The University of Manchester

SIMULATION OF 3D SENSORS

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2013

By
Ching-Hung Lai
School of Physics and Astronomy

Contents

List of Tables

List of Figures

Abstract

The Large Hadron Collider (LHC) at CERN has the highest energy and luminosity in the world. Radiation hardness is then a critical requirement for the inner tracker design. The inner tracker is important for identifying heavy quarks using high spatial precision detectors. Silicon detectors are now the primary technology for this application.

3D silicon sensors use a novel technology with penetrating electrodes and have excellent radiation hardness by design. It overcomes the signal loss with a low operation voltage by reducing the collection length compared to the current planar technology used in the ATLAS pixel detector.

The ATLAS insertable B-layer (IBL) is an upgrade to improve tracking resolution of the inner tracker and will be installed in 2013. It will be composed of 75% planar sensors and 25% 3D sensors in the large-η region. It is important to simulate the IBL tracking performance and to have a valid model for 3D sensors.

This thesis investigated the experimental data for heavily irradiated planar strip sensors and 3D sensors to develop a device simulator, in which impact ionisation has to be included. The modelling has found that the radiation induced effective doping concentration has two linear regimes with a smaller growth rate at high fluences. This shows the possibility to operate silicon sensors with a higher irradiation level.

The signal efficiency of each pixel is the basis to simulate the whole IBL response. A model and a code were developed to calculate the induced signal from electron-hole pairs generated by the traversing charge particles. This results in a 2D efficiency map used as an input of the 3D digitiser for the Geant4 simulation. This map was adopted by the IBL software team for the whole tracker simulation and has been validated by the test beam data. Key words: 3D silicon sensor, simulation, radiation damage

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Signed: Ching-Hung Lai
Date: 03/July/2013

Copyright

Acknowledgements

I deeply appreciate my supervisor, Prof. Stephen Watts, for his valuable and continuous guidance. This work cannot be finished without his patience and experience.

I would like to appreciate my advisor, Dr. Cinzia Da Vià, and Prof. Dalla Betta, Marco, Iain, Clara and Vlad for their essential assistance. I would also like to thank Prof. Loebinger, Sabah, Anne and all other people in the Manchester HEP group to provide a friendly and knowledgeable environment. I give my special thank to Marcello for our daily exchange of opinions.

I am grateful to my parents for their support which drives me forward. I really appreciate Prof. Zi-Yi Li and Guang-Chuang who broaden and deepen my thought. I am also happy that my friends, David, Mark, Angela, Anny, Kate, Daria, Elsa and Cheng-Ti, are always around and enrich my life. Nearly four years of studying, this important period makes differences.

Chapter 1

Introduction

3D pixel detectors are a novel technology for use in silicon trackers which offer radiation hard design with a low operation voltage for high luminosity hadron colliders. They will be installed in the ATLAS insertable B-layer (IBL) upgrade which is one motivation of this thesis. They also have potential for next generation linear colliders due to their fast collection time. The technology also has applications for biomedical micro-dosimetry (radiation therapy) and space dosimetry.

This chapter describes the reason to utilise 3D technology from physics motivations to detector designs. The thesis is then divided into three parts. The first part discusses the physics of semiconductor devices (Chapter 2) and defects (Chapter 3). The second part discusses the simulation algorithms for device modelling (Chapter 4) and carrier tracking (Chapter 5). The third part compares simulation results with laboratory characterisations (chapter 6) and test beam measurements (Chapter 7). Future modifications and applications will be also described.

1.1 The Standard Model

The Standard Model (SM) is a precise theory of particle interactions which has been widely tested up to 100 GeV and accepted over the past decades [131]. The LHC is now testing the SM up to TeV energies. It is a quantum field theory (QFT) which is consistent with (classical) special relativity and quantum mechanism. It divides elementary particles which are observed experimentally into two categories: spin-half quarks and leptons which constitute matter and integer-spin bosons which mediate fermion interactions [50].

Composite mesons and baryons are made of two and three quarks. Baryons like protons and neutrons can form atoms with electrons. The coupling constants $\alpha's$ determine the strength of interactions. The strong force ($\alpha_s \sim 1$) confines quarks by exchanging massless gluons. It also binds nuclei (protons and neutrons) tightly

regardless of the repelling electromagnetic force ($\alpha \sim 1/137$, the fine structure constant) which is mediated by massless photons.

Only the weak interaction ($\alpha_w \sim 10^{-6}$) can change quark flavours and violate P-symmetry and CP-symmetry. It is responsible for radioactive decay, such as β^- decays, $n \to p + e^- + \bar{\nu_e}$, through charged and neutral currents [95]. Particles are more likely to decay electromagnetically than through the weak interaction which is mediated by massive $W^{\pm}$ and Z bosons. The gravitational force ($\alpha_g \sim 10^{-39}$) is not included in the Standard Model.

The electro-weak gauge group, $U_Y(1) \otimes SU_L(2)$, relates the electric charge, weak iso-spin and weak hyper-charge with four massless gauge generators. $W^{\pm}$ and Z bosons were found to be massive experimentally and thus initiated further research. The Higgs mechanism was introduced and extended in around 1964 to generate boson masses through the spontaneous symmetry breaking (SSB) mechanics. Fermions can obtain masses through the Yukawa interaction (between a Dirac and scalar field) with the Higgs field.

1.2 LHC

The large hadron collider (LHC) is the largest particle accelerator ever constructed with the highest energy in the world. It was constructed at the Franco-Swiss border near Geneva in Switzerland from 1998 to 2008 by the European Organization for Nuclear Research (CERN). It was installed in the existing 26.7 km tunnel of the large electron-positron collider (LEP) which is around 45 to 170 m beneath the earth surface [38].

1.2.1 Motivation and Status

The Standard Model gives a satisfactory agreement with many experimental data, but the predicted Higgs boson has not been found for more than 50 years. The Tevatron at the Fermilab has been the energy and luminosity pioneer for a quarter of century [119], but it would need a long time to accumulate enough data for a Higgs search. The LHC was designed to have high beam luminosity and collision energy since the number of events is proportional to the luminosity and cross sections.

Figure 1.1 shows the combined Higgs results from the Tevatron and LEP which show that $158 < m_H < 175$ and $100 < m_H < 109$ GeV/c^2 are excluded. With only a few years of run after an accident in 2008, a wider range of the Higgs mass is excluded by the LHC in 2011. CERN announced that one Higgs-like particle was observed late in 2012. It has a mass of 126.5 and 125 GeV measured at ATLAS [21] and CMS [31] respectively and its additional properties are being studies. Physics beyond the Standard Model, such as super-symmetry and extra dimensions based on the string

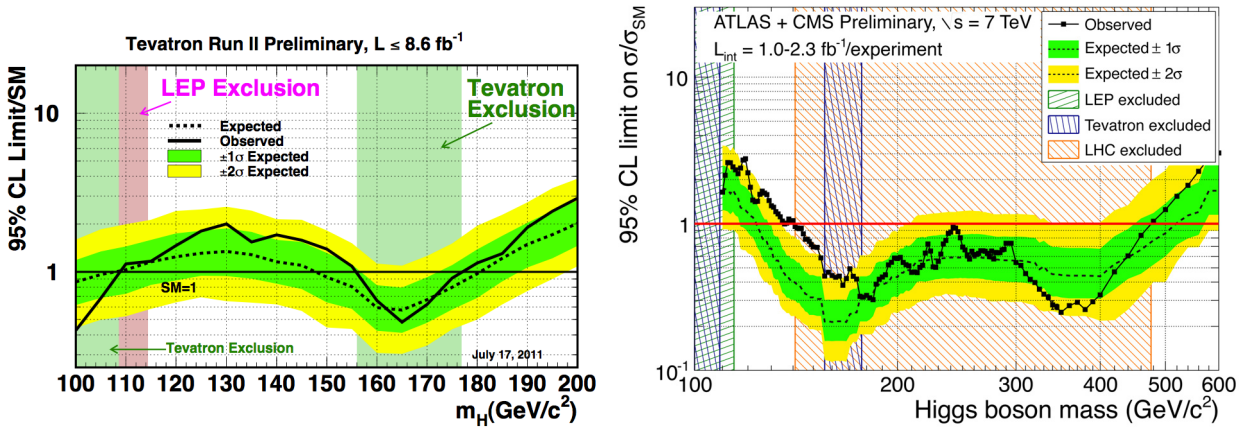


Figure 1.1: Exclusion of the Higgs boson mass for the combined CDF and D0 analyses (left) [119] and the LHC data (right) [11]. The observed data (solid) and expected without Higgs (dashed line) cross sections are plotted on the ratios to the Standard Model as a function of the Higgs mass. Data below the ratio of 1 are excluded regions. Data above the upper limits of 95% confidence level suggests possible Higgs mass regions.

theory, is also being explored by the LHC data.

1.2.2 Design

Unlike the Tevatron, which is a proton anti-proton collider, the LHC is a proton proton collider. Figure 1.2 shows the LHC scheme. Each proton beam is accelerated to 7 TeV which can provide a centre-of-mass energy $\sqrt{s}$ to 14 TeV. It is achieved by several stages: protons are sped up to 50 MeV, 1.4 and 26 GeV at the linear accelerator (Linac2), PS booster (PSB) and Proton Synchrotron (PS) sequentially. They are finally accelerated to 450 GeV in the Super Proton Synchrotron (SPS) and injected into the LHC ring [126]. This high energy significantly increases the interaction probability, especially for the Higgs boson, according to the pp cross sections.

The machine luminosity L is calculated as

$$L \approx \frac{N_b^2 n_b f_{rev}}{4\pi\epsilon_n} \tag{1.1}$$

for a beam which circulates at a frequency f_{rev} with n_b bunches per beam and N_b particles per bunch. ϵ_n is the normalized transverse beam emittance which is the average spread of particles as a function of the position and angle. The LHC has a revolution rate of 40 MHz which corresponds to a 25 ns bunch crossing time. It aims at the peak luminosity of 10^{34} $cm^{-2}s^{-1}$. This prohibits the use of anti-protons with a low production rate which has a merit of using the same proton ring [105].

There are four large detectors occupying half of the colliding sites on the LHC ring: a toroidal LHC apparatus (ATLAS) and the compact muon solenoid (CMS) are large

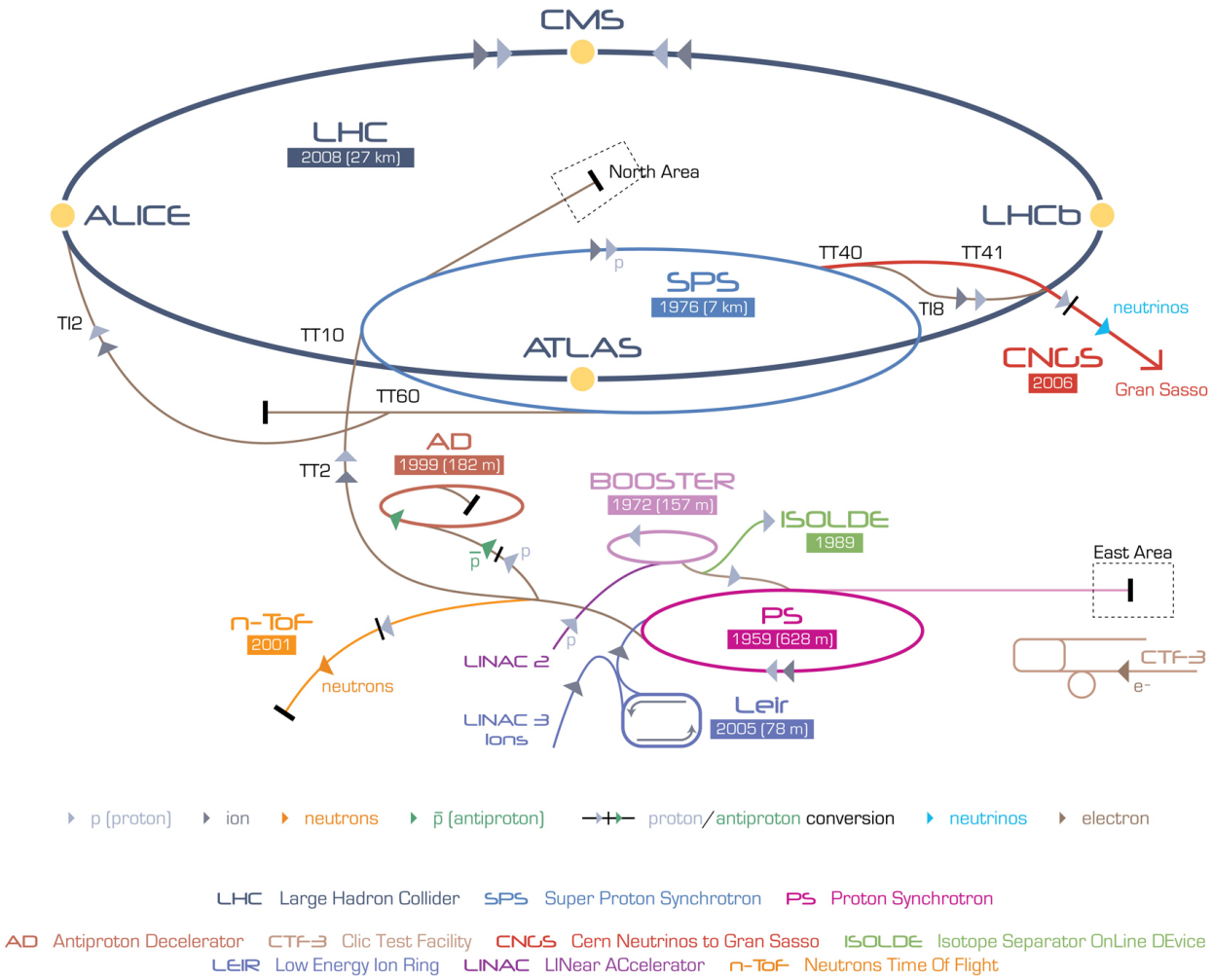


Figure 1.2: The accelerator complex [8] consists the linear accelerator (Linac2), PS booster (PSB), Proton Synchrotron (PS), Super Proton Synchrotron (SPS) and the LHC ring.

general purpose particle detectors. A large ion collider experiment (ALICE) and LHCb (b for beauty quark) act for more specific roles. Furthermore, the total elastic and diffractive cross section measurement (TOTEM), LHCf (forward) and the monopole and exotics detector at the LHC (MoEDAL) are much smaller for very specialized research.

1.3 ATLAS

ATLAS is designed to improve measurements of the Standard Model and search for the Higgs boson. It can also investigate new physics which may become detectable at the LHC energy. Physical events are reconstructed according to their azimuthal ϕ and polar angle θ measured in the detectors. The pseudo-rapidity η is used to replace θ

as $\eta = -ln[tan(\theta/2)]$. Momentum (energy) is conserved in the transverse plane since there is a Lorentz boost in the longitudinal direction.

The CMS experiment has a full silicon tracker with a 4 T solenoid field which gives a better momentum resolution compared to ATLAS (2 T, silicon sensors plus proportional chambers). The CMS calorimeters, placed inside the superconducting coil, are good for e/γ identification but bad for jet resolution. The ATLAS hadronic calorimeter has a better energy resolution and a longer interaction length compared to CMS with a compact design. This section discusses different components of the ATLAS detector from inner to outer.

1.3.1 Inner Detector

Inner detectors (IDs) are used for particle tracking which provide a precise momentum by measuring the curvature of charged particle tracks in the large magnetic field. There are approximately 23 overlapping interactions and 1000 generated particles per beam crossing [118][20]. Sensitive (higher granularity) and fast responding (25 ns) trackers are needed for such a high rate. More than three layers of detectors (three points determine a curvature) are needed since the tracking efficiency is not always 100% and there are gaps between neighbouring sensors.

The pixel detector, followed by the silicon micro-strip tracker (SCT), were designed to meet certain requirements. The pixel design increases the spatial resolution and also reduces the occupancy rate to around 10^{-4} of each pixel to avoid pile-up events. At a larger radius, straw-tubes of the transition radiation tracker (TRT) provide 36 tracking points. Figure 1.3 shows a schematic of the ATLAS inner detector.

The innermost pixel detector uses double-sided planar technology to obtain space points. Each pixel is a $n+$ implantation with a moderate p-spray for separation. The substrate uses oxygenated n-type silicon with a thickness of 280 μm. The RD48 Collaboration found that oxygenation reduces the build up space charge after type inversion [73]. Defect engineering is still being investigated by the RD50 Collaboration.

The SCT uses classic single-sided p-in-n technology, in which the strips record 1D positions. Two layers of strip detectors with an orthogonal orientation are used to obtain 2D information. The SCT, like the pixel detector, uses AC-coupled readout chips to deal with the large leakage current after high radiation doses. There are eight layers in the barrel SCT mounted at four radii.

The transition radiation tracker (TRT) is situated between the SCT and solenoid magnet. It is composed of thousands of polyimide drift straw-tubes with a 4 mm diameter. Straws are cathodes filled with a transition radiation gas mixture (70% Xe, 27% CO_2 and 3% O_2). Tungsten wires are anodes for signal collection.

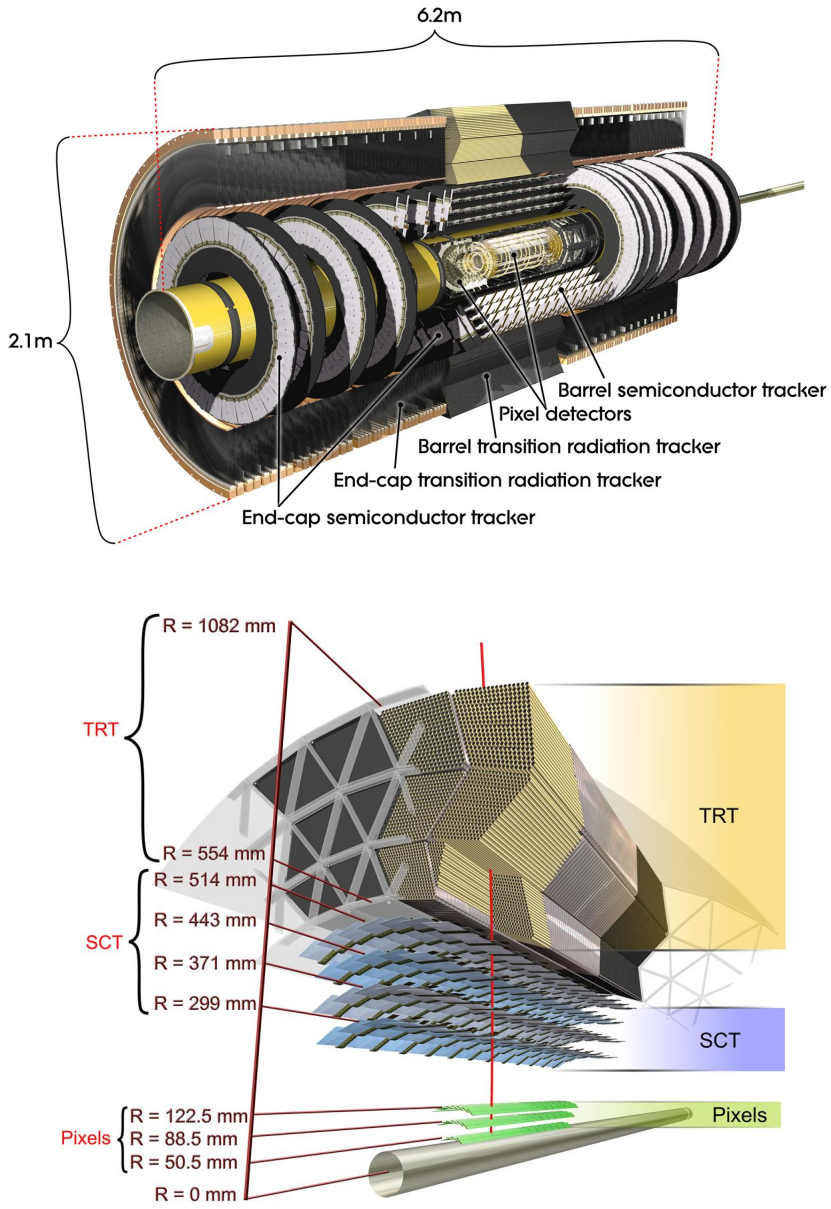


Figure 1.3: A cut-away view of the ATLAS inner detector. The superconducting solenoid provides a magnetic field of 2 T [1]. The pixel detector has three layers and the SCT has eight layers at four radii in the barrel. Multiple layers of detectors perverse the overall tracking efficiency after degradation by irradiation.

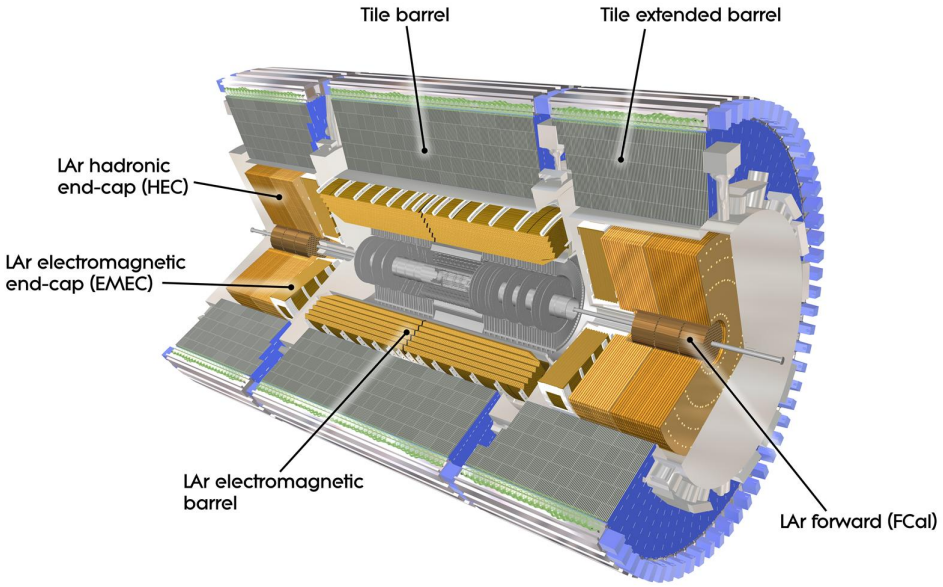


Figure 1.4: A cut-away view of the ATLAS calorimeters [1].

1.3.2 Calorimeter

The ATLAS calorimetry consists of the electromagnetic calorimeter (EMCal, $|\eta| < 2.475$) and electromagnetic end-cap calorimeter (EMEC, $1.375 < |\eta| < 2.475$). The hadronic barrel (or tile) calorimeter (HCal, $|\eta| < 1.7$), hadronic end-cap calorimeter (HEC, $1.5 < |\eta| < 3.2$), and forward calorimeters (FCal, $3.1 < |\eta| < 4.9$) are shown in Figure 1.4 [118][20]. They are all sampling calorimeters which provide a better spatial resolution but a worse energy resolution compared to a homogeneous design.

A pre-sampler detector ($|\eta| < 1.8$) is used to correct the energy loss in the inner material (IDs, cryostats and coils). The EM calorimeter is a lead/liquid-argon (LAr) detector. It has accordion-shaped kapton electrodes and lead absorbers with a complete ϕ-symmetry without azimuthal tracks. The EM calorimeter measures the energy of particles which interact electromagnetically. Photons, for example, interact primarily through pair production. The hadronic calorimeter has scintillation tiles as sampling medium and steel as absorbers. It stops hadronic particles through nuclear interactions and measures their energy.

1.3.3 Muon System

Muons are a minimum ionising particle (MIP), an energetic and heavy version of electrons, which deposit only a small fraction of their energy as they transverse the detectors. The ATLAS muon spectrometer uses a multi-layer muon chamber to increase the

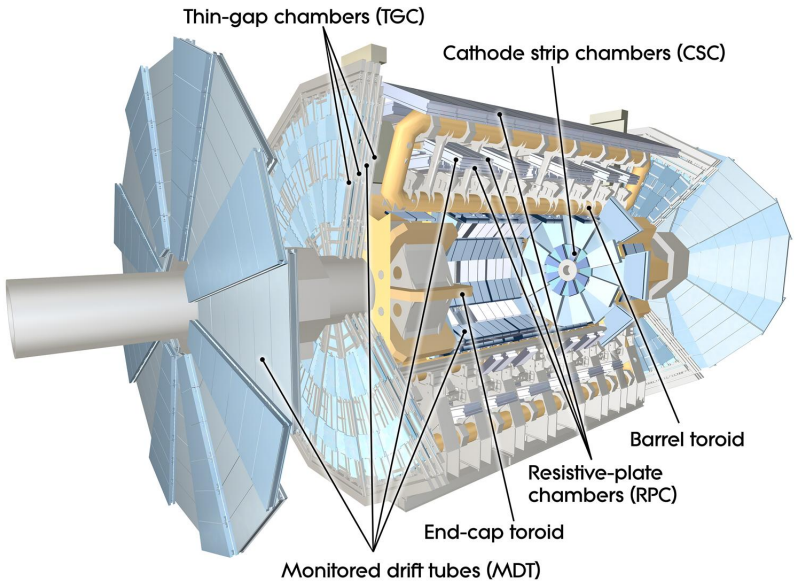


Figure 1.5: A cut-away view of the ATLAS muon detector [1] which makes up the size of the system for a total height of more than 12 m.

momentum precision. It has special toroidal magnets (barrel and endcap) which covers the whole η range. Toroidal magnets are more economic than solenoidal ones due to the coil shape. Muons at $|\eta| < 1.0$ and $1.4 < |\eta| < 2.7$ and the transition region of $1.0 < |\eta| < 1.4$ are deflected by the barrel, two end-cap magnets and the combination of the two respectively.

The muon detector consists of monitored drift tubes (MDTs), cathode strip chambers (CSCs, with a finer granularity), resistive plate chambers (RPCs) in the barrel and thin gap chambers (TGCs) in the end-cap region as shown in Figure 1.5. MDTs and CSCs measure the bending and direction. RPCs and TGCs provide the bunch-crossing identification and transverse momentum p_T thresholds for triggering [118][20].

1.4 Vertex Detectors

Particle trajectories were recorded visually in earlier particle physics experiments using bubble chambers or cloud chambers [48]. The multi-wire proportional chamber, which was invented by Charpak in 1968, gave an electronic viewpoint of particle trajectories. Trackers with a higher resolution, such as cylindrical proportional chambers and time-projection chambers (TPC), were developed quickly afterwards. Most tracker systems use silicon detectors nowadays which provide more precise and faster measurements.

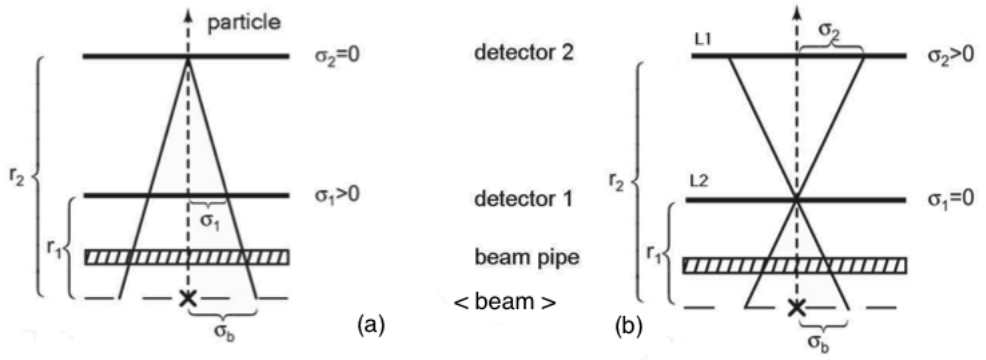


Figure 1.6: An ideal two layer detector system [129]. (a) has a perfect detector 2 and (b) has a perfect detector 1. The cross represents the interaction point.

1.4.1 Tracking Resolution

Tracking resolution can be understood by estimating the impact parameter error σ_b of a track as follows [129]: if detector 2 is perfect ($\sigma_2 = 0$) as shown in Figure 1.6 (a), the ratio of σ_b and σ_1 is written as $\sigma_b/\sigma_1 = r_2/(r_2 - r_1)$. If detector 1 is perfect ($\sigma_1 = 0$) in as Figure 1.6 (b), similarly the ratio is written as $\sigma_b/\sigma_2 = r_1/(r_2 - r_1)$.

Summing two contributions of σ_b and the multiple scattering term σ_{mp} yields

$$\sigma_b^2 = (\frac{r_2}{r_2 - r_1}\sigma_1)^2 + (\frac{r_1}{r_2 - r_1}\sigma_2)^2 + \sigma_{mp}^2 \ . \tag{1.2}$$

This indicates that the resolution of the innermost layer and its distance to the interaction point (IP) dominate σ_b. This leads to the replacement of wire chambers by silicon trackers due to a higher spatial resolution (σ_1).

For pixel detectors, each pixel has a minimum size in $(R - \phi) \times z$ of $50 \times 400\ \mu m^2$ which gives intrinsic accuracies of 10 μm $(r - \phi)$ and 115 μm (z) in the barrel [20]. For the SCT, each strip has a pitch of 80 μm and a length of 6.4 cm which gives an intrinsic accuracies of 17 μm $(r - \phi)$ and 580 μm (z) in the barrel. For the TRT, a spatial resolution of 130 μm per straw is obtained by measuring the drift time.

1.4.2 Hybrid Pixel Detector

Pixel detectors were proposed to provide position information without the ambiguity that occurs in multi-wire chambers or silicon strip detectors. Charge-coupled devices (CCDs) were the first pixel detector for linear colliders, e.g. the SLD experment at the Stanford Linear Accelerator Center (SLAC). However, they are too slow for the LHC due to a single amplifier for all channels which can provide an uniform image for astronomy telescopes.

The hybrid pixel detector uses its silicon substrate as an active region for particles

to deposit energy. Another readout chip, FE-I3 (front end IC version 3), is flipped and connected to the sensor which is mounted on the module by bump bonds. Indium (In) and PbSn are used to form contacts between two chips by AMS in Italy and IZM in Germany. The FE-I3 uses a standard 0.25 μm CMOS technology. It uses special layout rules to tolerate up to a total ionising dose of 50 $Mrad$ [93].

1.4.3 The IBL Upgrade

The LHC was shut down at the end of 2012 for upgrades on the accelerator (to a beam energy of 6.5 TeV) and large experiments. For ATLAS, the IBL detector will be placed just a few cm to the beam pipe. It is also a pixel detector with a smaller cell size in $(R-\phi)\times z$ of 50×250 μm^2. Both the shorter distance (r_1) and smaller pixel size will improve impact parameters and resolutions to < 10 μm $(r-\phi)$ and 72 μm (z). The number of channels will also increase from 18×160 [93] to 80×336 pixels [27].

After a long period of research and development, 3D sensors have been shown to be a feasible and durable candidate for the IBL upgrade. Planar sensors will contribute 75% of the IBL detector at low-η ranges (central part). 3D sensors will contribute 25% at high-η ranges (two ends) due their small Lorentz angle [34] and better z-resolution which will be discussed later. Two types of sensors are being fabricated and characterised and will be installed in 2013.

The original pixel detectors are able to operate to a total non-ionising dose of 10^{15} $n_{eq}cm^{-2}$ for ten years of run. The IBL sensor chip and FE-I4 readout chip have to tolerate five times more. Figure 1.7 shows the expected fluences for the pixel detectors at different radii. To meet this requirement, the planar collaboration reduces the substrate thickness to around 210 μm for a lower full depletion voltage and a higher signal efficiency. The 3D sensors have a low operation voltage intrinsically and use a 230 μm thick substrate to prevent sensors from breaking.

1.5 3D Technology

3D sensors, unlike conventional detectors, have electrodes that penetrate through the silicon substrate. This technology was proposed by Parker and Kenney in 1997 [92] for particle physics application. There are 18 institutions in the ATLAS 3D collaboration working on sensor characterisation and simulation. In addition, there are four facilities fabricating devices, SNF (Stanford Nano Fabrication Facility) in California in the USA, SINTEF (Stiftelsen for industriell og teknisk forskning) at Oslo in Norway, CNM (Centro Nacional de Microelectróica) at Barcelona in Spain and FBK (Fondazione Bruno Kessler) at Trento in Italy.

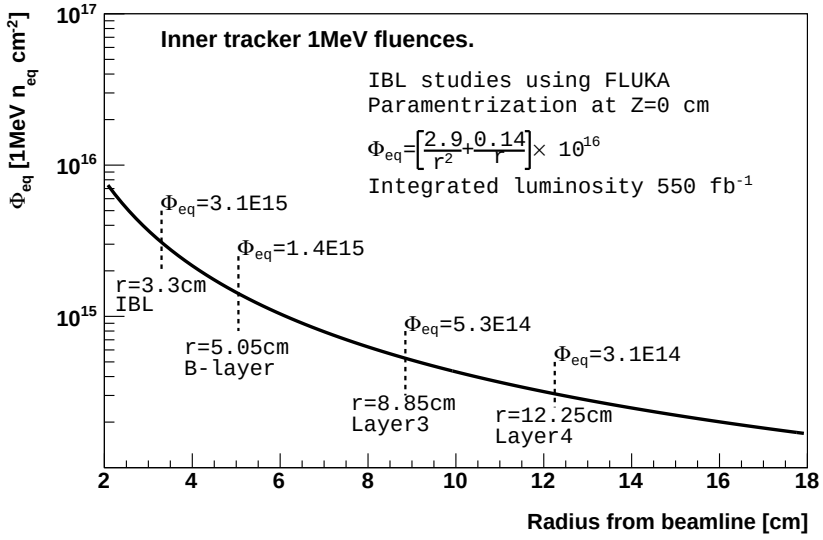


Figure 1.7: Expected fluences for the IBL detector and three current pixel detectors at different radii. The plot is based on [27].

1.5.1 Motivation

Radiation hardness and power consumption (which is controlled by the leakage current) are key criteria for upgrades like the IBL for ATLAS and the future high-luminosity LHC (HL-LHC) which is expected to give a total dose of 2×10^{16} $n_{eq}cm^{-2}$ for B-layer detectors from the targeted luminosity of 10^{35} cm^{-2}. New technologies are needed to preserve charge collection and operation conditions after heavy irradiation.

The 3D technology has a small inter-electrode length intrinsically due to the electrode orientation. Its shortest diagonal distance is 67 μm for the IBL design, while it is more than 200 μm for the planar technology. This results in good radiation hardness since a shorter drift length for carriers gives a faster collection time with less trapping. 3D sensors also have a smaller full depletion voltage but a larger capacitance which results in higher noise.

1.5.2 General Remarks

Figure 1.8 (a) shows general ideas of planar and 3D sensors. Penetrating electrodes with $n+$ and $p+$ dopants form a pn junction. Sensors are reverse biased with a negative voltage on the $p+$ electrodes. Readout ($n+$) electrodes collect electrons and bias ($p+$) electrodes collect holes. They interchange row by row and each $n+/p+$ electrode is surrounded by four other type of electrodes. Planar sensors have surface implant as electrodes and thus its inter-electrode distance is the substrate thickness.

The full depletion voltage V_{dep}, which can be estimated by solving the Possion

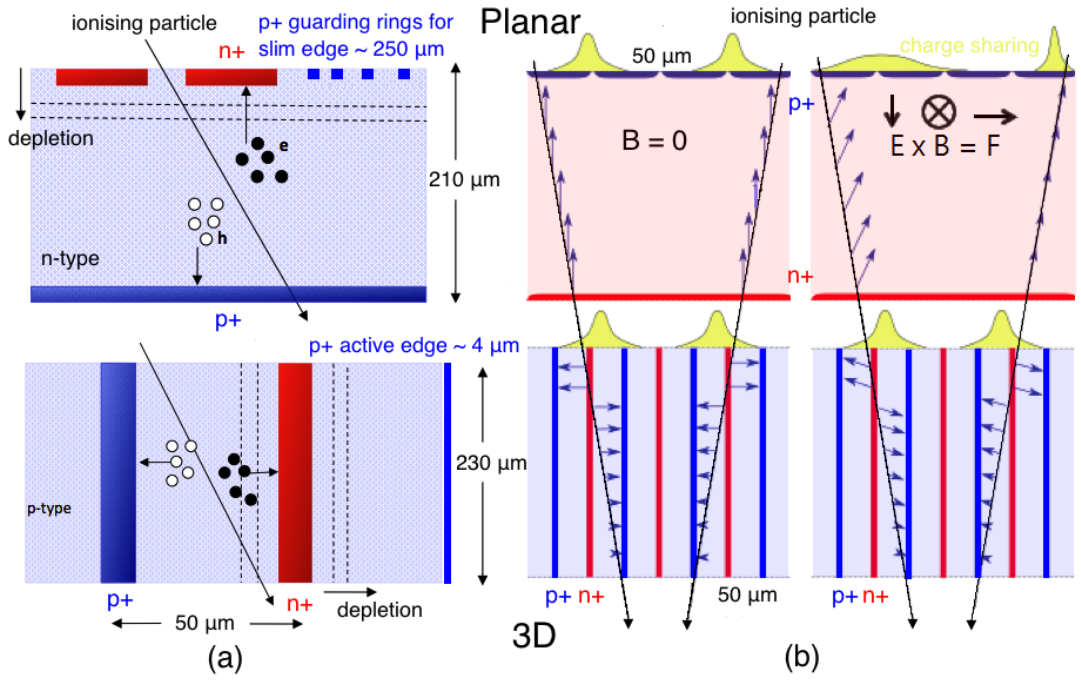


Figure 1.8: (a) a schematic of carrier generation and charge collection. (b) effects of the magnetic field on carrier movements towards collection electrodes for planar (upper) and 3D (lower) sensors respectively [47]. 3D sensors have electric field lines mostly parallel to the magnetic field and are less effected by the magnetic field. Even if the electric and magnetic field lines are orthogonal, carriers are bended in the vertical direction.

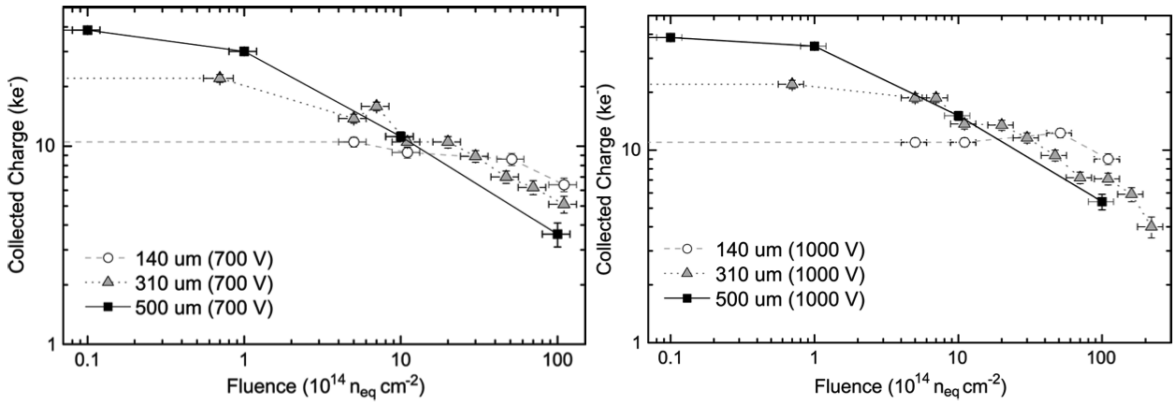


Figure 1.9: The depletion width and charge collection of planar sensors decrease with the fluence level at different thickness [66] at 700 (left) and 1000 (right) V. A thinned sensor (140 μm) can collect more than a thick one (310 μm) for a fluence greater than 5×10^{15} $n_{eq}cm^{-2}$ due to full depletion. There is also a hint of charge multiplication for the thinned sensor at a certain fluence.

equation, increases roughly with the square of the inter-electrode distance d_{ie} of 3D or the thickness d of planar sensors respectively. Though V_{dep} is larger for a coaxial (cylindrical) geometry is smaller than a planar structure, d_{ie} is still rather smaller d_{ie} and thus 3D have a smaller operation voltage. This is the reason that the planar collaboration reduced d of the IBL detector to halve V_{dep} and is trying to make further thinning for the substrate thickness to 140 μm. It will be a challenge for processing since a support wafer is needed to avoid bending (bowing). Figure 1.9 shows the depletion width and charge collection of planar sensors decrease with the fluence level at different thickness. A reduction in thickness means a reduction in signal, but thinned planar sensors can collect more than thick ones for a fluence greater than 5×10^{15} $n_{eq}cm^{-2}$ due to full depletion. However, this also results in larger capacitance as for 3D sensors.

Thick planar sensors may suffer from partial depletion as damage builds up with the operation time. They lose detection region gradually from the $p+$ plane where high-η events may be lower than the threshold. Even at full depletion, carriers generated by high-η incident particles experience more trapping. 3D sensors have a dead zone around $p+$ electrodes but remain active, due to their electrode and collection orientation, for high-η incident particles.

1.5.3 Process and Design

Electrodes of 3D sensors are drilled over a small radius to preserve detection area. Charge sensitive electrodes are of interest and under development (Chapter 7). Deep reaction ion etching (DRIE) is necessary to enhance a high aspect ratio (depth/width) of around 20 to 40. The Bosch process is one method of DRIE which uses pulsed etching to achieve nearly vertical edges. Figure 1.10 shows photos of electrodes fabricated by DRIE from SINTEF (upper) and Stanford (lower). Holes are well-formed and then filled with doped polycrystalline silicon. Electrodes can be also implanted with dopants directly without filling like the CNM and FBK sensors.

The substrate for 3D sensors is chosen to be a p-type float-zone (FZ) silicon since n-type materials were found to invert to p-type after irradation, namely more deep acceptors are introduced. Figure 1.11 shows the type inversion for different materials after neutron and proton radiation [74]. The inversion fluence is around 10^{12} to 10^{14} cm^{-2} which depends on material and radiation types [42][82][97]. Float-zone silicon is a single crystal grown from a seed using melted polycrystalline ingot by a moving RF coil. It has very low impurities, such as oxygen and carbon atoms, compared to single crystals grown by the Czochralski process.

3D sensors use $p+$ guard electrodes or active edges and the electric field drops to zero within a few cells (Chapter 6). Planar sensors use $p+$ guard rings at the edge which needs a larger area to remove the leakage current from dangling bonds. Figure

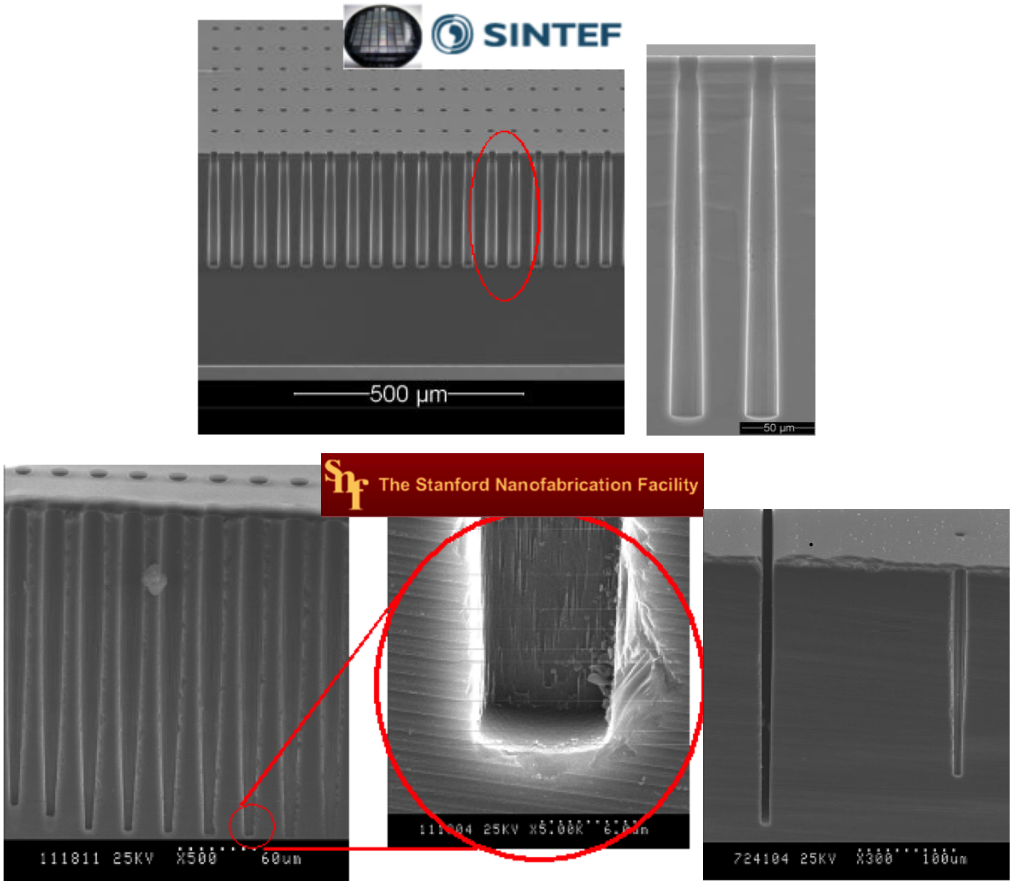


Figure 1.10: Photos of electrodes show integrity of pattern transfer using deep reaction ion etching technique [16].

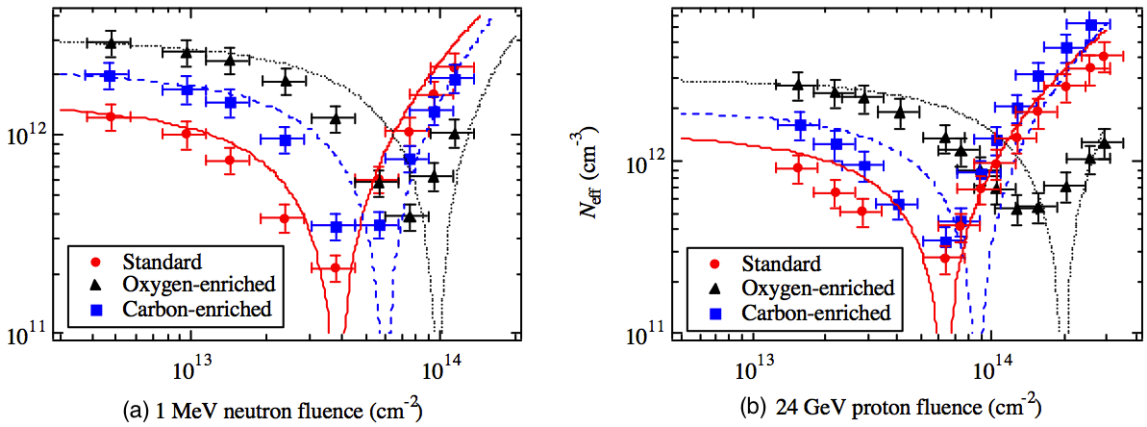


Figure 1.11: Type inversion for different materials after (a) neutron and (b) proton radiation [74]. The inversion fluence depends on material and radiation types.

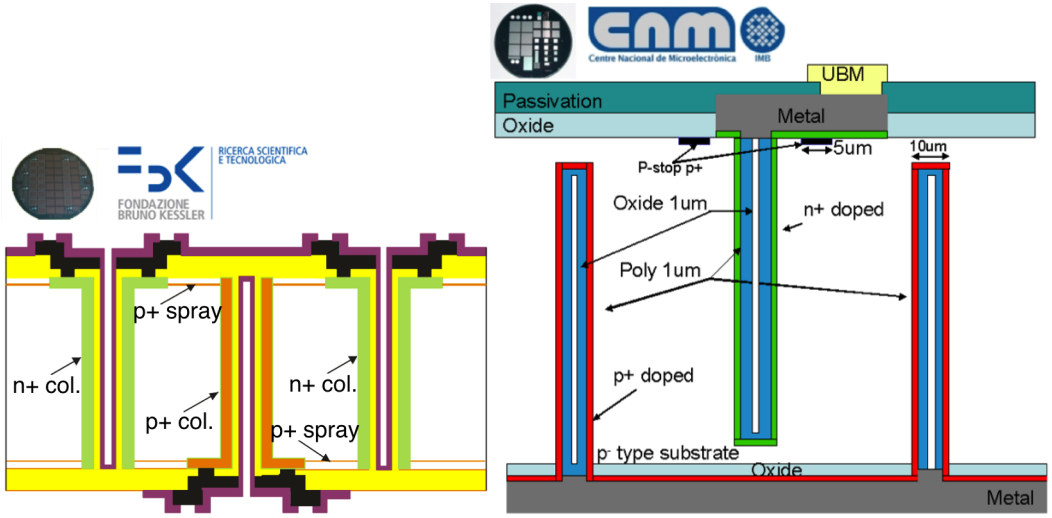


Figure 1.12: Sensor geometries for the FBK and CNM design [74]. FBK sensors have double-sided non-fully through columns for a lower depletion voltage, while CNM sensors have also double-sided but non-fully through column.

1.8 (b) shows that 3D sensors are less affected by the magnetic field [47] and thus a smaller Lorentz angle (Chapter 5).

3D sensors were developed and characterised on FE-I3 chips which can have two, three or four $n+$ electrodes (cells). The two electrode (2E) design was chosen for the FE-I4 chip which is a compromise between capacitance and the reduced inter-electrode spacing. SINTEF and Stanford use the full-3D design which forms electrodes on one side and needs a support wafer. This design also includes the full active edge which further reduces the edge area and is planned for the future upgrades

Figure 1.12 shows a horizontal cut view of FBK and CNM sensors. FBK fabricates double-sided fully through columns with floating $p+$ guard electrodes (referred to as full-through design/electrode). CNM fabricates double-sided non-fully through columns with biased $p+$ guard electrodes (referred to as non-full-through design/electrode). Both facilities use a $p+$ spray to isolate $n+$ electrodes from connecting through the electron layer which is induced by the positively charged oxide layer after irradiation. Experimental and simulation results for both 3D sensors will be discussed in the later chapters.

Chapter 2

Physics of Semiconductors

Semiconductor physics and devices have been investigated and studied for over 140 years since the first metal-semiconductor contact was discovered [117]. Semiconductor devices are widely used for research and commercial applications. Silicon sensors are a modification of *pn* junctions which were invented by Shockley in 1949 [109]. Readout chips use metal-oxide-semiconductor field-effect transistors (MOSFETs) which were invented by Kahng and Atalla in 1960. The silicon industry and related researches have had a large influence on society and will continue to do so.

This chapter describes the underlying semiconductor physics and material properties which are key inputs for device simulation. Basic characteristics of silicon can be understood from the crystal structure and the resulting band structure. Carrier concentrations and their relationship to the Fermi level are discussed. Generation and recombination processes are important in doped and irradiated regions. Semiconductor equations can describe the carrier dynamics for most cases. The physics of a *pn* junction diode in steady state is also discussed.

2.1 Semiconductor Material

Electronic devices are fabricated on silicon wafers which are slices of silicon ingot. The ingot is a solid cylinder grown from a crystalline seed. The wafer orientation depends on the cut surface across the ingot. Four common types of wafer are manufactured: $(1,0,0)$ or $(1,1,1)$ and p or n which are the Miller index and initial impurity doping respectively. These types are denoted by primary and secondary flats on the wafer circumference.

The substrate wafer for 3D sensors is chosen as high resistivity p-type silicon with crystal orientation $(1,0,0)$. The initial doping concentration is around $5\times\ 10^{11}\ cm^{-3}$ which is nearly intrinsic compared to the atom concentration of $5\times\ 10^{22}\ cm^{-3}$. $(1,0,0)$

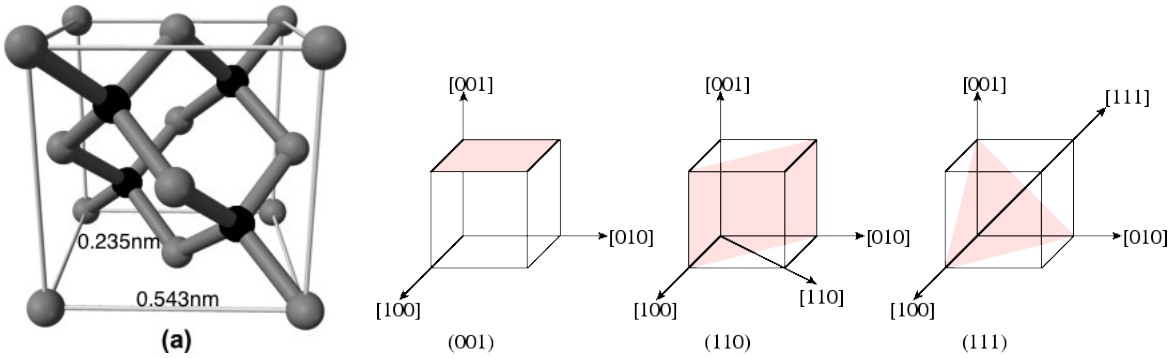


Figure 2.1: (a) diamond cubic crystal structure [60] and (b) three simple planes for a solid [37]. [hkl] represents crystal directions.

wafers have better surface quality with less dangling bonds and leakage current. MOSFETs also use $(1, 0, 0)$ wafers. Other applications, e.g. MEMs, may use $(1, 1, 1)$ wafers since they have better mechanical properties for motive sensing [69].

2.1.1 Crystal Structure

Silicon and germanium crystals have a diamond cubic structure with a lattice spacing of 0.54 and 0.566 nm. It has a repeating pattern as shown in Figure 2.1 (a) which has 8 atoms in a cell: 1 (1/8, 8 at corners) + 3 (1/2, 6 on faces) + 4 (1, 4 in the bulk). Each atom has four covalent bonds for silicon (diamond structure) and each bond has two tightly bound electrons. Diamond (carbon) has the same structure with a stronger $C - C$ covalent bond. Most of III-V compound crystals, e.g. gallium arsenide (GaAs), have a zincblende lattice which has a similar structure.

The Miller indices (h, k, l) are used to represent lattice planes given as

$$\mathbf{g}_{hkl} = h'\mathbf{a}_1 + k'\mathbf{a}_2 + l'\mathbf{a}_3 \tag{2.1}$$

using lattice vectors $\mathbf{a}_i$. (h', k', l') are intercepts of the plane on the three Cartesian axes. (h, k, l) is reduced from the reciprocal of (h', k', l') to integers. Figure 2.1 (b) shows three simple planes and two of them are common selections to slice silicon ingots into wafers.

The crystalline view illustrates the bonds of material atoms, but fails to account for quantum-mechanical effects. Some qualitative concepts, such as broken bonds and impurity sites [87], are well depicted. Thermal energy breaks some bonds of silicon (semiconductor) and creates some nearly-free electrons at higher temperatures. They become a current if an electric field is applied. For broken bonds or boron dopants, electrons can jump to adjacent vacant bonds which is equivalent to moving holes under an applied electric field. This is unlike diamond (insulator) which has no free electrons.

2.1.2 Band Structure

The band structure helps to understand electron motion in a semiconductor. The band model starts from the quantised energy levels for an electron orbiting a hydrogen-like (isolated) atom or ion. The electrostatic interaction, the Coulomb potential, determines the allowed energy levels E_n as [87]

$$E_n = -hcR_\infty \frac{Z^2}{n^2} = \frac{-13.6}{n^2}\ eV \tag{2.2}$$

where $R_\infty = m_e e^4/8\epsilon_0^2 h^3 c$ is the Rydberg constant. h, ϵ_0, c, n and Z are the Planck constant, vacuum permittivity, speed of light, principle quantum number and atomic number respectively.

According to Pauli's exclusion principle, only two electrons can be accommodated in one energy state. For two or more atoms interacting with each other, the original energy state must split to accommodate electrons around the same level. For a periodic lattice structure, atoms with different inter-distances determine the splitting. A closer distance results in a larger splitting since atoms have a stronger interaction. Splitting lines from neighbouring atoms form a band to a certain extent. Electrons can only move in partially filled bands.

There is a gap between energy bands for semiconductors and insulators (and for some metals). Band gaps for semiconductors enables various applications since electrons can be excited from the valence band to the conduction band and re-emit the gap energy. This rarely happens for insulators since their band gap is greater than 5 eV, e.g. 5.5 eV for diamond and 9 eV for SiO_2.

The exact wave function $\psi_{\mathbf{k}}(\mathbf{r})$ for the orbiting electron is obtained by solving the Schrödinger equation as [116]

$$[-\frac{\hbar^2}{2m} + V(\mathbf{r})]\psi_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}+\mathbf{G}}\psi_{\mathbf{k}}(\mathbf{r}) \tag{2.3}$$

where $\mathbf{r}$ and $V(\mathbf{r})$ are the position vector and electric potential. $\mathbf{k}$ is the reciprocal lattice vector (momentum space, k-space) which is the Fourier transform of the original lattice (spatial space) in the primitive cell. $\mathbf{G}$ is the reciprocal of the lattice defined in Equation 2.1. The 1D Kronig-Penney model assumes the electric potential $V(\mathbf{r})$ as a combination of periodic rectangular barriers (lattice atoms) [62]. Its band structure can be then obtained graphically or numerically.

According to the Bloch theorem, the wave (Bloch) function is written in a general form as

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} U_n(\mathbf{k}, \mathbf{r}) \tag{2.4}$$

where $U_n(\mathbf{k}, \mathbf{r})$ are the periodic coefficients for plane waves with different band index n.

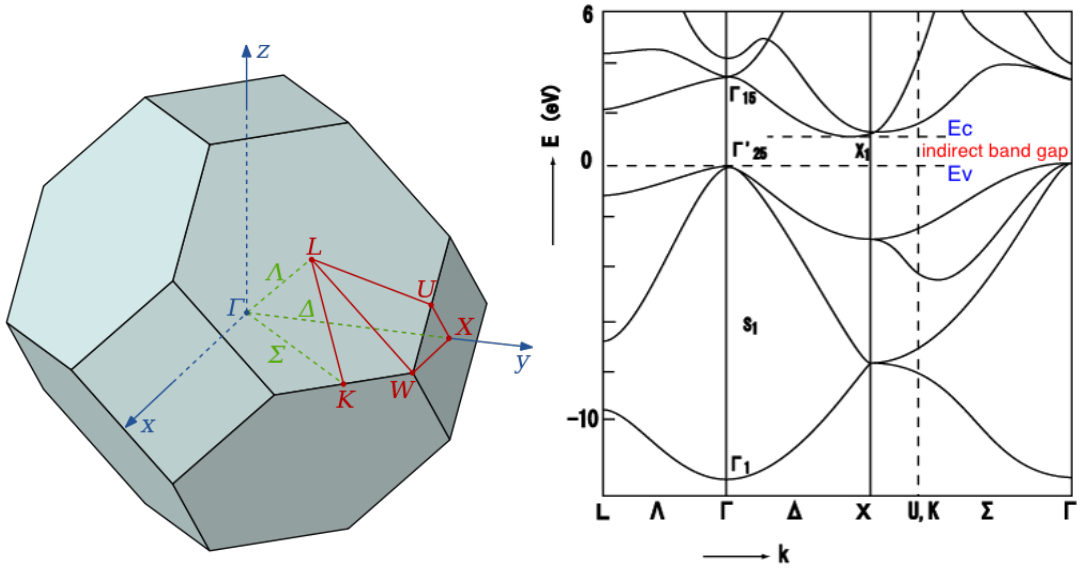


Figure 2.2: (a) reciprocal lattice of a diamond cubic crystal which is also the first Brillouin zone of a face-centered-cubic (FCC) lattice [3] and (b) band structure of silicon with an indirect band gap [10]. There are four possible reciprocal lattice vectors shown in the figure: Γ-L ($\vec{\Lambda}$), Γ-X ($\vec{\Delta}$), Γ-K ($\vec{\Sigma}$) and U-X. The energy-momentum E-k relationship is shown between different k points. For example, the leftmost region shows the E-k relationship on the vector Λ. Thus the energy states of valence electrons at point Γ are higher than those at point L. A momentum Δ (phonon) is needed for carrier excitations to the conduction band at the point X.

By applying boundary conditions, the solution gives a band gap which increases with the effective width of atoms [7][1]. The eigenvalues $E_{\mathbf{k}}$ are the allowed energy states which are periodic and determine the magnitudes of k.

The band structure is practically calculated with $\mathbf{k} \cdot \mathbf{p}$ perturbation theory which considers the interaction between the conduction and valance band. The momentum operator $\mathbf{p}$ is replaced by $\mathbf{p} + \mathbf{k}$ and the $\mathbf{k} \cdot \mathbf{p}$ term is obtained after expansion. The spin-orbit coupling can be also included as a perturbation. For each periodic k vector, the band structures vary as shown in Figure 2.2. The tight-binding model is also used which starts from an approximate superposition of the plane waves for isolated atoms. One electron mainly experiences a specific atom's electric potential and includes neighbouring atoms as a perturbation.

The band gap can be written as an empirical formula which depends on the lattice temperature given as [25]

$$E_g(T) = E_g(0K) - \frac{\alpha T^2}{T + \beta} \ . \tag{2.5}$$

The parameters are obtained by fitting experimental data as $E_g(0K) = 1.1696$ (1.519)

[1]For narrow and low barriers, electrons are free to move and experience no band gap which is the free electron model for metals.

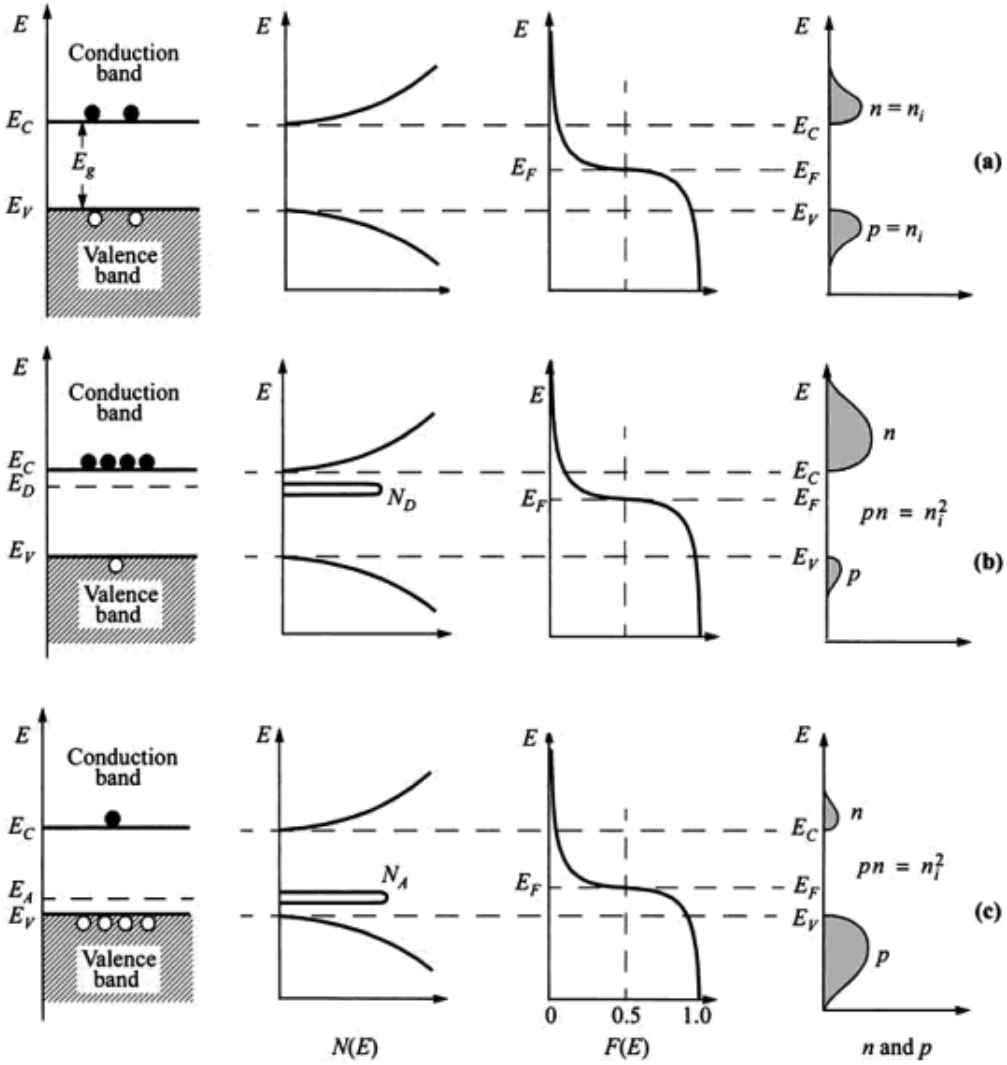


Figure 2.3: Scheme of the band diagram, density of states, Fermi-Dirac distribution and carrier concentrations for (a) intrinsic, (b) n-type and (c) p-type semiconductors [116].

eV, $\alpha = 4.73 \times 10^{-4}$ (5.405×10^{-4}) eVK^{-1} and $\beta = 636$ (204) K for silicon (GaAs). They give a band gap of 1.12 (1.42) eV for silicon (GaAs) at 300 K. This is used in device simulation since the intrinsic carrier concentration n_i depends on the band gap.

2.2 Carrier Concentrations

This section describes the occupancy function in an intrinsic silicon which has very few impurities compared with thermally generated carriers. Along with the band structure and the density of states, the carrier concentrations can be derived. The relationships between carrier concentrations and energy levels will be often used in later chapters.

2.2.1 Intrinsic Semiconductor

Electrons are fermions and thus the probability $f_{FD}(E)$ of finding an electron at a given energy E obeys Fermi-Dirac statistics

$$f_{FD}(E) = \frac{1}{1 + e^{(E-E_F)/k_BT}} \tag{2.6}$$

where k_B is the Boltzmann constant and T is the absolute temperature in Kevin. E_F is the Fermi level where the probability of occupation of an electron is exactly one-half at this energy. At $T = 0\ K$, the distribution is abrupt which is smeared by thermal energy as kinetic energy for carriers. Thus the conduction band is partially filled and the valance band is partially empty. At a certain temperature, the Fermi distribution function can be approximated as

$$\begin{aligned} f_{FD}(E) &\simeq e^{-(E-E_F)/k_BT} \approx 0.95 \quad \text{for} \quad E - E_F > 3k_BT \approx 0.08\ eV \\ f_{FD}(E) &\simeq 1 - e^{-(E-E_F)/k_BT} \approx 0.05 \quad \text{for} \quad E - E_F < 3k_BT\ . \end{aligned} \tag{2.7}$$

The electron density $n(E)$ is the product of the density of states $N_C(E)$ and $f_{FD}(E)$. The number density of electrons occupying the conduction band is given as

$$n = \int_0^{E_{top}} n(E)dE = \int_{E_C}^{E_{top}} N_C(E) f_{FD}(E) dE \tag{2.8}$$

where E_C and E_{top} are the bottom and top of the conduction band. Figure 2.3 relates the band diagram, density of states, Fermi-Dirac distribution and carrier concentrations for three types of semiconductor materials.

The density of states (DOS) can be approximated by the density near the bottom of the conduction band as [43]

$$N(E) = M_C \frac{\sqrt{2}}{\pi^2} \frac{\sqrt{E - E_C}}{\hbar^3} m_{de}^{3/2} \tag{2.9}$$

for low carrier concentrations and temperatures. $\hbar = h/2\pi$ is the reduced Planck's constant, M_C is the number of equivalent minima of the conduction band and m_{de} is the density-of-state mass of electrons which will be discussed later.

The bottom of the conduction band is assumed to have a parabolic shape as shown in Figure 2.2. The density of states is given as

$$E = E_C + \frac{\hbar^2 k^2}{2m_e^*}$$

where k and m_e^* are the momentum and effective mass of electrons. Electrons occupy the lowest energy level from the inner shell, the Fermi equal-energy surfaces, to higher

energies. Carriers transport at the bottom and top of the conduction and valence band respectively where is parabola-shaped.

By integrating over the density of states and probability of occupation for electrons described above, the electron concentrations and similarly for holes are calculated as

$$\begin{aligned} n &= N_C \tfrac{2}{\sqrt{\pi}} F_{1/2}(\tfrac{E_F - E_C}{k_B T}) = N_C e^{-(E_C - E_F)/k_B T} \\ p &= N_V \tfrac{2}{\sqrt{\pi}} F_{1/2}(\tfrac{E_V - E_F}{k_B T}) = N_V e^{-(E_F - E_V)/k_B T} \end{aligned} \tag{2.10}$$

where $F_{1/2}(x)$ is the Fermi-Dirac integral which is approximated to $\sqrt{\pi}e^x/2$ if E_F is smaller than E_C for non-degenerate semiconductors. N_C and N_V are the effective densities of states of the conduction and valance band.

