and TCAD doping comparison.*

Wafer 2	Nominal	SIMS	TCAD	error
^{11}B dose [cm^{-2}]	1.45×10^{13}	1.51×10^{13}	1.44×10^{13}	-4%
^{11}B absolute peak [cm^{-3}]	NA	1.35×10^{17}	1.24×10^{17}	-8%
^{11}B GL peak [cm^{-3}]	NA	3.79×10^{16}	3.41×10^{16}	-10%
^{11}B GL peak depth [μm]	NA	1.394	1.519	+9%

4.3.3 Electrical Simulation

The output structure obtained from SPROCESS was re-meshed to make it suitable for SDEVICE, the Sentaurus tool used for electrical simulation. Additional details on the meshing strategy used in this simulation phase will be provided in the final section of this chapter. DC and transient simulations were performed to investigate the effect of the doping of the gain layer on the IV and gain characteristics of the devices.

The DC simulation estimated the region of breakdown voltage. This allowed us to establish if, for a specific doping profile, a complete depletion of the active area of the device, including the gain layer, was reached before breakdown. A complete depletion is necessary to guarantee high gain and fast charge collection.

The transient simulation estimated the achievable gain of the LGAD device at different bias voltages. The gain is defined as the ratio of collected charge by the LGAD w.r.t. the charge collected by a PIN diode under the same biasing condition.

To reduce the computational time required for the simulation, a reduced 2D mesh of width smaller than that of the actual device was employed, both for LGAD and PIN, Figure 4.24. This mesh included the gain layer region and the lateral edge, including the JTE, p-stop and guard ring, to investigate the LGAD gain due to impact ionisation and avalanche breakdown triggered by the edge field. The reduced mesh size implied that the estimated current from the simulation did not represent the one obtained from real devices. However, it provided useful information on the expected operating voltage range of the LGAD device and methods to increase the breakdown.

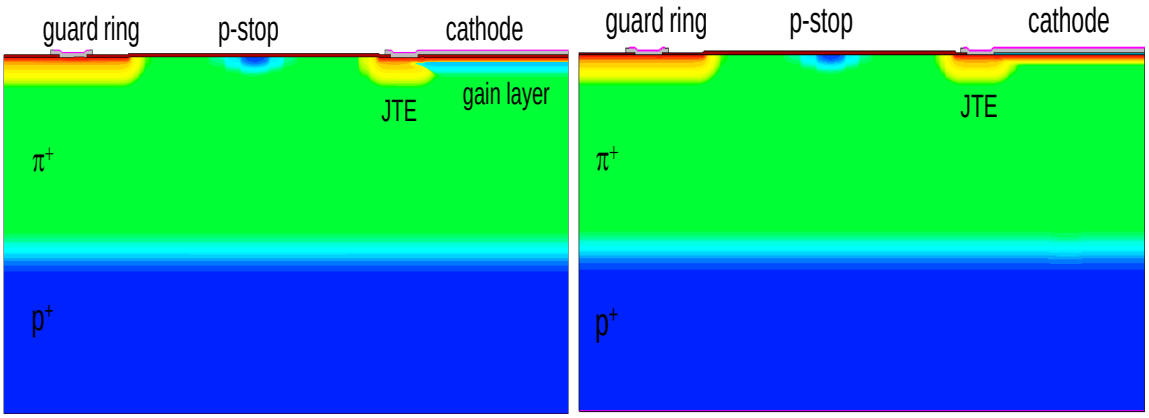


Fig. 4.24 *An example of the cross-section of an LGAD (left) and a PIN (right) used for breakdown and gain simulation.*

DC simulation

This simulation stage investigated the DC characteristics of the device, current versus voltage up to breakdown. As the LGAD devices exploit impact ionisation for their operation, proper modelling of such process, both in terms of physics and numerical methods used in the simulation, is crucial to estimate their performances with reasonable accuracy.

In general, the differential equations describing the current and including impact ionisation processes in the stationary case can be written in the form [80]:

$$\frac{dJ_n}{dx} = \alpha_e(x)J_n(x) + \alpha_h(x)J_p(x) \tag{4.1}$$

$$-\frac{dJ_p}{dx} = \alpha_e(x)J_n(x) + \alpha_h(x)J_p(x) \tag{4.2}$$

where J_n (J_p) is the electron (hole) current density and $\alpha_{e/h}$ is the ionisation rate, defined as the reciprocal of the mean free-path between consecutive scattering events that produce ionisation. Equations 4.1 and 4.2 show that the modelling of impact ionisation consists of two parts: one is about the ionisation rates and how they relate the carriers distribution function to physical quantities, like fields and applied potentials and the other one relates the ionisation rates to measurable values of current or charge.

Over the years, various models have been proposed in the literature to describe the experimentally measured ionisation rates to those theoretically deduced from basic assumptions related to the physics of the process [81]. However, recent works [82] on TCAD simulations on LGAD have shown significant differences between simulations and experimental results. The reasons behind such discrepancy are normally attributed to the difficulties in adequately modelling the internal electric field configuration of the devices, which the ionisation rates strongly depend on, and the inherent non-local process of the impact ionisation itself. The first cause of discrepancy can be eliminated, or at least much reduced, once an accurate description of the doping profile of the devices is implemented, as it has been attempted in this project, where a complete simulation of the fabrication process has been performed and results compared with actual structures. The second cause of discrepancy can be attributed to the difficulty in describing accurately the impact ionisation process in devices where a significant spatial and temporal non-uniformity of the electric field exists. This is expected to cause inaccuracies in LGAD device simulations due to the gain layer which alters the internal field configuration compared to the case of a simpler structure, like a PIN diode, with a more uniform internal field. The impact ionisation is inherently a non-local process, i.e. carriers do not gain energy at the same rate of the electric field spatial and temporal change [83], which

might cause inaccuracies in solutions of semiconductor equations based on drift-diffusion mechanisms, where only local variables are used. Also, higher-order semiconductor equation schemes, like the hydrodynamic model, rely on the local mean energy of the carriers [27], which implies that such schemes could lead to inaccurate predictions when impact ionisation effects are included. Besides the above, ionisation coefficients show exponential sensitivity to some of their parameters, which implies that very accurate knowledge of them is required to guarantee good predictions of device performances.

In SDEVICE there are six impact ionisation models available: the Van Overstraeten – de Man [84], the Okuto–Crowell [85], the Lackner [86], the University of Bologna [87], the New University of Bologna [88] and the Hatakeyama [89]. Only the first four models are relevant to this project since the Hatakeyama model describes the anisotropic behaviour in 4H-SiC power devices and the New University of Bologna model is an improvement developed to cover an extended temperature range between 25 °C and 500 °C. All the above models rely on the electric field as a local variable to calculate the impact ionisation rate, which is expressed by the general form (3.9) already described in Chapter 3 and reported here:

$$\alpha_v(F) = \alpha_v^{\infty} e^{\frac{-F_{th}^{v}}{F^{\delta}}} \tag{4.3}$$

Of the available models, the Van Overstraeten – de Man, the Lackner and the University of Bologna all have a probability of the type $e^{\frac{-1}{F}}$,i.e. with δ =1, whereas the Okuto-Crowell model follows a $e^{\frac{-1}{F^2}}$ law, i.e with $\delta = 2$. These two groups of laws represent complementary cases of impact ionisation processes, the one with $\delta = 1$ stemming from the Shockley theory of 'lucky' electrons [90] whereas the case with $\delta = 2$ is akin to Wolff's theory [53]. In the first case, the distribution function is peaked in the direction of the field whereas in the second case is nearly isotropic. In the model proposed by Baraff [57] a more realistic distribution function is a mixture of an isotropic part and a spike in the field's direction. Approximation of Baraff's model exists in the form of high-order polynomials [58] but they are not available by default in SDEVICE. Therefore, due to their complementarity, the ionisation models used in this project were:

- **University of Bologna :**

$$\alpha_v(F) = \frac{F}{a(T) + b(T) e^{\frac{d(T)}{F + c(T)}}} \tag{4.4}$$

with temperature-dependent coefficients:

$$a(T) = a_0 + a_1 t^{a_2} \qquad b(T) = b_0 \tag{4.5}$$

$$c(T) = c_0 + c_1 t + c_2 t^2 \qquad d(T) = \frac{\beta_\nu E_G}{e\lambda d_0}(d_0 + d_1 + d_2 t^2) \tag{4.6}$$

- **Okuto-Crowell :**

$$\alpha_\nu(F) = a(1 + c(T - T_0))F^\gamma e^{-(\frac{b(1+d(T-T_0))}{F})^\delta} \tag{4.7}$$

with

$$b_\nu = \frac{\beta_\nu E_G}{e\lambda} \tag{4.8}$$

where E_G is the energy bandgap, λ is the optical-phonon mean free path for the carrier and β_ν is a constant factor. Note that ν stands for electrons/holes.

Default parameter values of such models, as used in the device design phase, are reported in Appendix A.

The first step consisted of optimizing the PIN design to guarantee high breakdown voltage. Figure 4.25 shows an example of the optimization of the JTE design to increase the breakdown voltage by using different energies of the ^{31}P implants.

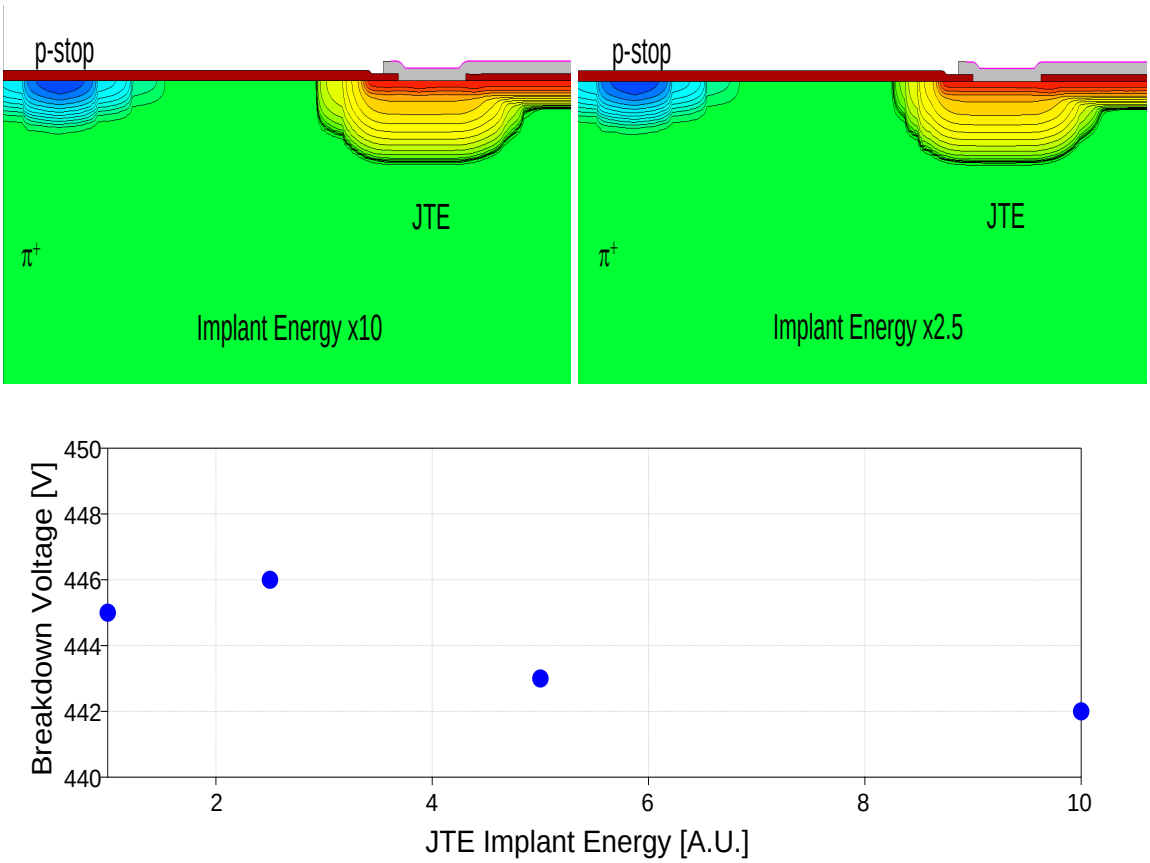


Fig. 4.25 *JTE optimization for PIN breakdown versus the energy of the implant. The top pictures show the contour lines of doping for two different energies of the ^{31}P implant. The bottom plot shows the predicted breakdown voltage peaked at around $\times$ 2.5 the reference ^{31}P implant energy.*

Several complete simulations of the fabrication process and electrical performances were executed to investigate the effect on the breakdown due to introducing a gain layer with different energy and dose configurations of the ^{11}B implant and additional thermal steps. An example of TCAD results for PIN and LGAD devices is shown in Figure 4.26, with values of breakdown voltages V_{bd} (defined as the point of maximum slope of current w.r.t. voltage) reported in Table 4.6. The predicted breakdown does not show a strong dependence on the impact ionisation model.

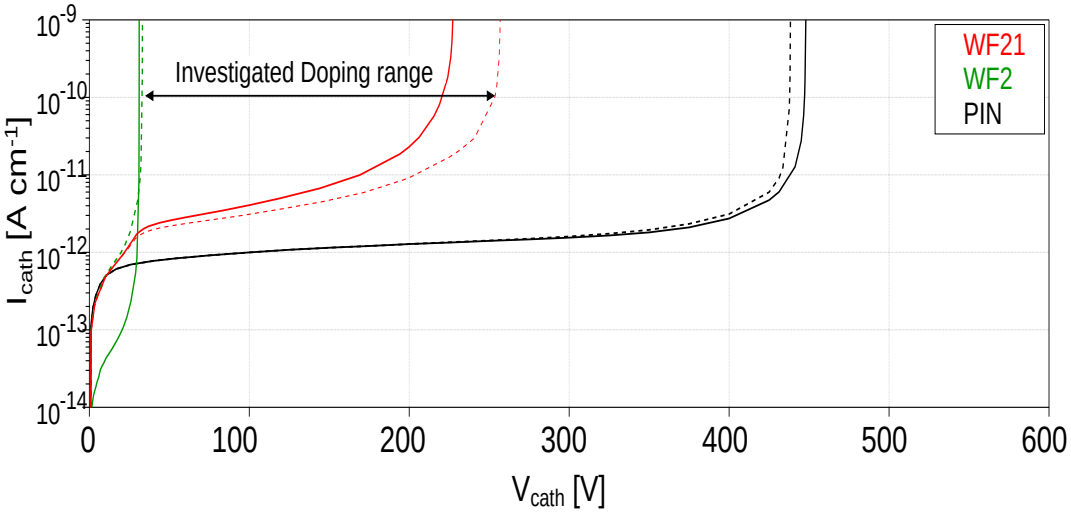


Fig. 4.26 *TCAD estimated breakdown voltage of PIN and LGAD of different gain layer doping. The IV curves for the Okuto-Crowell (plain) and the University of Bologna (dashed) models with default parameters are shown. T = 21 °C.*

Table 4.6 *Breakdown values from TCAD simulation in Volts.*

Ionisation Model	PIN	WF2	WF21
Okuto-Crowell	446	31	227
University of Bologna	437	33	256

Due to the uncertainties in impact ionisation modelling mentioned before, a conservative value of around half of the predicted PIN V_{bd} was set as the maximum allowed for the LGAD V_{bd}, as shown by the red curves in Figure 4.26, to minimise the possibility of breakdown ignited in the JTE region rather than the GL, which effectively would make the LGAD operate like a PIN diode. At the same time, a value corresponding to just the depletion of the gain layer was assumed as the minimum V_{bd}, as shown by the green curves in Figure 4.26. These two extremes of V_{bd} correspond to specific doping configurations of the LGAD gain layer that were then chosen as bounding limits for the ^{11}B implant and related thermal step. Subsequent values of dose and energy of the implant for the device fabrication were then selected within these boundaries and their values are reported in Table 4.2.

Figures 4.27 and 4.28 show in more detail the cross-section of LGAD devices of highest and lowest gain layer doping near breakdown. For the lowest doping, corresponding to WF21 in Table 4.2, the extension of the space charge follows the doping profile both in the GL and in the epitaxial region. At the same time, the predicted breakdown is lower by almost 50% than the one expected for the PIN diode. For the higher doping of WF2 in Table 4.2, the

extension of the space charge follows the gain layer doping profile in the proximity of the peak but does not cover it completely. This configuration was then assumed to be one of the highest doping of the gain layer, which would still guarantee a full depletion.

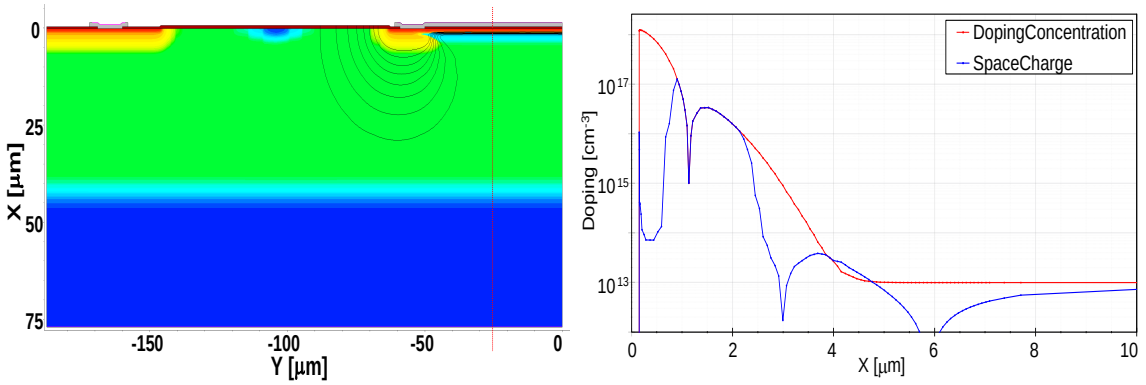


Fig. 4.27 *Left: cross-section of WF2 near breakdown, with equipotential lines. Right: the doping and space charge near breakdown along the cut line at Y = -25 µm (Okuto-Crowell impact ionisation model).*

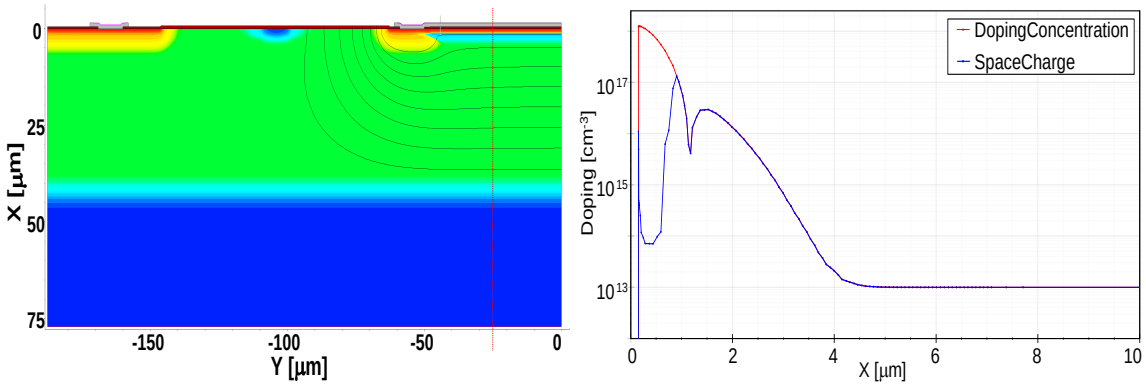


Fig. 4.28 *Left: cross-section of WF21 near breakdown, with equipotential lines. Right: the doping and space charge near breakdown along the cut line at Y = -25 µm (Okuto-Crowell impact ionisation model).*

For all these DC simulations the bias voltage was set to sweep from 0 up to 1 kV and applied to the cathode, while the substrate was grounded. The guard ring contact was kept floating. The physics model used for carrier mobility was the Philips mobility model [91], which unifies the description of majority and minority carrier bulk mobilities. Besides describing the temperature dependence of the mobility, the model takes into account electron–hole scattering, screening of ionised impurities by charge carriers and clustering of impurities. The different mobilities related to these effects are combined to give the bulk mobility for each carrier from Matthiessen's rule [92]. The model used to account for the

saturation velocity at high fields was the Canali mobility model [93], which includes also the modelling of saturation velocity with the parallel component of the electric field to the current as the driving force. To improve the numerical convergence, an interpolation of the driving force from the parallel electric field to the gradient of quasi-Fermi level was used [75]. The SRH model [94], with doping and temperature dependence, was used for the recombination. Additionally, due to the presence of a high field in the gain layer region of the LGAD, field enhancement recombination models were introduced in the form of the Hurkx model [95]. As the simulated structure included SiO_2, the effects on the mobility due to the presence of interface traps between Si and SiO_2 were also included in the simulations [96].

Transient simulation

This simulation step estimated the achievable gain from LGAD devices. In this context, the gain is conventionally defined as the ratio of collected charge by the LGAD, when the charge is injected via an ionisation process, compared to the charge collected by the PIN diode under the same operational conditions of ionisation, bias and temperature. The TCAD simulation set-up is shown in Figure 4.29.

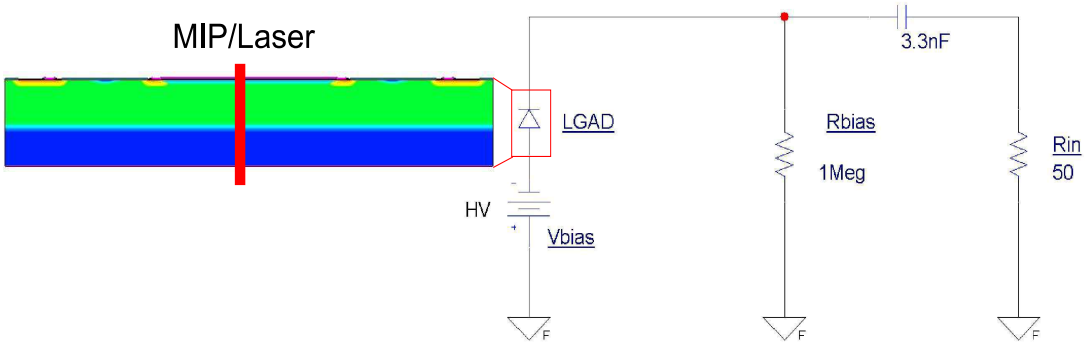


Fig. 4.29 *Transient simulation setup for charge collection. The RC network replicates the input stage of the custom-made Compact OPMD LGAD Amplifier (COLA) used for the tests.*

In the initial stage of transient simulation, the voltage bias is ramped up to a set value, expressed as a percentage of the predicted breakdown, due to the mentioned difficulty in predicting the correct breakdown voltage from simulation[3] In the transient simulation, after a settling time of 100 ns to guarantee the completion of any transient process, including capacitance charging effects, the charge is injected in the device using either monochromatic

[3]The reference doping chosen is of wafer 21 (WF21), Table 4.2.

light or charged particle injection. The resulting transient is followed for the subsequent 5 ns. The collected charge is estimated by integrating the current flowing through the device following the injection and for the duration of the transient. The simulation of charge injection in the device was carried out using two methods available in SDEVICE, Optical and HeavyIons.

The **Optical generation** method allows computation of optically created charge in the device from either a monochromatic source or from an input spectrum of different wavelengths. Details of the light source, including wavelength, angle of incidence, intensity, beam size and duration can all be set by the user. Various optical solvers can be selected, depending on the optical problem to be solved. In this project, a monochromatic source computation was used to match the generation of charge from an infrared laser that was used during the tests. The optical generation was based on the photon absorption using Beer's Law [97]:

$$G^{opt}(x,t) = J_0 F_t(t) F_{zy} \alpha(\lambda,x) e^{-\int_{x_0}^{x} \alpha(\lambda,\zeta) d\zeta} \tag{4.9}$$

where t is the time and $F(t)$ is the beam time behaviour function, a Gaussian of temporal width Δt and time spread σ_t. The non-uniform absorption coefficient along the vertical x-axis is described by $\alpha(\lambda,x)$ and $J_0 = \frac{P_0}{E_{ph}}$ where P_0 is the incident power density and E_{ph} the photon energy is the number of photons impinging onto the device surface. Finally, F_{zy} is the beam area factor. The extraction of the complex refractive index was performed along the line normal to the illumination window. Each separate region of silicon along the beamline has its complex refractive index [98], which was used to calculate the beam attenuation and the related charge generation.

The parameters used in the SDEVICE optical simulations are reported in Table 4.7. The beam wavelength, size and temporal characteristics of the parameters reflect those of the infrared (IR) laser setup used in the test. The power density was chosen to generate approximately 1 fC of charge in the device, i.e. similar to the average generated charge by a MIP when traversing a device of the same nominal 50 μm thickness.

Table 4.7 *Optical charge injection parameters used in the TCAD simulation.*

Wavelength [nm]	Power density [W/cm^2]	Beam area [μm^2]	Δt[ps]	σ_t[ps]
1024	3×10^3	20×20	39.5	0.25

In the region of the device where a charge is injected, either optically or using heavy ions, a high-resolution meshing is required, due to the high density of charge following ionisation. For the optical charge injection a meshing of 200 nm resolution, both along the radial and longitudinal beam direction, over a region of 15 μm width was applied as shown in Figure 4.30. During the initial stage of the transient, the time resolution was set to 1 ps. Such meshing parameters guaranteed adequate spatial resolution in the region covered by the laser beam and a temporal resolution high enough to avoid non-monotonicity issues in the solutions of the device equations. The meshing strategy is described in Section 4.4.

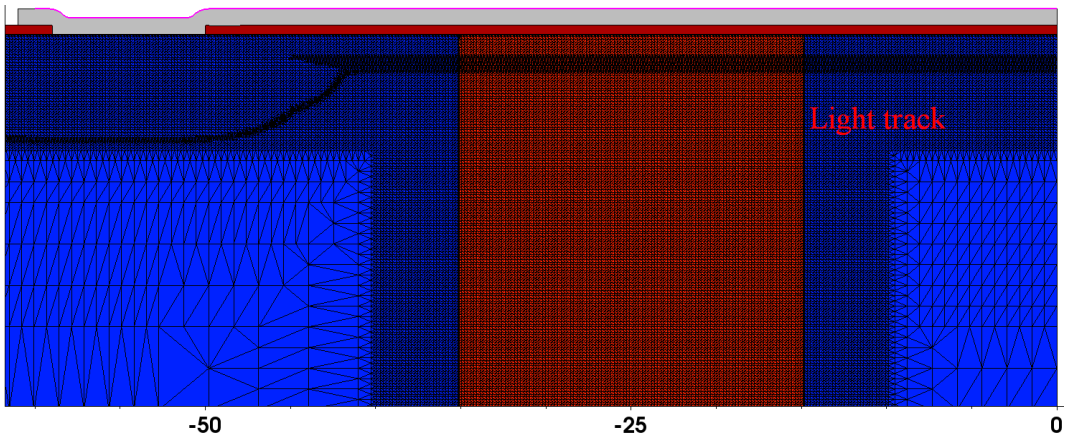


Fig. 4.30 *Device meshing for an optical charge injection in the PIN diode. The same meshing resolution was used for LGAD devices. Values in μm.*

For these simulations, the central vertical line of the beam was placed at Y = -25 μm as shown in Figure 4.30. Figures 4.31 and 4.32 show an example of charge density transient in an LGAD and PIN devices during the first 50 ps of the IR light pulse, using the parameters of Table 4.7. In the PIN case, Figure 4.32, the charge density increases uniformly across the active region during the IR pulse duration. In the LGAD case, Figure 4.31, there is an increase in charge density in the gain layer region, as a result of the impact ionisation process. However, during the transient, a higher concentration of holes, which exceeds the concentration of electrons, is seen accumulating near the bottom of the gain layer, while electrons are accumulating near the top. This process of charge accumulation in LGAD is opposite to the one observed in the PIN diode. Such charge distribution results in an internal electric field inside the gain layer which opposes the one from externally applied bias voltage, giving a screening which leads to gain suppression [99, 100]. The maximum charge density is found to be a few 10^{14} cm^{-3}, i.e. a small concentration compared to the typical dopants density in the gain layer region, usually in the range of 10^{16} - 10^{17} cm^{-3}. The effect on the space charge and electric field is of a small perturbation which might nonetheless

affect the ionisation coefficient, due to its exponential dependence on the field. The transient behaviour in Figure 4.33 shows that the electric field is not appreciably affected as a result of the changes in charge density. However, a close inspection of the ionisation coefficient α_e, shown in Figure 4.34 for the LGAD and for a transient of 1 ns, reveals that there is an initial drop in α_e in the first 100's ps of the transient. After around 1 ns the α_e coefficient is restored to almost the original value of the beginning of the transient. As this shielding effect is a consequence of charge creation following impact ionisation, it is then expected that a process of gain suppression might be an inherent feature of devices featuring an internal gain.

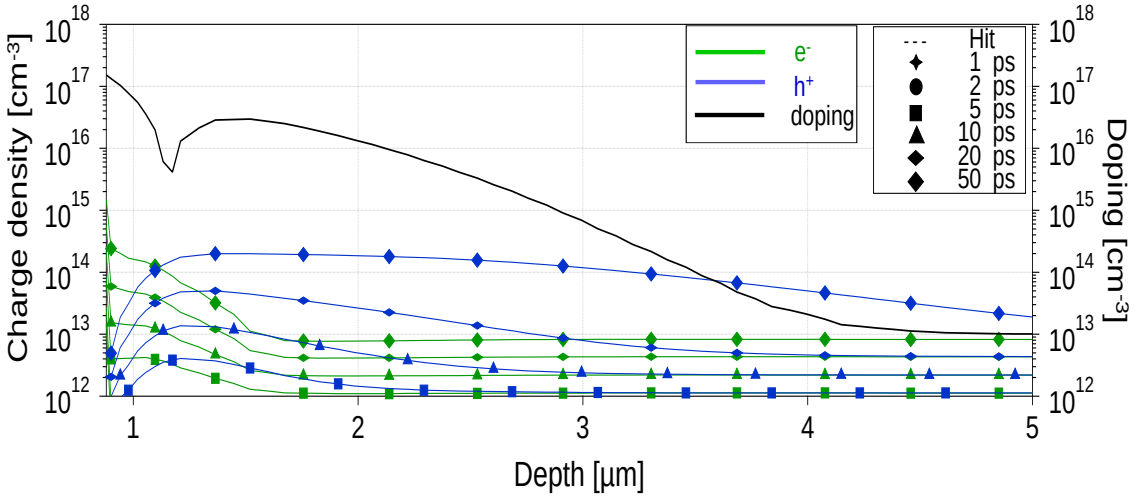


Fig. 4.31 *The transient charge density as a function of depth during an optical charge injection in an LGAD. The simulation uses the Okuto-Crowell impact ionisation model with default parameters. The charge densities are extracted at 90% of breakdown voltage and T = 21 °C.*

Figure 4.35 shows an example of the transient current for PIN and LGAD devices at different bias voltages, normalized with respect to V_{bd}, using the same setting parameters as in Table 4.7. The shape of the PIN transient current does not change appreciably with changes in bias voltage: after a fast rise, due to the Ramo's current, the pulse drops with a steep slope due to the electron's high mobility. After all electrons have been collected, at around 500 ps, the slope decreases, as the current is now due to the slower-moving holes. The entire transient pulse lasts around 1 ns. Conversely, for the LGAD device, there is a noticeable increase in the current slope with the bias voltage in the first stage of the transient, when electrons travel across the gain layer region until they get collected by the cathode. After all electrons have been collected, the charge multiplication stops and the current is now due to the motion of the holes to the opposite side of the detector.

