

TCAD Simulation of CMOS Single-Photon Avalanche Photodiode

Project Report HSP 2067

Tong Chang



A thesis submitted for the degree of Master of Engineering

The University of Edinburgh

August 2007

PROJECT MISSION STATEMENT

Project Title: Design of CMOS Single-Photon Avalanche Photodiodes (SPADs) detector for fast-speed and high-sensitivity imaging with high photon detection efficiencies at near Infrared wavelengths and investigation of minimum device pixel pitch to maximize the optical detection efficiency

Supervisor: Dr Robert K. Henderson

Student: Tong Chang

Subject Area: Technology Computer Aided Design for Photodetectors

Project Definition: The objective of this design project is to use TCAD tool Sentaurus to develop CMOS Single-Photon Avalanche Photodiodes with standard CMOS process, to find some approaches to reduce the disadvantage that silicon SPADs are of low detection efficiencies at longer wavelengths caused by the thin depletion layer with planar CMOS process, and to investigate how the variation of certain device dimension would affect the device performance.

Preparatory Tasks:

- Collect and study publications and references on SPADs, APD devices, passive and active quenching circuits
- Familiarize TCAD for process and device simulation and Cadence for device layout

Main Tasks:

- Carry out SPADs process and device simulation using Sentaurus TCAD
- Device layout using Cadence
- Integrated CMOS active and passive quenching circuits design, layout and simulation
- Device and system performance characterization

Scope of Extension:

- TDCs
- Investigate other CMOS processes, e.g. SiGe, etc

Background Knowledge:

- Optoelectronic device principles, device and process characterization and simulation
- Analog and mixed signal IC design

- Software tools, e.g. TCAD, Cadence, Hspice
- AMS 0.35 μ m HV design rule manual and design reference

Resources:

- IEEEExplore articles and other online publications
- MIT Lincoln Laboratory journals
- E. Charbon, EPFL, Switzerland
- S. Cova, POLIMI, Italy
- D. Stoppa, ITC, Italy
- Other groups work related

The supervisor and student are satisfied that this project is suitable for performance and assessment in accordance with the guidelines of the course documentation.

Signed

Student:

Supervisor:

Date:

Abstract

SPADs are avalanche photodiodes (APDs) operated at Geiger-mode (photon counting mode) for very low light level detection, which have an internal gain mechanism called avalanche multiplication. APDs have been commercially available for more than 20 years normally with a dedicated process, which do not allow monolithic integration with other electronic circuitry. To explore the huge commercial potential, several SPADs detectors with significant improved device performance have been reported in recent years with low costs using standard industrial CMOS processes [8]. Although several main influences of the technological, design parameters have been concluded, the reliable numerical models of the device operation is still under development, and the design rules of SPADs have not been settled, which is mainly because of the problems encountered in numerical description of the behaviour of devices working in very strong breakdown conditions [1].

Technology Computer Aided Design (TCAD) refers to using computer simulations to develop and optimize semiconductor processing technologies and devices. TCAD simulations are widely used throughout the semiconductor industry. As technologies become more complex, the semiconductor industry relies increasingly more on TCAD to cut costs and speed up the research and development. This aim of this project is to investigate the use of the state-of-the-art TCAD software tool Sentaurus™ to simulate Single-Photon Avalanche Photodiodes (SPADs) in CMOS technology, which allows the study of advanced physical effects via numerical simulation and design variations virtually, and helps to build up the numerical models in order to quantitatively assess the contributions of different phenomena on the device performance.

The first part of this work acts as the background of project, including Chapter 2 and Chapter 3. Chapter 2 introduces the physics and operation of the device under investigation, the variation of device structures, and design considerations; Chapter 3 describes TCAD simulations of semiconductor devices, device generation by structure editor or process simulation, refinement, physical models for device simulation, and some simulation strategies.

The second part of the project will include the application of TCAD simulation to the device, and the characterisation and measurements of the fabricated testing device. Comparison of the simulations and measurements will be made.

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the Department of Electronics and Electrical Engineering at the University of Edinburgh.

List your exceptions here and sign before your printed name.