By equating concentrations of n and p, the intrinsic carrier density n_i and intrinsic Fermi level E_i are obtained as

$$np = n_i^2 = N_C N_V e^{-(E_C - E_V)/kT} = N_C N_V e^{-E_g/kT} \tag{2.11}$$

$$E_F = E_i = \frac{E_C + E_V}{2} + \frac{kT}{2} ln(\frac{N_V}{N_C}) \tag{2.12}$$

where E_g is the band gap energy. The electron and hole density can be rewritten by n_i as

$$\begin{aligned} n &= N_C e^{-(E_C - E_i)/k_B T} e^{(E_F - E_i)/k_B T} = n_i e^{-(E_i - E_F)/k_B T} \\ p &= N_V e^{-(E_i - E_V)/k_B T} e^{(E_i - E_F)/k_B T} = n_i e^{-(E_F - E_i)/k_B T} \end{aligned} \tag{2.13}$$

under the complete ionization condition which states that the thermal energy is enough to ionise impurities in non-degenerate semiconductors. Note that the mass action law, $np = n_i^2$, holds for all cases at thermal equilibrium.

2.2.2 Effective Masses

The integrated effective density of states in the conduction band for electrons and with parameters for silicon are given as [117]

$$\begin{aligned} N_C &\equiv 2(2\pi m_{de} k_B T/h^2)^{3/2} M_C \\ N_C(m_e, T_e) &= 2.5094 \times 10^{19} (\frac{m_{de}}{m_0})^{3/2} (\frac{T_n}{300K})^{3/2} M_C \ cm^{-3} \ . \end{aligned} \tag{2.14}$$

The effective mass is a tensor with components m_{ij}^* defined as [116]

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2} \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j} \ . \tag{2.15}$$

The density-of-state mass for electrons m_{de} is a geometric mean of two transverse and one longitudinal masses is given as

$$m_{de} = (m_t^{*2} m_l^*)^{1/3} = [(\frac{0.1905 E_g(0K)}{E_g(T)} m_0)^2 (0.9163 m_0)]^{1/3} \tag{2.16}$$

which depends on the band gap or temperature indirectly. With $M_C = 6$ for silicon, $N_C(300K)$ is calculated as $2.80 \times 10^{19}\ cm^{-3}$.

Similarly, the integrated effective density of states in the valence band for holes and with parameters for silicon are given as [117]

$$\begin{aligned} N_V &\equiv 2(2\pi m_{dh} k_B T/h^2)^{3/2} \\ N_C(m_h, T_h) &= 2.5094 \times 10^{19} (\frac{m_{dh}}{m_0})^{3/2} (\frac{T_h}{300K})^{3/2}\ cm^{-3}\ . \end{aligned} \tag{2.17}$$

$N_V(300K)$ is calculated as $2.66 \times 10^{19}\ cm^{-3}$ which means E_f is slightly lower than mid-gap from Equation 2.12.

2.2.3 Fermi Level

The Fermi level is an upper bound for the most probable occupation of electrons whose density obeys the Fermi-Dirac statistics. It falls at mid-gap for an intrinsic case and $n = p = n_i$. When shallow donors or acceptors are introduced, which are fully ionised at room temperature to provide more charge carriers, the Fermi level adjusts to preserve charge neutrality.

For low temperature or a high doping concentration, the fraction of ionised donors and acceptors are given as

$$\begin{aligned} \frac{N_D^+}{N_D} &= \frac{1}{1 + g_D e^{E_F - E_D/k_B T}} \\ \frac{N_A^-}{N_A} &= \frac{1}{1 + g_A e^{E_A - E_F/k_B T}} \end{aligned} \tag{2.18}$$

where g_D is the degeneracy for donors which is 2 for silicon since a donor can accommodates one either spin and no electron. g_A is the degeneracy for acceptors and is 4 for silicon due to the doubly degenerate ground state as shown in Figure 2.2. The charge neutrality for a n-type material is written as

$$n + N_A^- \approx n = N_D^+ + p \tag{2.19}$$

and the Fermi level can be solved by replacing the carrier densities and ionised dopants

as

$$N_C e^{-\frac{E_C - E_F}{k_B T}} = N_D \frac{1}{1 + 2e^{E_F - E_D/k_B T}} + N_V e^{-\frac{E_F - E_V}{k_B T}} \ . \tag{2.20}$$

The Fermi level can be then solved graphically or numerically.

For room temperatures and a moderate doping concentration, dopants are assumed to be fully ionised. Along with the mass action law $np = n_i^2$, the carrier densities in a n-type material are given as

$$\begin{aligned} n_n = \frac{1}{2}[(N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2} \approx N_D \\ p_n = n_i^2/n_n \approx n_i^2/N_D \end{aligned} \tag{2.21}$$

and the Fermi level can be obtained as

$$E_F = E_C - ln(N_C/N_D) = E_i + ln(n_n/n_i) \ . \tag{2.22}$$

Similarly, the carrier densities in a p-type material are given as

$$\begin{aligned} p_p = \frac{1}{2}[(N_A - N_D) + \sqrt{(N_A - N_D)^2 + 4n_i^2} \approx N_A \\ n_p = n_i^2/p_p \approx n_i^2/N_A \end{aligned} \tag{2.23}$$

and the Fermi level can be obtained as

$$E_F = E_V + ln(N_V/N_A) = E_i - ln(p_p/n_i) \ . \tag{2.24}$$

2.3 Generation and Recombination

Generation is a process that an electron is excited to the conduction band due to thermal energy from the valance band where a hole is created. Recombination is an inverse process that an electron returns to the valence band and annihilates with a hole. At thermal equilibrium, the generation rate is equal to the recombination rate. When the thermal equilibrium condition is not satisfied, i.e. $np \neq n_i^2$, the material tends to preserve $np = n_i^2$ through these processes.

However, direct band-to-band transitions are unlikely to occur in silicon and germanium, except for high densities of electrons and holes. The minimum of the conduction band for group IV materials deviates from the maximum of the valance band by a k-vector in the Brillouin zone, i.e. an indirect band gap. Indirect transition requires phonon assistance which may be provided by stressed and/or thin silicon (band gap engineering).

2.3.1 Direct Recombination

Direct recombination, which is also called radiation recombination since it emits a photon, is important for a direct band gap material like GaAs. The recombination rate R_{dir} is given as

$$R_{dir} = C_{dir}(np - n_i^2) \tag{2.25}$$

where C_{dir} is the rate constant which is $10^{-10}\ cm^3 s^{-1}$ for GaAs [91]. It is usually not considered for silicon in simulation, but it becomes important at large injection values, e.g. power devices.

2.3.2 Auger Recombination

Auger recombination is similar to the direct combination, but it passes its energy to another carrier without emitting a photon. The carrier is then excited to a higher energy state or re-emitted in a form of thermal vibration. The recombination rate R_{Aug} is given as

$$R_{Aug} = (C_n n + C_p p)(np - n_i^2) \tag{2.26}$$

where C_n and C_p are transition coefficients which are 2.8×10^{-33} and $9.9 \times 10^{-32}\ cm^6 s^{-1}$ for silicon [91]. This is a three-particle process and only becomes important for large carrier densities at non-equilibrium conditions. The Auger effect, in which the third carrier is ejected from the orbit, is very rare to occur.

2.3.3 Indirect Recombination

Most indirect transitions occurs due to localised energy states created by lattice defects or doped impurities. Localised states act as intermediate steps in the band gap for carriers to undergo next transitions and they are called generation-recombination centres (recombination centres) [87]. Carriers interact with recombination centres through four processes which are depicted in Figure 2.4.

- Electron capture: an electron transits from the conduction band to an empty localised state. The total capture rate r_a is proportional to the electron density n in the conduction band and the density of empty localised states given as

$$r_a = C_n n N_t (1 - f(E_t)) = v_{nth} \sigma_n n N_t (1 - f(E_t)) \tag{2.27}$$

  where the electron capture probability C_n is proportional to the electron capture cross section σ_n and the electron thermal velocity v_{nth} at thermal equlibrium. $v_{nth}\sigma_n$ may be visualised as the volume swept by an electron per unit time [117].

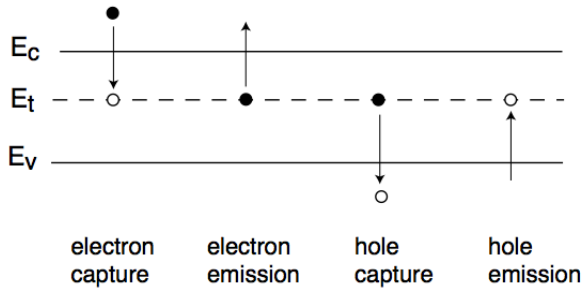


Figure 2.4: There are four processes for carriers to interact with with localised state E_t [87]. Filled and empty circles represent electrons and holes. Arrows represent transition directions.

The density filled recombination centres is the occupancy fraction $f(E_t)$ times the total density N_t and thus the empty localised states is given as $N_t(1-f(E_t))$.

- Electron emission: an electron transits from a localised state to the conduction band which is an inverse process of the electron capture. The total emission rate r_b is proportional to the occupied states $N_t f(E_t)$ given as

$$r_b = e_n N_t f(E_t) \tag{2.28}$$

where e_n is the electron emission probability. It is obtained by equating r_a and r_b at thermal equilibrium as

$$r_a = r_b = v_{nth}\sigma_n n N_t(1 - f(E_t)) = e_n N_t f(E_t)$$

which gives

$$e_n = v_{nth}\sigma_n n_i e^{(E_t - E_i)/k_B T} = v_{nth}\sigma_n N_C e^{(E_t - E_C)/k_B T} \ . \tag{2.29}$$

- Hole capture: an electron transits from an occupied localised state to the valence band and annihilates with a hole. The total capture rate r_c is given as

$$r_c = C_p p N_t f(E_t) = v_{pth}\sigma_p p N_t f(E_t) \tag{2.30}$$

where v_{pth} is the hole thermal velocity and σ_p is the hole capture cross section.

- Hole emission: an electron transits to an empty localised state from the valence band where a hole is created. The total emission rate r_d is given as

$$r_d = e_p N_t(1 - f(E_t)) \ . \tag{2.31}$$

where e_p is the hole emission probability. It is obtained by equating r_c and r_d at

thermal equilibrium and is

$$e_p = v_{pth}\sigma_p n_i e^{(E_i - E_t)/k_B T} = v_{pth}\sigma_p N_V e^{(E_V - E_t)/k_B T} \ . \tag{2.32}$$

2.3.4 Shockley-Read-Hall Recombination

The principle of detailed balance states that the number of electrons entering (r_b) and leaving (r_a) the conduction band must be equal in steady state [117]. For a non-equilibrium case in a n-type material where a uniform excitation G_l from illumination is applied, the time differentials of the electron and hole density in the conduction and valence band are given as

$$\begin{aligned} \frac{n_n}{dt} &= G_l - (r_a - r_b) = 0 \\ \frac{p_n}{dt} &= G_l - (r_c - r_d) = 0 \end{aligned} \tag{2.33}$$

in which

$$G_l = r_a - r_b = r_c - r_d \equiv R_{sp} - G_{sp} \equiv U \tag{2.34}$$

where U is the net recombination rate in steady state. R_{sp} and G_{sp} are the spontaneous recombination and generation rates which balance the external simulation. Note that an equilibrium case gives $G_l = 0$, $r_a = r_b$ and $r_c = r_d$.

Inserting equations of capture and emission rates for electrons and holes to Equation 2.34 gives

$$\begin{aligned} & v_{nth}\sigma_n n N_t(1 - f(E_t)) - N_t f(E_t) v_{nth}\sigma_n N_C e^{(E_t - E_C)/k_B T} \\ = \ & v_{pth}\sigma_p p N_t f(E_t) - N_t(1 - f(E_t)) v_{pth}\sigma_p N_V e^{(E_V - E_t)/k_B T} \end{aligned} \tag{2.35}$$

which is rearranged to give the trap occupancy function $f(E_t)$ as

$$f(E_t) = \frac{n + \frac{v_{pth}\sigma_p}{v_{nth}\sigma_n} N_V e^{\frac{E_V - E_t}{kT}}}{N_C e^{\frac{E_t - E_C}{kT}} + \frac{v_{pth}\sigma_p}{v_{nth}\sigma_n} N_V e^{\frac{E_V - E_t}{kT}} + n + \frac{v_{pth}\sigma_p}{v_{nth}\sigma_n} p} \ . \tag{2.36}$$

The net recombination rate is obtained by inserting $f(E_t)$ to the same equation as

$$U = \frac{v_{th}\sigma_n\sigma_p N_t(pn - n_i^2)}{\sigma_p[p + n_i e^{(E_i - E_t)/k_B T}] + \sigma_n[n + n_i e^{(E_t - E_i)/k_B T}]} \tag{2.37}$$

where $v_{th} = v_{nth} = v_{pth}$ and the band dependent terms are written in term of n_i and E_i. It can be written in a simple form as

$$U = \frac{pn - n_i^2}{\tau_n(p + p_0) + \tau_p(n + n_0)} \tag{2.38}$$

where $n_0 = n_i e^{(E_t - E_i)/kT} = N_C e^{(E_t - E_C)/kT}$ and $p_0 = n_i e^{(E_i - E_t)/kT} = N_V e^{(E_V - E_t)/kT}$ as shown in the previous section.

For low injection conditions, τ_n and τ_p are minority lifetimes for electrons in a p-type region written as

$$U \approx \frac{n - n_0}{\tau_n} \quad \longrightarrow \quad \tau_n = \frac{1}{v_{thn}\sigma_n N_t} \tag{2.39}$$

and for holes in a n-type region written as

$$U \approx \frac{p - p_0}{\tau_p} \quad \longrightarrow \quad \tau_p = \frac{1}{v_{thp}\sigma_p N_t} \ . \tag{2.40}$$

Note that $n_0 = p_0 = n_i$ for a mid-gap recombination centre. Kurata is a simulator using a representative mid-gap defect which will be described in Chapter 4. The net recombination rate can be further simplified by setting $\tau_n = \tau_p = \tau_0$ as

$$U = \frac{pn - n_i^2}{\tau_0[p + n + 2n_i Cosh(\frac{E_t - E_i}{kT})]} \ . \tag{2.41}$$

2.3.5 Surface Recombination

Dangling bonds on an abrupt semiconductor surface generate many localised states which enhance the generate rate. The surface recombination U_s is equivalent to the SRH recombination U_{SRH} at surface regions which can be expressed in the same form as [115]

$$U_s = \frac{pn - n_i^2}{(n_s + N_0)/s_p + (p_s + p_0)/s_n} \tag{2.42}$$

where s_n and s_p are the surface recombination rates for electrons and holes. It can be approximated for a low-injection p-type region as [117]

$$U_s \approx v_{th}\sigma_p N_{t,s}(p_s - p_0) \equiv s_l(p_s - p_0) \tag{2.43}$$

where s_l is the surface recombination rate for holes at low injections.

2.3.6 Total Recombination

In principle, three types of recombinations affect each other which cannot be modelled in current theories. The total recombination rate for electrons U_n or holes U_p is usually written in the sum of the three as

$$U = U_n = U_p = U_{dir} + U_{Aug} + U_{SRH} \tag{2.44}$$

in device simulation. Kurata considers only U_{SRH} and TCAD, which is a commercial device simulator, includes surface, Auger and other recombinations in addition.

2.4 Semiconductor Equations

Basic semiconductor equations describe the static (steady state) and dynamic (time dependent) behaviour of carriers in response to an external field. Three groups, the Maxwell equations, current-density equations and continuity equations, are discussed in this section which are the main body of device simulation [116][68].

2.4.1 Maxwell's Equations

Maxwell's equations describe the dynamics of electric charges and current if an electric field E or magnetic field B is present. Gauss' law, especially for pn junctions, is the most important and is used directly in device simulation. It is give as

$$\nabla \cdot \mathbf{D} = \rho(x, y, z) \quad \longrightarrow \quad \nabla \cdot (\epsilon_{Si}\mathbf{E}) \tag{2.45}$$

where $\mathbf{D}$ is the displacement vector and $\rho(x, y, z)$ is the space charge density. It can be related to the electric field $\mathbf{E}$ for static or low frequency cases.

It can also be written as a second-order partial differential equation

$$\nabla^2 V = -\nabla \cdot \mathbf{E} = -\frac{\rho(x, y, z)}{\epsilon_{Si}} \tag{2.46}$$

where V is the electric potential. This is Poisson's equation for silicon or other semiconductors by replacing the permittivity constant ϵ_{Si}.

2.4.2 Current-Density Equations

Current flows have two contributions which are the drift process caused by the electric field and the diffusion process caused by the carrier concentration gradient. The drift contribution dominates when there is a large electric field, and vice versa. The electron current density J_n and the hole are then given as

$$\begin{aligned} \mathbf{J}_n &= qD_n\nabla n + q\mu_n n\nabla V = q\mu_n(n\mathbf{E} + \frac{k_BT}{q}\nabla n) \\ \mathbf{J}_p &= -qD_p\nabla p + q\mu_p p\nabla V = q\mu_p(p\mathbf{E} - \frac{k_BT}{q}\nabla p) \end{aligned} \tag{2.47}$$

where D_n and D_p are the carrier diffusion constant for electrons and holes. They are given by the Einstein relation $D_n = (k_BT/q)\mu_n$ and $D_p = (k_BT/q)\mu_p$ where μ_n and μ_p are the mobilities for electrons and holes respectively. Carrier transport phenomenons, such as drift, diffusion and impact ionisation, will be discussed in Chapters 4 and 5. The total current density is the sum of the two, $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$.

2.4.3 Continuity Equations

Carrier concentrations remain constant in steady state. External stimulations, such as optical excitation or impact ionisation, disturb the equilibrium. Carriers will redistribute to balance the change and this process can be modelled by transient simulation. The time derivatives of the electron and hole concentration are the sum of their current densities ($\mathbf{J}_{n/p}$), generation $G_{n/p}$ and recombination $R_{n/p}$ rates given as

$$\begin{aligned} \frac{\partial n}{\partial t} &= \frac{1}{q}\nabla \cdot \mathbf{J}_n + (G_n - R_n) = 0 \\ \frac{\partial p}{\partial t} &= -\frac{1}{q}\nabla \cdot \mathbf{J}_p + (G_p - R_p) = 0 \end{aligned} \tag{2.48}$$

which are zero in steady state.

2.5 *pn* Junction Diode

A *pn* junction is the boundary between *p*-type and *n*-type regions in a single-crystal semiconductor. Doped regions are generated by ion implantation and dopant diffusion (Chapter 4) or epitaxy. *pn* junction diodes are very important for understanding and applying semiconductor physics. It is a basic element of advanced applications, such as silicon diode detectors, light-emitting diodes (LEDs), solar cells and transistors.

3D sensors use the detection concept of the planar technology which are a *p*+-*p*-*n*+ (current) or *p*+-*n*-*n*+ (early design) diode. The bulk uses high purity (resistivity) to reduce the full depletion voltage. For an un-irradiated case with a low generation rate, modelling of 3D sensors can be treated as an electrostatic problem at thermal equilibrium. The theoretical solution is derived and basic properties of a *pn* junction are discussed in this section.

2.5.1 Electrostatic Approximation

The Poisson equation in steady state gives an approximated solution for electrostatic studies of 3D or planar sensors. The charge density ρ is the sum of (fixed) doping and free carrier concentrations and gives

$$\nabla^2 V = -\frac{q}{\epsilon_s}(N_D - N_A + p - n) \tag{2.49}$$

where N_D and N_A are the structure dependent doping concentrations of donors and acceptors. n and p are the concentrations free electrons and holes which interact with the electric field E to reach an unique space charge distribution ρ_s.

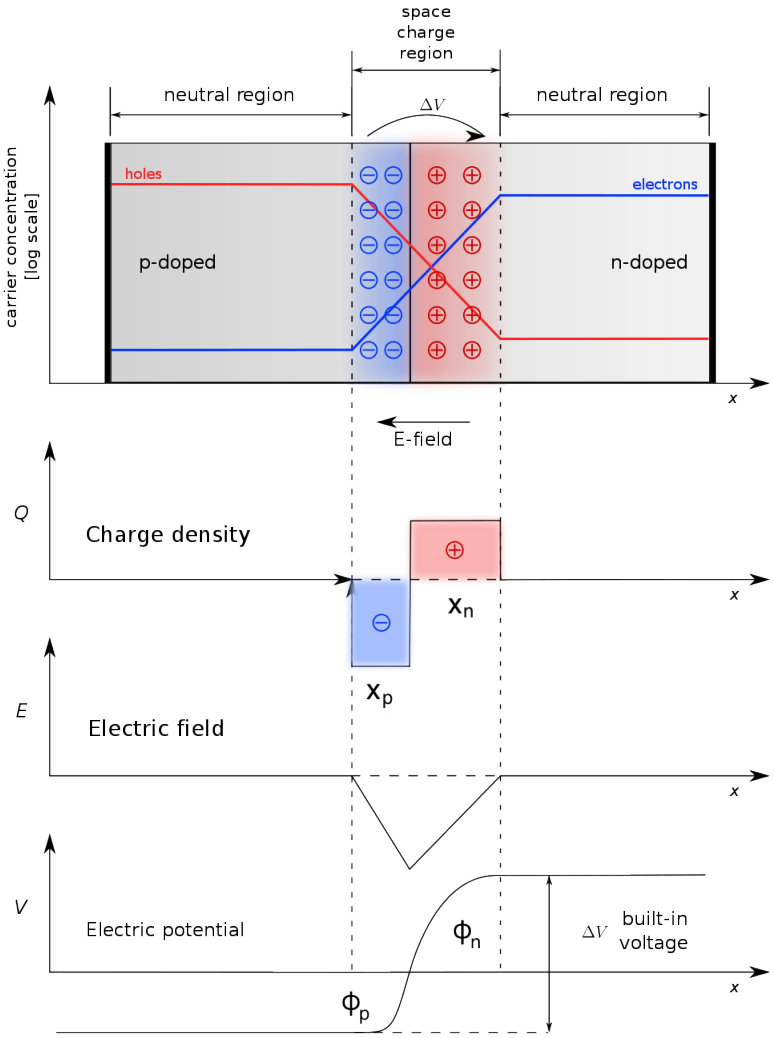


Figure 2.5: The depletion region of a *pn* junction diode and the corresponding built-in field (upper), charge distribution, electric field and electric potential (lower) [4]. The electric field tend to move electrons and holes.

2.5.2 Theoretical Model for a *pn* Junction Diode

Three regions are formed across a *pn* junction: the *p*-type ($p \gg n$) and *n*-type ($n \gg p$) are charge neutral regions where dopants provide free carriers. The depletion region is formed between two neutral regions by the fixed ions, donors and acceptors, since electrons transport from the *n*-type to *p*-type region. The space charge distribution, electric field and potential across a *pn* junction are shown in Figure 2.5.

The Fermi level E_F at doped regions deviates from mid-gap and the local potentials

$\phi_{n/p}$ are defined as

$$\begin{aligned} \phi_n &\equiv -\tfrac{1}{q}(E_i - E_F)\,|_{x \geq x_n} &&= \frac{kT}{q} ln(\frac{N_D}{n_i}) \\ \phi_p &\equiv -\tfrac{1}{q}(E_i - E_F)\,|_{x \leq -x_p} &&= -\frac{kT}{q} ln(\frac{N_A}{n_i}) \end{aligned} \tag{2.50}$$

for n-type and p-type regions with corresponding depletion width x_p and x_n (Figure 2.5) at thermal equilibrium. The Fermi level has to be identical across the diode which results in the energy band bending at the junction. ϕ_n and ϕ_p set up a potential difference as

$$V_{bi} = \phi_n - \phi_p = \frac{kT}{q} ln(\frac{N_A N_D}{n_i^2}) \; . \tag{2.51}$$

which is called the built-in potential V_{bi}.

Assuming the full depletion condition, $n = p = 0$, and an abrupt junction (the square doping approximation), the Poisson equation and the depletion widths are simplified as

$$\nabla^2 V = -\frac{q}{\epsilon_s}(N_D - N_A) \quad \text{and} \quad N_A x_p = N_D x_n \tag{2.52}$$

with the overall space charge neutrality. The total depletion width W is obtained by solving the Poisson equation in a function of the built-in potential as [116]

$$W = x_p + x_n = \sqrt{\frac{2\epsilon_s}{q}(\frac{N_A + N_D}{N_A N_D}) V_{bi}} \; . \tag{2.53}$$

Note that there is no current flow without a bias at thermal equilibrium.

If a bias is applied, the Fermi level on both sides is no longer the same but separated. For a reverse bias V_{rb} which increases (decreases) the voltage on $n+$ ($p+$) electrode, the depletion width is widened as

$$W = \sqrt{\frac{2\epsilon_s}{q}(\frac{N_A + N_D}{N_A N_D})(V_{bi} + V_{rb})} \; , \tag{2.54}$$

where the carriers are swept to the electrodes quickly. For a p-type bulk, the depletion capacitance C per unit area can be estimated as [116]

$$C = \frac{dQ}{dV} = \frac{d(qN_A W)}{d[(qN_A/2\epsilon_{So})W^2]} = \frac{\epsilon_{Si}}{W} = \sqrt{\frac{q\epsilon_{Si} N_A}{2}(V_{bi} + V_{rb})} \tag{2.55}$$

before the full depletion is reached[2]. Silicon sensors are operated in depletion, so there is always a capacitance which depends on the structure and geometry.

[2]This equation is also true for a forward bias with an additional diffusion capacitance.

Chapter 3

Physics of Defects

An ideal crystalline solid has a periodic crystal structure. A real solid unavoidably exhibits crystallographic defects which break the perfect periodicity. For particle physics applications, the radiation environment in experiments introduces a large amount of defects in detector materials. Radiation is especially severe for vertex detectors which are close to the interaction point. To predict the detector performance after irradiation, simulation has to incorporate the defect physics and semiconductor physics. This chapter describes the defect physics, defect modelling and their effects on detectors.

3.1 Radiation Interactions

There are four major categories of radiation [63]: heavy charged ions ($\lambda \sim 10^{-5}\ m$), fast electrons ($\lambda \sim 10^{-3}\ m$), neutrons ($\lambda \sim 10^{-1}\ m$) and X/γ rays ($\lambda \sim 10^{-1}\ m$), where λ is the characteristic distance of penetration for typical energies in solids. λ depends on the atomic number and radiation energy. Charge particle radiation has a much shorter penetration length due to its continuous interaction by the Coulomb force.

As charged radiation penetrates a material, it transfers energy to the atomic electrons and decelerates. Based on the collision dynamics (proximity), the deposited energy may be sufficient to excite atomic electrons to higher energy state (excitation), or even to remove electrons from their orbit (ionisation). Scintillators produce light emitted as electrons transit from excited states to lower states. Ionisation chambers detect the ionised gases (ions plus electrons) migrating to the electrodes in the large electric field.

The generation of electron-hole pairs in semiconductor materials can also be categorised as ionisation. The ionising energy is around 3.62 (300 K) and 2.96 (77 K) eV [63] for silicon and germanium respectively. A minimum ionising particle (MIP) creates around 80 electron-hole pairs in silicon per micron on average. However, this process may not be uniform but have a large variance (clumpiness) [128].

3.1.1 Stopping Power

The linear stopping power S for radiation particles is defined as dE/dx, which is called the specific energy loss. For charged particles, the energy loss can be written as [130]

$$\frac{dE}{dx} = -N_{tar}[S_e(E) + S_n(E) + S_r(E)] \tag{3.1}$$

where N_{tar} is the target density with contributions from electronic, nuclear and radiative interactions respectively.

The electronic stopping S_e is caused by the inelastic collisions between incident particles and material electrons. It is well defined by the Bethe formula for heavy ions[1] and dominates at energies higher than 100 keV. The Bethe formula (S_e) is a function with a broad minima and thus heavy ions with different energies have similar energy loss. MIPs are particles with energies and $\langle dE/dx \rangle$ corresponding to the minimised ionisation, e.g. muons with $1 < \beta\gamma < 10$ in copper.

The nuclear stopping S_n is caused by elastic collisions between radiation particles and target atoms which are more important for lower energies. For electrons and ions at very high energies or in high atomic number materials, the radiative stopping S_r needs to be considered. Electrons are decelerated by emitting photons which is called bremsstrahlung [63].

3.1.2 Displacement Energy

A radiation particle losses most of its energy via electronic stopping which does not deviate the particle path on average. It usually goes through many atomic layers until its energy is low, then it has a higher probability for collisions with the nuclei. Once an atom is removed from its lattice, its original lattice site and itself form a vacancy-interstitial pair, a Frenkel pair. This primary knock-on atom (PKA) will also produce a cascade of collisions if the energy transfer is enough. Figure 3.1 shows a schematic of a recoil atom which cascades in silicon and generates defect clusters [59][124]. The typical value of the displacement energy E_d for silicon is around 15∼25 eV.

[1] The Bethe-Bloch equation for charged particles is give as

$$-\langle\frac{dE}{dx}\rangle = K\frac{Z}{A}\frac{z^2}{\beta^2}[\frac{1}{2}ln(\frac{2m_ec^2\beta^2\gamma^2E_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2})] \tag{3.2}$$

where $\beta = v/c$ in special relativity gives a large $1/\beta^2$ factor for slow protons. The detailed parameters are listed in the Particle Physics booklet [120]. The Bethe formula describes the mean rate of energy loss with $0.1 < \beta\gamma < 10^4$. MIPs are particles at minima values by plotting the energy dependence of the Bethe energy loss.

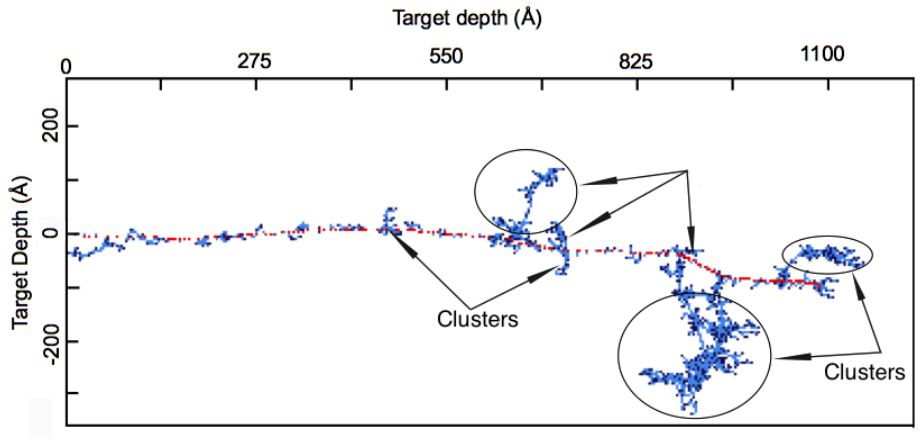


Figure 3.1: TRIM simulation of a damage cascade by a PKA with $E_R = 50\ keV$ which is the average energy transfer from a 1 MeV neutron. The PKA track and displaced silicon atoms (interstitials/vacancies) are shown in red and in blue respectively [59].

3.1.3 Binary Collision between Source and Target

The binary collision approximation (BCA) simplifies each collision process such that only two charged particles are involved. It is used in simulation to calculate the penetration depth and the corresponding energy transfer. A well known Monte Carlo simulator is SRIM/TRIM which was developed by Ziegler. He gave a semi-empirical formula for the screening function with Biersack and Littmark [17].

A non-relativistic primary atom is set to have a mass m_p, energy E_p and momentum p which is injected into a silicon target with a mass m_{Si}. The scattering angle θ and recoil angle ϕ in the laboratory system and the recoil energy E_R of the knock-on silicon atom are given as [130]

$$\theta = tan^{-1}(\frac{sin\Theta}{m_p/m_{Si} + cos\Theta}) \tag{3.3a}$$

$$\phi = cos^{-1}(sin\frac{\Theta}{2}) \tag{3.3b}$$

$$E_R = \frac{4m_p m_{Si}}{(m_p + m_{Si})^2} sin^2(\frac{\Theta}{2}) E_p\ . \tag{3.3c}$$

The scattering angle Θ in the centre-of-mass frame is given as

$$\Theta(p, E_p) = \pi - 2p \int_{r_{min}}^{\infty} \frac{dr}{r^2 \sqrt{1 - V(r)/E_p - p^2/r^2}} \tag{3.4}$$

where $V(r)$ is the repulsive Coulomb potential which treats repulsive potentials from other nuclei/electrons as screening.

The maximum energy transfer T_M is given as

$$T_M = \frac{2m_{Si}E_p(E_p + 2m_pc^2)}{(m_p + m_{Si})^2 + 2m_{Si}E_p} \tag{3.5}$$

which includes relativist effects. T_M are

$$T_{M,neutron} = \frac{4m_nm_{Si}E_n}{(m_n + m_{Si})^2} \quad \text{and} \quad T_{M,electron} = \frac{2(E_e + 2m_e)E_e}{m_{Si}} \tag{3.6}$$

for an incident neutron and electron respectively. Thus the minimum energy for neutrons or protons to create a Frenkel pair is about 110 eV, while it is about 260 keV for electrons due to their lighter mass.

3.2 Defect Estimation

Charged particles lose most of their energy interacting with electron clouds in reversible forms of excitation and ionisation. Charged particles can also create knock-on atoms due to the Coulombs interaction with the nuclei. This produces a single Frenkel pair. Protons and neutrons also interact with nuclei via the strong interaction and produce a PKA with significant energy which creates further knock-on atoms, and that are defect clusters. Thus neutrons tend to produce defect clusters while protons tend to generate defect clusters and point defects. The non-ionising energy loss (NIEL) scaling and the Kinchin-Pease model are important to estimate first-order defect production. The resulting radiation effects on sensors can be then modelled.

3.2.1 NIEL Scaling

The NIEL hypothesis relates the damage generated by any particle radiation to an equivalent 1 MeV neutron fluence Φ_{eq}^{1Mev} with linear coefficients as [125]

$$\Phi_{eq}^{1Mev} = \kappa\Phi = \kappa \int_0^\infty \phi(E)dE \tag{3.7}$$

where κ is the hardness parameter. It is defined as a total introduction rate ratio of an incident particle (labelled $_p$) with energy spectrum $\phi(E_p)$ to 1 MeV neutrons:

$$\kappa = \frac{1}{D_{neutron}(1MeV)} \frac{\int_0^\infty D_p(E_p)\phi(E_p)dE_p}{\int_0^\infty \phi(E_p)dE_p} \tag{3.8}$$

where D_p is the displacement damage cross section. Different reactions for three main isotopes of silicon, ^{28}Si, ^{29}Si and ^{30}Si with natural abundances 92.23%, 4.67% and 3.09% respectively, have to be considered.

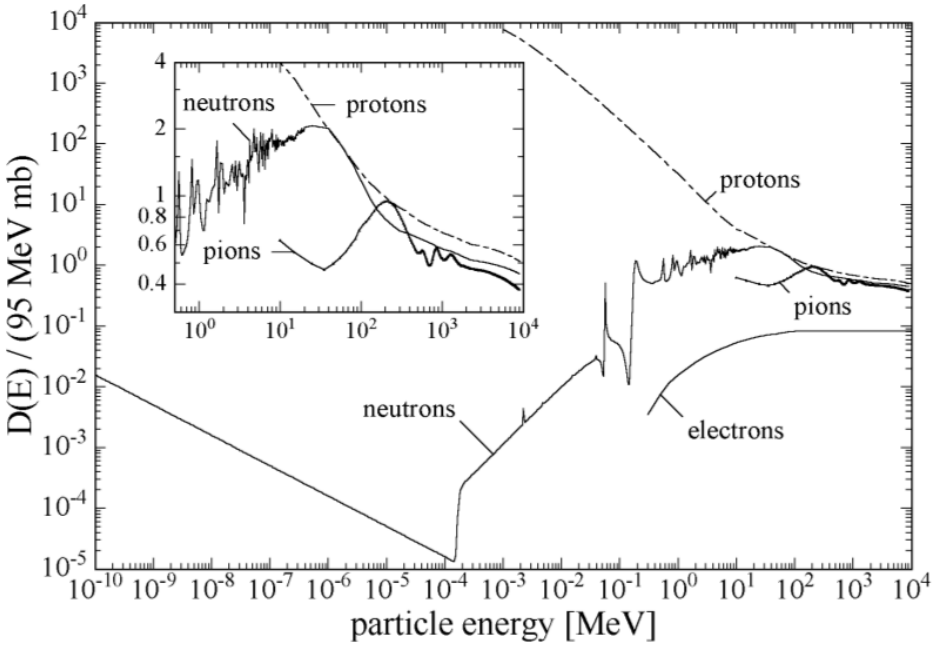


Figure 3.2: The displacement cross section of various projectiles at different energies and the inset is a zoomed display of the high energy region [84]. The cross sections are calculated theoretically and normalised by 1 MeV neutrons which then has $D(E)/(95MeVmb) = 1$. Neutrons with energies lower than 10^{-4} MeV are thermal neutrons and tend to be captured by nuclei for the fission process. This is useful for irradiation normalisation.

The displacement damage cross section D_p at an energy E_p is given as

$$D_p(E_p) = \Sigma_i \sigma_i(E_p) \int_{E_{R_{min}}}^{E_{R_{max}}} dE_R f_i(E_p, E_R) P(E_R) \ . \tag{3.9}$$

where $P(E_R)$ is the Lindhard partition function which describes the energy loss from a recoil nucleus at an energy E_R in non-ionising (displacement) processes. The index i sums over all possible interactions based on different probabilities f_i for a PKA generation as a function of E_p and E_R.

Moll [84] summarises the displacement damage for different energy regimes from various experiments. Figure 3.2 shows the normalised displacement damage function to 95 $MeVmbarn$ for neutrons, protons, pions and electrons. Note that the displacement damage for 1 MeV neutrons is given as $D_n(1MeV) = 95MeVmb$.

3.2.2 Modified Kinchin-Pease Model

Kinchin and Pease estimated the number of displacement atoms N_d (Frenkel pairs) generated as a function of the PKA energy [61]. Norgett and Robinson gave more explicit parameters [103][88]. The upper energy limit E_u, which specifies pure energy

loss to electron excitation, is absorbed into the energy transfer function. The original Kinchin-Pease model is summarised as

$$N_d = \begin{cases} 0\ , & 0 < E_\nu < E_d \\ 1\ , & 0 < E_\nu < 2E_d \\ E/2E_d\ , & 2E_d < E_\nu < E_u \\ E_u/2E_d\ , & E_\nu > E_u \end{cases} \tag{3.10}$$

which is linear and flat saturates smaller and larger than E_u. It is modified as

$$N_d = \begin{cases} 1\ , & E_d < E_\nu < 2.5E_d \\ 0.8E/2E_d\ , & E_\nu > 2.5E_d \end{cases} \tag{3.11}$$

with $E_d = 15\ eV$. The inelastic energy loss E_ν is calculated according to the integral method which was introduced by Lindhard [72] with a numerical approximation to the universal function $g(\epsilon_d)$ and reduced energy ϵ_d.

They are written explicitly as damage energy E_ν

$$E_\nu = \frac{E_R}{1 + k_d g(\epsilon_d)} \tag{3.12}$$

and other parameters [55]

$$k_d = 0.133745 \frac{Z_1^{2/3}}{a_1^{1/2}} \tag{3.13a}$$

$$g(\epsilon_d) = 3.4008\epsilon_d^{1/6} + 0.40244\epsilon_d^{3/4} + \epsilon_d \tag{3.13b}$$

$$\epsilon_d = (\frac{9\pi^2}{128})^{1/3} \frac{a_0}{Z_1 Z_2 q^2} (Z_1^{2/3} + Z_2^{2/3})^{-1/2} \frac{a_2 E_R}{a_1 + a_2} \xrightarrow{Z_2 = Z_1 (a_2 = a_1)} 0.0115 Z_2^{-7/3} E_R \tag{3.13c}$$

where a_0, $Z_1(a_1)$ and $Z_2(a_2)$ are the Bohr radius and atomic (mass) number of the projectile and target atom respectively. Figure 3.3 shows the number of Frenkel pairs generated as a function of the PKA energy with the Kinchin-Pease and modified model.

3.3 Types of Defects

Irregular defects can be classified as point, line, planar and bulk defects. Thermally-grown silicon wafers are well-crystallized and thus only point and bulk defects are discussed. Defects can be also categorised into shallow- and deep-levels by their energy levels or grouped as acceptors and donors by their charge states. Shallow defects are normally associated with dopants which minimally distorts the lattice. Deep defects

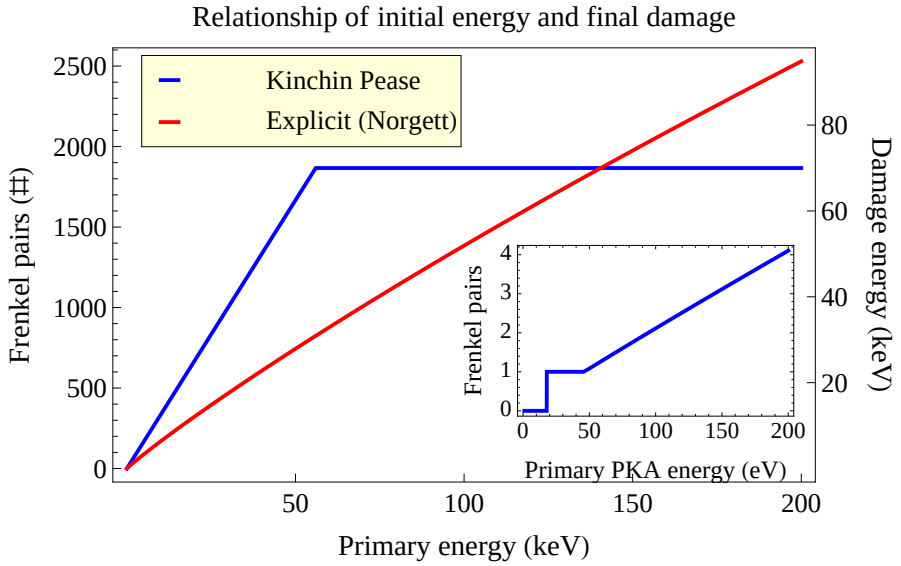


Figure 3.3: Frenkel pairs generated as a function of the PKA energy and the inset is a zoomed display of the low energy region. The Kinchin-Pease model has a simplified linear relationship with saturation and the modified model by Norgett is from theoretical calculation which gives a quite different behaviour.

break the lattice periodicity which results in different energy levels and charge states. Figure 3.4 shows a simple diagram of these, e.g. C_iO_i is a deep level and a donor; V_2 is also a deep level with more than one charge state.

3.3.1 Point Defects

Basic point defects are vacancies, self/foreign interstitials and substitutional atoms. Vacancies are lattice sites where their original atoms are removed. Interstitials occupy non-lattice sites in the crystal. Silicon self interstitials are referred to hereafter as interstitials. Figure 3.5 (upper) shows a knocked-out atom from its lattice site becomes an interstitial I and leaves a vacancy V. A nearby pair of V and I forms a Frenkel pair (defect). Atoms can be substituted by other atoms, e.g. carbon, boron or phosphorous atoms in silicon, which are called a substitutionally dissolved atom A_S or substitutional defect.

For electron or gamma radiation, the minimum incident energy is high but the recoil energy is low due relatively massive silicon atoms. They mainly induce points defects. Atoms can interchange their positions in a compound which forms anti-site defects, e.g. Ga_{As} and As_{Ga} in $GaAs$. Dangling bonds can also be considered as a defect. They form at the interface between the silicon lattice, at cut edges or Si/SiO_2 interface.

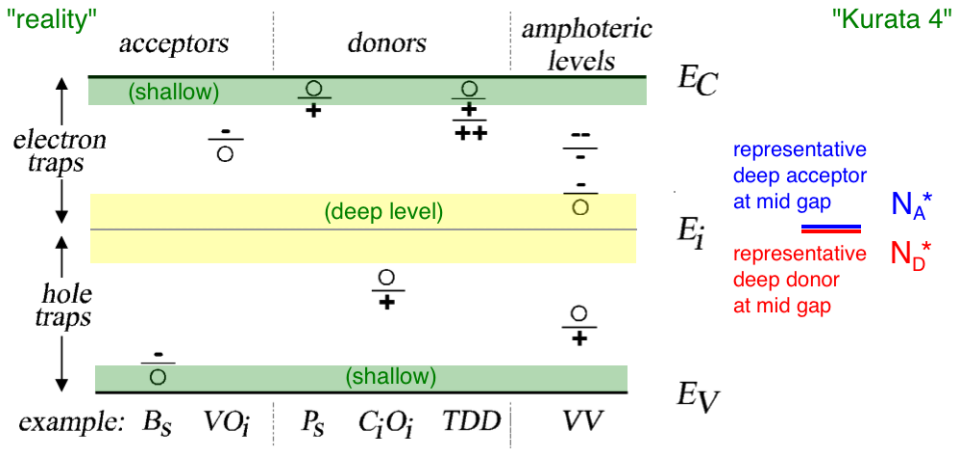


Figure 3.4: A schematic illustration of various electrical properties, energy levels and charge states of defects [84]. The idea of representative mid-gap deep levels is also shown.

3.3.2 Bulk Defects

For neutron and proton radiation, the recoil energy of a PKA is enough to create more Frenkel pairs, a PKA cascade. Higher order vacancies, V_2, V_3... V_n, exist and collect together to form a void region in the solid, a cluster of vacancies, as shown in Figure 3.5 (middle). Similarly, higher order interstitials, I_2, I_3... I_n, exist and form a dense precipitate in the solid, a cluster of interstitials as Figure 3.5 (middle) too. The cluster model was proposed by Gossick [45] to explain the high minority carrier recombination rate observed in irradiated devices by neutrons compared to electrons or gammas [84]. There is no cluster formed by gamma radiation which suggests the cluster region is the main source of the leakage current in neutron irradiated materials.

Divacancy-oxygen V_2O is considered as a key contributor of the effective doping concentration. Other defects (VO, V_2, C_i, C_iO_i, C_iC_s and IO_2) were investigated but no obvious correlation was observed [98]. Simulation shows that the V_2O production can be suppressed by a high oxygen concentration in the substrate [77][76]. Experiments observed this reduction is more prominent for proton irradiated devices where more point defects are generated [73]. This is due to low energy Coulomb interactions with the nuclei that do not occur with uncharged neutrons. By introducing more oxygen in the substrate, vacancies tend to form $VO's$ which are an acceptor away from mid-gap and is inactive as a generation centre.

3.3.3 Shallow- and Deep-Level Defects

Shallow defects usually refer to implanted or diffused dopants which are ionised at room temperature. Moll suggested that a defect level which is 70 meV below (above) the conduction (valence) band is a shallow-level based on the detectability of capacitance

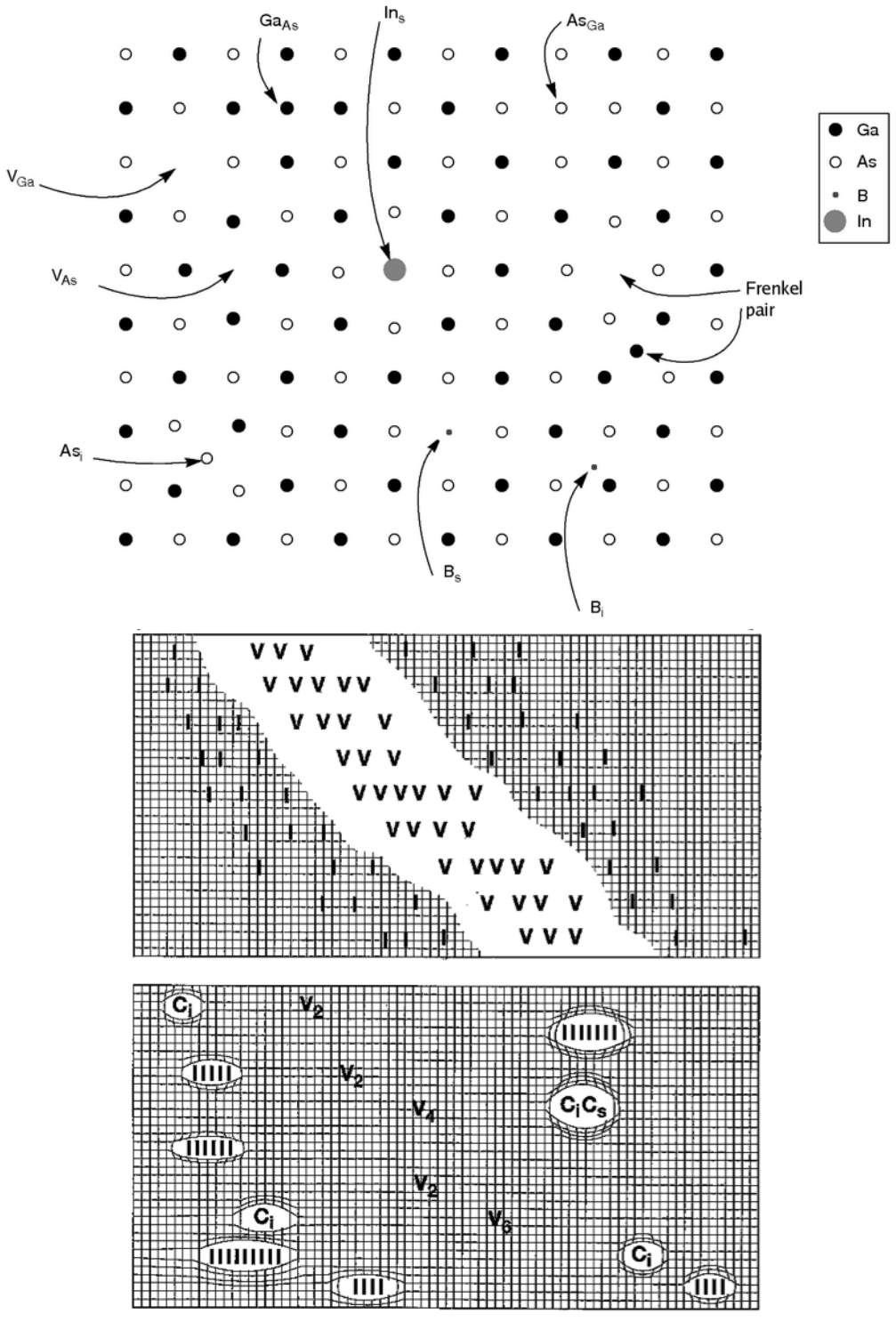


Figure 3.5: Schematic illustrations of point in a compound solid of *GaAs* [2] (top) and bulk defects in a single crystal [54] (middle and bottom). The top scheme depicts different vacancies, self/foreign interstitials and anti-site defects. The bottom schematic, which illustrates higher order vacancies and interstitials, is evolved from the middle one where a void region is formed by a displacement tree.

deep-level transient spectroscopy (C-DLTS) [84]. The smearing of the Fermi-Dirac distribution extends to around 20 meV at 300 K. Effective deep-level defects are located near mid-gap and dominate the Fermi level for unbiased materials.

Shallow defects, like phosphorous, can be treated as a hydrogen-like system. Four covalent bonds (long-ranged potential) are stable and ignored in the Hamiltonian. The extra electron from a donor is assumed to be orbiting the phosphorous atom and the ionisation energy is given as

$$E_D = (\frac{\epsilon_0}{\epsilon_{Si}})^2(\frac{m_n}{m_0})E_{Hydrogen} \tag{3.14}$$

where $E_H = -13.6/n^2\ eV$ is energy levels of the Bohr model. E_D is calculated and measured as around 25 meV for silicon. An acceptor also gives an extra orbiting hole.

Deep defects (with ionisation energy $\geq 3k_BT$) cannot be modelled by the hydrogen model and effective mass theory (approximation) due to the large lattice distortion as shown in Figure 3.5 (bottom). Deep defects usually have the opposite ordering, donors at the bottom half and acceptors at the top half of the band gap, due to their chemical nature. Their energy levels are calculated from first principles which considers the short-ranged potential (lattice distortion) and assisted by measured values. Note that shallow deep levels are possible, e.g. V_6 is very close to the conduction band.

3.3.4 Acceptors, Donors and Amphoteric Defects

Defects with energy levels in the band gap can capture and emit electrons and holes. An acceptor is negatively charged when occupied by an electron and neutral when empty. It is written as $(-/0)$ (filled/empty), e.g. boron is an acceptor. A donor is neutral when occupied by an electron and positively charged when empty. It is written as $(0/+)$ (filled/empty), e.g. phosphorous is a donor. Some defects are amphoteric defects which have more than one energy level and charge state, e.g. V_2 have several charge states ($--/-$, $-/0$ and $0/+$). According to the Fermi level, acceptors in the upper or donors in the lower half of the band gap are usually neutral which means that acceptors (donors) are empty (filled) with electrons.

3.4 Defect Mechanism

Many physical (e.g. diffusion) and chemical (e.g. compound formation) reactions occur as soon as defects are formed. Defects within clusters evolve with time and their final states depend on the temperature and various other parameters. The defect concentrations and mobilities determine the reaction speed which is also affected by the impurity concentrations. Main impurities are carbon and oxygen and their concentrations are

I Reactions	r_c (Å)	V Reactions	r_c (Å)	C Reactions	r_c (Å)
$I + V \longrightarrow \emptyset$	16.0			$I + C_S \longrightarrow C_I$	7.4
$I + I \longrightarrow I_2$	7.9	$V + V \longrightarrow V_2$	7.7	$C_i + C_s \longrightarrow C_iC_s$	
$I + I_n \longrightarrow I_{n+1}$		$V + V_n \longrightarrow V_{n+1}$		$C_i + O \longrightarrow C_iO$	
$I + O \longrightarrow IO$		$V + O \longrightarrow VO$	5.0	$V + ICC \longrightarrow CC$	8.6
$I + IO \longrightarrow I_2O$		$V + VO \longrightarrow V_2O$	8.4	$V + ICO \longrightarrow CO$	10.6
$I + I_2O \longrightarrow I_3O$		$V + V_2O \longrightarrow V_3O$		$I + CC \longrightarrow ICC$	14.2
$I + VO \longrightarrow O$	8.6	$V + IO \longrightarrow O$		$I + CO \longrightarrow ICO$	11.3
$I + V_2O \longrightarrow VO$	5.1	$V + I_2O \longrightarrow IO$			
$I + V_3O \longrightarrow V_2O$	11.7	$V + I_3O \longrightarrow I_2O$			

Table 3.1: The capture radii used in the Monte Carlo simulation for different reactions related to interstials, vacancies and carbons [56]. Values of r_c are from approaches with some unknowns, e.g. assumptions, DLTS measurements and molecular dynamics simulations. This table also shows most of possible reactions occurring in a silicon solid after irradiation.

written as $[C]$ and $[O]$ respectively.

3.4.1 Diffusion Mechanism

Defects mainly migrate by diffusion, including self and impurity processes, and they are summarized by Gösele [44]. Interstitially dissolved impurities A_i can move in a solid, such as metallic copper and iron atoms. This direct interstitial mechanism gives a faster A_i than for substitutionally dissolved atoms A_s, such as dopants.

A vacancy can exchange with an A_s or form a fast diffusing pair, $A_s + V \longleftrightarrow A_sV$. A self interstitial can also form a fast diffusing pair, $A_s + I \longleftrightarrow A_i$. This kick-out mechanism is important for silver, zinc and lead atoms. The Frank-Turnbull mechanism describes a dissociation from the lattice, $A_s \longleftrightarrow A_i + V$ for low temperatures.

3.4.2 Carbon and Compounds

Carbon C is a common and unavoidable impurity for silicon from the same group IV. An interstitial carbon C_i can diffuse directly or exchange with a lattice silicon atom and become a substitutional carbon C_s, vice versa. C_i's are highly mobile and can occupy a vacancy and form a C_s. Their combination, $C_s + C_i \longleftrightarrow C_SC_I$, is immobile. Carbon reactions can also assist interstitial diffusion.

Huhtinen gave systematic simulations for the non-ionising energy loss from different sources (gamma, neutrons and protons) and defect formation in silicon [56]. Most of possible reactions and their capture radii are listed in Table 3.1. Defects are assumed to be spherical and they react when two spheres overlap in the Monte Carlo simulation. The reaction rate is proportional to the diffusivity and capture radius which can be tuned to agree with experimental data.

3.4.3 Reaction Rate

Reaction rates is qualitatively described as

$$-\frac{dC_i}{dt} = k_i C_i \tag{3.15}$$

for first order reactions, e.g. diffusion; or as

$$-\frac{dC_i}{dt} = -\frac{dC_j}{dt} = k_{ij} C_i C_j = 4\pi R_{ij}(D_i + D_j) C_i C_j \tag{3.16}$$

for second order reactions, e.g. compound reactions. $N_{i,j}$ and $C_{i,j}$ are total number and concentration of element i and j. k, R and D are the rate constant, capture cross section and diffusion coefficient respectively. If an element (compound) is immobile or moves much slower, the reaction rate is dominated by the faster one.

3.4.4 Arrhenius Equation

The Arrhenius equation gives the temperature dependence of the reaction rate constant which is an empirical relationship for thermally induced processes. The reaction rate constant k is written as

$$k = k_0 e^{-E_A/k_B T} \tag{3.17}$$

where k_0 is the frequency factor and E_A is the activation energy for migration or dissociation. The diffusion constant D is written as

$$D = D_0 e^{-E_A/k_B T} \tag{3.18}$$

which is also a thermal process.

To extract the activation energy and pre-exponential factor from experiments, the Arrhenius equation is rearranged as

$$ln(k) = ln(k_0) - \frac{E_A}{k_B T} \tag{3.19}$$

and thus k_0 and E_A can be obtained from the intercept and slope from experimental data. Figure 5.3 in Chapter 5 will show an use of the Arrhenius plot to obtain the annealing time constant for the trapping lifetimes, $\tau = \tau_0 e^{E_A/k_B T}$.

3.4.5 Annealing Mechanism

Annealing starts right after defects are formed. It proceeds though migration and dissociation. Defects become mobile at different temperatures. They migrate in the silicon bulk until an immobile compound is formed, e.g. V_2 or $C_I C_S$. A recovery of

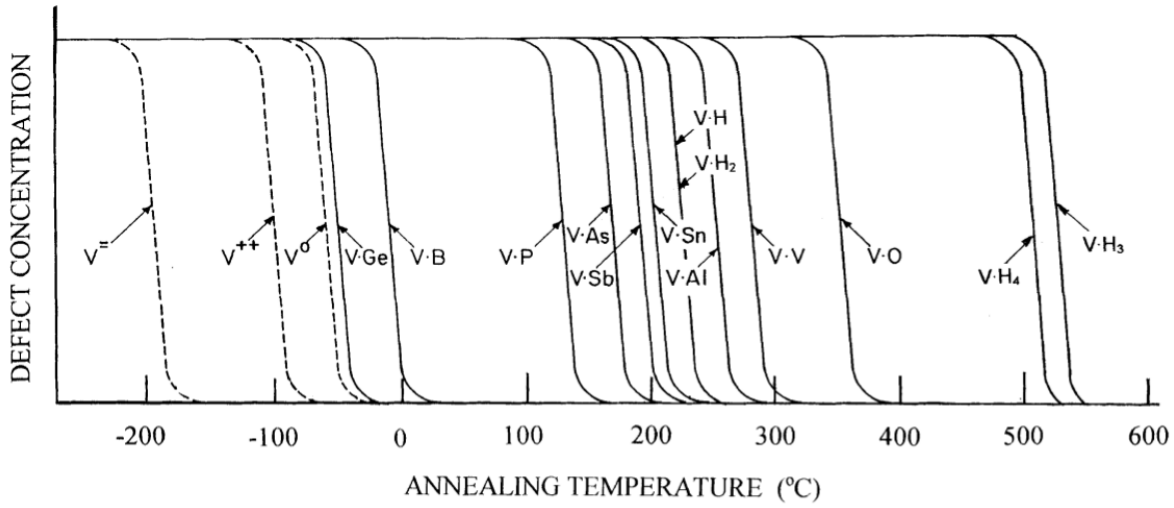


Figure 3.6: Annealing temperatures for various defects [127]. Defects migrate or dissociate to form other defects at their annealing temperatures.

the lattice structure, $I + V \longrightarrow \emptyset$ which is called the annihilation reaction, also occurs with a large probability.

A compound may dissociate if the vibrational (thermal) energy is larger than the binding energy. The decomposed components then start migrating and form different compounds as described above. Figure 3.6 shows the annealing temperatures of different defects when they start to migrate or dissociate. Normally defect concentrations increase after irradiation. After a period, bulk parameters tend to recover which is called annealing. However, sometimes bulk parameters degrade further due to the annealing in/out of some certain defects. This is referenced to as reverse-annealing which was modelled by the RD48 collaboration as follows.

Annealing changes the effective doping concentration N_{eff} as [85]

$$N_{eff}(\Phi_{eq}, T, t) = N_C(\Phi_{eq}) + N_A(\Phi_{eq}, T, t) + N_Y(\Phi_{eq}, T, t) \tag{3.20}$$

where N_C is the stable damage which depends only on the fluence Φ_{eq}. N_A and N_Y are beneficial (short-term) annealing and reverse (long-term) annealing which depend on the temperature and time. Figure 3.7 shows the beneficial and reverse-annealing which dominates after a long period.

3.5 Electrical Properties of Defects

The occupancy statistics for defects vary with their energy levels, temperature and Fermi level/electric potential. Occupancy function $f(E_t, E_F)$ is modelled by the quasi-Fermi level and SRH statistics. The changes in occupancy under a forward and reverse

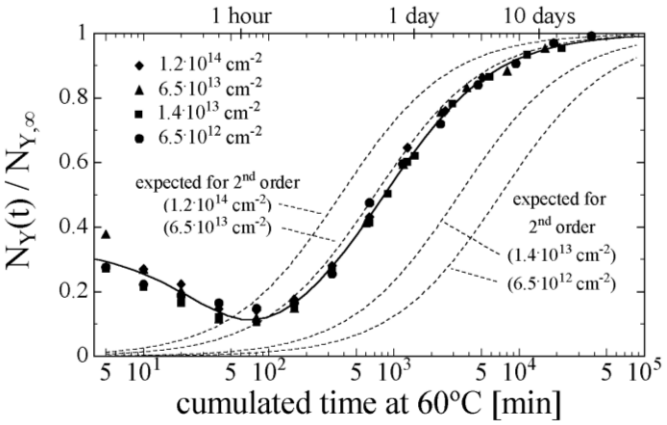


Figure 3.7: The beneficial and reverse annealing behaviour of the effective doping concentration in irradiated silicon diodes [84]. The effective doping concentration undergoes both beneficial and reverse annealing. The beneficial annealing also decreases the leakage current. See also Figure 4.3 which implies these two effects may be due to different deep defects.

bias are also discussed in this section.

3.5.1 At a Certain Fermi Level

Occupancy statistics for defects obeys the Fermi-Dirac statistics in Equation 2.6. For a given defect concentration N_t, the fractions for a defect level E_t occupied by electrons n_t (occupied) and holes p_t (empty) are given as

$$\begin{aligned} n_t &= N_t \frac{1}{1+e^{(E_t-E_F)/k_BT}} &&= N_t f(E_t) \\ p_t &= N_t \frac{1}{1+e^{-(E_t-E_F)/k_BT}} &&= N_t(1-f(E_t)) \end{aligned} \tag{3.21}$$

where $f(E_t)$ is the Fermi-Dirac distribution. Note that the totality condition holds, $n_t + p_t = N_t$.

Empty acceptors and filled donors have no contribution to the space charge distribution. The effective defect density from all types of deep acceptors $(-/0)$ which are occupied with electrons is given as

$$n_t^* = \sum N_{At} f(E_{At}) = \sum N_{At} \frac{1}{1+e^{(E_{At}-E_{Fn})/k_BT}} \tag{3.22}$$

where E_{Fn} is the quasi Fermi level for electrons. Similarly, the effective defect density from all types of deep donors $(0/+)$ which are not occupied with electrons (or occupied with holes) is given as

$$p_t^* = \sum N_{Dt}(1-f(E_{Dt})) = \sum N_{Dt} \frac{1}{1+e^{-(E_{Dt}-E_{Fp})/k_BT}} \tag{3.23}$$

where E_{Fp} is the quasi Fermi level for holes. The effective doping concentration affects the width of depletion region. Note that deep acceptors (donors) at the bottom (top) half of the band gap are more effective, but they are rare or non-existent among generated damage.

3.5.2 Shockley-Read-Hall Model

In reality, defects can capture or emit carriers in the conduction and valance and these processes tend to reach thermal equilibrium after a certain time. The standard way to determine the defect occupancy is the Shockley-Reed-Hall (SRH) relation which depends on the nature of defects. It provides a precise description for deep defects using their capture cross sections and thermal velocities as parameters. By utilising the SRH model model in Section 2.3, the defect occupancy functions for deep acceptors and donors are given as [113]

$$\begin{aligned} f_A(E_{At}) &= \frac{n + \frac{v_{pth}\sigma_{Ap}}{v_{nth}\sigma_{An}} n_i e^{\frac{E_i - E_{At}}{k_B T}}}{(n + n_i e^{\frac{E_{At} - E_i}{k_B T}}) + \frac{v_{pth}\sigma_{Ap}}{v_{nth}\sigma_{An}}(p + n_i e^{\frac{E_i - E_{At}}{k_B T}})} \\ &= \frac{v_{nth}\sigma_{An} n + v_{pth}\sigma_{Ap} n_i e^{-E_A/kT}}{v_{nth}\sigma_{An}(n + n_i e^{E_A/kT}) + v_{pth}\sigma_{Ap}(p + n_i e^{-E_{At}/kT})} \end{aligned} \tag{3.24}$$

and

$$\begin{aligned} f_D(E_{Dt}) &= \frac{\frac{v_{pth}\sigma_{Dp}}{v_{nth}\sigma_{Dn}} p + n_i e^{\frac{E_{Dt} - E_i}{k_B T}}}{(n + n_i e^{\frac{E_{Dt} - E_i}{k_B T}}) + \frac{v_{pth}\sigma_{Dp}}{v_{nth}\sigma_{Dn}}(p + n_i e^{\frac{E_i - E_{Dt}}{k_B T}})} \\ &= \frac{v_{pth}\sigma_{Dp} p + v_{nth}\sigma_{Dn} n_i e^{E_D/kT}}{v_{nth}\sigma_{Dn}(n + n_i e^{E_D/kT}) + v_{pth}\sigma_{Dp}(p + n_i e^{-E_D/kT})} \ . \end{aligned} \tag{3.25}$$

They are identical to Equation 3.22 and 3.23 if

$$v_{pth}\sigma_{Ap} = v_{nth}\sigma_{An} \quad , \quad v_{pth}\sigma_{Dp} = v_{nth}\sigma_{Dn} \quad \text{and} \quad E_{Fp} = E_{Fn} \tag{3.26}$$

for deep acceptors and donors. This has been implemented in Kurata version 6. See Chapter 4 for the version list.

3.5.3 Quasi-Fermi Level Approximation

An alternative model is the quasi-Fermi level approximation. It uses the relationship between quasi-Fermi levels and minority carrier concentrations in Equation 2.13 to replace E_{Fp} and E_{Fn} in Equation 3.22 and 3.23. The defect occupancy functions for

deep acceptors and donors are then given as [81]

$$\begin{aligned} f_A^*(E_{At}) &= \frac{1}{1+\frac{n_i}{n}e^{\frac{E_{At}-E_i}{k_BT}}} \\ f_D^*(E_{Dt}) &= \frac{1}{1+\frac{n_i}{p}e^{\frac{E_i-E_{Dt}}{k_BT}}} . \end{aligned} \tag{3.27}$$

In regard to the total charge density ρ used for solving the Poisson equation, the effective concentration N_{eff} from defects is given as [41]

$$N_{eff} = p_t^* - n_t^* = \sum_{donors} N_{Dt}(1 - f_D^*(E_{Dt})) - \sum_{acceptors} N_{At} f_A(E_{At}) \tag{3.28}$$

where the filled/empty deep-level acceptors are $(-/0)$ and donors are $(0/+)$.