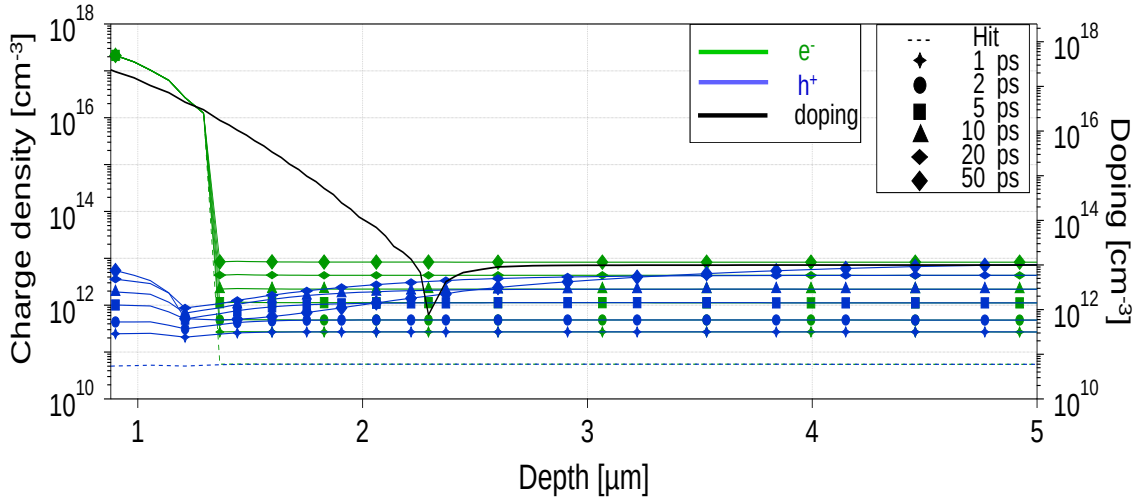


Fig. 4.32 *The transient charge density as a function of depth during an optical charge injection in a PIN diode. The simulation uses the Okuto-Crowell impact ionisation model with default parameters. The charge densities are extracted at 90% of breakdown voltage and T = 21 °C.*

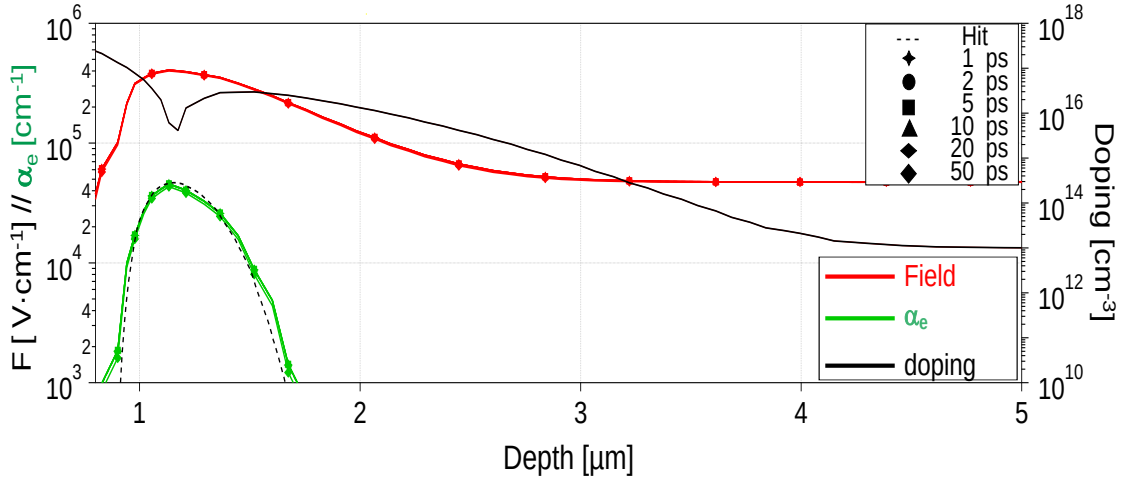


Fig. 4.33 *The transient electric field and ionisation coefficient α_e during an optical charge injection in an LGAD. The simulation uses the Okuto-Crowell impact ionisation model with default parameters. The variables are extracted at 90% of breakdown voltage and T = 21°C.*

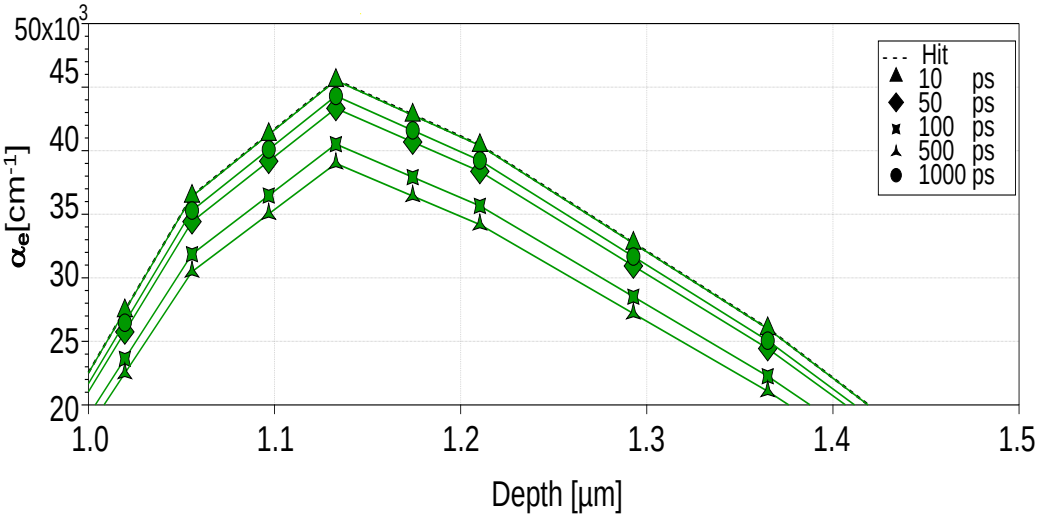


Fig. 4.34 *The transient ionisation coefficient α_e during an optical charge injection in an LGAD. The simulation uses the Okuto-Crowell impact ionisation model with default parameters. The ionisation coefficient is extracted at 90% of breakdown voltage and T = 21 °C.*

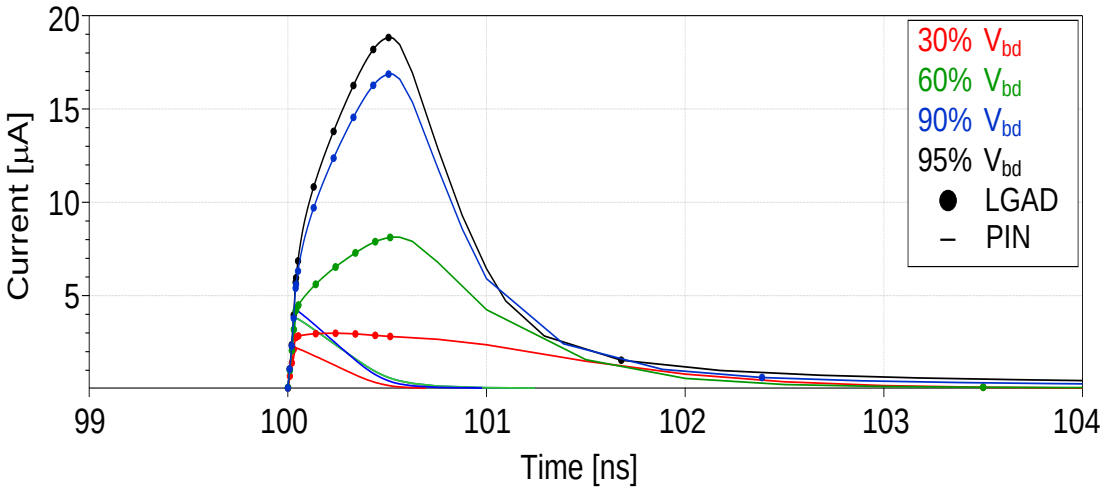


Fig. 4.35 *The transient current from optical charge injection in a PIN diode (line) and an LGAD from WF21 (line with markers) at different percentage of V_{bd}. The simulation uses the Okuto impact ionisation model with default values and T = 21 °C.*

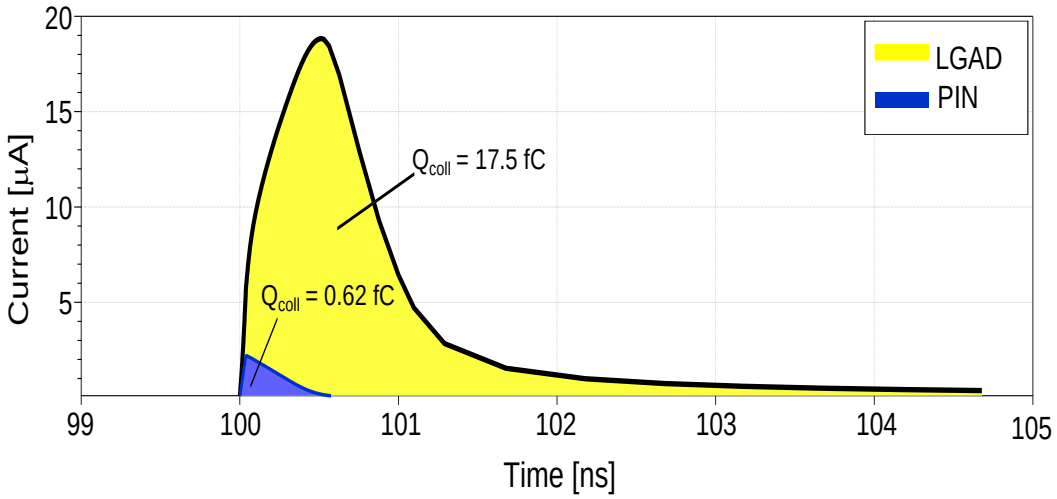


Fig. 4.36 *Maximum collected charge, in fC, for a PIN and an LGAD. The simulation uses the Okuto impact ionisation model with default values and T = 21 °C.*

The transient ends when all the holes have been collected by the anode, making the entire pulse duration last longer than in the case of PIN diode, i.e. around 2.5 ns. The ratio of the charge collected over the 5 ns of the transient simulation by the LGAD with respect to the PIN is the gain. An example is shown in Figure 4.36.

The Okuto-Crowell simulation of the collected charge normalized to the percentage of V_{bd}, for a PIN and a LGAD device from WF21, are reported in Table 4.8. In Table 4.9 results are reported using the University of Bologna model.

Table 4.8 *Optical charge injection Q_{coll} and Gain from TCAD for the Okuto - Crowell model at T = 21 °C.*

Breakdown voltage fraction [%]	30	60	90	95
I_{pk} PIN [μA]	1.9	2.1	2.2	2.2
I_{pk} LGAD [μA]	2.98	8.13	16.8	18.9
Q_{coll} PIN [fC]	0.60	0.60	0.61	0.62
Q_{coll} LGAD [fC]	4.8	8.8	15.7	17.5
Gain LGAD	8	14.6	25.7	28.2

Table 4.9 *Optical charge injection Q_{coll} and Gain from TCAD for the University of Bologna model at T = 21 °C.*

Breakdown voltage fraction [%]	30	60	90	95
I_{pk} PIN [μA]	1.9	2.1	2.2	2.2
I_{pk} LGAD [μA]	2.98	5.31	10.18	11.2
Q_{coll} PIN [fC]	0.60	0.60	0.61	0.62
Q_{coll} LGAD [fC]	3.8	6.6	12.7	14.01
Gain LGAD	6.33	11.01	20.81	22.59

An estimate of the slew rate of the sensor can be obtained from the slope of the transient current curves of Figure 4.35. At 50% of the peak current amplitude and with 95% of breakdown voltage bias, the slew rate is estimated to be 4.2 $\times$ 10^4 A$\cdot$ s^{-1} for the LGAD and a slightly higher value, around 9.5 $\times 10^4$ A$\cdot$ s^{-1}, for the PIN, for both the Okuto-Crowell and University of Bologna model. Despite a similar, or better, slew rate, the timing resolution achievable with a PIN is, in general, worse compared to an LGAD, because of the reduced jitter obtained from the higher signal in an LGAD, as shown in (3.27). For a Trans Impedance Amplifier (TIA) gain of 80 dB, i.e. 10^4, this would correspond to a measured slew rate of approximately 4.2 $\times$ 10^8 V$\cdot$ s^{-1}, to be compared with Figure 5.29.

The **HeavyIons generation** method provides a generation rate:

$$G(l,w,t) = G_{let}(l)R(w,l)T(t) \tag{4.10}$$

where G_{let} is the linear energy transfer (LET) generation density and $R(w)$ and $T(t)$ are functions describing the spatial and temporal variations of the generation rate. The LET generation density, $G_{let}(l)$, is given by:

$$G_{LET}(l) = a_1 + a_2 l + a_3 e^{a_4 l} + k[c_1(c_2 + c_3 l)^{c_4} + LET_f(l)] \tag{4.11}$$

where $LET_f(l)$ is a function of the length l. The spatial distribution $R(w,l)$ can be defined as either an exponential:

$$R(w,l) = e^{-\left(\frac{w}{w_t(l)}\right)} \tag{4.12}$$

or a Gaussian function:

$$R(w,l) = e^{-(\frac{w}{w_t(l)})^2} \tag{4.13}$$

where w is the perpendicular distance from the track and the characteristic distance $w_t(l)$ is a function of the length l. The default units for G_{LET}, w_t and l are pairs/cm^3, cm and cm respectively. They can be set to pC/μm and μm using the keyword *PicoCoulomb* in the HeavyIons directive. $T(t)$ defines the Gaussian describing the temporal window of the penetrating ion and is defined as:

$$T(t) = \frac{2e^{-(\frac{t-t_0}{\sqrt{2}s_{hi}})^2}}{\sqrt{2}s_{hi}\sqrt{\pi}(1+erf(\frac{t_0}{\sqrt{2}s_{hi}}))} \tag{4.14}$$

where t_0 is the time of penetration of the heavy ion and s_{hi} is the standard deviation value of the Gaussian. All the parameter values of the HeavyIon directive can be modified by the user to reflect the charge injection characteristics of interest.

To emulate the charge generation by a Minimum Ionising Particle (MIP), the parameters used for the heavy ion generation were obtained from Geant 4 simulations [101], using proton energy of 2.45 GeV impinging onto a 50 μm thick silicon. A constant LET of Gaussian shape was used, with parameters reported in Table 4.10.

Table 4.10 *MIP charge injection parameters used in TCAD simulations.*

Shape	LET [pC/μm]	w_t [μm]	t_0 [ns]	s_{hi} [ps]
Gaussian	1.79×10^{-5}	6.4×10^{-4}	100	0.1

The value used for the characteristic radial distance w_t required a higher resolution meshing for MIP compared to the optical injection case, with 0.3 nm minimum size along the radial direction from the track and 10 nm along the longitudinal direction, as shown in Figure 4.37. The time resolution was set to 100 fs during the initial stage of the transient simulation. In the simulation of MIP charge injection, the MIP track was generated at Y = -25 μm as shown in Figure 4.37 and with the parameters in Table 4.10. Figure 4.38 shows an example of a charge density transient in an LGAD during the first 50 ps of MIP charge injection. The corresponding transient in electric field and impact ionisation α_e are shown in Figure 4.39.

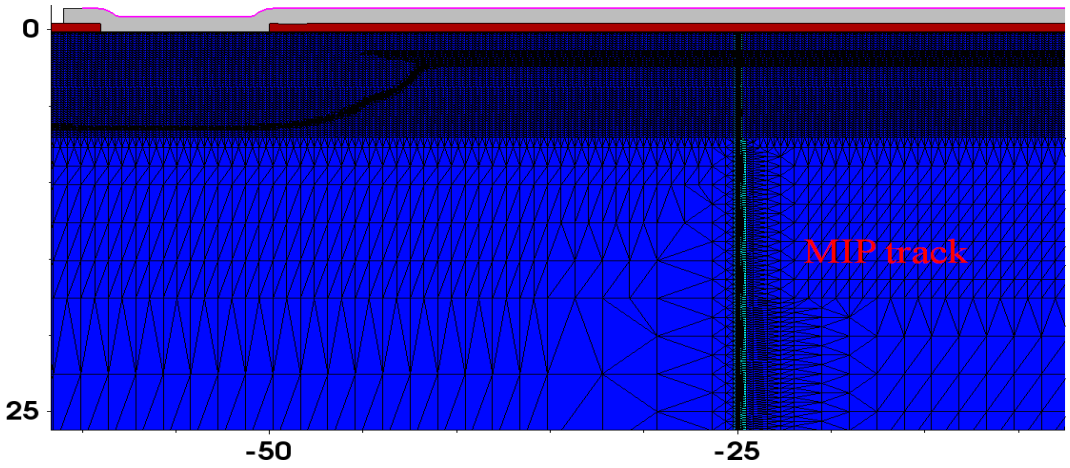


Fig. 4.37 *Device meshing for MIP charge injection. A meshing resolution of up to 0.3 nm was used in the radial direction along the MIP track. Values in µm.*

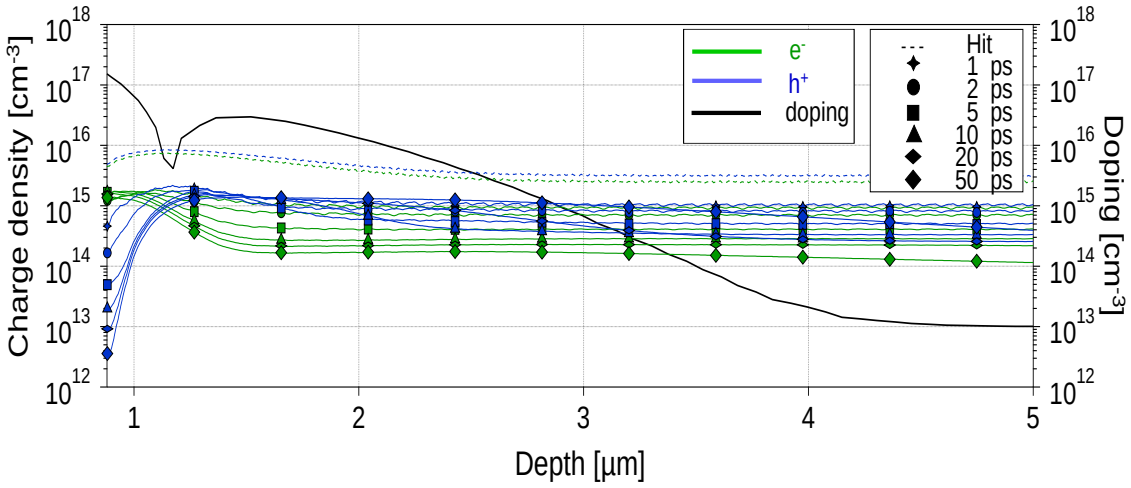


Fig. 4.38 *The transient charge from a MIP charge injection in an LGAD. The simulation uses the Okuto impact ionisation model with default values, 90% of breakdown voltage and T = 21 °C.*

The increase in charge density in the GL region is observed as in the case of optical charge injection, which was shown in Figure 4.31. However, the density of charge in the case of MIP charge injection is higher compared to the IR charge injection, which causes a higher gain suppression. This can be seen in Figure 4.40, where an example of the transient behaviour of the α_e ionisation coefficient is plotted as the function of the depth at different times of the transient following MIP charge injection. The initial drop in α_e is more pronounced compared to the case of optical charge injection but the value of α_e is restored at the end of

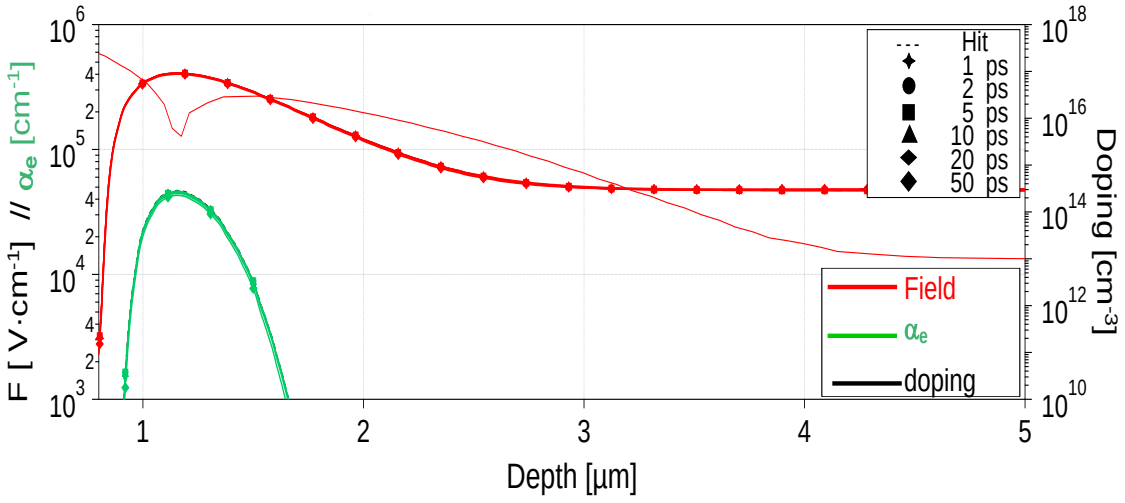


Fig. 4.39 *The transient electric field and impact ionisation α_e from a MIP charge injection in an LGAD. The simulation uses the Okuto impact ionisation model with default values, 90% of breakdown voltage and T = 21 °C.*

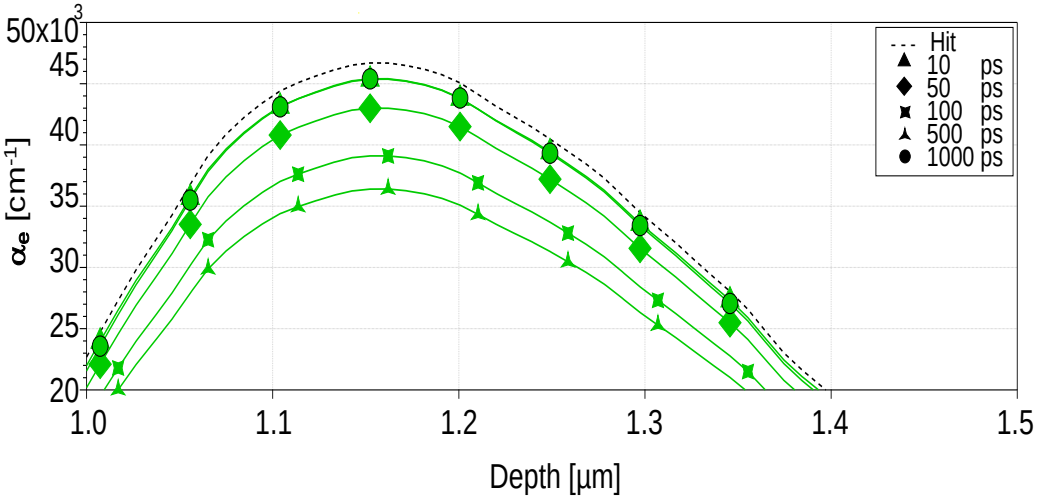


Fig. 4.40 *The transient ionisation coefficient α_e during MIP charge injection in an LGAD. The simulation uses the Okuto impact ionisation model with default values, 90% of breakdown voltage and T = 21 °C.*

the transient. Figure 4.41 plots the maximum value of ionisation coefficient α_e vs. time for both cases of IR and MIP injection. The minimum value of α_e is predicted at around 500 ps, i.e. when the pulse current reaches the peak. At this point, all electrons have been collected and the effect of gain suppression due to the created holes that are still remaining in the GL is at its maximum. For the MIP injection at 50% of the peak current amplitude and with 95% of breakdown voltage bias, the slew rate is estimated to be 3.04×10^4 A· s^{-1} for the LGAD and a higher value, around 2.5×10^5 A· s^{-1}, for the PIN, for both the Okuto-Crowell and University of Bologna model. The effect of gain suppression predicts a reduced slew rate in the case of the LGAD.

The simulation results of the collected charge for PIN and an LGAD device from WF21 using MIP charge injection and two ionisation models are reported in Tables 4.11 and 4.12. Figure 4.42 shows the simulation results of gain normalized to the percentage of V_{bd} using MIP and IR charge injection and two ionisation models for an LGAD device from WF21.

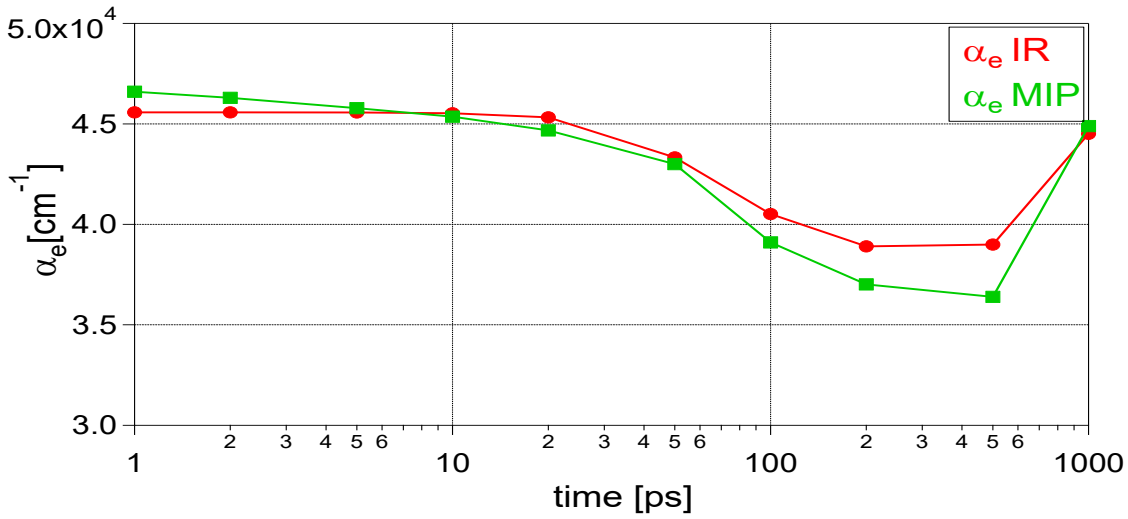


Fig. 4.41 *Maximum ionisation coefficient α_e vs. time during IR and MIP charge injection in LGAD. The simulations use the Okuto-Crowell ionisation models with default values at T = 21 °C.*

Table 4.11 *MIP charge injection Q_{coll} and Gain TCAD results for the Okuto - Crowell model at T = 21 °C.*

Breakdown voltage fraction [%]	30	60	90	95
I_{pk} PIN [μA]	2.8	3.1	3.2	3.2
I_{pk} LGAD [μA]	1.99	4.79	9	9.98
Q_{coll} PIN [fC]	0.87	0.87	0.88	0.89
Q_{coll} LGAD [fC]	3.3	5.4	9.5	10.6
Gain LGAD	4.1	6.75	11.87	13.1

Table 4.12 *MIP charge injection Q_{coll} and Gain TCAD results for the University of Bologna model at T = 21 °C.*

Breakdown voltage fraction [%]	30	60	90	95
I_{pk} PIN [μA]	2.7	3.1	3.2	3.2
I_{pk} LGAD [μA]	2.0	3.7	7.5	8.6
Q_{coll} PIN [fC]	0.87	0.87	0.88	0.89
Q_{coll} LGAD [fC]	2.6	4.1	7.3	8.6
Gain LGAD	3.25	5.12	9.12	10.71

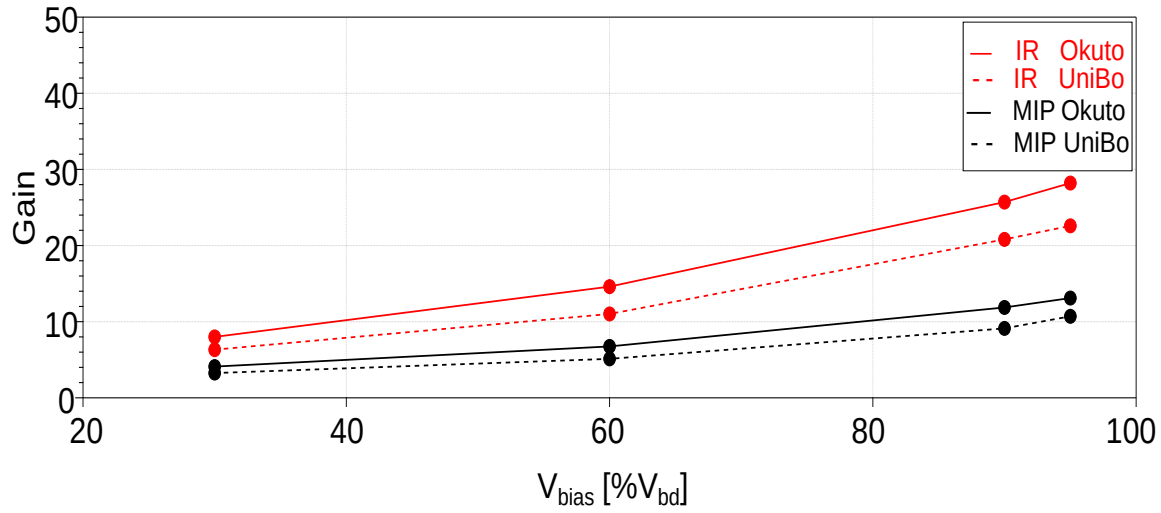


Fig. 4.42 *The simulation results of gain normalized to the percentage of V_{bd} using MIP and IR charge injection for an LGAD from WF21. The simulations use the Okuto-Crowell and University of Bologna ionisation models with default values at T = 21 °C.*

4.4 Meshing strategy

After the step of process simulation, the mesh structure obtained from Sentaurus Process needs to be re-meshed to make it suitable for electrical simulation, as the requirements of the device simulator differ from those of the process simulator. This step usually requires re-meshing with specific refinements in sensitive regions, for example near the *pn* junctions and, in this project, around the gain layer region of the LGAD devices. The meshing steps required are:

- Clear the previous process mesh;
- Set high-quality Delaunay [102] meshes;[4]
- Specify the refinement boxes in regions of interest, i.e. near the *pn* junction and, particularly, in the gain layer regions of the LGAD and around the JTEs;
- Specify the criteria used for the mesh spacings in these regions.

When solving the semiconductor equations using numerical methods, one must ensure that both n and p carrier concentrations are always positive to avoid unphysical solutions and to guarantee the monotonicity of the solutions to avoid spurious space oscillations [17]. The approach followed in this project was based on setting the spatial resolution in the refinement boxes of the order of the Debye length and, during the transient simulations, setting the time step intervals of the order of the dielectric relaxation time.

The first requirement on mesh size stems from the fact that the Debye length sets the length scale of charge redistribution at the interface between two regions of different doping. To resolve charge variations in space it is necessary to use a mesh resolution of this size. In equilibrium, the Debye length can be expressed as:

$$L_D = \sqrt{\frac{\varepsilon_s k_B T}{e^2 N}} \tag{4.15}$$

where N is the doping density, k_B the Boltzmann's constant, ε_S the semiconductor electrical permittivity, e the unit charge and T the absolute temperature. For example, in silicon at room temperature and with a doping of $N = 10^{16}$ cm^{-3}, the Debye length is approximately 40 nm and around 13 nm with $N = 10^{17}$ cm^{-3}. In non-equilibrium situations, such length can

[4]The Delaunay triangulation is often employed to generate meshes for space-discretised solvers for physics simulation because it maximizes the minimum angle of the resulting mesh and fast triangulation algorithms have been developed. Both aspects are important in the field of semiconductor device simulations.

vary depending on conditions of charge injections, which makes it then important to choose a mesh size, particularly in sensitive regions, smaller than this length. For this project, a minimum mesh size of 200 nm (0.3 nm) for the optical (MIP) charge generation was used.

The second requirement relates to the time discretization. To avoid non-monotonicity in the solutions, the time step during dynamic simulations should be smaller than the dielectric relaxation time, which sets the typical time for the majority carriers' fluctuations to decay under the field they produce. Such time is given by:

$$t_{dr} = \frac{\varepsilon_s}{eN\mu} \tag{4.16}$$

where μ is the mobility of the carriers and N their density. For example, in silicon at room temperature and with a doping N = 10^{16} cm^{-3}, the dielectric relaxation time is around 2.34 ps and drops to 0.3 ps with N = 10^{17} cm^{-3}. When running time-dependent analysis the time step should be chosen to be smaller than these values. In this project time steps as small as 1 ps during the optical charge generation and 50 fs during the MIP charge generation were used.

4.5 Validity of the local model

As it has been mentioned before, the impact ionisation process is inherently a non-local process. The related ionisation coefficients usually depend not only on the local value of the field but also on the previous history of the carriers. Only under relatively strict conditions the ionisation coefficients can be assumed to be independent of position and dependent only on the local value of the electric field [80]:

- fields spatially uniform and time-independent;
- long avalanche regions;
- carriers rapidly reaching a spatial steady state under the competing influences of the field and phonon collisions.

The validity of a local model approach can be assumed if:

$$\frac{1}{\alpha(F_m)} \frac{1}{F_m} \frac{dF}{dx} < 1 \tag{4.17}$$

where $\alpha(F_{\mathrm{m}})$ is the ionisation coefficient at the maximum field F_m, i.e. when the change in electric field over the minimum ionisation distance is less than the maximum value of the field and with the width of the avalanche region being greater than the minimum ionisation distance $\frac{E_{\mathrm{i}}}{eF_{\mathrm{m}}}$, with E_i being the ionisation threshold. The validity of the local model has been verified from TCAD simulation. In Figure 4.43, the electric field F, the derivative dF/dx and the ionisation coefficient α_e are shown, along with their maximum values as the function of the depth for the biasing condition of 95% of the breakdown voltage in an LGAD of WF21. It can be seen that, with the field derivative being of almost the same magnitude as the field in the gain layer region, the ratio with the maximum value of α_e and F satisfies condition (4.17), i.e. the ratio is $\ll 1$. The minimum ionisation distance, given by $\frac{E_{\mathrm{i}}}{eF_{\mathrm{m}}}$, is approximately 25 nm, with the values of E_i of around 1.5 eV [103] used in this project. As this distance is much less than the width extension of the gain layer region, it can be concluded that a local model for ionisation is valid for the simulation of the devices investigated in this project.

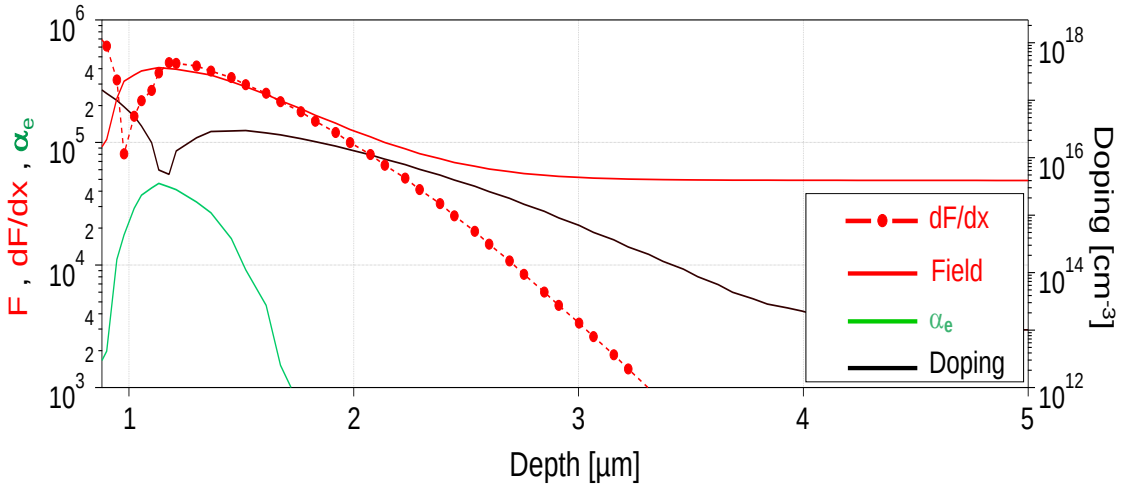


Fig. 4.43 *The electric field F [V·cm^{-1}], its derivative dF/dx [V·cm^{-2}], the electron ionisation coefficient α_e [cm^{-1}] at 95% V_{bd} and doping [cm^{-3}] for an LGAD of WF21.*

Chapter 5

Fabrication and Performances of Low Gain Avalanche Detectors

After processing at the Te2v semiconductor foundry the wafers were shipped to different test facilities, where the electrical performances were evaluated. The facilities included the Oxford Physics Microstructure Detector (OPMD, University of Oxford), Rutherford Appleton Laboratory (RAL) and the University of Birmingham. Additional tests were performed at the Centre of Medical Radiation Physics, Australia (CMRP, [104]). This chapter describes the fabricated devices, the experimental techniques and the setups used for their characterization and presents test results.

Section 5.1 provides details of the fabricated devices. The test setups used for the DC and AC tests, which included current versus voltage (IV), capacitance versus voltage (CV), gain and timing using a fast infrared (IR) laser and a radioactive source are shown in Section 5.2. In Section 5.3 the readout setups used for the gain and timing tests are presented. The IV and CV test results are shown in Section 5.4 and the gain and timing test results in Section 5.5. The last Section 5.6 describes the initial tests performed on LGAD devices to assess their potential use in radiation dosimetry for medical and space applications.

5.1 Description of fabricated devices

The first fabrication run of this project investigated the achievable performance using the available technology process from the silicon foundry, compared the predictions of the simulations with test results and selected the best-performing process for the design of the second run. The second fabrication run included a full array of 15 $\times$ 15 LGAD cells and it addressed the specifications of the ATLAS High Granularity Timing Detector, HGTD [10].

5.1.1 Run 1 devices

In the first fabrication run a total of eight 6" wafers were processed, with layouts described in Table 4.1 and gain layer characteristics reported in Table 4.2. A picture of one of the 6" wafers of the first run is shown in Figure 5.1.

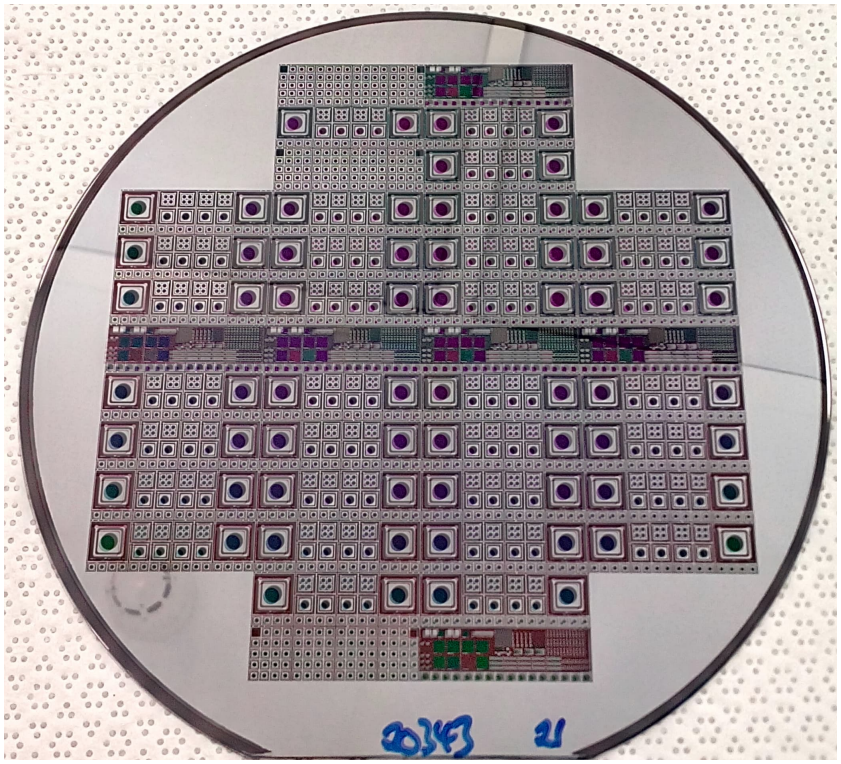


Fig. 5.1 *A picture of one of the 6" wafers of Run 1.*

The processing of the wafers did not include the backside metalization as the resistivity of the highly doped silicon substrate was low enough to provide a very good electrical contact to the chuck of the probe station used for the subsequent tests. Some of the diced devices used for gain and timing measurements were initially glued onto the bias pad of the amplifier board used for the test, see section 5.3, but it was found that using a double-sided Cu tape

provided an excellent electrical connection, without the need for the conductive glue.
After inspection and initial IV and CV tests on wafers, several devices, specifically the type 3 (1 mm) devices of Table 4.1, from wafer 2, wafer 8 and wafer 9 (WF2-WF8-WF9) were diced to perform IV, CV, gain and timing measurements.