Tong Chang

Table of Contents

Title Page	i
Mission Statement.....	ii
Abstract.....	iv
Declaration of Originality	v
Table of Contents.....	vi
List of Tables	viii
List of Figures	ix
List of Symbols	x
List of Abbreviations	xii
CHAPTER 1: INTRODUCTION	1
1.1 Introduction.....	1
1.2 Project Defination	1
1.3 Material and Methods	2
1.4 Project Organisation	3
1.4.1 Weekly Group Meetings	3
1.4.2 Gantt Chart.....	3
1.4.3 Laboratory Book	3
1.5 Thesis Structure	3
1.6 Conclusion	4
PART I: PROJECT BACKGROUND	5
CHAPTER 2: SPAD PHYSICS.....	6
2.1 Introduction.....	6
2.2 Principles of Photodetection	6
2.2.1 Photon Absorption	6
2.2.2 Quantum Efficiency and Responsivity	9
2.3 Avalanche Photodiodes (APDs)	11
2.3.1 Basic APD Operation.....	11
2.3.2 Impact Ionization	13
2.3.3 Multiplication Gain.....	14
2.3.4 Excess Noise	16
2.4 Single-Photon Avalanche Photodiodes (SPADs)	16
2.4.1 Basic SPAD Operation	16
2.4.2 SPAD Structures	17
2.4.3 SPAD Performance Parameters	21
2.4.3.1 Detection Probability	21
2.4.3.2 Dark Count Rate (DRC)	22
2.4.3.3 Afterpulsing	24
2.4.3.4 Timing Jitter.....	24
2.4.4 SPAD Design Considerations	24

2.5 Conclusion	25
CHAPTER 3: THE TCAD SIMULATIONS	26
3.1 Introduction.....	26
3.2 Device Generation	27
3.2.1 Device Structure Editor	27
3.2.2 Process Simulation.....	27
3.3 Refinement for Device Simulation	29
3.4 Device Simulation.....	30
3.4.1 Carrier Transport Models.....	30
3.4.1.1 The Drift-Diffusion Model	32
3.4.1.2 The Thermodynamic Model	32
3.4.1.3 The Hydrodynamic Model.....	33
3.4.2 Mobility Models	35
3.4.3 Carrier Generation	36
3.4.4 Carrier Recombination.....	37
3.4.4.1 Radiative Recombination.....	37
3.4.4.2 Non-Radiative Recombination	37
3.5 Other Simulation.....	38
3.6 Conclusion	38
REFERENCES	39
APPENDIX I	42
APPENDIX II	45

List of Tables

<i>Number</i>	<i>Page</i>
Table 1.1: Simple Gantt Chart for project	3
Table 2.1: Semiconductor photon absorption characteristics	7
Table 2.2: Pros and cons of thin and thick SPADs	11

List of Figures

<i>Number</i>	<i>Page</i>
Figure 1.1: Example of a general TCAD simulation flow	2
Figure 2.1: Absorption coefficient versus wavelength for various semiconductors.....	7
Figure 2.2: Photon absorption in an indirect bandgap semiconductor.....	8
Figure 2.3: Photon Absorption in a semiconductor	10
Figure 2.4: APD structure, net space charge density and field distribution	12
Figure 2.5: Multiplication of carriers in the avalanche region $\alpha_n \gg \alpha_p$	13
Figure 2.6: Reverse I-V curve of an APD under dark and light	15
Figure 2.7: Haitz's planar structure and McIntyre's reach-through structure	18
Figure 2.8: n^+/p substrate with deep diffused n^- guard ring, n^+/p^+ enrichment in a p-substrate virtual guard-ring and DJ-SPAD	18
Figure 2.9: BiCMOS SPAD.....	19
Figure 2.10: Processes for Rochas SPAD and Moloney SPAD	20
Figure 2.11: STI 0.18 μ m CMOS SPAD and the ISE-TCAD device simulation.....	21
Figure 3.1: A common process simulation strategy.....	29

List of Symbols

ρ_{net}	Net space charge density
E_g	Semiconductor Bandgap Energy
E_c	Semiconductor Conduction Band Energy
E_v	Semiconductor Valence Band Energy
k	Boltzmann constant
q	Elementary Charge
μ	Carrier Mobility
E_{ph}	Photon Energy
λ	Wavelength
α	Absorption Coefficient
δ	Penetration Depth
η	Quantum Efficiency
R	Fresnel Reflectivity
T	Transmittance
\mathcal{R}	Responsivity
α_n, α_p	Electron/Hole Impact Ionization Coefficients
\mathcal{E}	Electric Field
D_n, D_p	Electron/Hole Diffusion Coefficients
M_{ph}	Multiplication Gain
F	Excess Noise Factor
G	Generation Rate