Irradiated material usually has more than one defect with different energy levels and cross sections. One can introduce a representative defect at the mid-gap as a summation of all defects. This is the key concept of Kurata version 4. See also Chapter 4. Note that the total charge has to be zero written as

$$p - n + N_d - N_a + N_{eff} = 0 \tag{3.29}$$

at full depletion under a high bias.

3.5.4 Under a Reverse Bias

Applying a reverse bias across a diode widens the depletion region, the space charge region (SCR), where mobile carriers are swept to the electrodes. The time differential of the carrier concentration in Section 2.3 is rewritten as

$$\frac{dn_c}{dt} = r_a - r_b + r_c - r_d = -e_n n_t + e_p p_t \tag{3.30}$$

by dropping the capture terms since $n \approx p \approx 0$ in the depletion region. The change in carrier concentration is zero at steady state, $dn_c/dt = 0$, and thus the effective defect densities are given as [84]

$$\begin{aligned} n_t &= N_t \frac{e_p}{e_n+e_p} = \frac{N_t}{1+(\frac{c_n}{c_p})e^{\frac{E_t-E_i}{k_BT}}} \\ p_t &= N_t \frac{e_n}{e_n+e_p} = \frac{N_t}{1+(\frac{c_p}{c_n})e^{-\frac{E_t-E_i}{k_BT}}} \end{aligned} \tag{3.31}$$

for electrons (filled acceptors) and holes (empty donors) respectively.

The generation rate of electron-hole pairs G_t is given as

$$G_t = e_n n_t = e_p p_t = N_t n_i \frac{c_n c_p}{c_n e^{(E_t - E_i)/k_B T} + c_p e^{-(E_t - E_i)/k_B T}} \tag{3.32}$$

which is usually simplified to

$$G_t = \frac{N_t n_i c_n}{2Cosh(\frac{E_t - E_i}{k_B T})} \tag{3.33}$$

by assuming $c_n = c_p$. Silicon sensors are operated in depletion and thus there is always a leakage current which depends on the defects, especially near mid-gap.

3.5.5 Under a Forward Bias

Applying a forward bias injects a large number of free carriers into the diode. The capture rates will be larger than the emission rates. The time differential of the carrier concentration is written as

$$\frac{dn_c}{dt} = c_n n p_t - c_p p n_t \tag{3.34}$$

by dropping the emission terms, $e_n \approx e_p \approx 0$. Given $dn_c/dt = 0$ at steady state, the effective defect densities for electrons and holes are given as [84]

$$n_t = \frac{N_t}{1 + \frac{c_p p}{c_n n}} \quad \text{and} \quad p_t = \frac{N_t}{1 + \frac{c_n n}{c_p p}} \ . \tag{3.35}$$

Note that the total concentration N_t of any defect is always constant, $N_t = n_t + p_t$, for reverse and forward bias cases.

3.6 Effects of Defects

Moll [84] summarised major properties rescaled by radiation damages. The generation τ_g and trapping τ_t lifetimes are inversely proportional to the defect concentration N_t, $\tau_{g,t} \propto 1/N_t$. Three effects on the leakage current I_{leak}, charge collection efficiency (CCE) and depletion voltage V_{dep} are discussed in this section. They are written in short as

$$I_{leak} \propto N_t \propto \tau_g^{-1} \quad , \quad CCE \propto N_t^{-1} \propto \tau_t \quad \text{and} \quad V_{dep} \propto N_t \tag{3.36}$$

for reference. Note that these are only trends rather than a linear dependence. Equations here are written in terms of e_n and e_p which were defined in Section 2.3.

3.6.1 Generation Lifetime and Leakage Current

The leakage current is proportional to the generation rate which is the sum of emissions from electrons and holes under reverse bias. Defects near mid-gap are active recombination centres which enhance the generation rate since a two-step process with equal/similar steps is more probable. The generation time constant τ_g is given as

$$\frac{1}{\tau_g} = \sum_{traps} \frac{1}{\tau_{g,t}} = \sum_{traps} \frac{G_t}{n_i} = \sum_{traps} \frac{N_t}{n_i} \frac{e_n e_p}{e_n + e_p} \tag{3.37}$$

where G_t was defined in Equation 3.32. The bulk generation current density j_{bg} is just qn_i/τ_g.

3.6.2 Trapping Lifetime and Charge Collection Efficiency

The charge collection efficiency decreases with the trapping of electron-hole pairs since they are usually re-emitted around a few ms. The probability of trapping an electron is proportional to the capture cross section of electrons and the fraction of unoccupied defects. The trapping time constant τ_t is given as

$$\frac{1}{\tau_{t,n/p}} = \sum_{traps} c_{n/p} p/n_t = \sum_{traps} N_t c_{n/p} \frac{e_{n/p}}{e_n + e_p} \ . \tag{3.38}$$

3.6.3 Space Charge and Depletion Voltage

The total effective space charge density N_{eff} is rewritten as

$$N_{eff} = \sum_{donors} N_{Dt} \frac{e_{n,Dt}}{e_{n,Dt} + e_{p,Dt}} - \sum_{acceptors} N_{At} \frac{e_{p,At}}{e_{n,At} + e_{p,At}} \ . \tag{3.39}$$

The depletion voltage is a function of N_{eff} given as [78]

$$V_{dep} = \frac{q|N_{eff}|d^2}{2\epsilon_{Si}} \tag{3.40}$$

which increases with $|N_{eff}|$ and d^2. There is an upper limit for the bias voltage in operation. This is the reason to reduce the inter-electrode distance d since the total non-ionising dose and bias voltage are usually limited parameters.

$|N_{eff}|$ is usually determined experimentally by measuring the sensor capacitance. Their relationship is given as

$$|N_{eff}| = \frac{2V_{dep}C^2}{q\epsilon_{Si}} \tag{3.41}$$

which is a rearrangement of Equation 2.55.

3.7 Defect Identification

Aside from direct but sliced images obtained by transmission electron microscopy (TEM), defects are usually observed indirectly by their electrical properties: capacitance, current or charge. DLTS and TSCAP/TSC are discussed which measure the capacitance and current of defects. Electron paramagnetic resonance (EPR) spectroscopy, which excite and study spins of unpaired electrons, can be also used for defect identification. This was pioneered by Watkins in the 1960's. Figure 3.6 is due to Watkins using EPR to identify the key defects in silicon.

3.7.1 DLTS

Lang [70] introduced the deep-level transient spectroscopy (DLTS) technique for defect characterisation which investigates the capacitance transients of deep-levels. Trapping and emission of carriers gives a change of the sensor capacitance. The DLTS method assumes the C-t transient follows an exponential time dependence given as [108]

$$C(t) = C_0[1 - \frac{n_t}{2N_D}e^{-t/\tau_e}] \tag{3.42}$$

with a single time constant τ_e. It increases inversely with temperature as

$$\tau_e = \frac{e^{(E_C - E_t)/k_B T}}{\gamma_n \sigma_n T^2}$$

where $\gamma_n = 1.07 \times 10^{21}$ and $\gamma_p = 1.78 \times 10^{21}$ $cm^{-2}s^{-1}K^{-2}$ for n-type and p-type silicon respectively.

For a p+-n-n+ diode biased at a reverse bias V_R, traps can be filled by electrons for a majority pulse ($V_R < V_p < 0$) or by holes by an injection pulse ($V_p > 0$, a forward bias) experimentally. This procedure is repeated at different temperatures. The signal of defect emission is multiplied by a weighting function $w(t)$ which is usually a linear filter (a integrator or low-pass filter). For the Boxcar DLTS introduced by Lang, C-t waveforms are sampled at $t = t_1$ and $t = t_2$ where $w(t) = \delta(t - t_1) - \delta(t - t_2)$.

The capacitance difference δC between two times is calculated as

$$\delta C = C(t_1) - C(t_2) = \frac{n_t}{2N_D} C_0 (e^{-t_2/\tau_e} - e^{-t_1/\tau_e}) \tag{3.43}$$

and the maximum of δC occurs at a certain temperature T_{max}. Differentiating δC with respect to τ_e gives

$$\tau_e(T_{max}) = \frac{t_2 - t_1}{ln(t_2/t_1)} \ . \tag{3.44}$$

An Arrhenius plot extracted from such data is used to determine the defect level.

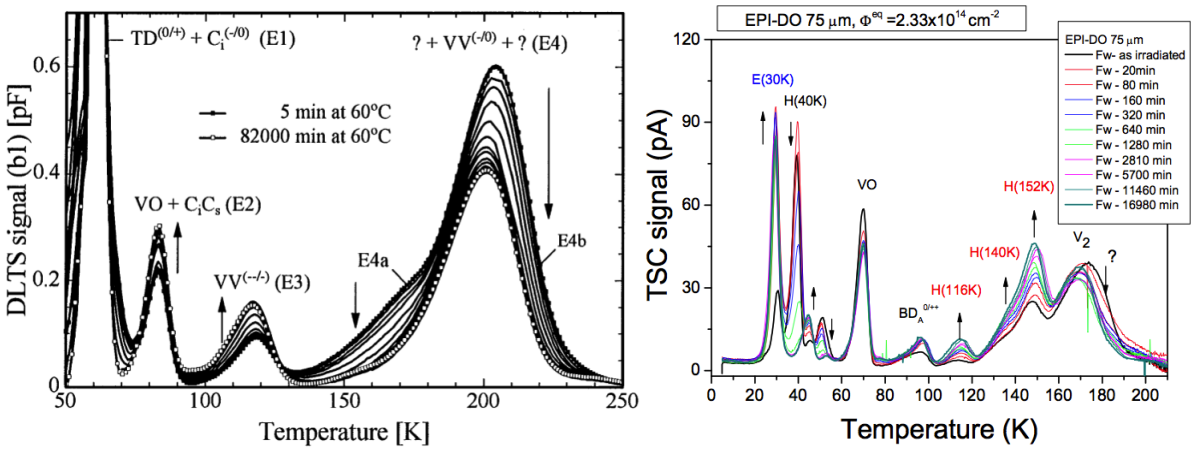


Figure 3.8: Typical results from the RD48 and RD50 collaboration using DLTS [74] (left) and TSC [121] (right) techniques. They both vary the temperature of instruments and measure the energy levels of defects according to their emission constants or activation energies.Identified defects are labelled in the plots, e.g. V_2 and VO.

3.7.2 TSCAP and TSC

Thermally stimulated capacitance (TSCAP) and current (TSC) can also identify deep-level defects. These techniques were used for insulators originally and later modified for semiconductors [108]. DLTS is very sensitive to electrically active defects, but it can only handle a fluence level up to 10^{12} cm^{-2} which is understood from the n_t/N_D term in Equation 3.42 for a common silicon material with an introduction rate of 1 cm^{-1}. TSC is then a better choice to deal with heavily irradiated sensors [98].

Traps are filled with majority carriers of the substrate type in a cooled device. They can also be filled with minority carriers by optical injection or a forward bias. The device is measured under reverse bias and heated gradually and constantly. Capacitance steps or current peaks appear as traps emit the occupied carriers. The activation energy $\triangle E$ is given as

$$\triangle E = kT_m ln(\frac{\gamma_n \sigma_n k T_m^4}{\beta(\triangle E + 2kT_m)}) \tag{3.45}$$

where β is the heating rate. $\triangle E$ is related to T_m, the mid point of capacitance steps in TSCAP or current peaks in TSC, which gives the defect level as $E_C - E_t$ or $E_t - E_V$. Figure 3.8 shows typical results for DLTS and TSC techniques.

Chapter 4

Device Simulation

Device modelling originated from Schockley's 1949 paper, in which he established the theoretical foundation for diodes and transistors [68]. It evolved over decades to full numerical simulation based on advanced computing power. Device simulation plays an important role for design and fabrication purposes nowadays. It is used in this work to understand and optimise 3D devices.

Commercial software packages, such as Synopsis (formerly ISE) [18] and Silvaco [15] TCAD (technology computer aided design), include the relevant semiconductor physics and different sophisticated models for silicon and gallium arsenide. They can provide information, such as the electric potential distribution, carrier concentrations and current densities, for transient or steady state cases. They can also simulate the fabrication process and circuit characteristics.

Research institutes sometimes build semiconductor simulators for general or specific purposes: EVEREST (European Venture for Research in Semiconductor Technology) aims to simulate transient or steady state behaviour for semiconductor devices in three dimensions. The John Hopkins group [112][114][113] built PIXELAV for the CMS pixel detector. The Manchester (formerly Brunel) group also developed Kurata to provide relevant physical properties in 1D for heavily irradiated devices.

Deep level defects are conventionally described by applying the SRH model individually. Robbins and Watts (Brunel) [81] alternatively modelled various deep levels in irradiated diodes by introducing a representative deep acceptor or donor at mid-gap. Their simulator, which uses the algorithm from Kurata [68] and is named after that, solves semiconductor equations for the electric potential and carrier concentrations. Osborne [90] modified Kurata to simulate avalanche photodiodes.

Kurata was written in Fortran for 1D structures originally. It was first rewritten in *Mathematica* [19] and C++ for 1D. It is extended further to 2D which enables modelling of many other structures. Final versions of Kurata are written in C++ since

compiled executable files[1] use much shorter computing time than *Mathematica*.

This chapter describes the algorithm and physics models for Kurata. The 2D extension is important for silicon sensor applications using strip and 3D structures. The leakage current and generation lifetime are discussed which are fluence dependent. The capacitance of 3D sensors is calculated and compared to experimental noise measurements in this chapter.

4.1 Numerical Method

Device simulators solve the coupled semiconductor equations simultaneously to obtain key properties. The finite difference (grid) and finite element method (FEM) are two common ways to digitise these linear equations. TCAD uses the FEM with a self-refining mesh which is suitable for all structures. Kurata uses the finite difference method which is mathematically simpler but less adequate for non-rectangular structures. This can be compensated by finer and well defined grids. The grid method is discussed in this section.

There are also other models for the continuity equations which can be chosen in TCAD simulations [115]. The thermodynamic model can account for self-heating which is suitable for high power devices. The hydrodynamic model gives a better description of hot electrons and non-stationary transport behaviour [71] for devices with a small active region. The Monte Carlo method solves the Boltzmann equation in which the carrier mobilities are band and momentum dependent.

4.1.1 Finite Difference Method

The semiconductor equations in Chapter 2 have three variables, ϕ (denoted as V previously) p and n, which are the electric potential, hole and electron concentration respectively. The drift and diffusion components for the continuity equations are proportional to the electric field $\nabla\phi$ and the gradient of carrier densities $\nabla n/p$ respectively. Scharfetter and Gummel proposed an integral form to combine them with two new terms which depend both on n/p and ϕ directly and simultaneously [107]. This can increase the solution stability for digitised linear equations.

Semiconductor equations in a partial derivative form are transformed to the finite difference form which treat derivatives as divisions, $\partial f(x)/\partial x$ as $\triangle f(x)/\triangle x = [f(x') - f(x)]/(x' - x)$. Variables are estimated at main points N with a spacing $h(M)$ and their derivatives are estimated at auxiliary points M with a spacing $h'(N + 1)$. The numerical details in the Kurata program based on [68] are now described.

[1] The C++ codes are compiled by the GNU compiler which is open-source and cross-platform [5].

By utilising the integral and finite difference form, the hole and electron current density are written as

$$\begin{aligned} J_p &= \tfrac{q}{h(M)}[\lambda_{p1}(M)p(N) + \lambda_{p2}(M)p(N+1)] \\ J_n &= \tfrac{q}{h(M)}[\lambda_{n1}(M)n(N) + \lambda_{n2}(M)n(N+1)] \ . \end{aligned} \tag{4.1}$$

They are inserted into the continuity equations for steady state as

$$\begin{aligned} \frac{1}{q}\frac{J_p(M) - J_p(M-1)}{h'(N)} - G(N) + R(N) &= 0 \\ \frac{1}{q}\frac{J_n(M) - J_n(M-1)}{h'(N)} + G(N) - R(N) &= 0 \ . \end{aligned} \tag{4.2}$$

The Poisson equation is digitised as

$$\begin{aligned} &\gamma_1\phi(N-1) + \gamma_2\phi(N) + \gamma_3\phi(N+1) \\ = -\frac{q}{\epsilon_{Si}}&[p(N) - n(N) + N_D(N) - N_A(N) + \sum N_D^*(N) - \sum N_A^*(N)] \end{aligned} \tag{4.3}$$

where $\sum N_D^*$ and $\sum N_A^*$ are the effective defect concentrations, namely the empty deep donors and filled deep acceptors respectively.

The Scharfetter-Gummel coefficient functions λ's are given as

$$\begin{aligned} \lambda_{p1,p2}(M) &= \mu_p(M)\tfrac{\phi(N)-\phi(N+1)}{1-e^{\mp\beta(M)}} \\ \lambda_{n1,n2}(M) &= \mu_n(M)\tfrac{\phi(N)-\phi(N+1)}{1-e^{\pm\beta(M)}} \end{aligned} \tag{4.4}$$

where

$$\beta(M) = \frac{1}{k_B T}[\phi(N) - \phi(N+1)] \ .$$

Note that the $\pm$ sign in the denominator leads to a conjugate property between $\lambda_{p/n}$ or $\lambda_{1/2}$ depending on the electric field[2]. The concentrations on two neighbouring grid

[2] λ's have a form of $x/(1-e^{\pm x})$ where $x \sim \beta$. It has properties given as

$$\begin{aligned} \lim_{x\to 0}\frac{x}{1-e^{x}} &= -1 \quad , \quad \lim_{x\to 0}\frac{x}{1-e^{-x}} = 1 \\ \lim_{x\to \infty}\frac{x}{1-e^{x}} &= 0 \quad , \quad \lim_{x\to \infty}\frac{x}{1-e^{-x}} \sim x \\ \lim_{x\to -\infty}\frac{x}{1-e^{x}} &\sim -x \quad , \quad \lim_{x\to -\infty}\frac{x}{1-e^{-x}} = 0 \ . \end{aligned} \tag{4.5}$$

At zero electric field, the continuity equations depend only on the two neighbouring concentrations. The direction and magnitude of this diffusion process is determined by the $\pm$ sign of $\lambda_{1/2}$ and the concentration difference respectively. As the electric field becomes larger, the continuity equations depend mostly on the concentration of one side. The direction and magnitude of this drift process is also determined by the $\pm$ sign of $\lambda_{1/2}$ and the potential gradient β respectively.

points and λ's describe the direction and magnitude of drift and diffusion. Two neighbouring concentrations, $p(N)/n(N)$ and $p(N+1)/n(N+1)$, and $\lambda_{1/2}$ contribute together at moderate electric fields. The coefficients γ are given as

$$\begin{aligned} \gamma_1(N) &= \frac{1}{h(M-1)h'(N)} \\ \gamma_2(N) &= -\frac{1}{h'(N)}[\frac{1}{h(M-1)} + \frac{1}{h(M)}] \\ \gamma_3(N) &= \frac{1}{h(M)h'(N)} \end{aligned} \tag{4.6}$$

which are denominators (spacings) of the second order derivatives.

4.1.2 The Runge-Kutta Method

The Runge-Kutta method is applied to solve non-linear and coupled partial differential equations. Variables are expressed in terms of the current values (denoted by a superscript 0) and associated increment quantities (denoted by a symbol ∂) of related variables by the truncated Taylor expansion, e.g. $\phi_{true} = \phi^0_{initial} + \partial\phi$.

Indirect variables, J_p, J_n, G and R, are linearised in terms of p, n and ϕ. Higher-order terms are neglected but the accuracy is monitored. Current densities J_p and J_n are given as

$$\begin{aligned} J_{p,n}(M) &= J^0_{p,n}(M) + \delta J_{p,n}(M) \\ &\approx J^0_{p,n}(M) + \frac{\partial J^0_{p,n}(M)}{\partial(p,n)(N)}\delta(p,n)(N) + \frac{\partial J^0_{p,n}(M)}{\partial(p,n)(N+1)}\delta(p,n)(N+1) \\ &\quad + \frac{\partial J^0_{p,n}(M)}{\partial\phi(N)}\delta\phi(N) + \frac{\partial J^0_{p,n}(M)}{\partial\phi(N+1)}\delta\phi(N+1) \ . \end{aligned} \tag{4.7}$$

The impact ionisation G and generation-recombination rate R are simplified with a summation notation as

$$G(N) = G^0(N) + \delta G(N) \approx G^0(N) + \sum_{l=p,n,\phi}\sum_{K=N-1}^{N+1}\frac{\partial G(N)}{\partial l(K)}\delta l(K) \tag{4.8}$$

$$R(N) = R^0(N) + \delta R(N) \approx R^0(N) + \sum_{l=p,n}\frac{\partial R(N)}{\partial l(N)}\delta l(N) \ . \tag{4.9}$$

The continuity equations are written in terms of increments based on the three

equations above and rearranged as

$$\begin{aligned}&\frac{1}{q}\frac{\partial J_{p,n}(M)-\partial J_{p,n}(M-1)}{h'(N)}\mp\partial G(N)\pm\partial R(N)\\&=\frac{1}{q}\frac{J^0_{p,n}(M)-J^0_{p,n}(M-1)}{h'(N)}\pm G^0(N)+\mp R^0(N)\end{aligned}\tag{4.10}$$

to separate the current and iterative values. The Poisson equation is linear originally and needs only rearrangement as

$$\begin{aligned}&\frac{q}{\epsilon}[\partial p(N)-\partial n(N)]+\gamma_1(N)\partial\phi(N-1)+\gamma_2(N)\partial\phi(N)+\gamma_3(N)\partial\phi(N+1)\\&=-\frac{q}{\epsilon}[\Sigma(N)+p^0(N)-n^0(N)]-\gamma_1(N)\phi^0(N-1)+\gamma_2(N)\phi^0(N)+\gamma_3(N)\phi^0(N+1)\ ,\end{aligned}\tag{4.11}$$

in which the current and iterative values are separated.

4.1.3 Matrix-Vector Representation

The three coupled equations can be written in a matrix-vector form as

$$\begin{aligned}A(N)\delta y(N-1)+B(N)\delta y(N)+C(N)\delta y(N+1)=F(N)\ ,\\2\leqslant N\leqslant L-1\end{aligned}\tag{4.12}$$

where $y(N)$ and $\delta y(N)$ are 3×1 matrices defined as

$$y(N)=\begin{bmatrix}p(N)\\n(N)\\\phi(N)\end{bmatrix}\quad\text{and}\quad\delta y(N)=\begin{bmatrix}\delta p(N)\\\delta n(N)\\\delta\phi(N)\end{bmatrix}$$

for main points from 1 to L. Note that δy represents the increment of each iteration which is obtained from the current values.

By comparing the matrix-vector equation and linearised semiconductor equations, the coefficients of $\delta y(N-1)$, $\delta y(N)$ and $\delta y(N+1)$ are 3×3 matrices given as

$$A(N)=\begin{bmatrix}-\frac{1}{qh'(N)}\frac{\partial J^0_p(M-1)}{\partial p(N-1)}-\frac{\partial G^0(N)}{\partial p(N-1)} & -\frac{\partial G^0(N)}{\partial n(N-1)} & -\frac{1}{qh'(N)}\frac{\partial J^0_p(M-1)}{\partial\phi(N-1)}-\frac{\partial G^0(N)}{\partial\phi(N-1)}\\ \frac{\partial G^0(N)}{\partial p(N-1)} & -\frac{1}{qh'(N)}\frac{\partial J^0_n(M-1)}{\partial n(N-1)}-\frac{\partial G^0(N)}{\partial n(N-1)} & -\frac{1}{qh'(N)}\frac{\partial J^0_n(M-1)}{\partial\phi(N-1)}+\frac{\partial G^0(N)}{\partial\phi(N-1)}\\ 0 & 0 & \frac{1}{h(M-1)h'(N)}\end{bmatrix}$$

$$B(N) = \begin{bmatrix} \frac{1}{qh'(N)}\frac{\partial J_p^0(M)-\partial J_p^0(M-1)}{\partial p(N)} - \frac{\partial G^0(N)}{\partial p(N)} + \frac{\partial R^0(N)}{\partial p(N)} & -\frac{\partial G^0(N)}{\partial n(N)} + \frac{\partial R^0(N)}{\partial n(N-1)} & \frac{1}{qh'(N)}\frac{\partial J_p^0(M)-\partial J_p^0(M-1)}{\partial \phi(N)} - \frac{\partial G^0(N)}{\partial \phi(N)} \\ -\frac{\partial G^0(N)}{\partial p(N)} - \frac{\partial R^0(N)}{\partial p(N)} & \frac{1}{qh'(N)}\frac{\partial J_n^0(M)-\partial J_n^0(M-1)}{\partial n(N)} + \frac{\partial G^0(N)}{\partial n(N)} - \frac{\partial R^0(N)}{\partial n(N)} & \frac{1}{qh'(N)}\frac{\partial J_n^0(M)-\partial J_n^0(M-1)}{\partial \phi(N)} + \frac{\partial G^0(N)}{\partial \phi(N)} \\ \frac{q}{\epsilon_{Si}} & -\frac{q}{\epsilon_{Si}} & -\frac{1}{h'(N)}[\frac{1}{h(M-1)} + \frac{1}{h(M)}] \end{bmatrix}$$

$$C(N) = \begin{bmatrix} -\frac{1}{qh'(N)}\frac{\partial J_p^0(M-1)}{\partial p(N+1)} - \frac{\partial G^0(N)}{\partial p(N+1)} & -\frac{\partial G^0(N)}{\partial n(N+1)} & -\frac{1}{qh'(N)}\frac{\partial J_p^0(M-1)}{\partial \phi(N+1)} - \frac{\partial G^0(N)}{\partial \phi(N+1)} \\ \frac{\partial G^0(N)}{\partial p(N+1)} & -\frac{1}{qh'(N)}\frac{\partial J_n^0(M-1)}{\partial n(N+1)} - \frac{\partial G^0(N)}{\partial n(N+1)} & -\frac{1}{qh'(N)}\frac{\partial J_n^0(M-1)}{\partial \phi(N+1)} + \frac{\partial G^0(N)}{\partial \phi(N+1)} \\ 0 & 0 & \frac{1}{h(M)h'(N)} \end{bmatrix}.$$

$F(N)$ is a 3×1 matrix[3] for initial or current values (real numbers) which is given as

$$F(N) = \begin{bmatrix} -\frac{1}{qh'(N)}(J_p^0(M) - J_p^0(M-1)) + G^0(N) - R^0(N) \\ -\frac{1}{qh'(N)}(J_n^0(M) - J_n^0(M-1)) - G^0(N) + R^0(N) \\ \{-\frac{q}{\epsilon_{Si}}[N_D(N) - N_A(N) + p^0(N)n^0(N)] \\ -\gamma_1(N)\phi^0(N-1) - \gamma_2(N)\phi^0(N) - \gamma_3(N)\phi^0(N+1)\} \end{bmatrix}.$$

Equation 4.12 represents a general relation for the unknown increment $\partial y(N)$. Equations at all grid points in the matrix-vector form can be expressed with a $(N-2) \times (N-2)$ coefficient matrix (from $N = 2$ to $N = L-1$), a $(N-2) \times 1$ increment matrix $\delta(N)$ and a $(N-2) \times 1$ number matrix $F(N)$ as

[3]Note that the matrix element in the curly brackets is one term and separated in two lines due to its long width.

$$\begin{bmatrix} B(2) & C(2) & \dots & & & & 0 \\ A(3) & B(3) & C(3) & & & & 0 \\ 0 & A(4) & B(4) & C(4) & & & 0 \\ \vdots & & & & & & \vdots \\ 0 & & & A(L-3) & B(L-3) & C(L-3) & 0 \\ 0 & & & & A(L-2) & B(L-2) & C(L-2) \\ 0 & & & & \dots & A(L-1) & B(L-1) \end{bmatrix} \begin{bmatrix} \delta y(2) \\ \delta y(3) \\ \delta y(4) \\ \vdots \\ \delta y(L-3) \\ \delta y(L-2) \\ \delta y(L-1) \end{bmatrix} = \begin{bmatrix} F(2) \\ F(3) \\ F(4) \\ \vdots \\ F(L-3) \\ F(L-2) \\ F(L-1) \end{bmatrix}, \tag{4.13}$$

in which each matrix element, $A(N)$, $B(N)$ and $C(N)$, is still a 3×3 matrix. Note that the equation is for 1D and there is no explicit representation for 2D or 3D. However, these relationships can be coded in C++ using matrices in higher dimensions, e.g. $B[M][N][x][y]$.

Note also that $[\partial J_{p/n}(N)/\partial p/n(N)]$'s and $[\partial J_{p/n}(N)/\partial \phi(N)]$'s are partial derivatives which are obtained from the digitised continuity equations. They are listed in [68] and their details are skipped here. So far, semiconductor equations, including the increment parts, are nine different variables of $p(N-1/N/N+1)$, $n(N-1/N/N+1)$ and $\phi(N-1/N/N+1)$.

4.1.4 Recursive Method

The recursive method is used to solve the block diagonal matrix in Equation 4.13. It requires more computing power to calculate three unknown vectors directly since $\delta y(N-1)$ and $\delta y(N+1)$ are obtained from the boundary conditions at the two ends and one can only have one at a time. Kurata introduced an efficient algorithm to reduce the unknowns from three down to two. The new recursion relation is written as

$$B'(N)\delta y(N) + C'(N)\delta y(N+1) = F'(N) \tag{4.14}$$

which can be rearranged as

$$\delta y(N) = B'(N)^{-1}F'(N) - B'(N)^{-1}C'(N)^{-1}\delta y(N+1)\ . \tag{4.15}$$

Replacing the $\delta y(N-1)$ term in the matrix-vector relation (Equation 4.12) with the primed one (Equation 4.15) gives

$$\begin{aligned}[B(N) - A(N)B'(N-1)^{-1}C'(N-1)]\delta y(N) + C(N)\delta y(N+1) \\ = F(N) - A(N)B'(N-1)^{-1}F'(N-1)\ .\end{aligned} \tag{4.16}$$

By comparing the coefficients of $\delta y(N)$ and $\delta y(N+1)$ with the primed matrix-vector relation (Equation 4.14), the recursion relations are obtained as

$$F'(N) = F(N) - A(N)B'(N-1)^{-1}F'(N-1) \tag{4.17a}$$

$$B'(N) = B(N) - A(N)B'(N-1)^{-1}C'(N-1) \tag{4.17b}$$

$$C'(N) = C(N)\ , \quad 3 \leqslant N \leqslant L-1\ . \tag{4.17c}$$

Each iteration procedure starts from $N = 2$ where $\delta y(1) = 0$ due to the fixed boundary condition. The corresponding recursion relations are given as

$$F'(2) = F(2)\ , \quad B'(2) = B(2) \quad \text{and} \quad C'(2) = F(2)\ .$$

For $N = 3$, the values from $N = 2$ are inserted into Equation 4.15 to obtain $\delta y(3)$ in terms of $\delta y(4)$. The procedure runs through all grids to $N = L-1$ to obtain all primed coefficients. By setting $\delta y(L) = 0$, $\delta y(L-1)$ is obtained. Then the procedure runs backwards from $N = L-1$ to $N = 2$ to have all unknown increments $\delta y(N)$. The increments are added to the present values, $y(N) = y^0(N) + \delta y(N)$, to become new current values for the next iteration.

Note that this modified recursion relation only needs matrices from the previous point. However, it is not free since a complete iteration runs though grid points to reach the last boundary condition and run backwards to obtain all the modified matrix elements. The truncated Taylor expansion considers only the first order differential which may require more steps to give convergence.

4.2 2D Extension

2D simulation is needed for a better understanding of overall properties since its results may be different from 1D simulation due to geometrical effects. For instance, a *pn* diode is usually modelled as a line across the junction since it has a perfect symmetry. However, 3D sensors cannot be modelled by its diagonal line across two electrodes in 1D since there is no translational symmetry in the orthogonal direction.

The algorithm discussed in the previous section has been extended to 2D. Variables are estimated at main grids (M, N) and auxiliary grids (M', N'). The spacings, $h_x(M)$

and $h_y(N)$, for main grids have to be chosen properly in the x and y directions at start. The spacings for auxiliary grids are then given as

$$h'_x(M) = \frac{1}{2}[h_x(M'-1) + h_x(M')] \quad \text{and} \quad h'_y(N) = \frac{1}{2}[h_y(N'-1) + h_y(N')] \ .$$

The grids are adjustable according to the complexity of geometry which has to balance the convergence stability and calculation time. A small step is also needed to ramp up voltages to the target bias. A small time step will be required for transient problems with a fast changing solution.

4.2.1 Scharfetter and Gummel's Method

The hole and electron currents are again written in the integral form [68] in the x and y directions as

$$\begin{aligned}
J_{px} &= \frac{q}{h'_x(M')}[\lambda_{px1}(M')p(M,N) + \lambda_{px2,nx2}(M')p(M+1,N) \\
J_{py} &= \frac{q}{h'_y(N')}[\lambda_{py1}(M')p(M,N) + \lambda_{py2,ny2}(N')p(M,N+1) \\
J_{nx} &= \frac{q}{h'_x(M')}[\lambda_{nx1}(M')n(M,N) + \lambda_{px2,nx2}(M')n(M+1,N) \\
J_{ny} &= \frac{q}{h'_y(N')}[\lambda_{ny1}(M')n(M,N) + \lambda_{py2,ny2}(N')n(M,N+1) \ .
\end{aligned} \tag{4.18}$$

The coefficient functions λ are given as

$$\begin{aligned}
\lambda_{px1}(M') &= \frac{\mu_p}{\theta}\frac{\beta(M')}{1-e^{-\beta(M')}} \ , \quad \lambda_{nx1}(M') = \frac{\mu_n}{\mu_p}\lambda_{px2}(M') \\
\lambda_{px1}(M') &= \frac{\mu_p}{\theta}\frac{\beta(M')}{1-e^{\beta(M')}} \ , \quad \lambda_{nx1}(M') = \frac{\mu_n}{\mu_p}\lambda_{px2}(M') \\
\lambda_{py1}(N') &= \frac{\mu_p}{\theta}\frac{\beta(N')}{1-e^{-\beta(N')}} \ , \quad \lambda_{ny1}(N') = \frac{\mu_n}{\mu_p}\lambda_{py2}(N') \\
\lambda_{py2}(N') &= \frac{\mu_p}{\theta}\frac{\beta(N')}{1-e^{\beta(N')}} \ , \quad \lambda_{ny2}(N') = \frac{\mu_n}{\mu_p}\lambda_{py1}(N')
\end{aligned}$$

where

$$\begin{aligned}
\beta(M') &= \theta(M')[\phi(M,N) - \phi(M+1,N)] \\
\beta(N') &= \theta(N')[\phi(M,N) - \phi(M,N+1)] \\
\theta(M') &= \frac{q}{k_B T(M')} \ , \quad \theta(N') = \frac{q}{k_B T(N')} \ .
\end{aligned} \tag{4.19}$$

Note that the potential gradient β determines the values of λ's. The carrier densities and corresponding λ's around a specific grid point describe the local transport

behaviour, diffusion and/or drift, which is discussed in Equation 4.5.

The divergence of current densities in 2D is approximated as a sum of differentials in the x and y directions. The continuity equations for the steady states are given as

$$\begin{aligned}\frac{1}{q}\frac{J_{px}(M')-J_{px}(M'-1)}{h'_x(M)}+\frac{1}{q}\frac{J_{py}(N')-J_{py}(N'-1)}{h'_y(N)}&=-U(M,N)\\ \frac{1}{q}\frac{J_{nx}(M')-J_{nx}(M'-1)}{h'_x(M)}+\frac{1}{q}\frac{J_{ny}(N')-J_{ny}(N'-1)}{h'_y(N)}&=U(M,N)\end{aligned} \tag{4.20}$$

where the avalanche induced current generation $G(M,N)$ is due to impact ionisation. It is assumed to be zero for the moment but its implementation is needed later.

The generation-recombination rate $U(M,N)$ (denoted as $R(M,N)$ previously) is assumed to obey the SRH model which is given as

$$U(M,N)=\frac{p(M,N)n(M,N)-n_i^2(M,N)}{\tau_n[p(M,N)+n_i(M,N)]+\tau_p[n(M,N)+n_i(M,N)]}\ . \tag{4.21}$$

Note that $U(M,N)$ is simplified for a mid-gap defect which does not require properties of the energy level or cross section. It determines the leakage current in the device simulation using the measured generation lifetime.

The Poisson equation is written in the finite difference expression as

$$\begin{aligned}&\frac{1}{h'_x(M)}\Big[\frac{\phi(M+1,N)-\phi(M,N)}{h'_x(M')}-\frac{\phi(M,N)-\phi(M-1,N)}{h'_x(M'-1)}\Big]\\ &+\frac{1}{h'_y(N)}\Big[\frac{\phi(M,N+1)-\phi(M,N)}{h'_y(N')}-\frac{\phi(M,N)-\phi(M,N-1)}{h'_y(N'-1)}\Big]\\ &=\frac{q}{\epsilon_{Si}}[N_D(M,N)-N_A(M,N)+p(M,N)-n(M,N)\\ &\qquad+\Sigma N_D^*(M,N)-\Sigma N_A^*(M,N)]\end{aligned} \tag{4.22}$$

where $N_{D/A}$ and $N^*_{D/A}$ are shallow and effective deep defects. 2D Kurata implements the standard SRH mode and quasi-Fermi level approximation for deep levels which were discussed in Section 3.5.2 and 3.5.3. They are also named as Kurata version 6 and version 4 which are summarised in Table 4.1.

4.2.2 2D Matrix-Vector Representation

Kurata suggested the successive over-relaxation (SOR) method, or the successive line over-relaxation (SLOR) method for 2D simulation specifically [68]. It solves each line as a 1D case and uses the solution as approximate initial values for iteration. It then calculates the coupling between three neighbouring lines $(N-1, N, N+1)$ and add increments to the original set of lines. The SLOR method reduces some calculation time, but increases the coding time. Instead, the 2D recursive method uses the same

Kurata	Language	Dimension	Defect occupancy model
version 4	Fortran/C++/ *Mathematica*	1D/2D	one representative acceptor for all deep acceptors and one representative donor for all deep donors energy level can be defined by defects and is normally set to mid-gap
version 5	Fortran	1D	one dominant deep SRH acceptor or donor (not used in the thesis)
version 6	C++	2D	a single deep SRH acceptor and a single deep SRH donor with given cross sections and energy levels

Table 4.1: Comparison of different Kurata versions.Kurata version 4 in 2D is the main simulator used in the thesis which includes more models and will be discussed in Chapter 6. Note that Kurata version 6 uses the same defect model as TCAD which will be compared later.

concept of the 1D recursion, namely the iteration runs through point by point on a line and then repeats the same process on the next line.

The truncated Taylor expansion for different components are similar to the 1D case and their details are skipped here. Similarly, the three coupled semiconductor equations are written in variables of $p(M,N)$, $n(M,N)$ and $\phi(M,N)$. They are rearranged in the matrix-vector form in 2D as

$$\begin{aligned} &A(M,N)\delta\theta(M,N-1)+B(M,N)\delta\theta(M,N)+C(M,N)\delta\theta(M,N+1)\\ &\quad +D(M,N)\delta\theta(M-1,N)+E(M,N)\delta\theta(M+1,N)=F(M,N)\ . \end{aligned} \tag{4.23}$$

The coefficients matrices for known increments $B(M,N)$, $A(M,N)$ and $C(M,N)$ in the y direction (or $B(M,N)$, $D(M,N)$ and $E(M,N)$ in the x direction), are similar to $B(N)$, $A(N)$ and $C(N)$ in the 1D case and their details are skipped here.

4.2.3 2D Recursive Method

The similar modified algorithm reduces the unknowns for $\delta\theta$ from five down to three. The new recursion relation is written as

$$\begin{aligned} &B'(M,N)\delta\theta(M,N)+C'(M,N)\delta\theta(M,N+1)\\ &\quad +E'(M,N)\delta\theta(M+1,N)=F'(M,N) \end{aligned} \tag{4.24}$$

which can be rearranged as

$$\begin{aligned} &\delta\theta(M,N)=B'^{-1}(M,N)F'(M,N)\\ &-B'^{-1}(M,N)C'(M,N)\delta\theta(M,N+1)-B'^{-1}(M,N)E'(M,N)\delta\theta(M+1,N)\ . \end{aligned} \tag{4.25}$$

The unknown $\delta\theta(M, N-1)$ and $\delta\theta(M-1, N)$ are obtained from the new recursion relation (Equation 4.25) as

$$\begin{aligned}\delta\theta(M,N-1) &= B'^{-1}(M,N-1)F'(M,N-1)\\ -B'^{-1}(M,N-1)C'(M,N-1)\delta\theta(M,N) &- B'^{-1}(M,N-1)E'(M,N-1)\delta\theta(M+1,N-1)\\ \delta\theta(M-1,N) &= B'^{-1}(M-1,N)F'(M-1,N)\\ -B'^{-1}(M-1,N)C'(M-1,N)\delta\theta(M-1,N+1) &- B'^{-1}(M-1,N)E'(M-1,N)\delta\theta(M,N)\ .\end{aligned} \tag{4.26}$$

By inserting Equation 4.26 back in Equation 4.23, the relation between the original and primed coefficient matrices are given as

$$\begin{aligned}&A(M,N)[B'^{-1}(M,N-1)F'(M,N-1)\\ -B'^{-1}(M,N-1)C'(M,N-1)\delta\theta(M,N) &- B'^{-1}(M,N-1)E'(M,N-1)\delta\theta(M+1,N-1)]\\ &+B(M,N)\delta\theta(M,N)+C(M,N)\delta\theta(M,N+1)\\ &+D(M,N)[B'^{-1}(M-1,N)F'(M-1,N)\\ -B'^{-1}(M-1,N)C'(M-1,N)\delta\theta(M-1,N+1) &- B'^{-1}(M-1,N)E'(M-1,N)\delta\theta(M,N)]\\ &+E(M,N)\delta\theta(M+1,N)=F(M,N)\ ,\end{aligned}$$

which can be rearranged as

$$\begin{aligned}&[B(M,N)-A(M,N)B'^{-1}(M,N-1)C'(M,N-1)\\ &-D(M,N)B'^{-1}(M-1,N)E'(M-1,N)]\delta\theta(M,N)\\ &+C(M,N)\delta\theta(M,N+1)+E(M,N)\delta\theta(M+1,N)\\ &=F(M,N)-A(M,N)B'^{-1}(M,N-1)F'(M,N-1)\\ &-D(M,N)B'^{-1}(M-1,N)F'(M-1,N)\\ &-A(M,N)B'^{-1}(M,N-1)E'(M,N-1)\delta\theta(M+1,N-1)\\ &-D(M,N)B'^{-1}(M-1,N)C'(M-1,N)\delta\theta(M-1,N+1)\ .\end{aligned} \tag{4.27}$$

By comparing the coefficients of the primed recursion relation (Equation 4.24) and Equation 4.27, the 2D recursion relations are obtained as

$$\begin{aligned}B'(M,N) = B(M,N) &- A(M,N)B'^{-1}(M,N-1)C'(M,N-1)\\ &-D(M,N)B'^{-1}(M-1,N)E'(M-1,N)\end{aligned} \tag{4.28a}$$

$$C'(M,N) = C(M,N) \tag{4.28b}$$

$$E'(M,N) = E(M,N) \tag{4.28c}$$

$$\begin{aligned} F'(M,N) = F(M,N) &- A(M,N)B'^{-1}(M,N-1)F'(M,N-1) \\ &- D(M,N)B'^{-1}(M-1,N)F'(M-1,N) \\ &- A(M,N)B'^{-1}(M,N-1)E'(M,N-1)\delta\theta(M+1,N-1) \\ &- D(M,N)B'^{-1}(M-1,N)C'(M-1,N)\delta\theta(M-1,N+1) \end{aligned} \tag{4.28d}$$

For 1D cases, two end points are usually fixed. For 2D cases, the circumferences (boundary conditions) also have to be set to known values to obtain δy. They can be fixed or floating which depend on their physical nature. Each iteration procedure starts forwards from $(M,N) = (2,2)$ with $\delta y(1,N) = 0$, $\delta y(L_x,N) = 0$ and $\delta y(M,1) = 0$. The program calculates across a line, $M = 2$ to $M = L_x - 1$, as in 1D at $N = 2$. This line will be the boundary for the next line at $N = 3$. The iteration procedure runs through all lines to $N = L_y - 1$ to obtain all primed coefficients. By setting $\delta y(M,L_y) = 0$, $\delta y(M,L_y - 1)$ is obtained. Since the bulk edges of the simulated region are floating boundaries, their solutions are set to the neighbouring lines which are equivalent to zero flux.

The procedure runs backwards from $N = L_y - 1$ to $N = 2$ in the y direction and $M = L_x - 1$ to $M = 2$ in the x direction to have all unknown increment $\delta y(M,N)$. The increments are added to the present values, $y(M,N) = y_0(M,N) + \delta y(M,N)$, for the next iteration. Similarly, it is not free to reduce the unknowns, loops from 2 to $L_x - 1$ and 2 to $L_y - 1$ are performed forward and backwards in 2D and its computing time can be roughly estimated as $(2L_xL_y/2L) \times (5/3)$ times longer than in 1D. The convergence speed is also much slower than in 1D due to complex coupling. A default tolerance fraction of 10^{-4} for the increments is used to control the solution stability. In another words, the simulator stops when the changes of all variables are less than 0.01%. Note that L and $L_{x/y}$ are the grid sizes in 1D and 2D (x/y) respectively.

4.3 Electrical Boundary Conditions

Dirichlet (first-type) and Neumann (second-type) are two main types of boundary conditions (BCs) for an ordinary or partial differential equation. They are discussed in this section and used in simulation. There are also the Cauchy boundary condition which applies the Dirichlet and Neumann BCs together and the mixed boundary condition which applies the Dirichlet BC at some region and the Neumann BC at another region. They are used for specific problems.

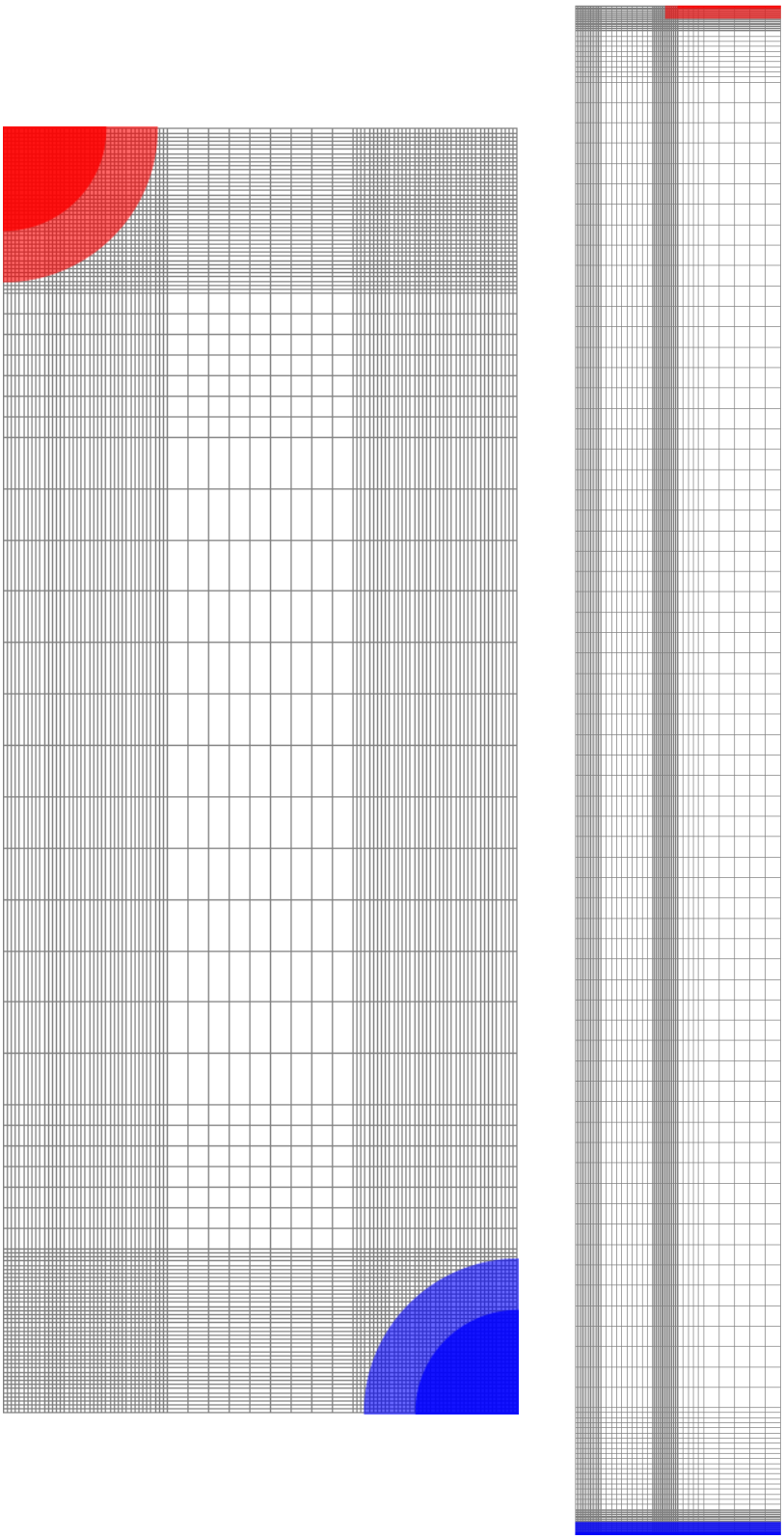


Figure 4.1: An example of grids chosen for a 3D and a planar structure. The red and blue region are $n+$ and $p+$ electrode whose doping concentrations charge rapidly. The grid lines have to be denser in those regions with significant changes in concentration (e.g. electrodes) or geometry (e.g. tips).

4.3.1 Dirichlet Boundary Condition

The Dirichlet boundary condition can also be described as a direct BC, since it specifies the values of the solution at the boundary. Ohmic contacts are Dirichlet BC in device simulation. The charge neutrality, $n_0 - p_0 = N_D - N_A$, has to hold since carriers are free to move without the barrier at Ohmic contacts[4]. The mass action law, $n_0 p_0 = n_i^2$, holds in the whole device. For Boltzmann statistics, variables are expressed as

$$\phi = \phi_F + \frac{kT_B}{q} sinh^{-1}(\frac{N_D - N_A}{2n_i}) \tag{4.29}$$

$$n_0 = \sqrt{\frac{(N_D - N_A)^2}{4} + n_i^2} + \frac{N_D - N_A}{2} \tag{4.30}$$

$$p_0 = \sqrt{\frac{(N_D - N_A)^2}{4} + n_i^2} - \frac{N_D - N_A}{2} \tag{4.31}$$

where ϕ_F is the Fermi potential. ϕ_F is also the applied bias voltage if the contact is not a resistive. The equation for ϕ is a general boundary condition. $sinh^{-1}[(N_D - N_A)/2n_i]$ can be approximated by Equation 2.13 for pn junctions. The doping profile for 3D and planar sensors are $n_0 = N_D$ and $p_0 = n_i^2/N_D$ at the $n+$ electrode and $p_0 = N_A$ and $n_0 = n_i^2/N_A$ in the p bulk or $p+$ electrode.

4.3.2 Neumann Boundary Condition

The Neumann boundary condition can also be described as a floating BC, since it specifies the derivative values of the solution at the boundary. For device simulation, edges of a semiconductor material and the symmetric axis (or plane) within the bulk are Neumann BC. Floating BCs are usually set to zero. The following normal derivatives,

$$\begin{aligned} \frac{\partial \phi}{\partial \mathbf{n}} = \nabla \phi \cdot \mathbf{n} = 0 \quad &\longrightarrow \quad E = 0 \\ \frac{\partial n}{\partial \mathbf{n}} = \nabla n \cdot \mathbf{n} = 0 \quad &\longrightarrow \quad J_n = 0 \\ \frac{\partial p}{\partial \mathbf{n}} = \nabla p \cdot \mathbf{n} = 0 \quad &\longrightarrow \quad J_p = 0 \ , \end{aligned} \tag{4.32}$$

state that there is no net electric field and also no flowing current across/on the simulation boundary. A zero (net) flow is due to the mirror/geometric symmetry on the two sides of the edge which implies the symmetry in the solution.

[4]Ohmic (metal) contacts have a resistance assumed to be 0.001 Ω when connected to a circuit in TCAD and zero otherwise. Ohmic contacts in semiconductors can be thought as a negligible depleted region with a highly doped junction.

4.4 Generation Lifetime

The generation lifetime defines the leakage current in the device through the SRH generation recombination process. The leakage current also depends on the depletion width w and thus the bias voltage V, since $w \propto \sqrt{V}$ for uniform doped diodes (not the case for graded junctions) before full depletion. The generation lifetimes and capture cross sections of deep levels in the SRH model affects their occupancy. Kurata defines the leakage current (by the recombination rate U) and occupancy functions (based on Section 3.5.2 and 3.5.3) separately. However they interact through the carrier densities and adapt the solution. It is essential to use a reasonable generation lifetime for numerical stability.

4.4.1 SRH Doping Dependence

The Scharfetter relation depends only on the doping concentration $N_{A/D}$ which is proportional to collision and recombination probability. It is given as [39][40]

$$\tau_{dop}(N_{A,0}, N_{D,0}) = \tau_{min} + \frac{\tau_{max} - \tau_{min}}{1 + (\frac{N_{A,0}+N_{D,0}}{N_{ref}})^1} \tag{4.33}$$

where τ_{max} is 1×10^{-5} and 3×10^{-6} s for electrons and holes.$\tau_{min} = 0$ s and $N_{ref} = 1 \times 10^{16}$ cm^{-3} are the same for both carriers.

4.4.2 SRH Temperature Dependence

A calculation using the low-temperature approximation of the multi-phonon theory gives the temperature dependence for carrier lifetime as [115]

$$\tau_{temp}(T) = \tau_{temp}(300K)(\frac{T}{300K})^{1.5} \tag{4.34}$$

The temperature dependence is strongly related to the nature of the recombination centre. The two SRH lifetimes are usually combined. There are also other models which depend on the electric field or tunnelling.

4.4.3 Experimental Generation Lifetimes

The two SRH dependences described earlier are only used for unirradiated cases. The experimental generation lifetime τ_g used in device simulation is obtained from leakage current measurements. The trapping lifetimes τ_t are obtained from the transient current technique (TCT) and used in the tracking simulation. The tracking simulation tracks electron-hole pairs in the device and will be discussed in Chapter 5

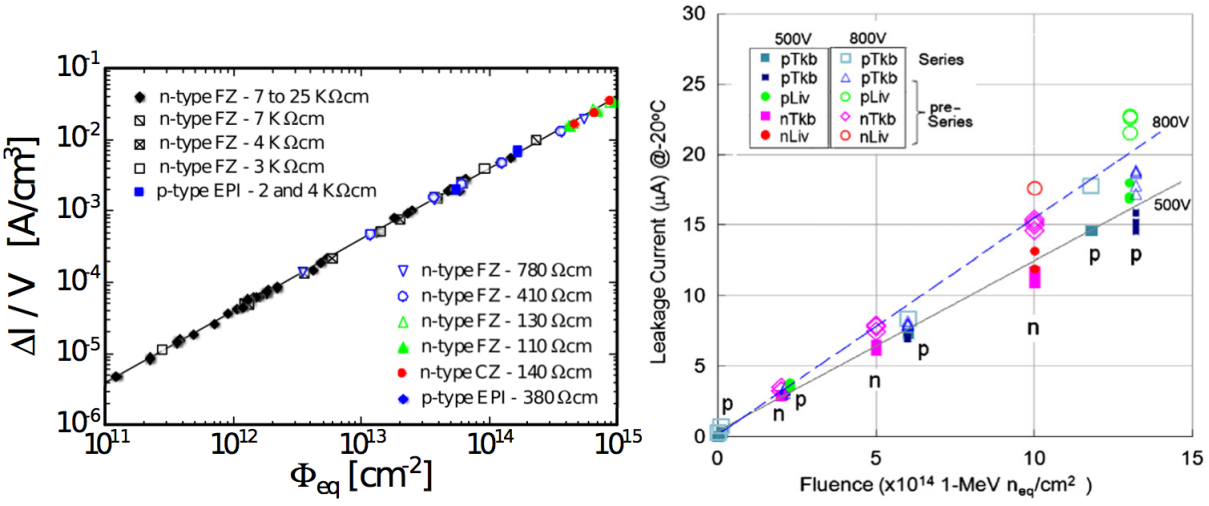


Figure 4.2: Leakage current is linear with the fluence level (left) [84] (right) [51].

The generation lifetime is related to the fluence as

$$\frac{1}{\tau_g} = \frac{1}{\tau_{0,g}} + \frac{\Phi_{eq}}{K_{\tau,g}} \approx \frac{\Phi_{eq}}{K_{\tau,g}} \tag{4.35}$$

where $K_{\tau,g}$ is the damage constant for the effective generation lifetime. The initial bulk background leakage current ($\sim 1/\tau_{0,g}$) is small and negligible after heavy irradiation.

The leakage current $\triangle I$ per unit of depleted volume V_{dep} is measured as a function of the fluence as

$$\alpha = \frac{\triangle I}{V_{dep}\Phi_{eq}} = \frac{qn_i}{2K_{\tau,g}} \tag{4.36}$$

where α is the leakage current damage constant. The leakage current is measured to be linear with fluence up to a few 10^{15} $n_{eq}cm^{-2}$ as shown in Figure 4.2. It is assumed to be linear for higher fluences.

α is recorded as 7×10^{-17} and $(3.99\pm0.03)\times10^{-17}$ Acm^{-1} at 293 K by the RD20 [81] and RD48 [74] collaborations respectively. The former is measured after irradiation, while the latter is measured after two weeks which is thought to be stable. It can be parametrised in a function of the annealing time given as [84]

$$\alpha(t) = \alpha_0 + \alpha_1 e^{-\frac{t}{\tau_1}} - \alpha_2 ln(\frac{t}{t_0}) \tag{4.37}$$

where the exponential and logarithm term are for the long term annealing at room temperature and high temperatures respectively. Parameters were obtained by fitting experimental data and are listed in [84]. Figure 4.3 shows that α decreases with time due to the beneficial annealing time. However, at some point reverse annealing occurs which increases N_{eff}, but the leakage current is not correlated to this change.

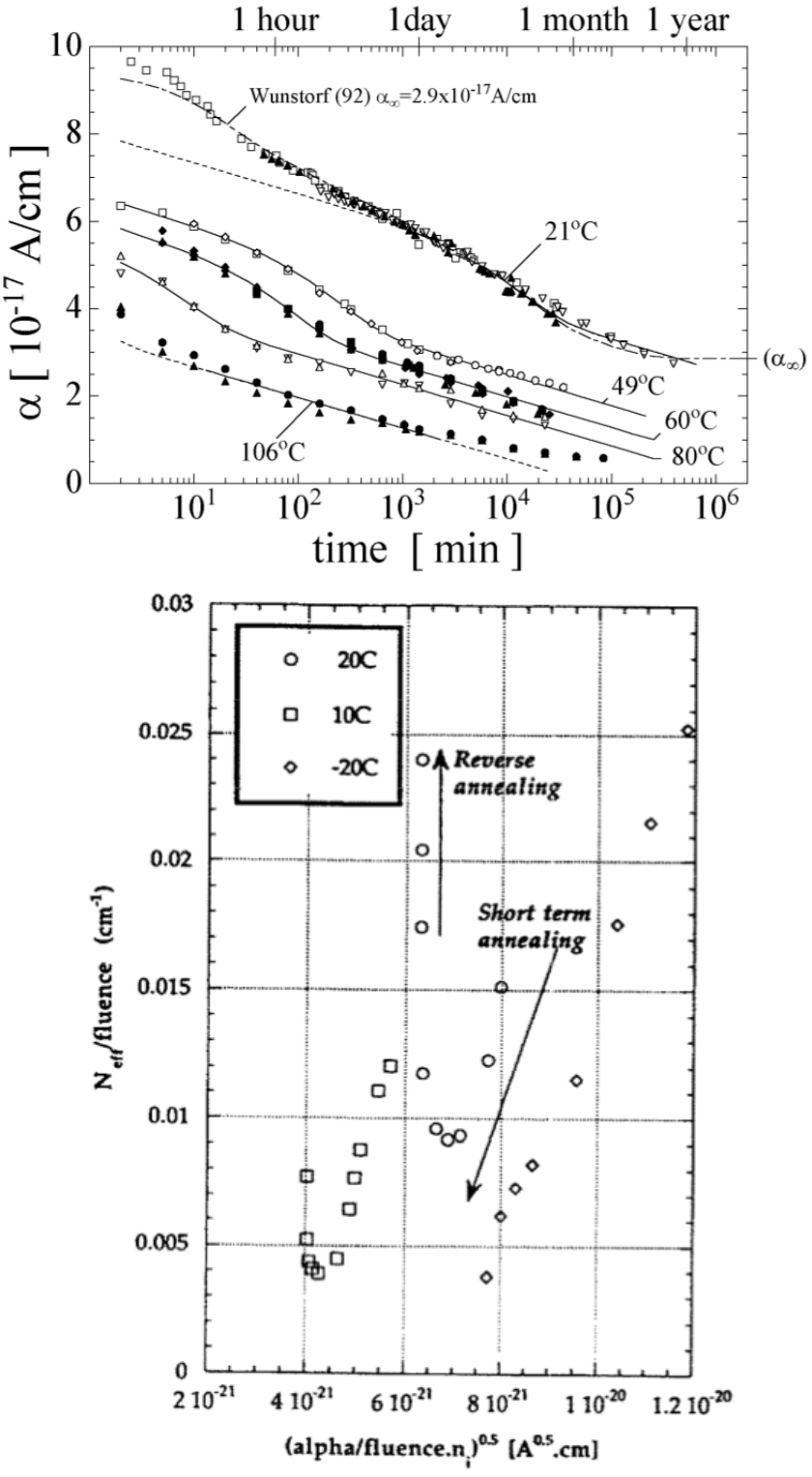


Figure 4.3: The unit leakage current decreases due to the beneficial annealing (top) [84] to a certain level, but the effective defect concentrations increases with time due to the reverse-annealing (bottom) [82]. This implies that one or more effective deep levels are generated while reverse annealing but the cluster structure is already stabilised.

4.5 Diffusion of Shallow Dopants

Doped regions are generated by ion implantation and dopant diffusion or epitaxy. The doping profile around the electrodes alters the term in the Poisson equation and affects the electric field around the junction.

4.5.1 Fick's Law

The microscopic view of dopant diffusion is that an impurity moves in a periodic potential formed by the lattice atoms [57]. Fick's first law states that the diffusion flux J is proportional to the concentration gradient given as

$$J = -D\frac{\partial C}{\partial x} \tag{4.38}$$

where D and C are the diffusion constant and concentration of the dopants respectively.

Fick's second law restates the dopants as a time evolving cloud given as

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{4.39}$$

by relating the flux as a time differential of the concentration. The solution to Fick's second law for a simple 1D can be obtained by solving the differential equation as

$$C(x,t) = C_0\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-s^2}ds = C_0 erfc(\frac{x}{2\sqrt{Dt}}) \tag{4.40}$$

where $erfc$ is the complementary error function[5].

4.5.2 Doping Profile

The time evolution of dopnats is not important for defining the doping profile in device simulation since it is assumed to be stable at the operation temperature. The doping concentration is approximated as a Guassian function which is written as

$$C(x) = C_1 e^{-\frac{1}{(X_j - R_p)^2} ln(\frac{C_1}{C_0})(x-R_p)^2} \tag{4.41}$$

where C_1 is the maximum implantation at the surface and C_0 is the minimum impurity of the bulk. C_1 is set as the injection depth R_p which is around 1 to 2 μm. C_0 is set as the junction depth X_j which is around 2 to 3 μm. The Gaussian doping profile is used in simulation which controls the junction position. It can also be shown that

[5]The error function is defined as $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-s^2}ds$ and the complementary error function is defined as $1 - erf(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-s^2}ds$

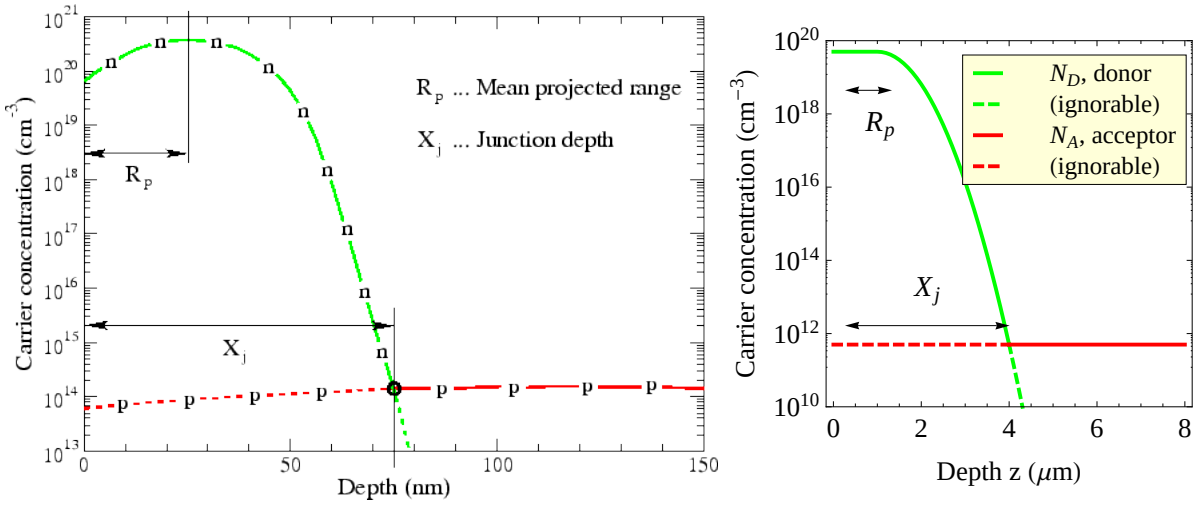


Figure 4.4: A realistic doping profile of a typical *pn* junction (left) [130] whose maximum concentration is at a few tens of *nm* beneath the surface. An approximate profile for the device simulation (right) where the Gaussian function is used. It is assumed as to be constant from the surface to the mean projected range. This does not alter the solution since the junction forms at the degraded doping region.

the difference between the error function and the Gaussian function is not large [57]. Figure 4.4 shows an error function and a Gaussian function shaped doping profile of a *pn* junction and a 3D electrode respectively.

4.6 Mobility of Free Carriers

The thermal energy of carriers is related to Boltzmann's constant and temperature which is also the kinetic energy without the electric field as

$$\frac{1}{2}m_{n/p}v_{th,n/p}^2 = \frac{3}{2}k_BT \tag{4.42}$$

where 3 represents the three degrees of freedom for carriers in the bulk. Carriers move through the silicon material and are randomly scattered by lattice silicon atoms and impurities. The average collision time τ_c, which corresponds to the collision probability per unit time $1/\tau_c$, is given as

$$\frac{1}{\tau_c} = \frac{1}{\tau_{c,lattice}} + \frac{1}{\tau_{c,impurity}} \ , \tag{4.43}$$

which is also the mean free time if no electric field is present. The corresponding mean free path λ is the average distance between collisions which has a typical value of around 10^{-5} cm as $\tau_c \sim 1$ ps. This value of τ_c is an important baseline for the time step choice which will be discussed in Chapter 5.

	Hole mobility μ_p	Electron mobility μ_n
N_{ref} (cm^{-3})	6.3×10^{16}	8.5×10^{16}
μ_{max} $(cm^2 s^{-1} V^{-1})$	495	1330
μ_{min} $(cm^2 s^{-1} V^{-1})$	47.7	65
E_{crit} $(V cm^{-1})$	1.95×10^4	8×10^3
α	0.76	0.72
β	1	2

Table 4.2: Fitting parameters of the Caughey-Thomas model for the electron and hole mobility.

When an electric field is applied, the drift velocity for electrons can be derived classically from momentum as

$$F = \triangle t = -qE\tau_c = m_n v_n \quad \text{and} \quad v_n = -(\frac{q\tau_c}{m_n})E \equiv -\mu_n E \tag{4.44}$$

where μ_n is the electron mobility. Similarly, the drift velocity for holes is $v_p = \mu_p E$ where μ_p is the hole mobility.