Two wafers (WF21 and WF2) were initially diced using a laser cutting system, operating at 1028 nm with a power of 10 W and with a laser spot size of 25 μm, which guarantees high cutting precision, as shown in Figure 5.2. It was however found that such a dicing process introduced defects in the devices which caused a lower breakdown voltage compared to the measurements on the wafer before dicing. The cause of the lower breakdown was attributed to defects created in the silicon crystal during the dicing process near the surface of the devices. Subsequent thermal annealing of the devices at 150 °C for 2 hours showed to fully recover the pre-dicing behaviour. However, due to the extra process required, it was decided to perform a more conventional saw dicing for all the subsequent wafers.

Figure 5.3 shows a die placed on the custom Compact OPMD LGAD Amplifier (COLA) used for the gain and timing tests, to be described next.

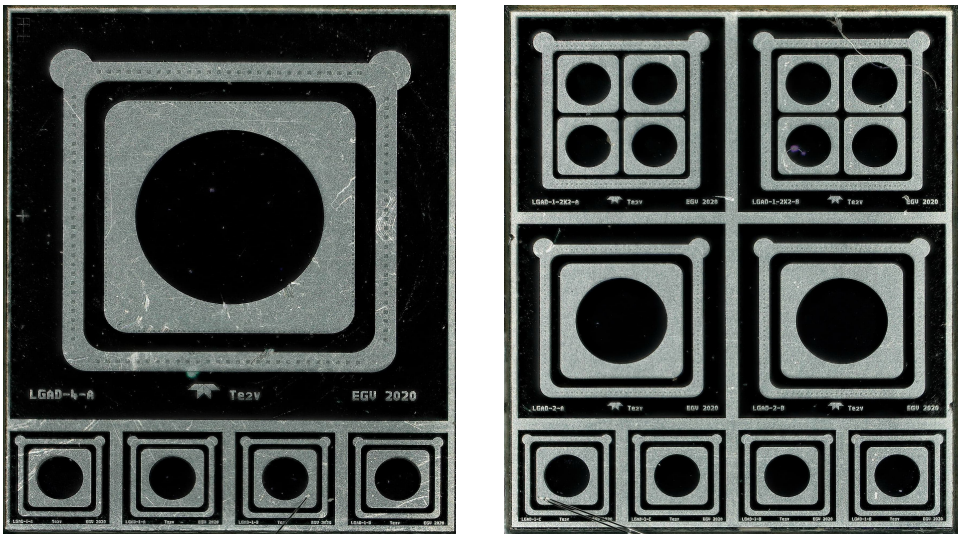


Fig. 5.2 *Laser cut dice from Run 1. Left: type 1 and 3 devices (4 mm and 1 mm cathode size). Right: Type 3, 2 and 4 devices (1 mm, 2 mm and 2 × 2 array of 1 mm cathode size). The circular aperture in the metal cathode is for the laser charge injection test.*

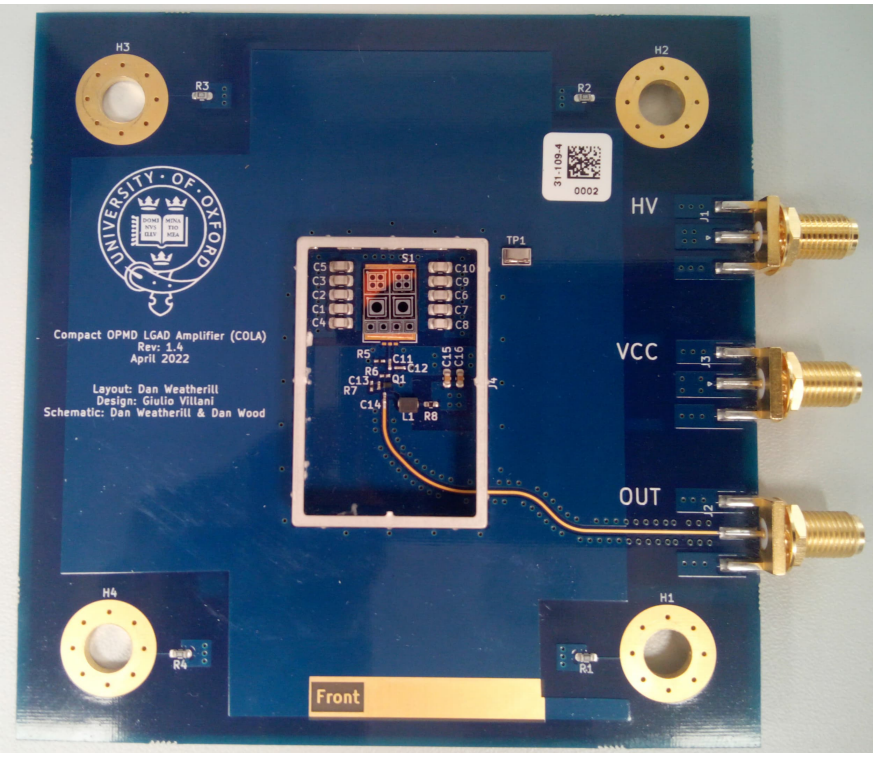


Fig. 5.3 *Run 1 die placed and bonded onto a COLA board for gain and timing measurements. The device is placed onto the biasing pad on board using a double-sided Cu tape.*

5.1.2 Run 2 devices

In the second fabrication run a total of four 6" wafers were processed. The layout and the gain layer doping of the devices are reported in Tables 4.3 and 4.4 respectively. A picture of one of the wafers is shown in Figure 5.4. Run 2 included 15×15 cell arrays, which were diced at the silicon foundry. A picture of one of the array and a detailed image near a corner is shown in Figure 5.5.

The readout of an entire Run 2 array would require a dedicated ASIC, which is not yet available. Therefore only IV and CV tests were performed, for an initial evaluation of the uniformity of the array, using a semiconductor analyser coupled to a switching matrix. The development of a custom ASIC for gain and timing measurements is ongoing and will be briefly mentioned in section 7.2 of Chapter 7.

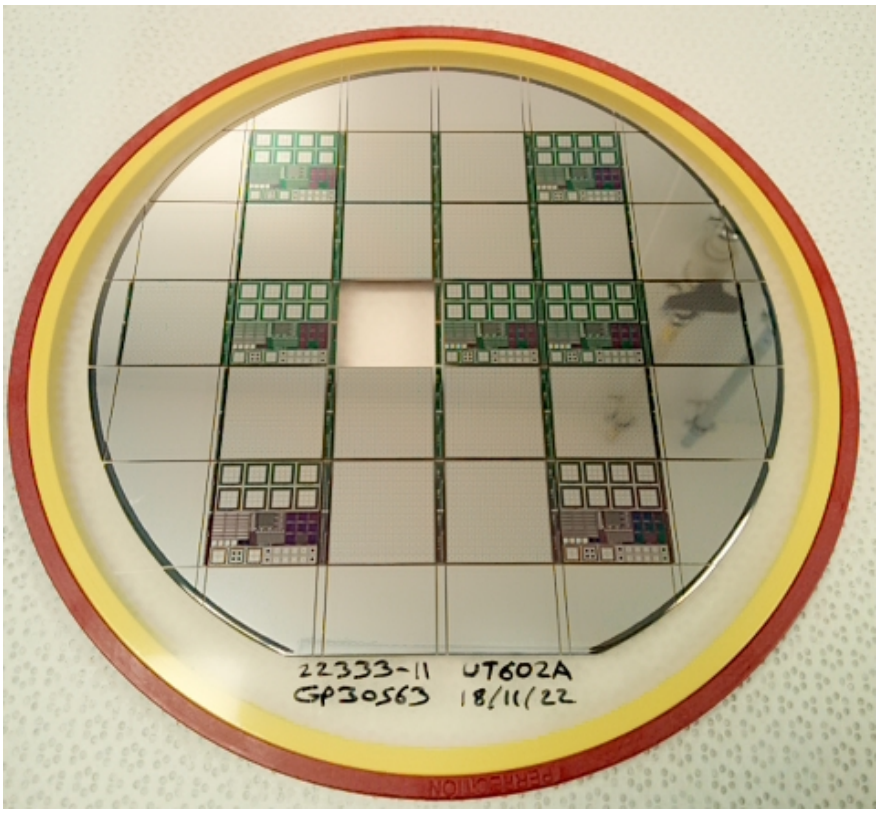


Fig. 5.4 *A picture of one of the 6" wafers processed in Run 2.*

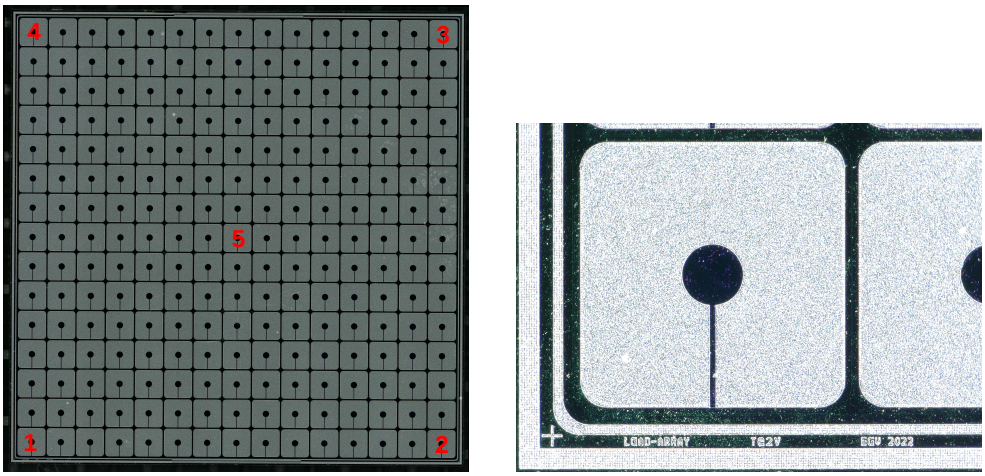


Fig. 5.5 *Left: A Run 2 array, consisting of 15 × 15 LGAD cells. Right: magnification near the bottom left corner, showing the central hole and vertical slit across the top metal layer for laser charge injection test. The five points shown on the left picture were used for the CV test.*

5.2 Experimental setups

Measurements of the IV and CV characteristics of the fabricated devices were performed at all the institutes involved in this project. The results presented here were obtained at OPMD, RAL, Birmingham and CMRP. The IV and CV tests were performed on wafers and dice using a probe station coupled to a high voltage (HV) unit. Gain and laser timing tests were performed using a laser setup with a readout based on a two-stage fast TIA (Transimpedance amplifier) and a fast scope for data acquisition (DAQ). Time resolution tests with charged particles were performed using a coincidence setup based on a ^{90}Sr source coupled to a readout of the same type as the one used for the laser tests. Measurements for dosimetry applications of LGADs were performed by CMRP using an 8 MeV ion microbeam at the IBIC facility at the Australian Nuclear Science and Technology Organisation (ANSTO).

5.2.1 IV and CV test setup

The IV and CV tests performed at OPMD and RAL used a Semiprobe SA12 probe station coupled to a CS4200 Keithley semiconductor analyser equipped with a Keithley 2657A for high voltage bias, Figure 5.6.

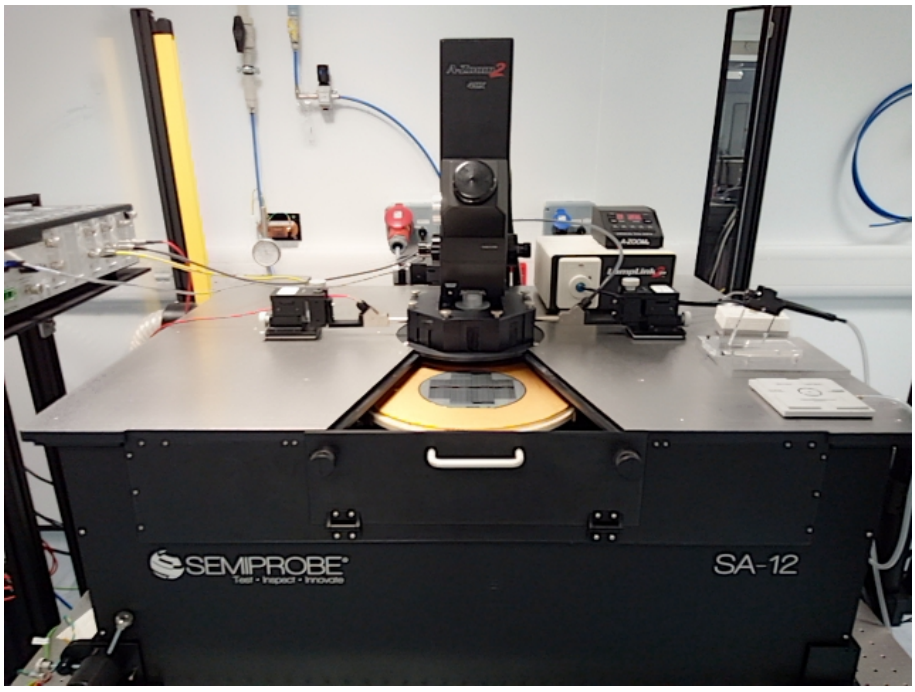


Fig. 5.6 *The probe station setup for IV and CV wafer test at OPMD. The grounded chuck can be cooled to a temperature of -30 °C for the tests of irradiated devices.*

5.2.2 Laser source test setup

A schematic diagram of the laser setup used for gain and timing measurements is shown in Figure 5.7 and the actual OPMD setup is in Figure 5.8. For these measurements, we used a laser pulse from a QD laser QCED106G-6410 [105] with wavelength 1064 nm, 15 ps pulse width and 200 fs jitter. As shown in Figure 5.7, the laser pulse, which provides the trigger signal to the DAQ, is coupled via an FC/APC connector (1) to a 50/50 beam splitter (2).

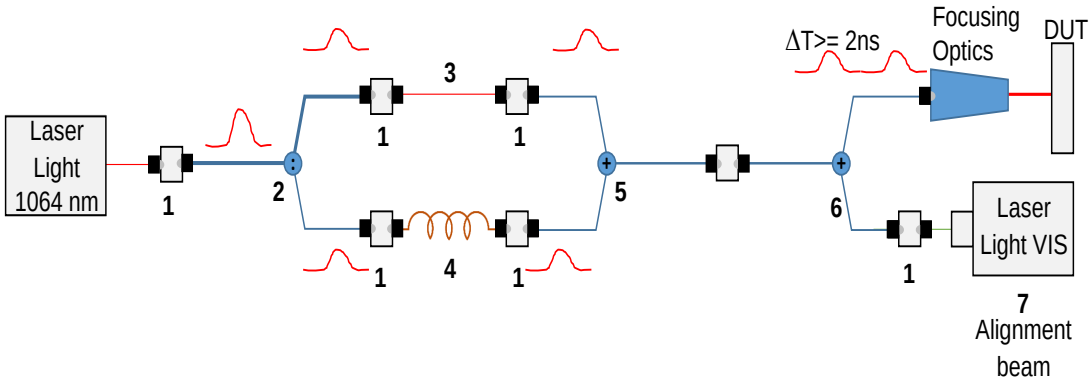


Fig. 5.7 *Diagram of the OPMD laser test setup for gain and timing measurement, based on the Split Delay Recombine (SDR) technique.*

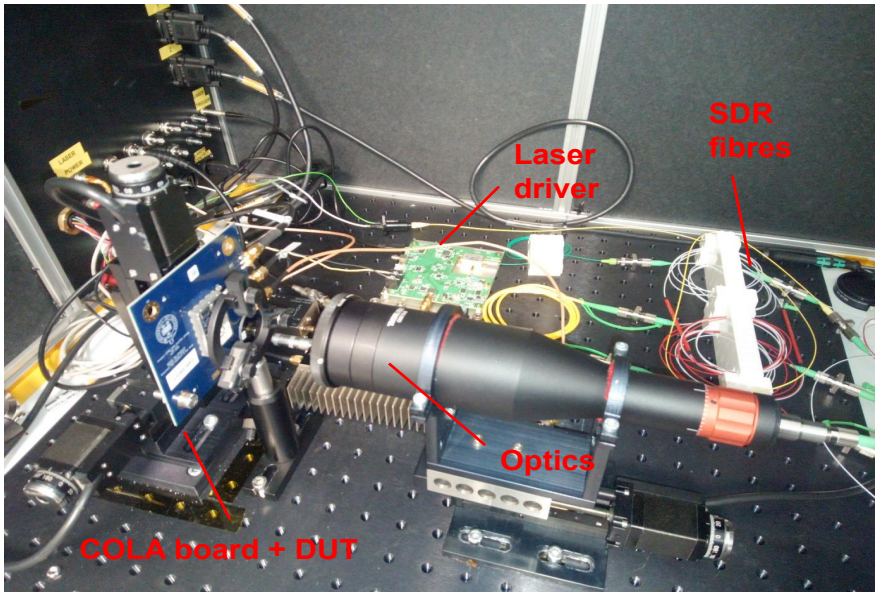


Fig. 5.8 *OPMD laser test setup for gain and timing measurement.*

One beam pulse travels along a fibre cable (3), whilst a fixed delay of approximately 5 ns, i.e. comparable to the predicted duration of the transient from simulation, as shown in Figure 4.35, is applied to the other beam via a long fibre cable (4). The two signals are then recombined (5) and collimated onto the device under test (DUT) via a beam telescope Thorlabs SM2D25D. The additional beam combiner (6) allows the introduction of a laser beam of visible wavelength into the optical path, to facilitate the process of alignment onto the DUT. The output pulse current from the DUT is then amplified and transformed into voltage by the COLA TIA and fed into the DAQ, an oscilloscope Keysight Infinium model UXR0134A operating at 128 GS/sec. The data obtained from the scope are then processed offline for analysis.

The timing resolution of the DUT can be obtained by measuring the standard deviation of the delay between the two pulses, the signal and its delayed copy:

$$\begin{aligned} \sigma_t^2 &= \sigma_1^2 + \sigma_2^2 + \sigma_{RO}^2 \\ &\cong \sigma_1^2 + \sigma_2^2 \\ &= 2\sigma_1^2 = 2\sigma_{DUT}^2 \end{aligned} \tag{5.1}$$

where the σ_1 and σ_2 are the jitters from the signal and the delayed copy respectively while σ_{RO} is the additional time uncertainty introduced by the readout (RO) chain, which includes the fast amplifier and the time to digital conversion (TDC) of the DAQ. The latter, which includes the fast amplifier, is negligible. Therefore:

$$\sigma_{DUT} = \frac{\sigma_t}{\sqrt{2}} \tag{5.2}$$

To compensate for the time walk, a constant fraction discrimination technique (CFD, [106]) was implemented in the analysis of the data. A signal snapshot that illustrates the process of timing measurement using the SDR method is shown in Figure 5.9. Subsequent measurements showed that the residual jitter of the QD laser output trigger was negligible ($<$1 ps) so the timing tests were then performed by measuring the standard deviation of the delay between the rising edge of the trigger and the DUT output pulse. The same laser setup was used to measure the device gain, i.e. the ratio of charge collected by the LGAD w.r.t. the PIN diode under the same biasing and temperature conditions. The total charge was obtained by integrating the current pulses.

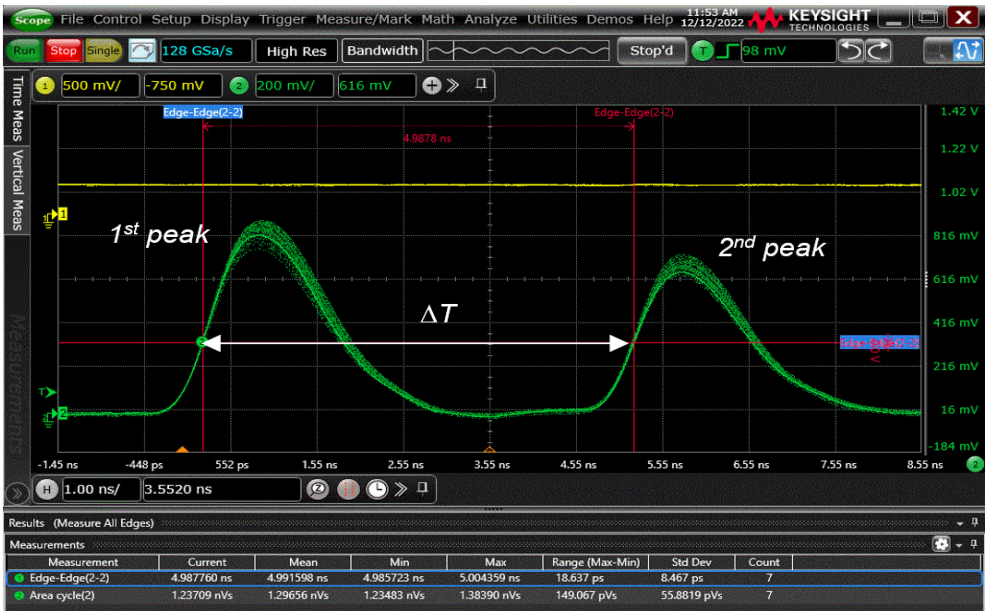


Fig. 5.9 *Example of timing measurement using the SDR method with a type 3 LGAD from WF2 (V_{bias} = 200 V, timescale 1 ns. CFD of 50%, T = 20 °C).*

A similar laser system was set up at RAL, Figure 5.10 and uses an oscilloscope Tektronix DPO73304SX as a DAQ. Such a system allows cooling of the DUTs to perform measurements on irradiated devices.

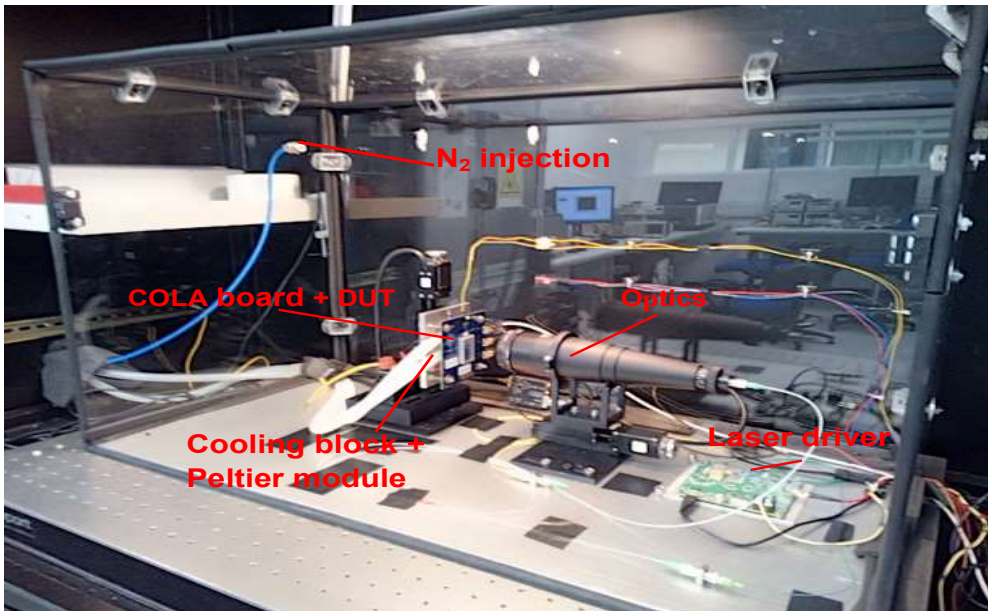


Fig. 5.10 *RAL laser test setup for gain and timing measurement. The sealed acrylic box can be filled with nitrogen, to prevent condensation when the DUT is cooled using a Peltier element to a temperature as low as T = -20 °C.*

5.2.3 Radioactive source test setup

While a laser injection setup might allow easy control of the amount of injected charge and where to deposit it, a charged particle setup has the advantage of providing particles similar to those at the LHC. Moreover, the Landau fluctuations in energy deposited across the detector while traversed by a minimum ionising particle (MIP), an effect which does not exist with laser charge injection, are an important factor in determining the achievable time resolution of LGADs [107]. A ^{90}Sr source provides beta decays with an energy of 0.55 MeV into ^{90}Y which then further beta decays to ^{90}Zr with an energy of 2.28 MeV and an average beta energy of 0.9336 MeV. Electrons with energy around or above $\approx$ 1 MeV are essentially MIPs and can be used for charge collection tests.

A picture of the ^{90}Sr test setup for timing measurements performed at OPMD is shown in Figure 5.11.

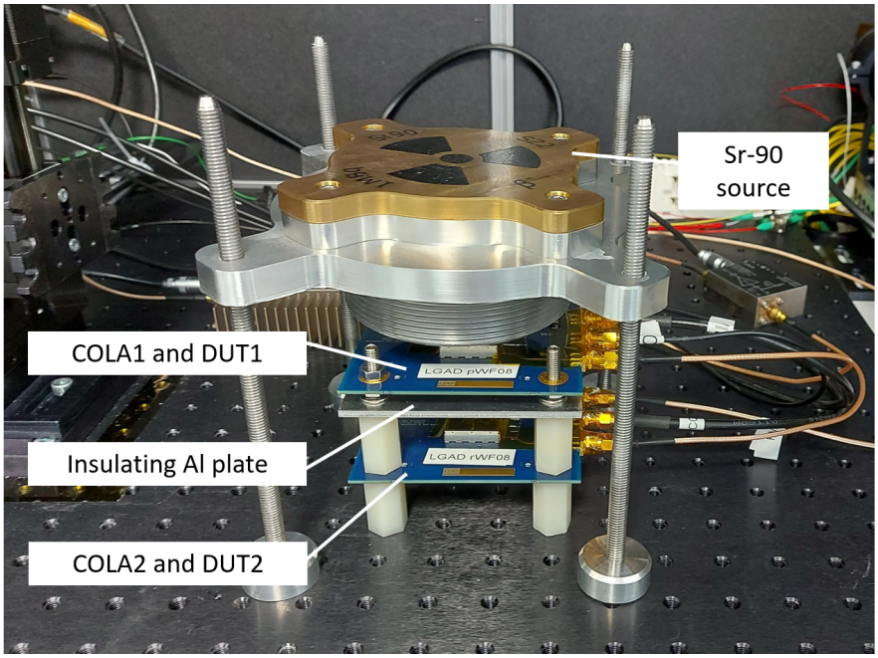


Fig. 5.11 *A picture of ^{90}Sr test setup for timing measurements.Two COLA boards are placed in line with a 1.40 MBq ^{90}Sr source such that two LGADs can detect the same electron.*

Two COLA boards, each equipped with an LGAD, are placed on top of each other and aligned. The top COLA board (COLA1) has a small hole drilled in the PCB, with an LGAD (LGAD1) mounted and bonded, to allow beta particles through and reach the second COLA board (COLA2), where the bottom LGAD (LGAD2) is placed. The output pulse signals from the two COLAs are fed into two external second-stage amplifiers. The outputs are then connected to the DAQ which records the pulses once an electron strikes one of the

sensors. The triggering is performed by COLA2, to reduce unnecessary data collection, since the LGAD1 sees more events than the LGAD2 because some of the 0.55 MeV electrons interacting with the COLA1 are absorbed before reaching the COLA2.

Similarly to the laser charge injection test, the timing resolution is determined from the standard deviation of the timing difference between the coincidence pulses. Due to the Landau fluctuations of the response, this is expected to be worse than in the case of jitter tests using laser charge injection. Since a coincidence can only occur when the paths of the electron intersect both LGADs, an offline analysis is performed to select events of interest. A histogram of the time difference is then obtained and fitted with a Gaussian function. The extracted standard deviation provides the time resolution:

$$\sigma_{meas}^2 = \sigma_{DUT}^2 + \sigma_{ref}^2 \tag{5.3}$$

where σ_{meas} is the standard deviation from the fitted Gaussian, σ_{ref} is the time resolution of the reference LGAD1 and σ_{DUT} is the timing resolution of the LGAD2 DUT. Like in equation (5.1) σ_{RO}, the uncertainty contribution coming from the readout circuitry, is found to be small and thus neglected in equation (5.3). If the two LGADs are of the same type and specifications, their timing resolution can be assumed to be the same and hence equation (5.3) simplifies to:

$$\sigma_{DUT} = \frac{\sigma_{meas}}{\sqrt{2}} \tag{5.4}$$

The CFD technique is applied to the data to compensate for the time walk effect. An example of coincidence pulses obtained using the setup is shown in Figure 5.12.

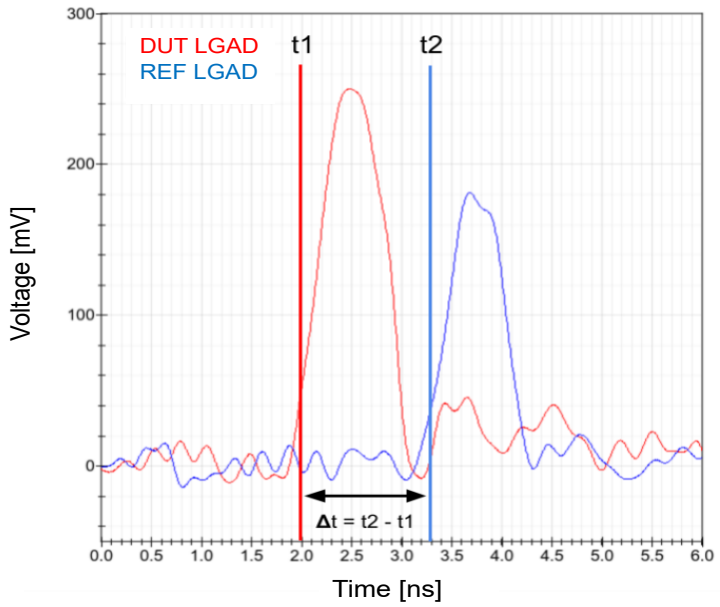


Fig. 5.12 *Example of a signal response from two LGADs detecting the same electron. A time can be assigned to each pulse at a constant fraction of the peak amplitude and the time difference between pulses calculated.*

5.3 Readout design and description

For all timing and gain measurements presented here, the readout consisted of a two-stage amplifier and an oscilloscope used as DAQ. The first amplifier stage, the Compact OPMD LGAD Amplifier (COLA), designed at OPMD, is a slightly modified version of the Santa Cruz Timing Board (SCTB, [108]). It contains a transimpedance amplifier (TIA) based on a single SiGe transistor which provides a gain of 50 dB and bandwidth over 1.2 GHz, with a layout board optimized to reduce ringing. The schematic diagram is shown in Figure 5.13 and an actual COLA board in Figure 5.14. The second stage consists of a DUPVA-60 variable gain amplifier (10 - 60 dB), a commercial RF amplifier from FEMTO [109], shown in Figure 5.15.

The characteristics of the amplifiers, for both COLA and DUPVA-60, were measured using an Infinium Oscilloscope UXR0134A and a signal generator Tabor LS1912B. The bandwidth was measured by applying a small signal amplitude of -30 dBm to the input, sweeping its frequency between 500 MHz to 1.5 GHz and measuring the corresponding output. The slew rate was measured by applying a 1 GHz sinewave signal to the input and increasing the input amplitude until the maximum slope at the output was observed. The jitter was determined by applying a -30 dBm sinewave input to the amplifier and measuring

variation in cycle time at the output at three different frequencies (500, 750 and 1000 MHz). The average of the three measurements provided an estimate of the jitter over the frequency range of interest. The broadband noise was determined from the rms value of the output when no input signal was applied. In the case of the COLA board, the input was connected to a type 3 run 1 LGAD from WF2 biased to full depletion, which provides an input load of C_{in} = 2.5 pF. An example of COLA rms output noise is shown in Figure 5.16. The measured characteristics of COLA TIA, including bandwidth, slew rate (i.e. the maximum rate of change of the voltage output), total RMS noise over a 1.5 GHz bandwidth and jitter are reported in Table 5.1. The measured characteristics of the DUPVA-60 are reported in Table 5.2.

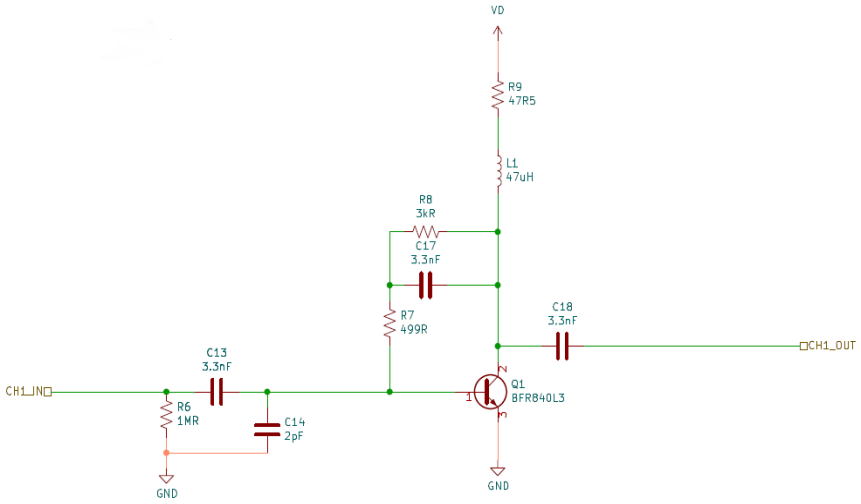


Fig. 5.13 *Schematic diagram of the Compact OPMD LGAD Amplifier (COLA).*

Table 5.1 *COLA specifications - T = 20 °C.*

Specifications	Values
Supply Voltage [V]	2.25
Gain [dB]	50
Bandwidth [GHz]	1.2
Slew Rate [GV·s^{-1}]	0.92
RMS Noise [μV]	155
Jitter [ps]	1.6

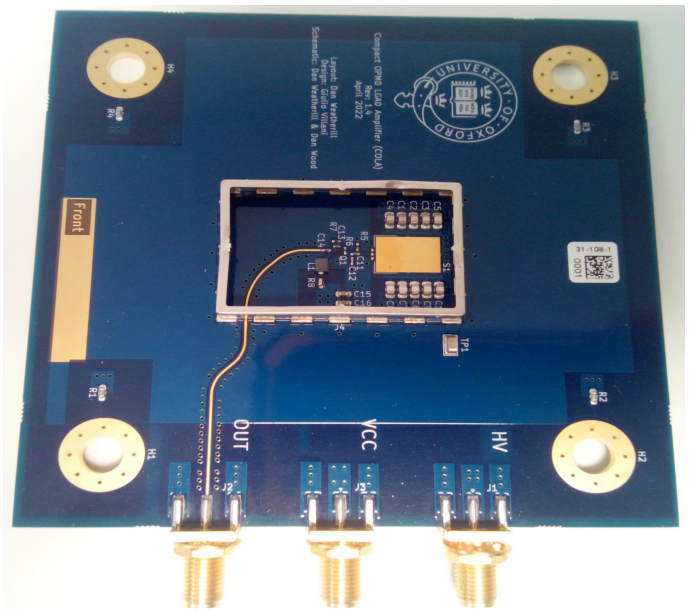


Fig. 5.14 *A picture of a COLA board. The size of the board is 10 × 10 cm^2. The central gold pad provides the HV biasing to the sensors and can accommodate devices of up to 5 × 10 mm^2.*

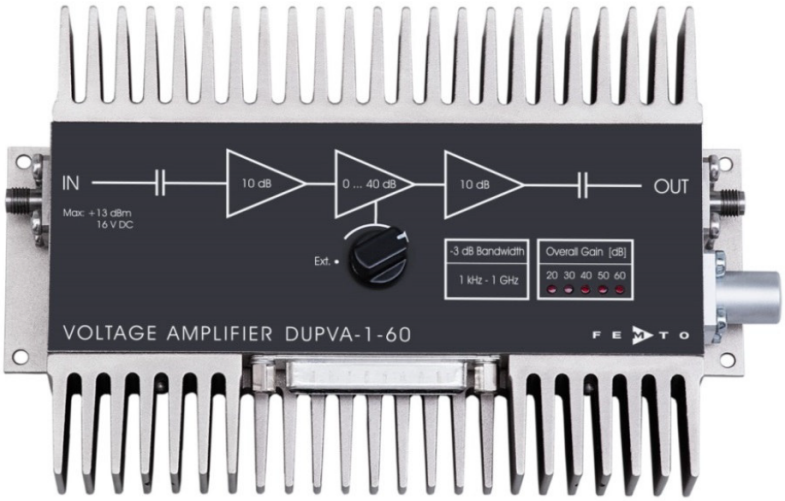


Fig. 5.15 *FEMTO DUPVA variable amplifier used as the second stage for LGAD tests.*

During the tests the gain of the second-stage amplifier was set to 30 dB, thus providing an overall gain of the amplifier chain of 80 dB over a bandwidth of approximately 0.8 GHz. The measured jitter of the amplifiers used, reported in Tables 5.1 and 5.2, are relatively small compared to the measured time uncertainties of the LGAD devices described in this thesis

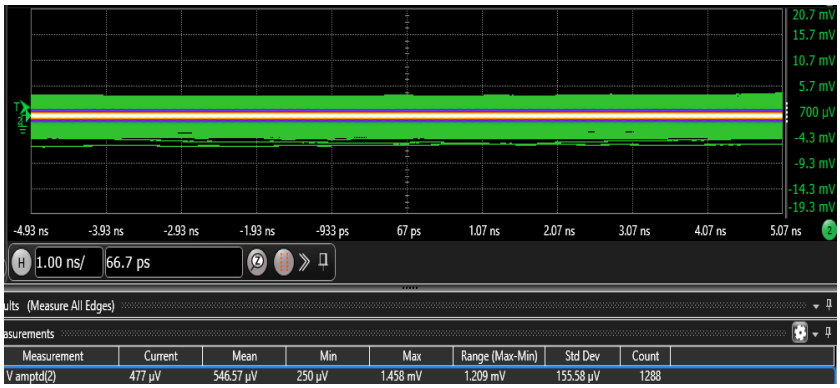


Fig. 5.16 *An oscilloscope snapshot of COLA rms output noise, with input load of 2.5 pF (type 3 LGAD fully depleted). Timescale 1 ns, 5 mV/division. A brick wall filtering of 1.5 GHz was applied to the signal.*

Table 5.2 *FEMTO DUPVA-60 specifications - Gain = 30 dB, T = 20 °C.*

Specifications	Values
Supply Voltage [V]	±15
Gain [dB]	30
Bandwidth [GHz]	1.2
Slew Rate [GV·s^{-1}]	2.67
RMS Noise [μV]	977
Jitter [ps]	3

and therefore could be neglected in the final analysis. Their measurements were nonetheless important to justify such approximation.