$\langle e_n \rangle, \langle e_p \rangle$	Electron/Hole Escape Rate from Traps
F_n	Quasi-Fermi Level
E_i	Intrinsic Fermi Level
E_t	Energy Level of Trap
ρ_{trap}	Charge Density Contribute by Traps and Fixed Charges
\vec{J}_n, \vec{J}_p	Electron/Hole Current Density Vector
R_{net}	Net Electron-hole Recombination Rate
\vec{F}	Electric Field Vector
Φ_n, Φ_p	Electron/Hole Quasi-Fermi Potential
P_n, P_p	Electron/Hole Absolute Thermoelectric Power
c_L	Lattice Heat Capacitance
κ	Thermal Conductivity
m_n, m_p	Electron/Hole Effective Mass

List of Abbreviations

APD	Avalanche Photodiode
BiCMOS	Bipolar Complementary Metal-Oxide-Semiconductor
CMOS	Complementary Metal-Oxide-Semiconductor
DCR	Dark Count Rate
DD	Drift-Diffusion
EHP	Electron-hole pair
FEM	Finite Element Method
HD	Hydrodynamic
IR	Infrared
PMT	Photomultiplier tube
QE	Quantum Efficiency
SNR	Signal Noise Ratio
SIMS	Secondary Ion Mass Spectrometry
SPAD	Single-Photon Avalanche Photodiode
SRH	Shockley-Read-Hall
STI	Shallow Trench Isolation
TAT	Trap Assisted Tunnelling
TCAD	Technology Computer-Aided Design
TD	Thermodynamic
UV	Ultraviolet

Chapter 1

Introduction

1.1 Introduction

This work deals with the simulation and design of Complementary Metal-Oxide-Semiconductor (CMOS) Single-Photon Avalanche Diodes (SPADs) using one of the commercial advanced Technology Computer-Aided Design (TCAD) software suite Sentaurus™. CMOS SPADs are variations of silicon p-n junctions reverse biased above breakdown, in an all-or-nothing counting mode similar to the way Geiger detectors are used in nuclear physics for particle counting. In this Geiger mode or photon counting mode, the diode can generate a self-sustaining discharge, whose current is used as an indicator for the generation of a photoelectron and thus an absorbed photon. SPADs have long been used as photon detectors in the visible and in the near infrared spectral range. [1] summarised that SPAD's are profitably used in a wealth of applications such as time-resolved spectroscopy, chemistry, physics, and biology, fluid velocimetry optical time-domain reflectometry, single molecule detection, astronomy, distributed sensing, optical modulators, investigations of quantum-mechanical phenomena, and studies of high field properties of semiconductors. [2] gave an excellent example of application in three dimensional imaging with SPADs. While there are other types of single photon detection devices fabricated using III-V compound semiconductor materials, there exists great interests in academia and industry to exploit the benefit of low-cost silicon CMOS process to integrate the device in an array with other processing electronic circuits, to meet the different market demand for low cost, high performance, reliable single-photon detectors. This trend puts increasing pressure on SPAD designers and manufactures to optimize the device at the design level to their full potential.

1.2 Project Definition

The initial phase of the project was to study the various SPADs device structures and operation principle, and semiconductor physics associated with the device. Several devices were laid out using Cadence Layout with Euro-practice AMS 0.35 μ m technology and sent

out for fabrication. Since there has not any previous device based on this process, it was the objective of this project to use TCAD tool Sentaurus™ to develop CMOS Single-Photon Avalanche Photodiodes with standard CMOS process, to find some approaches to reduce the disadvantage that silicon SPADs are of low detection efficiencies at longer wavelengths caused by the thin depletion layer with planar CMOS process, and to investigate how the variation of certain device dimension would affect the device performance.

1.3 Material and Methods

The full Synopsys Sentaurus™ TCAD software suite was available for use during the course of the project. Process simulation was carried out using both of the two process simulators Dios or Sentaurus Process, then the output of the process simulation file were remeshed using Mdraw or Sentaurus Mesh for device simulation. The 1D and 2D doping profile could be checked either during or after the process simulation. Sentaurus Device was used to carry out the device simulation, and the output of the device simulation could be visualised by Inspect (I-V curve) and Tecplot (doping profile and field distribution). The file flow of the simulation is clearly shown in Figure 3.