4.6.1 Caughey-Thomas Parameters

Carrier mobilities depend on many factors, such as the impurity concentration and electric field. The fitted empirical formula of the electron μ_n and hole mobilities μ_p by Caughey and Thomas are given as [30]

$$\mu_{p,n}(N, E) = [\frac{\mu_{max} - \mu_{min}}{1 + (\frac{|N|}{N_{ref}})^{\alpha} + f} + \mu_{min}] \cdot \frac{1}{[1 + (\frac{E}{E_{crit}})^{\beta}]^{1/\beta}} \tag{4.45}$$

where $f = [(pn)^{0.5}/2.04N_{ref}]^{\alpha}$ is a factor introduced as a correction for high injection levels as carrier-to-carrier scattering is important. The fitting parameters are given in Table 4.2.

Some later fitting models are implemented in Kurata based on the TCAD settings. If the bulk mobility is obtained from more than one model, they are usually combined by Matthiessen's rule as

$$\frac{1}{\mu} = \frac{1}{\mu_{Masetti}} + \frac{1}{\mu_1} + \frac{1}{\mu_2} ... \tag{4.46}$$

which is the same principle to bring lifetimes from different models together. Many models have been tested in 1D and 2D, but no significant difference is observed. It is more important for the tracking simulation since the collection time depends on the carrier mobilities and electric field, especially after heavy irradiation.

	Hole mobility μ_p	Electron mobility μ_n
μ_{min1} (cm^2/Vs)	52.2	44.9
μ_{min2} (cm^2/Vs)	52.2	0
μ_1 (cm^2/Vs)	43.4	29.0
P_c (cm^{-3})	0	9.23×10^{16}
C_r (cm^{-3})	9.68×10^{16}	2.23×10^{17}
C_s (cm^{-3})	3.43×10^{20}	6.10×10^{20}
α	0.680	0.719
β_0	1.109	1.213
β_{exp}	0.66	0.17
$v_{sat,0}$ (cms^{-1})	1.06×10^6	8.37×10^6
γ	0.87	0.52

Table 4.3: Fitting parameters of the Masetti and extended Canali model for the electron and hole mobility which are doping and electric field dependent respectively.

4.6.2 Mobility due to Phonon Scattering

The constant mobility model considers only phonon scattering which depends on only lattice temperature, given as [115]

$$\mu_{const} = \mu_{bulk}(\frac{T}{300})^{-\zeta} \tag{4.47}$$

where μ_{bulk} is 1417 and 470.5 cm^2s^{-1}, and ζ is 2.5 and 2.2 for electrons and holes respectively. This is similar to the temperature dependent generation lifetime.

4.6.3 Mobility due to Doping Degradation

The Masetti model is used to calculate the doping-dependence of mobility [80] as

$$\mu_{dop} = \mu_{min1}e^{\frac{-P_c}{N_{A,0}+N_{D,0}}} + \frac{\mu_{const}-\mu_{min2}}{1+(\frac{N_{A,0}+N_{D,0}}{C_r})^{\alpha}} - \frac{\mu_1}{1+(\frac{C_s}{N_{A,0}+N_{D,0}})^2} \ . \tag{4.48}$$

The fitting parameters are given in Table 4.3 [115].

4.6.4 Mobility due to Field Saturation

The high-field saturation models comprise three sub-models: the extended Canali mobility mode, the velocity saturation and driving force model [115]. The Canali model originates from the Caughey-Thomas formula [30] which was fitted for temperature up to 430 K as

$$\mu = \frac{(\alpha+1)\mu_{low}}{\alpha+[1+(\frac{(\alpha+1)\mu_{low}F_{hfs}}{v_{sat}})^{\beta}]^{1/\beta}} \tag{4.49}$$

where μ_{low} is the low field mobility and is from the Masetti model described above.

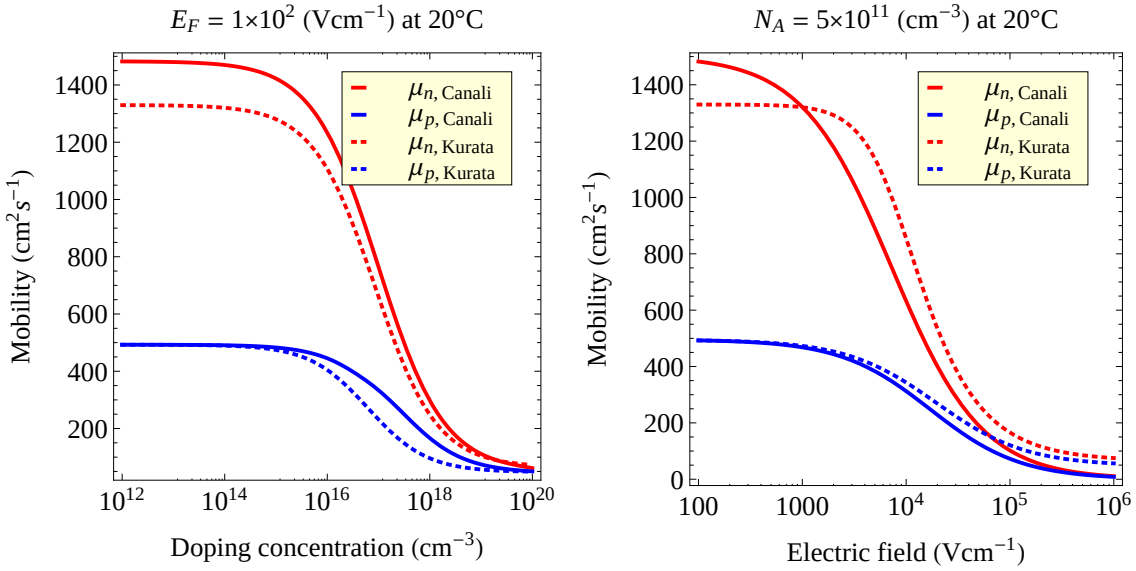


Figure 4.5: The electron mobility (red) and hole mobility (blue) are plotted as a function of the doping concentration and temperature (left, the lighter in colour the higher in magnitude) and electric field (right). The trends of two mobilities are similar which are higher at lower doping concentration and temperature, a lower collision probability. Mobilities saturate at a high electric field. The Canali model is more sensitive to the electric field with the bulk concentration $N_A = 5 \times 10^{11}\ cm^{-3}$ at 293 K.

The saturation velocity v_{sat} for silicon is given as

$$v_{sat} = v_{sat,0}(\frac{300}{T})^{\gamma} \tag{4.50}$$

which is temperature dependent. The driving force for electrons or holes is the gradient of the quasi-Fermi level, $F_{hfs,n/p} = |\nabla\phi_{n/p}|$. It is simply treated as the potential gradient in simulation, i.e. the electric field $|\nabla\phi|$. The fitting parameters are also given in Table 4.3.

The doping (left) and electric field (right) dependence are shown in Figure 4.5 that lower doping concentrations and electric fields give a larger mobility for both carriers. Lower temperatures also give a larger mobility which is not shown since temperature is assumed to be uniform over the device.

4.7 Impact Ionisation

Impact ionization is the most standard process for carrier generation (charge multiplication) [68]. A free carrier is accelerated and gains energy in a strong electric field, new carriers are generated whilst the absorbed energy is sufficient to excite an electron-hole

pair. This process can be written explicitly as [117]

$$E_0 = \frac{1}{2}m_e v_i^2 = E_g + 2(\frac{1}{2}m_e v_f^2) + \frac{1}{2}m_h v_f^2 \approx E_g + 3(\frac{1}{2}m_e v_f^2)$$
$$m_e v_i = 2m_e v_f + m_h v_f \approx 3m_e v_f \tag{4.51}$$

which use energy and momentum conservation. One can find that the initial energy $E_0 \geq 1.5E_g$. The actual energy required is around 3.6 and 5 eV for electrons and holes respectively. Generated carriers are also accelerated to excite further electron-hole pairs. Avalanche photo-diodes (APDs) facilitate this effect and usually have an internal gain around 100. Current generation from impact ionization is expressed as

$$G = \frac{1}{q}(\alpha_n |J_n| + \alpha_p |J_p|) \tag{4.52}$$

where α_n and α_p are ionization rates given as

$$\alpha = Ae^{-\frac{B}{|E|}} \tag{4.53}$$

$1/\alpha$ is a typical length that a carrier has to travel to obtain enough energy/probability for creating an electron-hole pair. To the lowest order of fitting, coefficient A is 3.80×10^6 and 2.25×10^7 cm^{-1} and B is 1.75×10^6 and 3.26×10^6 Vcm^{-1} for electrons and holes respectively. $\alpha_e > \alpha_h$ is due to the smaller collision energy required for electrons.

4.7.1 Crowell-Sze Model

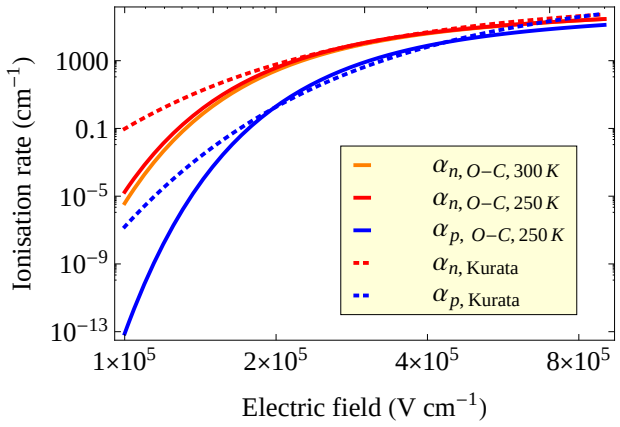
Crowell [33] and Sze [116] made a fit of Baraff's curve to second order. Sutherland [111] improved it with a third order term. The modified coefficients for electrons and holes are given as

$$\alpha_{n/p} = \frac{1}{\lambda_{n/p}} e^{[\frac{A\epsilon_i^2}{q^2E^2\lambda_{n/p}^2} + \frac{B\epsilon_i}{qE\lambda_{n/p}} + C]} \tag{4.54}$$

where

$$A = 11.5r^2 - 1.17r + 3.9 \times 10^{-4}\ , \quad B = 46r^2 - 11.9r + 1.75 \times 10^{-2}$$
$$C = -757r^2 + 75.5r - 1.92\ , \quad r = \frac{\langle \epsilon_r \rangle}{\epsilon_i} = tanh(\frac{\epsilon_r}{2k_BT})\ .$$

The activation energy ϵ_i is $3E_g/2$ and the Raman optical phonon energy ϵ_r is 0.063 eV. $\langle \epsilon_r \rangle$ is the mean energy loss per optical phonon collision which is approximated as $\epsilon_r tanh(\epsilon_r/2kT)$ on average. The optical phonon mean free path for electrons and holes is given by $\langle \epsilon_r \rangle / \lambda = \epsilon_r / \lambda_0$ as $\lambda = \lambda_0 tanh(\epsilon_r/2kT)$ where λ_0 is 76 and 55 $\mathring{A}$ for electrons and holes.



	α_n	α_p
$a\ (V^{-1})$	0.426	0.243
$b\ (Vcm^{-1})$	4.81×10^5	6.53×10^5
$c\ (K^{-1})$	3.05×10^{-4}	5.35×10^{-4}
$d\ (K^{-1})$	6.86×10^{-4}	5.67×10^{-4}

Figure 4.6: Ionisation rates for electrons (red) and holes (blue) as a function of the electric field. The Okuto-Crowell model gives smaller values at lower electric fields and provides temperature dependence, while Kurata model is independent of temperature. The graph also shows a lower ionisation rate at higher temperatures. The fitting parameters of the empirical formula for the electron and hole ionisation coefficient in the Okuto-Crowell model are given in the table on the right.

4.7.2 Okuto-Crowell Model

Okuto and Crowell [89] put in further effort fitting experimental curves using a simple formula. The ionisation rate is given as

$$\alpha(F_{avg}) = a(1 + c(T - 300))F_{avg}e^{(\frac{b(1+d(T-300))}{F_{avg}})^2} \tag{4.55}$$

which is used in simulation with parameters and curves given in Figure 4.6.

4.7.3 Partial Derivatives of the Generation Factor G

Current generation was ignored previously since sensors are not operated in this extreme condition. It gets important after heavy irradiation since the high effect doping concentration confined the depletion width and leads to a high electric field. Given $m_a = h(M)/2h'(N)$ and $m_b = h(M-1)/2h'(N)$, the derivatives with respect to the carrier densities in 1D are given as

$$\begin{aligned}
\frac{\partial G(N)}{\partial n(N-1)} &= \pm\frac{m_a}{h(M-1)}\lambda_{n1}(M-1)\alpha_n(N) \\
\frac{\partial G(N)}{\partial n(N)} &= \pm[\frac{m_a}{h(M-1)}\lambda_{n2}(M-1) + \frac{m_b}{h(M)}\lambda_{n1}(M)]\alpha_n(N) \\
\frac{\partial G(N)}{\partial n(N+1)} &= \pm\frac{m_b}{h(M-1)}\lambda_{n2}(M-1)\alpha_n(N)
\end{aligned} \tag{4.56}$$

where the $\pm$ sign depends on $Jn(N) \gtrless 0$ and $J_n(N) = m_a J_n(M-1) + m_b J_n(M)$. The electric field used for α_n is an average value given as $E(N) = m_a E(M-1) + m_b(M)$.

Derivatives with respect to the potential in 1D are given as

$$\begin{aligned}
\frac{\partial G(N)}{\partial \phi(N-1)} &= \frac{E(N)}{|E(N)|}\frac{m_a}{h(M-1)}\alpha_n(N)\frac{E_{n0}}{E(N)^2}\frac{|J_n(N)|}{q} + \frac{J_n(N)}{|J_n(N)|}\frac{m_a\alpha_n(N)}{q}\frac{\partial J_n(M-1)}{\partial \phi(N-1)} \\
\frac{\partial G(N)}{\partial \phi(N+1)} &= -\frac{E(N)}{|E(N)|}\frac{m_b}{h(M)}\alpha_n(N)\frac{E_{n0}}{E(N)^2}\frac{|J_n(N)|}{q} - \frac{J_n(N)}{|J_n(N)|}\frac{m_b\alpha_n(N)}{q}\frac{\partial J_n(M)}{\partial \phi(N)} \\
\frac{\partial G(N)}{\partial \phi(N)} &= -\frac{\partial G(N)}{\partial \phi(N-1)} - \frac{\partial G(N)}{\partial \phi(N+1)} \ . \qquad (4.57)
\end{aligned}$$

These derivatives are extended to 2D which is the most complicated part of the code and these details are not given here.

4.8 Results and Discussion

This short section shows the effects of including impact ionisation on planar strip structures and a comparison of simulations results for 3D structures between TCAD and Kurata.

4.8.1 Effect of Impact Ionisation on the Solution

A *p+-p-p+* planar strip sensor was simulated and its detailed structure will be shown in Figure 6.1. There will be more discussion about this structure in Chapter 6 and some results for the sensor irradiated at a fluence of 5×10^{14} $n_{eq}cm^{-2}$ show the important of including impact ionisation in the device simulation.

Figure 4.7 (b) and (c) show the electric field distribution with and without impact ionisation due to the high field in the device simulation. At low voltages, the electric field has not reached the energy barrier for impact ionisation to occur and thus the two simulated distributions are identical. At high voltages and high fluences, the electric field on the *n+* side, where there is a *pn+* junction, is strong enough for impact ionisation to occur and generates an extra current.

Without impact ionisation, the voltage drop is confined only to the *pn+* side. With generation in addition to the leakage current, the depletion width increases and the device tends to deplete which reduces the high electric field[6]. The full depletion voltage for an ideal *pn* diode can be estimated by Equation 3.40 which is 600 V for the selected fluence and structure. It is expected to be higher for a strip structure. However, the simulation without impact ionisation is not depleted even at 900 V while it is depleted by including the avalanche enhanced generation current.

[6]The integral of the electric field is a fixed value of voltage and thus an extension of the depletion width leads a drop of the peak field value.

The double junction model was proposed by the CMS collaboration to explain the data [113]. It requires the presence of both deep acceptors and donors in simulation as shown in Figure 4.8. The junction on the $n+$ side is due to type inversion as the CMS device are initially n-type bulk. Deep donor levels provide fixed and positive charges close to the $p+$ junction which results in the second $p+n$ junction. This phenomenon is not observed without impact ionisation in the device simulation and the full depletion is also needed.

4.8.2 Comparison of Kurata and TCAD

Kurata version 6 has the full SRH model for deep levels which is also used in TCAD, and thus they expected to generate similar results. Figure 4.9 shows the comparison between the Kurata and TCAD results of the electric field distribution for a 3D sensor with the IBL layout at a fluence of 5×10^{15} $n_{eq}cm^{-2}$. TCAD uses the three-level defect model [96] which has V_2 and V_3 for deep acceptors, while Kurata considers only V_3 since it dominates N_{eff}. Both use C_iO_i for deep donors which is less important due to its energy level.

The shape of the electric field from TCAD is similar to Kurata which suggests the reliability of the 2D Kurata code and the use of the SRH model. Kurata will give different results using the quasi-Fermi level model since its effective doping concentration is higher. A comparison between different defect and occupancy models will be discussed in Chapter 6.

4.9 Capacitance and Noise

Noise in the readout system increases with the device (input) capacitance and is expressed as an equivalent noise charge Q_n conventionally as [110]

$$Q_n^2 = i_n^2 F_i T_S + e_n^2 F_v \frac{C_i^2}{T_S} + F_{vf} A_f C_i^2 \tag{4.58}$$

where C_i is the sum of the input capacitance. i_n and e_n are input noise current spectral density and voltage respectively (current and voltage noise). F_i, F_v and F_{vf} are shape factors depending on the frequency or time response of the shaper. T_S is a characteristic time. The measured noise of the readout electronics is found to be related to the device capacitance which can be estimated from the electric field given by the device simulation. This section also discusses background information of the threshold, noise and the overdrive of FE-I3 chips.

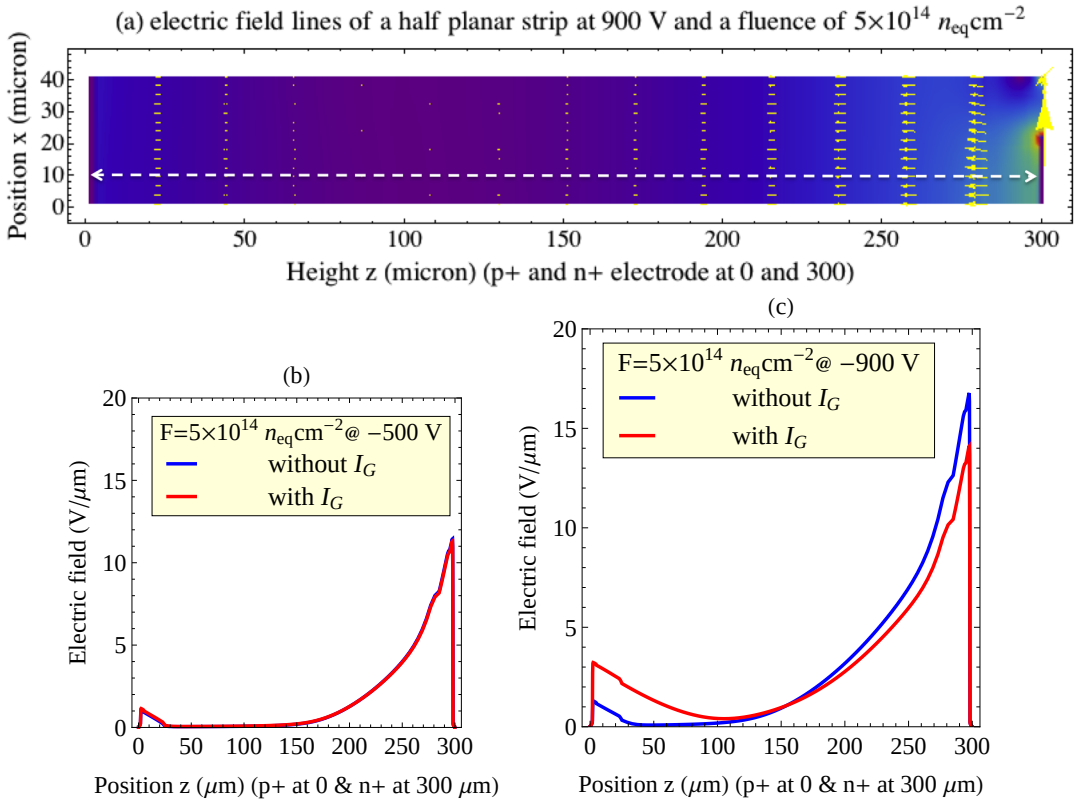


Figure 4.7: Comparison of simulation results with and without impact ionisation for a 300 μm thick n-in-p planar sensor at a fluence of 5×10^{14} $n_{eq}cm^{-2}$. (a) shows the electric field lines of a half-strip of a planar sensor, in which the largest electric field is between $n+$ electrode and p spray. (b) and (c) show the importance of including impact ionisation to generate an extra leakage current due to the high field. (b) has a low bias voltage and the electric field with and without generation current is the same since no impact ionisation occurs. (c) shows that the depletion width extends further if impact ionisation is included. Both simulations have an introduction rate of 1 and 0.2 cm^{-1} for deep acceptors and donors.

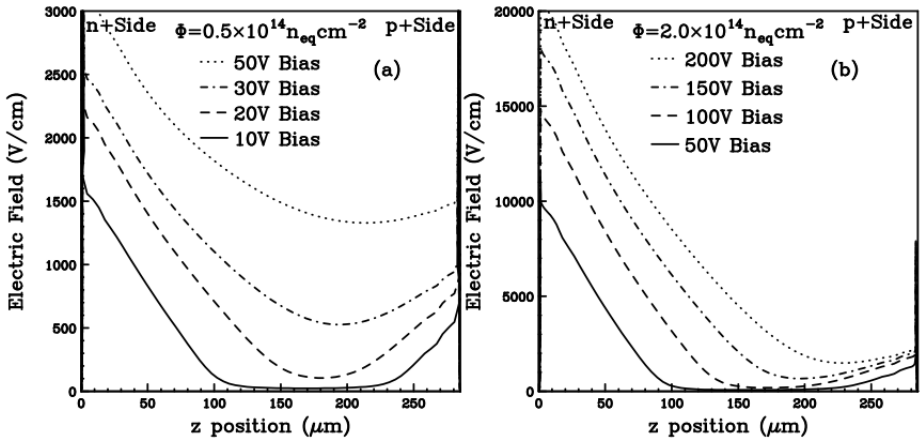


Figure 4.8: Simulation results of the double junction model for n-in-n strip sensors at fluences of 5×10^{13} and 2×10^{14} $n_{eq}cm^{-2}$ with various bias voltages [113].

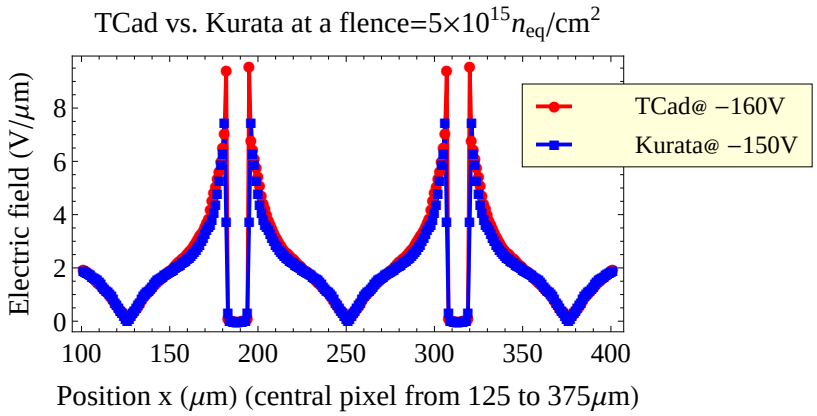


Figure 4.9: Comparison between TCAD and Kurata results of the electric field distribution for a 3D sensor with the IBL layout at a fluence of 5×10^{15} $n_{eq}cm^{-2}$. The TCAD result of the electric field was generated by Povoli at Trento [100].

4.9.1 Threshold and Noise

The threshold is set in the ATLAS readout chip to filter out background noise hits. Depending on individual sensors and irradiation levels, the bare threshold is set to around a few thousand electrons for best operation. Figure 4.10 shows the percentage of hits for different injection levels of electrons. The noise is ±20% around the 50% mean which is the set threshold. A higher threshold usually gives a lower noise.

Sensors are AC-coupled to the readout chips via bump-bonds. The induced signal goes to the amplifier on the readout chip. The discriminator outputs a signal only if the input is greater than the threshold. Signals with smaller amplitudes reach the threshold at a later time than signals starting at the same time but with higher amplitudes. This time difference of the rising edge is called the time-walk. Timings for the rising and falling edges are recorded and their difference is called the time over threshold (ToT). It is digitised to a few bits and is related to the collected charge. There is a precise formula to convert ToT values to electrons for FE-I3 chips which is approximately linear.

The hits suffering from time-walk (> 20 ns) may be associated to the next beam crossing (every 25 ns). The effective threshold is then higher and is also called the in-time threshold. The difference between these two thresholds is called the overdrive. A higher capacitance usually needs a higher overdrive since the rise time τ_r increases with the input capacitance C_i. It is given as [94]

$$\tau_r \approx (C_o C_f + C_o C_i + C_i C_f)(g_m C_f)^{-1} \approx C_i g_m^{-1} \tag{4.59}$$

where C_f is the feedback capacitance of the amplifier and C_f is the output capacitance of circuits after the amplifying stage. g_m is the transconductance of the input transistor.

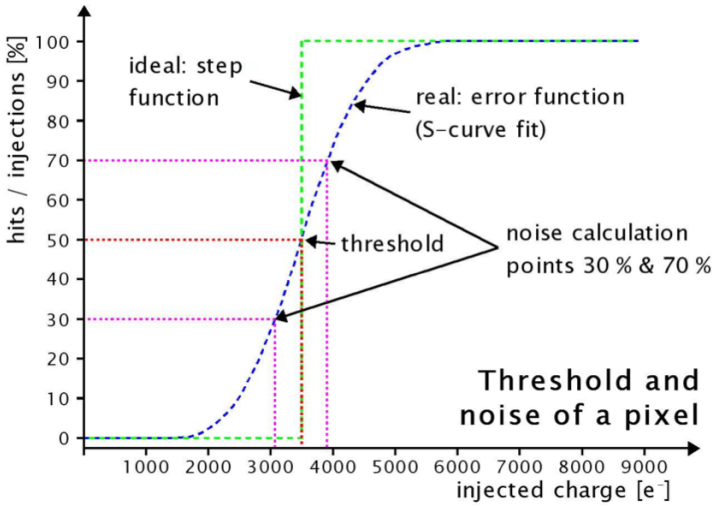


Figure 4.10: A schematic of threshold and noise. The mean of S-curve in a threshold scan is the corresponding threshold which is set as 3500 electrons here. The noise is ±20% around the 50% mean [94].

The rise time is mainly dependent on C_i as C_o is minimised.

4.9.2 Capacitance and Noise

The simplest capacitor is composed of two conducting plates separated by a dielectric with equal positive and negative charge. Separated and opposite charges induce an electric field which stores energy. For a pad structure, the energy stored U_{stored} is given as

$$U_{stored} = \int \frac{1}{2}\vec{D}\cdot\vec{E}d\vec{V_s} = \int \frac{1}{2}\epsilon_{Si}\vec{E}\cdot\vec{E}d\vec{V_s} = \int_0^Q \frac{q}{C}dq = \frac{Q^2}{2C} = \frac{1}{2}CV^2 \tag{4.60}$$

where $d\vec{V_s}$ represents the sensor volume. The sensor capacitance C is calculated by

$$C \approx \frac{2U_{stored}}{V^2} = \frac{2\epsilon_{Si}E^2}{V^2} \tag{4.61}$$

for a flat distribution. The integral form is needed for other non-flat field distribution. Diamond has a smaller dielectric constant of 5.7 ϵ_0 which gives smaller capacitance than silicon (11.7 ϵ_0) intrinsically. A device material using diamond rather than silicon will have a lower capacitance due to the smaller dielectric constant: 5.7 ϵ_0 compared to 11.7 ϵ_0 for carbon and silicon respectively.

Figure 4.11 (left) shows the capacitance calculated by Kurata and FlexPDE. The latter solves the approximate Poisson equation, which will be discussed in Chapter 5, for 2E (103 μm), 3E (71 μm) and 4E (56 μm) structures of 3D sensors with the FE-I3 layout. FlexPDE predicts a lower capacitance at depletion since it does not consider

the dopant diffusion of 2 μm and has a larger inter-electrode distance[7]. This is less effective for the 2E structure since it is relatively larger. FlexPDE also predicts a lower capacitance at low voltages since it uses an approximation for the depletion condition. There is more space charge which confines the depletion width and increases the field density.

Figure 4.11 (right) shows the linear relationship between the capacitance from simulations and the noise from measurements. Planar sensors have a very small capacitance. 3D sensors have a higher capacitance due to their parabolic field distribution[8]. However, planar sensors will also have parabolic field distribution after heavy irradiation as discussed in the previous section. This will also result in an increase in capacitance and thus higher noise.

Figure 4.12 (left) shows the measured capacitances for the 2E, 3E and 4E structure of 3D sensors. They (with the planar capacitance) are plotted against the measured overdrives at different input currents shown in Figure 4.12 (right). It shows that the overdrive is still needed at zero sensor capacitance which suggests an extra capacitance. The external capacitance source, around 80 fF, is thought to be from bump bonds which also results in a noise level around 150 electrons.

[7]By taking the diffusion length into account, Kurata gives the same electric field distribution to FlexPDE at depletion before irradiation which also suggests the reliability of the 2D simulator.

[8]The inequality of arithmetic and geometric means,

$$\frac{n}{\Sigma_{i=1}^{n}\frac{1}{x_i}} \leq \sqrt[n]{\Pi_{i=1}^{n}x_i} \leq \frac{\Sigma_{i=1}^{n}x_i}{n} \leq \sqrt{\frac{\Sigma_{i=1}^{n}x_i^2}{n}}\ , \tag{4.62}$$

shows that the squared sum of a parabola shape (3D sensors' electric field) is larger than a uniform distribution (planar sensors).

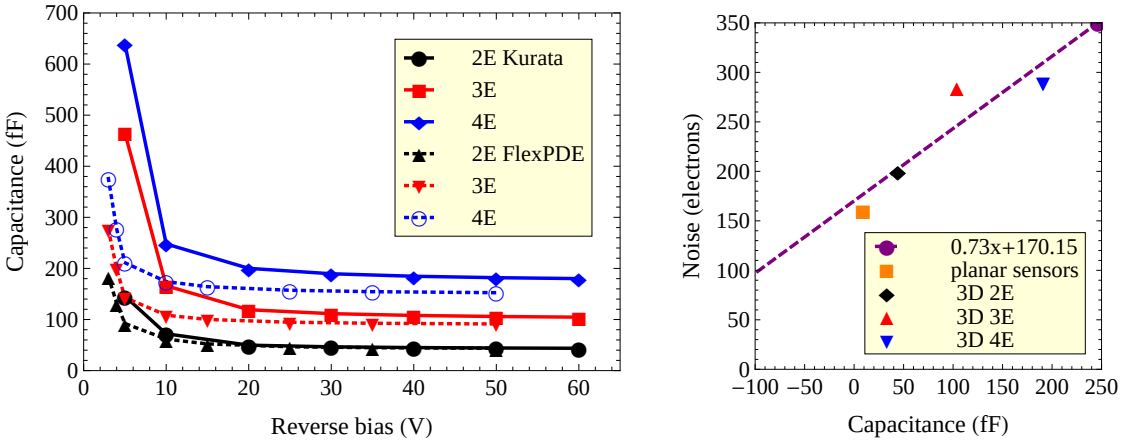


Figure 4.11: Capacitance calculated by Kurata (solid) and FlexPDE (dotted lines) for the 2E, 3E and 4E structure of 3D sensors respectively (left). The linear relationship between the capacitance and noise (right). The fitted function (purple dashed line) at the zero sensor capacitance still gives noise which implies there is an external capacitance source.

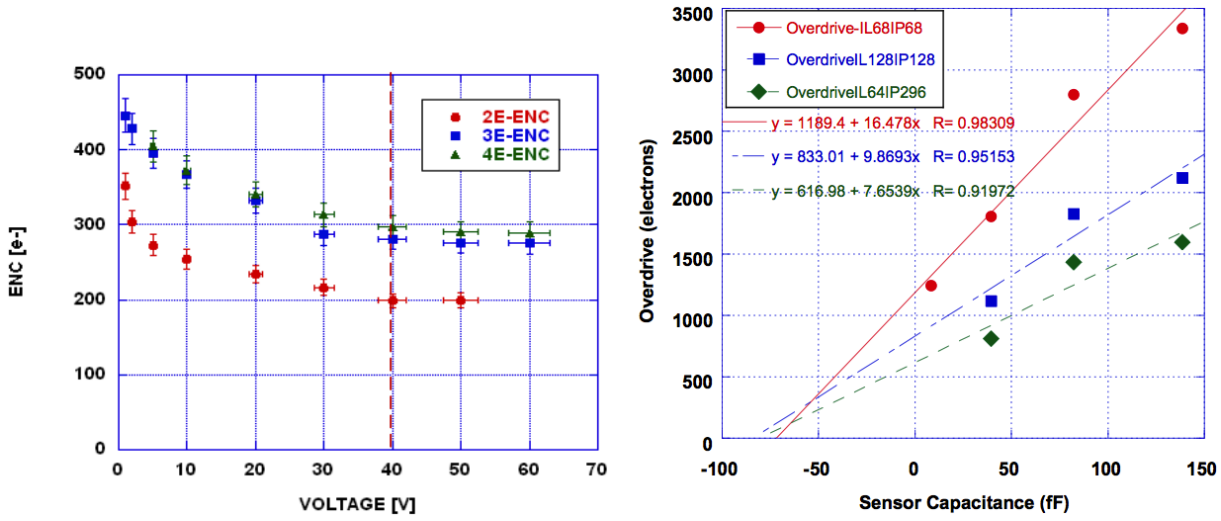


Figure 4.12: Noise measured for the 2E, 3E and 4E structure of 3D sensors respectively (left) and the linear relationship between the capacitance and overdrive (right) [128]. Three lines represent the standard (red, labelled "IP68"), 20% (blue, labelled "IP128") and 40% (green, labelled "IP296") increment of the applied current for the FE-I3 readout chip which shows that the overdrive decreases with the power used.

Chapter 5

Tracking Simulation

ToT and cluster size distributions are first plots to characterise sensors. Hit and efficiency maps can provide further information after track reconstruction. Simulation aims to reproduce these properties at different biases and fluences. This chapter describes the idea of the tracking simulation, where "tracking" refers to modelling of the carrier trajectories within the device. These electron-hole pairs are generated by the charged particles as they traverse the device. They are usually travelling then collected within a cell for 3D sensors. Three processes can be modelled classically or using the Monte Carlo method. They are described in the sections labelled with a " * " symbol.

The tracking simulation uses the classical Lorentz force equation which is a function the electric field distribution in a device and the external magnetic field. Electron-hole pairs migrate along the electric field lines and are slightly deflected by the magnetic field for 3D sensors. The induced current and collected charge in the amplifier are calculated by Ramo's theorem. The efficiency map is used by a Geant4 simulator which considers the whole sensor and tracking system. This Geant4 simulation is described in detail in another thesis [26].

5.1 Signal Formation

Signal formation is simulated by tracking electron-hole pairs separately and summing the induced currents. Trapping due to radiation damage and impact ionisation due to the high electric field are considered. Charge sharing is automatically included by applying Ramo's theorem. The Ramo field is usually small in the regions of a device, e.g. pixel edges, where carriers can diffuse to the neighbouring pixels. The carriers collected in the neighbouring pixels give a small and negative contribution to the target pixel. This is compensated by the positive contribution from carriers collected correctly. The carriers generated in the neighbouring pixels can also be collected in the target pixel. This gives a small tracking efficiency and is considered as charge sharing.

5.1.1 Tracking Carriers in Devices

Electrons and holes in the bulk are tracked using the classic Lorentz force

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) = m^*_{n/p} a \tag{5.1}$$

where $m^*_{n/p}$ are the effective masses for electrons and holes. The electric field E provides energy (drift velocity v_d) to carriers via the drift process given as

$$\vec{v}_{d,n/p} = \mu_{n/p} \vec{E} \tag{5.2}$$

which are proportional to their mobilities.

Carrier positions are shifted firstly by the drift process, $\triangle\vec{x} = \mu\vec{E}\triangle t$, and then deflected by the magnetic at each time step. This is a continuous process and can only be modelled discretely and thus a small time step is needed. The magnetic field B does not accelerate carriers directly but alters their direction. The velocity in Equation 5.1 is obtained by the time differential of two consecutive positions, $\vec{v} = (\vec{x_2} - \vec{x_1})/dt$, and then the cross product is applied to obtain the acceleration given as

$$\vec{v} \times \vec{B} = \begin{bmatrix} i & j & k \\ v_x & v_y & v_z \\ B_x & B_y & B_z \end{bmatrix} = \frac{m^*_{n/p}}{q} \vec{a} \ .$$

The deviation from the calculated acceleration $\vec{a}$ is added to the carrier trajectories by $\triangle\vec{x} = \vec{a}\triangle t^2/2$, which assumes a straight path in each segment.

5.1.2 Ramo's Theorem

Ramo's theorem is a way to calculate the induced current at the electrodes versus time due to the moving carriers. According to Ramo's theorem, the induced current is given as

$$i_c = q\vec{v} \cdot \vec{E_w} = q\mu_{n/p}\vec{E} \cdot \vec{E_w} \tag{5.3}$$

where E_w is the weighting potential. Note that carriers induce signal as soon as they move in the field.

5.1.3 Hecht's Equation*

Defects are formed after a heavy irradiation which can trap both electrons and holes. Carriers are trapped and re-emitted, but the emission time is larger than the collection time or amplifier shaping time. Trapped carriers stop contributing to the induced signal and cause signal loss. For an initial charge q_0 generated, the charge collected at

electrode is given as

$$q(t) = q_0 e^{-t_c/\tau} \tag{5.4}$$

where t_c and τ are the collection time and trapping lifetime of electrons or holes. This is the Hecht equation which can be derived from a basic differential equation with effective trapping time constant as

$$\frac{d(n/p)}{dt} = -c_{n/p}(n/p) \quad \longrightarrow \quad (n/p) = e^{-c_{n/p}t} \tag{5.5}$$

since the capture (trapping) process is proportional to the capture cross section $c_{n/p}$ and carrier density n/p.

5.1.4 Induced Charge

By implementing the Hecht equation in the Ramo theorem, the total induced current is given as

$$i_c(t) = q(t)\vec{v} \cdot \vec{E_w} = q_0(e^{-t_n/\tau_n}\mu_n \vec{E} \cdot \vec{E_w} + e^{-t_p/\tau_p}\mu_p \vec{E} \cdot \vec{E_w}) \ . \tag{5.6}$$

which considers trapping to both electrons and holes. The induced current is integrated by summing over all small segments with time to give the signal response. Trapping reduces the carrier densities travelling to the electrodes and causes a significant loss in the induced current for heavily irradiated sensors.

For instance, a simple pad structure has a Ramo field which is $1/w$ for a thickness w. The field and mobility (doping and field dependent) are assumed to be uniform for an abrupt junction. The induced current and total induced signal are calculated as

$$\begin{aligned} i_c(t) &= q_0 e^{-t/\tau}\mu \frac{V}{w}\frac{1}{w} \\ q_c &= q_0 \int_0^{t_c} e^{-t/\tau}\mu \frac{v}{w}\frac{1}{w} dt = q_0 \frac{\tau\mu V}{w^2}(1 - e^{-\frac{w^2}{\tau\mu V}}) \end{aligned} \tag{5.7}$$

where $t_c = w/v = w/(\mu V/w) = w^2/\mu V$ is the collection time. Electron-hole pairs are assumed to be generated at the bias electrode in this case and only electrons contribute to the signal. However, there is no analytical formula for real structures, such as 3D sensors with a complicated Ramo potential.

5.1.5 Avalanche Multiplication*

The avalanche gain across a high field region can be derived analytically [46]. At steady state, an electron current $i_e(0)$ is injected into a depletion region of width w at $x = 0$. This region has a strong enough field to produce impact ionisation. By ignoring

thermal- or photo-generation of carriers, the generation rate is given as

$$\frac{di_e(x)}{dx} = \alpha_n(x) i_e(x) + \alpha_p(x) i_h(x) \tag{5.8}$$

for the electron and hole current. The total current from carriers is constant across the device, $i_e(x) + i_h(x) = I$. There is no hole injection at $x = w$ and thus $i_e(w) = I$. The generation rate is rewritten as

$$\frac{di_e(x)}{dx} - (\alpha_n - \alpha_p) i_e(x) = \alpha_p I$$

whose solution is given as

$$i_e(x) = \frac{i_e(0) + \int_0^x \alpha_p I e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx}{e^{-\int_0^x (\alpha_n - \alpha_p) dx'}} \ .$$

The multiplication factor M_e of the injected current is given as

$$M_e = \frac{I}{i_e(0)} = \frac{i_e(w)}{i_e(0)} = \frac{i_e(0) + i_e(w) \int_0^w \alpha_p e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx}{i_e(0) e^{-\int_0^w (\alpha_n - \alpha_p) dx}}$$

which is rearranged as

$$M_e = \frac{1 + M_e \int_0^w \alpha_p e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx}{e^{-\int_0^w (\alpha_n - \alpha_p) dx'} dx}$$

$$\frac{1}{M_e} = e^{-\int_0^w (\alpha_n - \alpha_p) dx} - \int_0^w \alpha_p e^{-\int_0^w (\alpha_n - \alpha_p) dx'} \ .$$

By using the fact that

$$e^{-\int_0^w (\alpha_n - \alpha_p) dx} = 1 - \int_0^w (\alpha_n - \alpha_p) e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx \ ,$$

the multiplication factor is calculated as

$$M_e = \frac{1}{1 - \int_0^w \alpha_n e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx} \tag{5.9}$$

which is also called the impact ionisation integral. The breakdown condition is given as

$$\int_0^w \alpha_n e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx = 1 \tag{5.10}$$

when the de-numerator vanishes. The avalanche gain is also integrated by summing all small segments of impact ionisation coefficients $\alpha_{e/h}$ in the two integrations as in the formula.

5.1.6 Thermal Diffusion*

Carriers diffuse at temperatures above absolute zero with thermal energy $k_BT/2$ for each degree of freedom. Electron-hole pairs generated will experience numerous collisions as they move randomly in the crystal which is similar to Brownian motion. Carriers move to the collection electrodes gradually due to the electric field. Thermal diffusion of generated carriers, multiple scattering of incident particles and delta electrons cause charge sharing. The latter two contributions are simulated by the Geant4 simulation developed in collaboration with Borri at Manchester.

Thermal diffusion is considered in the tracking code by giving a random kick which is arbitrary in azimuthal and polar angle at each time step. The time step has to be around the same order as the mean free time τ, which represents the average time between two collisions. The diffusion magnitude of each kick is a fixed amount as

$$l_{n/p} = \tau_{n/p} v_{th,n/p} \tag{5.11}$$

which is the mean free path for electrons and holes. The thermal velocity is derived by relating the kinetic energy to $m^* v_{th}^2/2 = k_BT/2$ in 1D and $3k_BT/2$ in 3D. The thermal velocity for electrons is around $2 \times 10^7\ cms^{-1}$ in 3D [115] and usually the same values applies for holes.

The time step τ can be estimated as

$$\begin{aligned} l_{n/p} v_{th,n/p} &= \frac{m^*_{n/p} v^2_{th,n/p}}{q} \frac{q\tau_{n/p}}{m^*_{n/p}} = \frac{k_BT}{q} \mu_{n/p} \equiv D_{n/p} \\ \tau_{n/p} &= \frac{D_{n/p}}{v^2_{th,n/p}} \end{aligned} \tag{5.12}$$

which is around 10^{-12} to 10^{-13} s for both carriers.

5.2 Input Tables and Parameters

FlexPDE (see below) and Kurata are used to obtain the electric field distribution for a given geometry, but they have different limitations which are noted . The electric field distribution is exported as a discrete table. Interim positions are obtained by interpolation which is discussed in Appendix II. The trapping lifetime is another input and a key factor that affects the signal efficiency.

5.2.1 Numerical Model for a *pn* Junction Diode

For 2D and 3D structures, the Poisson equation can be solved by FlexPDE which is a commercial software package for solving (coupled) partial differential equations using

the finite element method (FEM) [14]. The mesh is created automatically in a form of irregular triangles according to a defined geometry. More complex geometries lead to smaller triangles which is equivalent to a finer grid.

For a *pn*-junction (n/p at the left/right of the origin) under electrostatic condition, the electron density in n-type region is defined by the Fermi level as [67]

$$n(x < -x_n) = N_C e^{[E_F - E_C(x<-x_n)]/k_B T} \tag{5.13}$$

where E_C is the band edge energy given as

$$E_C(x) = E_C(x < -x_n) - q(\phi(x) - \phi(x < -x_n)) \ . \tag{5.14}$$

The electron density can be approximated as

$$n(x) = N_C e^{[E_F - E_C(x<-x_n) + q(\phi(x) - \phi(x<-x_n))]/k_B T} = n(x < -x_n) e^{q(\phi(x) - \phi(x<-x_n))/k_B T} \tag{5.15}$$

Similarly the density of holes is given as

$$p(x) = p(x > x_p) e^{q(\phi(x>x_p) - \phi(x))/k_B T} \ . \tag{5.16}$$

3D sensors are a *p+-p-n+* diode, but the same approximation for a *pn* junction can also be applied. At full depletion, carrier densities fall quickly away from the electrodes. The modified Poisson equation is given as

$$\nabla^2 \phi = -\frac{q}{\epsilon_s}[N_D - N_A - N_{D(n+)} e^{q(\phi(x) - \phi(n+))/k_B T} + N_{A(p+)} e^{q(\phi(p+) - \phi(x))/k_B T}] \ . \tag{5.17}$$

Note that this equation is only valid for unirradiated sensors.

5.2.2 Electric Field Map from Kurata

The algorithm of Kurata using the grid method was described in Chapter 4. The original differential equations for carrier transport are transformed into integral forms using the Scharfetter-Gummel discretisation. The Runge-Kutta method, a multi-variable version of the Euler method, is used to increase numerical stability. Simulation starts from an initial solution and iterates to arrive at convergence. However, Kurata can only cope with 2D structures and the grid method is less suitable for complicated shapes. Kurata is used for both unirradiated and irradiated devices.

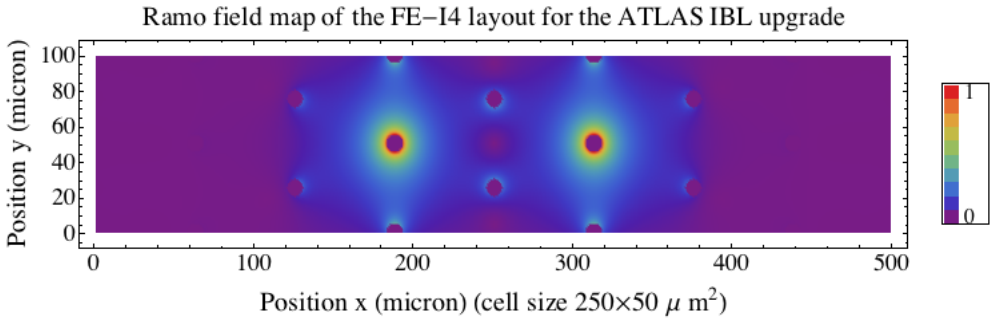


Figure 5.1: A 2D plot for the Ramo potential of a 3D sensor with the FE-I4 layout generated by FlexPDE.

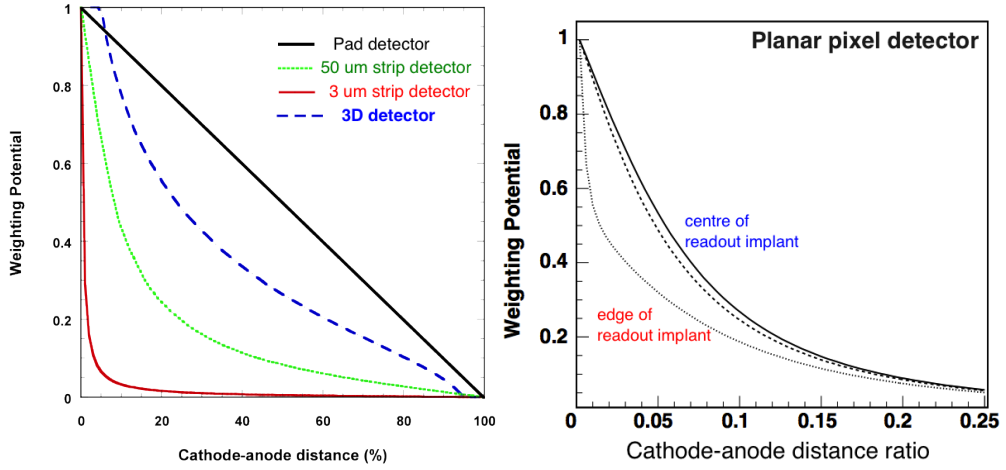


Figure 5.2: The Ramo potential on the inter-electrode line for pad, strip and 3D structures (left) [128] and planar detectors (right) [58].

5.2.3 Ramo Field Map from FlexPDE

The Ramo weighting field is obtained by solving the Laplace equation using FlexPDE which can define complicated geometry and provide accurate solution in 2D and 3D. The Dirichlet boundary conditions are set to 1 for the collection electrodes which are the only two $n+$ electrodes in the targeted pixel[1]; and 0 for all other electrodes. Figure 5.1 shows a 2D plot fof the Ramo potential oc a 3D sensor with the FE-I4 layout.

Figure 5.2 shows the Ramo potential on the inter-electrode line for pad, strip, 3D (left) and planar (right) detectors. Their gradients are the Ramo weighting fields which is flat for pad detectors. For planar strip and pixel detectors, the Ramo field has a very large peak at the readout electrode. 3D sensors also have more symmetric peak fields at both electrodes.

[1]This is a 2E structure that each sensor pixel has two 3D cells which are connected to the same amplifier in a readout pixel.

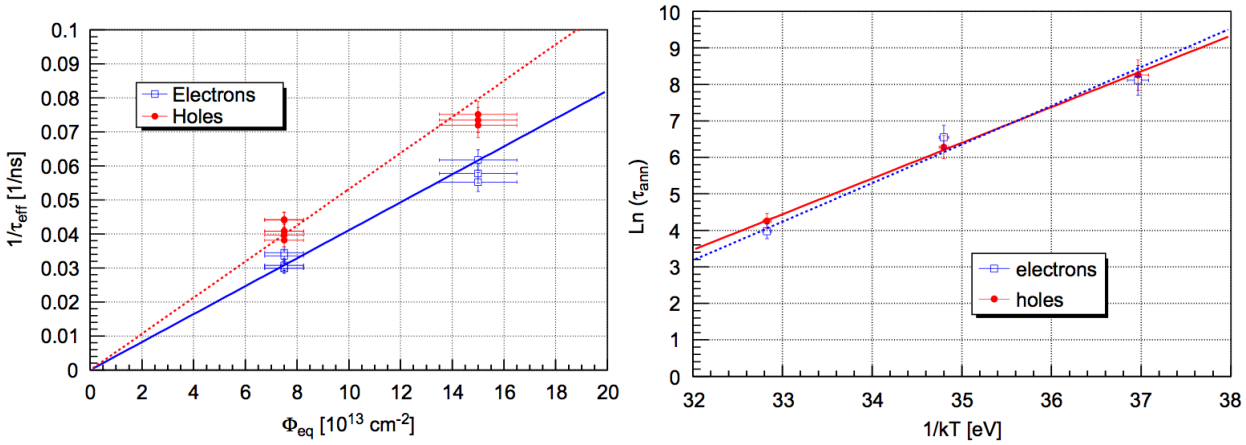


Figure 5.3: The trapping lifetime is a function of the fluence and the activation energy of annealing can be obtained from the Arrhenius plot [65].

5.2.4 Experimental Trapping Lifetimes

Similarly to the generation lifetime, the effective trapping times are expressed as a function of the fluence Φ_{eq} as

$$\frac{1}{\tau_{t,e/h}} = \frac{1}{\tau_{0,t,e/h}} + \beta_{e/h}\Phi_{eq} \approx \beta_{e/h}\Phi_{eq} \tag{5.18}$$

where $\beta_{\tau,t,e/h}$ is the damage constant for the effective trapping lifetimes. There are many experimental values for $\beta_{e/h}$ which depend on the material properties, but they have similar order of magnitudes. Values from Kramberger are used which give $\beta_e = (3.2 \sim 3.7) \times 10^{-16}$ and $\beta_h = (3.5 \sim 5.7) \times 10^{-16}$ $cm^2 ns^{-1}$ [32][79][65][28] for electrons and holes respectively. For instance, the trapping lifetimes for electrons and holes are $(2.7 \sim 3.1)$ and $(1.8 \sim 2.9)$ ns at a fluence of 1×10^{15} $n_{eq}cm^{-2}$.

However, these experiments have fluences less than 2×10^{14} $n_{eq}cm^{-2}$ whilst the experimental conditions in later chapters are up to a few 10^{16} $n_{eq}cm^{-2}$. The trapping lifetimes are still assumed to be inversely linear with fluence and further experiments are needed.

5.3 Simulation of Physical Processes

Analytical equations (macroscopic) are often used to estimate trapping and impact ionisation. They can also be modelled by stochastic processes using a full Monte Carlo simulation (microscopic) since only one carrier is simulated at a time. Thermal diffusion has to be modelled by the Monte Carlo method since an analytical formula is more suitable for a cloud of carriers as shown in this section.

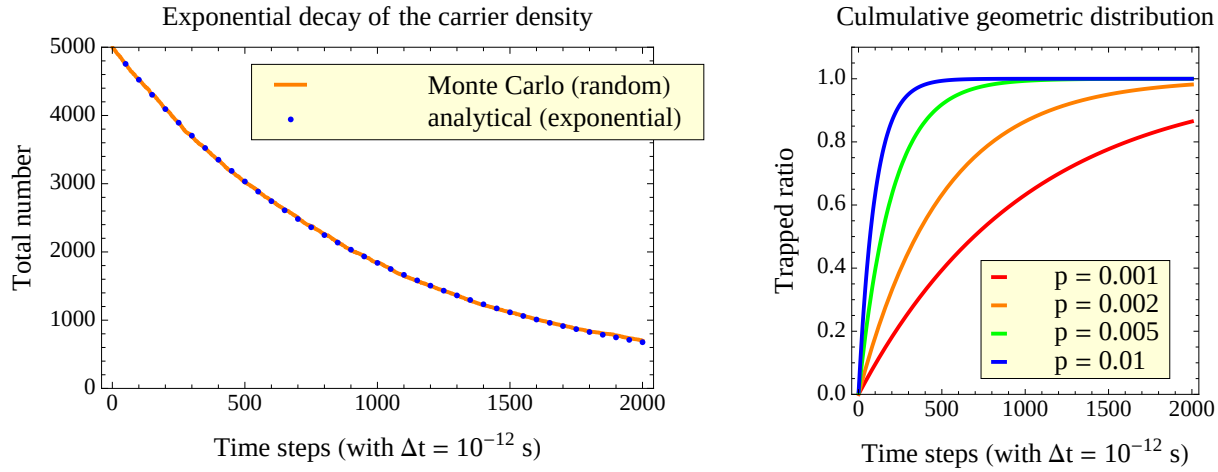


Figure 5.4: Trapping of carriers simulated by random events compared with the analytical formula in Equation 5.4 (left). The cumulative geometrical distribution also shows the effective lifetime decreases as the trapping probability increases.

5.3.1 Trapping

The trapping process is characterised by different lifetimes for electrons and holes. Trapped carriers are usually re-emitted after a long time compared to the average collection time. This process can be modelled by a geometric distribution which describes a successful Bernoulli trial after $k-1$ failed trials. Its probability and accumulative distribution are given as

$$\begin{aligned} P_{geo}(k) &= (1-p)^{k-1}p \\ C_{geo}(k) &= 1-(1-p)^{k} \end{aligned} \tag{5.19}$$

where p is the probability of getting a successful trial.

Carriers have a probability p to get trapped which can be obtained from their average lifetime τ. If the time interval is set to $\triangle t$ in simulation, the average trapping probability is $p \simeq \triangle t/\tau$ as $\tau/\triangle t$ is large. Figure 5.4 (left) shows a comparison between an analytical formula and a Monte Carlo simulation for 5000 initial carriers. τ and $\triangle t$ are set to 10^{-9} and 10^{-12} s respectively which subjects to a trapping probability $p = 0.001$. One thousand time steps, $k = 1000$, correspond to the half-life time, $\triangle T = k\triangle t = \tau = 1$ ns, and thus carriers experience a $1/e$ loss. Figure 5.4 (right) shows the the cumulative geometrical distribution with different trapping probabilities. Carriers get fully trapped with a trapping ratio of 1. The effective lifetime decreases as the trapping probability increases and this controls the effective drift length for carriers.

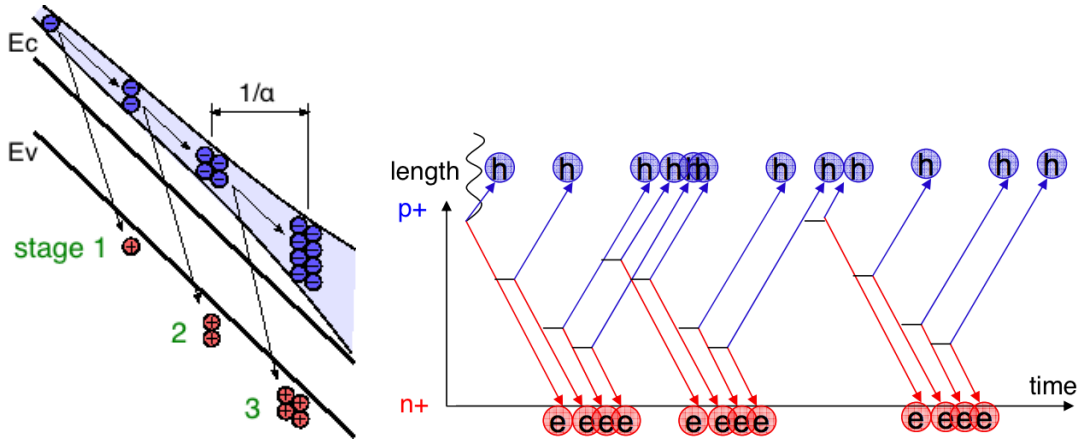


Figure 5.5: A schematic of discrete (left) [122] and continuous (right) [75] impact ionisation processes which correspond to Table 5.1 (c) and (e) respectively.

length (μm)	avalanche integral (a) $\alpha_p = 0$	avalanche integral (b) $\alpha_p \neq 0$	discrete Poisson (c) $\alpha_p = 0$	continuous Poisson (d) $\alpha_p = 0$	continuous Poisson (e) $\alpha_p \neq 0$	exponential approximation (f) $\alpha_p = 0$
1	2.72	2.79	2.00±1.00	2.70±2.13	2.76±2.28	2.72
2	7.39	8.77	3.97±2.44	7.31±6.77	$\mu = 8.68$, $\sigma = 9.15$	7.39
3	20.1	47.9	8.01±5.18	19.7±19.2	$\mu = 43.7$, $\sigma = 62.3$	20.1

Table 5.1: Multiplication factors for different simulations. The impact coefficient α_n is set to 1 μm^{-1} and the corresponding α_p is calculated for (b) and (e). (d) and (e) have 10^5 events.

5.3.2 Impact Ionisation

Carriers are accelerated in the large electric field and obtain enough energy to further generate electron-hole pairs. Once the kinetic energy is above the band gap, the probability to excite one or more electron-hole pairs becomes significant. This phenomenon can be modelled by a Poisson process. The Poisson distribution describes the probability for k events to occur in a fixed time interval with an chraterstic average rate. Its probability is given as

$$P(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k} \tag{5.20}$$

where λ is the average rate per time interval. The average occurrence rate is analogous to the ionization rates which gives an average charge generation over a unit interval. For a higher rate or a longer time interval, one tends to get more events. The Poisson distribution approximates to a Gaussian distribution as $\lambda > 10$.

Discrete Poisson Process

For the discrete case, $\alpha_n = 1\ \mu m^{-1}$ ($\lambda = 1$) and drift lengths of 1, 2 and 3 μm as shown in Figure 5.5 (left), one can calculate all possible combinations for multiplication

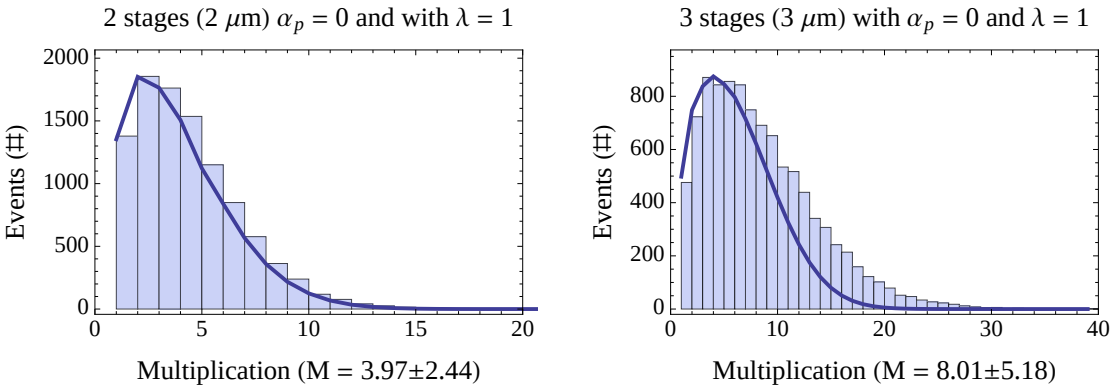


Figure 5.6: Calculation of all combinations (blue curve) and Monte Carlo simulation (histogram) for multiplication for discrete two or three stage Poisson processes.

distributions. For example, a total multiplication of 3 at the third stage (μm) can be permutations of (0,1,1) or (0,0,2), which represents multiplications at each stage[2]. Figure 5.6 shows a comparison between the probability combinations (blue curves) and full Monte Carlo simulations (histograms) for the 2 and 3 μm case. It is easy lose some combinations or need to drop less probable ones which underestimates the multiplication factor. The Monte Carlo mean values of multiplication are shown in Table 5.1 (c) as expected, but they are different from the analytical values.

Continuous Poisson Process

The reason is that this process should be continuous to be realistic. Impact ionisation occurs continuously and randomly as shown in Figure 5.5 (right) with a smaller probability at each step. Avalanche generation from holes is much smaller and it can be included or not as following.

- Impact ionisation from electrons and holes ($\alpha_n \neq 0$ and $\alpha_p \neq 0$)

  Figure 5.7 shows the Monte Carlo simulations for a length of 2 and 3 μm. A continuous Poisson process is equivalent to a many-stage discrete Poission process. It has a mean λ controlled by the length step $\triangle l$ as

  $$\lambda_{n/p} = \alpha_{n/p} \triangle l \tag{5.21}$$

  where $\triangle l$ is set to 0.01 (instead of 1) μm which gives $\lambda = 0.01$ and a hundred times more stages.

[2]Note that probability $f(0,1,1) = P(0,1)_1 * P(1,1)_2 * [C_1^2 * P(1,1) * P(1,0)]_3$ is different from $f(1,0,1) = P(1,0)_1 * [P(1,0) * P(1,0)]_2 * [C_1^2 * P(1,1) * P(1,0)]_3$.

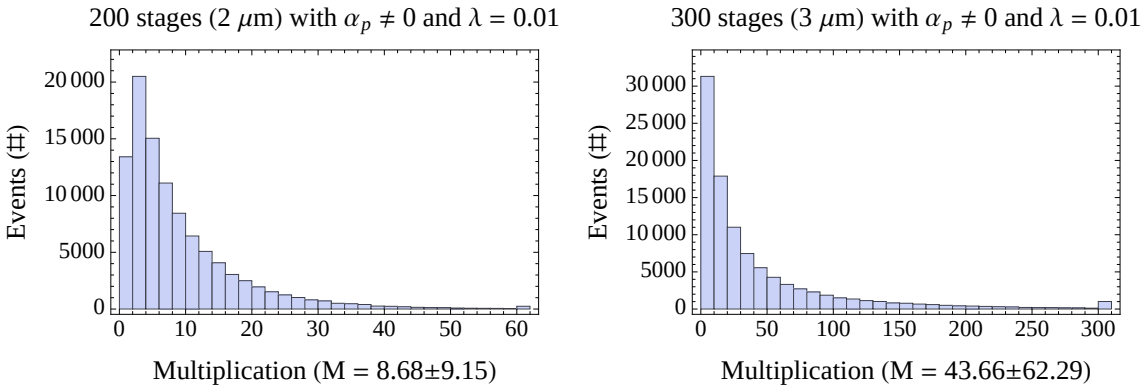


Figure 5.7: The Monte Carlo simulation results for multiplication from electrons and holes of continuous 2 or 3 μm Poisson processes. Long tails show larger dispersion and deviations.

Holes are also created along with the pairs from impact ionisation ($\alpha_p \neq 0$). They travel backwards and create some further electron-hole pairs in which electrons travel forwards and create further electron-hole pairs too. This is repeated seven times in the simulation and the contribution gets smaller each time. Values from the Monte Carlo simulation are listed in Table 5.1 (e) which are very similar to the theoretical values in Table 5.1 (b).

- Impact ionisation from electrons only ($\alpha_n \neq 0$ and $\alpha_p = 0$)

  Figure 5.8 also shows the Monte Carlo simulations of 2 and 3 μm lengths for a continuous Poisson process. It has the same parameters as the previous case but impact ionisation from holes is neglected. Values from the Monte Carlo simulation are listed in Table 5.1 (d) which are also very similar to the theoretical values in Table 5.1 (a). This validates the inference of Poisson processes for impact ionisation. They are larger than the discrete case in Table 5.1 (c) which validates the assumption of a continuous process.

  However, standard deviations shown after the " ± " symbol for impact ionisation processes are very large and may exceed sample means due to large dispersion of simulation results. This implies that the use of avalanching for signal gain may not be a good idea since it results in higher noise. A design which only uses avalanche generation from electrons is better by comparing Figures 5.7 and 5.8, in which the avalanche gain is close to its standard deviation.

Exponential Approximation

It is useful to calculate the avalanche gain by the Monte Carlo simulations since carrier trajectories are not straight lines but have random components due to thermal diffusion.

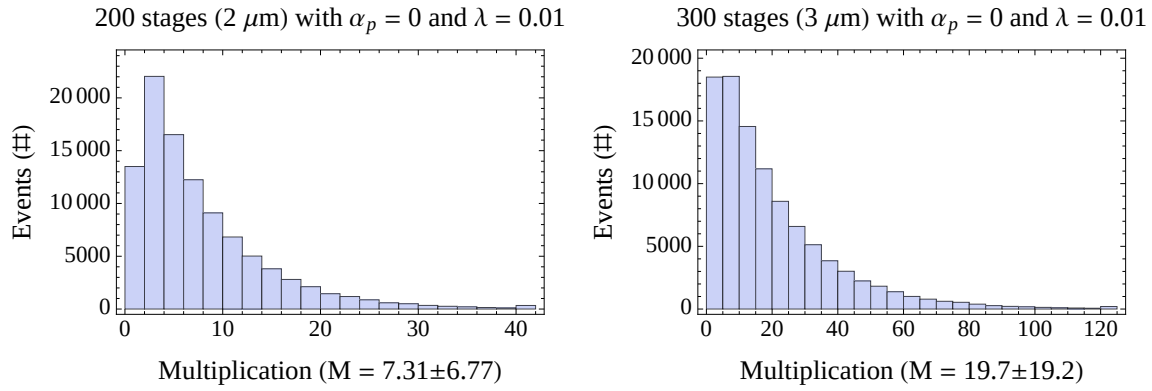


Figure 5.8: The Monte Carlo simulation results for multiplication from electrons only of continuous 2 or 3 μm Poisson processes. They have smaller multiplications than the previous case with $\alpha_p \neq 0$ in Figure 5.7 and also less dispersion (noisy).