5.4 Test results: IV and CV characteristics

Using the test setup described in section 5.2 the IV and CV plots of the fabricated devices were obtained, both on wafers and diced devices. The tests carried out on devices of Run 1 were restricted to the type 3 devices (1 mm cathode size) as their size closely resembles the size of the cells planned for the HGTD (1.3 mm cathode size). For these tests, the device's backside was grounded to the chuck of the probe station and a voltage sweep was applied to the cathode via a probe needle. The guard ring was kept floating during the measurements. To investigate the uniformity of characteristics across the wafer, the devices to test were taken from 5 different field locations shown in Figure 5.17. All measurements were taken at temperature T = 21.0 ± 0.5 °C, which contributes to an error of around 3% in the leakage current measurements.

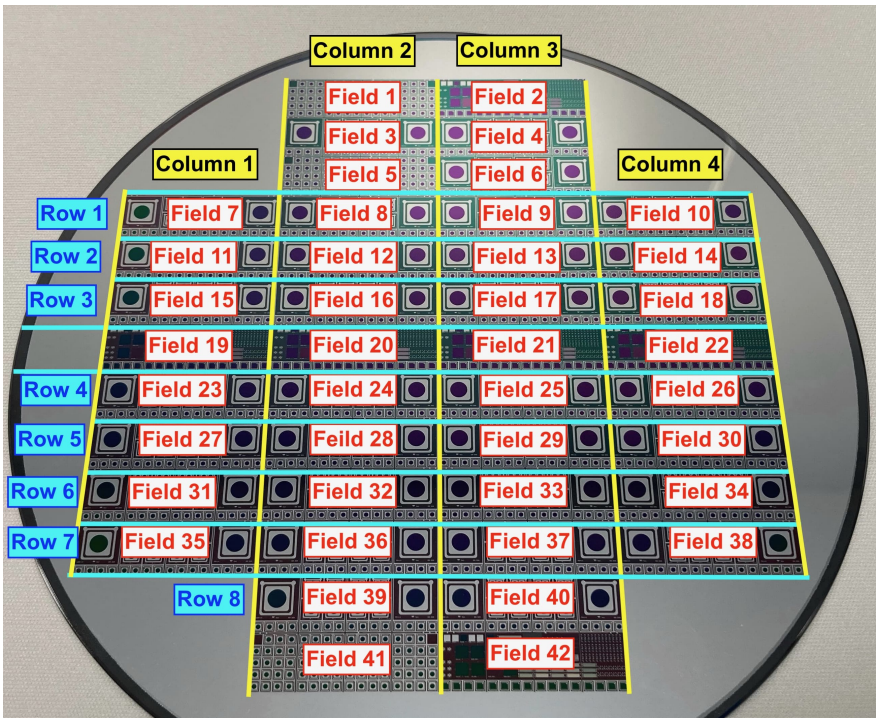


Fig. 5.17 *A picture of a wafer map showing the field locations. To investigate the uniformity of characteristics, devices from different locations on the wafer, i.e. fields 7, 10, 38, 35 and central, 25, were tested.*

5.4.1 IV plots

The IV plots of five type 3 LGAD and PIN devices are shown for the three different wafers WF2, WF8 and WF9 of Run 1 in Figure 5.18. The PIN diodes used were taken from WF2. A voltage sweep of 700 V in 1 V steps, with a hold time of 0.75 s and integration time of 1 s, was used for the test with current compliance set to 1 μA. The breakdown is here defined as the voltage V_{bd} at which the current exceeds 200 nA. No significant breakdown voltage differences were measured among the LGADs of various guard rings to p-stop distance. From the IV curves in Figure 5.18, it was found that the PIN diodes, not featuring the gain layer, show a typical V_{bd} over 500 V. All the three LGAD flavours tested from Run 1 showed a typical V_{bd} lower than that of the PIN diode, indicating that the avalanche is initiated in the GL region and not from a high lateral electric field at the cathode's edge.

Initial tests were done on one of the LGAD arrays from Run 2, WF19 from table 4.4. The breakdown voltage map is shown in Figure 5.19.

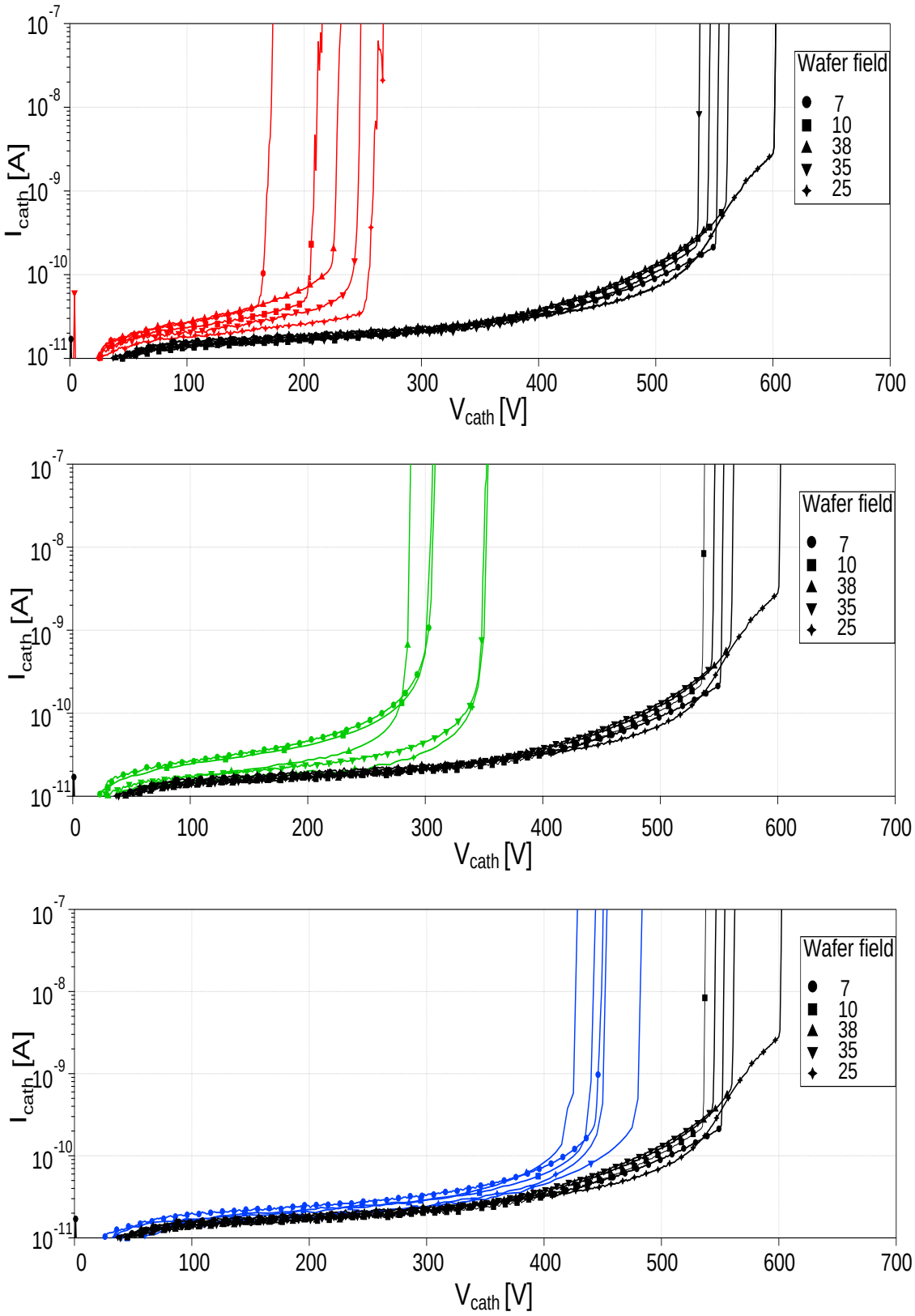


Fig. 5.18 *The IV plots of WF2 (top, red), WF8 (middle, green) and WF9 (bottom, blue) LGAD devices and PIN diodes (all, black).*

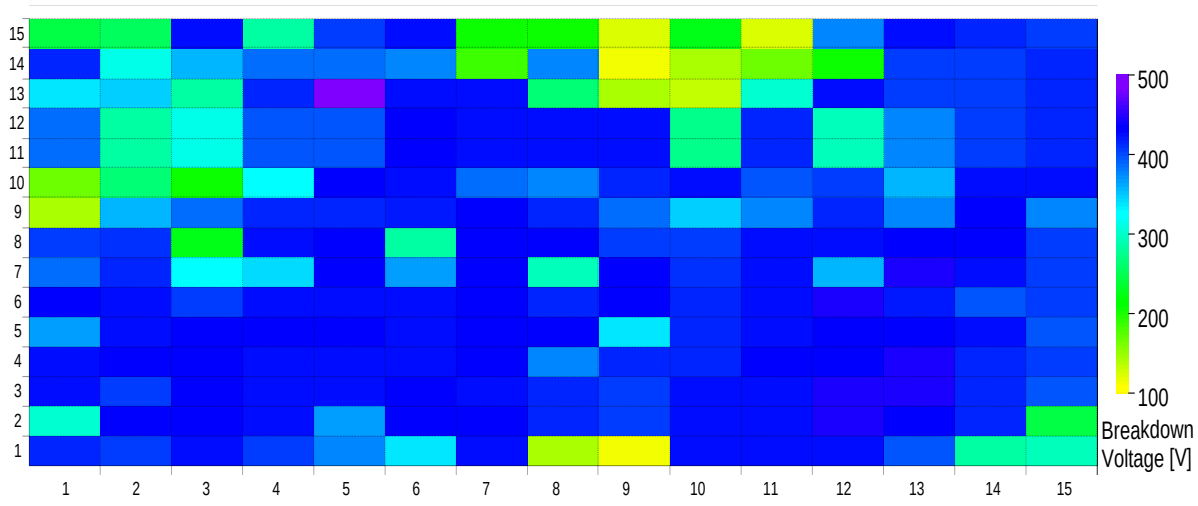


Fig. 5.19 *An example of IV map of WF19 15 × 15 LGAD array from Run 2.*

5.4.2 CV plots

The CV and $1/C^2$ measurements were conducted in a bias voltage range between 0 and 50 V, i.e. a bias voltage high enough to guarantee full depletion, in 200 mV steps at a frequency of 100 kHz with an amplitude of 30 mV rms. The results of these measurements for five Run 1 LGADs of type 3 are shown in Figure 5.20 and for PIN in Figure 5.21. From the CV measurements, the gain layer depletion voltage V_{gl} and full depletion voltage V_{fd} of the LGAD devices were determined from the point of interception of the fitting lines before and after the kink, as shown in the top plot of Figure 5.20. As expected from simulations, devices from WF2, the top plot of Figure 5.18, show the lowest V_{bd} voltage and highest V_{gl}, due to the deeper gain layer implant obtained by the high energy implantation used for their fabrication [110], shown in Table 4.2. At the other extreme, devices from WF9, the bottom plot of Figure 5.18, with a shallower gain layer implant, show a higher V_{bd} and lower V_{gl}.

The IV and CV test results for the three LGAD types and PIN diodes of Run 1 are summarized in Table 5.3, where it is shown that the average V_{bd} of the LGADs increases by about 100 V between WF2 and WF3, and by another 100 V between WF3 and WF9. The spread in breakdown is substantially the same for all flavours of the LGADs therefore, expressed as a percentage of the V_{bd}, it is smaller for the LGAD of higher V_{bd}.

In Table 5.4 some of the required electrical characteristics of the HGTD ATLAS sensor [10] are reported for comparison with the fabricated LGADs.

Table 5.3 *IV and CV test results for Run 1 type 3 LGAD and PIN.*

Flavour	<V_{bd}>[V]	<V_{gl}>[V]	<V_{fd}>[V]	<I_{lk}>@95% V_{bd}[nA]	<I_{lk}>[nA·cm^{-2}]
WF2	228.4^{+36}_{-48}	$27.1^{+1.9}_{-0.7}$	$29.0^{+0.4}_{-1.3}$	$0.1^{+0.01}_{-0.5}$	10
WF8	321.8^{+31}_{-34}	$26.6^{+1.3}_{-1.5}$	$28.7^{+0.3}_{-1.0}$	$0.18^{+0.1}_{-0.05}$	18
WF9	452.4^{+30}_{-22}	$24.7^{+2.4}_{-1.0}$	$26.4^{+1.9}_{-1.2}$	$0.11^{+0.2}_{-0.01}$	11
PIN	560.0^{+41}_{-22}	N.A.	$30.1^{+0.8}_{-0.6}$	$0.26^{+2.0}_{-0.01}$	26

Table 5.4 *Required electrics properties of the single pad of the HGTD sensors at room temperature.*

Pad leakage current (V_{bd} condition) Device total leakage current I_{lk}	< 500 nA $< 20\ \mu$A·cm^{-2} at bias voltage $< V_{bd}$
Breakdown voltage V_{bd}	$V_{bd} > V_{fd} + D^* \times 2.0$ V·μm^{-1}
Gain layer depletion V_{gl}	24 V $< V_{gl} <$ 55 V
Full depletion V_{fd}	$<$ 70 V
Inter-pad isolation Pad capacitance	p-stop $<$ 4.5 pF (to the backplane and inter-pad)

*D is the active thickness of the sensor, $50 \pm 5\ \mu$m.

The average leakage current I_{lk} measured at 95% of V_{bd} for all LGAD samples corresponds to a current density of around 13 nA·cm^{-2}, well within the specifications of Table 5.4. For all LGADs tested the measured breakdown voltage V_{bd} at room temperature also complies with the specifications of Table 5.4. The measured GL average depletion voltage V_{gl} is between 24.7 V and 27.1 V, with full depletion voltage V_{fd} achieved at around 29 V max, again in full compliance with the required values of Table 5.4. The average capacitance of the LGAD devices at full depletion is measured to be approximately 2.5 pF. Since the size of the type 3 cell (1.0×1.0 mm^2) is smaller than the size of the HGTD sensors (1.3×1.3 mm^2) an increase in capacitance is expected for the nominal size LGAD cell.

An example of a CV plot of five cells of the WF4 full LGAD sensor array is shown in Figure 5.22. The average capacitance of the cell is approximately 4.7 pF, measured with the bias voltage applied only to the cell under test and keeping the neighbouring cells floating, whereas the correct biasing condition would require all the cathode of the cells kept at the same potential. As discussed in the final chapter, more tests are already planned for this LGAD array. The drop in capacitance at around 30 V also indicates the depletion of the GL, therefore suggesting a correct operation of the array.

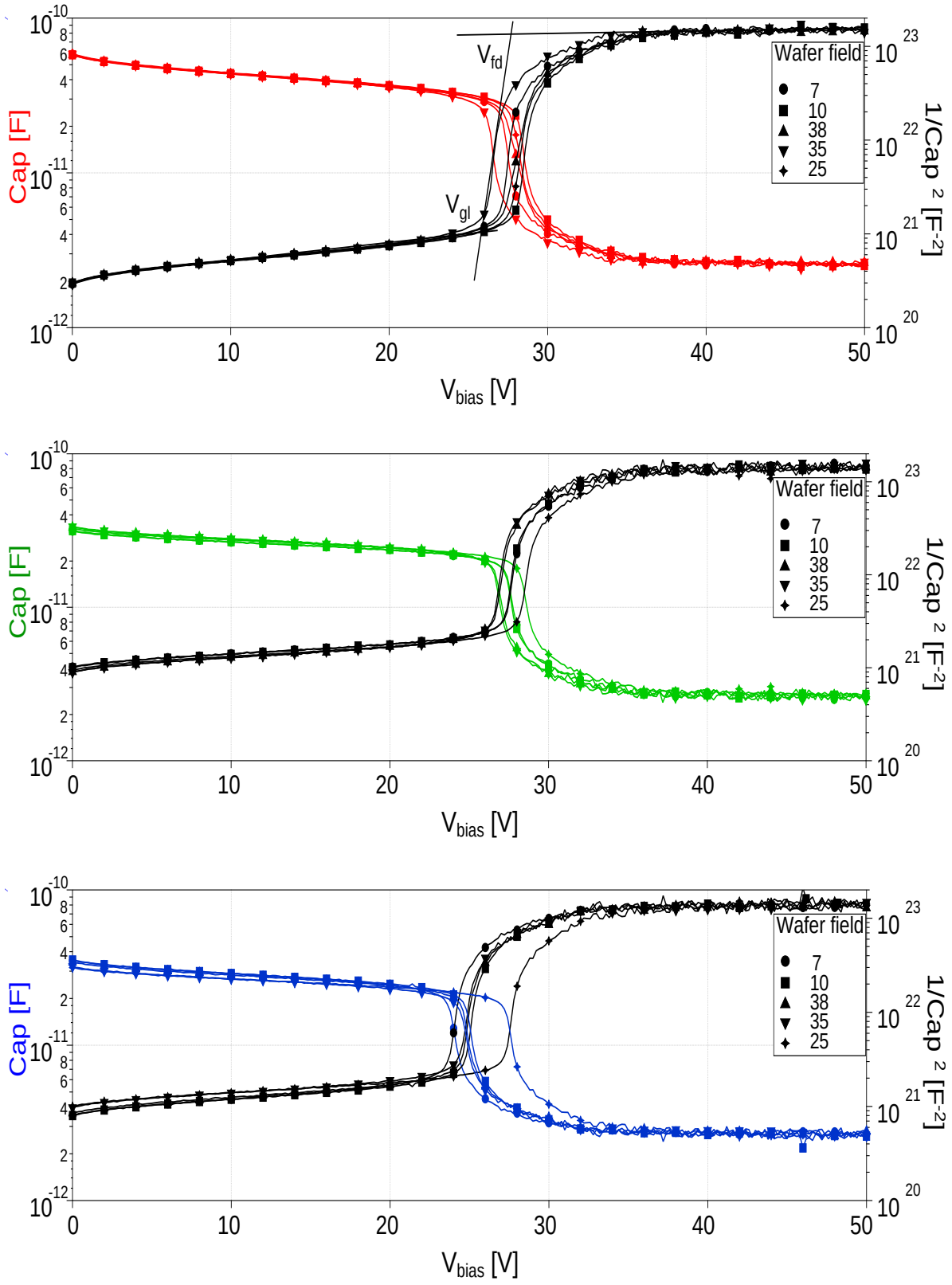


Fig. 5.20 *The CV plots of WF2 (top, red), WF8 (middle, green) and WF9 (bottom, blue) LGAD devices. In each plot, the corresponding $1/C^2$ is shown in black.*

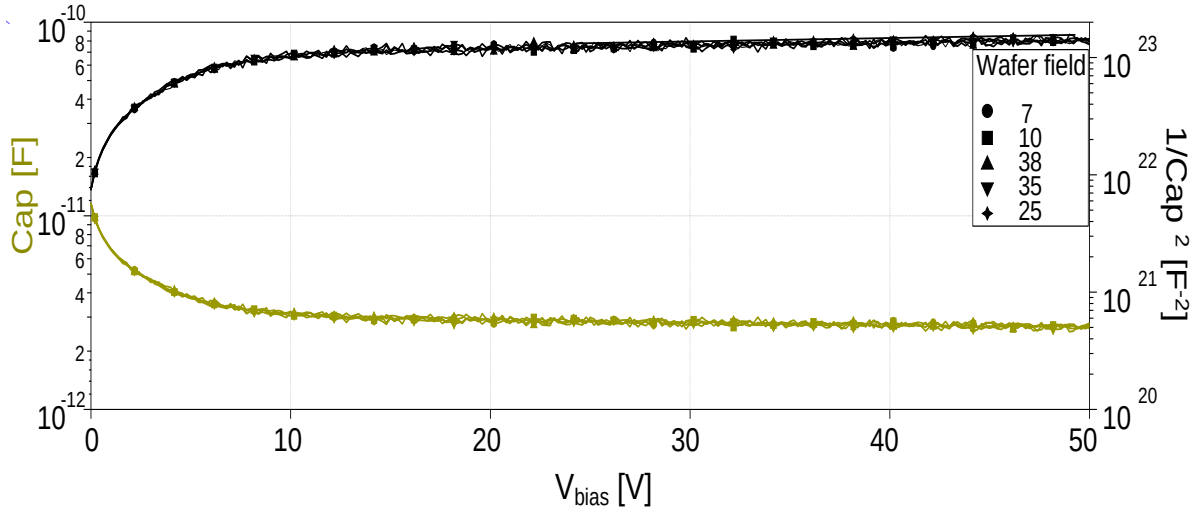


Fig. 5.21 *The CV and 1/C^2 plot of WF2 PIN diodes. The guard ring was kept floating during the test.*

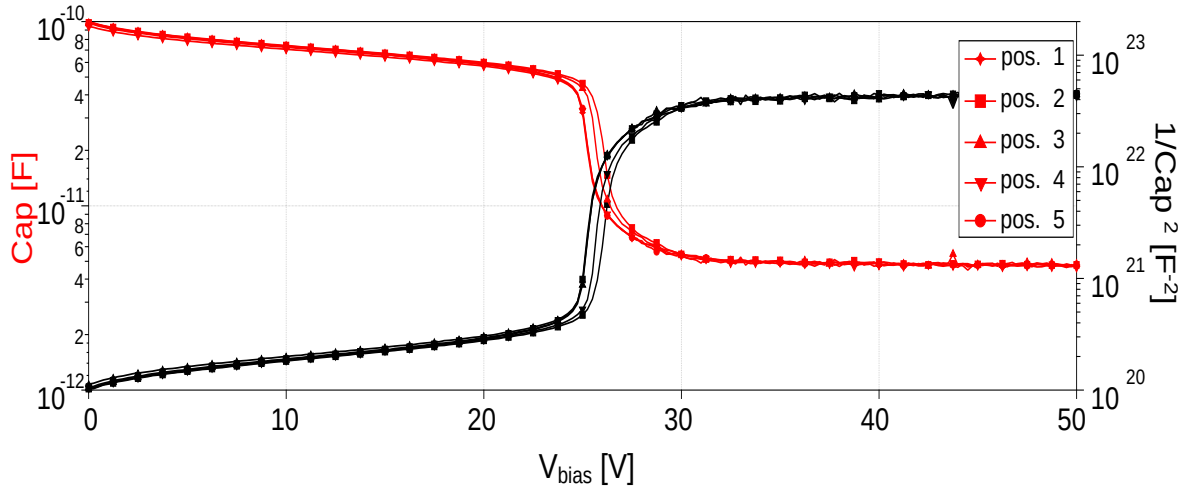


Fig. 5.22 *A plot of CV and 1/C^2 of LGAD cells from a Run 2 array. The tested cells are shown in Figure 5.5.*

5.5 Test results: Time Resolution and Gain

For gain and timing tests the devices were placed onto the COLA board described in section 5.3 and the test setups described in section 5.2.2 and 5.2.3 were used.

5.5.1 Gain

For the gain measurements using the laser charge injection described in 5.2.2, the laser setup was calibrated to provide approximately 1 fC of collected charge in the PIN diode, used as a reference device, i.e. similar to the average generated charge by a MIP when traversing a device of the same nominal 50 μm thickness and to match the value of injected charge in the simulations. Table 5.5 shows the QD laser parameters used for the tests. Figure 5.23 shows an example of the output laser pulse measured using an optical probe.

Table 5.5 *Table of QD laser settings used for gain and timing test.*

Laser I_{bias}[mA]	Laser V_{th}[V]	Trigger Frequency [MHz]	Laser diode Temperature [°C]
9	0.830	0.1	12

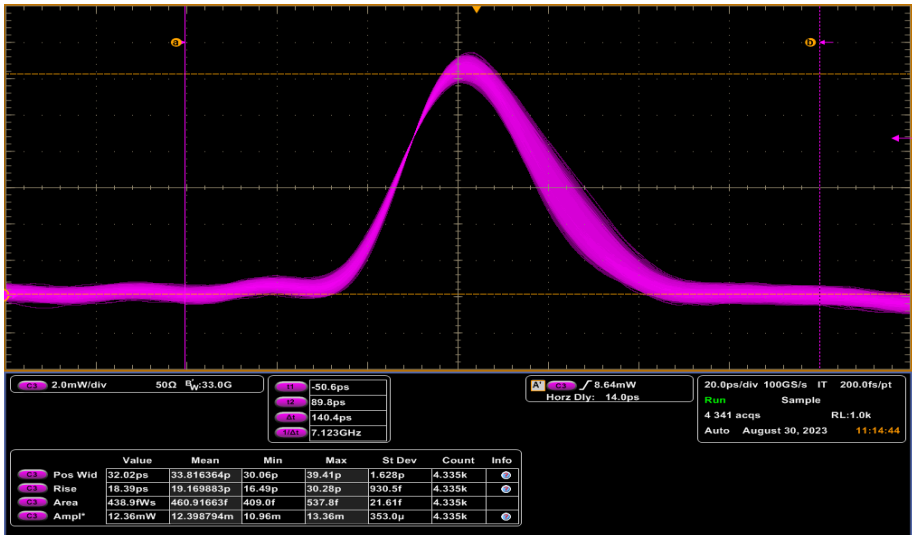


Fig. 5.23 *Example of a QD laser pulse signal measured with a Tektronix oscilloscope DPO73304SX set to 100 GS/s and equipped with a 33 GHz optical probe NXD33.*

The collected charge was measured by integrating the averaged current pulses provided by the PIN diode at the output of the readout chain and dividing it by the measured gain of

the two-stage amplifiers. The gain was then measured as the ratio of the LGAD over PIN collected charge for different bias voltages. An example is shown in Figure 5.24. The total collected charge from the PIN diode, once biased to full depletion, did not show significant differences with increasing bias voltage, as reported in Table 5.6 and compares well with the simulation results, reported in Table 4.8.

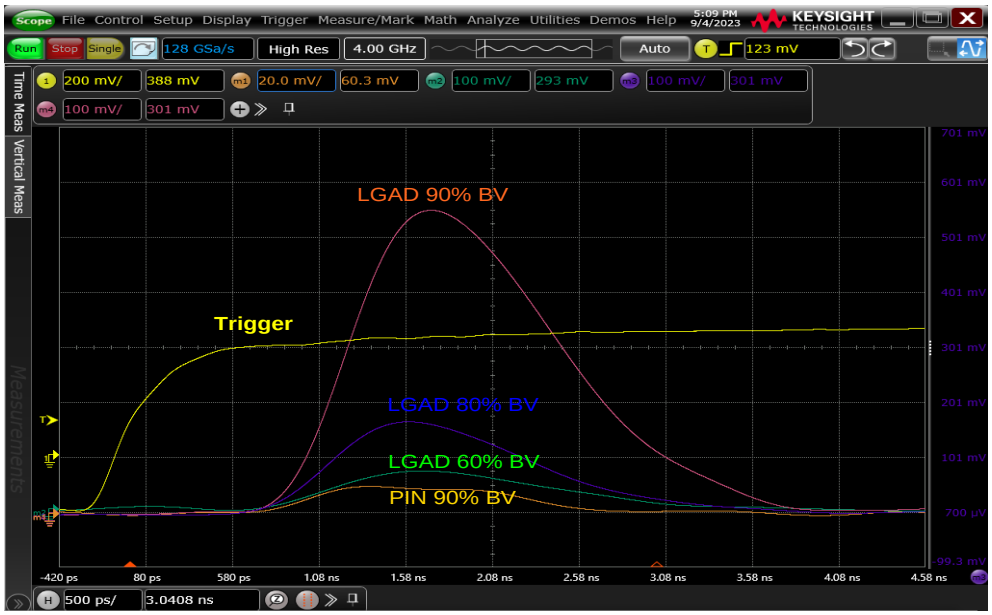


Fig. 5.24 *An example of PIN and LGAD WF2 pulse signal output for laser charge injection at different bias voltages. Scope settings: 2 ps time resolution, 1.5 GHz low pass filtering, 100 averages.*

Table 5.6 *PIN collected charge.*

V_{bias}	<V_{amp}>[mV]	<Area>[pV·s^{-1}]	<Q>[fC]
50	9.9	8.9	0.62 ± 0.01
100	9.9	8.9	0.64 ± 0.05
200	9.9	8.9	0.66 ± 0.08
300	10.1	9	0.68 ± 0.08
400	9.9	8.9	0.69 ± 0.09
500	10.2	9.1	0.71 ± 0.09

The gain and charge collection results for the LGAD devices versus V_{bias} are shown in Figures 5.25, 5.26 and 5.27. In Figure 5.28 the gain and collected charge of the LGAD devices are plotted versus the bias voltage normalized to the breakdown voltage V_{bd} of each device.

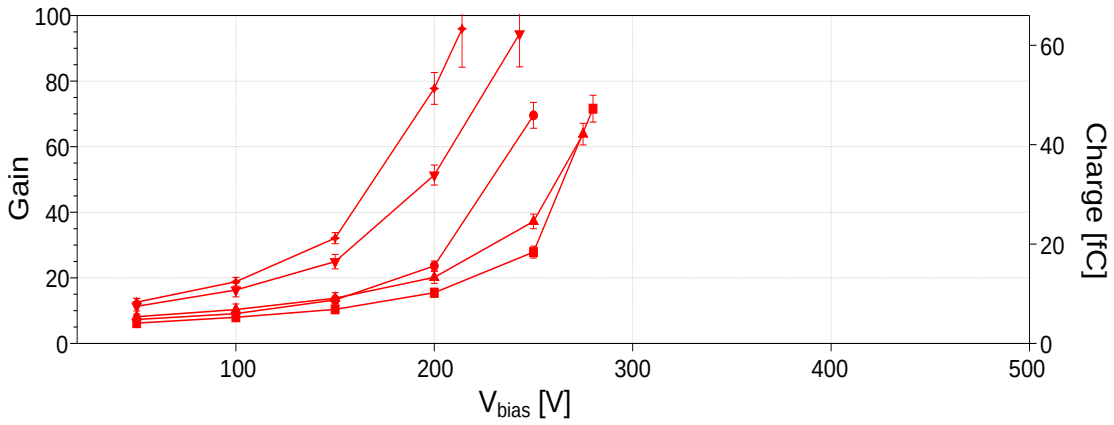


Fig. 5.25 *Gain and charge collected for five WF2 LGAD devices.*

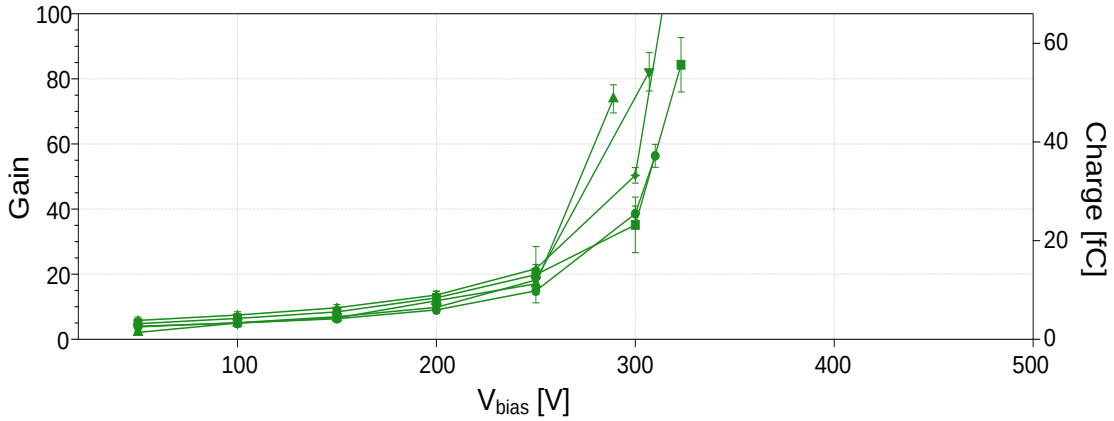


Fig. 5.26 *Gain and charge collected for five WF8 LGAD devices.*

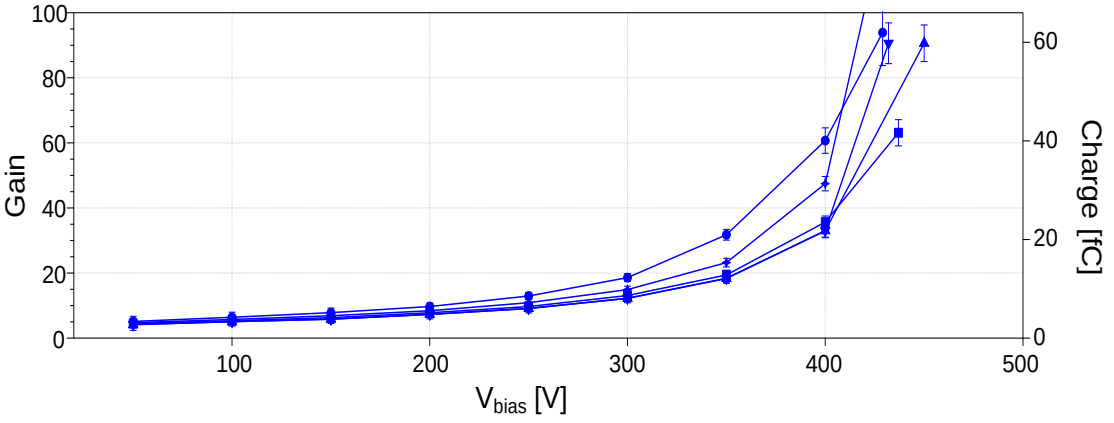


Fig. 5.27 *Gain and charge collected for five WF9 LGAD devices.*

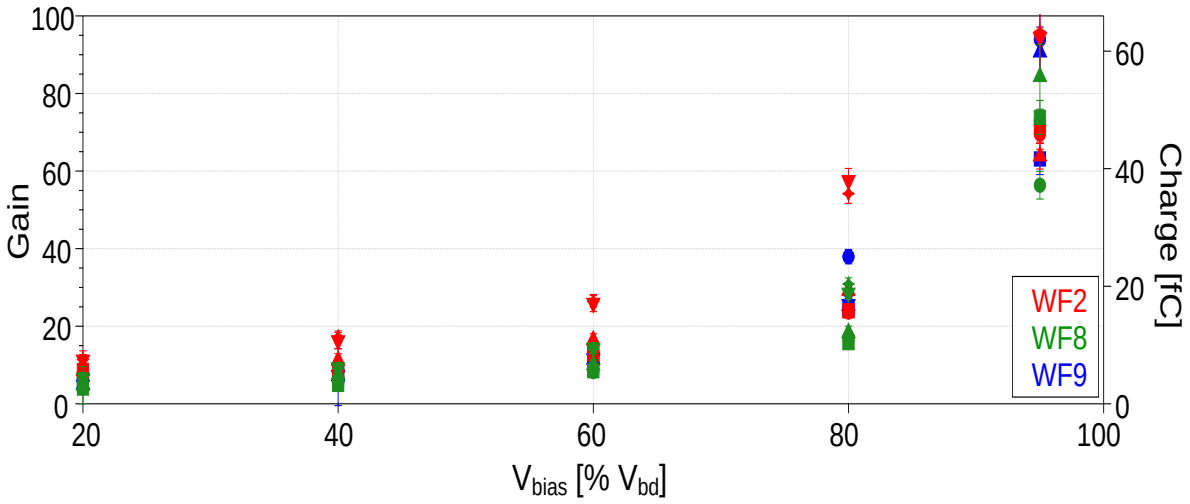


Fig. 5.28 *Gain and collected charge plots of WF2, WF8 and WF9 LGAD devices versus bias voltage as a percentage of breakdown voltage V_{bd}.*

Figure 5.25, 5.26 and 5.27 show the general increase in gain versus bias voltage for all the three LGAD types, due to the enhanced impact ionisation with an increase in bias. The gain plot from Figure 5.28 shows that, when normalised to a fraction of the breakdown voltages, the three flavours tested showed similar gain performances, with minor differences that could be attributed to the variability among individual devices. Comparing Figure 5.28 with Figure 4.42 shows that the TCAD simulations predict the LGAD gain reasonably well up to approximately 80% of the breakdown voltage V_{bd}, with lower values predicted when using the University of Bologna model. At higher values of the bias, the gain estimate is lower than the measurements. This would suggest that further investigation on the modelling of impact ionisation is needed and that accurate knowledge of the device doping might not be sufficient to obtain reliable predictions when operating the devices near breakdown[1]. Table 5.7 summarizes the normalized gain results for the infrared laser charge injection tests.