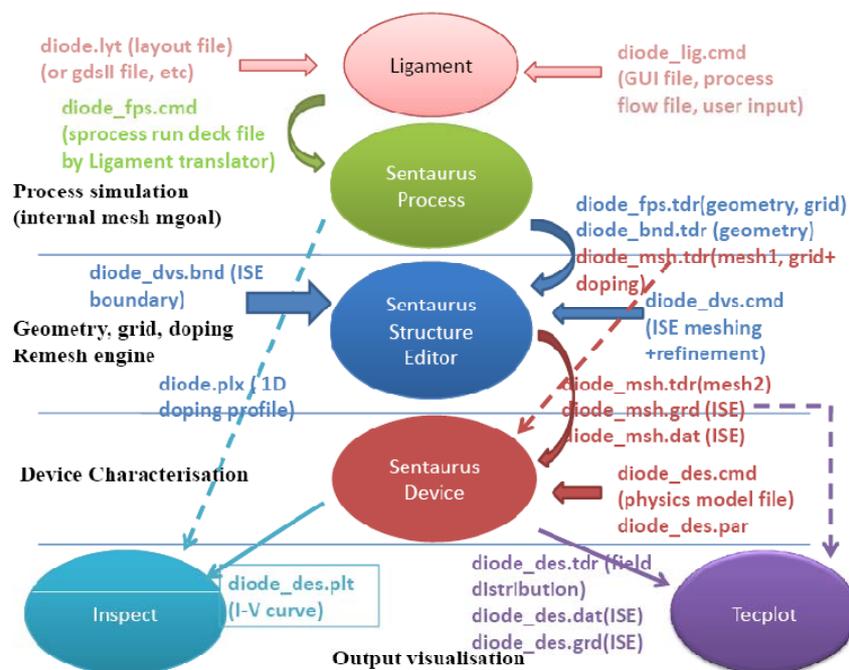


Figure 1.1 Example of a general TCAD simulation flow

1.4 Project Organisation

1.4.1 Weekly Group Meetings

Weekly meetings were arranged to meet with group members to review published papers and discuss activities carried out by each member, including progress made, difficulties encountered and short-term objectives for the following week.

1.4.2 Gantt Chart

A Gantt chart was devised to provide a general overview of long-term objectives and short-term targets. This chart was referred to regularly to ensure the project was progressing satisfactorily and to revise targets in light of new data. Figure 1.0 below is an example of the Gantt Chart structure used.

Week	2	4	6	8	10	12	14	16
Device Layout								
Process Simulation								
Device Simulation								

Table 1.1 Simple Gantt Chart for project

1.4.3 Laboratory Book

Throughout the project, a detailed laboratory book was kept to document experiments, results and conclusions on a daily basis. As the project progressed and the laboratory book was relied upon more heavily, detailed comparison between simulations and devices testing results could be compared.

1.5 Thesis Structure

The first part of this work acts as the background of project, including Chapter 2 and 3. Chapter 2 introduces the physics and operation of the device under investigation, the

variation of device structures, and design considerations; Chapter 3 describes TCAD simulations of semiconductor devices, device generation by structure editor or process simulation, refinement, physical models for device simulation, and some simulation strategies.

The second part of the project will include the application of TCAD simulation to the device, and the characterisation and measurements of the fabricated testing device. Comparison of the simulations and measurements will be made.

1.6 Conclusion

The demand for low cost, reliable, and high-performance single photon detectors, combined with the advantage of low costs of industrial standard CMOS process have given SPADs designer motivation to improve their design strategies, to speed up the research and development, and to reduce the cost. This project aims to reduce some of the current CMOS SPAD bad characteristics, such as high Dark Count Rate/dark current, etc, in an attempt to improve the overall performance of the device and to increase the market share of CMOS technologies in this field.

The project proved to be very instructive, challenging, with a wealth of new skills and in-depth knowledge being gained.

Part I

Project Background

Chapter 2

SPAD Physics

2.1 Introduction

This chapter presents the physics of avalanche photodiode and then introduces the Geiger-mode avalanche photodiode. Section 2.2 gives a general explanation of the photodetection process including photon absorption, quantum efficiency, and responsivity. Section 2.3 describes the operation of an avalanche photodiode, including impact ionization, multiplication gain, and excess noise. Section 2.4 reviews the development of Geiger-mode avalanche photodiodes, some recent CMOS SPAD structures and fabrication process, performance parameters, and design considerations.