However, it requires a large number of events for averaging and thus a longer computing time. It is also very difficult to include avalanche generation from holes in the tracking code. With absence of hole multiplication, the avalanche gain can also be calculated as

$$M = (1 + \alpha_n \triangle l)^{\frac{l}{\triangle l}} \tag{5.22}$$

with values listed in Table 5.1 (e). They have exactly the same values with the previous calculations using the impact ionisation integral and Monte Carlo method as $\alpha_p = 0$ in Table 5.1 (a) and (d). Both the exponential approximation and impact ionisation integral are implemented in the tracking code. $\alpha_{n/p}$ are field-dependent and thus a small time step is needed. Due to the small peak field in depleted sensors values, α_p is small ($\alpha_p \approx 0.035$ as $\alpha_n = 1$) and these two models give similar multiplication factors.

5.3.3 Thermal Diffusion

Thermal diffusion of individual carriers can be simulated microscopically by the thermal velocities $v_{th,n/p}$ and collision times $\tau_{n/p}$ as discussed previously. For a large group of carriers, the diffusion radius, which is the standard deviation σ obtained by fitting the carrier cloud using a Gaussian distribution, can be estimated macroscopically using the mean squared displacement and Einstein relation for charged particles as

$$\sigma = \sqrt{2D_{n/p}\triangle T} = \sqrt{2\frac{k_B T}{q}\mu_{n/p}\triangle T} \tag{5.23}$$

in 1D where $\triangle T = \triangle t \times n$ (steps).

Figure 5.9 shows that the diffusion distribution obeys the Gaussian distribution

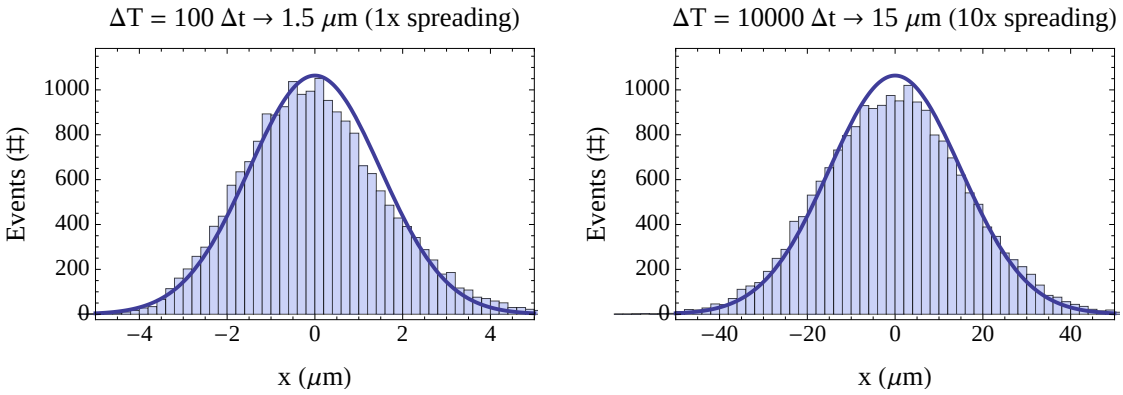


Figure 5.9: The diffusion radius of a carrier cloud can be predicted by the mean squared displacement. The size of the carrier cloud is the square root of the total time $\triangle T$. Histograms from the Monte Carlo simulations are fitted with a Gaussian distribution (curve) for the cloud radius and a longer total times corresponds to a larger cloud radius.

whose radius σ is proportional to $\sqrt{\triangle T}$. The average electron collision time τ_n is 2×10^{-13} s from Equation 5.12. Figure 5.10 shows that using this value in the Monte Carlo simulation agrees with the estimation that $\sigma_{1D} = 0.5$ μm from Equation 5.23. Note that simulations were performed in 3D and then projected to 1D.

To summarise, trapping and impact ionisation were implemented in the analytical way since it is faster than the Monte Carlo method. Thermal diffusion was included using random processes since individual carriers are tracked, not the carrier cloud (radius). A full Monte Carlo tracking simulator is achievable but the settings above are preferred.

5.4 Results and Discussion

Carrier tracks and the device structure are visualised in three-dimension. Simulation is compared with the Lorentz angle data in this section. Different time steps are simulated using Geant4 to show its effect on the cluster size and charge sharing.

5.4.1 Visualisation

The tracking code was written for three-dimensional structures which are defined by the field maps and electrode geometry. Figure 5.11 (a) visualises the trajectories of three electron-hole pairs generated at different positions in a cell. Electrons (red) are attracted to the central column while holes (black) are attracted to different bias electrodes according to their initial positions. Figure 5.11 (b) shows an incident particle (black line) and its track crosses three cells at this angle. The final positions are where electrons (red) and holes (blue) are collected. Figure 5.11 (c) is a horizontal projection

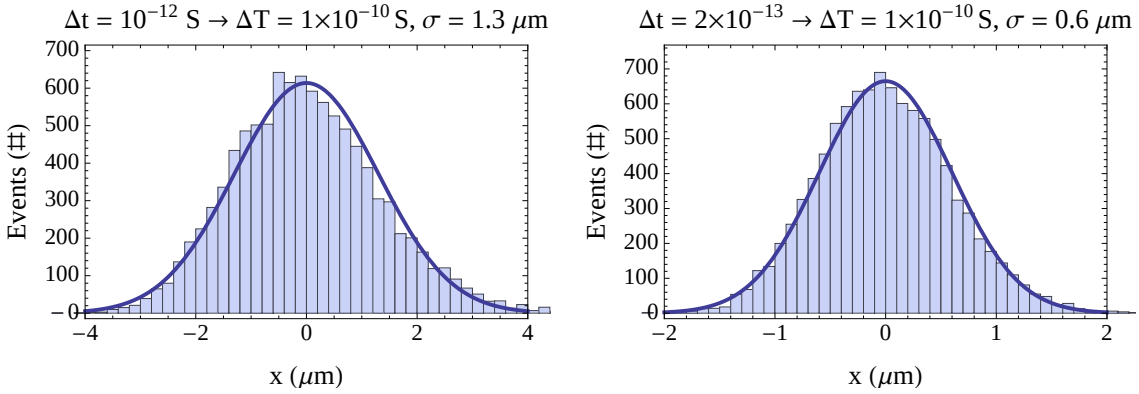


Figure 5.10: The times step can also be estimated by the Monte Carlo simulation (histogram) and fitted by a Gaussian curve. (b) $\triangle t = 0.2\ ps$ is the theoretical value which also gives a correct estimation of the mean squared displacement for the electron cloud. (a) a longer mean free time τ_n will overestimate the cloud radius. These four figures of diffusion have 10^4 events.

across the sensor plane. Four cells are fired in this case and the carrier numbers that they collect can be visualised and estimated.

5.4.2 Lorentz Angle

Depending on the relative sensor positioning to the interaction point, there will be various incident angles with respect to the sensor plane. The cluster size shows the number of fired pixels for one crossing event. There is a strong magnetic field in the tracker for momentum measurement and charge identification. The generated carriers are also deflected as they drift to the electrodes. This bending angle is called the Lorentz angle or drift angle.

The cluster size is minimised when the incident angle equals the Lorentz angle. One can thus determine the Lorentz angle by tilting the sensor at different angles experimentally and extracting the angle minimising the cluster size. It is usually a few degrees for planar sensors since the cluster size may be larger or smaller due to the magnetic distortion. The sensor was rotated along the long pixel direction and the magnetic field is in the same direction. The experimental setup of the telescope will be given in Chapter 7

Figure 5.12 (left) shows the test beam data of the overall charge sharing probability as a function of incident angle with a 1.6 T magnetic field [47]. 3D sensors from FBK and Stanford (labelled "STA") have a negligible Lorentz angle, while planar (labelled "PLA") ones have around eight degrees. Figure 5.12 (right) shows the simulated cluster size at different incident angles with the same trend as the data. It was generated in a similar condition (2 T) but without a threshold (no readout chips). 3D sensors are almost unaffected by the magnetic field up to 2 T as will be discussed later.

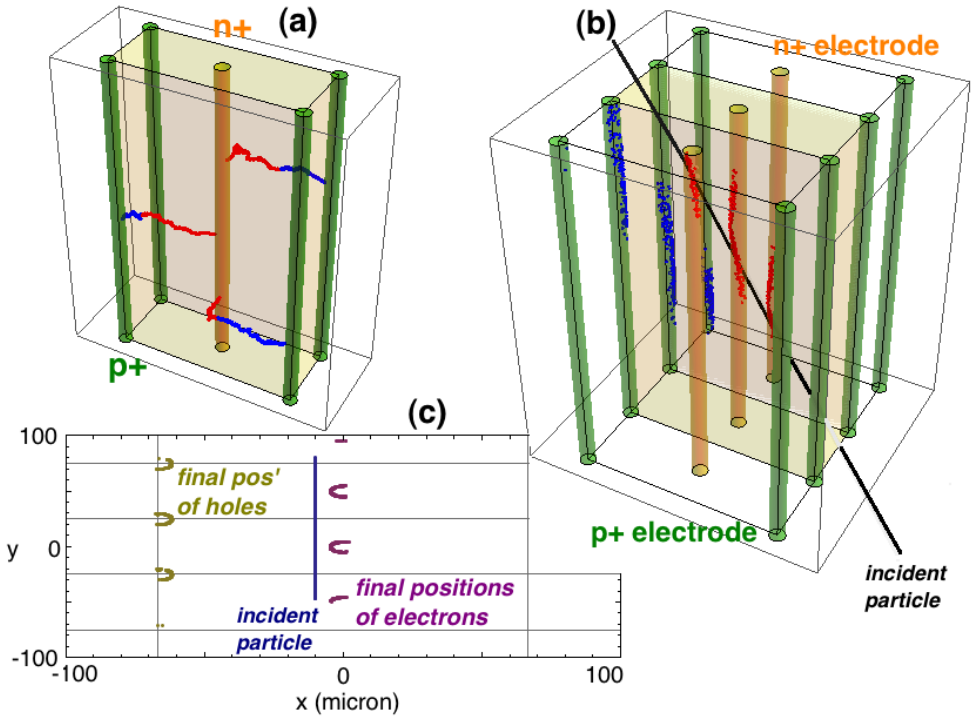


Figure 5.11: Visualisation of tracking for electron-hole pairs and clustering for an incident particle. (a) shows the carrier trajectories for electrons (red) and holes (blue). (b) shows the end points of carriers, namely the positions that they are collected, for electrons (red) and holes (blue). (c) shows a horizontal projection across the sensor plane of (b). This is the FE-I3 layout with a cell size of $133.3 \times 50\ \mu m^2$.

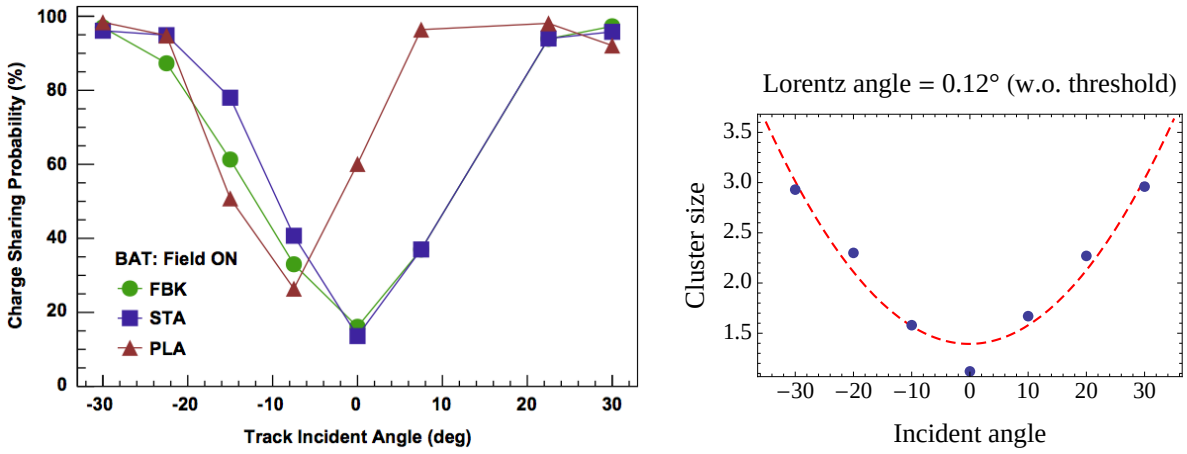


Figure 5.12: Overall charge sharing of different sensor designs with a 1.6 T magnetic field [47] and the simulation of the averaged cluster size with similar conditions. Note that the cluster size in simulation is the number of cells with charge over a set value and the charge is the number electrons collected in each cell.

5.4.3 Time Step

The time step is essential to obtain the correct level of charge sharing which depends only on diffusion in the tracking simulation. It is also important as the magnetic field is included since large steps give incorrect results when using straight line approximation discussed earlier. From the theoretical estimation and simulation above, the correct average collision time is 2×10^{-13} s for electrons and 1.5×10^{-13} s for holes. However, a time step around 1 ps agrees better with the test beam data of charge sharing and thus $\triangle t = 1 \times 10^{-12}$ s is chosen for simulations. The discrepancy is not understood for the moment.

Figure 5.13 shows the cluster size distribution (top), average cluster size (middle) and charge sharing maps (bottom) for different time steps [26]. The time step needs to be larger in simulation to explain the cluster size data. This may be due to the difference in carrier mobilities since the MIP-generated cloud is a dense cluster. A shorter time step also directly leads to a longer computing time for the same amount of events which was also investigated in [26].

5.5 Digitiser of 3D Sensors

The full response of the ATLAS tracker from all sub-sensors is modelled by the Athena Framework. The digitiser is proposed for the pixel detector to understand the signal response and track reconstruction. The planar community simulates the signal response for each particle that crosses the device which consumes more computing time. The 3D collaboration uses a 2D electric field map from Kurata which is best suited to the FBK fully-through design. It is planned to use 2D field maps from TCAD for CNM sensors to include the non-fully-penetrating electrodes. More structures, such as the p-spray, p-stop, oxide and metal layers, can also be considered in TCAD.

5.5.1 Signal Efficiency Map

Signal efficiency maps, before and after irradiation, are used in the Geant4 simulation as an input to define the detector response. Generated particles by Geant4 read signal efficiency values according to their relative positions in the map. Multiple scattering and delta electrons are also included. Geometries and coordinates of all the sensor in the tracker are also specified. Pixels are assumed to be identical in design and response in the Geant4 simulation. Temperature and irradiation level are also assumed to be uniform in the device and tracking simulation.

A four-pixel area is simulated to obtain an efficiency map for a pixel. Hundreds of electron-hole pairs are randomly generated per 1 μm^2 area which is the specified resolution of the map. The signal response which depends on the thickness and averaged

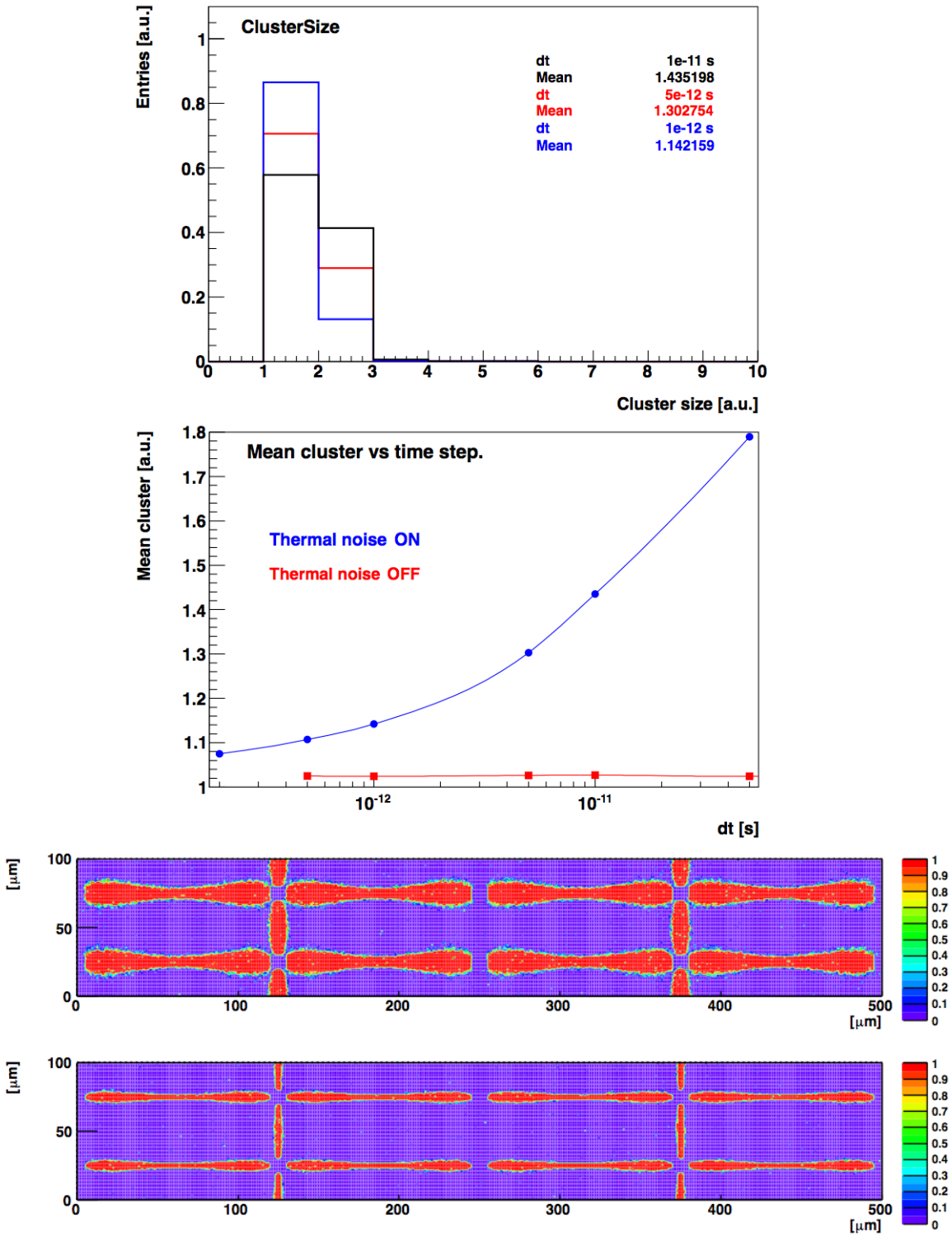


Figure 5.13: Cluster size distribution (top), average cluster size (middle) and charge sharing maps (bottom) for different time steps [26]. Time steps are 5×10^{-12} and 1×10^{-12} s for the charge sharing maps whose smearing decreases with $\triangle t$.

electron-hole pairs is normalised to the signal efficiency at each 1 μm^2 bin. Charge sharing is obtained by the carriers generated in the neighbouring pixel but collected in the central cell, and vice versa. The signal response of FBK sensors is assumed to be independent of depth. One 2D map is accessed every 10 μm in depth. Some maps can be replaced by depth-dependent ones for CNM sensors.

Figure 5.14 shows signal efficiency maps with $\triangle t = 5 \times 10^{-12}$ s for electrons (a), holes (b) and their sum (c). The summed map is used for unirradiated sensors in the 3D digitiser for the IBL. The contribution ratio from electrons and holes depends on the hit position. Charge sharing is mainly from electrons as shown in Figure 5.14 (a) and thus it is controlled by the average electron collision time. A smaller time step, $\triangle t = 5 \times 10^{-13}$ s, reduces smearing at the cell edges as shown in Figure 5.14 (d).

5.5.2 Effect of the Magnetic Field

The effects of the magnetic field can also be visualised in the 2D efficiency maps. Figure 5.15 shows that 3D sensors are not affected by the magnetic field up to 2 T (e.g. ATLAS). The charge sharing regions (cell edges) widen in the y-direction for a 4 T magnetic field (e.g. CMS) in the x-direction, and vice versa. This shows one merit of the 3D technology.

5.5.3 Parametrisation of the Signal Efficiency Map

The signal response can also be parametrised as a function of hit position. The S-curve (sigmoid curve as $1/(1+e^{-x})$) is used to describe the smearing response of the pixel edges and electrode circumferences. Different values of the smearing widths σ are tested to fit the tracking simulation results shown in Figure 5.17 (red).

The collection probability in the bulk is set to 1 (without trapping) for an unirradiated sensor. All the boundaries are multiplied together, four edges (left, right, up and down), two $n+$ electrodes and six $p+$ electrodes, to give a hit map. It is written in an analytical form as

$$\prod_{electrodes}^{2n+,\ 6p+} \frac{1}{1+e^{(\frac{d_{perimeters}}{\sigma_{electrodes}})}} \times \prod_{edges}^{l,\ r,\ u,\ d} \frac{1}{1+e^{(\frac{d_{edges}}{\sigma_{edges}})}} \ . \tag{5.24}$$

Note that d's are the distance to each edges and each component gives one as $d > \sigma$.

Smearing of $\sigma_{left} = \sigma_{right} = 2.5$ μm, $\sigma_{upper} = \sigma_{down} = 3$ μm at pixel edges and $\sigma_{n+} = \sigma_{p+} = 1$ μm at electrode circumferences are selected. The selected σ values (blue) and double the selection (orange) are plotted for comparison in Figure 5.17. Charge sharing is proportional to σ and is larger along the long edges. The parametrisation function roughly agrees with the map method using the tracking simulation.

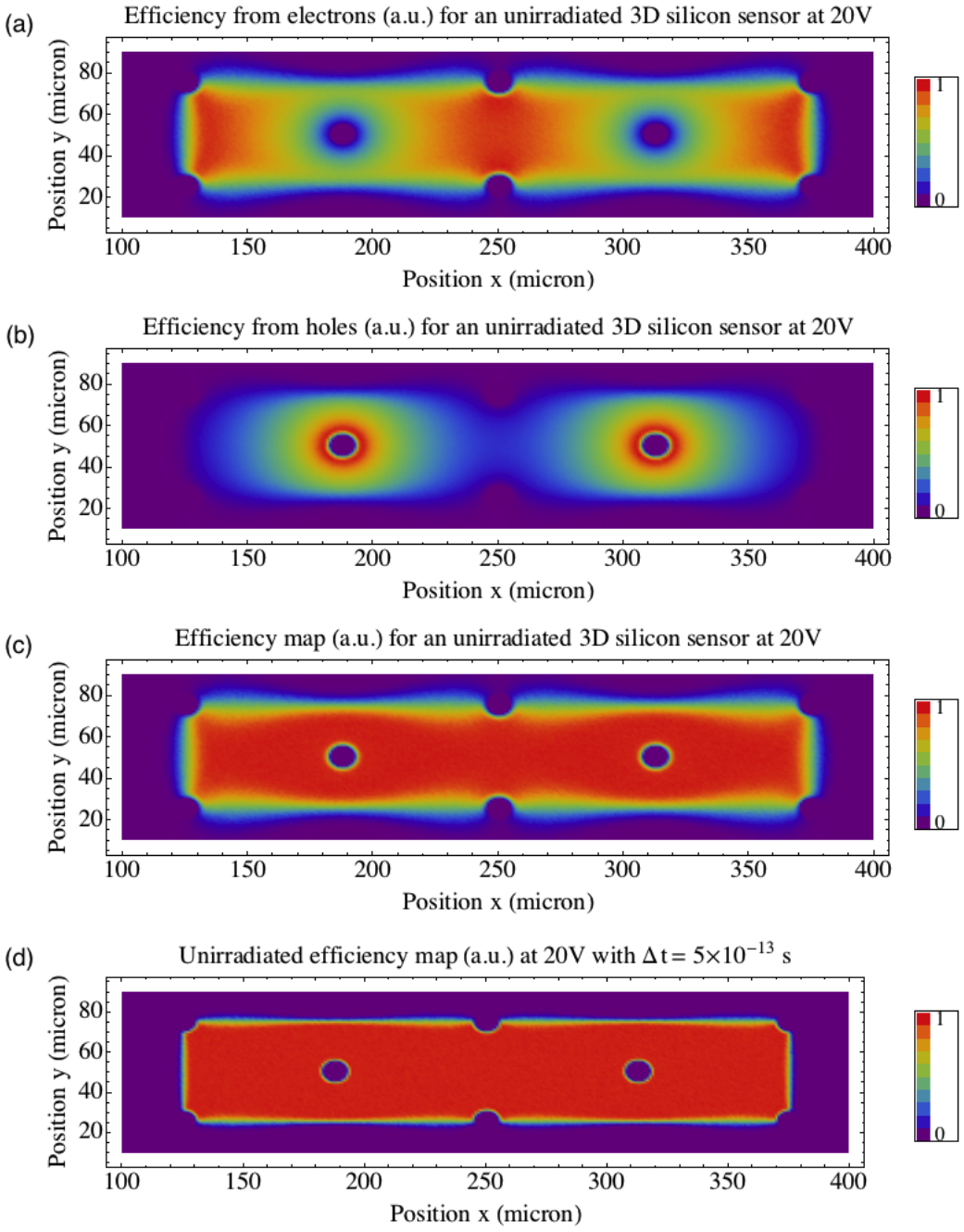


Figure 5.14: Signal efficiency maps from electrons (a), holes (b) and their sum (c). A shorter time step of 5×10^{-13} s gives much smaller charge sharing (d) compared to other three plots with $\triangle t = 5 \times 10^{-12}$ s. A time step between 0.5 to 1 ps agrees better with the cluster size data.

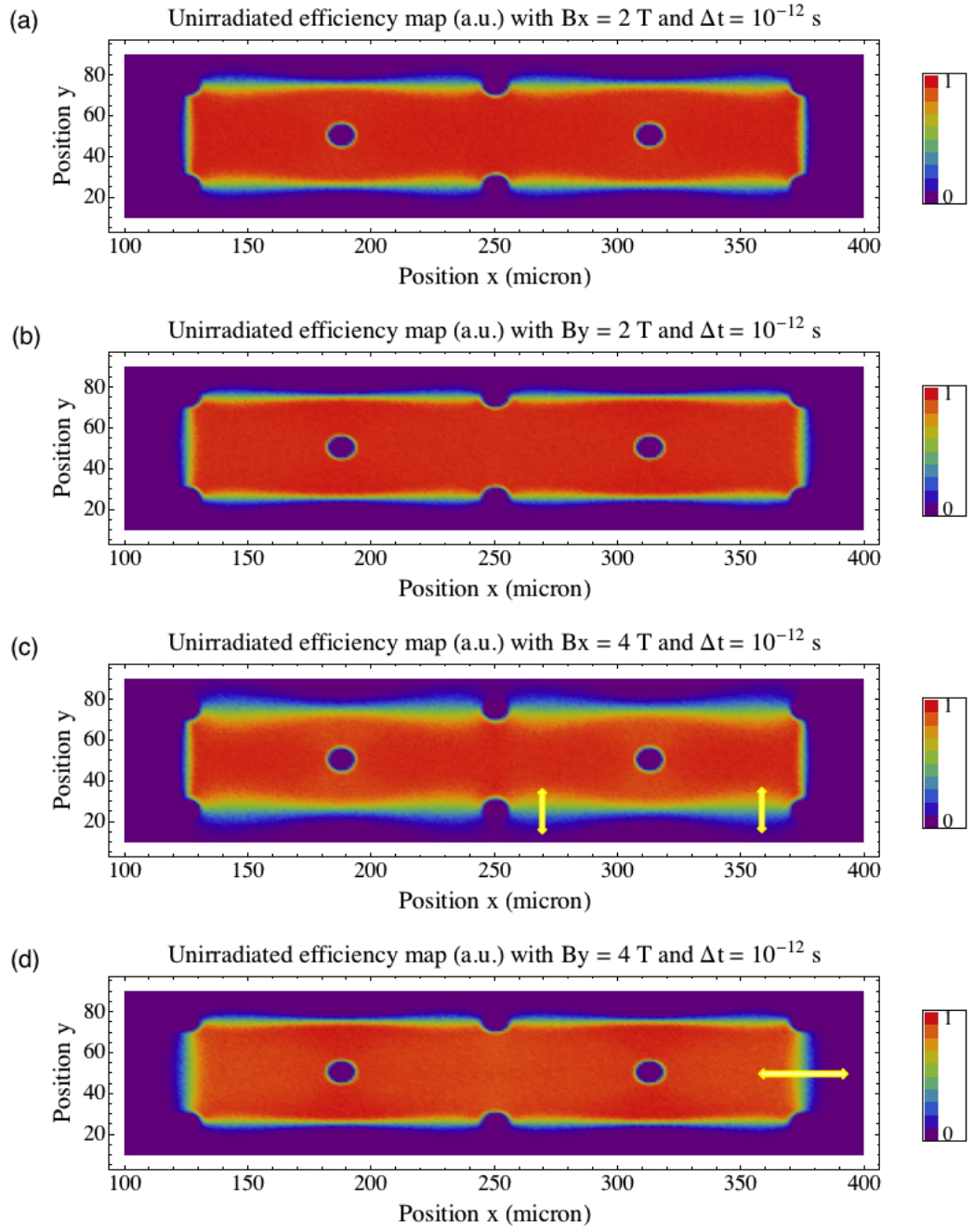


Figure 5.15: 3D sensors are not affected by the magnetic field up to 2 T. The charge sharing regions widen in the y-direction for a 4 T magnetic field in the x-direction, and vice versa. Yellow arrows show the direction of expansion for the 4 T case with respect to the 2 T one.

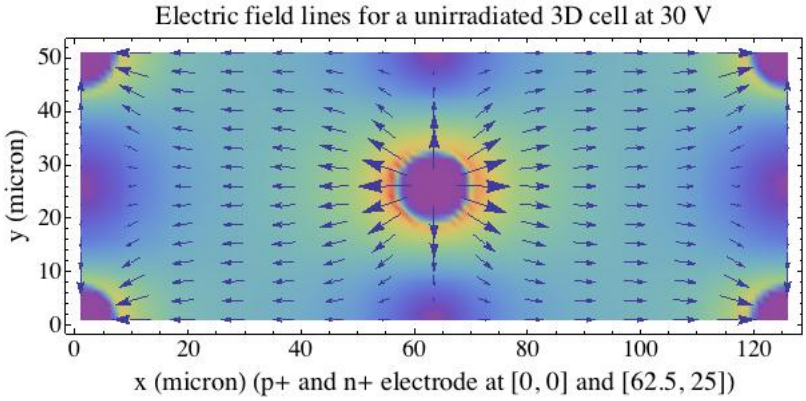


Figure 5.16: Electric field lines of a cell of a 3D FBK sensor.

The bell-shaped bands in Figure 5.14 (c) on the long edges are not reproduced and the widest regions are chosen to be fitted for σ. The electric field lines are parallel to the cell edges due to the narrow pixel shape as shown in Figure 7.19 and thus carriers there are more affected by thermal diffusion.

5.5.4 Parametrisation of the Total Signal Efficiency

Watts has parametrised the signal efficiencies as a function of fluence which can predict the performance of irradiated sensors [35]. For a pad diode, the induced signal and integrated signal are given as

$$\begin{aligned} \frac{dS}{dt} &= \pm q \frac{dV_w}{dx} \frac{dx}{dt} e^{-\frac{x}{\lambda_{p/n}}} \\ S &= \frac{\lambda_{p/n}}{h}(1 - e^{-\frac{x}{\lambda_{p/n}}}) \end{aligned} \tag{5.25}$$

where $\lambda_{p/n} = v_{d,p/n}\tau_{t,p/n}$ and x is the distance to the collection electrode. Note the integral is taken from the end point to the start to evaluate the trapping term correctly. Figure 5.18 shows the signal induced by electrons and holes.

The induced signal saturates in a few tens of microns since $\lambda_{p/n}$ is small at high fluences which also indicates the merit of 3D sensors. The total signal is lower close to the $n+$ region since the signal is mainly due to with the short hole lifetime. However, the Ramo weighting potential and electric field are larger close to the $pn+$ junction for a strip structure due to the confined geometry and effective doping concentration respectively. It is also not suitable for a 3D structure due to the complicated field distributions. The map method is needed for irradiated samples and higher statistics is certainly preferred.

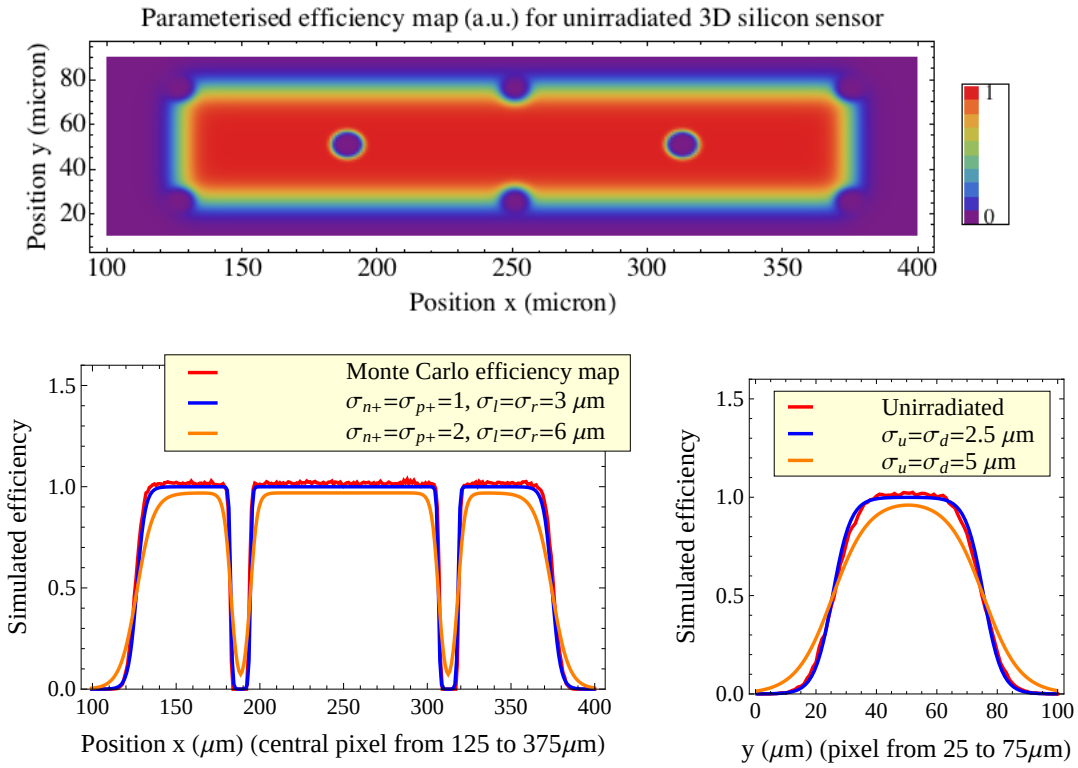


Figure 5.17: Parametric function of central pixels of a 3D FBK sensor and compared with the map method (red). The bell-shaped bands on the long edges are not reproduced. The 1D cuts across the device in the y (left) and x (right) direction show that the blue set of parameters agree with the simulation while the orange ones overestimate smearing.

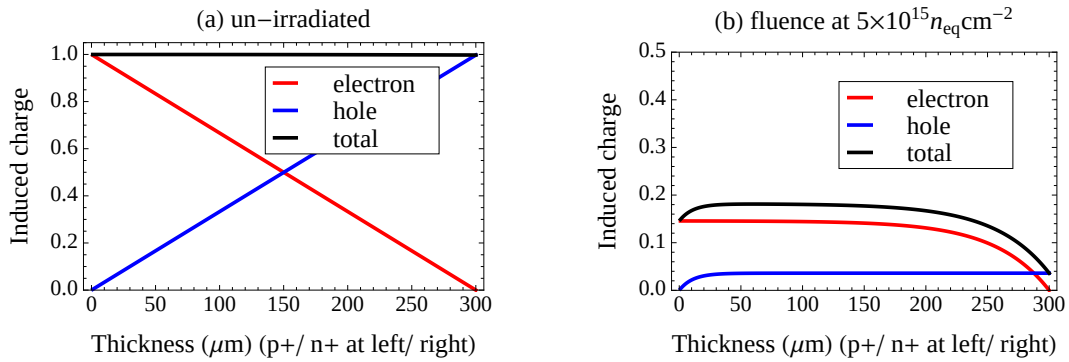


Figure 5.18: Signal induced in a pad diode by electrons and holes for unirradiated, in which the contributions from electrons and holes are identical due to their relatively long carrier lifetimes, and irradiated cases at difference positions.

Chapter 6

Signal Response of Irradiated Silicon Sensors to Radiation Sources

One method to characterise the detector performance is to use the energetic particles from decay processes of radiation sources, such as ^{90}Sr, ^{60}Co and ^{241}Am. This can give information about signal response and noise for pixels (strips) of the device under study. There is an important and unexpected phenomenon observed in these experimental data, in which the irradiated silicon devices have signal gain, i.e. more electron-hole pairs are generated than expected [28]. Sometimes the signal response is even larger than before irradiation [79][64]. The reason is that the deep level defects are filled with carriers and create more space charge, which confines the voltage drop, increases the electric field and initiates the avalanche effect.

Strip detectors were simulated and compared with the experimental data in this chapter. Strips are manufactured with two technologies, planar strips and 3D cells. Strip detectors are the second layer of the ATLAS tracker which provides less accuracy but a lower cost. 3D pixel detectors for the IBL were also simulated and compared with the experimental data. FBK and CNM use different geometries in processing which leads to differences in avalanche behaviour. This chapter will conclude by assessing possible designs to generate avalanche gain in unirradiated structures.

6.1 Planar Strip Detectors

The detectors that were modelled have n-in-p planar strips which are 80 μm wide. Around half of the strip area is covered by a $n+$ implant as the readout electrode and an uniform $p+$ implant as the bias electrode on the back. The structure is sliced vertically

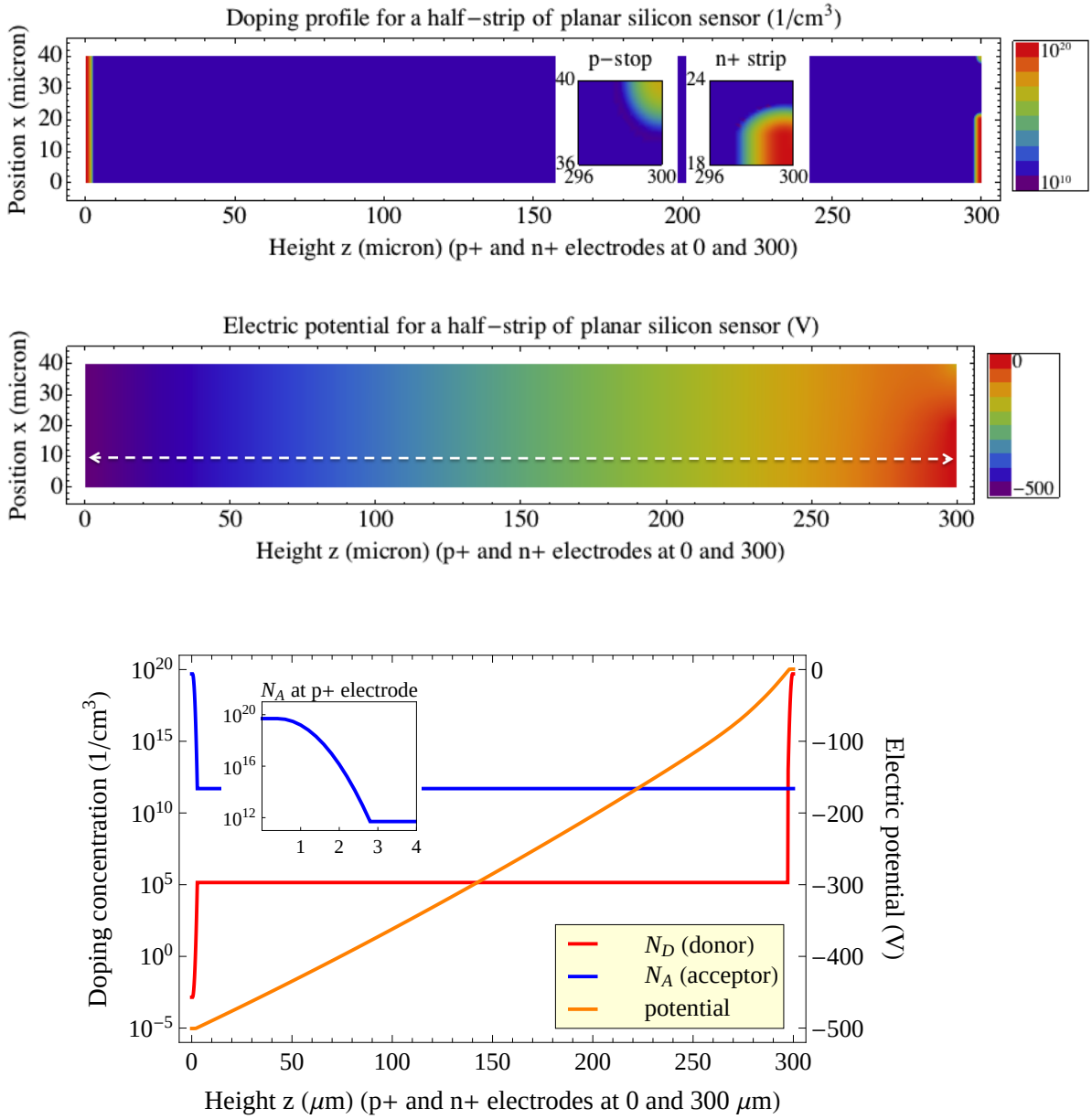


Figure 6.1: The doping profile used in simulation (top) and the electric potential distribution obtained (middle) from simulation. The inset shows the zoomed doping profile of the $n+$ electrode, p-stop (top) and $p+$ electrode (bottom). The 2D structure is sliced horizontally across the two electrodes which is a half-strip of a planar strip sensor. The 1D curve is a line across the two electrodes (bottom) which shows the electric potential is roughly a straight line as expected for the p-i-n junction structure.

across the sensor plane for 2D simulation. The width of the highly $n+$ doped region was set to 40 μm in the simulation and a p-stop was used between strips to prevent the $n+$ electrodes from shorting especially after irradiation. The doping concentrations of the substrate, junction electrodes and p-stop were set to 5×10^{11}, 5×10^{19} and 2×10^{17} $(atoms)cm^{-3}$ respectively.

Figure 6.1 shows the doping profile and electric potential for a half-strip of a planar strip sensor which is the simulated area. The inset shows a zoomed view of the concentration gradient on the p-stop, $p+$ and $n+$ electrode. The electric potential is almost a straight line as expected for the p-i-n junction structure with a larger voltage drop on the $n+$ electrode side due the use of a p-type substrate. Kurata uses different models to explain the data in this section.

6.1.1 Experimental Results

Casse provided a complete set of measurements for planar strip sensors at different fluences and thickness [28]. Measurements were carried out using β particles from a ^{90}Sr source and these electrons (0.546 MeV) are close to being minimum ionising particles (MIPs) in terms of energy deposition. The readout was based on the SCT128 40 MHz analogue chip.

The experimental data for strip sensors with a 300 μm thickness was chosen for simulation. These samples were irradiated with reactor neutrons. There are also samples irradiated with 24 GeV and 26 MeV protons with thickness of 140 and 300 μm in the paper, which show the advantage of thinned planar devices which are easier to deplete and give a better signal efficiency.

Figure 6.2 shows the collected charge of irradiated strip sensors versus reverse bias at different fluences up to 2×10^{16} $n_{eq}cm^{-2}$. The signal response increases with the bias voltage. It is proportional to the depletion width when trapping is ignored. Simulation aims to explain the relationship between voltage and depletion width.

For unirradiated and lightly irradiated sensors ($< 5 \times 10^{14}$ $n_{eq}cm^{-2}$), the signal response shows an initial rise followed by a plateau[1]. The rising region is due to partial depletion with the depletion width proportional to the square root of voltage, while the device is fully depleted in the plateau region. For heavily irradiated sensors ($> 5 \times 10^{14}$ $n_{eq}cm^{-2}$), the device are never fully depleted due to the large substrate thickness. The signal response shows a linear increase with voltage which does not agree with Equation 2.54. This may imply an increment in signal at high voltages.

[1]Figure 6.2 shows a clear plateau up to a fluence of 2×10^{14} $n_{eq}cm^{-2}$ and a small plateau at the very end of the fluence of 5×10^{14} $n_{eq}cm^{-2}$.

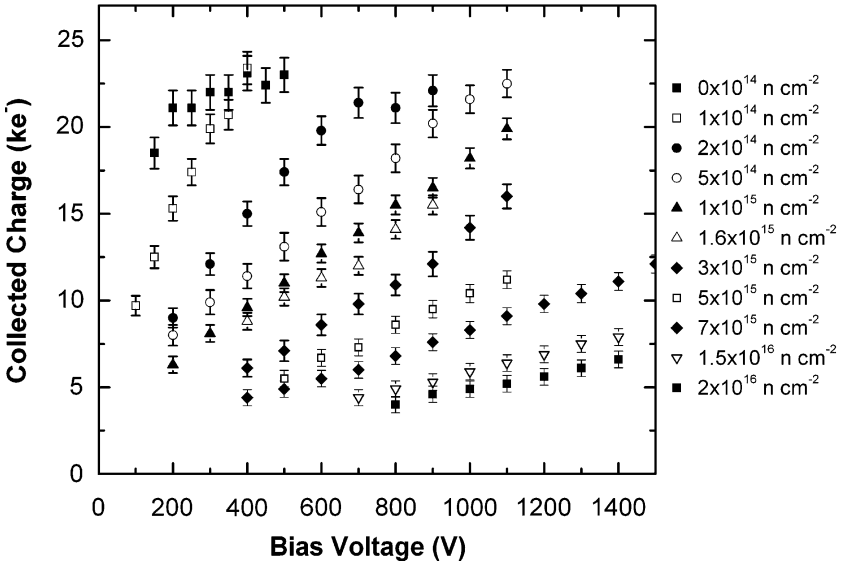


Figure 6.2: Measurements of neutron irradiated and 300 μm thick planar strip sensors at different fluences using ^{90}Sr source [28]. The signal response increases with the bias voltage which indicates that the depletion region widens.

6.1.2 Simulation Overview

The "signal response" in Chapter 6 and 7 is defined as the collected charge at electrodes where carriers suffer from trapping and recombination. The "signal efficiency" is defined as the "normalised" collection efficiency which is the final charge collected over initial charge generated. This is not the "tracking efficiency" used for track reconstruction.

The signal efficiency across the device is obtained by two steps with the algorithms described earlier. The device simulation in Chapter 4 provides electric field maps as an input for the tracking simulation in Chapter 5 which calculates the induced signal. The calculated efficiencies at each (binned) position in a device is the sum over all signal induced at the amplifier divided by the initial number of electron-hole pairs. The efficiency is normalised to a value from 0 to 1. One corresponds to 100% collection of the generated carriers.

The experimental data shown in Figure 6.2 were measured by amplifiers which read the current or voltage increments and convert to the values in ToT or collected charge. To compare simulations with experiments, the unirradiated simulation result (from 0 to 1) is scaled up to the same magnitude as the unirradiated experimental data in this chapter. Thus a scaling factor of 25 ke^- from signal efficiency to charge was used for planar strips in this section.

The measured signal response is an overall property which is affected by various factors. For example, the trapping lifetime of carriers affects the overall signal magnitude

Model	Parameter	Description
G1	$\alpha_{n/p} = 0$ in device simulation $\alpha^T_{n/p} = 0$ in tracking simulation	no avalanche effect for leakage current no avalanche gain for signal response
G2	$\alpha_{n/p} = 0$ in device simulation $\alpha^T_{n/p} \neq 0$ in tracking simulation	no avalanche effect for leakage current with avalanche gain for signal response
G3	$\alpha_{n/p} \neq 0$ in device simulation $\alpha^T_{n/p} = 0$ in tracking simulation	avalanche enhanced leakage current no avalanche gain for signal response
G4	$\alpha_{n/p} \neq 0$ in device simulation $\alpha^T_{n/p} \neq 0$ in tracking simulation	avalanche enhanced leakage current with avalanche gain for signal response

Table 6.1: Combinations of with or without impact ionisation in the device and tracking simulation. The avalanche enhanced leakage current and signal gain are additional parameters in the device and tracking simulation using the same impact ionisation model discussed previously. The model G2 and G3 are not reasonable but are simulated for comparison.

in the tracking simulation (exponential trapping term), but not the shape of efficiency versus voltage. This will be shown in Section 6.3.

The generation lifetime represents the leakage current which changes the electric field distribution in the device simulation. The region with a non-zero electric field is usually referred to as the depletion region which strongly affects the signal efficiency in the tracking simulation. Thus the generation lifetime affects both the device and tracking simulation.

Impact ionisation and the deep donor concentration are important which significantly change the electric field distribution in the device simulation. Impact ionisation is also considered in the tracking simulation which gives extra gain and changes the shape of the signal response versus voltage. There are four combinations to turn on/off the avalanche effect in the device and tracking simulation as listed in Table 6.1. They will be simulated and compared in this section first. The impact ionisation coefficients for both simulations are from the Crowell-Sze model in Chapter 4.

There is also experimental evidence of a non-linear relationship between the total dose and effective doping concentration at high fluences. This will be confirmed by comparing the quasi-Fermi level approximation and the SRH model. This section will conclude by investigating the effects of the leakage current and deep donor concentration. Further modifications will be needed to fully explain the data.

6.1.3 Effects of Impact Ionisation in the Device and Tracking Simulation

Impact ionisation affects the solution of the device simulation and also increases the signal response in the tracking simulation. The lifetime parameters, $\beta_n = 3.7 \times 10^{-16}$ and $\beta_p = 5.7 \times 10^{-16}\ cm^2 ns^{-1}$ for electrons and holes respectively, are used in this

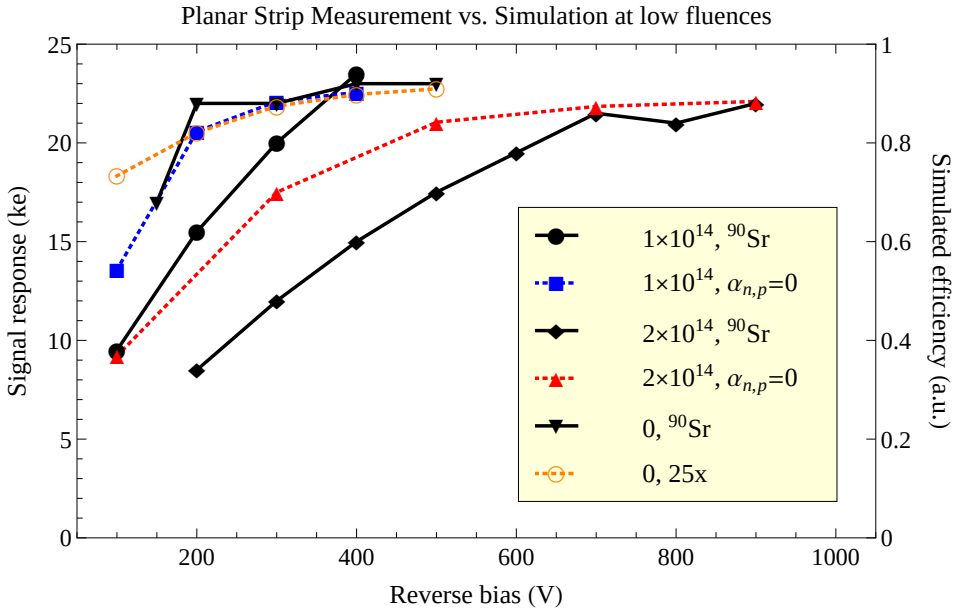


Figure 6.3: Experimental data and simulated signal efficiencies for planar strip sensors at low fluences of 1×10^{14} (labelled "1×10^{14}") and 2×10^{14} (labelled "2×10^{14}") $n_{eq}cm^{-2}$. The solid lines are the data while the dotted lines are simulations.

section. Carrier mobilities in this chapter depend on the electric field and doping concentration as discussed in Chapter 4.

Simulation Results of Low Fluences

Figure 6.3 shows the simulation results for planar strip sensors at fluences of 1×10^{14} and 2×10^{14} $n_{eq}cm^{-2}$. The simulation predicts the same signal efficiency as the experimental data at high voltages where saturation occurs. However, the simulation predicts a much higher signal efficiency at low voltages. The peak value of the electric field at low fluences is not strong enough for impact ionisation to occur due to early depletion. The avalanche effect on the leakage current was not simulated. The change in the signal efficiencies is also negligible (less than 1%) up to this irradiation level. Thus simulation results are plotted without impact ionisation (labelled "$\alpha_{n,p} = 0$"). The full depletion voltage is thought to be smaller due to the use of a 2D simulation, whereas the device is three dimensional. The simulated device reaches plateau earlier than the data.

Simulation Results at High Fluences without Avalanche Enhanced Generation Current

Figure 6.4 shows the simulation results for planar strip sensors at fluences of 5×10^{14} (top, labelled "5×10^{14}") and 1×10^{15} (bottom, labelled "1×10^{15}") $n_{eq}cm^{-2}$. Blue curves do not consider the avalanche enhanced generation current in the device simulation

(labelled "$\alpha = 0$"). The avalanche gain can be included (labelled "$\alpha^T \neq 0$", G2) or not (labelled "$\alpha^T = 0$", G1) in the tracking simulation, in which impact ionisation increases the signal response.

The simulated signal efficiency versus voltage is initially parabolic at a fluence of 5×10^{14} $n_{eq}cm^{-2}$ since the depletion width increases with the square root of voltage. It deviates from the data at higher voltages and tends to saturate. The simulated curve becomes linear when impact ionisation is included in the tracking simulation since the avalanche gain increases exponentially with voltage. The simulation shows a large gain at a fluence of 1×10^{15} $n_{eq}cm^{-2}$ and has a similar slope as the data. However, the data are much better than the simulation in total amplitude which cannot be explained by the trapping lifetimes.

The reason is that the depletion width becomes very narrow due to the high space charge density and thus the voltage drop is confined around the $n+$ electrode. The simulated electric field is higher than to 40 $V\mu m^{-1}$ at 1000 V which exceeds the breakdown voltage of silicon. Avalanche gain enhances the signal response which gives better agreement with the data. This is included in the tracking simulation but not in the device simulation. This is not reasonable and thus motivated the implementation of impact ionisation for the generation current in Kurata version 4.

Simulation Results at High Fluences with Avalanche Enhanced Generation Current

The red curves in Figure 6.4 show the simulation results at the same fluences but include impact ionisation in the device simulations (labelled "$\alpha \neq 0$"). The updated results show better agreement with the experimental data at a fluence of 5×10^{14} $n_{eq}cm^{-2}$. The avalanche gain (between G4 and G3) in the tracking simulation as a percentage is smaller than the previous results (between G2 and G1). This is because the device tends to be fully depleted by including the avalanched enhanced generation current. The extra generation current increases the depletion width through the defect occupancies in the Poisson equation. The peak field also decreases since the total voltage is the integral of the electric field.

It is crucial for sensors with a 300 μm substrate to introduce the avalanche enhanced generation current at fluences above 5×10^{14} $n_{eq}cm^{-2}$ to give a reasonable depletion voltage. The measured full depletion voltage is around 800 V at a fluence of 5×10^{14} $n_{eq}cm^{-2}$ from Figure 6.6. The full depletion voltage in the device simulation without impact ionisation is around 1500 V, while it is between 800 to 900 V with the avalanched enhanced generation current included as shown in Figure 4.7 (c). This can also be observed in the simulated signal efficiency (G3) which has a plateau at high voltages.

The results with impact ionisation included are also needed to explain the data at

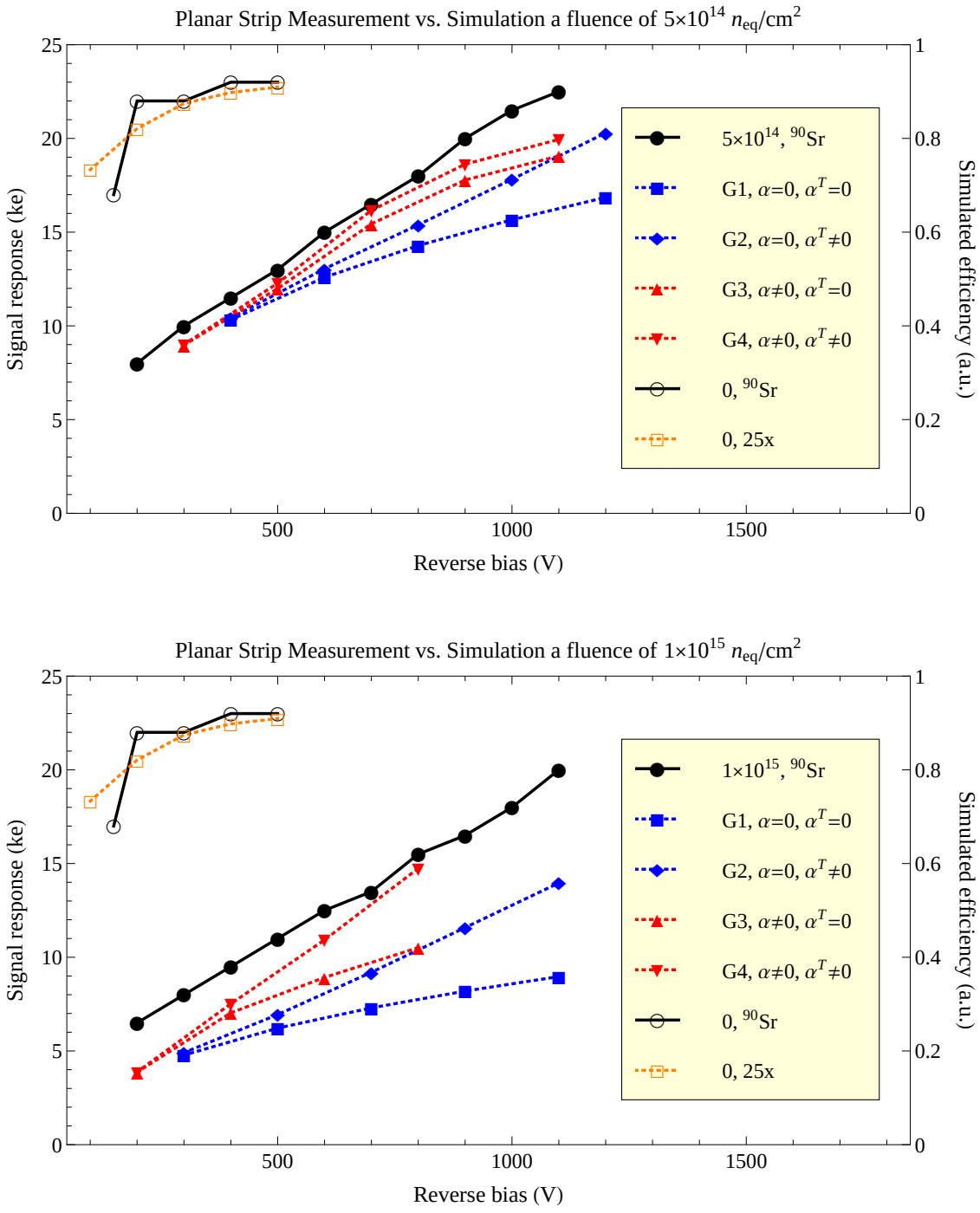


Figure 6.4: Experimental data and simulated signal efficiencies for planar strip sensors at fluences of 5×10^{14} (top, labelled "5×10^{14}") and 1×10^{14} (bottom, labelled "1×10^{15}") $n_{eq}cm^{-2}$. Impact ionisation on the generation current is ignored (blue, labelled "$\alpha = 0$") or included (red, labelled "$\alpha \neq 0$") in the continuity equations of the device simulation. Avalanche gain of the signal response is ignored (labelled "$\alpha^T = 0$") or included (labelled "$\alpha^T \neq 0$") in the tracking simulation. The solid lines are the data while the dotted lines are simulations.

a fluence of 1×10^{15} $n_{eq}cm^{-2}$. They agree well up to 800 V where early breakdown occurs due to the high generation current. The highest field occurs between the $n+$ electrode and p-stop as shown in Figure 4.7 (a). Kurata fails to handle local overflows which leads to breakdown since its mesh is not adaptive. When the simulation still breaks down with a finer mesh, it suggests that this is the maximum operation voltage for such condition.

To summarise, impact ionisation in the tracking simulation causes the simulated signal to be linear since the original parabolic shape of the signal-voltage dependence is combined with an exponential dependence. The avalanched enhanced generation current in the device simulation has to be included since it changes the field shape and predicts an reasonable depletion voltage. It also alters the depletion width which is directly related to the signal efficiency. By including impact ionisation in both the device and tracking simulation, the results are in good agreement with the data. If one adds avalanching to the signal response, one must also add it to the leakage current.

6.1.4 Comparison of Defect Models

The simulation results using the quasi-Fermi level approximation with the model G4 predict lower signal efficiencies than the experimental data as shown in Figure 6.4. Simulations using the SRH model for V_2O were performed for comparison since the defect model strongly changes the sensor behaviour. Kurata version 4 and 6 were validated earlier and used to generate results for the quasi-Fermi level approximation and the SRH model respectively.

Deep levels are generation-recombination centres which contribute to N_{eff} and the leakage current in Kurata version 6. The experimental generation lifetime is used in Kurata version 4. Depending on the Fermi level and carrier concentrations, some of them are electrically active and contribute to the space charge. The difference of the defect occupancies between the two models were discussed in Chapter 3.

The detailed parameters for the SRH model from the literatures are summarised in Table 6.3 [86][96]. The ratio of effectiveness for deep acceptors V_2O (M4) and V_3 (M5) and deep donor C_iO_i is approximately $100 : 50 : 1$ which is mainly dependent on their energy levels with respect to mid-gap.

Simulation Results at High Fluence

The irradiated data reach the full collection efficiency bounded by the unirradiated data at a fluence of 5×10^{14} $n_{eq}cm^{-2}$, while the simulation result using the defect model for V_2O (M4) saturates at around 800 V as shown in Figure 6.5 (blue, labelled "5×10^{14}, M4"). There is avalanche gain in the data, but no gain is observed in the simulation using the model M4. The simulation using M4 predicts roughly the same magnitude of

Model	Name	Description	β (cm^{-1})	at Fluences F/F' $(n_{eq}cm^{-2})$
M2	$N_{eff,0}$	original RD20 model $N_{eff,0} = 0.016F + 0.66 \times 10^{11}$	1	validated up to 1×10^{14} assumed up to 5×10^{14}
M3	N_{eff}	β rescaling $N_{eff} = N_{eff,0} + 0.0016F'$	interpolation 0.1	between 5×10^{14} and 2×10^{15} after 2×10^{15} $(F = 5 \times 10^{14})$

Table 6.2: Some properties about the quasi-Fermi level model used in the simulation. One representative deep level for at mid-gap [81] for all deep acceptors and donors each. β is the introduction rate.

Model	Name	Two-level	Defects	σ_n (cm^2)	σ_p (cm^2)	β (cm^{-1})	Ref.
M4	V_2O	$E_C - 0.55$	$V_2O^{(-/0)}$	1×10^{-15}	1×10^{-16}	0.08	[86]
		$E_V + 0.36$	$CO^{(0/+)}$	1×10^{-16}	1×10^{-15}	1	
M5	V_3	$E_C - 0.46$	$V_3^{(-/0)}$	5×10^{-15}	5×10^{-14}	0.9	[96]
		$E_V + 0.36$	$C_iO_i^{(0/+)}$	2.5×10^{-14}	2.5×10^{-15}	0.9	

Table 6.3: The parameters of of two-level defects models suggested in literatures which were used in simulation. σ_n and σ_p are the capture cross sections of electrons and holes for deep level defects respectively. Two-level defect models consider only one dominant defect for deep acceptors and donors respectively for simplicity, while TCAD can consider more deep levels as suggested in the reference [86][96]. Two dominant deep acceptors are V_2O and V_3 which have different effects in terms of the space charge. Note that the introduction rate of V_2O is quite low since it is a secondary product which is mainly formed from V and VO.

the signal efficiency compared to the quasi-Fermi level approximation (M2) in Figure 6.4 (top) which are at depletion. But M4 depletes earlier than M2 which suggests that M4 gives less space charge.

The irradiated data at a fluence of 1×10^{15} $n_{eq}cm^{-2}$ has a similar shape and slope to the 5×10^{14} $n_{eq}cm^{-2}$ data. It should also have avalanche gain but the overall magnitude of the signal efficiency is limited by the short trapping lifetime. The simulation result (red, labelled "1×10^{15}, M4") using the model M4 has a similar shape and slope as the data, but its overall magnitude is smaller. This suggests the predicted depletion width is smaller than reality. M2 gives smaller signal efficiencies in Figure 6.4 (bottom) compared to M4 which also suggests M4 gives less space charge (N_{eff}).

Combining the two simulation results suggests that the defect model needs some modification. The SRH defect model (M4) estimates less space charge at a fluence of 5×10^{14} $n_{eq}cm^{-2}$, while it estimates more at a fluence of 1×10^{15} $n_{eq}cm^{-2}$ than the data. This is indirect evidence from simulation that the effective doping concentration may not be linear with irradiation level.

The high signal gain at high voltages in the data are not observed in the simulations using M4. The signal gain could be generated at the edges/corners of the $n+$ implants since three-dimensional structures cannot be simulated in Kurata. The regions between the $n+$ electrode and p-stop are known to have a large electric field which could increase after irradiation. This is due to the silicon dioxide SiO_2 layer on the surface which

becomes positively charged and attracts an electron layer. The surface charge/damage can be simulated using three-dimensional TCAD.

6.1.5 Rescaling of the Introduction Rate

There is also direct evidence that the effective doping concentration N_{eff} may not be linear with the irradiation level at very high fluences [28]. Figure 6.6 shows two introduction rates which are $\beta = 0.023\ cm^{-1}$ for fluences lower than $5 \times 10^{14}\ n_{eq}cm^{-2}$ and $\beta' = 0.0022\ cm^{-1}$ for fluences greater than $2 \times 10^{15}\ n_{eq}cm^{-2}$. There is a transition between these two linear regions at fluences of $(0.5 \sim 2) \times 10^{15}\ n_{eq}cm^{-2}$. The fit at low fluences β is not clearly shown in the figure in the linear scale. It is based on the RD48 collaboration [73] and the Kramberger group [32] which are well-known to give introduction rates of around 0.02 to 0.06 cm^{-1} depending on the substrate type and impurity levels. The fit at high fluences β' is emphasised to give an introduction rate smaller by an order.

The deep acceptor concentration N_A is a parameter in Kurata using the quasi-Fermi level approximation and the introduction rate is set to 1 cm^{-1}. Deep acceptors results in an effective doping concentration which is linear with the fluence F and its relationship can be fitted as $N_{eff,0} = 0.016F + 0.66 \times 10^{11}$. This agrees well with the experimental data up to a fluence of $1 \times 10^{14}\ n_{eq}cm^{-2}$ [81] and it is assumed to be linear with fluences up to $5 \times 10^{15}\ n_{eq}cm^{-2}$. This is the model M2.

To incorporate the two-stage behaviour observed in the experiments, another term $0.0016F'$ with $\beta' = 0.1\beta$, is added for fluences above $2 \times 10^{15}\ n_{eq}cm^{-2}$. The original F is for fluences up to $5 \times 10^{14}\ n_{eq}cm^{-2}$. This is the model M3. The simulations for planar strip sensors using $N_{eff,0}$ (M2) up to a fluence of $1 \times 10^{15}\ n_{eq}cm^{-2}$ as the original Kurata model is sufficient. The experimental data for 3D sensors have fluences up to $5 \times 10^{15}\ n_{eq}cm^{-2}$ and thus the introduction rate rescaling using N_{eff} (M3) will be used later. M2 and M3 are also listed in Table 6.2.

Selected values at different fluences F_{exp} (ϕ_{exp}) and N_{eff} are listed in Figure 6.6 which are the recorded fluences and corresponding N_{eff} measured by the capacitance-voltage method respectively. The corresponding N_A used in Kurata is also listed in Figure 6.6 which is calculated from F_{exp} and β (β'). For example, a sensor irradiated to a fluence of $5 \times 10^{15}\ n_{eq}cm^{-2}$ is simulated with $N_A = 1.3 \times 10^{15}\ cm^{-3}$ which gives $N_{eff} = 2.6 \times 10^{13}\ n_{eq}cm^{-2}$.