Table 5.7 *Laser charge injection average gain $\pm$ 1 σ for WF2, WF8 and WF9.*

LGAD flavor	V_{bd} 20%	V_{bd} 40%	V_{bd} 60%	V_{bd} 80%	V_{bd} 95%
WF2	9.44 $\pm$ 1.83	12.5 $\pm$ 4.01	18.97 $\pm$ 6.73	37.72 $\pm$ 16.76	79.06 $\pm$ 14.99
WF8	5.24 $\pm$ 1.27	7.38 $\pm$ 1.91	11.06 $\pm$ 2.99	22.13 $\pm$ 7.05	96.01 $\pm$ 35.73
WF9	5.42 $\pm$ 0.55	7.75 $\pm$ 0.95	12.52 $\pm$ 1.58	28.25 $\pm$ 5.59	104.66 $\pm$ 32.64

[1]There is also an increased risk of Single Event Burnout (SEB) from single beam particles when operating the devices near breakdown. To reduce such risk, a bias voltage that guarantees a field not exceeding 11 V·μm^{-1} should be applied [111]

5.5.2 Timing

Figure 5.29 shows the measured jitters and slew rate using the laser charge injection method and a 25% CFD value.

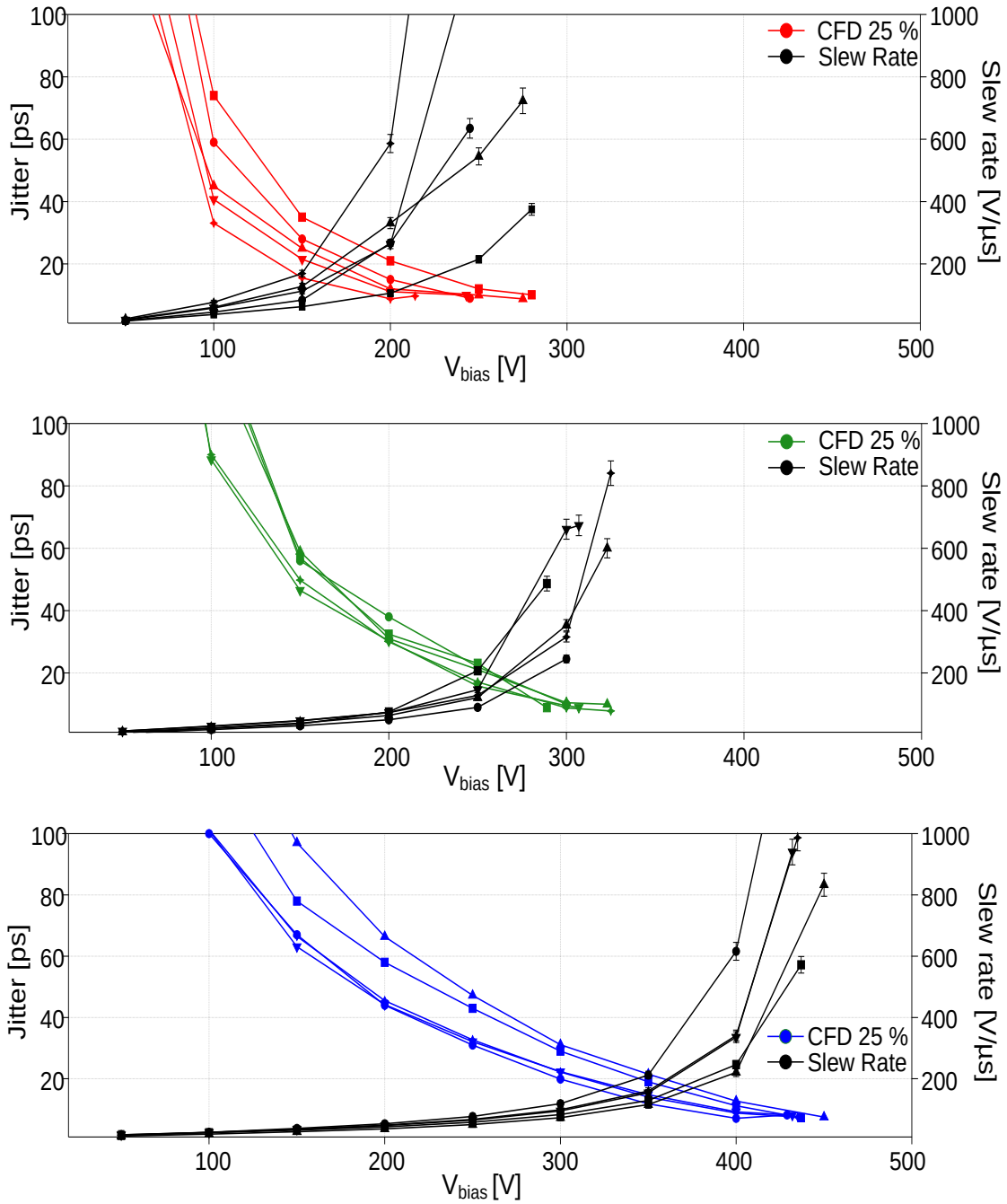


Fig. 5.29 *Jitter for LGAD devices of WF2 (top, red), WF8 (middle, green) and WF9 (bottom, blue). The slew rate for the same devices is shown in black.*

The measured jitters decrease as the bias voltage is increased, due to the gain increase. In Figure 5.30, the jitters for the three LGAD types are plotted versus the percentage of breakdown voltage, using a CFD value of 25%.

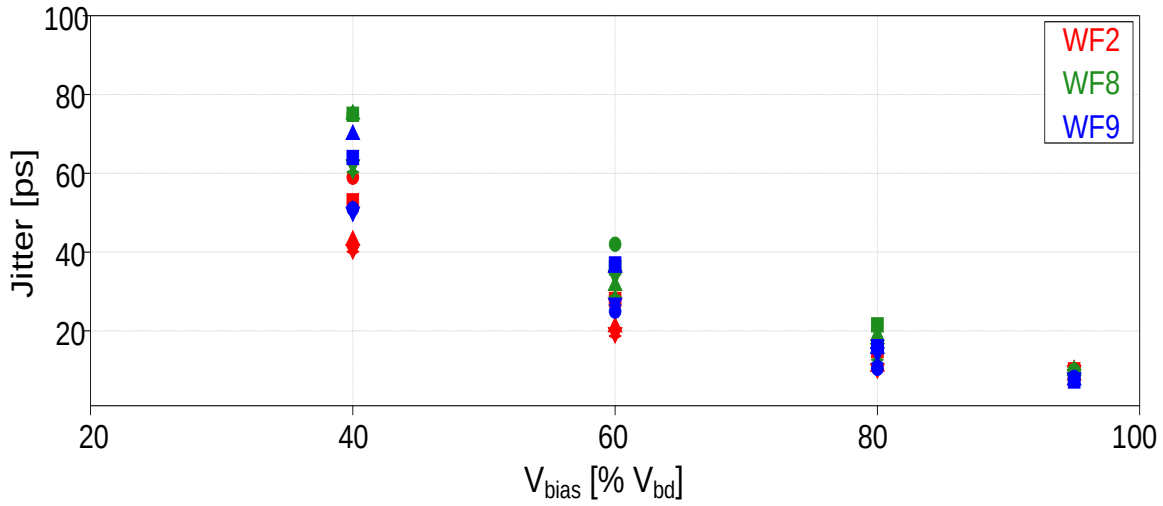


Fig. 5.30 *Jitter of WF2, WF8 and WF9 LGAD devices versus bias voltage as a percentage of breakdown voltage V_{bd}.*

In tables 5.8, 5.9 and 5.10 the jitter values are reported versus percentage of breakdown voltage and for different CFD values. The minimum jitter is reached at a CFD of around 25%, which corresponds to the point of maximum steepness in the curve signal.

Table 5.8 *Laser charge injection average jitter ± 1 σ [ps] for LGAD devices of WF2.*

CFD value	V_{bd} 20%	V_{bd} 40%	V_{bd} 60%	V_{bd} 80%	V_{bd} 95%
25%	173.2 ± 34.1	45.56 ± 6.02	22.06 ± 3.7	11.56 ± 1.95	9.72 ± 0.8
50%	157 ± 9.34	47.36 ± 6.35	22.82 ± 4.08	11.88 ± 2.36	10.46 ± 0.99
90%	172.4 ± 22.07	66.98 ± 7.86	53.6 ± 6.61	50.36 ± 4.89	49.82 ± 2.76

Table 5.9 *Laser charge injection average jitter ± 1 σ [ps] for LGAD devices of WF8.*

CFD value	V_{bd} 20%	V_{bd} 40%	V_{bd} 60%	V_{bd} 80%	V_{bd} 95%
25%	192.50 ± 31.51	69.5 ± 7.56	34.46 ± 5.03	17.11 ± 3.34	8.91 ± 0.77
50%	201.26 ± 33.13	76.56 ± 15.67	43.49 ± 13.48	20.66 ± 11.69	9.51 ± 0.86
90%	192.6 ± 27.89	79.5 ± 8.15	62.46 ± 9.75	50.07 ± 5.17	48.35 ± 2.51

Table 5.10 *Laser charge injection average jitter ± 1 σ [ps] for LGAD devices of WF9.*

CFD value	V_{bd} 20%	V_{bd} 40%	V_{bd} 60%	V_{bd} 80%	V_{bd} 95%
25%	131.4 ± 25.23	57.22 ± 9.08	30.36 ± 5.71	13.66 ± 2.4	7.72 ± 0.41
50%	128.72 ± 27.11	55.92 ± 7.97	31.46 ± 6.13	14.82 ± 2.82	8.61 ± 0.59
90%	144.26 ± 17.86	74.14 ± 10.32	57.23 ± 4.6	50.54 ± 3.3	48.44 ± 1.72

Initial timing results obtained using the radioactive source are shown in Figure 5.31. The timing results using the source setup have been so far obtained for one LGAD flavour only, WF2 and show time resolution of the order of 40 ps. This is a factor of 3 - 4 times worse than the results on jitter obtained from laser charge injection, a consequence of the Landau fluctuations during the process of ionisation from charged particles that do not exist in the case of photons. Further tests are planned and will be performed soon.

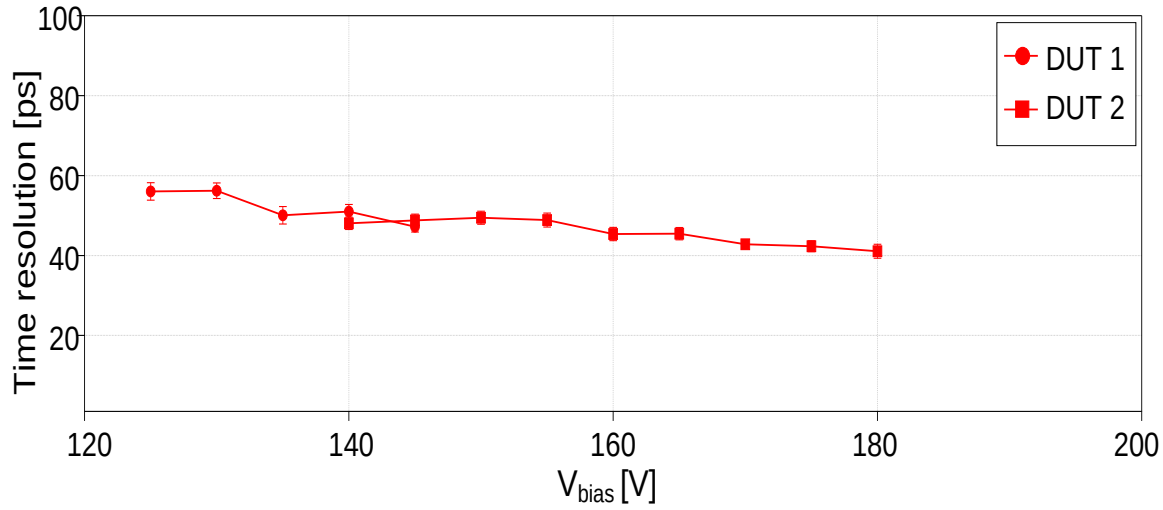


Fig. 5.31 *An example of time resolution of type 3 LGAD devices from WF2 .*

5.6 LGAD for dosimetry

Reduction of out-of-field doses is one of the challenges to improving the long-term survival of patients following radiotherapy [112]. Moreover, the radiation environment of space consists of a considerable amount of solar particle events (SPE), essentially low bursts of low LET particle [113]. In particle radiotherapies and the space radiation environment, low LET and high energy delta electrons (0.1 keV$\cdot\mu$m^{-1}) contribute significantly to the dose accumulated. Thus, particles in this LET region must be considered for radiobiological effects (RBE) in particle therapy and the average quality factor Q for the radiation protection of astronauts. Silicon-based microdosimeters are used for dose monitoring but they suffer from poor signal-to-noise ratio for low LET particles, which can contribute to a considerable fraction of the accumulated dose [114]. By virtue of their intrinsic gain, LGAD devices are a possible solution to improve the signal-to-noise ratio for low LET dosimetry applications. For this purpose, gain tests were performed on LGAD devices using the ion beam-induced charge (IBIC) facility at the Australian Nuclear Science and Technology Organisation (ANSTO) which uses the Australian National Tandem Research (ANTARES) accelerator to produce a scanning microbeam as small as 1 μm.

5.6.1 Gain results

To investigate the gain characteristics of these LGADs for different induced spatial charge carrier densities, the devices were irradiated using 1, 1.8, 3 and 8 MeV protons and with bias voltages ranging from 0 to 250 V. In all the tests the resolution of the scanning microbeam was 1 μm. The energy deposited in the detector was calibrated using 1 MeV and 1.8 MeV protons incident on a fully depleted 300 μm thick fully depleted Hamamatsu PIN photodiode.

Two-dimensional energy deposition spectra were obtained by considering the median energy collected in each spatial bin. Figure 5.32 shows an example of the spectra obtained for different proton energies and with a constant bias voltage of 180 V. The central region, where the gain layer of the device is located, shows different values of collected charge, hence gain, for different energies of the incident protons. The border region shows no gain, as it corresponds to the JTE region, used to improve the breakdown characteristics of the device but with no gain layer underneath. The central metalization aperture, which consists of 1 μm thick aluminium, visibly affects the response to 1 and 1.8 MeV protons, as they have a projected range of 15.8 μm and 39.2 μm respectively [115]. Hence, protons of this energy passing through the aluminium layer deposit less energy in the active region of silicon.

The gain of the devices was found from the peak amount of charge generated in the regions with and without gain with the devices operating at full depletion. The position

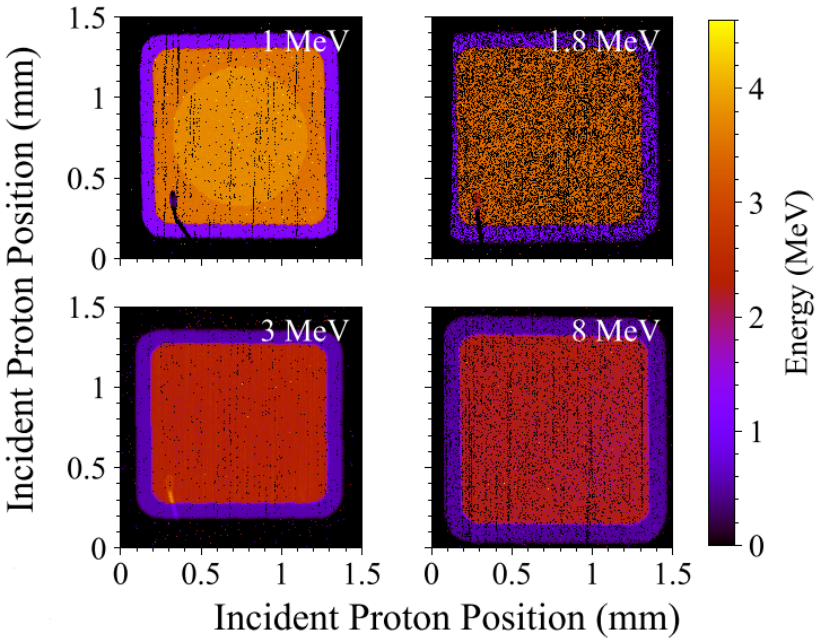


Fig. 5.32 *Median energy deposition maps for protons of varying energy incident on the type 3 LGAD from WF2 measured at CMRP. The bias voltage was 180 V.*

of the peaks in the energy spectra was determined by first finding the peaks in a smoothed spectrum by the implementation of a moving average. The peak properties were determined by fitting a Gaussian to data points within 2σ of the peaks. In Figure 5.33 the obtained gain for different protons energy is shown. For increasing bias above approximately 50 V the gain monotonically increased in the detector. The gain however was found to be lower for lower energy protons. This is an implication of the higher LET and increasing charge density of the track structure of these protons as they traverse the silicon, leading to gain suppression [116]. Moreover, the gain is not always monotonically increasing with bias voltage, with a dip between 30 V and 100 V bias for the 1 MeV and 1.8 MeV protons, an effect that could be due to two competing factors. In the first one, there is an initial reduced lateral spread of the induced plasma column as charge carriers which diffuse at the edge of the plasma will be more quickly accelerated towards the electrodes. This results in an increased rate of recombination at the core of the plasma column due to stronger electric field screening. In the second one, there is an increase in charge multiplication as a result of the increased electric field with increasing bias. A similar explanation was proposed in [117]. The second mechanism of increasing gain with increasing reverse voltage is the usual effect due to impact ionisation, upon which LGAD devices are based. The former of these mechanisms has been measured for 3 MeV particles incident on a Medipix detector, showing a decrease in the cluster size of induced charge for increasing bias above full depletion. This is consistent with

the approximate plasma time being proportional to the square of the linear plasma density and the square of the electric field [118].

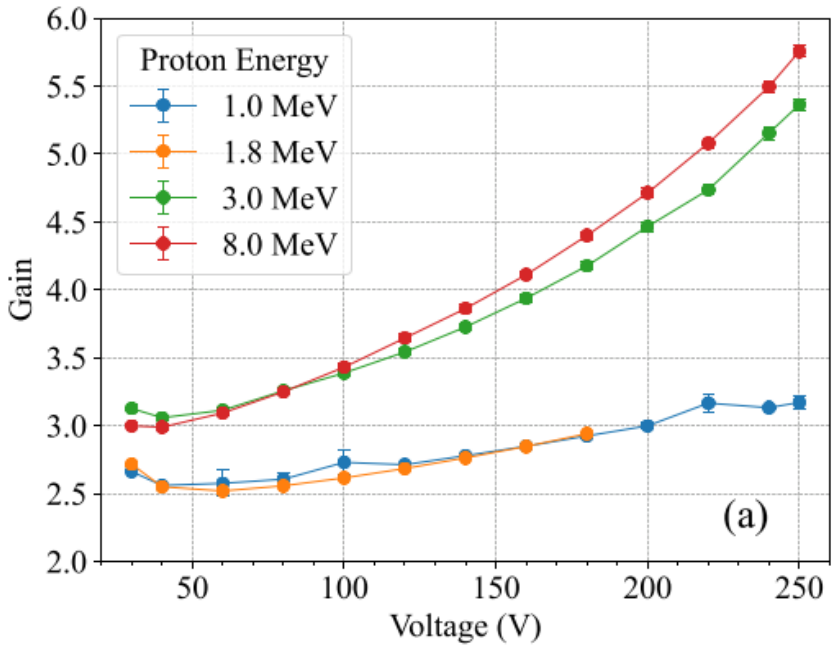


Fig. 5.33 *Resulting gain for protons of different energy incident on the type 3 LGAD from WF2 measured at CMRP.*

5.7 Discussion

The measurements presented in this chapter demonstrate that the fabricated LGADs show good performances. The IV and CV tests of section 5.4 compare well with the specifications of Table 5.4. The results shown in Figure 5.28 indicate that using a calibrated IR charge injection, a gain over 20 is obtained already at 80% of V_{bd} and the collected charge is about 20 fC. Also, the jitter results shown in Figure 5.30 demonstrate a resolution of around 10 ps using the same injection method. These results compare well with what has been reported in various publications on LGADs [46, 119, 120].

However, it is not straightforward to relate the results obtained from the laser and source charge injection, since the ionisation density significantly differs even if the same amount of charge has been deposited, affecting the LGAD response due to the gain suppression mechanism. The time resolution is also affected by the Landau fluctuations in the latter case. The initial results shown in Figure 5.31 where a timing resolution of around 40 ps is obtained at room temperature are however suggesting that the fabricated LGADs provide timing

capabilities close to the pre-irradiation required specifications of the HGTD. As no provision has been made in these runs of fabrication to compensate for the effects of radiation damage it is however expected that the performances of these devices will be significantly degraded at high radiation fluence. Plans for the complete evaluation of the fabricated LGADs and the irradiation campaign are discussed in section 7.2 of Chapter 7. Finally, the additional results presented in Section 5.6 demonstrate the interesting potential applications of LGADs in the field of radiation dosimetry.

Chapter 6

High Granularity LGAD

As discussed in previous chapters, the motivation behind proposing LGAD devices was precise timing resolution for MIP through internal charge amplification. The typical structure of LGAD includes a highly doped region just below the junction of a PIN diode sensor, as shown in Figure 3.11. When properly biased this doping region creates an electric field strong enough to ignite impact ionisation during the process of charge collection. This effect leads to amplification of the pulse signal that enhances the signal-to-noise ratio and achieves high timing resolution. Published results indicate timing resolution of the order of 30 ps and below for MIP [121] with frame rates in excess of 500 MHz [122].

The first planned application of LGADs is for the High Luminosity LHC (HL-LHC), in the ATLAS timing layer HGTD [10] and CMS End-Cap timing layer [123], to suppress the pile-up which would lower the efficiency in tracking and vertex reconstruction. However, the segmentation technique presented in Section 3.4.2 of Chapter 3, which includes JTE and p-stop, while being effective, has the downside of decreasing the fill factor, i.e. it reduces the gain area of the device, as the region between the pads provides no gain. This currently limits the achievable spatial granularity of these devices to a millimetre scale, a value nonetheless acceptable for the current HL-LHC applications. Future HEP applications would however benefit from devices [124, 125] achieving a granularity of 100 μm or less. The next generation of light sources would also require X-ray imaging of similar granularity [126].

In this chapter, after a summary of the proposed and investigated solutions on how to increase the fill factor in a segmented LGAD, a technology solution based on Deep Junction is presented along with TCAD simulations, which could be used to increase LGAD granularity. Section 6.1 discusses the issue arising in high segmentation LGADs using the conventional approach and briefly summarizes the proposed and investigated solutions to achieve a high fill factor. In section 6.2.1 a novel technological solution based on graded epitaxial growth (GELGAD) to obtain a buried gain layer during the epitaxial process growth of silicon is

introduced. Section 6.3 provides a full description of TCAD simulations performed on the proposed GELGAD structure, including process fabrication, gain and charge collection and realistic implementation of termination structures.

6.1 LGAD segmentation techniques

The typical fill factor for a PIN diode sensor can be in principle 100%, since under full depletion the total area is sensitive to incoming particles, making it possible to achieve high segmentation without significant fill-factor loss. The typical layout of devices with internal gain, like LGADs, has a Junction Terminating Extension (JTE) region at the edge of the GL, as shown in Figure 6.1a. The JTE, which is introduced to increase the breakdown voltage, reduces the region of the devices with gain. Spatially segmented LGAD sensors will then have a reduced fill factor because the inter-pad region will have no gain. The typical extension of this area is around $\approx 100\ \mu$m which limits the granularity of practical LGAD sensors to around 1 mm. Several solutions have been proposed and are currently being investigated to obtain higher granularity, [127, 128], a summary is reported here.

In the AC coupled LGAD design, [129, 130], the n^{++} implant well is replaced by a more resistive n^{+} layer, with electrodes that are AC coupled to it via a thin dielectric layer, as shown in Figure 6.1b.

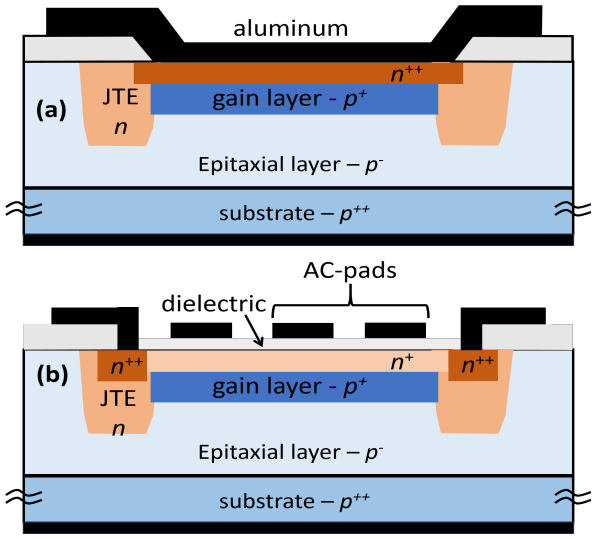


Fig. 6.1 *(a) Sketch of a section of a single-pad standard LGAD; (b) sketch of a section of a segmented AC-LGAD (not to scale).*

This approach eliminates the need for a JTE structure except at the very edge of the sensor, thus providing a non-segmented GL. A fill factor of nearly 100% can be achieved without complicating the fabrication process. The AC coupled signal and the intrinsic charge sharing in the resistive n^+ layer might however make it difficult to achieve a high repetition rate.

The so-called Inverse LGAD (iLGAD) or double-sided LGAD [131] implements a non-segmented GL on the opposite side of the sensor with respect to the conventional LGAD, as sketched in Figure 6.2. This approach allows segmentation techniques to be used at the readout surface, without the loss of gain between signal electrodes. The fabrication process requires however wafer processing on both sides, thus relatively thick wafers, which might also limit the achievable time resolution.

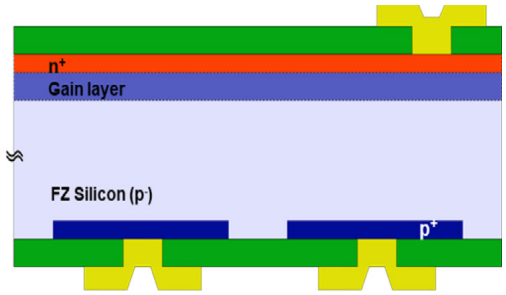


Fig. 6.2 *Sketch of a section of an inverted or double-sided LGAD, with the gain layer on the opposite side of the electrodes (not to scale).*

In the Trench-Isolated LGAD (TI LGAD) approach [132] the isolation among neighbouring pads is achieved through deep trenches, as shown in Figure 6.3.

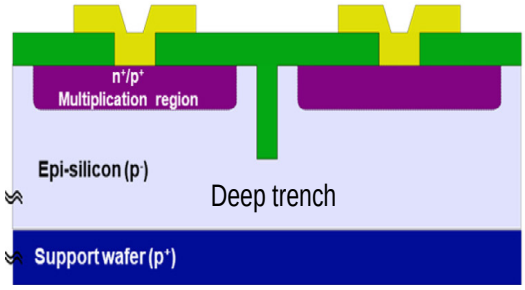


Fig. 6.3 *Sketch of a section of a Trench-Isolated (TI) LGAD (not to scale).*

In TI LGADs the GL is the same as in conventional LGAD but the termination and isolation structures (p-stop and JTE) are replaced by a deep trench, less than a μm wide, a process widely used for pixel isolation in CMOS imagers and Silicon Photomultipliers

(SiPM). This isolation considerably reduces the extension of the no-gain region between pads, increasing the fill factor.

Another approach pursued to increase the granularity without sacrificing the fill factor is the deep-junction (DJ-LGAD) [133], schematically shown in Figure 6.4. A deep junction, made by a large area of uniform n^+ and p^+ gain implants, sits a few microns from the surface where n^+ DC coupled electrodes are placed. In this design the GL region is contained between the buried n^+ and p^+ gain layer, while everywhere else the electric fields are well below the threshold for impact ionisation but high enough to guarantee velocity saturation of carriers. Electrons multiplied in this region then drift to be collected by the n^+ electrodes at the interface while holes drift towards the back. The buried GL is obtained either by abutting two wafers onto each other after gain layer implantation or by implanting it on a base wafer and then adding an epitaxial deposition to form the avalanche region [134]. If the buried implant is patterned to terminate inside the cathode the edge JTE termination can be avoided, which might be beneficial for high granularity detector design.

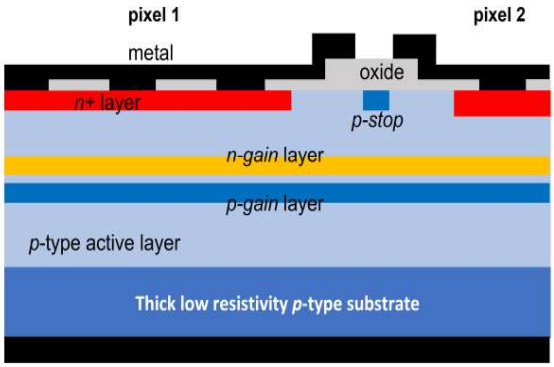


Fig. 6.4 *Sketch of a section of a Deep-Junction (DJ) LGAD (not to scale).*

6.2 Graded Epitaxial Layer LGAD (GELGAD)

Here an alternative approach is proposed, that still makes use of a buried GL but is obtained by graded epitaxy. This solution directly embeds the GL in the silicon wafer during the epitaxial growth, a solution which should provide a uniform gain layer throughout the entire cell array. Directly embedding the GL during the wafer growth also avoids the additional process step of implantation and annealing needed for its creation and the complicated process of wafer-to-wafer bonding. Formation of the GL via graded epitaxy should also avoid issues related to defects in the epitaxial created by implantation [134], which can increase the leakage current and decrease the gain. The proposed approach described here, presented with full TCAD simulations, uses the LGAD layout of Run 2 described in Chapters 4 and 5, i.e. 1.3 mm size cells to be used in the ATLAS HGTD. The same fabrication process could however be employed for the fabrication of smaller size LGADs.

6.2.1 Graded Epitaxy LGAD structure

The proposed Graded Epitaxy LGAD (GELGAD) structure is schematically shown in Figure 6.5 and 6.6, which refer to the two fabrication options investigated in this thesis. The GL consists of the n^+p^+ junction located within the epitaxial layer at some depth. The two solutions presented differ in the depth at which the GL is located, to investigate its effects on the LGAD performance [135].

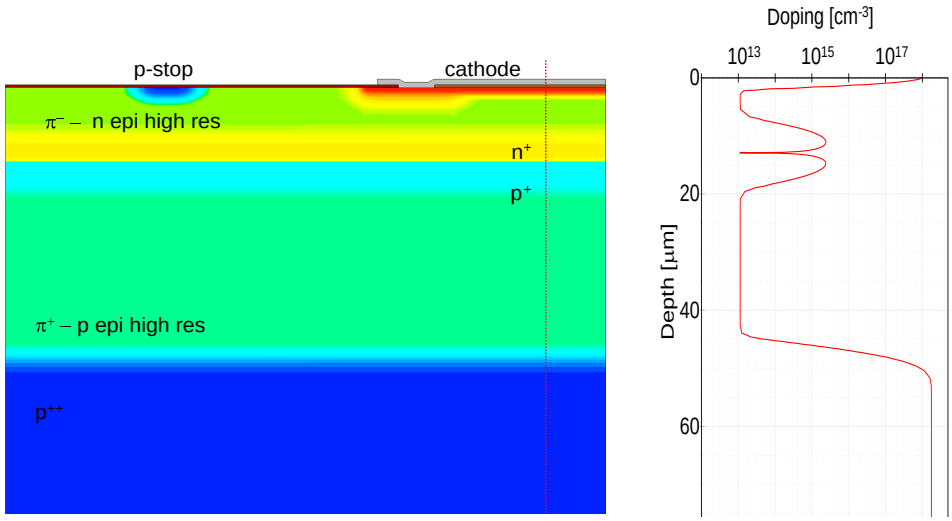


Fig. 6.5 *Left: Shallow GELGAD cross-section. Right: Doping profile.*

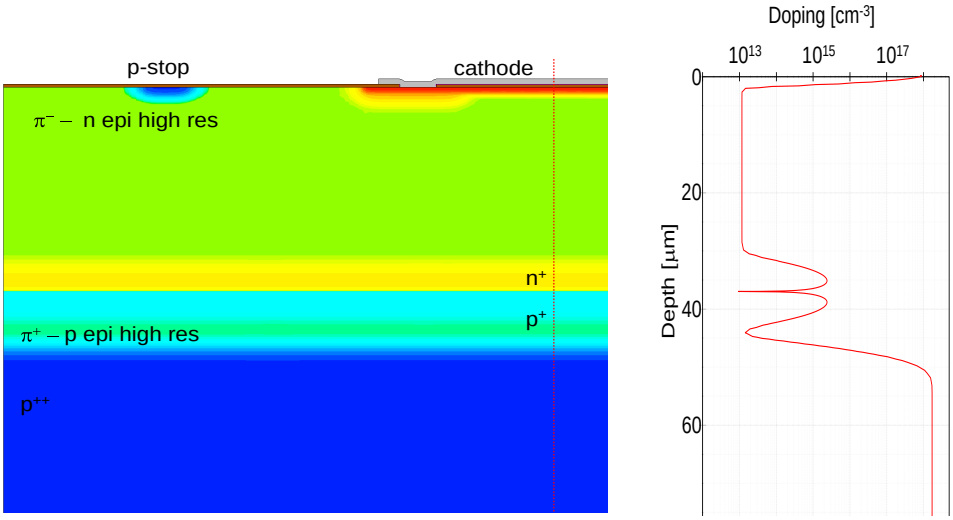


Fig. 6.6 *Left: Deep GELGAD cross-section. Right: Doping profile.*

The shallow and deep GELGAD have a junction at a depth of 13 μm and 37 μm respectively. The total epitaxial layers thickness is 50 μm, the same used for the fabrication of the Te2v LGADs presented before. Figure 6.7 shows the 2D section of the LGAD array, described in section 4.2.2 of Chapter 4, which is used in the TCAD simulations presented in this chapter.