2.2 Principles of Photodetection

2.2.1 Photon Absorption

Photodetection based on photon absorption in semiconductor materials, which can occur in three situations. Intrinsic band-to-band absorption occurs when the photon energy E_{ph} is greater than the material bandgap energy E_g , and it is the dominant absorption mechanism in most semiconductors used for photodetection. The requirement that the photon energy ($E=h\nu$) be sufficient to create an electron-hole pair (EHP) can be expressed as

$$E_{ph} = h\nu = h\frac{c}{\lambda} > E_g \quad (2.1)$$

The upper cut-off wavelength λ_g for photo-generative absorption is therefore determined by the bandgap energy E_g of the semiconductor

$$\lambda_g(\mu m) = h\frac{c}{E_g} = \frac{1.24\mu m}{E_g(eV)} \quad (2.2)$$

The bandgap energy, the corresponding maximum usable wavelength, and the typical operating wavelengths for several semiconductor materials are listed in Table 2.1 [9].

Material	Bandgap Energy (eV)	Upper cut-off wavelength (μm)	Typical operating range (μm)
Si	1.12	1.11	0.5-0.9
Ge	0.67	1.85	0.9-1.3
GaAs	1.43	0.87	0.75-0.85
$In_xGa_{1-x}As_yP_{1-y}$	0.38-2.25	0.55-3.26	1-1.6

Table 2.1 Semiconductor photon absorption characteristics [9]

Incident photons with wavelengths shorter than the upper cut-off wavelength λ_g become absorbed as they travel in the semiconductor. The photon intensity decreases exponentially with the depth in the semiconductor

$$N_{ph}(x) = N_{opt}e^{-\alpha x} \quad (2.3)$$

The depth in the semiconductor is represented as x and α is the absorption coefficient, which depends on the photon energy or wavelength and semiconductor material as illustrated in Figure 2.1 [9]. Most of the photon absorption (63%) occurs over a depth $\delta = \frac{1}{\alpha}$ called the penetration depth. Penetration depth is defined at the point at which e^{-1} of the optical power remains.

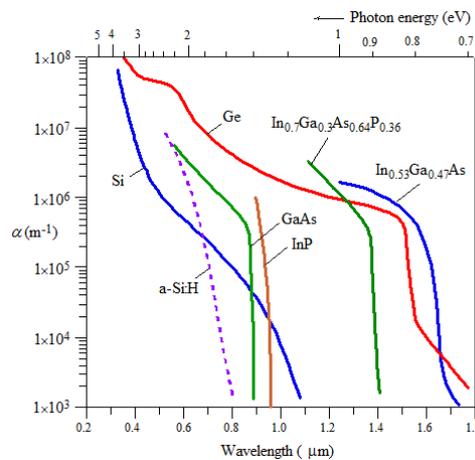


Figure 2.1 The absorption coefficient versus wavelength for various semiconductors [9]

In indirect bandgap semiconductors such as Si and Ge, the photon absorption for photon energy near E_g requires the absorption and emission of lattice vibrations (phonons) during the absorption process as shown in Figure 2.2. The absorption process is called indirect as it depends on lattice vibrations which in turn depend on the temperature. Since the interaction of a photon with a valence electron needs a lattice vibration, the probability of photon absorption is not as high as in direct transition as those in III-V semiconductors (e.g. GaAs, InAs, InP, GaSb) and their alloys (e.g. InGaAs, GaAsSb). The onset of the absorption does not exactly coincide with bandgap energy E_g , but typically it is very close to E_g as long as the phonon energy is small ($<0.1\text{eV}$).

At the wavelength of radiation, the absorption occurs over a depth covering the depletion layer so that the photo-generated EHPs can be separated by the field and collected by the electrodes. If the absorption coefficient is too large then absorption will occur very near the surface of the p^+ layer which is outside the depletion layer. First, the absence of a field means that the photo-generated electron can only make it to the depletion layer to cross to the n-side by diffusion. Secondly, photo-generation near the surface invariably lean to rapid recombination due to surface defects that acts as recombination centres. On the other hand, if the absorption coefficient is too small, only a small portion of the photons will be absorbed in the depletion layer and only a limited number of EHPs can be photo-generated.