6.1.6 Leakage Current and Deep Donors

The generation lifetime used in the simulations previously gave a leakage current I_L larger than the experimental value by a factor of three. The parameters and models used to produce results in Figure 6.7 are listed in a table of the parameters listed.

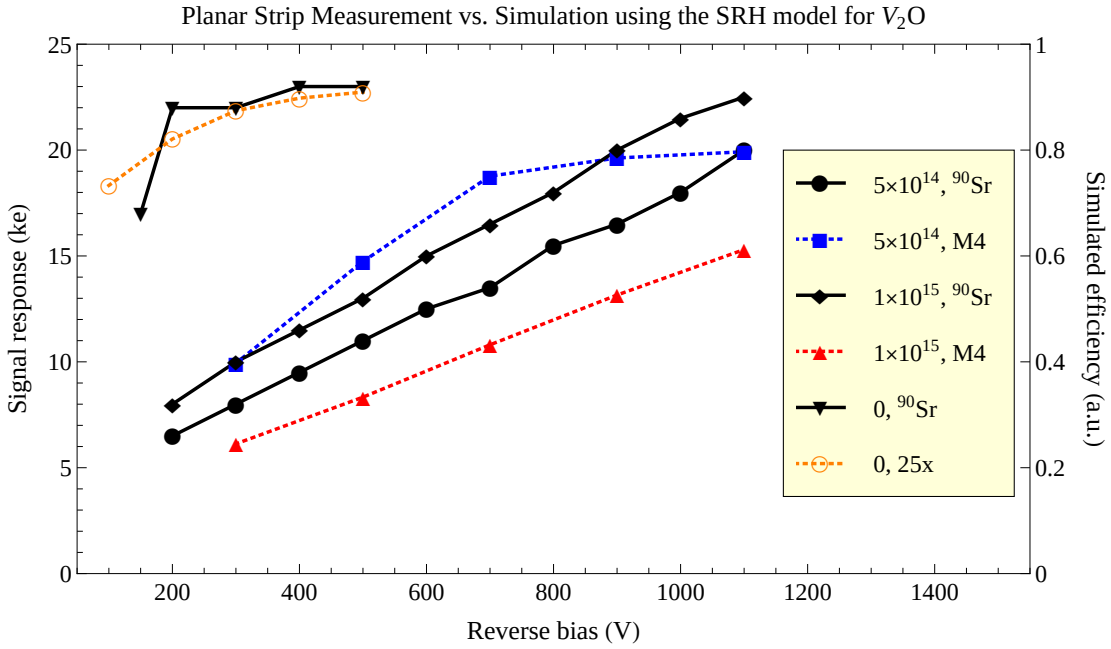


Figure 6.5: Experimental data and simulated signal efficiencies for planar strip sensors at fluences of 5×10^{14} (blue, labelled "5×10^{14} M4") and 1×10^{14} (red, labelled "1×10^{15} M4") $n_{eq}cm^{-2}$. The solid lines are the data while the dotted lines are simulations.

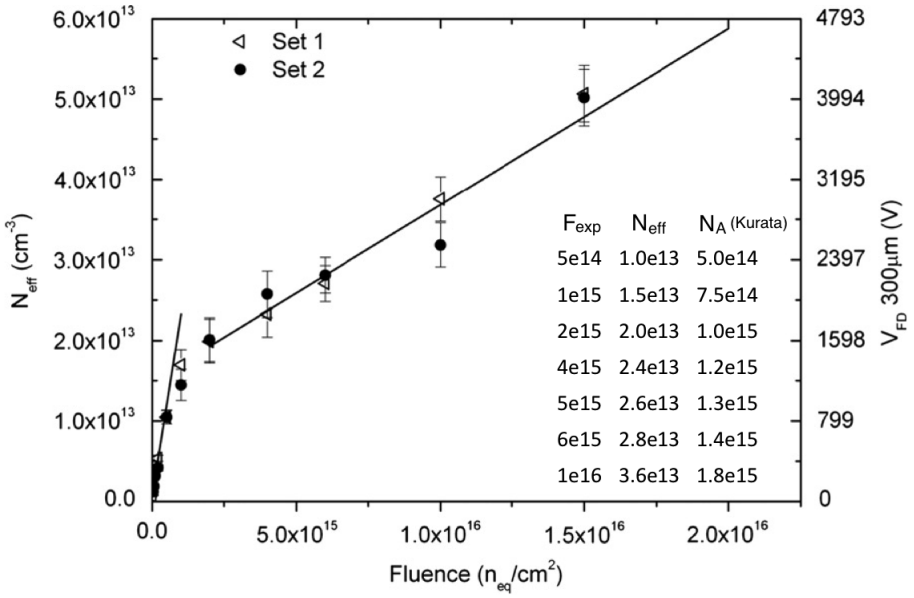


Figure 6.6: Measured effective doping concentrations N_{eff} and the corresponding full depletion voltages V_{FD} of planar strip sensors at difference fluences [28]. Thin planar strip sensors of a 140 μm thickness were measured using the capacitance-voltage (CV) method while V_{FD} were calculated for sensors of a 300 μm thickness. The fit at low fluences is based on earlier experiments which indicated a slope of this order.

N_A/N_D is the ratio of the representative deep acceptor and donor concentration (or the introduction rate equivalently) which are at mid-gap as shown in Figure 3.4.

Figure 6.7 (a) shows the difference between the previous (high I_L) and new (low I_L) value of leakage current for strip sensors at a fluence of 5×10^{14} $n_{eq}cm^{-2}$. The simulation result with the corrected leakage shows a large electric field on the $p+$ electrode side and a very low field in the bulk. This field distribution will give a very low signal efficiency which is not observed in the data.

Figure 6.7 (b) shows the effects of different deep donor concentrations on strip sensors at a fluence of 5×10^{14} $n_{eq}cm^{-2}$. Deep donor levels are known to increase the junction field on the $p+$ side due to the positive space charge. The $p+$ field has a similar magnitude to the $n+$ field in the simulation with $N_A/N_D = 5$. The donor level has to be lowered to give a reasonable electric field on the $p+$ side. $N_A/N_D = 20$ is the lowest ratio needed to explain the data of strip sensors at all voltages.

Figure 6.7 (c) shows a comparison between a pn diode and a strip sensor at a fluence of 1×10^{15} $n_{eq}cm^{-2}$. This proves that the unexpected effect from the deep donor level with a low leakage current is due to geometry. The pn structure is symmetric and gives a reasonable field distribution with $N_A/N_D = 5$, while the strip structure has a very large peak on the $p+$ side. This can be understood by the semiconductor equations which have symmetric acceptor and donor terms.

The suitable deep donor concentration is determined by simulating planar strip sensors which enhances the effectiveness of deep donors at the $p+$ electrode. This also corresponds to the two or three-level model used in TCAD. The effective doping concentration from C_iO_i is much smaller than V_2O or V_3 due to its large energy difference from mid-gap.

Figure 6.7 (d) shows the effect of leakage current on strip sensors at a fluence of 1×10^{15} $n_{eq}cm^{-2}$. This field distribution will predict a very low signal efficiency but a very high gain which is not realistic. There is also no obvious field on the $p+$ side without including impact ionisation in the device simulation.

To summarise, the low leakage current in Figure 6.7 (a) and avalanche enhanced generation current in Figure 6.7 (d) enhance the effectiveness of deep donors in simulation. The high leakage current increases the depletion width in simulation with or without impact ionisation and the latter is shown in Figure 6.7 (d).

6.1.7 Summary

The initial "G1, $\alpha_{n/p} = 0$ & M2" model was found to have problems. The modified "G4, $\alpha_{n/p} \neq 0$ & M2" model, which includes impact ionisation in both the device and tracking simulation, was found to agree better with the experimental data for a fluences of 5×10^{14} and 1×10^{15} $n_{eq}cm^{-2}$. Avalanche generation in leakage current and signal

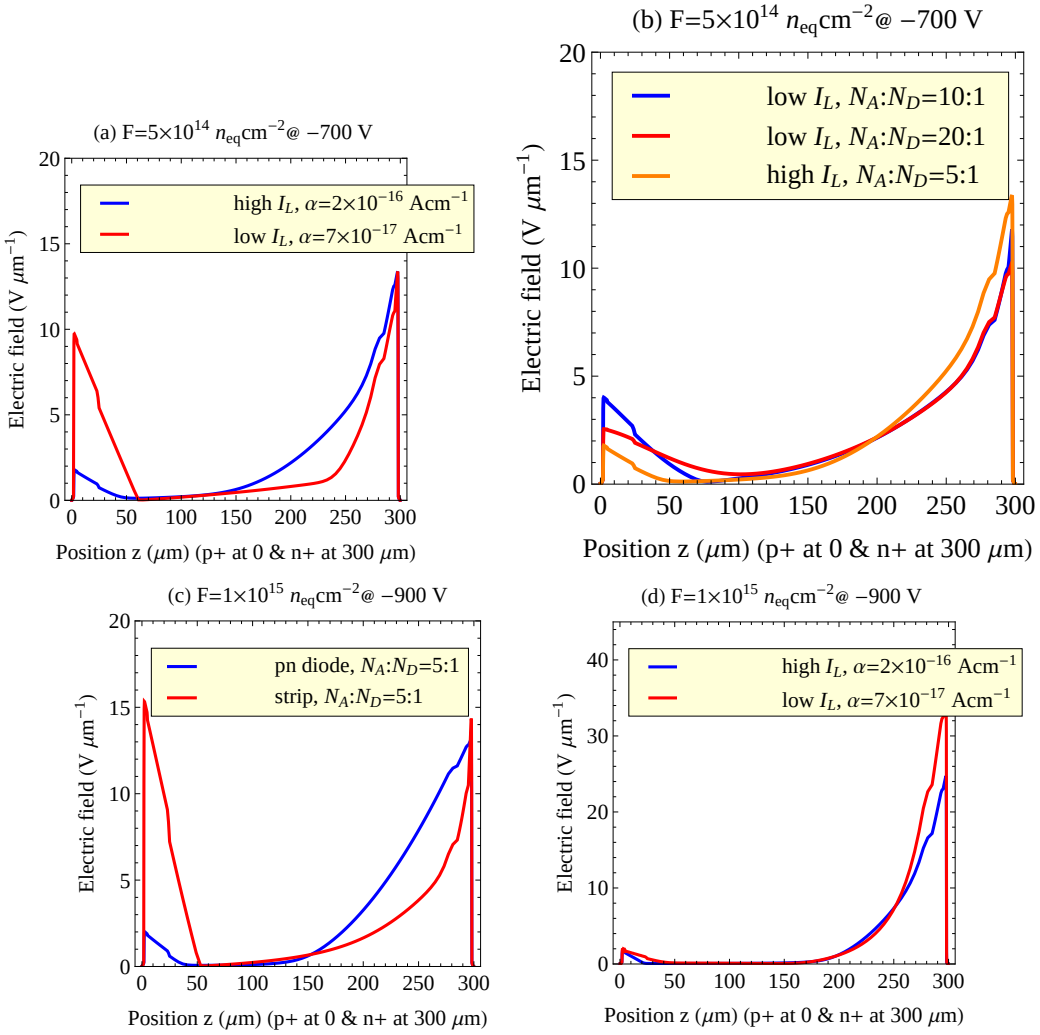


Figure 6.7: Comparison of various parameters for strip sensors. See text for detailed explanation and the parameters are listed as

	$(a.1)^*$	(a.2)	(b.1)	$(b.2)^{**}$	(b.3)	(c.1)	(c.2)	(d.1)	(d.2)
F_{exp} (ϕ_{exp})	$5e14$	$5e14$	$5e14$	$5e14$	$5e14$	$1e15$	$1e15$	$1e15$	$1e15$
N_A/N_D	5	5	10	20	5	5	5	5	5
I_L/α	3	1	1	1	1	1	1	3	1
device model	G4	G4	G4	G4	G4	G4	G4	G1	G1

where $(a \sim d)$ and $(1 \sim 3)$ are four plots and different lines in each plot respectively.

The deep acceptor has an introduction rate of 1 cm^{-1} according to the recorded fluence F_{exp} (ϕ_{exp}) in a unit of $n_{eq}cm^{-2}$. N_A/N_D gives the ratio between the deep acceptor and donor concentration. I_L is the leakage current with the current density $\alpha = 7 \times 10^{-17}$ Acm^{-1}. All simulations were performed using the M2 defect model. The columns labelled * and ** show the parameters used for the previous and final simulations of planar strip sensors. The final parameters with a lower leakage current and deep donor level are closer to the measured values.

response is critical for strip structures, but less important for 3D structure due to the short inter-electrode distance.

The leakage current was initially set too high and then reduced in later simulations. The simulation results with a low leakage current did not agree with the data. It is necessary to alter the N_A/N_D ratio which has to be reasonably small to give the correct field distribution. This is also less obvious for 3D structures due to the symmetric geometry of their cylindrical electrodes.

The introduction rate of deep acceptors needs some rescaling using the "G4, $\alpha_{n/p} \neq 0$ & M3" model which will be shown to agree better with the data than M2 for 3D sensors in later sections. Their signal efficiencies will also be recalculated using these new parameters.

Figure 6.8 shows the simulation results with the corrected leakage current (low I_L, $\alpha = 7 \times 10^{-17}\ Acm^{-1}$) and the lowered deep donor concentration ($N_A/N_D = 20$) without rescaling the deep acceptor concentration (N_{eff0}, M2). The signal-voltage dependence is parabolic initially and reaches a plateau. However, the data show a linear trend with voltage which suggests some avalanching.

Signal gain is not observed in the simulation results using the modified parameters. The results for low fluences in Figure 6.3 show an earlier depletion than expected. Both may be due to the need for a full three-dimensional simulator, such as TCAD, which can include surface effects and edge/corner structures. They may also imply some other physics at very high fluences and voltages.

To summarise, implementing impact ionisation in both the device and tracking simulation is crucial. Leakage current and the introduction rates for deep acceptors and donors are also important parameters at all fluences and voltages for planar strip sensors. α and N_A are determined by the measured values in simulation. N_A/N_D may be found by further experiments.

6.1.8 Charge Sharing

Figure 6.9 shows the charge sharing at different voltages. Higher voltages have larger depletion regions and carriers are less affected by thermal diffusion. This plot depends strongly on the time step $\triangle t$ in simulation. It was set to $5 \times 10^{-12}\ s$ which exaggerates charge sharing for clarity.

6.2 3D Strip Sensor

Strip detectors are traditionally manufactured using planar technology. One can also use 3D technology for better radiation hardness. CNM has constructed such devices. The $p+$ column electrodes form square cells of 80 μm^2 with a $n+$ electrode in the

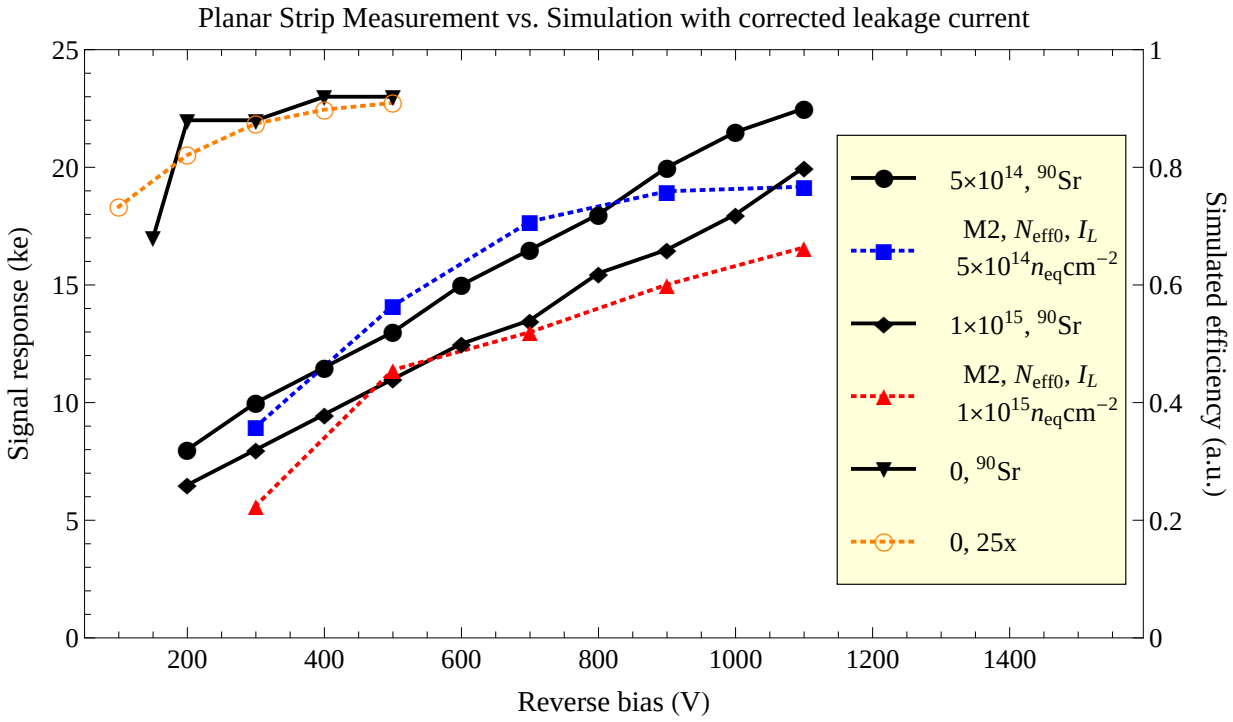


Figure 6.8: Experimental data and simulated signal efficiencies for planar strip sensors at fluences of 5×10^{14} (blue, labelled "M2, N_{eff0}, I_L") and 1×10^{14} (red) $n_{eq}cm^{-2}$. The solid lines are the data while the dotted lines are simulations.

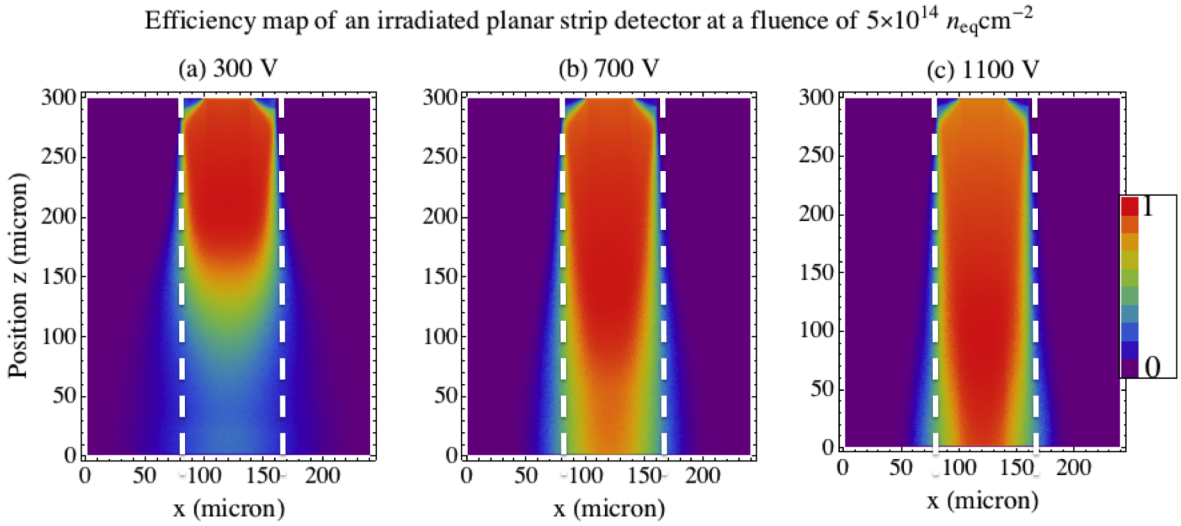


Figure 6.9: Charge sharing of planar strip sensors at a fluence of 5×10^{14} $n_{eq}cm^{-2}$ for different voltages. The time step $\triangle t$ is 5×10^{-12} s in these simulations using the parameters in the caption of Figure 6.7 (a.1). The dotted white lines indicate the strip edges and charge sharing is contributed from the colour regions.

centre on a p-type substrate. This is a double-sided n-in-p design with non-fully-through electrodes. The $n+$ electrodes on a row are metallised and connected as a readout strip, while all the $p+$ electrodes are connected together as a bias plate.

6.2.1 Experimental Results

The experimental data of CNM 3D strip sensors were chosen for simulation. Measurements were performed using β particles from a ^{90}Sr source and pion beams from the CERN SPS. There is a small difference between the two which is most likely due to the multiple scattering of the β particles. The readout was based on the CMS tracker APV25 chip with a 50 ns shaping time at 40 MHz. The samples were irradiated with 25 MeV protons at the Karlsruhe Compact Cyclotron in Germany [64].

Figure 6.10 shows the measurements and simulation results by the CNM group [23]. There is a large avalanche gain at a fluence of 2×10^{15} $n_{eq}cm^{-2}$ whose highest magnitude is almost double compared to the unirradiated data. However, the TCAD simulation only predicts a small gain at a higher voltage and overestimates the signal response at low voltages. The data and simulation results at a fluence of 2×10^{16} $n_{eq}cm^{-2}$ have different operation ranges, but the overlapping voltages are in agreement.

6.2.2 Simulation Results

Figure 6.11 shows the simulation results at fluences of 2×10^{15} and 2×10^{16} $n_{eq}cm^{-2}$ using the 2D electric field maps generated from TCAD by Povoli at Trento [100]. The experimental data from the ^{90}Sr source scan, which has a lower signal efficiency than the pion beam test, is compared with simulation. The scaling factor of 26 ke^{-} from signal efficiency to charge was used for 3D strip sensors in this section.

The large gain observed in the experimental data at fluences of 2×10^{15} $n_{eq}cm^{-2}$ is considered to occur around the tips of the electrode columns, since CNM detectors have a non-fully-penetrating design. There is only a small gain at the highest voltage since Kurata is a 2D simulator and cannot model the electrode tips.

TCAD can model the three-dimensional electrode tips but it also fails to describe the data. This may due to the low effective doping concentration the use of the V_3 model and thus the device tends to deplete earlier and has a smaller peak electric field. This may also imply some other physics at very high fluences and voltages.

6.3 FBK Pixel Detector

FBK is one provider of 3D pixel detectors for the IBL. The IBL layout has 80×336 pixels on a 2 cm^2 detector. The pixel size is designed as 250×50 μm^2 to enhance the resolution in the azimuthal direction. Each pixel has two cells (2E design) and each

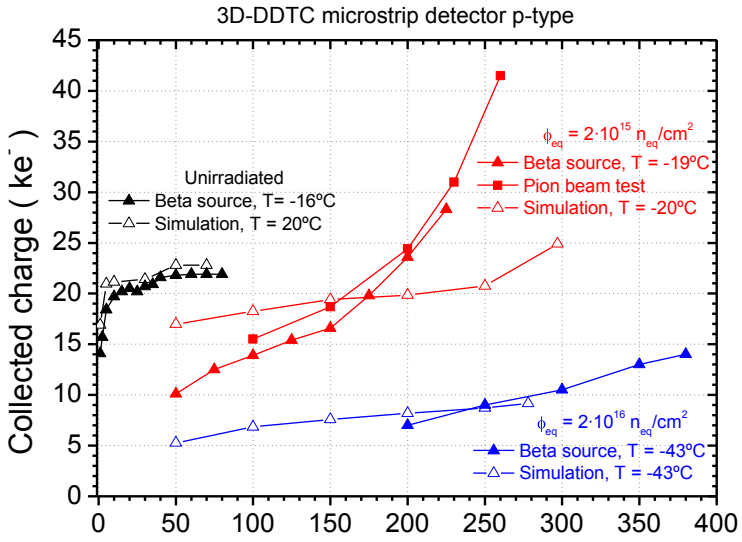


Figure 6.10: Measurement using a ^{90}Sr source and the CERN SPS of 3D strip sensors at fluences of 2×10^{15} (red) and 2×10^{16} (blue) $n_{eq}cm^{-2}$. The filled and open scatters are the data and simulation respectively [23].

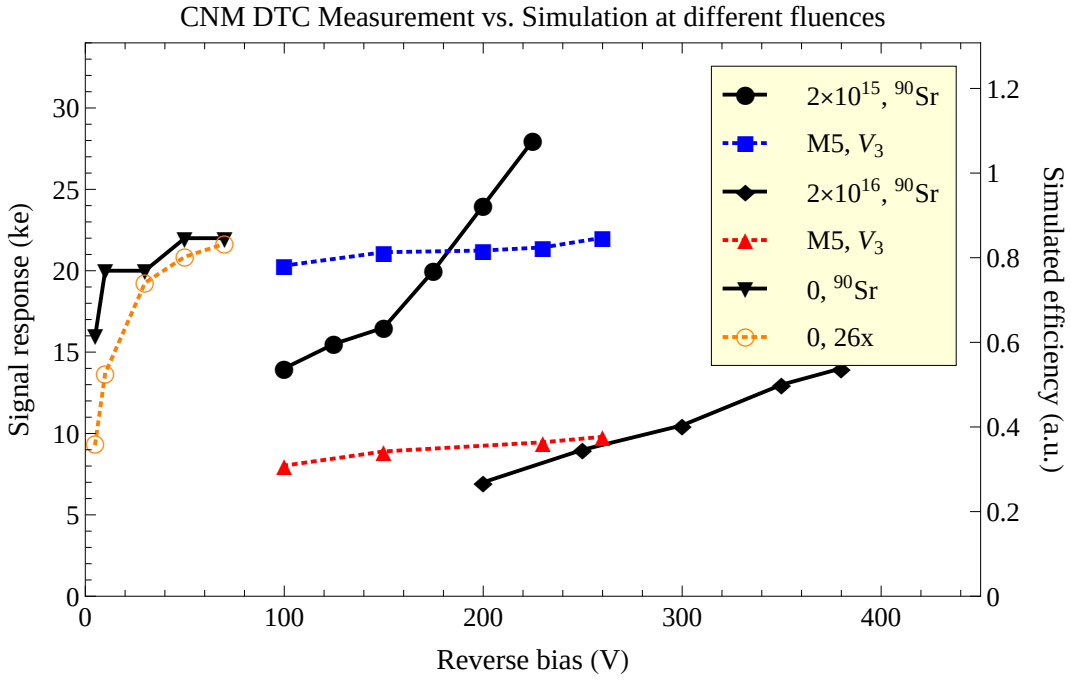


Figure 6.11: Experimental data and simulated signal efficiencies of CNM 3D strip sensors at fluences of 2×10^{15} and 2×10^{16} $n_{eq}cm^{-2}$. The former shows a clear sign of avalanching which cannot be modelled by 2D simulations. The latter shows a larger depletion at lower voltages due to the V_3 defect model. The solid lines are the data while the dotted lines are simulations.

cell is 125×50 μm^2. This design is a compromise between radiation hardness (shorter inter-electrode distance and larger signal efficiency) and noise (lower capacitance). 3D pixel sensors use n-in-p technology which has a faster collection time due to the higher mobility of electrons collected at the n electrodes. These $n+$ readout electrodes are in the centre of cells, while the $p+$ bias electrodes are at the cell corners.

Measurements were performed by Micelli at Udine in Italy using β particles from a ^{90}Sr source [83]. The silicon sensor chip is assembled with the FE-I4 readout chip using flip-chip technology. Bump-bonds are used to connect the two chips. The FE-I4 and sensor assembly is mounted on a single chip card and connected to a USBPix board (adapter card plus multi-IO board) by an Ethernet cable. The whole setup is linked to a computer by a USB cable and controlled by the STcontrol software for device operation and data acquisition [6].

A quarter-cell region was modelled in the device simulation to save computing time and then expanded to a larger area using symmetry. A four-pixel area was modelled in the tracking simulation to calculate the charge collection efficiency. The charge sharing simulation will be compared to the cluster size data in the next chapter. Simulations were in 2D using a horizontal slice across the sensor and thus the p-stop, p-spray and silicon dioxide around the surfaces were not included. The doping concentrations of the substrate and junction electrodes were set to 5×10^{11} and 5×10^{19} $(atoms)cm^{-3}$ respectively. Figure 6.12 shows the doping profile and electric potential for a quarter-cell and a cell of a 3D FBK sensor.

6.3.1 Comparison of Irradiation Levels and Trapping Lifetimes

The effective doping concentration changes the electric field distribution which affects device operation, since the signal is mainly contributed by carriers generated in the depletion region. The trapping lifetime also alters the signal efficiency since it controls the probability for carriers to be trapped.

Figure 6.13 shows the simulation results using different parameters for 3D sensors. N_A is the deep acceptor concentration in the device simulation. $\beta_{long/short}$ are two sets of the carrier trapping lifetimes in the tracking simulation. β_{long} corresponds to the damage constants of 3.2×10^{-16} and 3.5×10^{-16} $cm^2 ns^{-1}$ for electrons and holes which were used for 3D sensors, while β_{short} corresponds to $\beta_n = 3.7 \times 10^{-16}$ $cm^2 ns^{-1}$ and $\beta_p = 5.7 \times 10^{-16}$ $cm^2 ns^{-1}$.

The signal efficiencies drop significantly as the deep acceptor concentration increases due to partial depletion. The absolute difference between two schemes (red and blue curves) is at most 5% in the signal amplitude which is not that critical. This shows the importance of having a correct effective doping concentration.

If the rescaling of the introduction rate at high fluences is correct, F_{exp} (ϕ_{exp}) is the

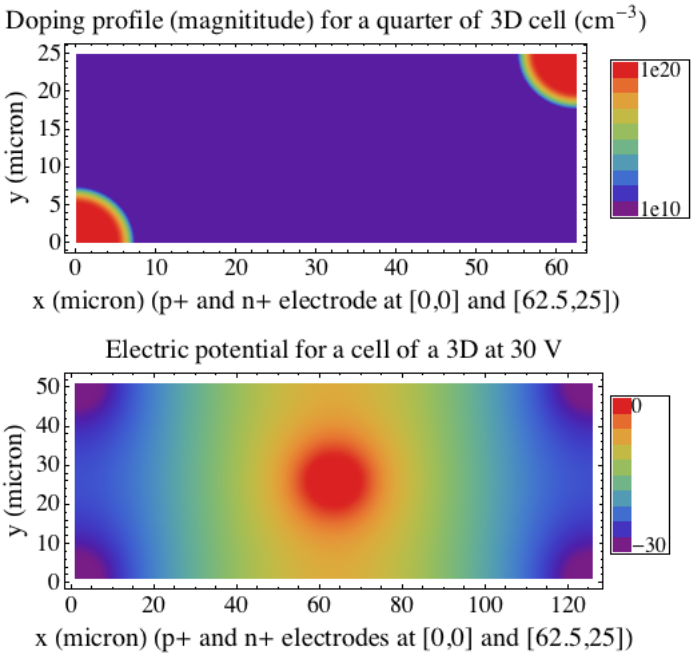


Figure 6.12: Doping profile used in (upper) and electric potential distribution obtained (lower) from simulation for a quarter-cell and and a cell of a 3D strip sensor.

corresponding experimental fluence from the model M3. This implies that 3D sensors will have acceptable signal response at a fluence of 2×10^{16} $n_{eq}cm^{-2}$ which is the expected dose for the HL-LHC. This rescaling is also preferred by the experimental data and is discussed in this section.

6.3.2 Comparison of Leakage Current and Defect Models

The effects of the introduction rate rescaling and leakage current in Table 6.2 using the quasi-Fermi level approximation are simulated for 3D sensors. The rescaling is preferred by the experimental data. Leakage current gives different effects compared to the strip structures. Two real deep defects in Table 6.3 for the SRH model are also compared. The scaling factor of 10.5 from signal efficiency to ToT was used for 3D FBK sensors in this section.

Quasi-Fermi Level Approximation

The simulation using the quasi-Fermi level approximation need the leakage current and deep defect concentrations as input parameters. The higher and corrected leakage current were compared for planar strip sensors in the previous section and the corrected one showed a smaller depletion width. They are also compared for 3D FBK sensors with the leakage current damage constant $\alpha = 2 \times 10^{-16}$ (labelled "I_0") and $\alpha = 7 \times 10^{-17}$

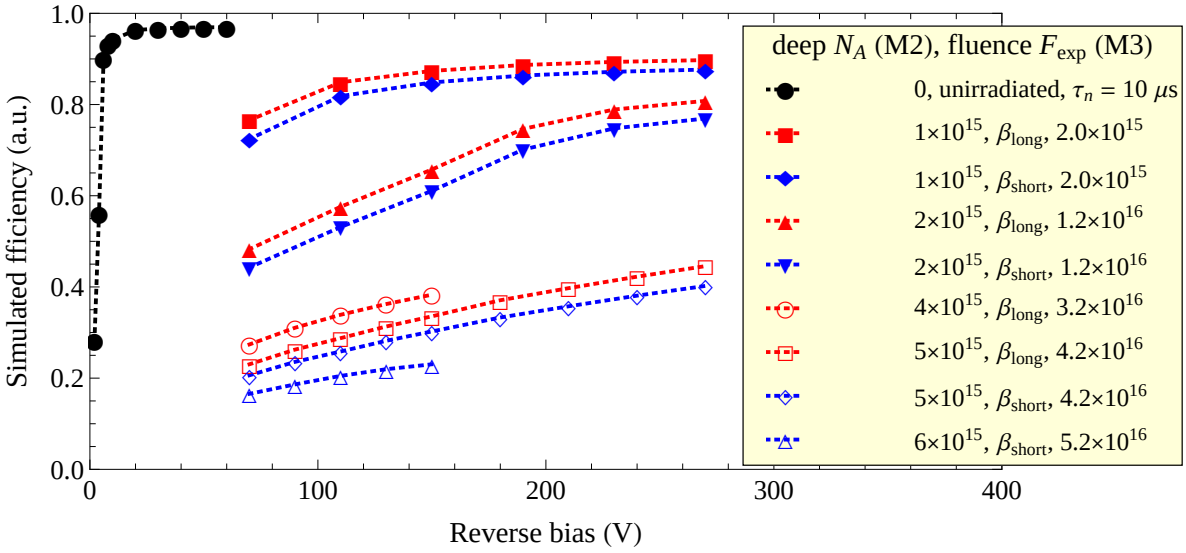


Figure 6.13: Simulated signal efficiency for FBK 3D pixel detectors at different deep acceptor concentrations N_A which are calculated from the fluence and introduction rates. Red (labelled "β_{long}") and blue (labelled "β_{short}") curves are two schemes for the damage constants which are used to calculate the trapping lifetimes by the fluence. β does not affect the signal-voltage shape but only the magnitude. F_{exp} (ϕ_{exp}) is the recorded fluence in experiments if the rescaling of model M3 is correct. Medium voltages around 150 V are used for operation.

(labelled "I_L") Acm^{-1} respectively (denoted as "high" and "low I_L" previously).

Figure 6.14 (top) shows that the simulation results with the corrected leakage current have similar but larger signal efficiencies than with a higher one (green curves) for both the model M2 and M3. The simulation results for 3D sensors are different from planar strip sensors which may be due to their symmetric geometry. Note that two results for I_L and I_0 use $N_A/N_D = 20$ and 5 respectively whereas the former is needed for planar strip sensors as discussed earlier.

The introduction rate rescaling was observed in experiments as shown in Figure 6.6. The model M2 (N_{eff0}) and M3 (N_{eff}) use the original and scaled representative acceptor concentration with an introduction rate of 1 and 1/0.1 (before/after a fluence of $5 \times 10^{14}/2 \times 10^{15}$ $n_{eq}cm^{-2}$) cm^{-1} respectively. This deep acceptor level represents the combination of all real defects with various energy levels. The experimental data were measured in ToT which is assumed to be linear with the collected charge (but not exactly true for the FE-I4 chip). Figure 6.14 (top) shows that the model M3 agrees better with the data than M2 since the linear relationship with fluence overestimates the effective doping concentration.

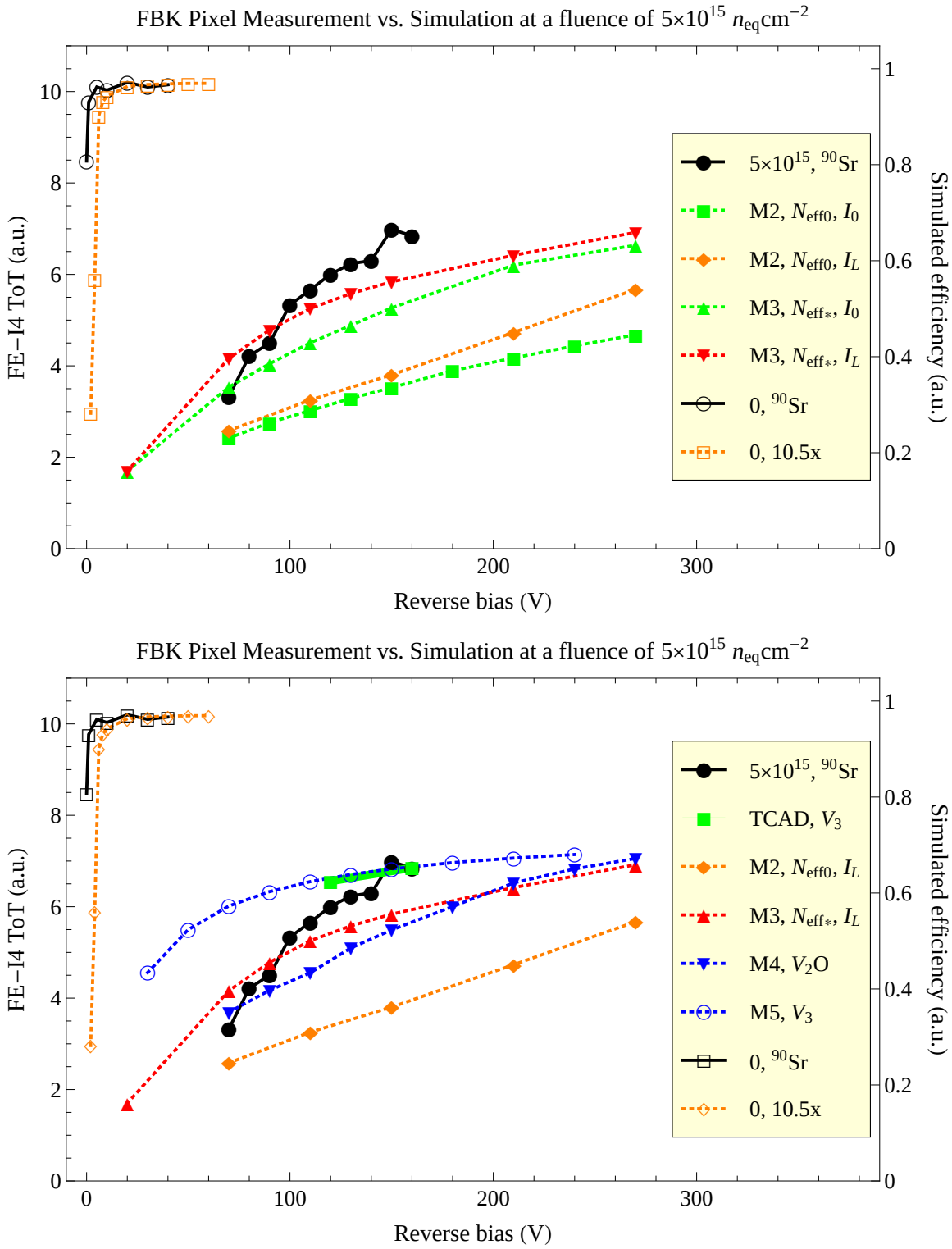


Figure 6.14: Experimental data and simulated signal efficiency of FBK 3D pixel detectors at a fluence 5×10^{15} $n_{eq}cm^{-2}$ with the quasi-Fermi level approximation (M2 and M3) and the SRH model (M4 and M5). Different leakage currents affect the depletion widths and thus the signal efficiencies (upper). Different defect models also affect N_{eff} and thus the depletion widths (lower). The solid lines are the data while the dotted lines are simulations.

Shockley-Read-Hall Model

In addition, the SRH model for the occupancy function is simulated and compared since it is conventionally used in TCAD. Figure 6.14 (bottom) (labelled "M2", "M4", "M5" and "TCAD") shows the summary of the simulation results at a fluence of 5×10^{15} $n_{eq}cm^{-2}$. The model M5 (V_3) and M4 ("V_2O") set a higher and a lower limit for the experimental data respectively. This is due to their energy levels with respect to mid-gap discussed earlier and thus they give different space charge. M4 and M5 only have good agreement at high and low voltages respectively.

The green segment (labelled "$TCAD$") uses the TCAD output generated by Povoli using the three-level defect model (with V_3 dominant) in the tracking simulation [100]. It is on top of the blue curve (labelled "M5") which uses the Kurata output with the same V_3 model for tracking. This also validates the Kurata modelling for the Shockley-Read-Hall statistics.

These three models, M3, M4 and M5, converge above 200 V where the device reaches full depletion. The signal efficiencies are 40% less compared to the unirradiated data due to large trapping at very high fluences. This is much larger than at a fluence of 2×10^{15} $n_{eq}cm^{-2}$. M3 also agrees well with the data at medium voltages and thus it is considered a good model to generate further results.

Summary

The introduction rate rescaling for deep acceptors significantly reduces the effective doping concentration and thus full depletion can be reached with reasonable bias at around 200 V. A fluence of 5×10^{15} $n_{eq}cm^{-2}$ after rescaling gives as much space charge as a fluence of about 1.5×10^{15} $n_{eq}cm^{-2}$ with the original linear relationship. This rescaling for deep acceptors is preferred by the data as one consistent model M3 applies for all voltages. The defect kinetics at high fluences may be more complicated. For instance, defect clusters may become involved in reactions for different vacancies and interstitials.

There is no experimental evidence of a reduction in the leakage current which is crucial and assumed to be linear with fluence. The inverse trapping lifetime is also assumed to be linear with fluence. Further experiments are needed to clarify the use of the linear relationship for the generation and trapping damage constant.

However, no model can perfectly predict the data as experiments always give a steeper signal-voltage dependence and this is usually due to avalanching. No significant signal gain is observed in the simulation results for 3D sensors due to the low full depletion voltage after rescaling. Although FBK sensors are a fully-penetrating design, a 2D horizontal cut across the sensor plane may not be a good approximation for the whole sensor.

6.4 CNM Pixel detector

CNM is another provider of 3D pixel detectors for the IBL using the same layout and FE-I4 readout chip. They are like CNM strip detectors to use a non-fully-penetrating design. Measurements were performed by Grinstein at Barcelona using β particles from a ^{90}Sr source. The TCAD simulation results are generated by Balbuena at Barcelona [23]. A 2D horizontal cut across the sensor plane was used for simulation as the FBK results. This is only valid for the middle part of the CNM electrodes.

6.4.1 Experimental and Simulation Results

Figure 6.15 shows a comparison of the experimental data and TCAD simulation results at fluences of 2×10^{15} and 5×10^{15} $n_{eq}cm^{-2}$. The data at a fluence of 2×10^{15} $n_{eq}cm^{-2}$ reach the simulation result at high voltages, while the simulation overestimates the signal response at a fluence of 5×10^{15} $n_{eq}cm^{-2}$.

The low fluence data reach full depletion at high voltage, while the sensor with a higher fluence breakdowns before full depletion. This suggests that the defect model used in TCAD gives less space charge than reality. This was found for FBK sensors in the previous section but the discrepancies are larger for CNM sensors. This is enhanced due to the non-fully-penetrating electrode design used by CNM and the regions around $p+$ electrodes are not depleted.

The TCAD simulation results for CNM strips sensors in Figure 6.10 and pixel detectors in Figure 6.15 at a fluence of 2×10^{15} $n_{eq}cm^{-2}$ are similar at full depletion. The former recovers the signal response by avalanching while the latter has no signal gain for both fluences. This could be due to the smaller (20%) inter-electrode distance for 3D strip sensors.

Simulation Results from Kurata

A scaling factor of 18.6 ke^- to convert from signal efficiency to charge was used for CNM sensors in this section. This is smaller than strip sensors (300 μm) since pixel detectors are thinner (230 μm). Figure 6.16 (top) shows a comparison between the quasi-Fermi level approximation (labelled "M2" and "M3") and the SRH model (labelled "M5") at a fluence of 2×10^{15} $n_{eq}cm^{-2}$ using Kurata. The data fall between the model M2 and M3 while M5 overestimates the signal response as the TCAD result in Figure 6.15. There is no significant gain shown by any model or the data.

Figure 6.16 (bottom) shows a comparison of the quasi-Fermi level approximation (labelled "M2" and "M3") and the SRH model (labelled "M4" and "M5") at a fluence of 5×10^{15} $n_{eq}cm^{-2}$ using Kurata. The data also fall between M2 and M3 or M4, while M5 largely overestimates the signal response as TCAD in Figure 6.15. There is also no

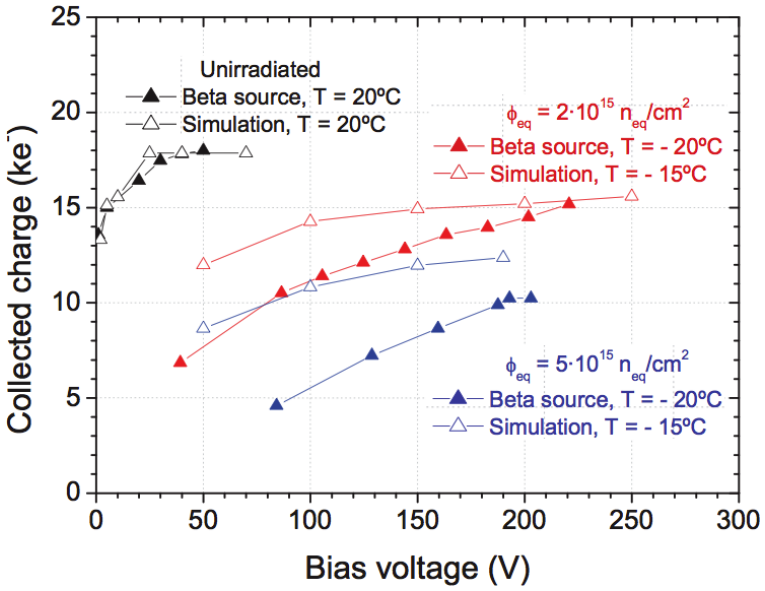


Figure 6.15: Experimental data and simulated signal efficiencies of CNM 3D pixel detectors at fluences of 2×10^{15} and 5×10^{15} $n_{eq}cm^{-2}$. Even 3D TCAD simulations cannot fully explain the data. The filled and open scatters are the data and simulation respectively [23].

significant gain observed.

The Kurata results using the V_3 defect model give good agreement with the TCAD results using the three-level model (with V_3 dominant). The signal efficiencies in both cases are 17% less compared to the unirradiated data due to trapping at a fluence of 2×10^{15} $n_{eq}cm^{-2}$. The loss is 31% and 33% due to larger trapping at a fluence of 5×10^{15} $n_{eq}cm^{-2}$ for Kurata and TCAD respectively. The small difference between the 2D Kurata and 3D TCAD simulation is due to early full depletion using the V_3 defect model so that the effect of the non-fully-through electrodes is less apparent. These also validate the Kurata modelling for the SRH statistics.

6.4.2 Summary

The large discrepancies between the CNM data and simulations using the model M5 suggest that the V_3 defect model is not appropriate. It agrees better for FBK sensors since they are easier to deplete with the fully-through electrodes.

The Kurata results using M3 and M4 agree better with the CNM data, but M3 was preferred by the FBK data. This is due to the limitation of 2D Kurata which cannot simulate the CNM electrodes and predicts larger signal efficiencies. This shows the crucial requirement to use a three-dimensional simulator.

M1 is still ruled out since it cannot reach full depletion at the highest voltage. This is important for the use of the introduction rate rescaling assumption. Further investigations on cluster formation and defect evolution are needed to understand the

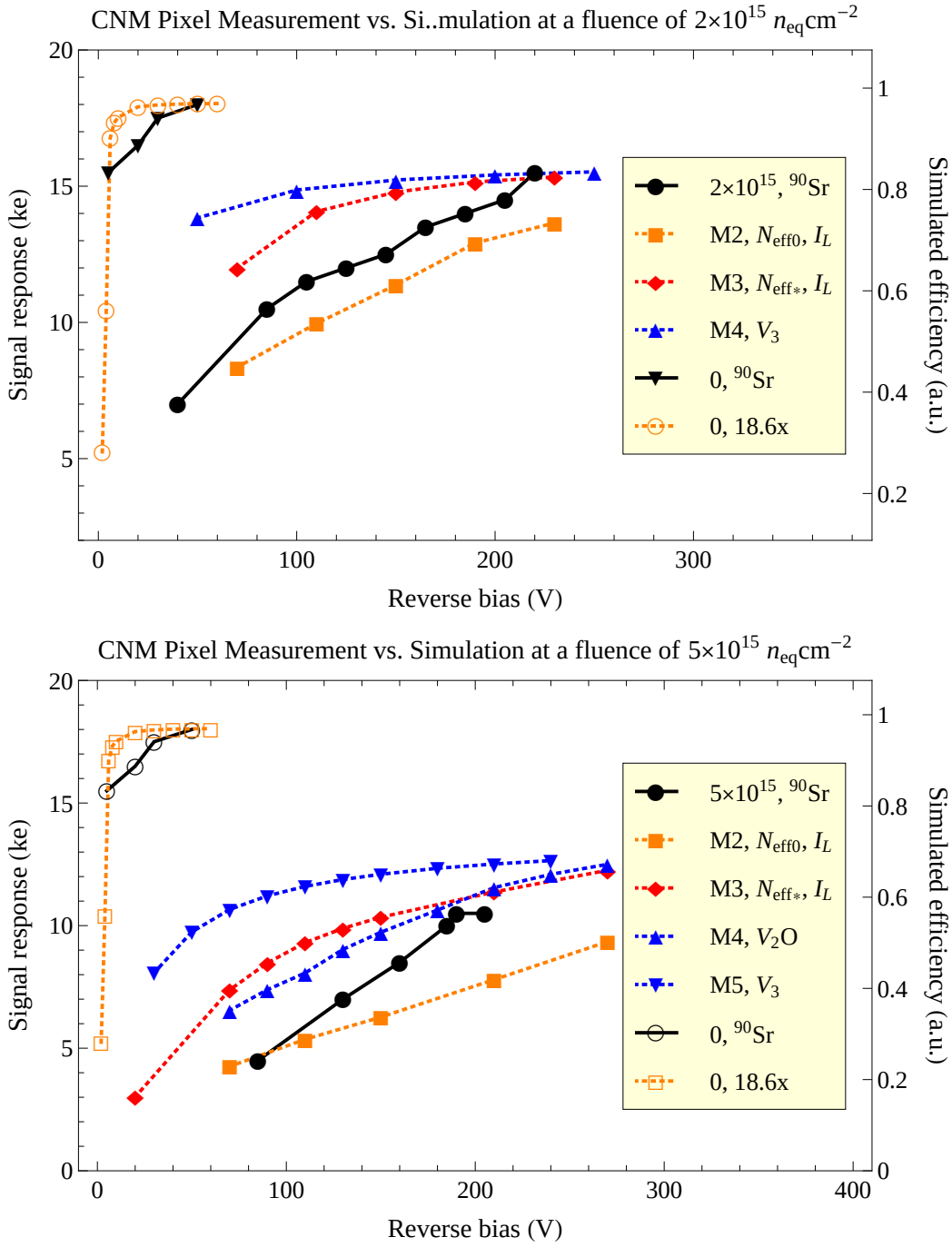


Figure 6.16: Experimental data and simulated signal efficiencies of CNM 3D pixel detectors at fluences of 2×10^{15} (upper) and 5×10^{15} $n_{eq}cm^{-2}$ (lower). The solid lines are the data while the dotted lines are simulations. Simulation results give an upper limit for the signal efficiencies due to the limitation of 2D simulation, which means the simulations are valid at depletion where the trapping lifetimes are the main factor for the signal efficiencies. Thus the model M3 give a larger prediction than the CNM data.

two regimes of the introduction rate.

6.4.3 Summary for 3D Sensors

Figure 6.17 shows the FBK and CNM data and the Kurata simulation results using the scaled quasi-Fermi level model (labelled "M3"). The CNM data have smaller signal response due to partial depletion around the $p+$ electrodes, but they have a higher breakdown voltage due to the different p-spray design to FBK. The model M3 has good agreement for the FBK data, while it is reasonable for the CNM data.

Concerning the avalanche gain, there is no experimental evidence for this in FBK and CNM pixel detectors to date. It has been observed for 3D strip sensors and may occur in planar strip sensors. 3D strip sensors have a shorter inter-electrode distance and also have the electrode tips that may be avalanching. Further experiments may be needed since strip sensors were measured using different readout chips.

Most discrepancies are due to the defect models. Some discrepancies could be due to three-dimensional geometries, such as the p-stop, p-spray and silicon dioxide layer, which affect the full depletion and breakdown voltage. Some other physics may be needed at very high fluences and voltages to provide a large electric field for avalanching.

2D slices across the sensor plane are a good approximation for FBK 3D sensors and also for CNM 3D sensors operated in full depletion. This validates the use of 2D signal efficiency maps for the 3D digitiser. Full three-dimensional simulators (TCAD) using a reasonable defect model, e.g. M3, can always improve the simulation results.

6.5 Avalanche Diodes

Avalanche photodiodes (APDs) can increase the sensitivity to detect ionisation induced current by magnifying the signal response by avalanching. It requires a special design for the doping profile to create a very high peak field. This is conventionally achieved by a $p+$-π-p-$n+$ or a $n+$-π-n-$p+$ diode, in which an additional asymmetric pn junction is used and π stands for the intrinsic collection region. This design has been applied on planar strip sensors [29] by modifying the implant of the readout strips. The same concept can be also used for 3D sensors.

The signal gain can be either from the doping profile or geometrical effects. A robust (but not necessarily large) gain is desired and thus complicated structures, e.g. avalanching from electrode tips, are not preferred. One APD doping design and one geometrical avalanching design for 3D sensors are discussed in this section. Other possible designs based on geometrical effects can be found in [36].

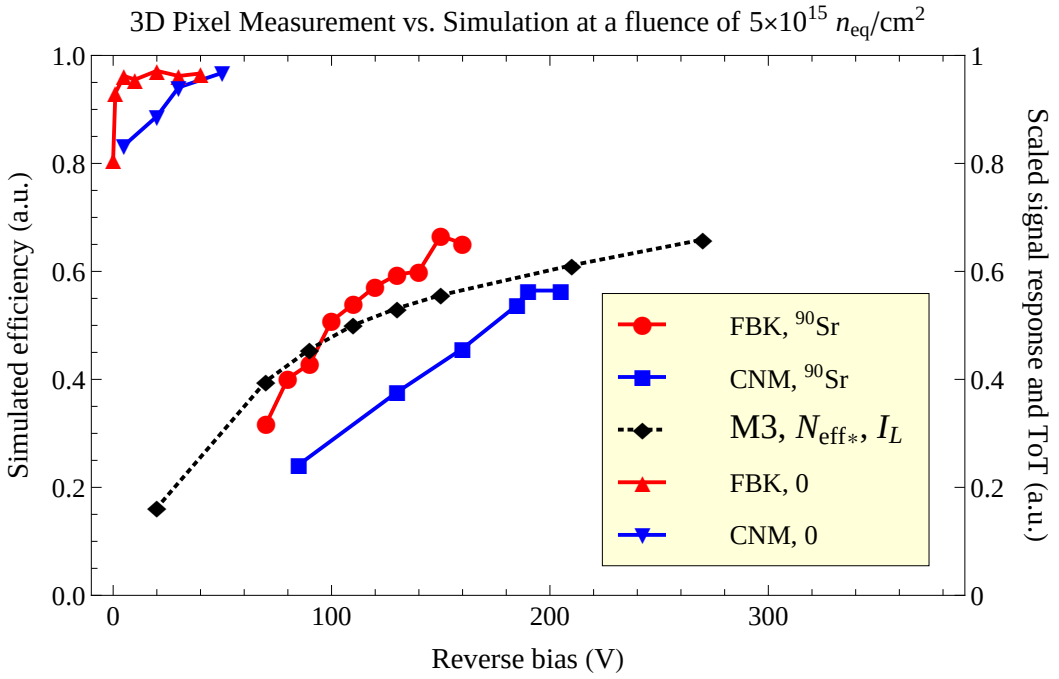


Figure 6.17: Experimental data and simulated signal efficiencies for two 3D pixel sensors at a fluence of 5×10^{15} $n_{eq}cm^{-2}$. No simulation can fully agree with the data, but Kurata with the experimental values is alternatively a good way to model silicon devices. The solid lines are the data while the dotted line is simulation.

6.5.1 3D Avalanche Diode

It is possible to combine the APD doping profile on the 3D electrodes, which is also a p+-π-p-n+ diode. Figure 6.18 shows a representative doping profile (left) and the corresponding electric field distribution (right) of a 3D avalanche diode with the APD doping profile. The cell size is set to 50×50 μm^2 which is planned for the next ATLAS upgrade of the silicon tracker system of the HL-LHC.

Selected combinations for the diffusion length and concentration of p dopants are listed in Table 6.4 and their electric fields on the inter-electrode line are shown in Figure 6.19 (upper). Once the highest bias voltage is applied, their electric field distributions tend to be similar, except for $D_7 - P_{16}$ case which has a very confined peak field.

However, there are two limitations for fabrication: once the p dopants are implanted, they form a boron glass. No further implantation, e.g. n+ electrodes with phosphorous, can be performed. This could be solved by using a n+-π-n-p+ structure or by removing the boron glass. It is difficult to control the doping profile since dopant diffusion depends on many factors. Thus it is difficult to realise a successful avalanche device using doping.

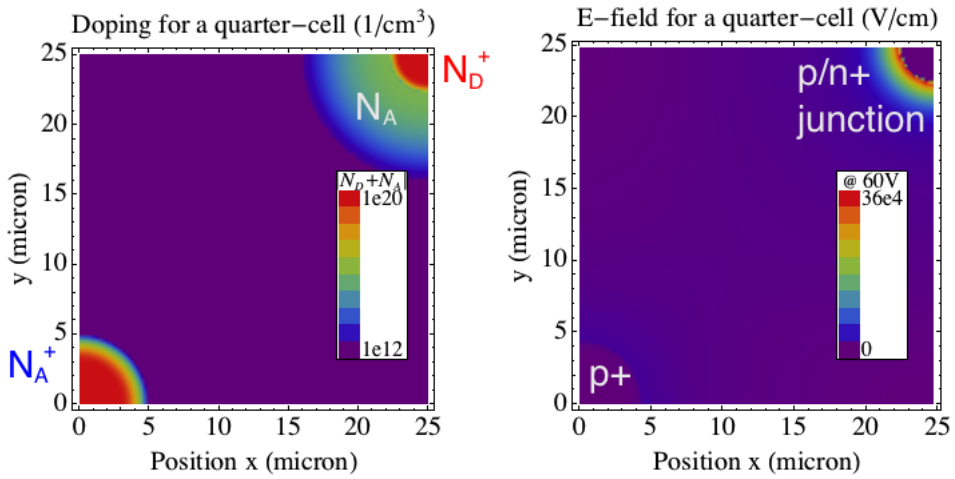


Figure 6.18: Doping profile (left) and electric field distribution (right) for a 3D avalanche diode with the APD design.

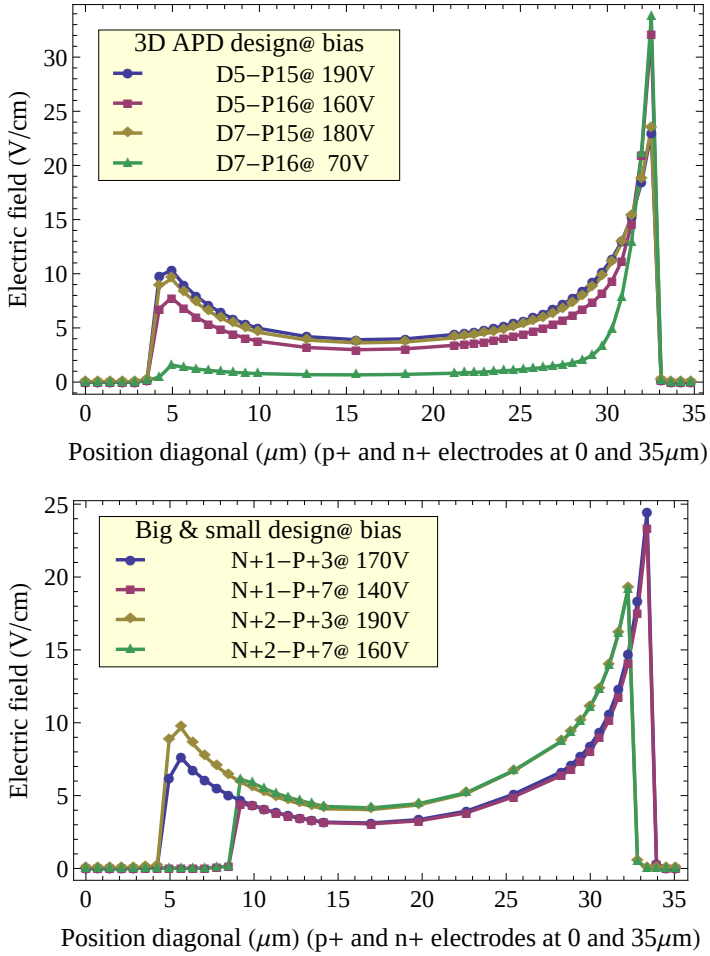


Figure 6.19: Simulation results for 3D avalanche diodes with the APD doping profile (upper) and "big and small electrodes" design (lower).

Design name	p diffusion (μm)	p concentration (cm^{-3})	Peak gain	Bias voltage (V)
$D_5 - P_{15}$	5	10^{15}	2.21	190
$D_5 - P_{16}$	5	10^{16}	15.86	160
$D_7 - P_{15}$	7	10^{15}	2.45	180
$D_7 - P_{16}$	7	10^{16}	31.17	70

Table 6.4: Simulation results for the selected combinations for the diffusion length and concentration of p dopants for 3D avalanche diodes. The radius for the $n+$ and $p+$ electrode are 2 and 3 μm respectively to preserve active area. The diffusion length for the $n+$ and $p+$ dopants are 1 and 2 μm respectively.

6.5.2 Big and Small Electrodes

One possible 2D design is to combine a big $p+$ electrode with a small $n+$ electrode. This gives a large voltage drop at the $n+$ side due to the larger curvature. Figure 6.20 shows the TCAD simulation of a 24×24 μm^2 cell. The radii are 1.5 and 3.5 μm for the $n+$ and $p+$ electrode respectively. The diffusion lengths of dopants are 1.5 (red) and 2.5 (black) μm for comparison. It is clear that the diffusion length controls the junction position. A signal gain of around 2 at 120 V is estimated. The electric field is confined at the $n+$ junction side (blue) after heavy irradiation as expected.

The TCAD results can validate the "big and small" design. However, the cell size for the TCAD simulation is difficult to realise. It was modified to 50×50 μm^2 as in the previous section and simulated using Kurata.

Selected combinations for the radii of the $p+$ and $n+$ electrode are listed in Table 6.5 and the electric field on the inter-electrode line are shown in Figure 6.19 (lower). The radius and diffusion length of the $n+$ electrode is as small as possible, while they are expected to be large for the $p+$ electrode but a large active area is also needed. Once the highest bias voltage is applied, their electric field distributions tend to be similar as above.

Due to variations in the processing, a signal gain of around 2 is expected using the impact ionisation integral. However, a 3D electrode with a radius of 1 μm corresponds to a maximum substrate thickness of less than 80 μm using the Bosch process (FBK can achieve an aspect ratio of 40). The signal gain can partly compensate the low signal response due to the thin substrate. Further design and testing are needed for 3D avalanche diodes.

The avalanche effect enhances the signal response and reduces the substrate thickness which gives less multiple scattering. This also reduces the total capacitance for 3D sensors. The $N_{+2} - P_{+3}$ case is a sensible compromise between the thickness and active area. This design will benefit from a large "active" and filled $p+$ electrode, in which carriers may escape from the polysilicon region. This will be discussed in the next chapter with different electrode processes.

Design name	$n+$ radius	$p+$ radius (μm)	Peak gain	Bias voltage (V)
$N_{+1} - P_{+3}$	1	3	3.76	170
$N_{+1} - P_{+7}$	1	7	2.55	140
$N_{+2} - P_{+3}$	3	3	1.68	190
$N_{+2} - P_{+7}$	3	7	1.67	160

Table 6.5: Simulation results for selected combinations for 3D big and small diode. The diffusion length for the $n+$ and $p+$ dopants are 1 and 3 μm respectively.

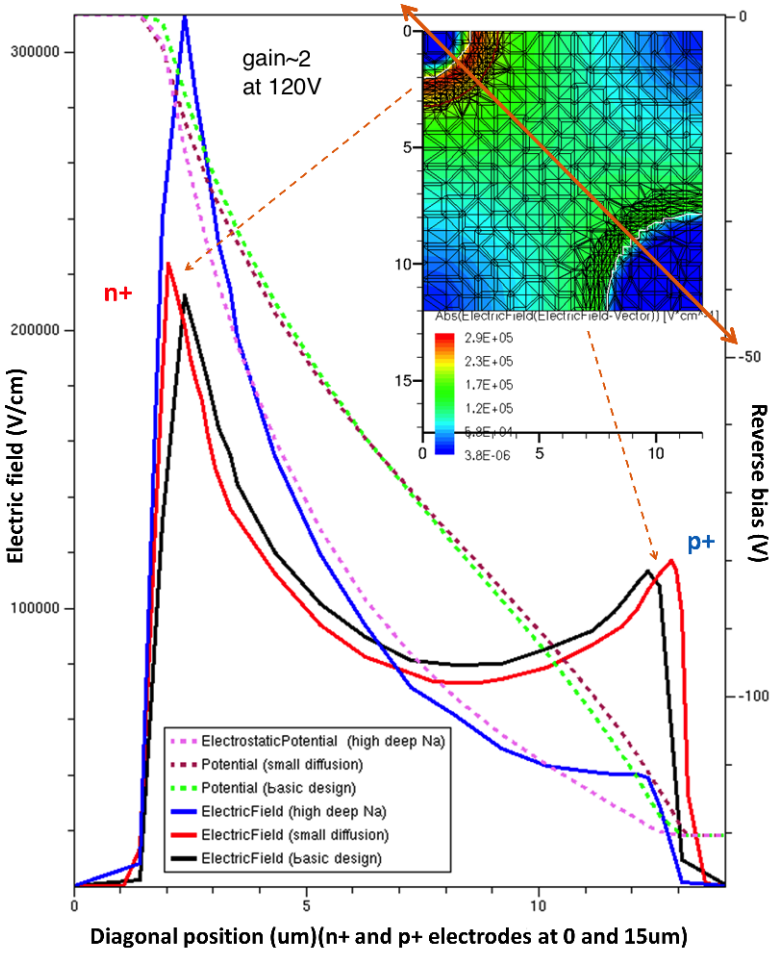


Figure 6.20: TCAD simulation results with different diffusion lengths of dopants. The inset shows the simulated quarter cell and the values on the red cut line across two electrodes are plotted here. The radii are 1.5 and 3.5 μm for the $n+$ and $p+$ electrode respectively. The black and red curve use the diffusion lengths of 1.5 and 2.5 μm respectively which change the junction positions. The blue curve shows that the electric field is very confined at the $n+$ junction side as expected after heavy irradiation at a fluence of 1×10^{16} $n_{eq}cm^{-2}$ using the SRH model. The dashed lines are the correspondent electric fields for three conditions.

Chapter 7

Test Beam Data and Simulations for 3D Sensors

Another method to characterise the detector performance is to use particle beams from accelerators or light from a synchrotron. Sensors were tested using X-ray, electron and pion beam sources located at Diamond, CERN and DESY respectively. Results from Diamond and the CERN SPS are discussed in this chapter. The DESY results are not shown because electrons, due to their light mass, suffer from multiple scattering which degrades the spatial resolution.

Two multi-gate 3D diode devices, with all cells connected together, were measured at Diamond. The aim was to study the relationship between the signal response and electrode processing. The experimental data do not show much improvement for either $n+$ and $p+$ electrodes. Simulations were performed to estimate the carrier trapping lifetime in the electrodes.