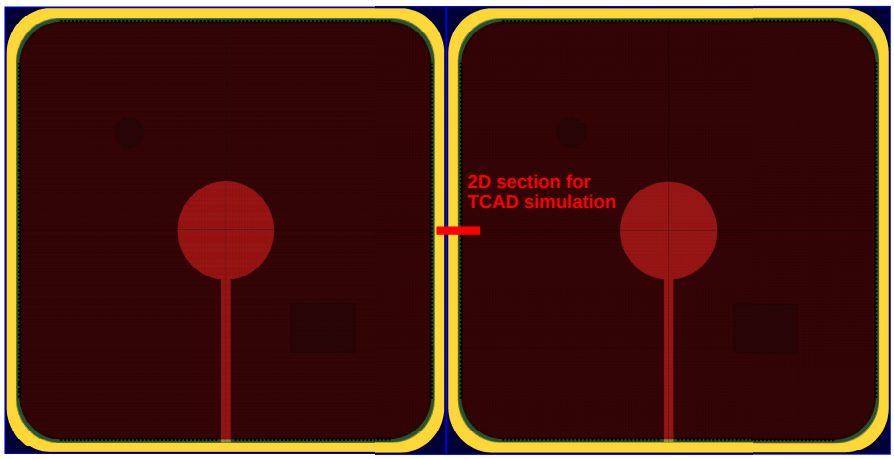


Fig. 6.7 *2D section of Te2v Run 2 cell array used for TCAD simulation of GELGAD.*

By directly embedding the GL in the wafer this solution only requires the common planar processing of detectors to fabricate a sensor and, as such, might be more easily implemented in a standard fabrication process by most silicon foundries. This could eventually lead to a monolithic integration of LGAD and readout on the same silicon chip.

6.3 GELGAD simulation

Full TCAD simulations, which included fabrication steps, DC electrical characteristics and charge collection efficiency, have been carried out to investigate the performances of GELGAD and its potential advantages over the conventional LGAD structure.

6.3.1 GELGAD process simulation

The fabrication process of the proposed GELGAD is shown in Figure 6.8. All the fabrication steps have been implemented in the SPROCESS simulator to obtain a realistic GELGAD structure. Full details of each fabrication step cannot be disclosed as they are proprietary information owned by the semiconductor foundry, nonetheless, a summary description of the proposed manufacturing process is provided below.

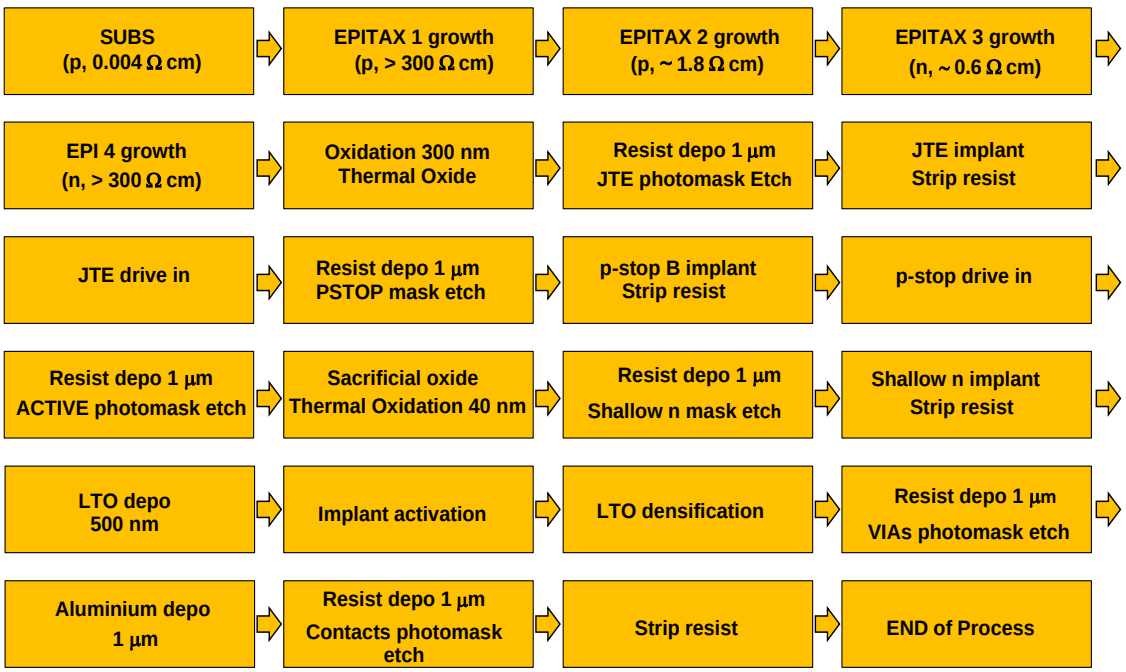


Fig. 6.8 *GELGAD fabrication process flow.*

Each of the steps of the fabrication process has been discussed and checked with the semiconductor foundry [136], including the steps of graded epitaxial layers growth, their attainable thickness and doping accuracy, to guarantee the feasibility of the approach. The epitaxial growth of the GL is performed on a low resistivity substrate with $\rho \approx 0.004$ Ω·cm. In the proposed fabrication process the graded epitaxial layers have a minimum thickness of 3 μm and a resistivity ρ of 0.6 (n) and 1.8 (p) Ω·cm respectively. The high resistivity *n*

and p epitaxial layers, above and below the GL, have a resistivity $\rho > 300$ Ω·cm. The total thickness of the active region, i.e. the combined thickness of the epitaxial layers, is kept at 50 μm. After the GL growth, shallow or deep, the subsequent process steps implement the JTE structure. In this design the JTE is retained to increase the breakdown voltage and, due to the buried location of GL in GELGAD, its presence should not affect the gain uniformity. Design of higher granularity detectors, of reduced cathode width, might avoid the JTE structure. Next, the process of implantation of p-stop implants for the isolation of neighbouring cells, followed by the n-well cathode implantation is performed. The n-well implantation is performed through a thin layer of oxide, to suppress the channeling of the implanted ions. The steps of p-stop and n-well annealing used in this design widen the buried layers of the GL, according to the simulation, up to a total width of ≈ 15 μm. This width could be further reduced by modifying the fabrication steps or using a rapid thermal annealing (RTA) process if a thinner than 50 μm active region of the detector was targeted. The resulting doping profile obtained from SPROCESS using the flowchart 6.8 for the two investigated GELGAD, shallow and deep, are shown in Figure 6.9 and Figure 6.10.

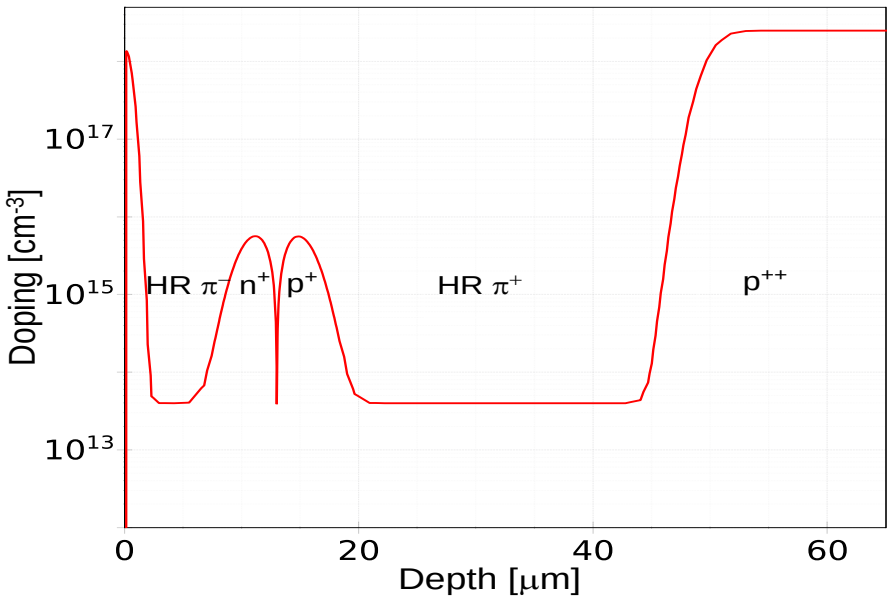


Fig. 6.9 *Shallow GELGAD doping profile obtained by SPROCESS using the process flow of Figure 6.8.*

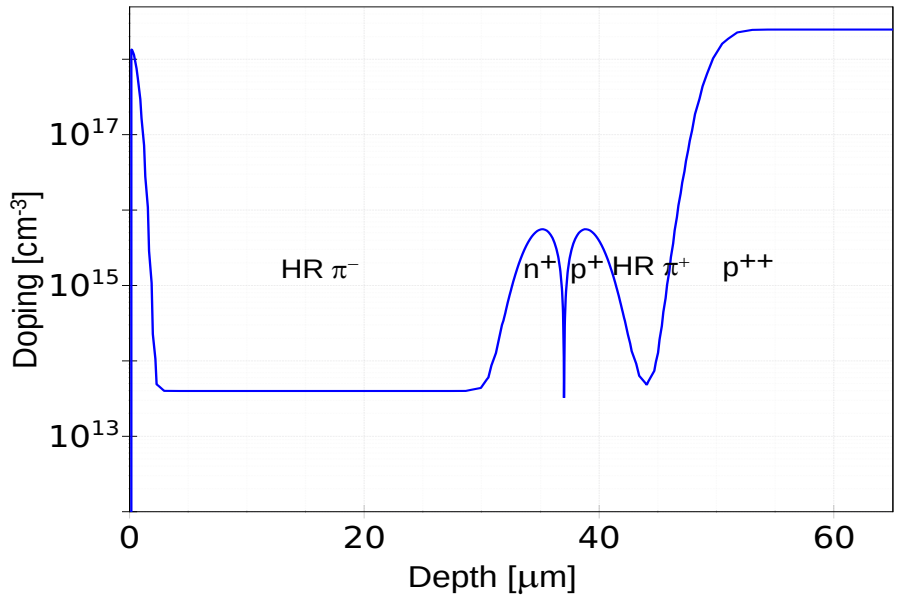


Fig. 6.10 *Deep GELGAD doping profile obtained by SPROCESS using the process flow of Figure 6.8.*

6.3.2 GELGAD DC electrical simulations

Current vs. voltage (IV) simulation results for the shallow and deep GELGAD structures are shown in Figure 6.11.

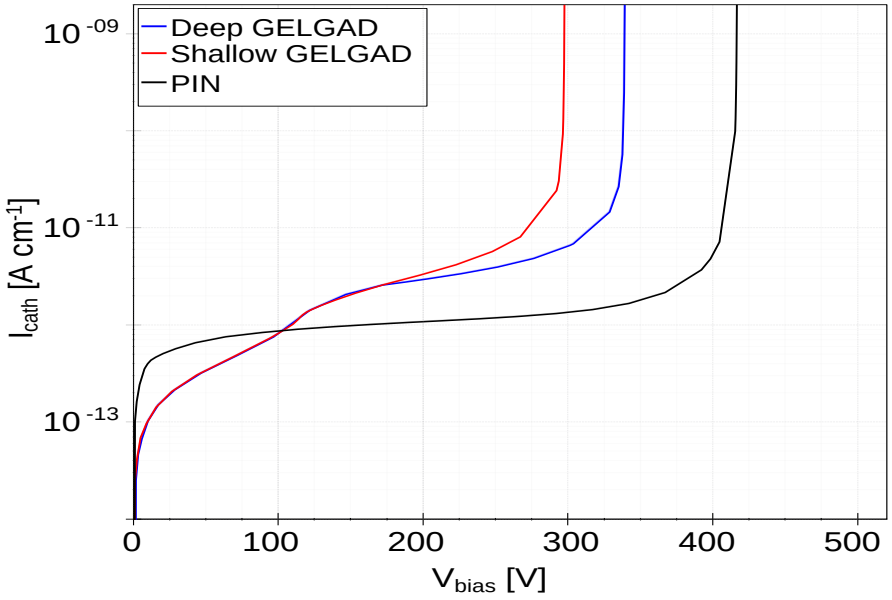


Fig. 6.11 *Simulation of current vs. voltage of shallow and deep GELGAD. The plot of PIN diode current of the same total epitaxial thickness is also shown. T = 21 °C.*

The same SDEVICE settings as described in section 4.3.3 of Chapter 4 were used in these DC simulations. The substrate was kept to ground and a voltage sweep was applied to the cathode to locate the point of breakdown. The impact ionisation model used was the University of Bologna model [87]. The simulations predict a breakdown for both GELGAD structures lower than that of the PIN diode, i.e. 297 V for the shallow GELGAD, 337 V for deep GELGAD and 418 V for the PIN diode. Nonetheless as discussed in section 4.3.3 of Chapter 4, the uncertainty on the impact ionisation process modelling can lead to inaccuracies in the prediction of the breakdown voltages V_{bd} absolute value.

Figures 6.12 and 6.14 show a 2D map of the equipotential lines and the x-component of the electric field, i.e. the field component along the vertical direction[1] at 95% of breakdown voltage V_{bd} (from here on indicated as V_{b95}) for the deep and shallow GELGAD. The maximum value of the field is seen in the buried GL region, in both structures is around 3×10^5 V·cm^{-1}, high enough to ignite impact ionisation. The remaining region of the bulk shows a weaker field of around 10^4 V·cm^{-1} sufficiently high to guarantee fast drift of charge. Figures 6.13 and 6.15 show the fields along the cut lines C1 and C2 shown in Figures 6.12 and 6.14, i.e. along the vertical direction through the middle of the p-stop and from within the cathode n-well, for a bias of V_{b95}. The peak value is 2.99×10^5 V·cm^{-1} and 3.04×10^5 V·cm^{-1} for the shallow and deep GELGAD structures, which correspond to a value of the ionisation coefficient α_e of 7.0×10^3 cm^{-1} and 8.3×10^3 cm^{-1} respectively.

In both structures along the C1 cut line a drop in field strength is observed, which will be investigated in the simulation of multi-cell structure in the next section of this chapter.

[1] In the structures generated by SPROCESS the x coordinate indicates the vertical direction, i.e. the depth. The same convention is used here.

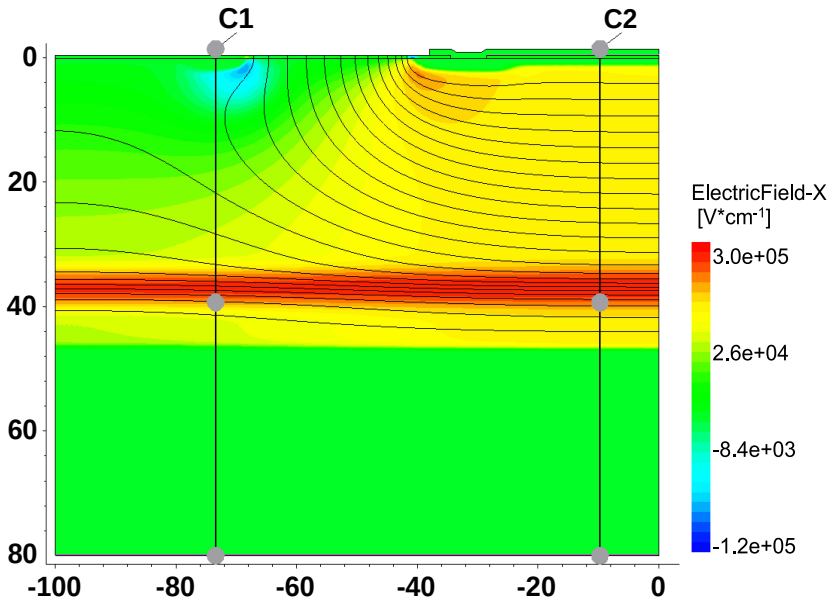


Fig. 6.12 *Deep GELGAD electric field at V_{b95} bias. The contours show the equipotential lines, indicating that the field is directed towards the cathode. Sizes in µm.*

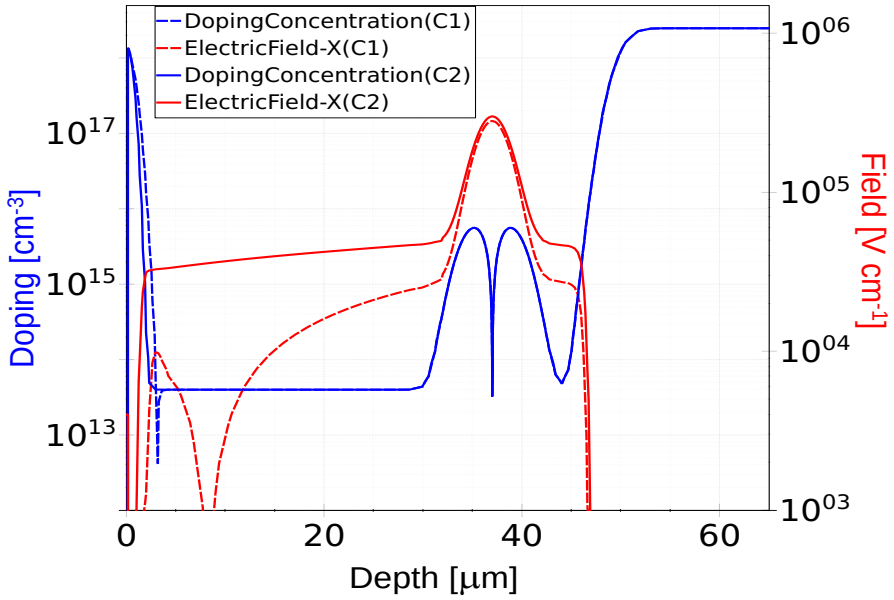


Fig. 6.13 *Doping concentration and electric field profile of deep GELGAD at V_{b95} bias along the two cut lines C1 and C2. T = 21 °C.*

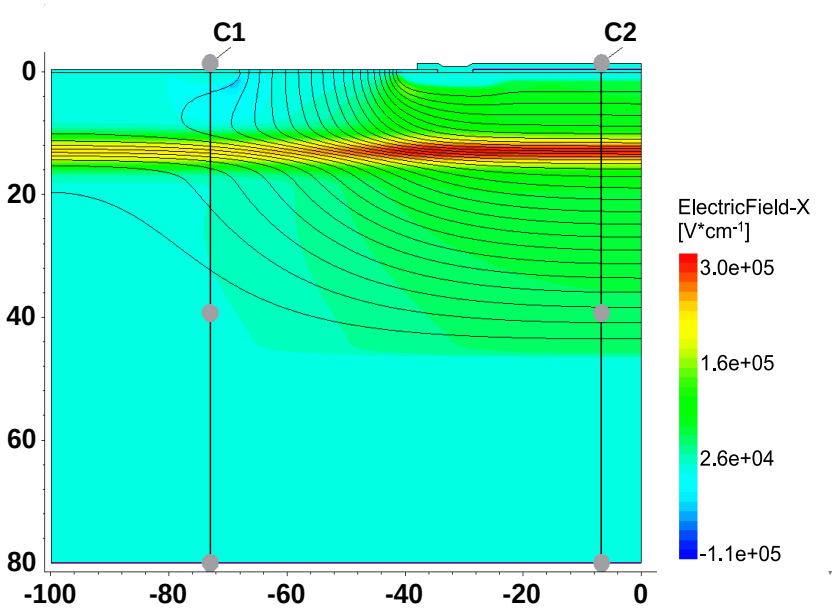


Fig. 6.14 *Shallow GELGAD electric field at V_{b95} bias. The contours show the equipotential lines, indicating that the field is directed towards the cathode. Sizes in µm.*

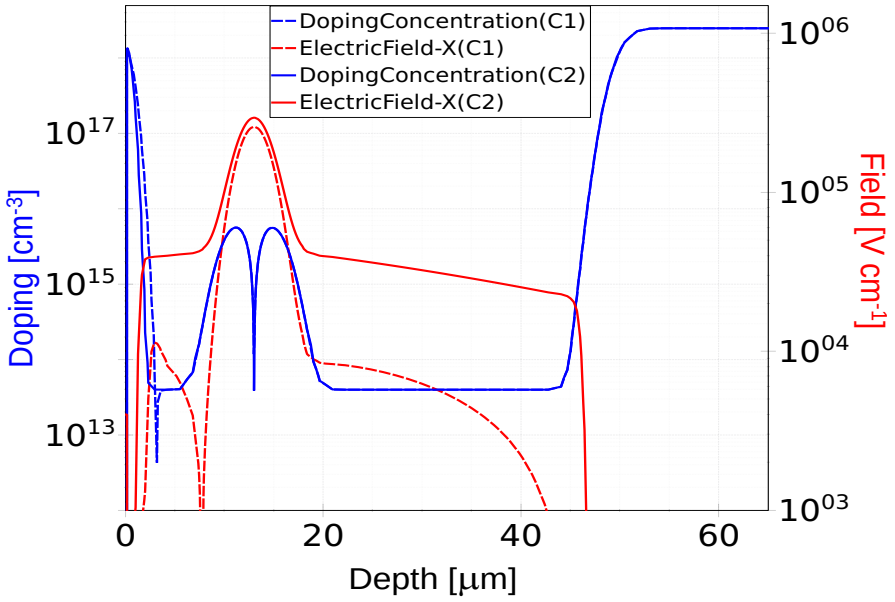


Fig. 6.15 *Doping concentration and electric field profile of shallow GELGAD at V_{b95} bias along the two cut lines C1 and C2. T = 21 °C.*

6.3.3 GELGAD charge collection simulations

Charge collection simulations were performed on the GELGAD structures using the same setup of MIP charge injection with the SDEVICE heavy ion statement from the TCAD package described in section 4.3.3 of Chapter 4. Single and dual GELGAD cells were simulated to investigate gain versus bias voltage, gain uniformity and cross talk among neighbouring cells.

GELGAD single cell charge collection

Figure 6.16 shows an example of transient current of a shallow GELGAD and a PIN diode at a bias voltage normalized to the breakdown for a MIP central hit, i.e. along cutline C2 in Figure 6.14. In the simulation, the MIP particle hits the device at t = 100 ns and the transient is followed for the next 5 ns.

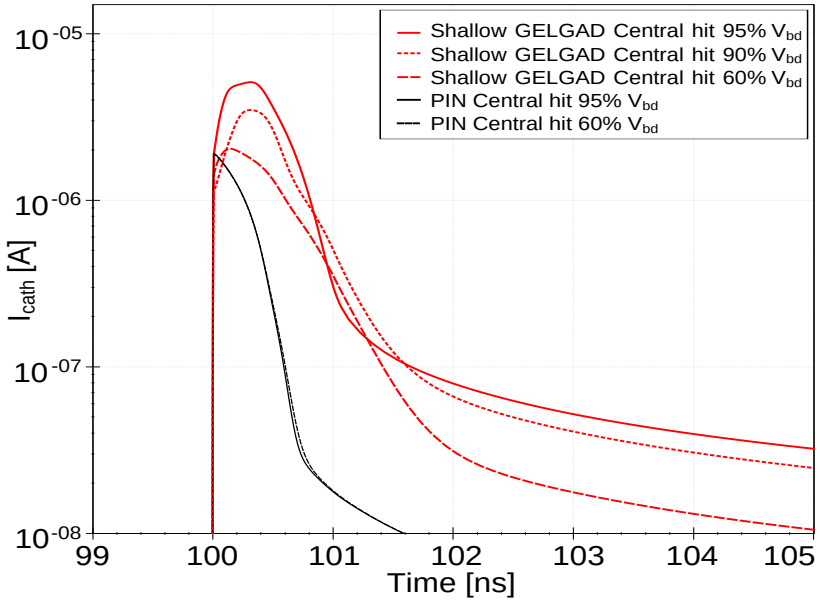


Fig. 6.16 *Shallow GELGAD transient current for C2 MIP hit. T = 21 °C.*

The PIN pulse current does not show a relevant change in magnitude as the bias voltage is increased, as expected. Conversely the GELGAD current increases in magnitude as a result of the higher gain. The slope of the GELGAD pulse current reaches a maximum of 25.5×10^3 $A \cdot s^{-1}$ for a bias of V_{b95}. The first part of high slope pulse current, of duration of $\approx$ 200 ps, is due to electrons crossing the gain layer and it ends when all electrons have traversed it. The next part of the transient is due to electrons drifting towards the cathode and it lasts until they have all been collected. The total duration of these transients due to the drifting electrons is

approximately 400 ps. After this time the pulse current is due to the holes drifting towards the backside and it lasts approximately 1.1 ns. The gain and collected charge of the shallow GELGAD, calculated as described in section 4.3.3 of Chapter 4 by taking the ratio of current integrals, are plotted against the bias voltage expressed as a fraction of the breakdown in Figure 6.17.

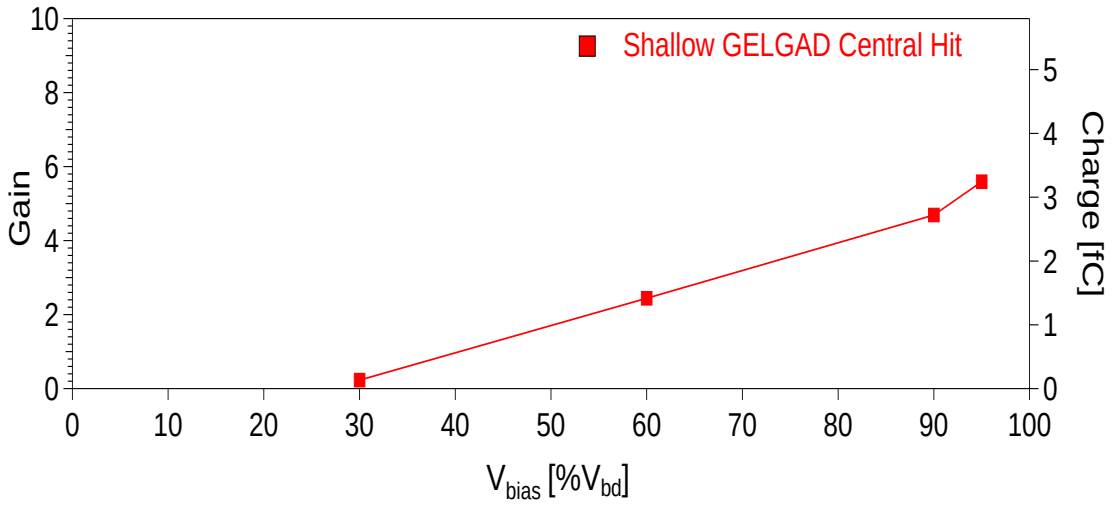


Fig. 6.17 *Shallow GELGAD gain and collected charge as a function of bias voltage expressed as a percentage of breakdown.*

As noted in section 4.3.3 of Chapter 4, different impact ionisation models available in SDEVICE predict different values of breakdown and gain, with lower values obtained using the University of Bologna model. Using this model, the GELGAD simulations estimate the gain to be ≈ 6 at a bias voltage of V_{b95}. The estimated collected charge exceeds 3 fC.

The same simulations using the same MIP injection setup were performed for the deep GELGAD structure of Figure 6.12. The transient currents simulations along the cutline C2 are shown in Figure 6.18 and gain with collected charge results in Figure 6.19.

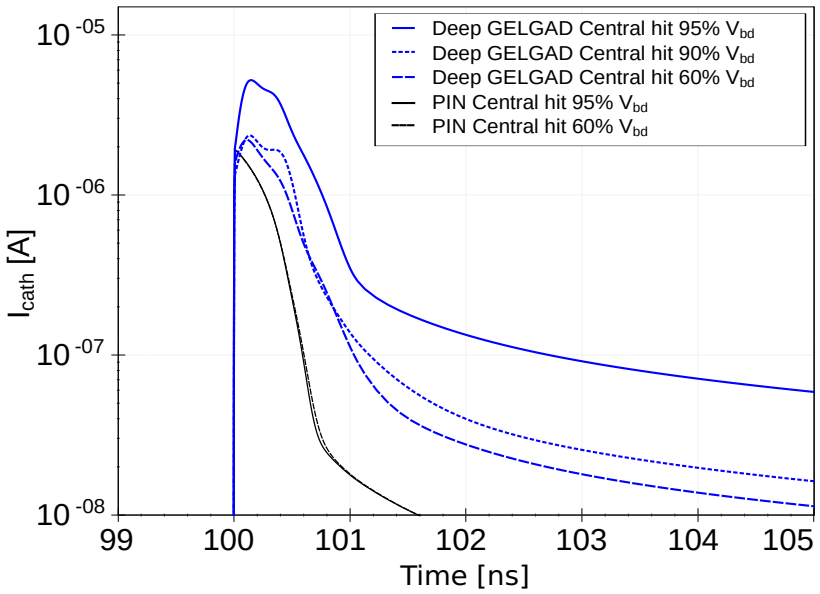


Fig. 6.18 *Deep GELGAD transient current for C2 MIP hit. T = 21 °C.*

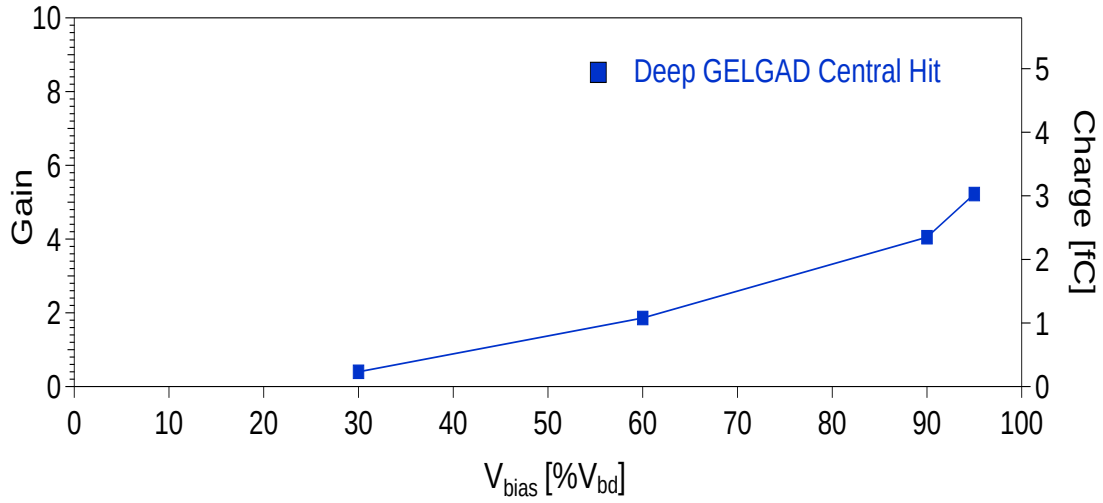


Fig. 6.19 *Deep GELGAD gain and collected charge as a function of bias voltage expressed as a percentage of breakdown.*

Figure 6.20 compares the pulse current for the two GELGAD structures, shallow and deep, for the first 1 ns of the transient following a MIP hit. While the peak values of the pulse current are predicted to be the same, the slope of the deep GELGAD current reaches a maximum of 35.4×10^3 A·s^{-1} for a bias of V_{b95}, a slightly higher value compared to the 25.5×10^3 A·s^{-1} of the shallow GELGAD. This stems from the fact that the peak value of the electric field in the deep GELGAD is predicted to be higher than in the shallow GELGAD,

as shown in Figures 6.13 and 6.15, leading to a higher ionisation coefficient. However, as the gain transient is shorter in the deep GELGAD, the overall gain is lower than in the shallow GELGAD, Figure 6.21.

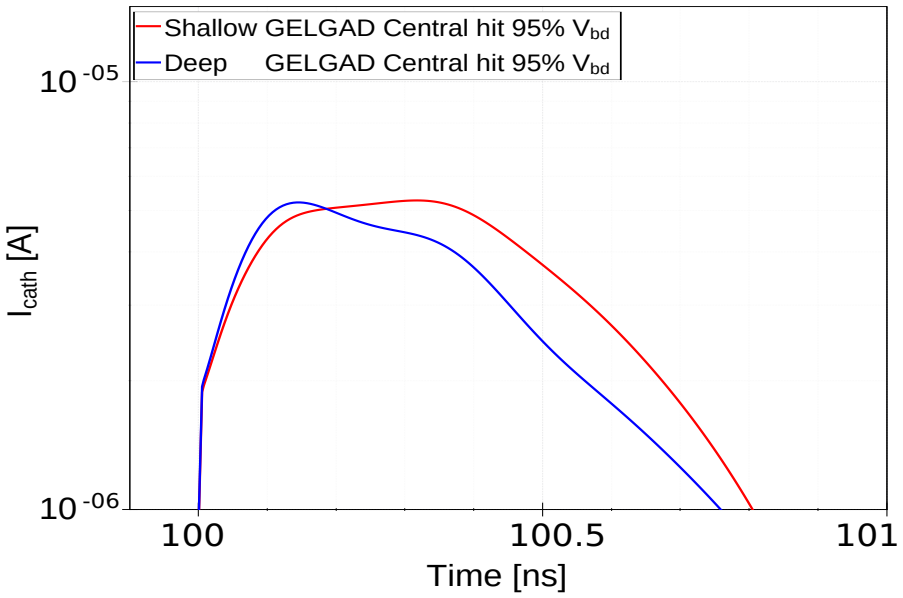


Fig. 6.20 *Shallow and Deep GELGAD pulse current for a central hit.*

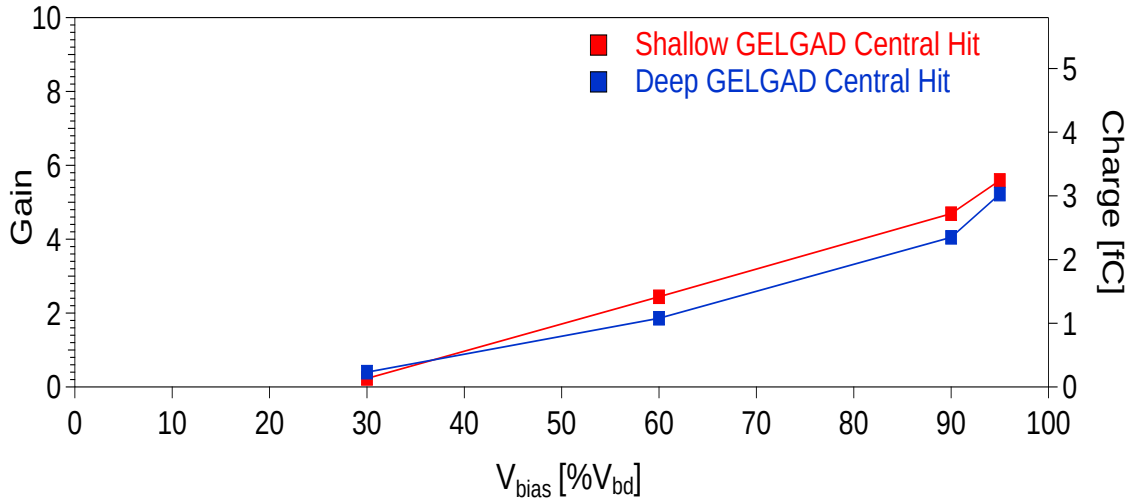


Fig. 6.21 *Shallow and Deep GELGAD gain and collected charge bias voltage expressed as a percentage of breakdown.*

GELGAD dual cell charge collection

Shallow and deep GELGAD structures consisting of two neighbouring cells were also simulated to investigate gain uniformity and cross talk. The gain uniformity over the device surface was investigated for the structures of Figures 6.22 and 6.23 by performing MIP charge injections along the directions C1, C2 and C3 (i.e. in between cells, halfway between cathode and p-stop and within the cathode n-well), integrating the pulse current of cathode 1 to measure the charge collected and comparing it with the collected charge of an equivalent PIN diode.