The lower cut-off wavelength is determined by the charge collection efficiency of electrons and holes near the surface of the semiconductor. Since the penetration depth is inversely related to the absorption coefficient, the higher energy photons do not travel as far as the lower energy photons.

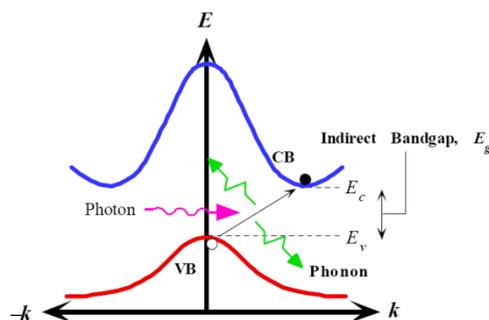


Figure 2.2 Photon absorption in an indirect bandgap semiconductor [9]

2.2.2 Quantum Efficiency and Responsivity

Since not all the incident photons are absorbed to create free EHPs that generate photocurrent. There is a basic metric called quantum efficiency (QE), defined as the number of free carriers (EHPs) produced per incident photon

$$\eta = \frac{I_{ph}}{q\Phi} = \frac{I_{ph}}{q} \left(\frac{hv}{P_{opt}} \right) \quad (2.4)$$

where I_{ph} is the photocurrent, $\Phi = \frac{P_{opt}}{hv}$ is the incident photon flux, and P_{opt} is the incident optical power. The ideal quantum efficiency is unity. In reality, the reduction of quantum efficiency is due to the current loss by recombination, incompleteness of absorption, reflection, etc [4]. It depends on the absorption coefficient α of the semiconductor at the wavelength of interest and on the structure of the device. Thus, QE can be increased by reducing the reflections at the semiconductor surface, increasing absorption within the depletion layer and preventing the recombination or trapping of carriers before they are collected. This QE defined is known as the external QE.

The internal QE is the number of free carriers per *absorbed* photon and is quite high for many devices. The incident photon will continue to lose energy to the semiconductor crystal lattice as it propagates through the semiconductor, as illustrated in Figure 2.3. There is a reflection loss due to the differences in index of refraction at the semiconductor surface. The Fresnel reflectivity for an optical signal at normal incidence to an interface between two materials is given by

$$R = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2} \quad (2.5)$$

where: n_1 = index of refraction of first material (for air $n_1 \approx 1$)

n_2 = index of refraction of second material

Reflection losses can be reduced with anti-reflection coatings to less than 1%.

Transmittance T relates the intensity of the transmitted photons to that of the incident photons, which adds the light reflected to unity, i.e. $T+R=1$. The transmittance for an optical signal at normal incidence to an interface is given by

$$T = \frac{4n_1n_2}{(n_1+n_2)^2} \quad (2.6)$$

For the air, silicon dioxide and silicon surface system, the common model used to calculate the wavelength-dependent transmittance $T(\lambda)$ from air to silicon through a thin silicon dioxide layer for a normal incidence optical signal is given by [8]

$$T(\lambda) = \frac{4n_{Si}}{\left[(n_{Si}+1)\cos\delta - \frac{k_{Si}}{n_{ox}}\sin\delta \right]^2 + \left[\frac{n_{Si}+n_{ox}}{n_{ox}}\sin\delta - k_{Si}\cos\delta \right]^2} \quad (2.7)$$

The silicon dioxide layer is considered as a non-absorbing medium with a refractive index n_{ox} , n_{Si} and k_{Si} are the real and imaginary part of the complex silicon refraction $n_{Si} - ik_{Si}$ and $\delta = \frac{2\pi}{\lambda}n_{ox}d$ is the phase change on a traversal of the silicon dioxide, where d is the thickness of the silicon dioxide layer.

The amount of power absorbed in the semiconductor as a function of distance is give by

$$P_{absorb}(x) = P_{opt}(1 - R)(1 - e^{-\alpha x}) = P_{opt}T(1 - e^{-\alpha x}) \quad (2.8)$$

Thus, the number of photocarriers generated per number of incident photons for a specific semiconductor material with reflectivity R and absorption coefficient α is the internal quantum efficiency of the photodetector.