Several sets of pixel detectors using planar and 3D technology for the IBL upgrade were measured at DESY and CERN SPS using the EUDET telescope. The data were analysed using track reconstruction to obtain the signal response in terms of ToT and charge sharing in terms of cluster size. Simulations were performed to estimate the signal efficiency and charge sharing for 3D sensors. Simulations were made for similar conditions to the experiments and compared with the data [9].

7.1 Diamond Test Beam

The Diamond Light Source is a national synchrotron facility located in Oxfordshire in the UK. Diamond can provide intense light beams with wavelengths from infra-red to X-ray. Two measurements were performed in 2011 and 2012 at Beamline 16 at Diamond. This beamline is used to test new detector technology by providing both white and monochromatic X-rays [12]. The beam radius is supposed to be around 2 to

3 μm at 15 keV.

The sample was mounted inside a shielded metal box with an opening for the X-ray beam. The opening is covered by a thin aluminium foil to ensure the box is electromagnetically shielded. The light source is a focused and fixed beam, while the box is movable to measure the signal response over the whole sensor. The box was mounted on two linear precision motors which moved vertically and horizontally. The motor repeatability was better than a micrometer.

The signal response was read out directly from the electrodes and was magnified with a current amplifier. The gain, from current to voltage, ranges from 10^6, 10^7 to 10^8 VA^{-1}. There is also an upper limit for the current to protect the electronics. The gain was set to 10^7 VA^{-1} for both measurements. There was no triggering and the signal was continuously read out and averaged over time. The stepper motor moved the box every 1 to 2 seconds .

7.1.1 Experimental Result of a Stanford 3D MBC Silicon Sensor

Figure 7.1 shows a microscope image of the Stanford 3D MBC sensor (abbreviated as "MBC") tested in 2011. This device is a p-in-n design and was biased at +30 V. A 0.5 mm^2 area (at the right-lower corner in the photo) was measured to cover almost nine cells and the cell size is 150×150 μm^2. The step size was set to 5 μm which allows the area of interest to be scanned in a reasonable time.

Figure 7.2 shows the experimental data using the X-ray light. This resolution is not sufficient to distinguish the electrode circumference consistently. The electrode efficiencies are lower than 5% (nearly inactive) which suggests the trapping lifetimes are short. This is an old prototype device with a n-type substrate. There is a small accumulation of charge which increases with the scanning time (a). There is no obvious accumulation for a short scan (α).

7.1.2 Experimental Result for a Stanford 3D 2E Silicon Sensor

Figure 7.3 shows a microscope image of the Stanford 3D 2E sensor (abbreviated as "2E") tested in 2012. This device is a n-in-p design and was biased at $-20V$. A 0.5×0.25 mm^2 area (at the right-upper corner in the photo) was measured to cover six cells where the cell size is 125×50 μm^2 (the IBL layout). The inset figure (at the right-lower corner) shows the magnified image of a floating $p+$ guard strip, $n+$ and $p+$ electrode strips. Two step sizes, 15 and 4 μm, are used to have a quick and thorough scan. The floating $p+$ guard strip labelled in the inset is invisible in the data.

Figure 7.4 shows the experimental data using the X-ray light. The data were plotted in 3D to give a better idea of the signal magnitude. There is an unexpected accumulation of background signal (leakage current) for the long run with a 4 μm

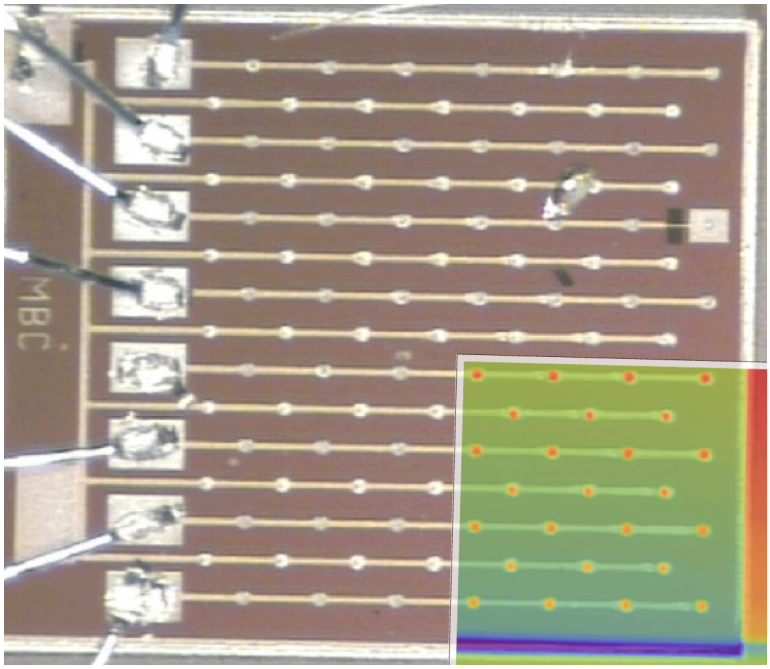


Figure 7.1: Microscope image of Stanford 3D sensor. The right-lower corner was measured. The uppermost line is one of the ohmic strips which connect $p+$ electrodes together. It is biased at the same voltage as the sensor edge. The active area (brown region) is around $1.2 \times 1.2\ mm^2$.

step. There is also a small accumulation of background signal for the short run. The electrode radius is also much larger than expected.

7.1.3 Accumulation of Background Signal

The varying background rate is thought to be related to the beam intensity, scanning time and bias voltage. Unfortunately the beam intensity was not recorded in 2011 and the beam radii were not recorded in both experiments. The signal response, total minus background current, is around 0.2 and 0.15 V (amplified from A to V) in 2012 and 2011 respectively. The beam intensity in 2011 should be smaller than the nominal 2012 one according to the signal response from a similar substrate thickness. The 2012 beam intensity was measured by an ion chamber and was 7.5×10^{11} $(photons)s^{-1}$ per 250 mA which is the electron current in the Diamond synchrotron.

The background signal has a strong dependence on the scanning time. For the 2011 data, the corresponding background signal increases from 0.04 to 0.1 V after a long scan (around 20000 points and 2 s per point). There is no accumulation for the short scan (around 1000 points) as shown in the inset of Figure 7.2 (α). For the 2012 data, the corresponding background signal rises from around 1.23 to 1.3 V which is around 5% after a long scan (around 10000 points), while it increases just 1% after a short scan (around 1000 points).

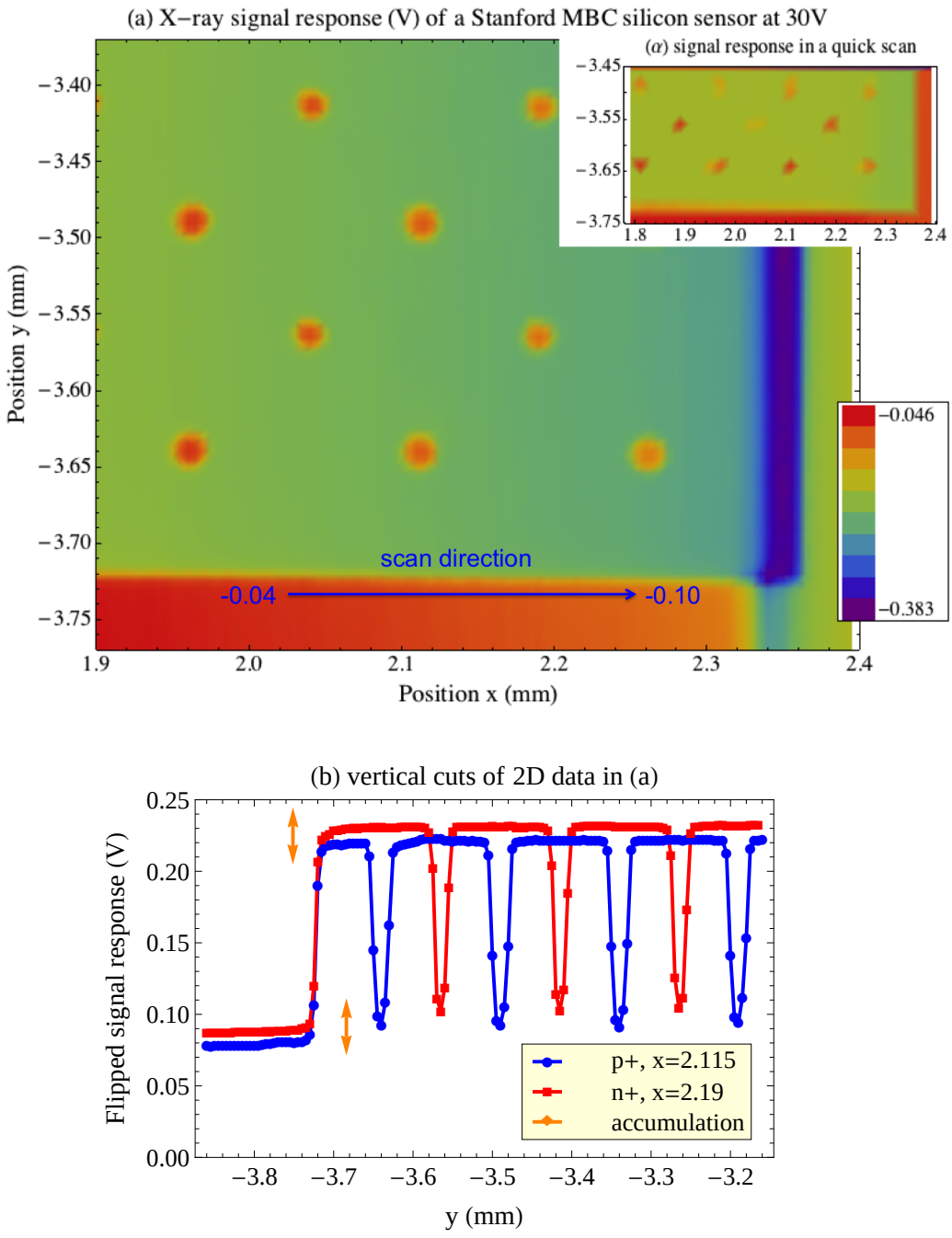


Figure 7.2: Signal response map (a) and vertical cuts (c) of the Stanford MBC sensor. Original X-ray signal response were taken at a step size of 5 (upper) and 20 (inset) μm. The vertical cuts are the flipped X-ray signal response (c) through $p+$ and $n+$ electrodes. An accumulation around 0.06 V was observed after a long run (a) while there is no accumulation for the short one which is shown in the inset (α). The electrode radius is around 10 μm.

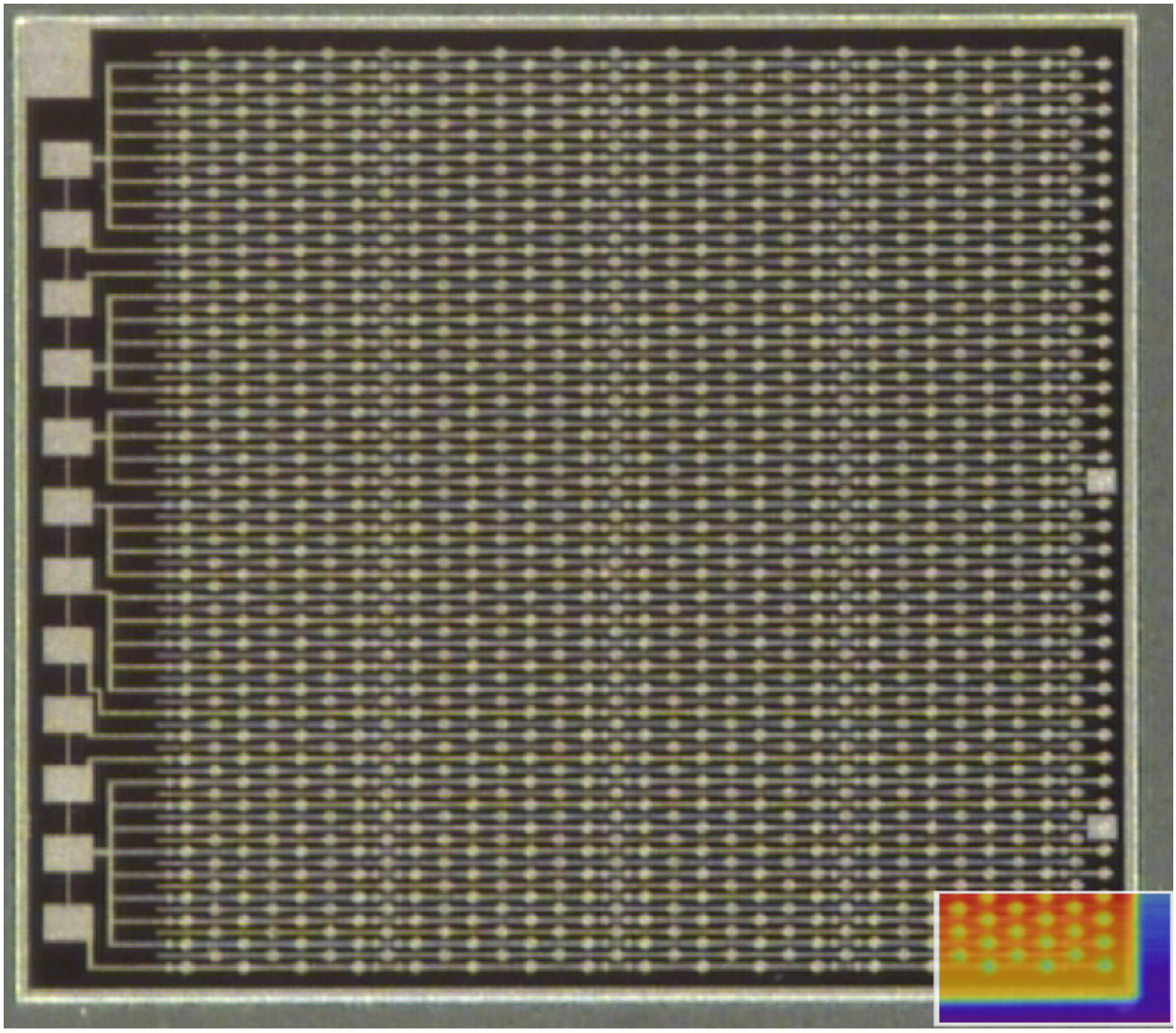


Figure 7.3: Microscope image of the Stanford 3D 2E sensor. The right-lower corner was measured. The right-upper corner shows a floating $p+$ guard strip, a $n+$ and a $p+$ electrode strips from the top. This is a test device with the same cell size to the ATLAS FE-I4 layout. The active area (black region) is around $2 \times 2\ mm^2$.

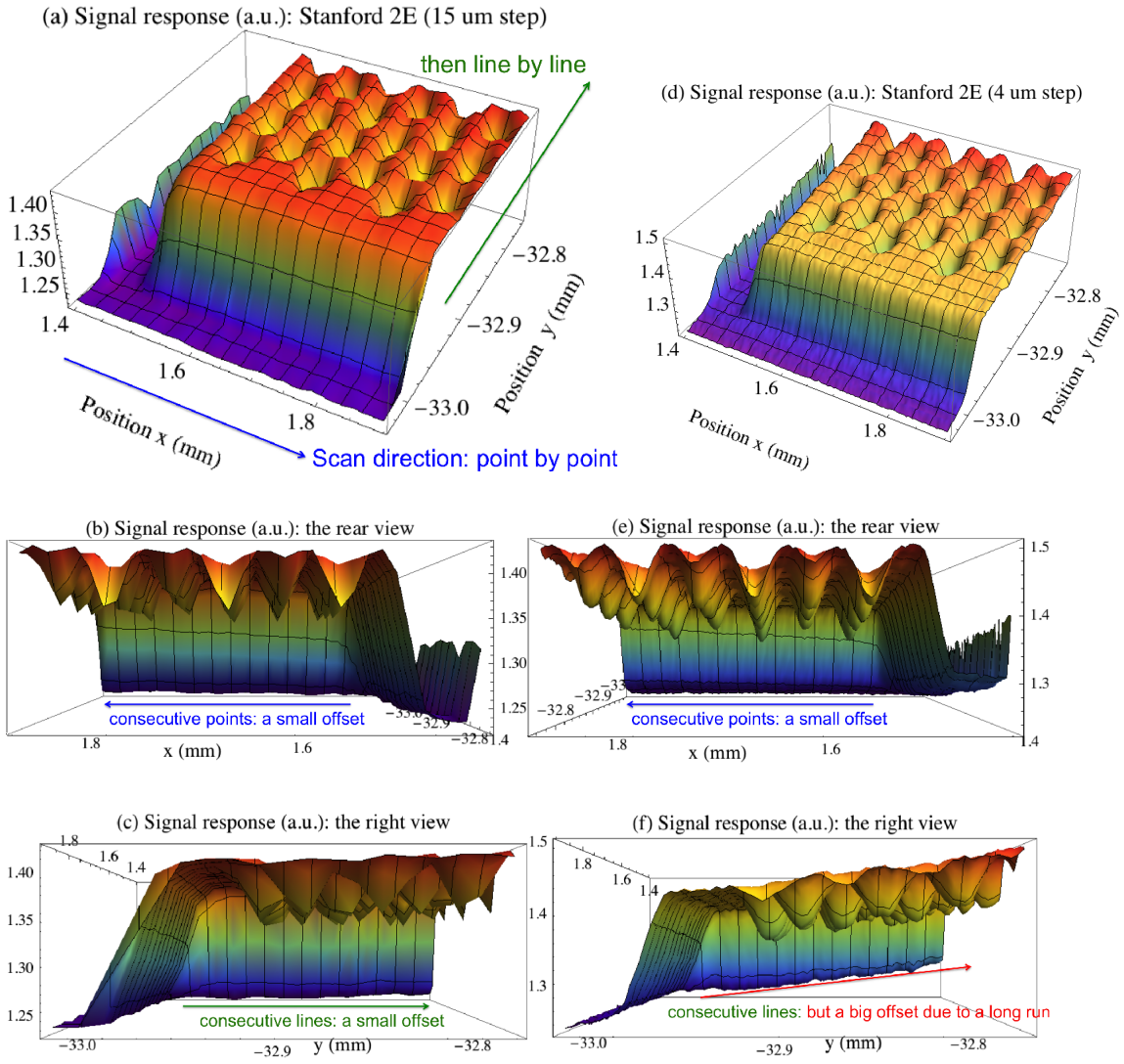


Figure 7.4: Three-dimensional signal response maps of the Stanford 3D 2E sensor in 2012. The left (a, b and c) and right (d, e and f) columns were taken with steps of 15 and 4 μm respectively. This shows that the background signal increases for the 4 μm step measurement due to the long scanning time. The upper row (a and d) shows an overall view of the both data sets. The middle row (b and e) shows the rear view and the baseline represents the first line scan in the x direction where the offset is small. The bottom row (c and f) shows the edge view and the baseline represents one end of all line scans in the y direction where an offset starts to accumulate. This offset can be described and subtracted by a linear function of x and y positions (or time).

The background signal is also affected by the bias voltage which is related to the leakage current. One edge of the same Stanford 2E device was scanned at a step size of 10 μm under a bias from -10 to -40 V. Figure 7.5 shows that the background signal also increases with the bias voltage, but the signal response remains around 0.2 V.

Figure 7.6 shows the IV measurement of this Stanford 2E device at room temperature. There is an extra source of leakage current at low voltages. The background signal measurements at the Diamond test beam are also plotted in the same diagram. The voltage scan at the edge shows a similar trend with the IV curve, while the long scan at the corner shows an increase due to accumulation.

The 5% increase is thought to be a thermal generation due to heating by the X-ray beam. The ratio of leakage currents from T_1 to T_2 is given as [110]

$$\frac{I_{leak}(T_2)}{I_{leak}(T_1)} = (\frac{T_2}{T_1})^2 e^{-\frac{Eg}{2}(\frac{T_1-T_2}{T_1T_2})} \ . \tag{7.1}$$

It requires just 0.7 K to generate a 5% increase. The temperature increase can be characterised by the heat equation given as

$$\nabla \cdot (k\nabla T) + Q - hT = c\rho\frac{\partial T}{\partial t} \tag{7.2}$$

where k is the thermal conductivity, h is the convection heat transfer coefficient, c is the specific heat capacity and ρ is the density. The thermal conductivity of silicon substrate, around 149 $Wm^{-1}K^{-1}$, is relatively high and makes it a good thermal conductor. The sensor was attached to a PCB board and then mounted on a aluminium box. The surrounding air is assumed to be uniform at room temperature. As long as the input heat Q from the X-ray beam is greater than the outgoing heat by convection, there will be an increase in the device temperature. This could be the reason for the varying background current.

7.1.4 Correction to the Signal Response

The electrode efficiencies are around 60% which may suggest that the trapping lifetime for the MBC device is longer the 2E device. The 2012 data need to be corrected due to the varying background. It is assumed that the accumulation rate is constant so that one can subtract the background for the long scan.

The original and corrected signal response are shown in Figure 7.7 (a) and (b). There is a time gap between each horizontal line in Figure 7.7 (c) due to the small step size. The shift in background signal was corrected in Figure 7.7 (d) and it shows that the real signal response does not increase with the background. The efficiency for the $p+$ electrode is slightly larger than the $n+$ one which are around 61% and 56%

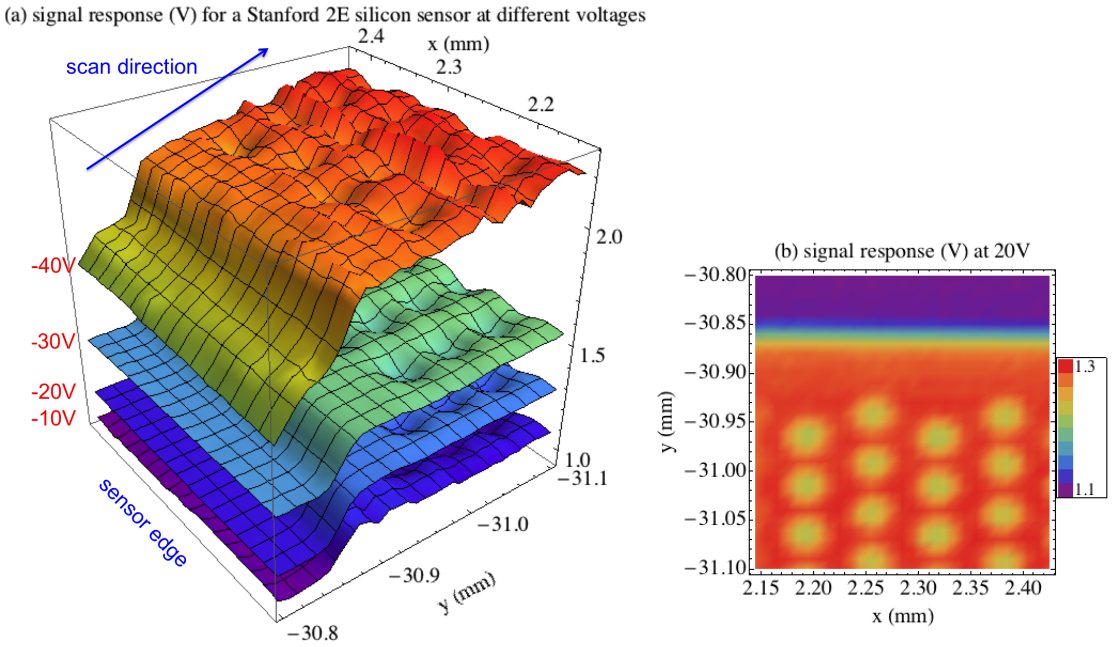


Figure 7.5: Signal response of the Stanford 3D 2E sensor at different bias voltages plotted in 3D (left) and at -20 V in 2D (right). One edge was scanned at a step size of 10 μm. The background signal increases with the bias and has a similar trend to the leakage current measured in the lab. The IV measurement of this Stanford 2E device at room temperature is performed by Haughton at Manchester

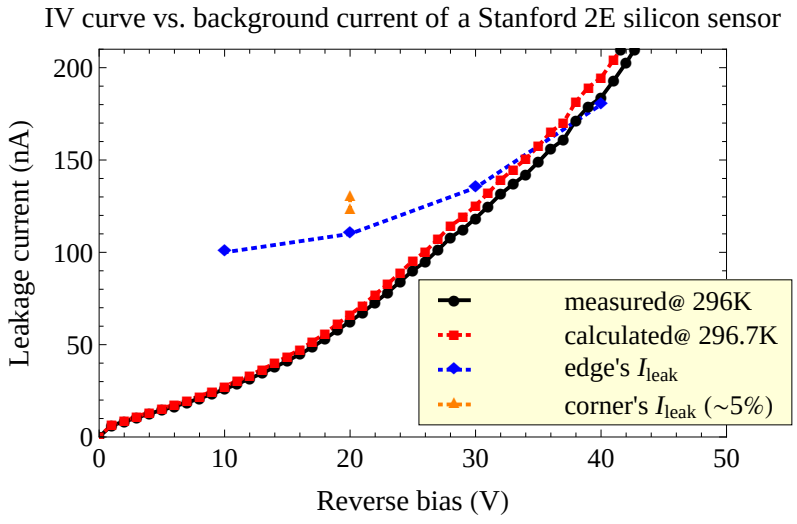


Figure 7.6: The IV curve and background signal of the Stanford 3D 2E sensor. An IV curve was measured at room temperature and another curve at plus 0.7 K is calculated from the data. The background signal at the edge and corner of this device are also plotted which represent the accumulation after a short and long scan respectively. At low voltages, there is a non-device related leakage current which decreases with the bias (the difference between the blue and black curve) and the reason to that is not understood for the moment.

respectively. This may be an indication of active electrodes in which carriers can escape from the boundary.

7.1.5 Process and Signal Efficiencies of the Electrodes

The FBK and CNM 3D sensors have unfilled (empty) electrodes which are drilled by deep reactive ion etching (DRIE). The Stanford and SINTEF 3D sensors have electrodes filled with polycrystalline silicon which is deposited through low pressure chemical vapour deposition (LPCVD). Silane is used to provide silicon atoms as

$$SiH_4 \overset{900K}{\longrightarrow} Si + 2H_2 \ . \tag{7.3}$$

and phosphorus pentoxide ($n+$) and boron trioxide ($p+$) are used to provide dopants as

$$2P_2O_5 + 5Si \longrightarrow 4P + 5SiO_2 \quad \text{and} \quad 2B_2O_3 + 3Si \longrightarrow 4B + 3SiO_2 \ . \tag{7.4}$$

The deposited polysilicon surface is less than 1 μm in thickness.

Annealing follows the deposition and causes dopants to diffuse into the surrounding single-crystalline silicon bulk and form junctions. This process will also form a highly doped glass on the wafer surface and within the electrodes. After removing the glass (which may not be completely removed), the electrodes are also filled by depositing doped gases which are phosphoryl chloride ($POCL_3$, for $n+$) and boron tribromide ($BBr_3 + O_2$, for $p+$).

However, this will result in a high content oxygen and may form defect clusters which trap carriers. Oxygen precipitates, or bulk micro defects (BDM) as defined in the figure, are fixed positive charge clusters which deplete the bulk locally by a certain radius. Porrini assumes that all oxygen forms precipitate clusters uniformly as shown in Figure 7.8 (left) [99]. These clusters will bend the local band and trap carriers. The trapping lifetime is inversely proportional to the concentration and squared radius (effective cross section). Porrini found that the experimental lifetimes agreed with the model as shown in Figure 7.8 (right).

The aim was to design sensitive electrodes and thus an oxygen-free process was implemented. The Diamond beam experiment can test whether different processing improves the electrode efficiency. The doping gases are replaced with (oxygen-free) phosphine ($n+$) and diborane ($p+$) as

$$4PH_3 + 3Si \longrightarrow 4P + 3SiH_4 \quad \text{and} \quad 2B_2H_6 + 3Si \longrightarrow 2B + 3SiH_4 \ . \tag{7.5}$$

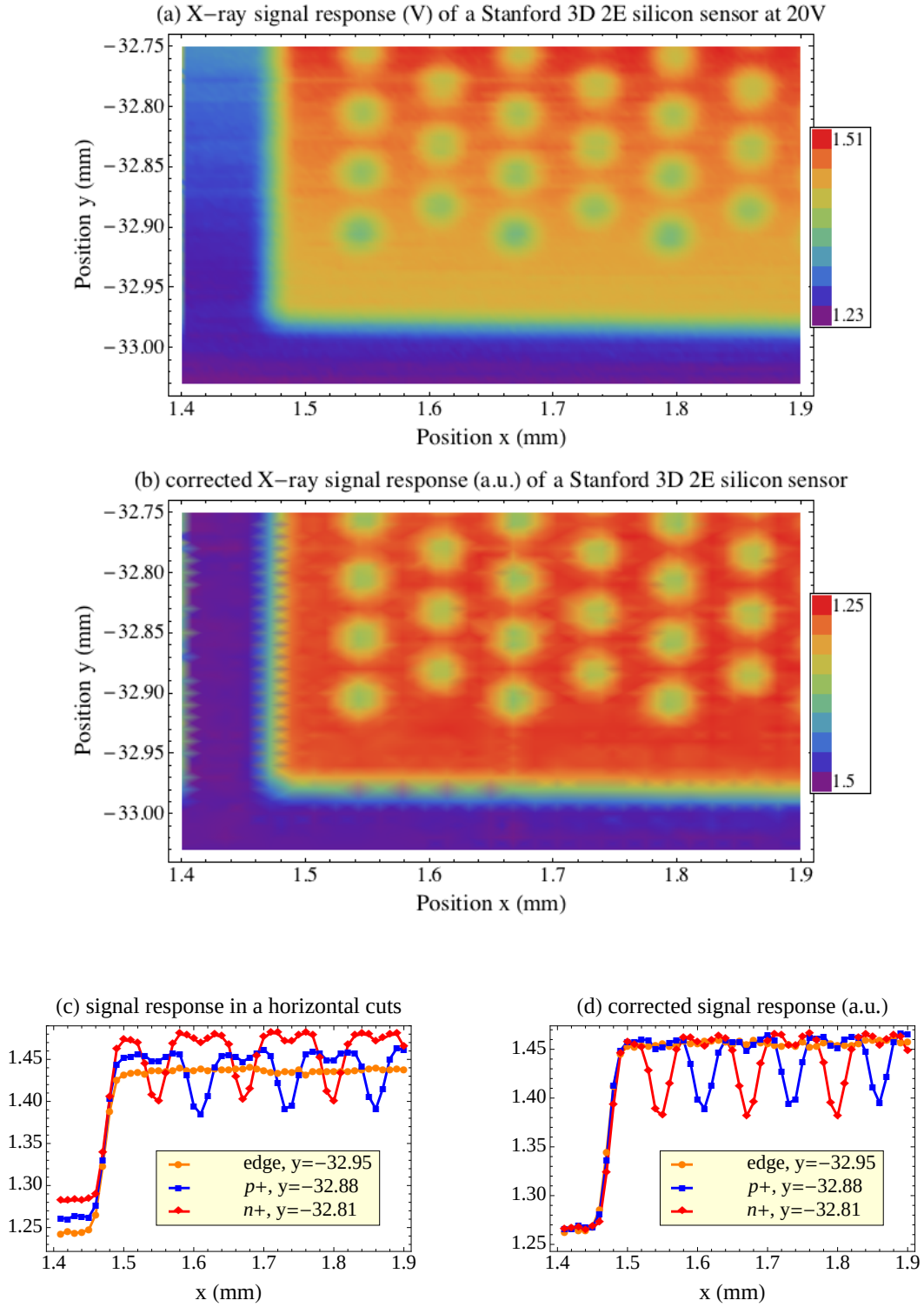


Figure 7.7: Original (upper) and corrected (middle) signal response maps of the Stanford 3D 2E sensor. Horizontal cuts (lower) on (a) and (b) through the edge, $n+$ and $p+$ electrodes show a background accumulation in (c) and an identical magnitude of the signal response after subtraction in (d).

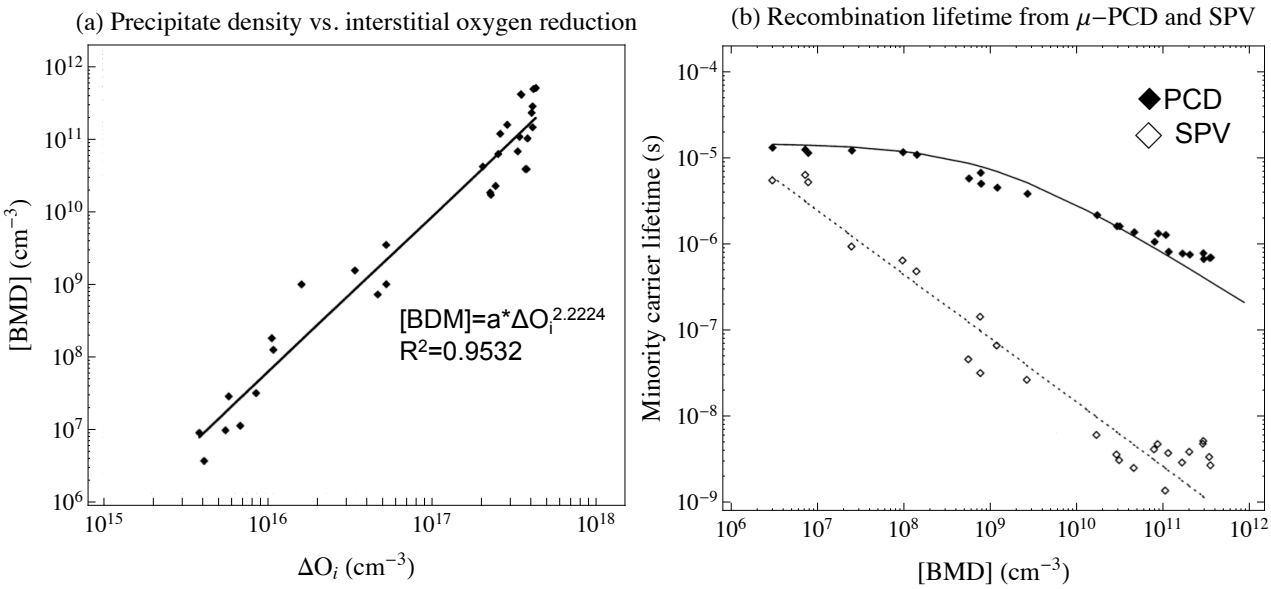


Figure 7.8: The relationship between the recombination carrier lifetime and the oxygen precipitates. (a) The interstitial oxygen reduction increases linearly with the oxygen precipitates. (b) The minority carrier lifetime decreases with the concentration of oxygen precipitates [99] μ-PCD technique operates at a high injection level, while it is low for SPC.

7.1.6 Simulation of Lifetime in the Electrodes

The simulation aims to reproduce the experimental data using the Monte Carlo method and the tracking simulation in Chapter 5. There is no electric field within the electrodes and thus diffusion (time step) and trapping lifetimes (escaping probability) are crucial for the electrode efficiencies.

The X-ray beam size is the first parameter to determine since it affects the response distribution. The sensors edges are assumed to have an abrupt difference in signal efficiency. The signal response of a scan across the edge is a convolution of a Gaussian distribution[1] (X-ray) and a step function (active sensor edge). This can be calculated analytically or using the Monte Carlo method.

Figure 7.9 shows convolutions of a step function and Gaussian distributions with different standard deviations for the Stanford MBC and 2E sensor. For the 2011 data, the beam size (standard deviation σ) is fitted to be 3 μm (red) and the 5 μm curve (blue) is an exaggeration. For the 2012 data, $\sigma = 3$ μm (blue) cannot explain the wide edge measured and $\sigma = 11$ μm (red) is needed to fit the data. This beam size was much larger than expected. The beam was set up by a Diamond beamline scientist

[1]Electron-hole pairs are generated according to a 2D normal distribution to model the Gaussian-shaped X-ray beam.

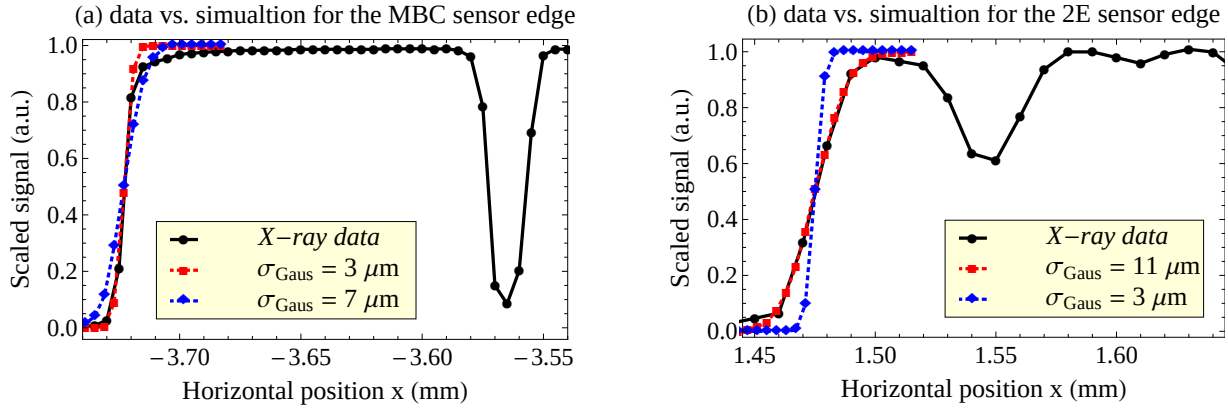


Figure 7.9: Convolutions of a step function and Gaussian distributions with different standard deviation for the Stanford MBC and 2E sensor. The beam size of the Stanford MBC and 2E sensor are 3 and 11 μm respectively according to the edge fit.

and its profile was checked by a gold wire detector. But it was not checked by the user independently after a long time of use.

The MBC Data in 2011

By using the beam size (3 μm) from the edge fit, the response of electrodes can be estimated. The $n+$ and $p+$ electrode are set to be inactive (empty) initially to estimate the electrode radius using the Monte Carlo method. Data were taken at a certain step size, the simulation also has a 4 μm spacing between points. One point was set in the centre of electrode as an ideal alignment. This is a reasonable assumption since the step size is much smaller than the electrode size and the variation of beam coverage between the neighbouring points is small.

Figure 7.10 shows that the simulation results of the electrode radii of 7 and 8 μm for the $n+$ and $p+$ electrode respectively agree well with the data. With a fine tuning for the beam size (4 μm), the electrode radii of 9 and 10 μm agree better with the data for the $n+$ and $p+$ electrode respectively. These are reasonable values since the designed electrode radius is around 5 μm and the junction usually forms a few microns further due to diffusion of the dopants.

The 2E Data in 2012

By using the beam size (11 μm) from the edge fit, the response of inactive electrodes can be estimated. Figure 7.11 shows the simulation results of the electrode radii of 7, 11 and 15 μm for the $n+$ and $p+$ electrode respectively. The result with a 7 μm radius

gives a large signal efficiency than the data due to the large beam compared to the electrode which is excluded. The result with a 11 μm radius agrees well the data (with a tunable range). The result with a 15 μm radius shows that the electrodes have to be active to explain the data.

The electrode radius of around 11 μm is close to the value given by Hasi who fabricated these batch of devices at Stanford [52]. The apparent electrode efficiency is mainly due to the large beam at this electrode radius. One can infer that the beam size is large from the signal drop at the $p+$ ($n+$) electrode position on the $n+$ ($p+$) electrode cut. A large beam covers some electrode regions whilst a small beam is fully in the bulk region. However, both the beam and electrode radius cannot be independently fitted since the data points were not exactly at the centre of the electrodes.

To allow the possibility of active electrodes, some simulations were performed for a 15 μm electrode radius and polysilicon with a longer trapping lifetime. The carrier lifetime in electrodes is discussed in the next section.

Active Electrodes

The electrodes are deposited polysilicon and have grains of different sizes. The lattice is mismatched between the grains due to their different orientation. Depending on the position of electron-hole pair generation, carriers generated in a grain have to transit three domains: grains, grain boundaries and junctions between poly and single crystalline silicon. It is complicated to model all possible arrangements of grains of different sizes. Instead, one representative trapping lifetime is used in the simulation to model different lifetimes in these three domains in the tracking simulation.

Figure 7.12 shows the simulation results for the electrode radius of 15 μm and polysilicon with different trapping lifetime for the Stanford 2E sensor. The trapping lifetimes that $\tau_p \approx 20$ and $\tau_n \approx 10$ ns agree better with the data for holes and electrons respectively. This is due to the higher mobility for electrons. This trapping lifetime is large compared to a highly doped silicon estimated by Equation 4.33. However, it is unlikely that polysilicon has a longer trapping lifetime than single crystalline silicon with the same doping concentration. The response shape is also narrower and less similar to the data.

There were also early data for Stanford MBC sensors taken by Hasi [53] which showed signal efficiencies of 57% and and 34% for the $n+$ and $p+$ electrodes respectively. These data need to be reanalysed to check the beam size and electrode radius. One can claim to obtain active electrodes only if the beam is smaller than the electrode.

Tilting of the device is possible since it shifts the beam by around 5 μm between the top and bottom surface for each degree of the incident angle. However, this will result in a measured elliptical shape for the electrodes while this is not obvious in Figure 7.7

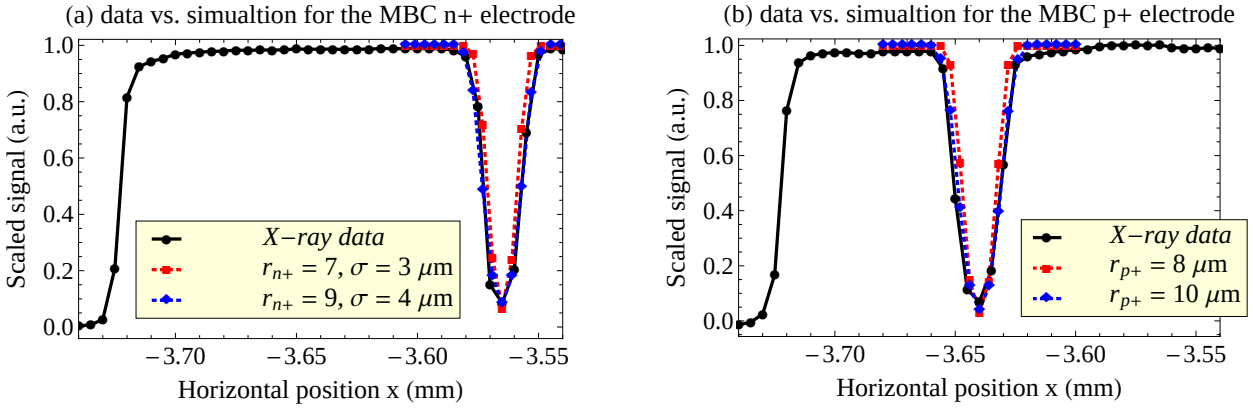


Figure 7.10: Simulation results of the electrode radii of 7 and 8 μm for the $n+$ and $p+$ electrode for the Stanford MBC sensor.

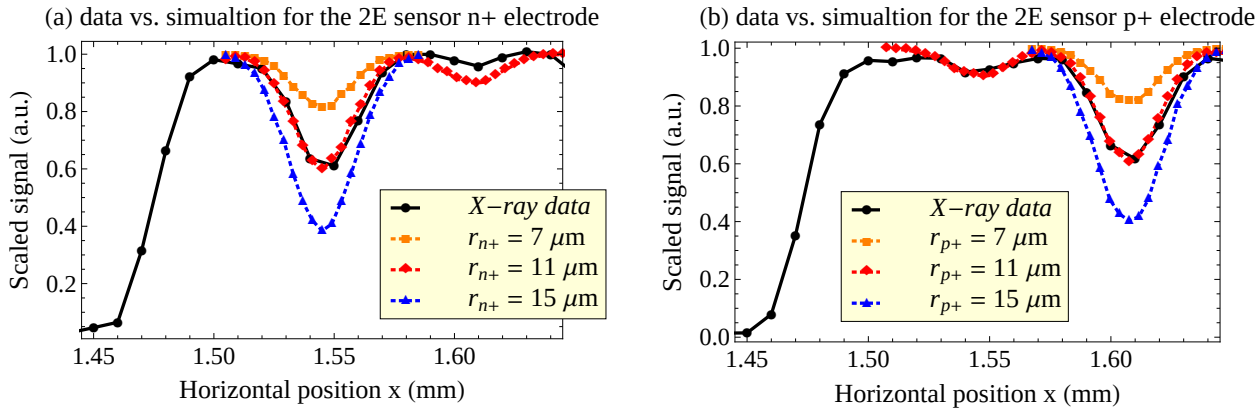


Figure 7.11: Simulation results of different electrode radii for the $n+$ and $p+$ electrode for the Stanford 2E sensor.

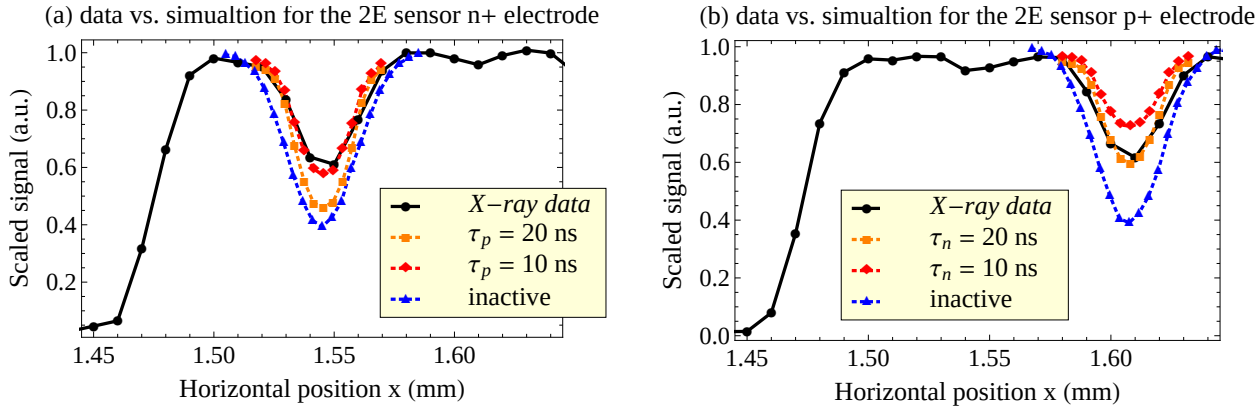


Figure 7.12: Simulation results for the electrode radius of 15 μm and polysilicon with different longer trapping lifetime for the Stanford 2E sensor.

(a) or (b) and thus only a very small angle is possible.

The Stanford MBC and 2E sensor have roughly the same thickness. The latter has 33% more in signal response which suggests that the beam flux may be larger in 2012. This may explain the larger accumulation in 2012 since the saturation temperature for heating is controlled by the input thermal energy (flux).

To summarise, the oxygen-free process does not appear to make a significant improvement in the signal response for the $n+$ or $p+$ electrodes. The large signal response measured for the Stanford 2E device may be due to a broader X-ray beam than thought. The large electrode radius fitted by simulation may be caused by a poor aspect ratio of 10 for DRIE which results in a 15 μm electrode radius for a 300 μm thick substrate. Further experiments are needed to clarify the effect of the oxygen-free process.

7.1.7 A Lesson Learnt from the Diamond Test Beam

The beam profile and flux should be monitored before any measurement since they alter the signal response and underlying background. The electrode radius should also be recorded during fabrication since it determines the diffusion for carriers in the electrodes. A reference sample from the previous beam time is needed for a quick check on-line. An appropriate flux should be used to reduce the underlying background current, but it also reduces the signal response. A small beam radius and step size are always preferred to understand the signal efficiency within the electrode.

7.2 Corner Scan at CERN

The beam test data at DESY and CERN were taken by the high resolution EUDET telescope. It has four scintillators, two in front and two at the back, which are used for triggering. Six Monolithic Active Pixel Sensors (MAPS) are in a row to provide a resolution less than 3 μm [104]. This enables detailed investigations of the sensor performance. The device under test (DUT) was placed in the middle in a plastic box to keep the sensors inside at low temperature. Figure 7.13 (left) shows the experimental set-up of scintillators, MAPs and DUT's used in the test beam.

The MAPS is a CMOS sensor called MIMOSA which has one million pixels in a 2 cm^2 chip and each pixel is around 17 μm^2 [49]. Figure 7.13 (right) shows the performance of MIMOSA. Both planar and 3D sensors bump-bonded with the FE-I4 readout chip were measured using the EUDET telescope. An unirradiated planar sensor is usually used as the reference for alignment and track reconstruction. The sensor and readout assembly was mounted on a USBPix board to record data. Data were reconstructed with the TBMon software for hit efficiencies, track residuals, charge collection and charge sharing [106].

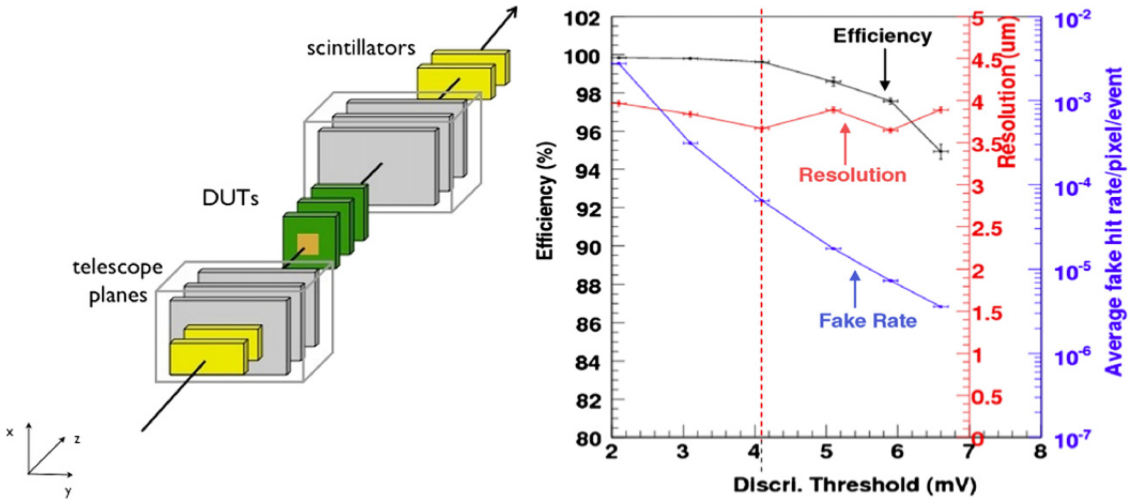


Figure 7.13: A schematic of the EUDET telescope (left) and the performance of MIMOSA (right). The DUT is mounted in the middle. Three MAPS are placed on each side of the DUT for track reconstruction. Two scintillators are placed at the end of each arm for triggering. The fake rate decreases with the threshold, but the sensor efficiency also decays.

Pion and electron beams have an area of about 1 cm^2 which can cover around a quarter of the sensor area. The signal response of the sensor edge was extracted from the test beam data which is used to examine the performance of the guard structures at the edge. Dangling bonds at the cutting surface will induce unwanted leakage current which may cause an early breakdown or discharge. Different guard structures are used to isolate and confine the electric field lines from the edge pixels. Some field lines cross the geometrical cell boundaries by design and their effects were tested by the edge signal response.

7.2.1 Experimental FBK Corner Result

The slim edge fence design is used for FBK sensors which consists of a few rows of ohmic columns ($p+$ electrodes). They can reduce the electric field gradually and prevent a lateral depletion region from reaching the cut surface [22]. Figure 7.14 shows the experimental data of the edge pixel. Two sensors irradiated to a fluence of 2×10^{15} and 5×10^{15} $n_{eq}cm^{-2}$ were measured at 60 and 140 V. The higher biased sensor (lower) shows a wider active area since the depletion region extends further. The $p+$ guard electrodes confine the carrier movement and thermal diffusion occurs at the edge around the depletion region. Note that the data were taken with the sample rotated by 15° (angle) referenced to the long pixel line. This does not affect the diffusion width in the horizontal direction (long pixel).

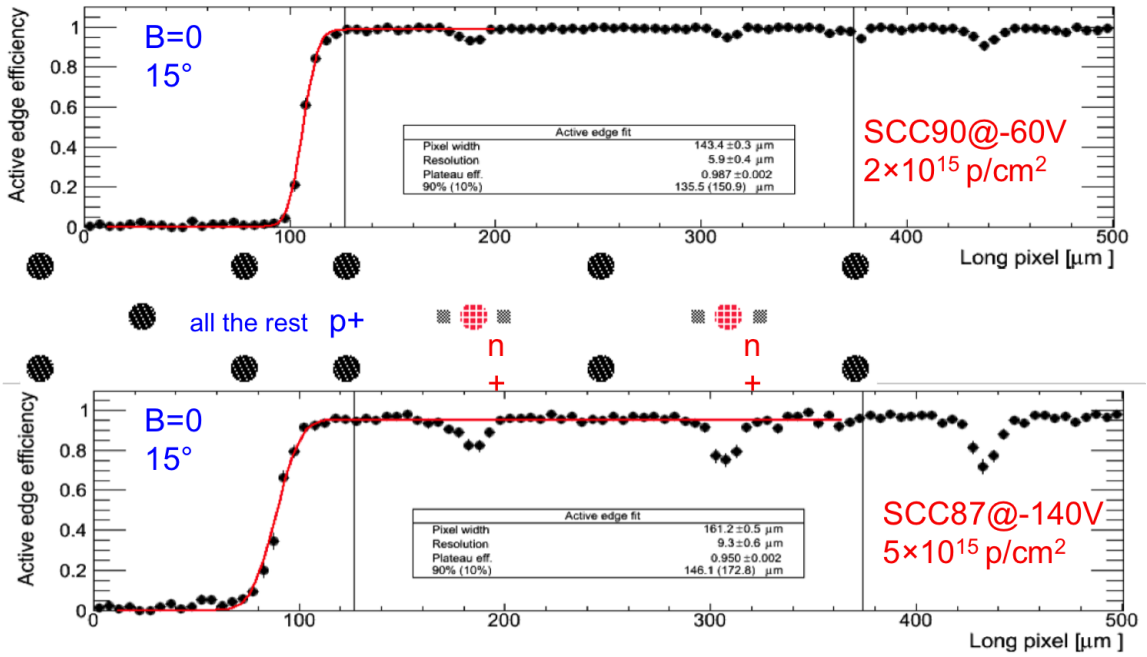


Figure 7.14: Experimental signal response of an edge pixel of 3D FBK sensors. The devices under test are proton-irradiated at fluences of 2 (upper) and 5×10^{15} $n_{eq}cm^{-2}$ (lower) biased at 60 and 140 V. Charge sharing extends the effective collection area for around 10 and 40 μm in the data respectively [123].

7.2.2 Simulated FBK Corner Result

Figure 7.15 (a) shows the design layout and the simulated field distribution. The simulation was performed by Povoli using TCAD who also was worked on the sensor fabrication [100]. The simulation result shows an effective confinement (decreasing electric field) from the guard fence. Figures 7.15 (b) and (c) show the efficiency maps from the tracking simulation for unirradiated and irradiated (at fluence of 2×10^{15} $n_{eq}cm^{-2}$) devices at 20 and 60 V respectively. There is a high efficiency at the region (marked "2" in the figure) next to the long side of the edge pixel (marked "1") due to electrode arrangement for both devices. There is still some efficiency for the unirradiated sensor at the region (marked "3") outside of the two edge pixels due to thermal diffusion.

The simulation at a fluence of 2×10^{15} $n_{eq}cm^{-2}$ gives the signal efficiency around 0.7 on average due to trapping and the small electric field close to $p+$ electrodes as shown in Figure 7.15 (c). Note that the data are the tracking efficiency, not the signal efficiency. An event is 100% effective if its signal in charge is greater than the threshold. For a 230 μm thick substrate, a collection probability of around 0.2 will give enough signal for a threshold set to 3200 electrons.

The extension of active area for edge pixels is focused. The tracking efficiencies from the experimental data and the signal efficiencies from simulations at the region

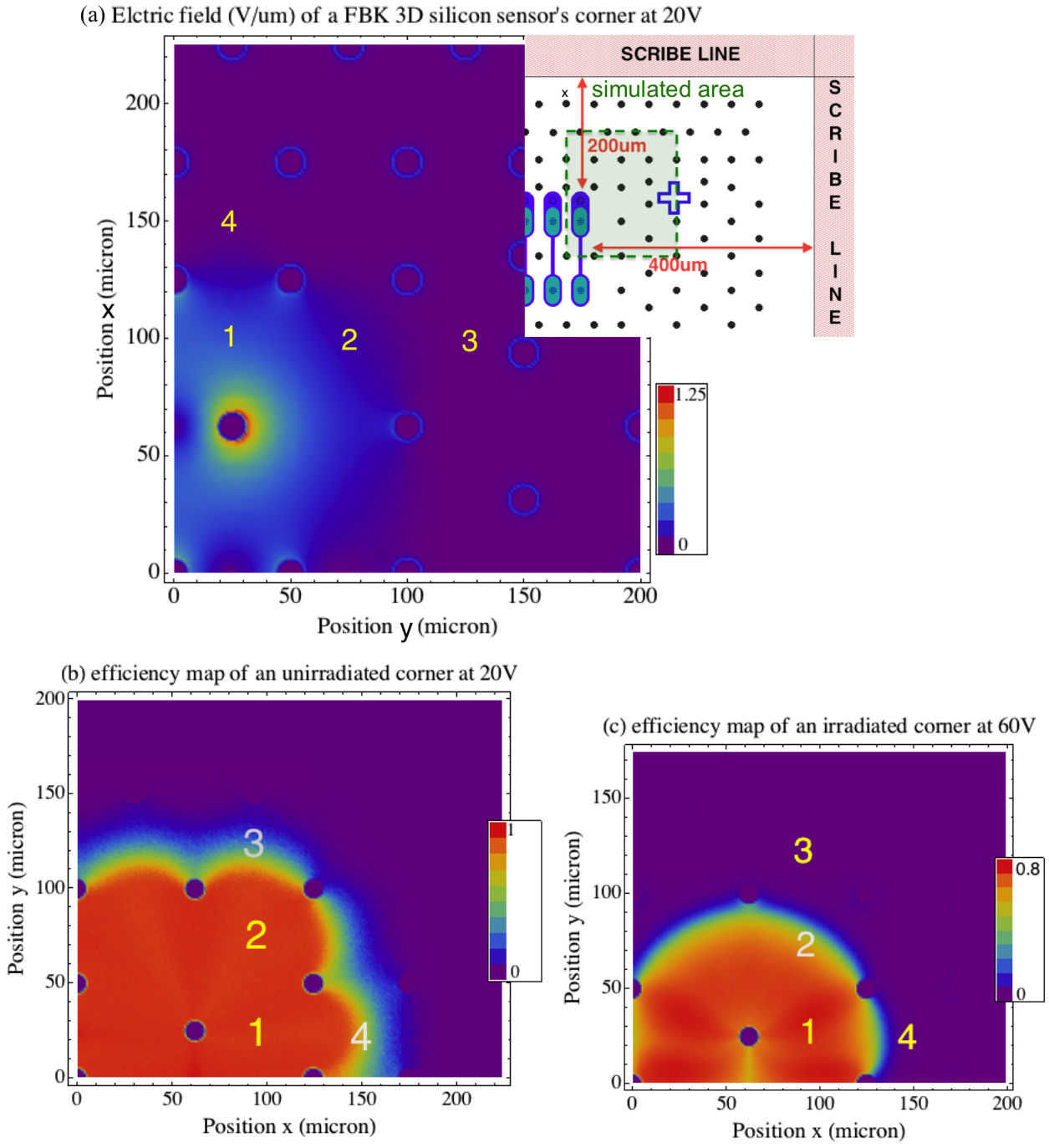


Figure 7.15: Simulated electric field (upper) and signal efficiency (lower) for different fluences at the corner of a FBK 3D sensor. Rows/columns of $p+$ electrodes (blue circles in the upper figure) are used as the guard fence [101].

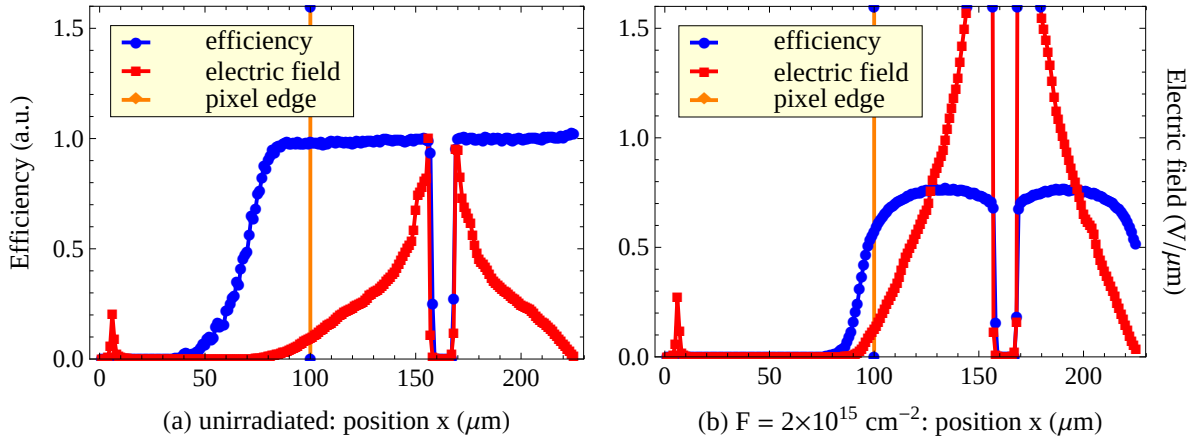


Figure 7.16: Electric field and signal response on a horizontal cut through the $n+$ electrode of the corner pixel of 3D FBK sensors. The unirradiated and irradiated device are biased at 20 and 60 V respectively. The time step $\triangle t$ is 1×10^{-12} s in the tracking simulation. The small peaks at around 0 μm are due to the concentration gradient of junctions.

(marked "4") are compared as follows:

- There is no simulation with similar conditions to the data at a fluence of 5×10^{15} $n_{eq}cm^{-2}$ in Figure 7.14 (lower) as the TCAD result was not available. But a similar behaviour can be found in the simulation in Figure 7.16 (a) for an unirradiated device. They both have an extension of collection area along the long pixels for around 40 μm. The reason is that this irradiated device was measured at 140 V and it is expected to be fully depleted to the cell edge as the unirradiated device.

- The data at a fluence of 2×10^{15} $n_{eq}cm^{-2}$ in Figure 7.14 (upper) shows an extension of the collection area along the long pixels for around 10 μm, while the simulation in Figure 7.16 (b) with matched conditions also gives an extension of 10 μm by considering the threshold. The reason is that the device was measured at 60 V which is not enough to be fully depleted and thus the active area is smaller than the previous case.

7.2.3 Experimental CNM Corner Result

Figure 7.17 shows the mask layout and experimental data of a CNM sensor. The tracking efficiencies drop rapidly at the pixel edge compared to the FBK sensors discussed above. The reason is that CNM sensors have the slim guard ring design which has a grounded $n+$ electrode between biased fence $p+$ electrodes next to each edge pixel.

Ohmic ($n+$) electrodes are connected as a slim 3D guard ring [22] and thus charge sharing is small and not much dependent on the bias voltage. The sample was also rotated by a 15° angle referenced to the long pixel line. No simulation exists since TCAD and the exact sensor design were not available at the time.

7.3 Pixel Scan at CERN

The signal response across a pixel can indicate the depletion area and how it is affected by radiation damage. The experimental data were reconstructed from pion trajectories which are less sensitive to multiple scattering due to their higher energy. The signal magnitude is proportional to the time over threshold (ToT) and is digitised as a number from 0 to 14 (4 bits). In addition to the corner scan of 3D sensors, individual pixel scans are discussed in this section.

The tracking simulation and 3D digitiser in Chapter 5 are used to predict the sensor performance. Some parameters and defect models, such as the time step, leakage current and rescaling of the effective doping concentration, are validated and preferred by the test beam data from the signal response or cluster size/charge sharing. Some tuning for the edge efficiencies is needed to explain the residual distribution which is discussed in another thesis [26].

7.3.1 Simulated FBK Pixel Result

Simulation was performed on a 2D cross section parallel to the sensor plane since Kurata is limited to 2D structures. A four-pixel area was simulated with a bin size of 1 μm^2 which is enough to compared with the data from the EUDET telescope with a resolution of a few microns. This simulation is compared and validated with the FBK data due to its fully through design which can be approximated as a 2D geometry. A full three-dimensional digitiser for 3D (and also planar) sensors using TCAD models and results is under development.

Figure 7.19 shows the electric field maps of the FE-I4 layout for unirradiated and irradiated (at a fluence of 5×10^{15} $n_{eq} cm^{-2}$) devices. Two defect models, M3 and M5 in Chapter 6, give different depletion areas. Figure 7.20 shows the corresponding efficiency maps. The model M3 is considered to give better agreement with the data using the quasi-Fermi level approximation and rescaling of the effective doping concentration. M5 is commonly used in TCAD by FBK and is included for comparison. M3 gives a smaller efficiency due to its partial depletion at the cell edge which results in a larger collection time and trapping.

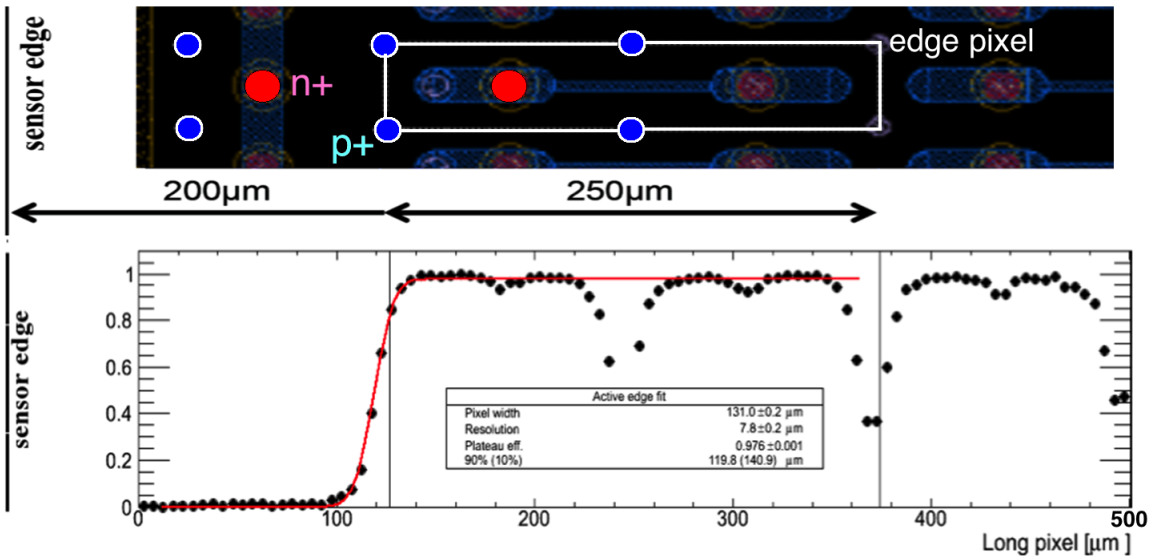


Figure 7.17: Experimental signal response of the edge pixel of a 3D CNM sensor. The device under test is neutron-irradiated at fluences of 5 $\times 10^{15}$ $n_{eq}cm^{-2}$ biased at 160 V [22][106]. A schematic layout is also shown for reference: red and blue circles represent the $n+$ and $p+$ electrodes respectively and blue pads are metal contacts.