Figures 6.24 and 6.25 show the pulse current results for the two GELGAD structures and PIN diodes biased at V_{b95}. Pulse currents for cathode 1 and 2 are individually shown for the GELGAD in case of charge injection along C2 and C3. For the PIN diode only cathode 1 pulse currents are shown due to the negligible crosstalk. The GELGAD pulse current for the neighbouring cathode 2 for charge injection along C2 and C3 is due to capacitive coupling and it is significantly smaller than the cathode 1 current. After a transient of less than 1 ns, the current integrates to zero, indicating that there is no charge collected by cathode 2. The pulse current for C1 injection in the middle of the two cells is of course the same for both electrodes C1 and C2.

In all three injections cases C1, C2 and C3, the GELGAD pulse current for both structures is higher than the corresponding one for the PIN diode, indicating that there is charge gain. The magnitude of the gain decreases in moving from C3 to C1, i.e. transversally along the sensor surface, due to the fact the peak value of the electric field in the buried gain layer decreases along the same direction, as shown in Figure 6.15 and 6.13. This is due to the potential drop in the high resistivity epitaxial layer. Decreasing the inter-cell distance, for instance in the case of an array of cells of smaller size, would reduce this effect, making the gain more uniform. Figure 6.26 shows the gain and the collected charge for the two GELGAD structures versus transverse position at points C1, C2 and C3 at V_{b95} bias.

For the C1 injection, the deep GELGAD structure shows higher gain compared to the shallow GELGAD, leading to a better gain uniformity. The lowest gain is predicted to be ≈ 2.5 at the C1 position, which corresponds to a drop of approximately 50% from the maximum value of 5.35. For the C1 injection, the deep GELGAD shows an initial pulse current of a higher slope compared to the shallow GELGAD, which is essentially flat for the first 200 ps, until the gain electrons arrive at the collecting cathode. This is a consequence of Ramo's theorem [137]. The product $ev_d \cdot F_w$, with v_d the electrons drift velocity and F_w the weighting field, has smaller values for the shallow GELGAD due to the large angle subtended between the vertical direction, where v_d is directed as a result of the large field in the GL, and the surface of the cathode 1. In the case of deep GELGAD, such an angle is

smaller, which results in a higher current. A reduction of the inter-cell distance would reduce such effect, without suppressing it completely. A deep gain layer structure seems therefore more advantageous in terms of achieving a high slew rate, which is beneficial for timing measurements.

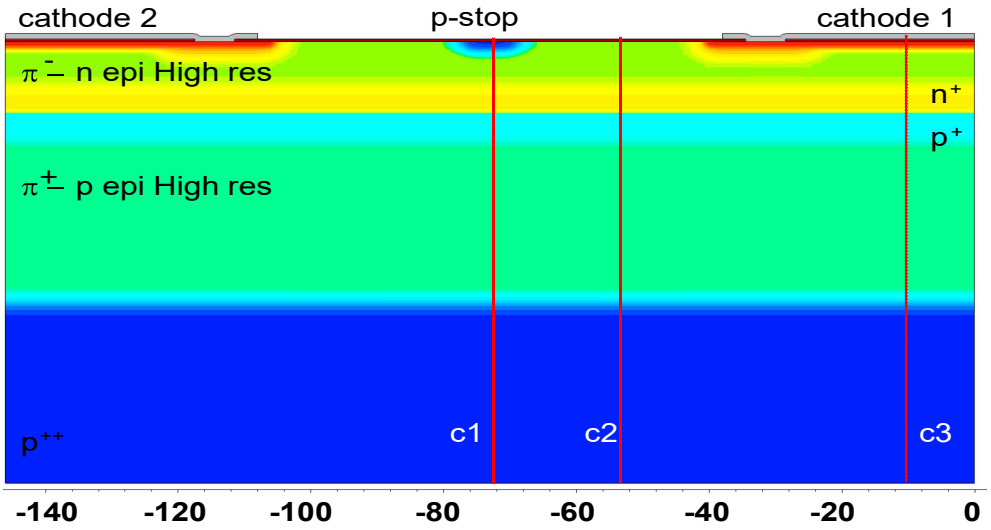


Fig. 6.22 *Shallow GELGAD two cells structure. Sizes in µm.*

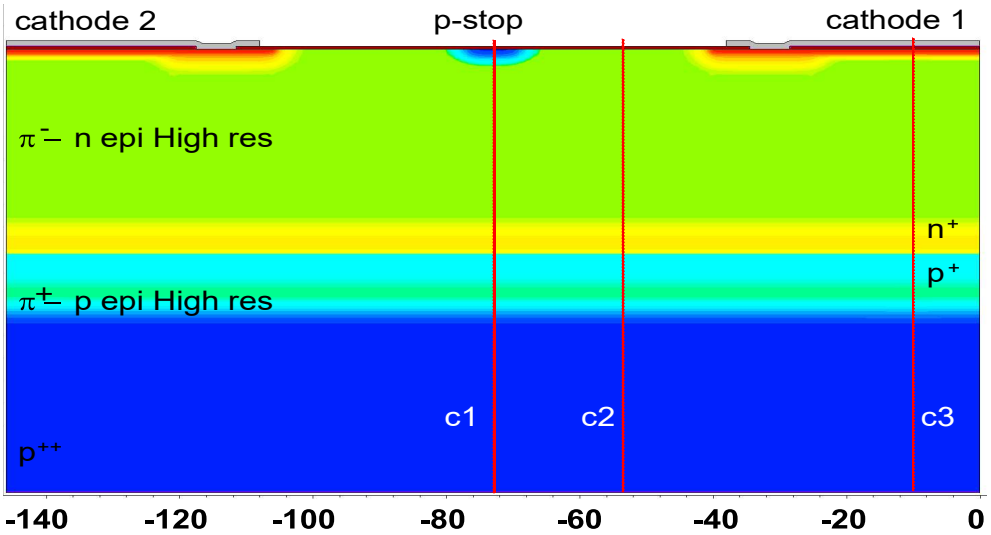


Fig. 6.23 *Deep GELGAD two cells structure. Sizes in µm.*

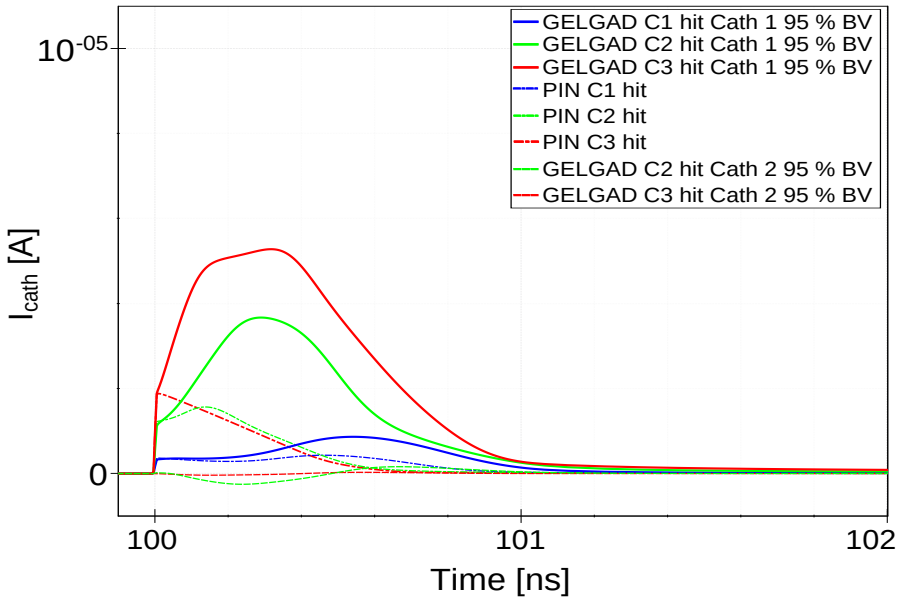


Fig. 6.24 *Transient current generated by MIPs in a two-cell shallow GELGAD and a PIN diode. T = 21 °C.*

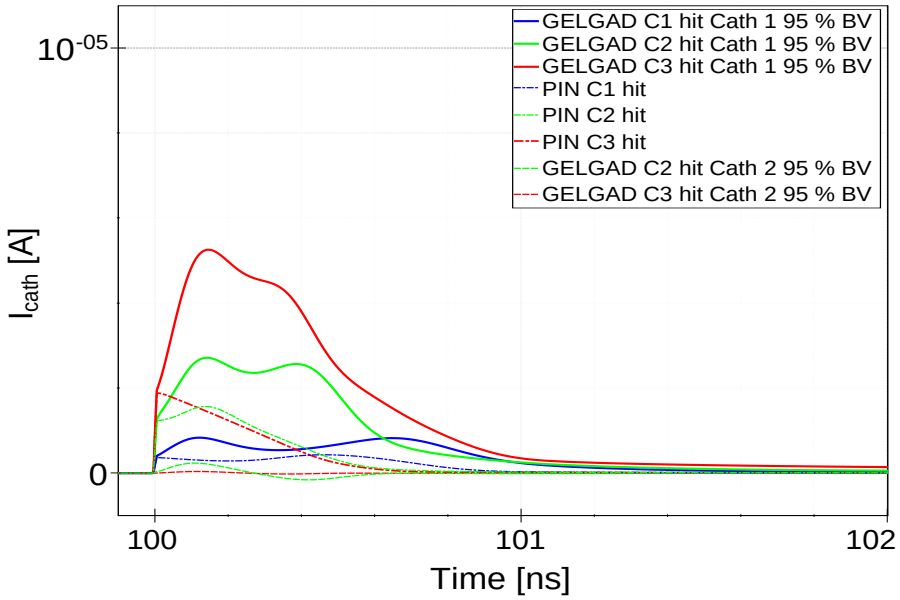


Fig. 6.25 *Transient currents generated by MIPs in a two-cell deep GELGAD and a PIN diode. T = 21 °C.*

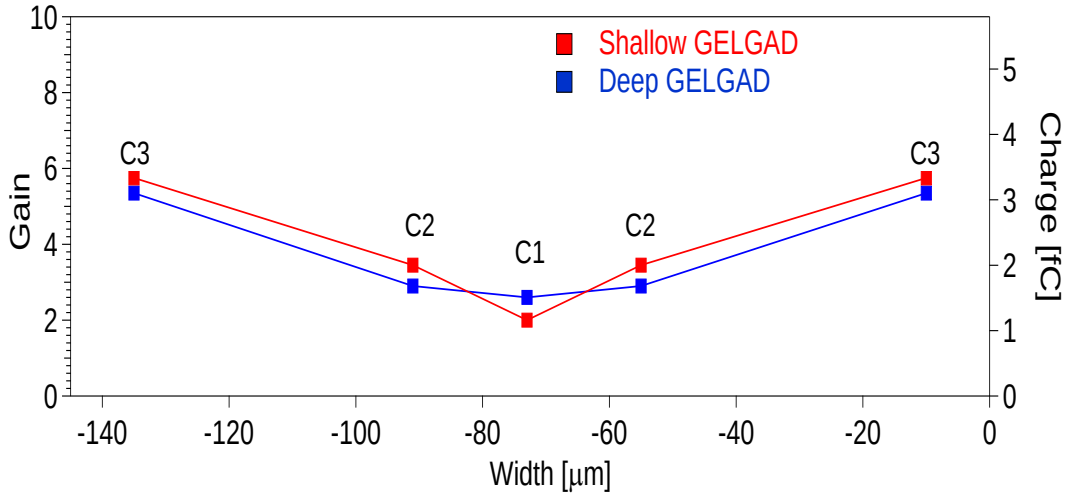


Fig. 6.26 *GELGAD two cells MIP gain vs. transverse position. T = 21 °C.*

Tables 6.1 and 6.2 summarize the gain, charge collection and slew rate for the two GELGADs and the three injection points C1, C2 and C3.

Table 6.1 *Shallow GELGAD results - T = 21 °C.*

Specifications	C1	C2	C3
Gain [a.u.]	2.0	3.45	5.75
Charge [fC]	1.2	2.05	3.35
Slew Rate [A·s^{-1}]	0.01×10^4	1.25×10^4	2.8×10^4

Table 6.2 *Deep GELGAD results - T = 21 °C.*

Specifications	C1	C2	C3
Gain [a.u.]	2.53	2.9	5.34
Charge [fC]	1.5	1.72	3.1
Slew Rate [A·s^{-1}]	0.53×10^4	1.4×10^4	3.6×10^4

6.3.4 GELGAD implementation

In the proposed GELGAD structure, the buried gain layer is obtained during the epitaxial growth of silicon. Therefore, a high-field semiconductor junction extends beyond the periphery of the cells to the wafer boundary. To investigate the effects this termination of the junction has on the leakage current and breakdown, TCAD simulations were performed on the deep and shallow GELGAD structures of Figures 6.27 and 6.28. The cross-sections shown refer to the layout of the Run 2 LGAD termination structure described in section 4.2.2 of Chapter 4 and include a guard ring and a cell boundary. The same fabrication steps discussed in section 4.3.1 of Chapter 4 were used here to obtain the guard ring and cell boundary termination structures.

Figures 6.29 and 6.30 show the electron ionisation coefficient α_e and depletion region obtained at V_{b95} for the two GELGADs. In both cases, the highest value of α_e is located in the buried GL, with the onset of ionisation appearing in the JTE region at the cathode's edge. The depletion region in case of deep GELGAD fully extends beyond the guard ring up to the cell boundary. A reduced depletion is however observed in the case of shallow GELGAD, with the fully depleted region not extending beyond the p-stop. In terms of charge collection efficiency and timing performances, this would suggest that the deep structure is preferable, as it guarantees fast drift of charge over the entire pixel region.

IV plots are shown in Figure 6.31. The predicted breakdown voltages do not substantially differ from those of Figures 6.11, indicating that the termination region does not cause an early breakdown or excessive increase of leakage.

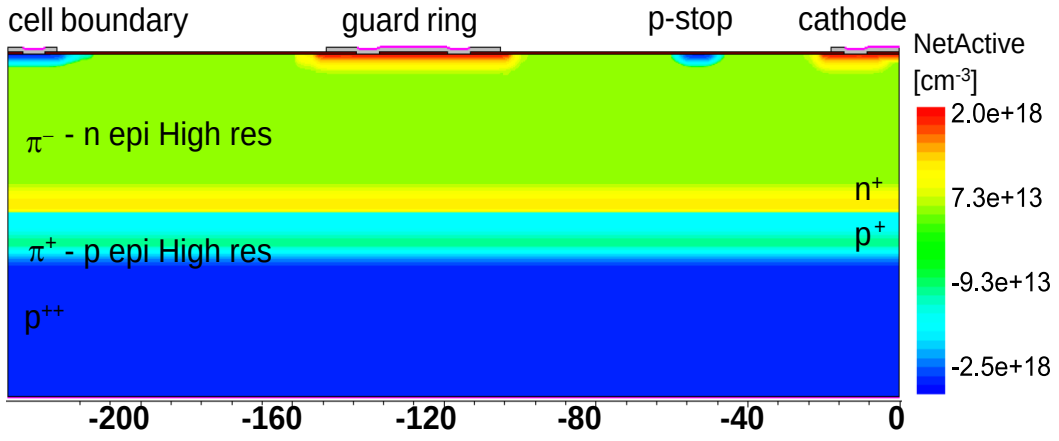


Fig. 6.27 *Deep GELGAD termination structure. Sizes in µm.*

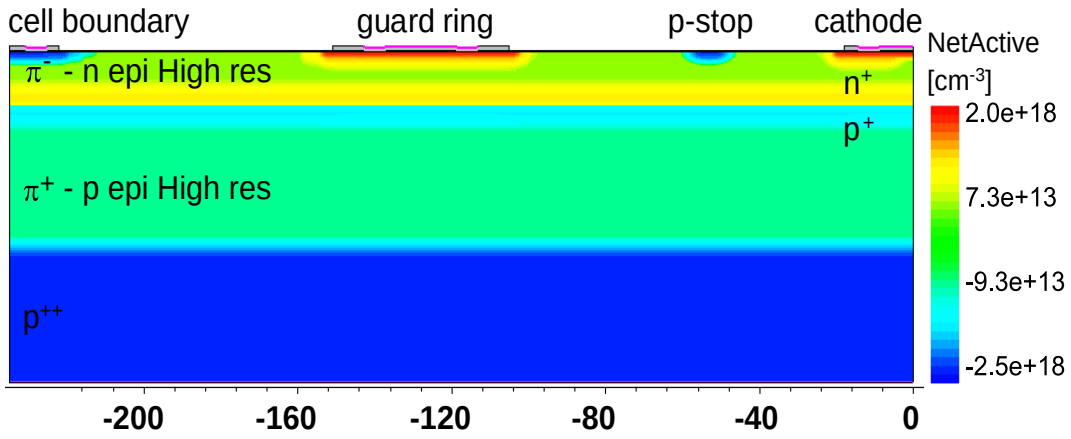


Fig. 6.28 *Shallow GELGAD termination structure. Sizes in µm.*

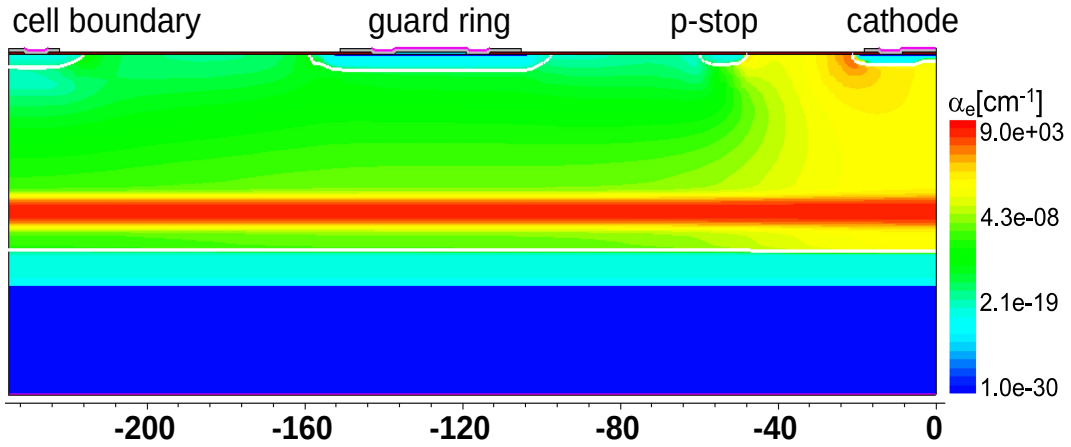


Fig. 6.29 *Deep GELGAD periphery at V_{b95} bias showing the electron ionisation coefficient α_e. The depletion region is shown in thick white lines. Sizes in µm.*

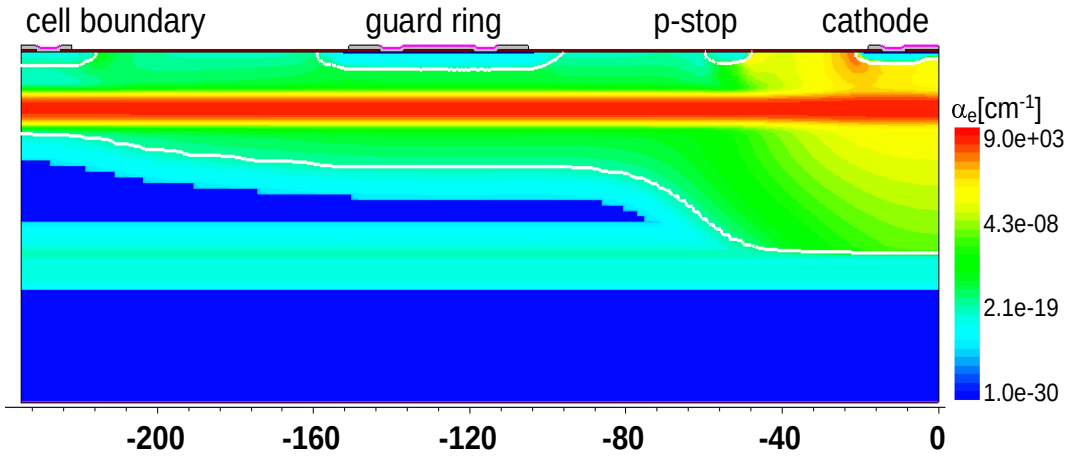


Fig. 6.30 *Shallow GELGAD periphery at V_{b95} bias showing the electron ionisation coefficient α_e. The depletion region is shown in thick white lines. Sizes in µm.*

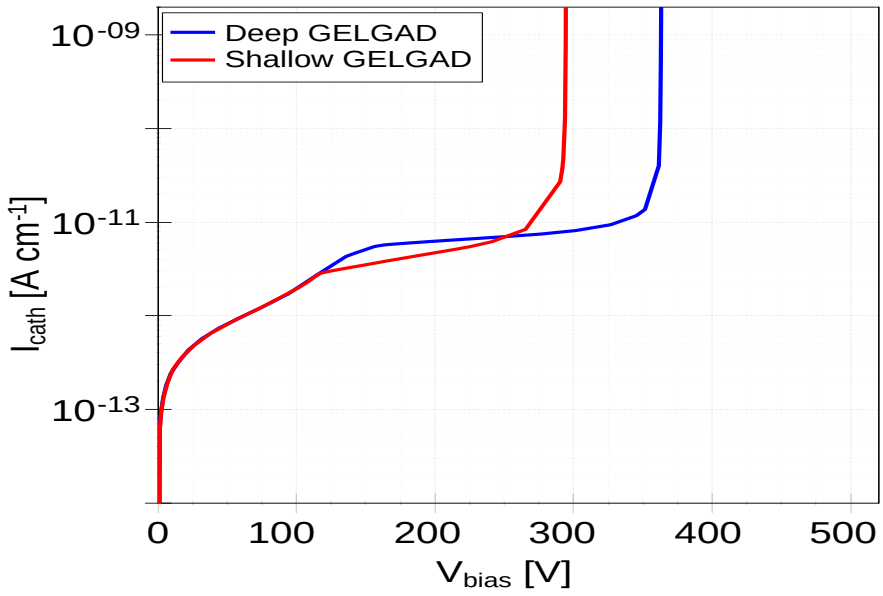


Fig. 6.31 *Current vs. voltage of the shallow and deep GELGADs of Figure 6.27 and 6.28. Besides a slightly higher leakage, due to the bigger semiconductor volume simulated, the V_{bd} is the same as in Figure 6.11. Temp = 21 °C.*

As mentioned in section 6.3.1, the quoted uncertainty in attainable doping of the graded epitaxial layers is around ±0.2 Ω·cm. Additional simulations have been performed to assess the sensitivity of GELGAD to doping variations of the buried gain layer. Figures 6.32, 6.33, 6.34 and 6.35 show the field profile of deep and shallow GELGAD structures obtained for doping variations of ± 0.2 Ω·cm respectively when the structure is biased at V_{b95} along the same cutlines C1 and C2 of Figures 6.12 and 6.14.

For the deep GELGAD structure, increased doping of the GL still exhibits the peak amplitude of the field located in the buried layer with the rest of the bulk which remains fully depleted, as shown in Figure 6.32. A more significant drop along C1 is observed for the shallow GELGAD, as illustrated in Figure 6.33.

A lower doping leads to a minor decrease in the peak field in the GL for both GELGADs and an increase of the depletion region for the shallow GELGAD, Figures 6.34 and 6.35. There is also a visible difference in the peak value and shape of the field compared to the case of higher doping, with the deep GELGAD showing better field uniformity in the GL compared to the shallow GELGAD.

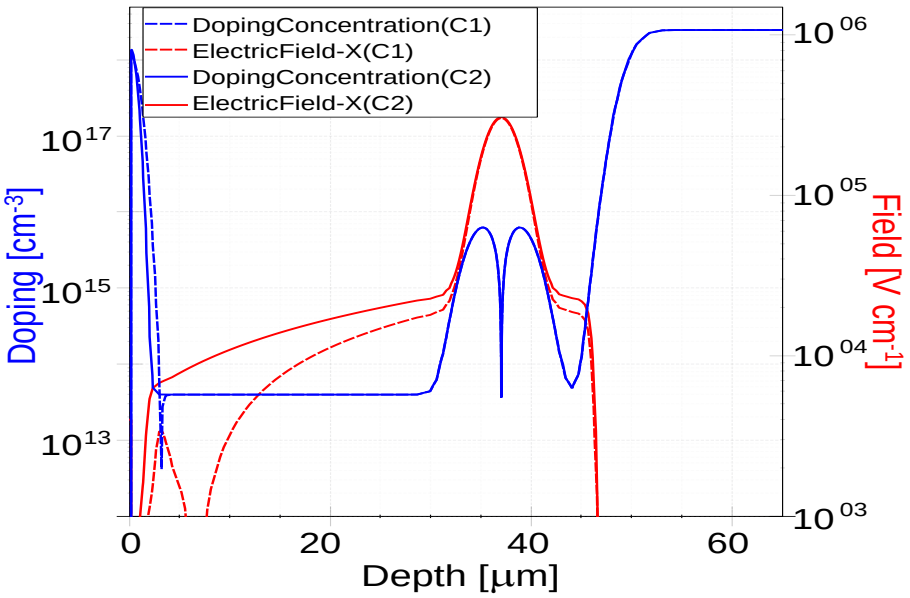


Fig. 6.32 *Doping concentration and electric field profile of the high doping deep GELGAD at a bias voltage of V_{b95} along the two cut lines C1 and C2.*

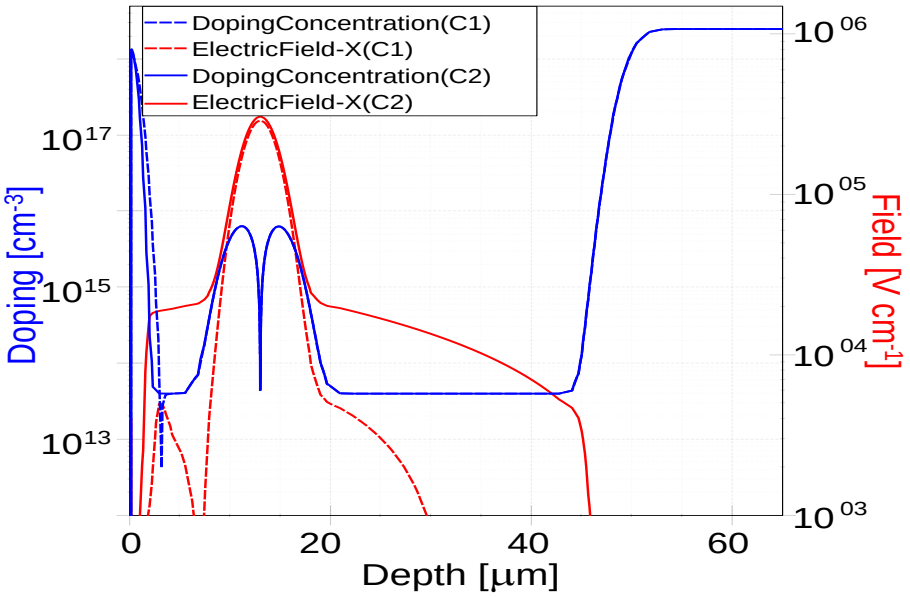


Fig. 6.33 *Doping concentration and electric field profile of the high doping shallow GELGAD at a bias voltage of V_{b95} along the two cut lines C1 and C2.*

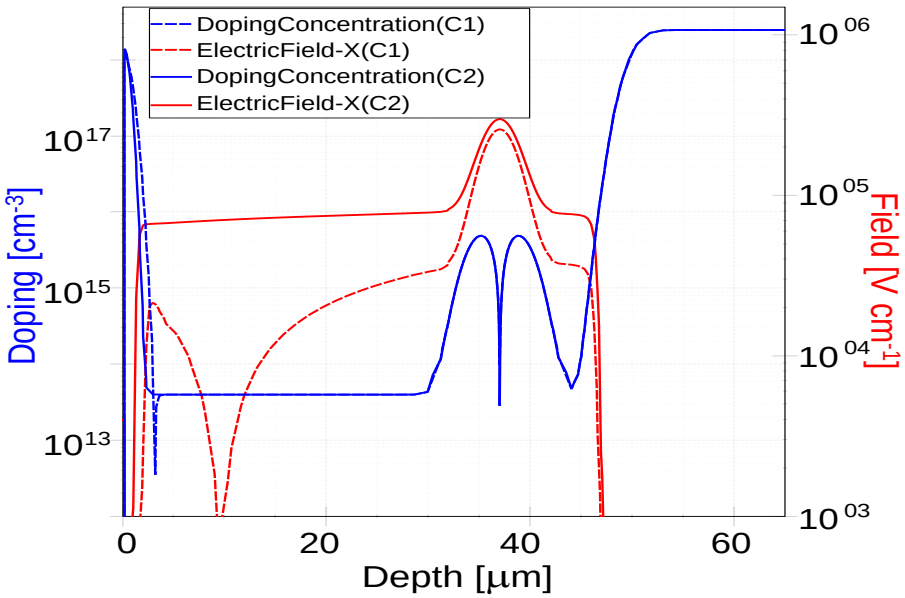


Fig. 6.34 *Doping concentration and electric field profile of the low doping deep GELGAD at a bias voltage of V_{b95} along the two cut lines C1 and C2.*

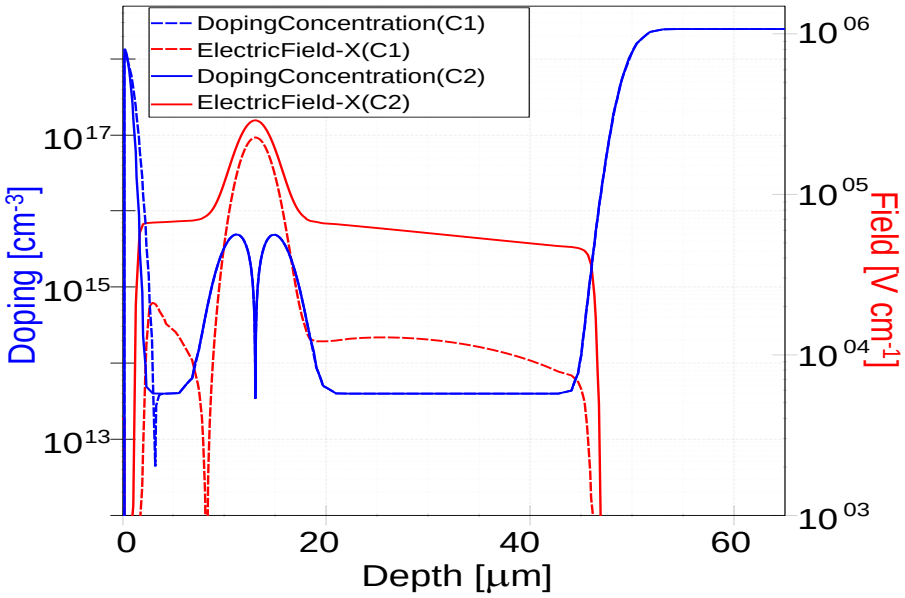


Fig. 6.35 *Doping concentration and electric field profile of the low doping shallow GELGAD at a bias voltage of V_{b95} along the two cut lines C1 and C2.*

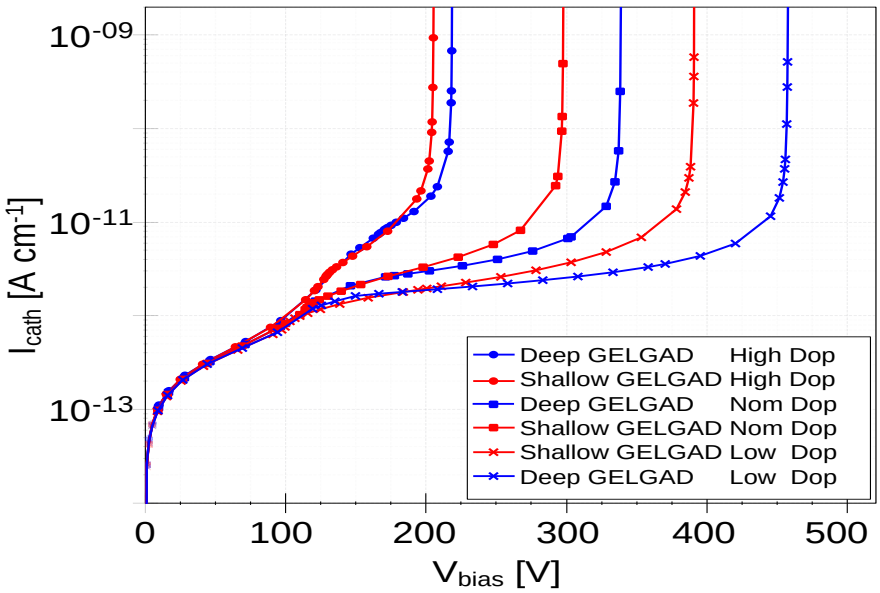


Fig. 6.36 *Current vs. voltage of the shallow and deep GELGADs for higher and lower doping.*

Figure 6.36 shows the IV plots for the nominal, lower and higher doping cases mentioned. Comparing these curves with those of Figure 6.11 it is seen that the lower doping considered here would shift the IV curves to the region of PIN breakdown, i.e. ≈ 400 V. The higher and nominal doping would however keep the curves significantly below this limit, suggesting that, in line with what was discussed in Chapter 4, a conservative GELGAD design would set the nominal as the maximum doping to be used in GELGAD fabrication. Owing to the novelty of the fabrication process and its inherent uncertainties, a more conservative design, which would include a range of doping of the epitaxial layers, would be necessary to guarantee a reasonable yield of fabricated working GELGAD devices.

Chapter 7

Conclusions and future work

7.1 Conclusions

The design simulation, fabrication and characterisation of LGAD detectors have been presented in this thesis.

In Chapter 1 the advantages of fast timing detectors in Particle Physics experiments, with a focus on HL-LHC applications, were discussed and, in particular, the ATLAS HGTD, based upon LGAD sensing technology, was introduced. The fundamental principles of semiconductor physics, related to their use as particle detectors, were discussed in Chapter 2, along with some aspects of device manufacturing. In Chapter 3 the timing detectors were introduced and their implementation using LGAD was discussed. Aspects of the physics of impact ionisation, upon which LGAD rely for their operation, were also presented in this chapter. The design and simulation of LGAD devices using Synopsys TCAD simulator was the subject of Chapter 4. A description of the layouts of the two runs of device fabrication was given, along with details of the simulation process, which included a simulation of manufacturing and electrical performances. Comparison of simulated and measured doping profiles obtained using SIMS techniques were shown and discussed. Optimization of the JTE profile, to improve breakdown, was shown and an estimate of the charge gain for LGADs was presented. In Chapter 5 the test results on the fabricated LGADs were presented and discussed. A description of the test setups used was given. Results of charge collection, gain and timing were shown. A novel technological solution, aiming at improving the fill factor of LGADs that could potentially be implemented in the design of higher granularity LGADs was introduced in Chapter 6. After a summary of the currently investigated solutions for high-granularity LGAD arrays, results obtained from TCAD simulations on this novel structure were shown and discussed.

As no previous experience on LGAD fabrication using the Teledyne e2v technology was available at the start of the project, particular attention was paid to the TCAD simulations in this thesis. To reduce as much as possible errors coming from incorrect knowledge of the doping profiles, TCAD simulations were compared with SIMS results made available by the foundry. It was found that MC simulations provided the best matching to SIMS results compared to analytical profiles when using ^{11}B implant.

Because LGADs rely for their operation on the impact ionisation process, the validity of the local ionisation model for the description of such effect was investigated. It was found that the validity condition expressed by equation (4.17), which requires the variation of the electric field dF/dx over the minimum mean free path between ionising events $\alpha(F_m)^{-1}$ to be smaller than the maximum field F_m, is verified for the LGAD devices described in this thesis. The spatial resolution of the mesh used in the device simulation was set to a fraction of the Debye length, as discussed in section 4.4 of Chapter 4. Similarly, the temporal resolution used in dynamic simulations was set to values of the order of the dielectric relaxation time, to avoid errors arising from the non-monotonicity of the solution variables, i.e. the LGAD output current. The simulations of charge collection and gain, which included infrared laser and MIP charge injection, demonstrated gain suppression for MIP charge injection.

The fabricated LGADs were tested using different setups. Their IV and CV characteristics were measured using a probe station. The leakage current for Run 1 devices of type 3 was found to be about 10 nA·cm^{-2} at full depletion. The typical capacitance was around 2.5 pF. The infrared laser test setup operating at 1064 nm used for gain and timing measurement was designed, assembled and tested to confirm it had the required characteristics to evaluate LGAD performances. The laser was calibrated to inject approximately 0.7 fC of charge in a PIN reference diode, to emulate a MIP charge release. The RO included two-stage amplifiers and a fast oscilloscope operating at 128 GS/sec. The first stage amplifier was designed to provide the required bandwidth, slew rate and jitter to properly evaluate LGADs. Test results on the Run 1 LGAD devices showed a maximum gain of approximately 60 when operating near breakdown and a residual jitter from the laser charge injection test of the order of 10 ps, regardless of the LGAD flavour tested. A comparison with TCAD simulations showed good matching with the gain measurements up to approximately 80% of the breakdown voltage. The normalized plots of the measured gain showed similar trends, with differences among LGAD flavours that could be due to the variability among individual devices. Timing tests performed using a ^{90}Sr source showed a time resolution of approximately 40 $\pm$ 2 ps. Initial tests were done on the LGAD array from Run 2 and a 2D breakdown voltage map was obtained. Preliminary tests done on Run 1 LGADs demonstrated their potential use for the detection and dosimetry of low LET particles.

The novel process for LGAD device fabrication proposed in section 6.3 of Chapter 6 consists of a buried gain layer that could improve the fill factor and homogeneity of LGAD response. As it relies on epitaxial growth of silicon to create the gain layer it would avoid issues related to defects creation from implantation and considerably simplify the fabrication process and related costs of manufacturing. The TCAD simulations performed included the full process of fabrication, based upon information provided by a commercial silicon foundry, and showed a predicted gain of the order of 6-7 using the existing LGAD layout developed for Run 2. The proposed process could potentially be used for the planar fabrication of smaller cells, for the design of high granularity LGADs.

7.2 Future works

Some of the fabricated LGADs from Run 1 have been already irradiated with neutrons and protons and will be tested to evaluate their gain and timing performances following irradiation. The tests will be conducted at temperatures of -20 °C or below using the developed test setups for this project. Investigation of their degradation due to the gain layer reduction following boron inactivation will be performed and results compared with what is already available in published literature.

The LGAD array from Run 2 will require a more advanced setup for its full characterization besides the initial IV and CV maps shown in section 5.4 of Chapter 5 of this thesis. Currently, the development of a custom ASIC using a TSMC 28 nm CMOS process targeting the LGAD RO is ongoing. The TSMC 28 nm process is one of the last true CMOS nodes before more expensive and exotic technologies such as FinFETs [138] become necessary. The 28 nm technology offers a smaller feature size than the 65 nm node enabling a factor of 4 to 5 higher circuit density. It achieves a factor of two higher speeds and lower power consumption. In addition, this technology has also been found to provide better Total Ionising Dose (TID) radiation performance than the 65 nm node, with radiation hardness improving with shorter transistor lengths [139]. This project, funded by STFC under the framework of the Centre for Instrumentation (CfI) programme, consists of two runs of device fabrication. The first run, the design of which was completed in September 2023, includes a dual channel transimpedance amplifier, a variable threshold comparator and digitized Current Mode Logic (CML) output integrated on a 1×1 mm^2. The second run will include a four-channel amplifier with a Time to Digital Converter (TDC). The first Run 1 ASICs were delivered to RAL in January 2024. The LGAD array will be bump bonded to a custom PCB, already designed and currently being manufactured, which will act as an interface between the array and the ASIC. The tests will be performed at RAL and OPMD using the setups already

described in section 5.3 of Chapter 5 of this thesis. An irradiation campaign to evaluate the degradation of performances of the LGAD array and to test the radiation hardness of the developed ASICs is planned. The submission of the second run is envisaged for October 2024.

As mentioned in Chapter 5, the LGAD technology is now being investigated for its potential benefits for a range of other scientific applications besides Particle Physics. Uses in the detection of Low LET particles and radiation dosimetry will continue in collaboration with the Centre of Medical Radiation Physics (CMRP). An additional CfI-funded project, currently ongoing, seeks to investigate the use of LGADs for X-ray science. Results already published [140] demonstrated interesting applications for LGADs in photon science. In this project, which consists of two fabrication runs, multi-project LGAD wafers will be produced in collaboration with Micron Semiconductor [141], to investigate the uses of LGADs as X-ray detectors. The first run will have a selection of devices and use the established technology employing junction termination extensions (JTE) and p-spray isolation between pixels, at the expense of a non-gain area between pixels. The second run will have no doping implanted at Micron to form the pixel isolation but will use a trench isolation method that was recently successfully demonstrated at the Scottish Microelectronics Centre (SMC). This solution will yield high fill factor pixel devices and push the technology to higher granularity. The ultimate goal is to combine such a trench isolation technique with the proposed method of buried gain layer fabrication described in Chapter 6 and to demonstrate that such a solution could eventually be employed for the fabrication of monolithic LGADs.

Besides the novel design and technology solutions described above, the simulation results presented in Chapter 4 suggest that further investigation on the TCAD modelling of the impact ionisation process is necessary. The results obtained in this thesis showed a good matching between predicted and measured LGADs gain up to around 80% of the breakdown voltage but differ considerably at higher bias voltages. Also, the prediction of LGADs' absolute breakdown voltages differs from measurements. Assuming the validity of the local model for impact ionisation, as discussed in section 4.5 of Chapter 4, it is believed that better modelling might come from optimizing the ionisation coefficients of the models. The Okuto-Crowell and University of Bologna models used in this thesis did not predict substantially different results but, as mentioned in Chapter 3, implementation of more sophisticated models, like the Baraff's ionisation model, using a Physical Model Interface (PMI), a TCAD software features that allows implementation of custom physical models, might improve the accuracy of the simulations. Further investigations are needed to improve the modelling of the impact ionisation processes. Moreover, additional work, both theoretical and experimental, will be needed to implement effects like gain reduction resulting from radiation damage. An

additional interesting simulation topic is the prediction of the timing resolution of LGAD devices, besides the gain. This would require the dynamic modelling of LGAD internal noise during the process of charge collection, which would include effects due to thermal, shot and excess noise factor and, in case of the signal generated by a charged particle, the implementation of Landau fluctuations during the process of charge generation.

References

[1] H. Spieler, *Semiconductor Detector Systems*, Oxford University Press, 2005.

[2] F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*, Springer, 2009.

[3] https://library.oapen.org/bitstream/handle/20.500.12657/50882/1/9781420083002.pdf.

[4] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214, 2012.

[5] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235, 2012.

[6] G. Apollinari, A. I. Béjar, O. Brüning, P. Fessia, M. Lamont, L. Rossi, L. Tavian, *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, URL: http://cds.cern.ch/record/2284929, 2017.

[7] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 S08003, doi:10.1088/1748-0221/3/08/S08003, 2008.

[8] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, Technical Report CERN-LHCC-2017-005, ATLAS-TDR-025, 2017.

[9] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, tech. rep. CERN-LHCC-2017-021, CERN, 2018, URL: https://cds.cern.ch/record/2285585.

[10] ATLAS Collaboration, *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, tech. rep. CERN-LHCC-2020-007, CERN, 2020, URL: https://cds.cern.ch/record/2719855.

[11] C. de La Taille, S. Callier, S. C. Di Lorenzo et al., *ALTIROC0, a 20 pico-second time resolution ASIC for the ATLAS High Granularity Timing Detector (HGTD)*, PoS, vol. TWEPP-17, p. 006, 2018.

[12] G. Pellegrini, P. Fernández-Martínez, M. Baselga et al., *Technology developments and first measurements of low gain avalanche detectors (LGAD) for high energy physics applications*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 765, p. 12 – 16, 2014.

[13] RD50 collaboration, URL: https://rd50.web.cern.ch/rd50.

[14] S. Sze, *Physics of Semiconductor Devices*, New York: Wiley, 2nd edition, 1981.

[15] https://cleanroom.groups.et.byu.net/EW_orientation.phtml.

[16] G. Calefato, D. Creanza, M. de Palma et al., *A comparison on radiation tolerance of <100> and <111> silicon substrates of microstrip detectors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 476, Issue 3, p. 744-750, 2002.

[17] D. Vasileska, S. M. Goodnick, G. Klimeck, *Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation*, CRC press, ISBN-13: 978-1-4200-6484-1, 2010.

[18] G. Lutz, *Semiconductor radiation detectors: device physics*, Springer, 1999.

[19] C. Kaimao, W. Zhongan, *Deep levels related to copper in silicon*, J. of Electron. (China) 5, 285–293, https://doi.org/10.1007/BF02778709, 1988.

[20] S.S. Li, *Semiconductor Physical Electronics*, 2nd edition, ISBN 13: 978-0387-28893-2, 2006.

[21] W. Van Roosbroeck, *Theory of flow of electrons and holes in germanium and other semiconductors*, Bell System Technical Journal, vol. 29, p. 560-607, 1950.

[22] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag Wien New York, 1984.

[23] T. M. Apostol, *Calculus, Vol. II: Multi-Variable Calculus and Linear Algebra*, (Blaisdell, Waltham, MA), Chap. 1, 1969.

[24] H. K. Gummel, *A self-consistent iterative scheme for one-dimensional steady state transistor calculation*, IEEE Trans. Electron. Devices, 11, p. 455–465, 1964.

[25] D. Vasileska, *Drift-Diffusion Model, Part B: Solution Details*, https://nanohub.org/resources/1512, 2006.

[26] K. Blotekjaer, *Transport Equations for Electrons in Two-Valley Semiconductors*, IEEE Transactions on Electron Devices, 17(1): p. 38–47, 1970.

[27] C. Jungemann, T. Grasser, B. Neinhüs, and B. Meinerzhagen, *Failure of Moments-Based Transport Models in Nanoscale Devices Near Equilibrium*, IEEE Transactions on Electron Devices, 52: p. 2404–2408, 2005.

[28] K. Hess, *Advanced Theory of Semiconductor Devices*, Wiley-IEEE Press, ISBN: 9780470544105, 2000.

[29] C. Grupen, *Particle Detectors*, Cambridge University Press, 2nd edition, ISBN: 978-0521187954, 2011.

[30] W. Leo, *Techniques for Nuclear and Particle Physics Experiments, A How-to Approach*, 2nd edition, 1994.

[31] S. Ramo, *Currents Induced by Electron Motion*. Proceedings of the IRE. 27 (9): p. 584–585, 1939.

[32] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001, 2018.

[33] S. Meroli, D. Passeri and L. Servoli, *Energy loss measurement for charged particles in very thin silicon layers*, J.Instrum., 6:P06013, 2011.

[34] F. Scholze et al., *Determination of the electron–hole pair creation energy for semiconductors from the spectral responsivity of photodiodes*, Nucl. Instrum. Methods A439, 208, 2000.

[35] F. Berghmans, B. Brichard, A. Fernandez et al, *An Introduction to Radiation Effects on Optical Components and Fiber Optic Sensors*, 10.1007/978-1-4020-6952-9_6, 2007.

[36] J. Kemmer, *Fabrication of low noise silicon radiation detectors by the planar process*, Nuclear Instruments and Methods, Volume 169, Issue 3, p. 499-502, ISSN 0029-554X, https://doi.org/10.1016/0029-554X(80)90948-9, 1980.

[37] https://www.hitachi-hightech.com/global/en/knowledge/semiconductor/room/about/silicon.html.

[38] https://smithsonianchips.si.edu/ice/cd/CEICM/SECTION7.pdf.

[39] K. Kakimoto, *Development of Crystal Growth Technique of Silicon by the Czochralski Method*, Vol. 124 ACTA PHYSICA POLONICA A, http://dx.doi.org/10.12693/APhysPolA.124.227, 2013.

[40] T.F. Ciszek, T.H. Wang, *Silicon defect and impurity studies using float-zone crystal growth as a tool*, Journal of Crystal Growth, 237, p. 1685-1691, 2002.

[41] N. Teranishi, G. Fuse, M.A. Sugitani, *Review of Ion Implantation Technology for Image Sensors*, Sensors (Basel). 20;18(7):2358. doi: 10.3390/s18072358. PMID: 30036992; PMCID: PMC6068500, 2018.

[42] Justin K. Wang, Dean R. Denison, *Advanced techniques for interlayer dielectric deposition and planarization*, Proc. SPIE 2090, Multilevel Interconnection: Issues That Impact Competitiveness; https://doi.org/10.1117/12.156535, 1993.

[43] S.B. Inayat, A.P. Nayak, V.J. Logeeswaran et al., *Dry Etching Processes*, Bhushan, B. (eds) Encyclopedia of Nanotechnology. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6178-0_353-2.

[44] https://www.axcelis.com/wp-content/uploads/2019/02/Ion_Implantation_in_Silicon_Technology.pdf.

[45] C. K. Hu and J. M. E. Harper, *Copper Interconnections and Reliability*, Mater. Chem. Phys., vol. 52, p. 5-16, 1998.

[46] H. F.-W. Sadrozinski, A. Seiden, N. Cartiglia, *4-Dimensional Tracking with Ultra-Fast Silicon Detectors*, https://doi.org/10.48550/arXiv.1704.08666, 2017.

[47] M. Ferrero, R. Arcidiacono, M. Mandurrino et al., *An Introduction to Ultra-Fast Silicon Detectors*, https://doi.org/10.1201/9781003131946, 2021.

[48] S. Garbolino, S. Martoiu, A. Rivetti et al., *Implementation of Constant-Fraction-Discriminators (CFD) in sub-micron CMOS technologies*, DOI:10.1109/NSSMIC.2011.6154364, 2011.

[49] S. Ramo, *Currents Induced by Electron Motion*. Proceedings of the IRE. 27 (9): p. 584–585, 1939.

[50] C. Canali, C. Jacoboni, F. Nava et al., *Electron drift velocity in silicon*, https://doi.org/10.1103/PhysRevB.12.2265, 1975.

[51] C. Scharf, R. Klanner, *Measurement of the drift velocities of electrons and holes in high-ohmic <100> silicon*, https://doi.org/10.1016/j.nima.2015.07.057, 2015.

[52] W.T. Tsang, *Semiconductors and semimetals*, Volume 22m Lightwave communication technology, A.P. Inc, ISBN 0-12-752153-4, 1985.

[53] P.A. Wolf, *Theory of electron multiplication in Silicon and Germanium*, Phys.Rev., 95, 6, p.1415-1420, 1954.

[54] Y. Okuto and C. R. Crowell, *Ionization coefficients in semiconductors: A nonlocalized property*, Phys. Rev. B 10, 4284, https://doi.org/10.1103/PhysRevB.10.4284, 1974.

[55] K. K. Thornber, *Applications of scaling to problems in high-field electronic transport*, Journal of Applied Physics 52, 279–290, https://doi.org/10.1063/1.328490, 1981.

[56] W. Shockley, *Problems related to p-n junctions in silicon*, Solid State Electronics, Volume 2, Issue 1, p. 35,IN9,61-60,IN10,67, 1961.

[57] G. A. Baraff, *Distribution functions and ionization rates for hot electrons in semiconductors*, Physical Review, vol. 128, no. 6, p. 2507-2517, 1962.

[58] A. Sutherland, *An improved empirical fit to Baraff's universal curves for the ionization coefficients of electron and hole multiplication in semiconductors*, IEEE Transactions on Electron Devices, vol. 27, no. 7, p. 1299-1300, 1980.

[59] https://www.synopsys.com/silicon/tcad.html.

[60] P. P. Webb and R. J. McIntyre, *Large area reach-through avalanche diodes for x-ray spectroscopy*, IEEE Transactions on Nuclear Science, vol. 23, p. 138–144, Feb 1976.

[61] A. Akindinov, A. Martemianov, P. Polozov et al., *New results on mrs APDs*, Nuclear Instruments and Methods in Physics Research Section A: Accel- erators, Spectrometers, Detectors and Associated Equipment, vol. 387, no. 1, p. 231 – 234, 1997.

[62] C. Piemonte et al., *Performance of NUV-HD Silicon Photomultiplier Technology*, IEEE T.Electron Dev., 63(3):1111, 2016.

[63] M. Ferrero et al., *Study and characterization of low gain avalanche diode*, In TREDI2017: 12th "Trento" Workshop on Advanced Silicon Radiation Detectors, 2017.

[64] H. Richter, L. Andricek, T. Gebhart et al., *Strip detector design for ATLAS and HERA-B using two-dimensional device simulation*, Nucl.Instrum.Meth.A 377, p. 412-421, 1996.

[65] L. Evensen, A. Hanneborg, B. Sundby Avset et al., *Guard ring design for high voltage operation of silicon detectors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 337, Issue 1, p. 44-52, 1993.

[66] P. Saxena, K. Ranjan, A. Bhardwaj et al., *Development of multi-guard ring-equipped p+–n Si microstrip sensors for the SiD detector at the ILC*, Semicond. Sci. Technol. 25 105012, 2010.

[67] A. van Der Ziel, *Noise:Sources,Characterization, Measurements*, Englewood Cliffs, NJ, Prentice Hall, 1973.

[68] Sh. Kogan, *Electronics noise and fluctuations in solids*, Cambridge University Press, 1996.

[69] G-F. Dalla Betta et al., *Design and TCAD simulation of double-sided pixelated low gain avalanche detectors*, Nucl. Inst.Meth.A, 796:154, 2015.

[70] R.R. Spencer and M.S. Ghausi, *Introduction to electronic circuit design*, Upper Saddle River NJ: Prentice Hall/Pearson Education, Inc. p. 533. ISBN 0-201-36183-3, 2003.

[71] A. Rivetti, *CMOS: Front-End Electronics for Radiation Sensors*, DOI: 10.1201/b18599, 2015.

[72] Hochbaum, Dorit S. et al., *Rating Customers According to Their Promptness to Adopt New Products.*, Operations Research, vol. 59, no. 5, p. 1171–83, 2011.

[73] https://www.tcl.tk/about/language.html.

[74] https://www.teledyne-e2v.com.

[75] https://www.synopsys.com/silicon/tcad/process-simulation/sentaurus-process.html.

[76] https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html.

[77] R. F. K. Herzog and F. P. Viehböck, *Ion source for mass spectrography*, Phys. Rev., vol. 76, p. 855–856, Sep 1949.

[78] T. Clarysse, D. Vanhaeren, I. Hoflijk, W. Vandervorst, *Characterization of electrically active dopant profiles with the spreading resistance probe*, Materials Science and Engineering: R: Reports, Volume 47, Issues 5–6, 2004.

[79] C. W. Magee, D. Jacobson and H.J. Gossmann, *Accuracy of SIMS in determining ion implanted B doses as confirmed by nuclear reaction analysis*, 1998 International Conference on Ion Implantation Technology. Proceedings (Cat. No.98EX144), Kyoto, Japan, p. 42-45 vol.1, doi: 10.1109/IIT.1999.812047, 1999.

[80] M. Levinshtein, J. Kostamovaara, S. Vainshtein, *Breakdown Phenomena in Semiconductors and Semiconductor Devices*, https://doi.org/10.1142/5877, September 2005.

[81] F. Capasso, *Chapter 1: Physics of Avalanche Photodiode*, Editor(s): W.T. Tsang, Semiconductors and Semimetals, Elsevier, Volume 22, Part D, p. 1-172, 1985.

[82] M. Moll et al., *TCAD simulation of ionization in non-irradiated LGADs*, 40th RD50 workshop https://indico.cern.ch/event/1157463/, 2022.

[83] G. J. Rees1, J. P. R. David, *Nonlocal impact ionization and avalanche multiplication*, J. Phys. D: Appl. Phys. 43 243001, DOI: 10.1088/0022-3727/43/24/243001, 2010.

[84] R. van Overstraeten and H. de Man, *Measurement of the Ionization Rates in Diffused Silicon p-n Junctions*, Solid-State Electronics, vol. 13, no. 1, p. 583–608, 1970.

[85] Y. Okuto and C. R. Crowell, *Threshold Energy Effect on Avalanche Breakdown Voltage in Semiconductor Junctions*, Solid-State Electronics, vol. 18, no. 2, p. 161–168, 1975.

[86] T. Lackner, *Avalanche Multiplication in Semiconductors: A Modification of Chynoweth's Law*, Solid-State Electronics, vol. 34, no. 1, p. 33–42, 1991.

[87] M. C. Vecchi and M. Rudan, *Modeling Electron and Hole Transport with Full-Band Structure Effects by Means of the Spherical-Harmonics Expansion of the BTE*, IEEE Transactions on Electron Devices, vol. 45, no. 1, p. 230–238, 1998.

[88] E. Gnani et al., *Extraction method for the impact-ionization multiplication factor in silicon at large operating temperatures*, in Proceedings of the 32nd European Solid-State Device Research Conference (ESSDERC), Florence, Italy, p. 227–230, September 2002.

[89] T. Hatakeyama et al., *Physical Modeling and Scaling Properties of 4H-SiC Power Devices*," in International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Tokyo, Japan, p. 171–174, September 2005.

[90] W. Shockley, *Problems related to p-n junctions in silicon*, Solid State Electronics, Volume 2, Issue 1, p. 35,IN9,61-60,IN10,67, Jan. 1961.

[91] D. B. M. Klaassen, *A Unified Mobility Model for Device Simulation—I. Model Equations and Concentration Dependence*, Solid-State Electronics, vol. 35, no. 7, p. 953–959, 1992.

[92] A. Matthiessen, *On the Electric Conducting Power of the Metals*, Philosophical Transactions of the Royal Society of London. 148: 383–387, doi:10.1098/rstl.1858.0020, 1858.

[93] C. Canali et al., *Electron and Hole Drift Velocity Measurements in Silicon and Their Empirical Relation to Electric Field and Temperature*, IEEE Transactions on Electron Devices, vol. ED-22, no. 11, p. 1045–1047, 1975.

[94] M. S. Tyagi and R. Van Overstraeten, *Minority Carrier Recombination in Heavily-Doped Silicon*, Solid-State Electronics, vol. 26, no. 6, p. 577–597, 1983.

[95] G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, *A New Recombination Model for Device Simulation Including Tunneling*, IEEE Transactions on Electron Devices, vol. 39, no. 2, p. 331–338, 1992.

[96] F. Moscatelli et al., *Measurements and TCAD simulations of bulk and surface radiation damage effects in silicon detectors*, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), San Diego, CA, USA, 2015, p. 1-6, doi: 10.1109/NSS-MIC.2015.7581944, 2015.

[97] A. Beer, *Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten [Determination of the absorption of red light in colored liquids]*, Annalen der Physik und Chemie (in German), doi:10.1002/andp.18521620505, 1852.

[98] B. Soumyadipta, L. Bong Jae, Z. Zhang, *Infrared Radiative Properties of Heavily Doped Silicon at Room Temperature*, Journal of Heat Transfer-transactions of The Asme - J HEAT TRANSFER. 132. 10.1115/1.4000171, 2010.

[99] E. Curras, M. Fernandez, M. Moll, *Gain suppression mechanism observed in Low Gain Avalanche Detectors*, arXiv:2107.10022v1 [physics.ins-det] 21 Jul 2021.

[100] M. Carmen Jiménez-Ramos, J. García López, A. García Osuna et al., *Study of Ionization Charge Density-Induced Gain Suppression in LGADs*, Sensors 22, 1080. https://doi.org/10.3390/s22031080, 2022.

[101] S. Agostinelli, J. Allison, K. Amako et al., *GEANT4—a simulation toolkit*. In: Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506.3, p. 250–303, 2003.

[102] A. Okabe, B. Boots, and K. Sugihara, *Spatial Tessellations*, John Wiley and Sons Ltd, 1992.

[103] E. Cartier, M. Fischetti, E. A. Eklund et al., *Impact ionization in silicon*, Appl. Phys. Lett. 21; 62 (25): 3339–3341. https://doi.org/10.1063/1.109064, June 1993.

[104] CMRP, https://www.uow.edu.au/engineering-information-sciences/research/centre-medical-radiation-physics/.

[105] https://www.aptechnologies.co.uk/products/laser-diodes/singlemode-laser-diodes/1020-1180nm-dfb.

[106] S. Garbolino et al., *Implementation of Constant-Fraction-Discriminators (CFD) in sub-micron CMOS technologies*, DOI:10.1109/NSSMIC.2011.6154364, 2011.

[107] S. Meroli, D. Passeri, and L. Servoli., *Energy loss measurement for charged particles in very thin silicon layers*, JINST 6(06):P06013, 2011.

[108] C. Agapopoulou et al, *Performance in beam tests of irradiated Low Gain Avalanche Detectors for the ATLAS High Granularity Timing Detector*, JINST 17 P09026, 2022.

[109] https://www.femto.de/en/products/ghz-wideband-amplifiers/variable-gain-1-2-ghz-dupva.html.

[110] M. Ferrero, R. Arcidiacono, M. Mandurrino, V. Sola, N. Cartiglia, *An Introduction to Ultra-Fast Silicon Detectors*, ISBN: 978-0-367-64629-5, 2021.

[111] R. Heller, *Studies of LGAD mortality using the Fermilab Test Beam*, https://indico.cern.ch/event/1029124/contributions/4411279/.

[112] A. Rosenfeld, *Novel detectors for silicon based microdosimetry, their concepts and applications*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 809, Advances in detectors and applications for medicine, p. 156–170, 2016.

[113] J. C. Chancellor, G.B. Scott, and J. P. Sutton., *Space radiation: the number one risk to astronaut health beyond low earth orbit*, Life 4.3, p. 491–510, 2014.

[114] A. Wroe, A. Rosenfeld, and R. Schulte, *Out-of-field dose equivalents delivered by proton therapy of prostate cancer*, Medical physics34.9, p. 3449–3456, 2007.

[115] M. J. Berger, J. S. Coursey, M. A. Zucker et al., *Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions*, NIST Standard Reference Database 124, NIST 2017.

[116] M. Carmen Jimenez-Ramos, Javier Garcia Lopez, Adrian Garcia Osuna et al., *Study of ionization charge density-induced gain suppression in LGADs*, Sensors 22.3, p. 1080, 2022.

[117] A. Crnjac, M. Jakšić, M. Rodriguez-Ramos et al., *Spatiotemporal dynamics of charge transport in LGAD probed with focused ion beams*, Paper presented at the 45th RD50 Worshop - Sevilla Spain. 2022.

[118] M. Campbell, E. Heijne, T. Holý et al., *Study of the charge sharing in a silicon pixel detector by means of α particles interacting with a Medipix2 device*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 591.1, Radiation Imaging Detectors 2007, p. 38–41, 2008.

[119] G. Paternoster et al., *Developments and first measurements of Ultra-Fast Silicon Detectors produced at FBK*, Journal of Instrumentation 12.02, p. C02077–C0207, 2017.

[120] M. Carulla, A. Doblas, D. Flores et al.,*50 μm thin Low Gain Avalanche Detectors (LGAD) for timing applications*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 924, p. 373-37, 2019.

[121] N. Cartiglia, A. Staiano, V. Sola et al., *Beam test results of a 16ps timing system based on ultra-fast silicon detectors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 850, p. 83 - 88, 2017.

[122] Z. Galloway, C. Gee, S. Mazza, H. Ohldag, R. Rodriguez, H.-W. Sadrozinski et al.,*Use of LGAD ultra-fast silicon detectors for time-resolved low-kev x-ray science*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 923, p. 5-7, 2019.

[123] CMS collaboration, *A MIP Timing Detector for the CMS Phase-2 Upgrade*, Tech. Rep. CERN-LHCC-2019-003. CMS-TDR-020, CERN, Geneva, Mar, 2019.

[124] H. F.-W. Sadrozinski, A. Seiden and N. Cartiglia, *4d tracking with ultra-fast silicon detectors*, Reports on Progress in Physics 81, 026101, 2017.

[125] D. Dutta, *The LHCb VELO Upgrade*, PoS, vol. VERTEX2018, p. 022, 2019.

[126] Z. Wang, *On the single photon counting (spc) modes of imaging using an xfel source*, JINST 10 C12013, 2015.

[127] N. Cartiglia, *Issues in the design of Ultrafast silicon detectors*, TREDI2015, Trento, Italy . https://indico.cern.ch/event/351695/contributions/828366/, 2015.

[128] G. Paternoster, G. Borghi, R. Arcidiacono et al., *Novel strategies for fine-segmented Low Gain Avalanche Diodes*, Nuclear Inst. and Methods in Physics Research, A 987 164840, 2021.

[129] A. Apresyan, W. Chen, G. D'Amen et al., *Measurements of an AC-LGAD strip sensor with a 120 GeV proton beam*, JINST 15 P09038, [2006.01999], 2020.

[130] G. Giacomini, W. Chen, G. D'Amen, A. Tricoli, *Fabrication and performance of AC-coupled LGADs*, https://doi.org/10.1088/1748-0221/14/09/P09004, 2019.

[131] E. Curras, M. Carulla, M. Centis Vignali et al., *Inverse low gain avalanche detectors (ilgads) for precise tracking and timing applications*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 958 16254, 2020.

[132] G. Paternoster, G. Borghi, M. Boscardin et al., *Trench-isolated low gain avalanche diodes (ti-lgads)*, IEEE Electron Device Letters 41 884-887, 2020.

[133] Y. Zhao, S. Ayyoub, W. Chen et al., *A new approach to achieving high granularity for silicon diode detectors with impact ionization gain*, J. Phys. Conf. Ser. 2374, 012171, 2022.

[134] A. Apresyan et al, *Buried Layer Low Gain Avalanche Diodes*, J. Phys.: Conf. Ser. 2374 012166, 2022.

[135] S. M. Mazza et al. *Tuning of gain layer doping concentration and Carbon implantation effect on deep gain layer*, J. Phys. Conf. Ser. 2374.1, p. 012173. doi: 10.1088/1742-6596/2374/1/012173. arXiv: 2201.08933 [physics.ins-det], 2022.

[136] https://www.plutosemi.net/.

[137] H. Zhong, *Review of the Shockley–Ramo theorem and its application in semiconductor gamma-ray detectors*, Nuclear Instruments and Methods in Physics Research A 463 250–267, 2001.

[138] T.J.K Liu, *FinFET: History, Fundamentals and Future*, University of California, Berkeley. Symposium on VLSI Technology Short Course (June 11, 2012).

[139] C. Zhang, et al. *Bias dependence of total ionizing dose effects on 28-nm bulk MOSFETs*, 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC). IEEE, 2018.

[140] M. Andra, et al., *Development of low-energy X-ray detectors using LGAD sensors*, J. Synchroton Rad., 26, 1226-1237. https://doi.org/10.1107/S1600577519005393, 2019.

[141] http://www.micronsemiconductor.co.uk/.

Appendix A

Impact Ionisation Models

Impact ionisation models with parameter values used in the simulations are reported here.

- **University of Bologna**

  The model reads:

$$\alpha_v(F) = \frac{F}{a(T) + b(T)e^{\frac{d(T)}{F+c(T)}}} \tag{A.1}$$

  with temperature-dependent coefficients:

$$a(T) = a_0 + a_1 t^{a_2} \qquad b(T) = b_0 \qquad c(T) = c_0 + c_1 t + c_2 t^2 \tag{A.2}$$

  and

$$d(T) = \frac{\beta_v E_G}{e\lambda d_0}(d_0 + d_1 + d_2 t^2) \tag{A.3}$$

  with E_G the band gap and $t = T$ in K.

The default parameter values are:

Table A.1 *University of Bologna Impact Ionisation parameters for silicon.*

Symbol	Parameter name	Default value: Electrons	Default value: Holes	Unit
a_0	ha0	4.3383	2.376	V
a_1	ha1	-2.42×10^{-12}	1.033×10^{-2}	V
a_2	ha2	4.1233	0	1
b_0	hb0	0.235	0.17714	V
b_1	hb1	0	-2.178×10^{-3}	V
c_0	hc0	1.6831×10^{4}	9.47×10^{-3}	$V \cdot cm^{-1}$
c_1	hc1	4.3796	2.4924	$V \cdot cm^{-1}$
c_2	hc2	0.13005	0	$V \cdot cm^{-1}$
d_0	hd0	1.2337×10^{6}	1.4043×10^{6}	$V \cdot cm^{-1}$
d_1	hd1	1.2039×10^{3}	2.9744×10^{3}	$V \cdot cm^{3}$
d_2	hd2	0.56703	1.4829	$V \cdot cm^{-1}$
λ	lambda	62×10^{-8}	45×10^{-8}	cm
β	beta	0.680414	0.56214	1

- **Okuto-Crowell**

  The model reads:

$$\alpha_\nu(F) = a(1 + c(T - T_0))F^\gamma e^{-(\frac{b(1+d(T-T_0))}{F})^\delta} \tag{A.4}$$

  with

$$b_\nu = \frac{\beta_\nu E_G}{e\lambda} \tag{A.5}$$

  where T_0 = 300 K, E_G is the energy bandgap, λ is the optical-phonon mean free path for the carrier and β_ν is a constant factor, (ν = electrons/holes).

The default parameter values are:

Table A.2 *Okuto-Crowell Impact Ionisation parameters for silicon.*

Symbol	Parameter name	Default value: Electrons	Default value: Holes	Unit
a	a	0.426	0.243	V^{-1}
b	b	4.81×10^{5}	6.53×10^{5}	$V \cdot cm^{-1}$
c	c	3.05×10^{-4}	5.35×10^{-4}	K^{-1}
d	d	6.86×10^{-4}	5.67×10^{-4}	K^{-1}
γ	gamma	1	1	1
δ	delta	2	2	1
λ	lambda	6.82×10^{-8}	45×10^{-8}	cm
β	beta	0.265283	0.261395	1

Appendix B

List of Publications and Presentations

- E. Giulio Villani, P. Allport, K. Ball, R. Bell, D. Bortoletto, M. Gazi, L. Gonella, D. Hynds, D. Jordan, I. Kopsalis, S. McMahon, J. Mulvey, R. Plackett, D. Weatherill, *Design and development of Low Gain Avalanche Detectors using Teledyne e2v process*, JINST 19 C01038 doi: 10.1088/1748-0221/19/01/C01038, 2024.
- M. Gazi, P. Allport, D. Bortoletto, L. Gonella, D. Hynds, D. Jordan, I. Kopsalis, S. McMahon, J. Mulvey, R. Plackett, K. Stefanov, E.G. Villani, *Low Gain Avalanche Detector development for the LHC's High-Luminosity upgrade at Teledyne e2v*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 1037, 166876, ISSN 0168-9002, https://doi.org/10.1016/j.nima.2022.166876, 2022.
- J. W. Archer, E. G. Villani, A. Rosenfeld et al., *A Two-Dimensional Characterization of Low-Gain Avalanche Diodes for Low-LET Microdosimetry*, IEEE Transactions on Nuclear Science, vol. 71, no. 3, pp. 342-351, doi: 10.1109/TNS.2024.3359249, 2024.
- Presentation at 16th Topical Seminar on Innovative Particle and Radiation Detectors (IPRD23) 25-29 September 2023, Siena, Italy.