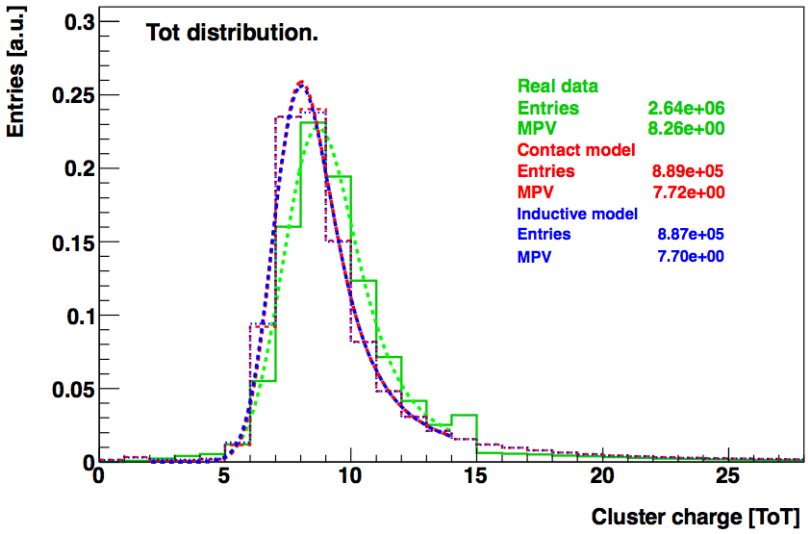


Figure 7.18: Clustered ToT distribution for an unirradiated FBK (SCC-13) 3D sensor [26]. Each histogram is fitted by a Landau distribution convoluted with a Gaussian one and. The most probable value (MPV) and mean value are extracted from the fitted distribution. ToT has a unit of the beam crossing time which is 25 ns. The threshold was set to 1600 electrons.

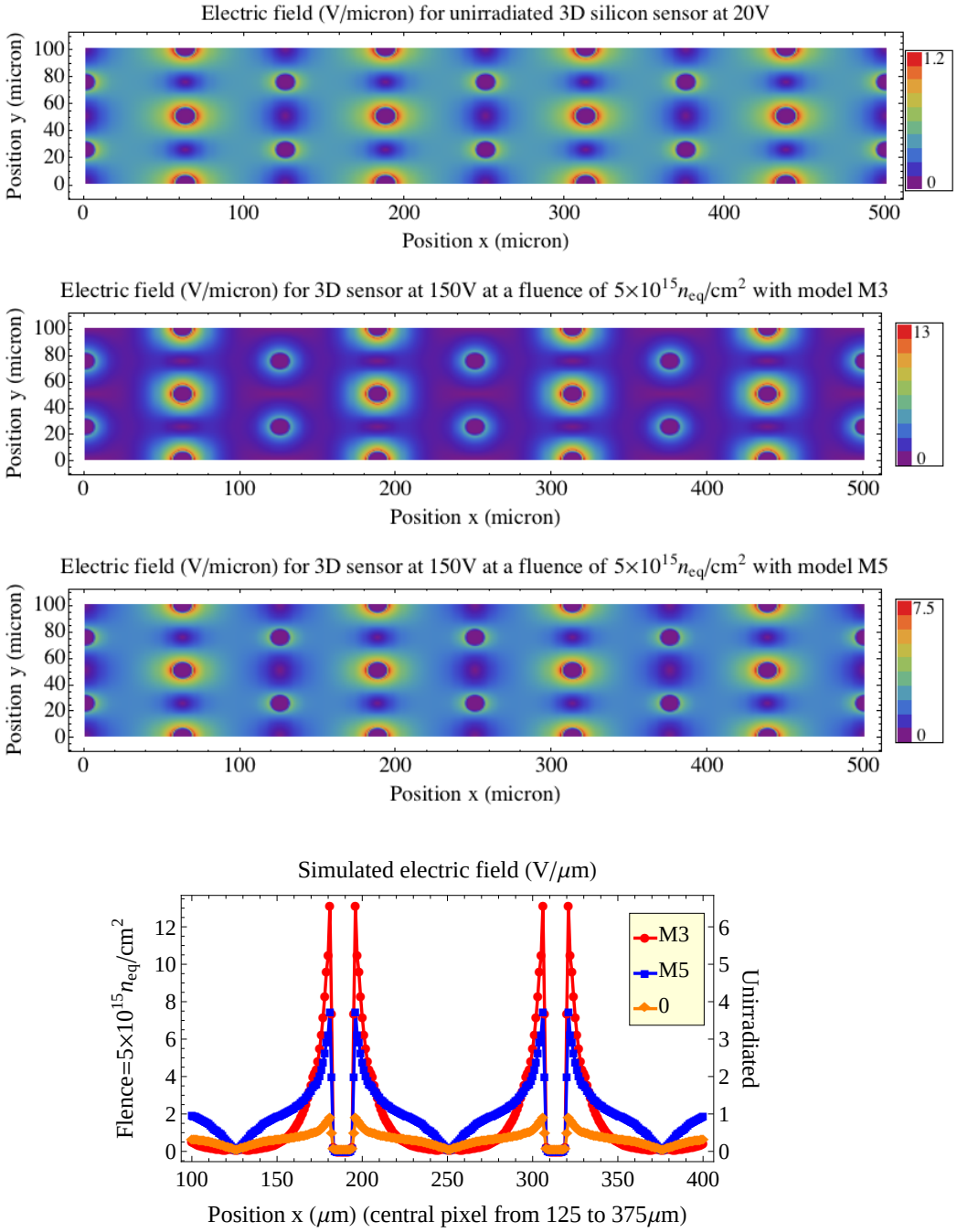


Figure 7.19: Electric field maps of a FBK 3D sensor. Both model M3 and M5 are fully depleted, but model M3 has a higher effective doping concentration and its electric field is much smaller on the $p+$ side.

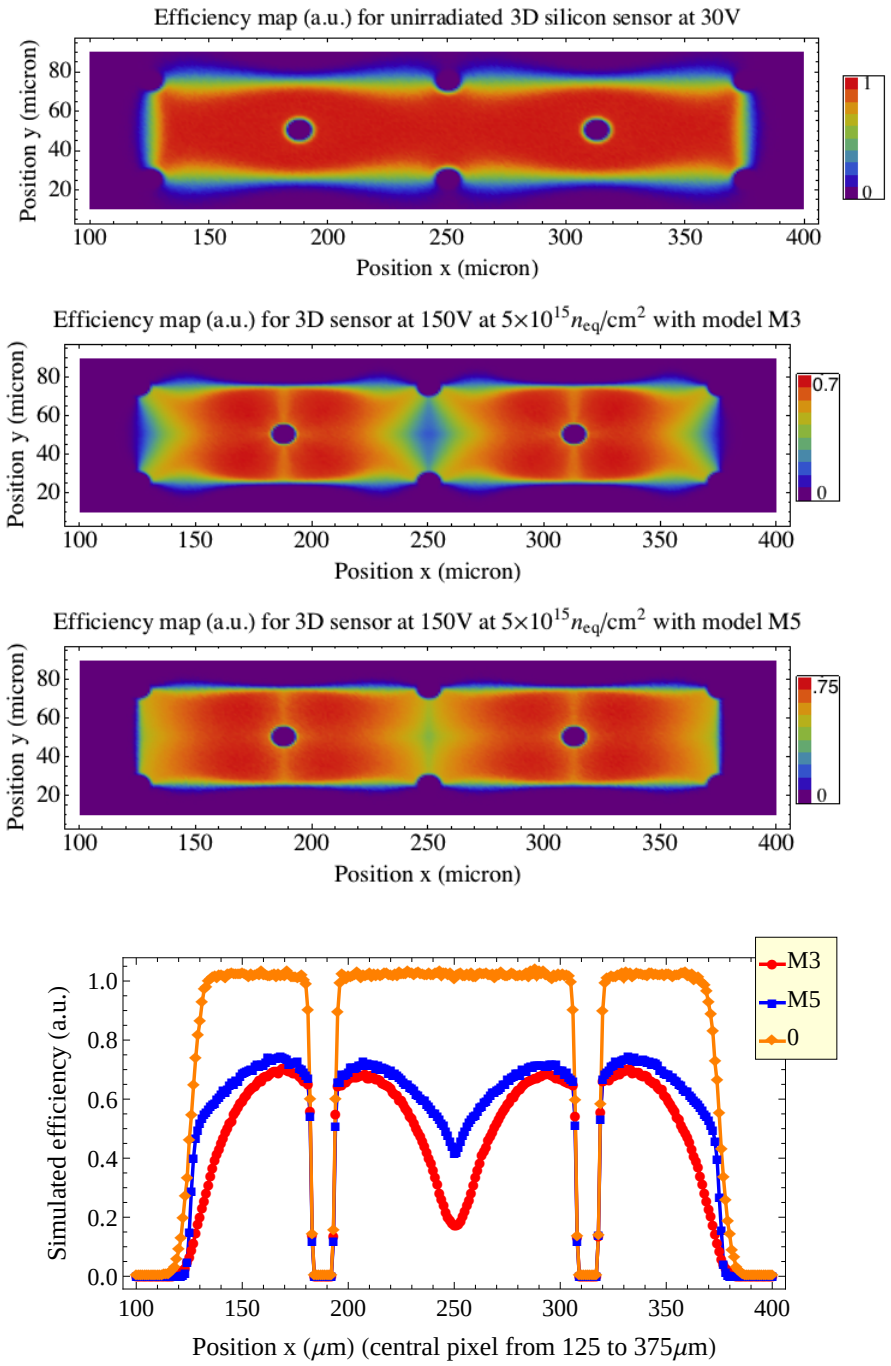


Figure 7.20: Signal efficiency maps of a FBK 3D sensor with a readout at the centre. The model M3 has a smaller signal response since its electric field is not as uniform as M5 and it requires more charge collection time at the regions close to the $p+$ electrodes with a larger trapping probability.

7.3.2 Experimental FBK Pixel Result

Efficiency maps from the tracking simulation are used as an input to the digitiser for 3D sensors. The 3D digitiser was developed in collaboration with Borri at Manchester who coded a Geant4 simulator in C++ to predict the data. It is implemented in the ATHENA framework for the ATLAS IBL system for a full simulation of the sensor performance. Further developments, e.g. defect models and incident angles (high-η[2]) for irradiated devices, will be worked on by the 3D collaboration in the future.

Figure 7.18 shows the clustered time over threshold (ToT) distribution for an unirradiated FBK 3D sensor biased at 20 V. Clustered ToT measures all energy deposited in a cluster generated by a particle. The clustered ToT distribution gives a Landau distribution. The most probable value (MPV) in the data is slightly larger than the simulation which may be due to the non-optimal conversion relation between charge and ToT. Figure 7.21 shows the cluster size distribution. The digitiser agrees with the total cluster size distribution, but has some minor discrepancies in the x and y direction. This is a validation of the 3D digitiser for an unirradiated device.

7.3.3 Experimental CNM Pixel Result

Tracking efficiency and ToT plots from experiments provide more position dependent information which can also be compared to the efficiency maps obtained by the tracking simulation. Figure 7.22 shows the 2D ToT plots for an unirradiated and a proton irradiated (at a fluence of 5×10^{15} $n_{eq}cm^{-2}$) CNM 3D sensor biased at 20 and 140 V respectively at normal incidence. ToT plots are a direct representation of the charge collection probability.

The red regions with larger ToT values (unirradiated) in Figure 7.22 (upper) correspond to the bell-shaped charge sharing region of the efficiency map in Figure 5.14 (c). This may be due to the non-optimal conversion relation between charge and ToT since charges generated are almost evenly distributed between two neighbouring pixels. The regions with higher ToT values (irradiated) in Figure 7.22 (lower) are around the $n+$ electrodes after irradiation since the device depletes from here. The low ToT regions at the cell edges are noted and the simulation shows similar effects.

Figure 7.23 shows the tracking efficiency maps of the two sensors and a neutron irradiated (at a fluence of 5×10^{15} $n_{eq}cm^{-2}$) sensor in addition. Four low efficiency circles in Figure 7.23 (upper) are the $p+$ electrodes. Figure 7.23 (middle) shows many lower efficiency regions while this is only observed around the $p+$ electrodes in Figure 7.23 (lower). The reason is that the threshold is higher and the bias voltage is lower for the proton irradiated sensor (middle) than the neutron irradiated one (lower). Thus the latter tends to deplete and low efficiency regions appear at the cell edges.

[2]Preliminary simulations at high-η regions for unirradiated devices agree with the data.

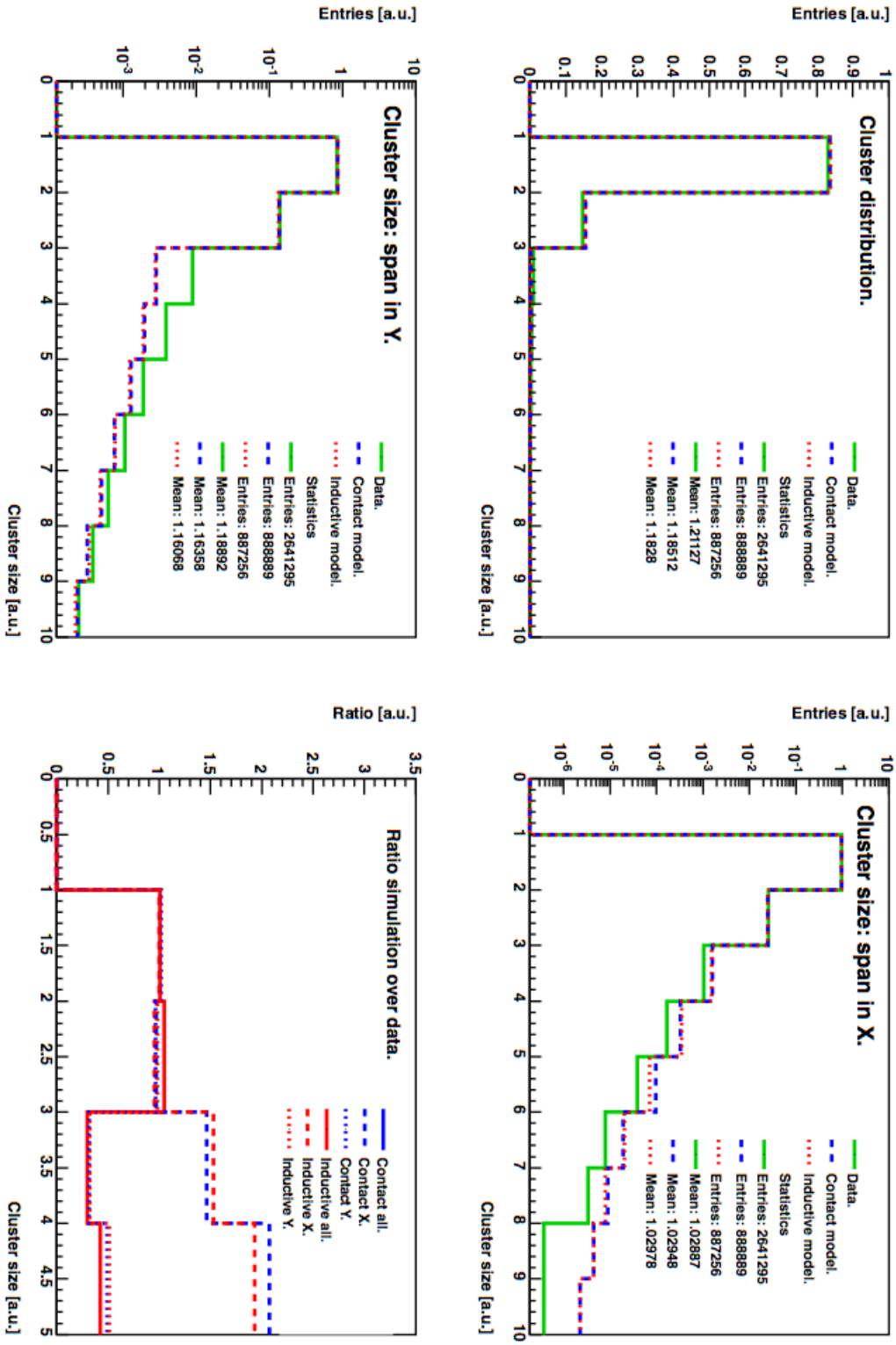


Figure 7.21: The clustered ToT distribution for an unirradiated (SCC-13) FBK 3D sensor [26]. The digitiser agrees with the total cluster size, but has some minor discrepancies in the x (slightly over-) and y (under-estimated) direction.

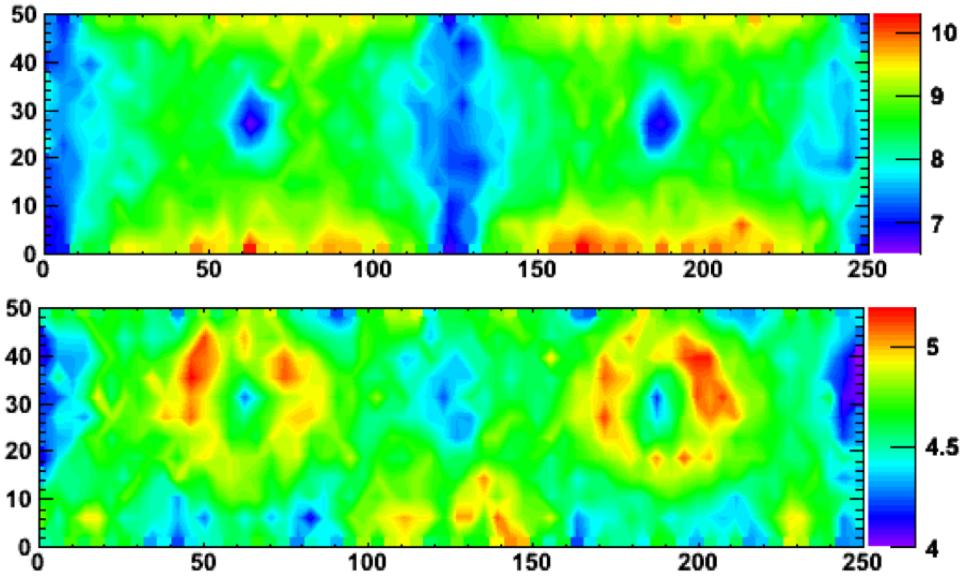


Figure 7.22: ToT maps for a pixel of an unirradiated (SCC-55, upper) and a irradiated (SCC-34, lower) CNM 3D sensor [123]. See Figure 7.23 for the experimental conditions.

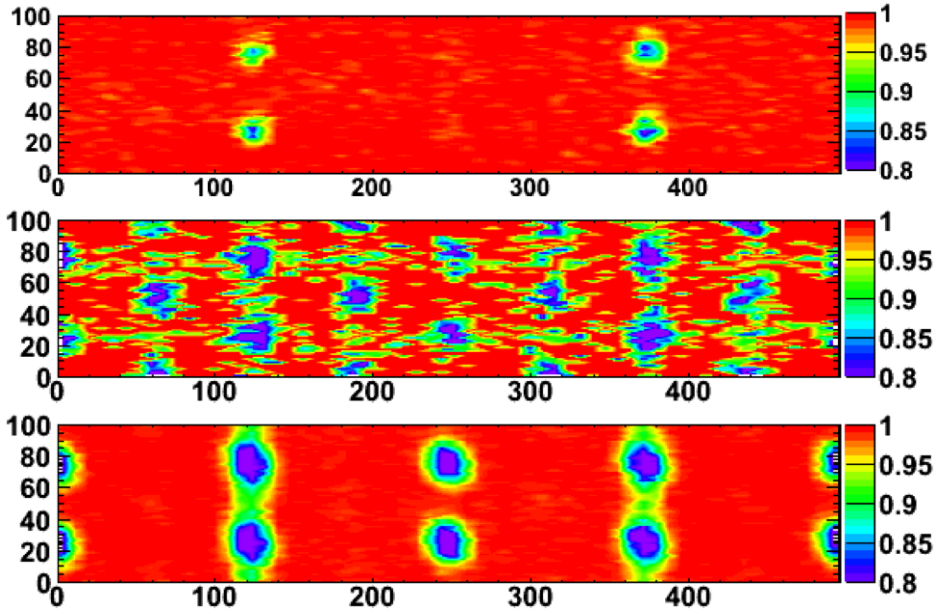


Figure 7.23: Signal efficiency maps for a pixel of an unirradiated (SCC-55, upper), a proton irradiated (SCC-34, middle) and a neutron irradiated (SCC-81, lower) CNM 3D sensor [123]. The irradiated sensors have a fluence of 5×10^{15} $n_{eq}cm^{-2}$ from 23 MeV protons at Karlsruhe and reactor neutrons at Ljubljana respectively. The threshold were set to 1600, 3200 and 1500 electrons respectively. The sensors were biased at 20, 140 and 160 V respectively. All sensors were measured at normal incidence.

Chapter 8

Summary

The goal of the thesis was to simulate and predict the physical properties of 3D sensors. To correctly model the circular electrodes, one has to extend the 1D device simulator to 2D. This is a crucial step since geometry significantly changes the Ramo weighting field and electric field which affect charge collection and the signal response.

Strip detectors were investigated to give predictions of the signal efficiencies after heavy irradiation. To obtain better agreement with the data, different models and parameters were tested. The avalanche enhanced generation current is important for carrier density and breakdown voltage estimation. The effective doping concentration N_{eff}, based on the direct experimental data and indirect simulation results, has two linear dependences for a fluence before 5×10^{14} and after 2×10^{15} $n_{eq}cm^{-2}$. The growth of N_{eff} with fluence is much reduced at high fluences.

The final and optimal parameters for the device simulation are listed in Table 8.3. Device and defect model are preferred by the data. Leakage current and introduction rates of deep acceptors are determined by the experiments. Some key figures are repeated in this chapter.

Sensor Performance after Heavy Irradiation

With 2D simulation and sensible parameters, the experimental data for planar and 3D sensors can be reasonably well predicted. Figure 8.1 shows that the simulation results agree with the data for planar strip sensors at fluences of 5×10^{14} and 1×10^{15} $n_{eq}cm^{-2}$. Figure 8.2 shows the simulation results agree with the data for CNM and FBK 3D sensors at a fluence of 5×10^{15} $n_{eq}cm^{-2}$. These results confirm the validity of the 2D Kurata simulator using the quasi-Fermi level approximation.

In summary, radiation damage has three main effects on silicon devices: an increase of the effective doping concentration and a decrease of the generation and trapping lifetimes. The generation lifetime is linear up to high fluences based on the leakage

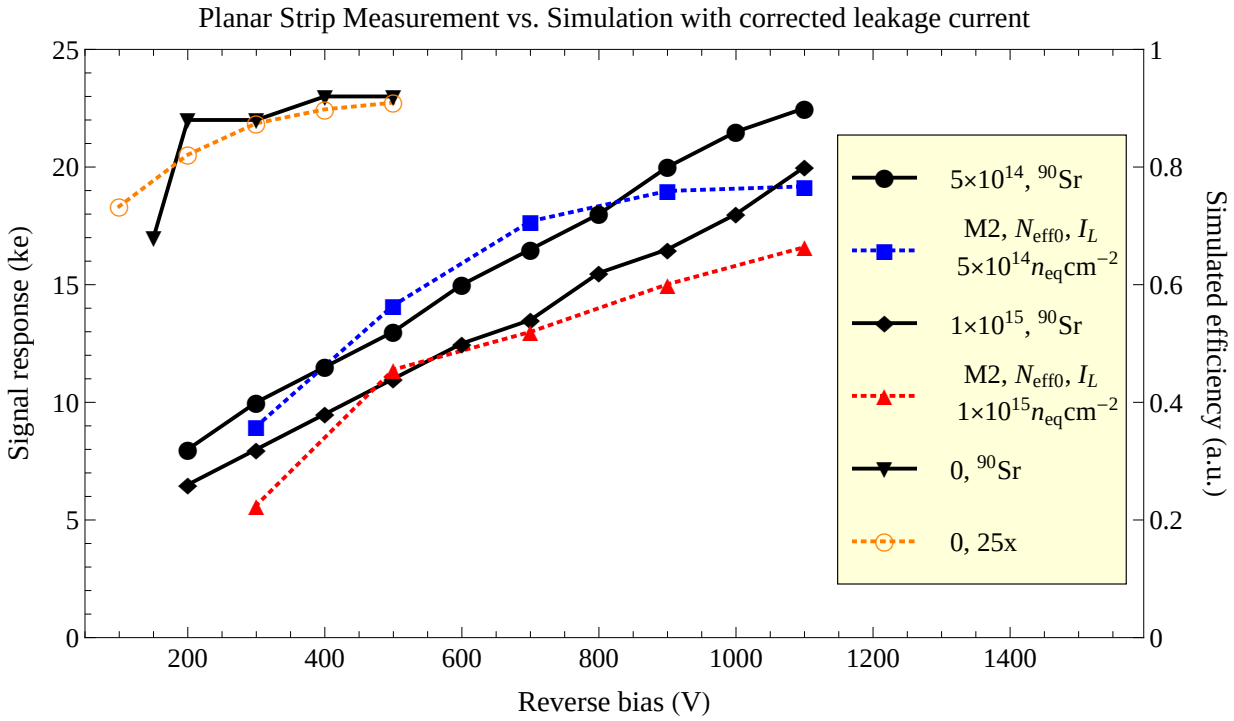


Figure 8.1: Comparison of three defect models with simulated signal efficiency of planar strip sensors at a fluence of 5×10^{14} and 1×10^{15} $n_{eq}cm^{-2}$. The solid lines are data while the dotted lines are simulations.

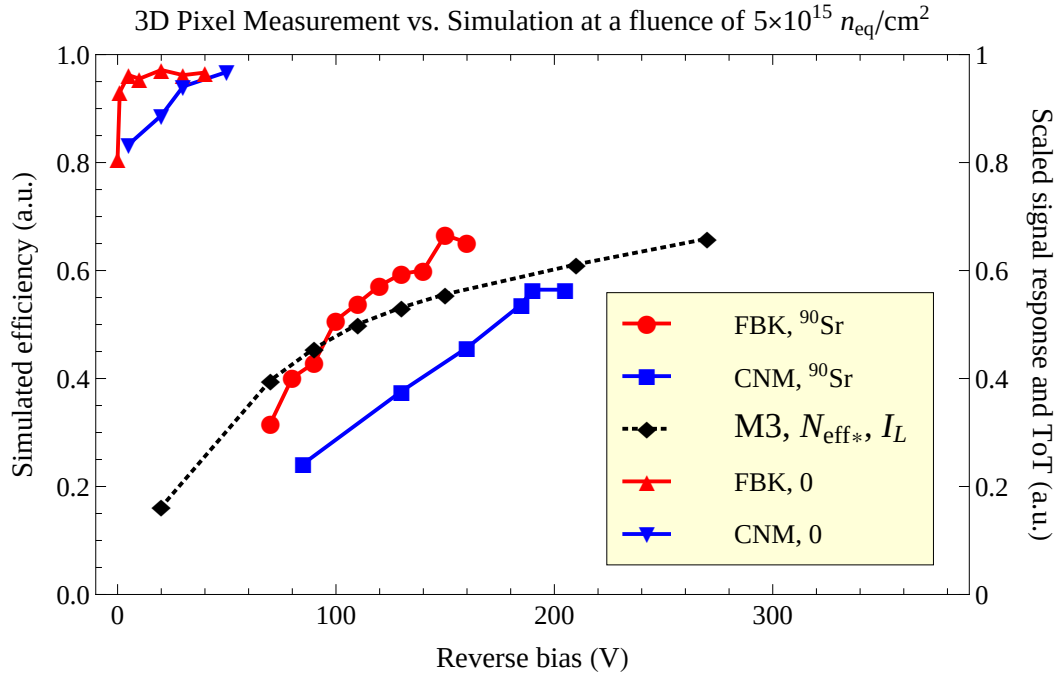


Figure 8.2: The experimental data and simulated signal efficiencies for two 3D pixel sensors at a fluence of 5×10^{15} $n_{eq}cm^{-2}$. The solid lines are the data while the dotted line is simulation.

current measurement. The trapping lifetime is also assumed to be linear based on the transient current technique. The effective doping concentration has a smaller damage constant at higher fluences which suggests that silicon sensors can survive to higher doses than one would have expected previously.

2D Efficiency Map

The efficiency maps for unirradiated and irradiated sensors at different fluences and voltages are obtained by the tracking simulation. This is closely related to the previous plots for planar and 3D sensors, in which the data represents the sum of position dependent efficiencies. The simulation time step has to be around 0.2 to 1 *ps* to model charge sharing correctly. Figure 8.4 shows the efficiency map for an un-irradiated 3D FBK sensor which is used in the 3D digitiser.

Electrode Efficiency

There is no apparent improvement for the electrode efficiencies using the new oxygen-free process. Active electrodes will preserve the effective area for the 3D technology, especially for 3D avalanche sensors whose electrodes cover more area with respect to the sensor size. Further experiments are needed to confirm the use of oxygen-free processes. The beam size and electrode radius are the most important parameters for characterising the carrier trapping lifetimes in polysilicon. The beam size and step size need to be smaller than the electrode radius.

Sensor Capacitance

Large capacitance is one weakness for 3D sensors since this leads to higher noise. Figure 8.5 shows the simulated capacitance and its relationship with the measured noise. Planar sensors have much lower noise with the current design. Both technologies plan to decrease the sensor thickness to reduce multiple scattering, but this also decreases the signal response. A reduction in thickness is crucial for planar sensors since it enhances the radiation hardness, but this also increases the capacitance and decreases the S/N ratio. A reduced thickness decreases the capacitance for 3D sensors which results in smaller overdrive in the front-end electronics and thus less power consumption. Thinned 3D sensors have similar S/N ratio and depletion voltage compared to the original substrate.

In conclusion, this thesis described the algorithms for the device and tracking simulation which explain the experimental data for both technologues using one generalised model. Improvement in the design and device processing are ongoing. Further studies

Model	Parameter	Description
G4	$\alpha_{n/p} \neq 0$ in device simulation $\alpha^{T}_{n/p} \neq 0$ in tracking simulation	avalanche enhanced leakage current with avalanche gain for signal response

Model	Description	at Fluences F/F' ($n_{eq}cm^{-2}$)
M3	$N_{eff,0} = 0.016F + 0.66 \times 10^{11}$ $N_{eff} = N_{eff,0} + 0.0016F'$ interpolation	$F < 5 \times 10^{14}$ (M2) $F' > 2 \times 10^{15}$ ($F = 5 \times 10^{14}$) between 5×10^{14} and 2×10^{15}

Leakage current	$\alpha = 7 \times 10^{-17}\ Acm^{-1}$	Ratio of β at mid-gap	$N_A/N_D = 20$

Figure 8.3: Optimal parameters for the device simulation. The introduction rate for deep acceptors is $\beta = 1\ cm^{-1}$. Both the representative deep acceptor and donor are at mid-gap.

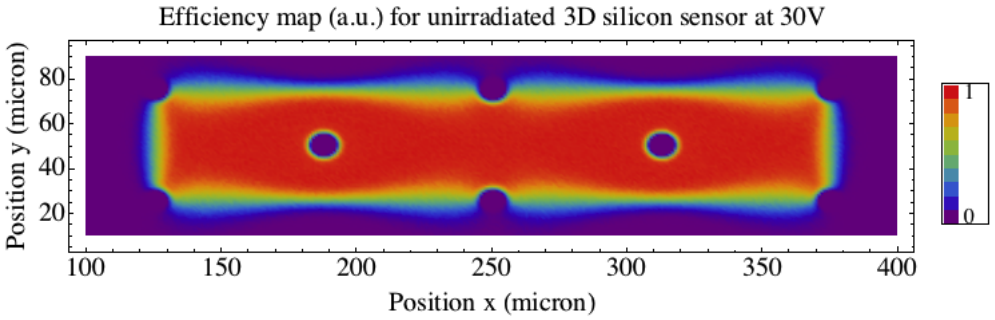


Figure 8.4: Signal efficiency maps of a FBK 3D sensor with a readout at the centre. M3 has a smaller signal response since its electric field is not as uniform as M5 and it requires more time for charge collection at the region close to $p+$ electrodes, i.e. a larger trapping probability.

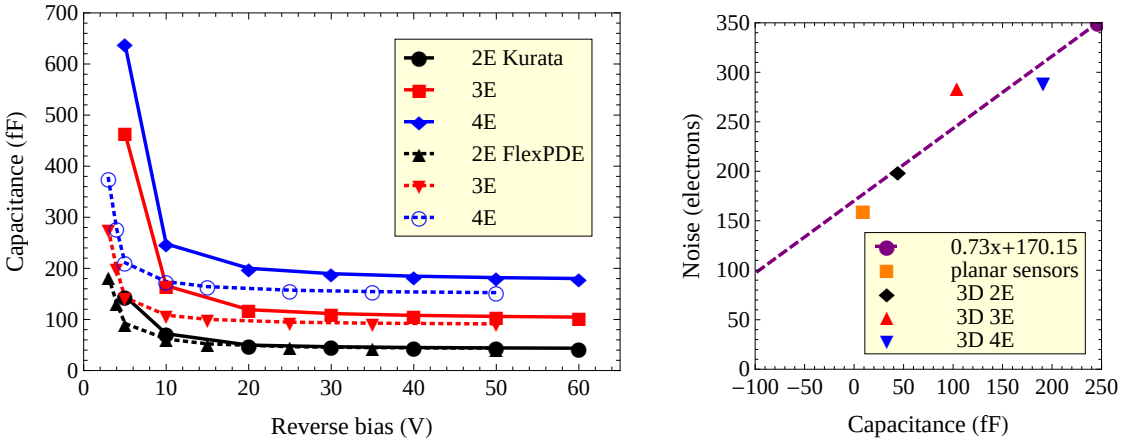


Figure 8.5: Capacitance calculated by Kurata and FlexPDE for 2E, 3E and 4E structures of 3D sensors respectively (left). The linear relationship between capacitance and noise.

of defect formation, such as defect evolution in dense clusters and effective trapping/-generation centres, will be essential to understand the radiation hardness.

Publications in Collaboration

- Charge Multiplication by Design in Thin Silicon 3D Sensors. Marco Povoli, Gian-Franco Dalla Betta, , Cinzia Da Via, Ching Hung Lai, Lucio Pancheri, Stephen Watts. Mar 2013. Submitted to *Nuclear Instruments and Methods*.

- Recent developments and future perspectives in 3D silicon radiation sensors. G. -F. Dalla Betta, C. Da Via, M. Povoli, S. Parker, M. Boscardin, G. Darbo, S. Grinstein, P. Grenier, J. Hasi, C. Kenney, A. Kok, C. -H. Lai, G. Pellegrini and S. Watts. Oct 2012. Published in *JINST* 7 C10006, DOI:10.1088/1748-0221/7/10/C10006.

- Prototype ATLAS IBL Modules using the FE-I4A Front-End Readout Chip. ATLAS IBL Collaboration (J. Albert et al.). Sep 2012. 45 pp. Published in *JINST* 7 (2012) P11010, DOI: 10.1088/1748-0221/7/11/P11010, e-Print: arXiv: 1209.1906.

- 3D-FBK pixel sensors: Recent beam tests results with irradiated devices. A. Micelli, K. Helle, H. Sandaker, B. Stugu, M. Barbero, F. Hugging, M. Karagounis, V. Kostyukhin, H. Kruger, J.W. Tsung et al. Jan 2011. 8 pp. Published in *Nuclear Instruments and Methods* A650 (2011) 150-157, SLAC-REPRINT-2012-065, DOI: 10.1016/j.nima.2010.12.209, Conference: C10-09-06.2 Proceedings.

- Test Beam Results of 3D Silicon Pixel Sensors for the ATLAS upgrade. ATLAS 3D Collaboration (P. Grenier et al.). Jan 2011. 7 pp. Published in *Nuclear Instruments and Methods* A638 (2011) 33-40, SLAC-PUB-14401, DOI: 10.1016/j.nima.2011.01.181, e-Print: arXiv:1101.4203.

Bibliography

[1] http://atlas.ch/photos/index.html. 23, 24, 25

[2] http://en.wikipedia.org/wiki/crystallographic_defect. 58

[3] http://en.wikipedia.org/wiki/electronic_band_structure. 36

[4] http://en.wikipedia.org/wiki/p-n_junction. 48

[5] http://gcc.gnu.org/. 71

[6] http://icwiki.physik.uni-bonn.de/twiki/bin/view/systems/usbpix. 143

[7] http://lamp.tu-graz.ac.at/ hadley/ss1/kronigpenney/kronigpenney.php. 36

[8] http://public.web.cern.ch/public/en/lhc/lhc-en.html. 21

[9] https://ppes8.physics.gla.ac.uk/twiki/bin/view/atlas/puuka/ibltestbeam. 156

[10] http://www2.warwick.ac.uk/fac/sci/physics/current/postgraduate/regs/mpags/ex5/bandstructure/. 36

[11] http://www.atlas.ch/news/2011/atlas-cms-combination-plot.html. 20

[12] http://www.diamond.ac.uk/. 156

[13] http://www.math.nsysu.edu.tw/statdemo/centrallimittheorem/centrallimit.html. 202

[14] http://www.pdesolutions.com/. 107

[15] http://www.silvaco.com/products/tcad/. 70

[16] http://www.sintef.no/home/. 31

[17] http://www.srim.org/. 52

[18] http://www.synopsys.com/tools/tcad/pages/default.aspx. 70

[19] http://www.wolfram.com/. 70

[20] G. Aad and the ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. 22, 24, 25, 26

[21] G. Aad and the ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. 19

[22] J. Albert. Prototype ATLAS IBL modules using the FE-I4A front-end readout chip. Technical Report ATL-COM-INDET-2012-022, CERN, Geneva, May 2012. 171, 175, 176

[23] J. P. Balbuena. *Radiation Damage in Silicon Particle Detectors*. PhD thesis, Universitat Autònoma de Barcelona, 2007. 141, 142, 148, 149

[24] R. J. Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series)*. WileyBlackwell, reprint edition, 1989. 201

[25] W. Bludau, A. Onton, and W. Heinke. Temperature dependence of the band gap of silicon. *Journal of Applied Physics*, 45(4):1846–1848, 1974. 36

[26] M. Borri. *Characterization of 3D Silicon Assemblies for ATLAS Pixel Upgrade*. PhD thesis, University of Manchester, 2013. 102, 118, 119, 175, 176, 180

[27] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer technical design report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-019, CERN, Geneva, Sep 2010. 27, 28

[28] G. Casse, A. Affolder, P. P. Allport, H. Brown, I. McLeod, and M. Wormald. Evidence of enhanced signal response at high bias voltages in planar silicon detectors irradiated up to $2.2 \times 10^{16} n_{eq}/cm^2$. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 636(1, Supplement):S56–S61, 2011. 109, 125, 127, 128, 135, 136

[29] G. Casse, D. Forshaw, T. Huse, I. Tsurin, M. Wormald, M. Lozano, and G. Pellegrini. Charge multiplication in irradiated segmented silicon detectors with special strip processing. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 699(0):9–13, 2013. Proceedings of the 8th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors. 151

[30] D. Caughey and R. Thomas. Carrier mobilities in silicon empirically related to doping and field. *Proceedings of the IEEE*, 55(12):2192–2193, Dec 1967. 90, 91

[31] S. Chatrchyan and the CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. 19

[32] V. Cindro, G. Kramberger, M. Lozano, I. Mandić, M. Mikuž, G. Pellegrini, J. Pulko, M. Ullan, and M. Zavrtanik. Radiation damage in p-type silicon irradiated with neutrons and protons. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 599(1):60–65, 2009. 109, 135

[33] C. R. Crowell and S. M. Sze. Temperature dependence of avalanche multiplication in semiconductors. *Applied Physics Letters*, 9(6):242–244, 1966. 93

[34] C. Da Via, E. Bolle, K. Einsweiler, M. Garcia-Sciveres, J. Hasi, C. Kenney, V. Linhart, S. I. Parker, S. Pospisil, O. Rohne, T. Slavicek, S. J. Watts, and N. Wermes. 3D active edge silicon sensors with different electrode configurations: Radiation hardness and noise performance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 604(3):505–511, 2009. 27

[35] C. Da Via and S. J. Watts. The geometrical dependence of radiation hardness in planar and 3D silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 603(3):319–324, 2009. 123

[36] G.-F. Dalla Betta, C. Da Via, M. Povoli, S. I. Parker, M. Boscardin, G. Darbo, S. Grinstein, P. Grenier, J. Hasi, C. Kenney, A. Kok, C.-H. Lai, G. Pellegrini, and S. J. Watts. Recent developments and future perspectives in 3D silicon radiation sensors. *Journal of Instrumentation*, 7(10):C10006, 2012. 151

[37] S. Dhar. *Analytical Mobility Modeling for Strained Silicon-Based Devices*. PhD thesis, Technische Universität Wien, 2007. 34

[38] L. Evans and P. Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001, 2008. 19

[39] J. G. Fossum and D. S. Lee. A physical model for the dependence of carrier lifetime on doping density in nondegenerate silicon. *Solid-State Electronics*, 25(8):741–747, 1982. 85

[40] J. G. Fossum, R. P. Mertens, D. S. Lee, and J. F. Nijs. Carrier recombination and lifetime in highly doped silicon. *Solid-State Electronics*, 26(6):569–576, 1983. 85

[41] K. Gill, G. Hall, and B. C. MacEvoy. Bulk damage effects in irradiated silicon detectors due to clustered divacancies. *Applied Physics Letters*, 82(1):126–136, 1997. 65

[42] K. Gill, G. Hall, S. Roe, S. Sotthibandhu, R. Wheadon, P. Giubellino, and L. Ramello. Radiation damage by neutrons and protons to silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 322(2):177–188, 1992. 30

[43] D. Gong. *GSS Simulator Manual*. University of Science and Technology of China, Jul 2007. 38

[44] U. Gösele, D. Conrad, P. Werner, Q.-Y. Tong, R. Gafiteanu, and T. Y. Tan. Point defects, diffusion and gettering in silicon. *MRS Proceedings*, 469:13–24, 1997. 60

[45] B. R. Gossick. Disordered regions in semiconductors bombarded by fast neutrons. *Journal of Applied Physics*, 30(8):1214–1218, Aug 1959. 57

[46] J. Gowar. *Optical communication systems*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2nd edition, 1993. 104

[47] P. Grenier and the 3D Collaboration. Test beam results of 3D silicon pixel sensors for the ATLAS upgrade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 638(1):33–40, 2011. 29, 32, 116, 117

[48] C. Grupen and B. A. Shwartz. *Particle Detectors*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 2nd edition, 2008. 25

[49] T. Haas. A pixel telescope for detector rd for an international linear collider. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 569(1):53–56, 2006. Proceedings of the 14th International Workshop on Vertex Detectors. 170

[50] F. Halzen and A. D. Martin. *Quarks and Leptons*. 1985. 18

[51] K. Hara, A. A. Affolder, P. P. Allport, R. Bates, C. Betancourt, J. Bohm, H. Brown, C. Buttar, J. R. Carter, G. Casse, H. Chen, A. Chilingarov, V. Cindro, A. Clark, N. Dawson, B. DeWilde, F. Doherty, Z. Dolezal, L. Eklund, V. Fadeyev, D. Ferrere, H. Fox, R. French, C. García, M. Gerling, S. G. Sevilla, I. Gorelov, A. Greenall, A. Grillo, N. Hamasaki, H. Hatano, M. Hoeferkamp, L. Hommels, Y. Ikegami, K. Jakobs, J. Kierstead, P. Kodys, M. Köhler, T. Kohriki, G. Kramberger, C. Lacasta, Z. Li, S. Lindgren, D. Lynn, P. Maddock, I. Mandić,

F. Martinez-McKinney, S. M. i Garcia, R. Maunu, R. McCarthy, J. Metcalfe, M. Mikestikova, M. Mikuž, M. Miñano, S. Mitsui, V. O'Shea, U. Parzefall, H. F.-W. Sadrozinski, D. Schamberger, A. Seiden, S. Terada, S. Paganis, D. Robinson, D. Puldon, S. Sattari, S. Seidel, Y. Takahashi, K. Toms, D. Tsionou, Y. Unno, J. V. Wilpert, M. Wormald, J. Wright, and M. Yamada. Testing of bulk radiation damage of n-in-p silicon sensors for very high radiation environments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 636(1, Supplement):S83–89, 2011. 7th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors. 86

[52] J. Hasi. Private communication. 168

[53] J. Hasi. *3D - The next step for Silicon Particle Detectors*. PhD thesis, Brunel University, 2004. 168

[54] A. Holmes-Siedle and L. Adams. *Handbook of Radiation Effects*. Oxford science publications. OUP Oxford, 2002. 58

[55] A. Hössinger. *Simulation of Ion Implantation for ULSI Technology*. PhD thesis, Technische Universität Wien, 2000. 55

[56] M. Huhtinen. Simulation of non-ionising energy loss and defect formation in silicon. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 491(1-2):194–215, 2002. 60

[57] S. W. Jones. *Silicon Integrated Circuit Process Technology*, chapter Diffusion in Silicon. IC Knowledge LLC, 2008. 88, 89

[58] I. Jung, H. Krawczynski, S. Komarov, and L. Sobotka. Simulation studies of *czt* detectors as γ-ray calorimeter. *Astroparticle Physics*, 26(2):119–128, 2006. 108

[59] A. Junkes. *Influence of radiation induced defect clusters on silicon particle detectors*. PhD thesis, Universität Hamburg, 2011. 51, 52

[60] A. Khoei and H. DorMohammadi. Validity and size-dependency of cauchy - born hypothesis with tersoff potential in silicon nano-structures. *Computational Materials Science*, 63(0):168–177, 2012. 34

[61] G. H. Kinchin and R. S. Pease. The displacement of atoms in solids by radiation. *Reports on Progress in Physics*, 18(1):1, 1955. 54

[62] C. Kittel. *Introduction to Solid State Physics*. John Wiley & Sons, Inc., New York, 6th edition, 1986. 35

[63] G. F. Knoll. *Radiation detection and measurement*. Wiley, New York, NY, 4th edition, 2010. 50, 51

[64] M. Köhler, L. Wiik, R. Bates, G.-F. Dalla Betta, C. Fleta, J. Harkonen, K. Jakobs, M. Lozano, T. Maenpaa, H. Moilanen, C. Parkes, U. Parzefall, G. Pellegrini, and L. Spiegel. Beam test measurements with planar and 3D silicon strip detectors irradiated to sLHC fluences. *IEEE Transactions on Nuclear Science*, 58(3):1308–1314, Jun 2011. 125, 141

[65] G. Kramberger, M. Batiç, V. Cindro, I. Mandić, M. Mikuž, and M. Zavrtanik. Annealing studies of effective trapping times in silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 571(3):608–611, 2007. 109

[66] G. Kramberger, V. Cindro, I. Mandić, M. Mikuž, M. Milovanović, M. Zavrtanik, and K. Žagar. Investigation of irradiated silicon detectors by edge-tct. *IEEE Transactions on Nuclear Science*, 57(4):2294–2302, Aug 2010. 29

[67] K. M. Kramer and W. N. Hitchon. *Semiconductor Devices: A Simulation Approach*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 1997. 107

[68] M. Kurata. *Numerical analysis for semiconductor devices*. Lexington Books (Lexington, Mass.), 1982. 46, 70, 71, 76, 78, 79, 92

[69] D. Kwak, J. Kim, S. Park, H. Ko, and D.-I. Cho. Why is (111) silicon a better mechanical material for MEMS: Torsion case. *ASME Conference Proceedings*, 2003(37211):259–264, 2003. 34

[70] D. V. Lang. Deep-level transient spectroscopy: A new method to characterize traps in semiconductors. *Journal of Applied Physics*, 45(7):3023–3032, Jul 1974. 68

[71] S.-C. Lee and T.-W. Tang. Transport coefficients for a silicon hydrodynamic model extracted from inhomogeneous monte-carlo calculations. *Solid-State Electronics*, 35(4):561–569, 1992. 71

[72] J. Lindhard, V. Nielsen, M. Scharff, and P. Thomsen. Integral equations governing radiation effects. *Mat. Fys. Medd. Dan. Vid. Selsk.*, 33(10):1–42, 1963. 55

[73] G. Lindström and the RD48 Collaboration. Developments for radiation hard silicon detectors by defect engineering - results by the CERN RD48 (ROSE) collaboration. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 465(1):60–69, 2001. 22, 57, 135

[74] G. Lindström, S. J. Watts, and F. Lemeilleur. RD48 status report: the ROSE collaboration (R&D on silicon for future experiments). Technical Report CERN-LHCC-2000-009, CERN, Geneva, Dec 1999. 30, 31, 32, 69, 86

[75] A. Macchiolo, L. Andricek, M. Beimforde, H.-G. Moser, R. Nisius, R. Richter, and P. Weigell. Performance of thin pixel sensors irradiated up to a fluence of and development of a new interconnection technology for the upgrade of the ATLAS pixel system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 650(1):145–149, 2011. International Workshop on Semiconductor Pixel Detectors for Particles and Imaging 2010. 111

[76] B. C. MacEvoy. Defect evolution in silicon detector material. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 388(3):365–369, 1997. Radiation Effects on Semiconductor Materials, Detectors and Devices. 57

[77] B. C. MacEvoy, G. Hall, and K. Gill. Defect evolution in irradiated silicon detector material. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 374(1):12–26, 1996. 57

[78] B. C. MacEvoy, A. Santocchia, and G. Hall. Defect-engineering rad-hard particle detectors: the role of impurities and inter-defect charge exchange. *Physica B: Condensed Matter*, 273-274:1045–1049, 1999. 67

[79] I. Mandić, V. Cindro, A. Gorisek, G. Kramberger, M. Mikuž, and M. Zavrtanik. Observation of full charge collection efficiency in heavily irradiated n+p strip detectors irradiated up to $3 \times 10^{15} n_{eq}/cm^2$. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 612(3):474–477, 2010. Proceedings of the 7th International Conference on Radiation Effects on Semiconductor materials, Detectors and Devices - RESMDD 2008. 109, 125

[80] G. Masetti, M. Severi, and S. Solmi. Modeling of carrier mobility against carrier

concentration in arsenic-, phosphorus-, and boron-doped silicon. *IEEE Transactions on Electron Devices*, 30(7):764–769, Jul 1983. 91

[81] J. Matheson, M. Robbins, and S. J. Watts. The effect of radiation induced defects on the performance of high resistivity silicon diodes. Technical Report RD20-TN-36, CERN, Geneva, Jan 1995. 65, 70, 86, 134, 135

[82] J. Matheson, M. Robbins, S. J. Watts, G. Hall, and B. C. MacEvoy. A microscopic explanation for type inversion and the annealing behaviour of radiation damaged silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 371(3):575–577, 1996. 30, 87

[83] A. Micelli. Private communication. 143

[84] M. Moll. *Radiation Damage in Silicon Particle Detectors*. PhD thesis, Universität Hamburg, 1999. 54, 57, 59, 63, 65, 66, 86, 87

[85] M. Moll, E. Fretwurst, and G. Lindström. Investigation on the improved radiation hardness of silicon detectors with high oxygen concentration. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 439(2-3):282–92, 2000. 62

[86] F. Moscatelli, A. Santocchia, D. Passeri, G. M. Bilei, B. C. MacEvoy, G. Hall, and P. Placidi. An enhanced approach to numerical modeling of heavily irradiated silicon devices. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 186(1-4):171–175, 2002. 133, 134

[87] R. S. Muller and T. I. Kamins. *Device Electronics for Integrated Circuits*. John Wiley & Sons, Inc., 1986. 34, 35, 42, 43

[88] M. Norgett, M. Robinson, and I. Torrens. A proposed method of calculating displacement dose rates. *Nuclear Engineering and Design*, 33(1):50–54, 1975. 54

[89] Y. Okuto and C. Crowell. Threshold energy effect on avalanche breakdown voltage in semiconductor junctions. *Solid-State Electronics*, 18(2):161–168, 1975. 94

[90] M. D. Osborne, P. R. Hobson, and S. J. Watts. Numerical simulation of neutron radiation effects in avalanche photodiodes. *IEEE Transactions on Electron Devices*, 47(3):529–536, Mar 2000. 70

[91] V. Palankovski. *Simulation of Heterojunction Bipolar Transistors*. PhD thesis, Technische Universität Wien, 2000. 42

[92] S. I. Parker, C. J. Kenney, and J. Segal. 3D - a proposed new architecture for solid-state radiation detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 395(3):328–343, 1997. Proceedings of the Third International Workshop on Semiconductor Pixel Detectors for Particles and X-rays. 27

[93] I. Perić, L. Blanquart, G. Comes, P. Denes, K. Einsweiler, P. Fischer, E. Mandelli, and G. Meddeler. The FEI3 readout chip for the ATLAS pixel detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 565(1):178–187, 2006. 27

[94] I. Perić and N. Wermes. *Design and Realisation of Integrated Circuits for the Readout of Pixel Sensors in High Energy Physics and Biomedical Imaging*. PhD thesis, Bonn Universität Hamburg, 2004. Presented 2004. 98, 99

[95] D. H. Perkins. *Introduction to High Energy Physics*. Cambridge University Press, 2000. 19

[96] M. Petasecca, F. Moscatelli, D. Passeri, and G. U. Pignatel. Numerical simulation of radiation damage effects in p-type and n-type FZ silicon detectors. *IEEE Transactions on Nuclear Science*, 53(5):2971–2976, Oct 2006. 96, 133, 134

[97] I. Pintilie, E. Fretwurst, G. Lindström, and J. Stahl. Second-order generation of point defects in γ-irradiated float-zone silicon, an explanation for "type inversion". *Applied Physics Letters*, 82(13):2169–2171, 2003. 30

[98] I. Pintilie, G. Lindström, A. Junkes, and E. Fretwurst. Radiation-induced point- and cluster-related defects with strong impact on damage properties of silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 611(1):52–68, 2009. 57, 69

[99] M. Porrini and P. Tessariol. Minority carrier lifetime of p-type silicon containing oxygen precipitates: influence of injection level and precipitate size/density. *Materials Science and Engineering: B*, 73(1-3):244–249, 2000. 164, 166

[100] M. Povoli. Private communication. 98, 141, 147, 172

[101] M. Povoli. *Development of Enhanced Double-Sided 3D Radiation Sensors for Pixel Detector Upgrades at HL-LHC*. PhD thesis, Università degli Studi di Trento, 2013. 173

[102] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: the art of scientific computing*. Cambridge University Press, New York, NY, USA, 2nd edition, 1992. 202, 203

[103] M. T. Robinson. The energy dependence of neutron radiation damage in solids. *Nuclear Fusion Reactor Conference*, pages 364–378, 1969. 54

[104] P. Roloff. The eudet high resolution pixel telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 604(1-2):265–268, 2009. Proceedings of the 8th International Conference on Position Sensitive Detectors. 170

[105] L. Rossi. Superconducting magnets for accelerators and detectors. *Cryogenics*, 43(3-5):281–301, 2003. 20

[106] I. Rubinskiy. Irradiation and beam tests qualification for ATLAS IBL pixel modules. Technical Report ATL-INDET-PROC-2012-007, CERN, Geneva, Mar 2012. 170, 176

[107] A. Santocchia, B. C. MacEvoy, G. Hall, G. M. Bilei, F. Moscatelli, D. Passeri, and G. Pignatel. A comprehensive analysis of irradiated silicon detectors at cryogenic temperatures. In *Nuclear Science Symposium Conference Record, 2002 IEEE*, volume 1, pages 628–632, Nov 2002. 71

[108] D. K. Schroder. *Semiconductor Material and Device Characterization*. Wiley-Interscience, 2006. 68, 69

[109] W. Shockley. The theory of p-n junctions in semiconductors and p-n junction transistors. *The Bell System Technical Journal*, 28(4):435–489, Jul 1949. 33

[110] H. Spieler. *Semiconductor Detector Systems*. Oxford University Press, 2005. 96, 162

[111] 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*, 27(7):1299–1300, Jul 1980. 93

[112] M. Swartz. A detailed simulation of the CMS pixel sensor. Technical Report CMS-NOTE-2002-027, CERN, Geneva, Jul 2002. 70

[113] M. Swartz, V. Chiochia, Y. Allkofer, C. Amsler, D. Bortoletto, L. Cremaldi, S. Cucciarelli, A. Dorokhov, C. Hörmann, D. Kim, M. Konecki, D. Kotlinski, K. Prokofiev, C. Regenfus, T. Rohe, D. A. Sanders, S. Son, and T. Speer. Simulation of heavily irradiated silicon pixel detectors. Technical Report physics/0605215, May 2006. 64, 70, 96, 97

[114] M. Swartz, D. Fehling, G. Giurgiu, P. Maksimovic, and V. Chiochia. A new technique for the reconstruction, validation, and simulation of hits in the CMS pixel detector. *PoS*, Vertex 2007(CMS-NOTE-2007-033), Jul 2007. 70

[115] I. Synopsys. *Sentaurus Device User Guide*, Jun 2012. 45, 71, 85, 91, 106

[116] S. M. Sze. *Physics of Semiconductor Devices*. Wiley, 2nd edition, 1981. 35, 37, 39, 46, 49, 93

[117] S. M. Sze. *Semiconductor Devices: Physics and Technology*. John Wiley & Sons, Inc., 2002. 33, 39, 40, 42, 44, 45, 93

[118] The ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999. 22, 24, 25

[119] The CDF and D0 Collaborations. Combined CDF and D0 upper limits on Standard Model Higgs boson production with up to 8.6 fb^{-1} of data. 2011. 19, 20

[120] The Particle Data Group. Particle physics, Jul 2012. 51

[121] The RD50 Collaboration. RD50 status report 2008 - radiation hard semiconductor devices for very high luminosity colliders. Technical Report CERN-LHCC-2010-012. LHCC-SR-003, CERN, Geneva, Sep 2010. 69

[122] O. Triebl. *Reliability Issues in High-Voltage Semiconductor Devices*. PhD thesis, Technische Universität Wien, 2012. 111

[123] S. Tsiskaridze. Beam test performance of 3d pixel detectors for the IBL upgrade. Master's thesis, Universitat Autònoma de Barcelona, Jan 2012. 172, 181

[124] V. A. J. Van Lint. *Mechanisms of radiation effects in electronic materials*, volume 1 of *Mechanisms of Radiation Effects in Electronic Materials*. Wiley, 1980. 51

[125] A. Vasilescu. The NIEL scaling hypothesis applied to neutron spectra of irradiation facilities and in the ATLAS and CMS SCT. Technical Report ROSE/TN/97-02, http://cern.ch/rd48, 1999. 53

[126] M. Vretenar, G. Bellodi, R. Garoby, F. Gerigk, K. Hanke, A. M. Lombardi, S. Maury, M. Pasini, C. Rossi, and E. Z. Sargsyan. Liner accelerator design for the upgrade of the CERN proton injector complex. volume 4, pages 529–533. Raja Ramanna Centre for Advanced Technology (RRCAT), Asia Particle Accelerator Conference (APAC), 2007. 20

[127] G. D. Watkins. Intrinsic defects in silicon. *Materials Science in Semiconductor Processing*, 3(4):227–235, 2000. 62

[128] S. J. Watts. Private communication. 50, 101, 108

[129] N. Wermes. Pixel vertex detectors. Technical Report Physics/0611075, Universität Bonn, Nov 2006. 26

[130] R. Wittmann. *Miniaturization Problems in CMOS Technology- Investigation of Doping Profiles and Reliability*. PhD thesis, Technische Universität Wien, 2007. 51, 52, 89

[131] T. Wyatt. High-energy colliders and the rise of the standard model. *Nature*, 448(7151):274–280, Jul 2007. 18

- Note that the numbers (hyperlinks in the PDF file) after each citation are the pages where the citation appear.
- Private communication includes comments, suggestions and unpublished data obtained from other researchers.

Appendix A

Kurata and the Tracking Code

The final versions of Kurata and the tracking code were written in C++. There are changeable parameters and geometry definitions in both codes. This section describes the parts in the two codes which can be modified for more specific problems.

A.1 Kurata

- Geometry and Mesh

  The geometry is defined by the main grid points through the spacing matrix $h(M, N)$ and program calculates the auxiliary spacings $h'(M, N)$. The geometry grid is important since a suitable mesh will reduce the computing time and enhance the numerical stability.

- Doping and Defect Concentrations

  The doping profile is defined for the electrodes and bulk. The deep acceptor and donor concentration are also defined according to the introduction rate β and fluence F_{exp} which were discussed in the thesis. These concentrations may need to be manually ramped up step by step for convergence of the initial solution.

- Voltage and Electric Field

  For a new run, the initial bias voltage is usually set to 1 V. The successive runs will be usually set to 2 V or so, and manually increase by a small step. Once the solution is stable, an automatic run us used with a specified final voltage. Voltage step and iteration times are also specified on demand.

- Statistics and Export

  A tolerance is specified as discussed in Chapter 4. Once the error $(\delta y/y)$ is smaller than the tolerance, matrices (tables) of various variables (electric potential, electron and hole concentration) will be exported. These values are on the main grid

points and electric field is calculated based on the varying spacings. The tables will be imported to *Mathematica* to enlarge the covered region for the tracking simulation using symmetry since only part of the structure is modelled in the device simulation.

A.2 Tracking

- Geometry and Magnetic Field

  The geometry is defined by the map dimension and electrode positions. Device structures are included in the Ramo and electric field maps in the device simulation. The magnetic field is defined by the components in the three axes in a unit of T. Other parameters, such as time step and thermal velocity, use different values as templates to meet the experimental data.

- Statistics and Export

  The generation of electron-hole pairs can be either a point or a Gaussian beam which is usually set as a few μm. The bin size is defined as 1 μm^2 for the efficiency map. Events in each beam and bin can be specified on demand. Contributions from electrons and holes, with and without impact ionisation are exported.

Appendix B

Miscellaneous Information

B.1 Numerical Methods

Some numerical methods are discussed in this section. The beam profile is usually a Gaussian distribution. Gaussian events are generated using the central limit theorem. Tracking is based on the continuous positions using the x-y coordinates. Electrical properties are recorded on grid points as 2D tables. Values of positions between grid points are obtained by interpolation.

B.1.1 Gaussian and Uniform Distribution

The Gaussian or normal distribution is a bell-shaped curve symmetric about its mean $x = \mu$. Its width is controlled by the standard deviation σ given as [24]

$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad . \tag{B.1}$$

The uniform distribution gives $1/(b-a)$ over a region from a to b. It has a mean of $(a+b)/2$ and a variance of $(a+b)^2/12$. For a random number generator in programming compliers, it corresponds $a = 0$ and $b = 1$ and gives a random number in between with equal probability.

B.1.2 Central Limit Theorem

According to the central limit theorem (CLT), a large number of sample means from a known probability distribution will approximate to a Gaussian distribution. This theorem was proved and written in different representations. The Lyapunov CLT is used for simplicity.

For a distribution with a mean of μ and standard variation of σ, X is defined as

the sum of n independent variables x as [13]

$$X = \Sigma_{i=1}^{i=n} x_i \overset{n\to\infty}{\longrightarrow} \mu \ . \tag{B.2}$$

Each sample sum X is a Gaussian event if their number N is large. This is given as

$$\begin{gathered} \delta_N^2 = \Sigma_{i=1}^{N} \sigma_i^2 \\ \frac{1}{\delta_N} \Sigma_{i=1}^{N} (X_i - \mu) \overset{N\to\infty}{\longrightarrow} P(x;0,1) \end{gathered} \tag{B.3}$$

where $P(x;0,1)$ is the Gaussain probability function defined above.

For simplicity, the uniform distribution from 0 to 1 is selected for X and 12 events are selected since it variance is $1/12$. The sum of the 12 random numbers described above minus 12 times of its mean 0.5 gives a Gaussian event.

B.1.3 Interpolation

Variables are simulated and calculated discretely on the grid points. Interpolation gives more accurate values of positions between these grid points. A random point (x, y) is mapped on to the 2D Cartesian coordinates as described in [102]

$$x[i] \leq x < x[i+1] \ , \ y[j] \leq y < y[j+1] \tag{B.4}$$

where $x[i]$ and $y[j]$ are the grid points in the x and y direction.

The simplest interpolation in 2D is the bilinear interpolation, in which the weightings of values on four corners are dependent on their inter-distance inversely. The distant ratios are defined as

$$s \equiv \frac{x - x[i]}{x[i+1] - x[i]} \quad , \quad t \equiv \frac{y - y[j]}{y[j+1] - y[j]} \ . \tag{B.5}$$

The interpolated value f of a given point (x, y) in the first order is given as

$$f(x,y) = (1-s)(1-t)\ f_{11} + s(1-t)\ f_{21} + (1-s)t\ f_{12} + st\ f_{22} \tag{B.6}$$

where 11, 21, 12 and 22 are left-lower, right-lower, left-upper and right-upper respectively according to the given point.

Interpolation is needed for the electric and Ramo fields in the tracking simulation. There are other higher-order methods that can provide a better accuracy, but they are not used since the bin size is small.

B.2 Convolution and Deconvolution

The laser and X-ray measurements were performed to obtain the structure of 3D diodes where all the $p+/n+$ electrodes are connected to a gate pn diode. The laser beam size is not an ideal point but a Gaussian distribution source with a certain sigma. The measured signal at each position integrates over the cells that the beam covers. It is about a few hundreds μm^2 and thus the image is "defocused", or more precisely, convoluted by a Gaussian lens.

B.2.1 Convolution

The convolution of f and g is written as $f * g$ and is given as

$$f * g(t) \equiv \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau \ . \tag{B.7}$$

The apparatus usually has a certain response function and sampling rate and thus the discrete convolution is given as [102]

$$f * g[n] \equiv \Sigma_{k=-M/2+1}^{M/2} f[n-k]g[k] \tag{B.8}$$

where f is the ideal signal and g is the finite impulse response (FIR) with a finite duration M. M is an even integer, the number of the non-zero response function g.

B.2.2 Deconvolution

Deconvolution is the reverse of convolution to recover the original signal from the measured signal if the FIR is known. Deconvolution is performed by solving the convolution equation intuitively, similar to the inverse Fourier transformation, an inverse function f^-1 is used. However, there is always noise appearing in the measurement and Equation B.7 becomes $f * g + \epsilon = h$ with noise ϵ.

B.2.3 Convolution Matrix

The signal matrix is convoluted with the response function matrix which is a Gaussian distribution on a 2D discrete grid. Figure B.1 is a schematic diagram for a perfect electrode image convoluted by a Gaussian apparatus. It is a normalised Gaussian matrix as shown in Equation B.9 whose elements add up to 1. It is chosen according to the beam size.

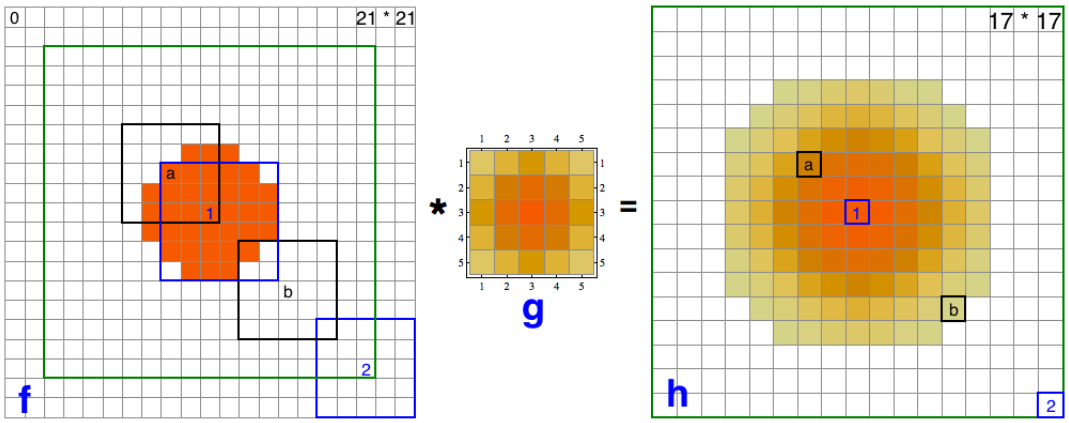


Figure B.1: A convolution of an electrode convoluted by a Gaussian apparatus.

$$\begin{pmatrix} 0.0216319 & 0.0328729 & 0.0380683 & 0.0328729 & 0.0216319 \\ 0.0328729 & 0.0499552 & 0.0578505 & 0.0499552 & 0.0328729 \\ 0.0380683 & 0.0578505 & 0.0669936 & 0.0578505 & 0.0380683 \\ 0.0328729 & 0.0499552 & 0.0578505 & 0.0499552 & 0.0328729 \\ 0.0216319 & 0.0328729 & 0.0380683 & 0.0328729 & 0.0216319 \end{pmatrix} \tag{B.9}$$

The marker labelled as "1" in Figure B.1 at the pixel centre is surrounded by 24 pixels and is multiplied with the 5×5 Gaussian matrix. The marker "a" and "b" indicate the last pixel of the real electrode and the seen electrode by apparatus respectively. By the inverse Gaussian matrix, one can deconvolve the measured signal back to the original image. However, deconvolution will give an even worse result if the image quality is low.