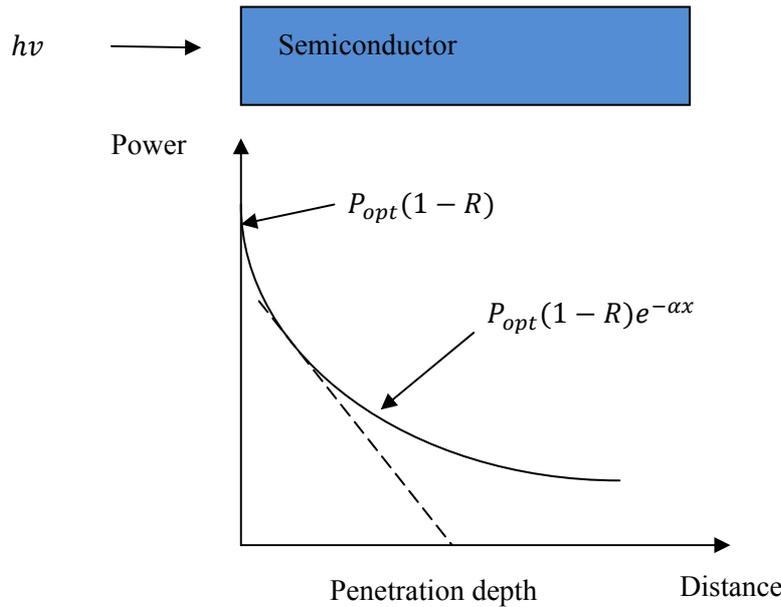


Figure 2.3 Absorption in a semiconductor

Another similar metric is the responsivity, which characterises the performance in terms of the generated photocurrent I_{ph} per incident optical power P_{opt} at a given wavelength,

$$\mathcal{R} = \frac{I_{ph}}{P_{opt}} = \frac{\eta q}{h\nu} = \frac{\eta\lambda(\mu m)}{1.24} \quad A/W \quad (2.6)$$

2.3 Avalanche Photodiodes (APDs)

2.3.1 Basic APD Operation

Avalanche photodiodes (APDs) are widely used in optical communications due to their high speed and internal gain. A simplified generic schematic diagram of a silicon reach-through APD is shown in Figure 2.4. The thin p^+ side is connected to anode and illuminated to optical signal. There are three other layers of different doping levels to suitably modify the field distribution across the diode: the thick lightly p-type doping (almost intrinsic) π layer, the thin p-type layer and the thin n^+ layer connected to cathode. The diode is reverse biased to increase the fields in the depletion regions. The net space charge distribution across the diode is shown in Figure 2.4 (b). The depletion region in the p-layer will not extend across to the π -layer under zero bias, but it widens to reach-through to the π -layer when a sufficient reverse bias is applied. The field extends from the exposed positively charged donors in the thin depleted region in the n^+ layer all the way to the exposed negatively charged acceptors in the thin depleted region in the p^+ layer. The electric field is the integration of the net space charge density ρ_{net} across the diode subject to the reverse bias voltage, which is shown in Figure 2.4 (c). The field is maximum at the n^+p junction, then decreases fast through the p-layer and slightly through the π -layer as the net space charge density is very small in this layer. The field vanishes at the end of the narrow depleted region in the p^+ layer. The absorption of photons (photogeneration) takes place mainly in the almost intrinsic π -layer, where the nearly uniform field separates the EHPs and drifts them at velocities near saturation towards n^+ and p^+ sides respectively. When the drifting electrons reach the p-layer, they experience even greater fields and therefore acquire sufficient kinetic energy ($> E_g$) to impact-ionize some of the silicon covalent bonds and release EHPs as illustrated in Figure 2.5. These generated EHPs can also be accelerated by the high fields to gain sufficient energies to further cause impact ionization and release more EHPs, which is termed as avalanche of

impact ionization. The APD is said to possess an internal gain mechanism in that a single photon absorption leads to a large number of EHPs.

The avalanche multiplication is a statistical process, which leads to carrier generation fluctuation (excess noise) in the avalanche multiplication process. This is minimized if impact ionization is restricted to the carrier with the highest impact ionization efficiency which in silicon is the electron ($\alpha_n > \alpha_p$). This is also the reason for keeping the photon absorption (photogeneration) within the nearly intrinsic π -layer in the structure so that it only allows the photogenerated electrons to drift to the avalanche region but not the photogenerated holes.

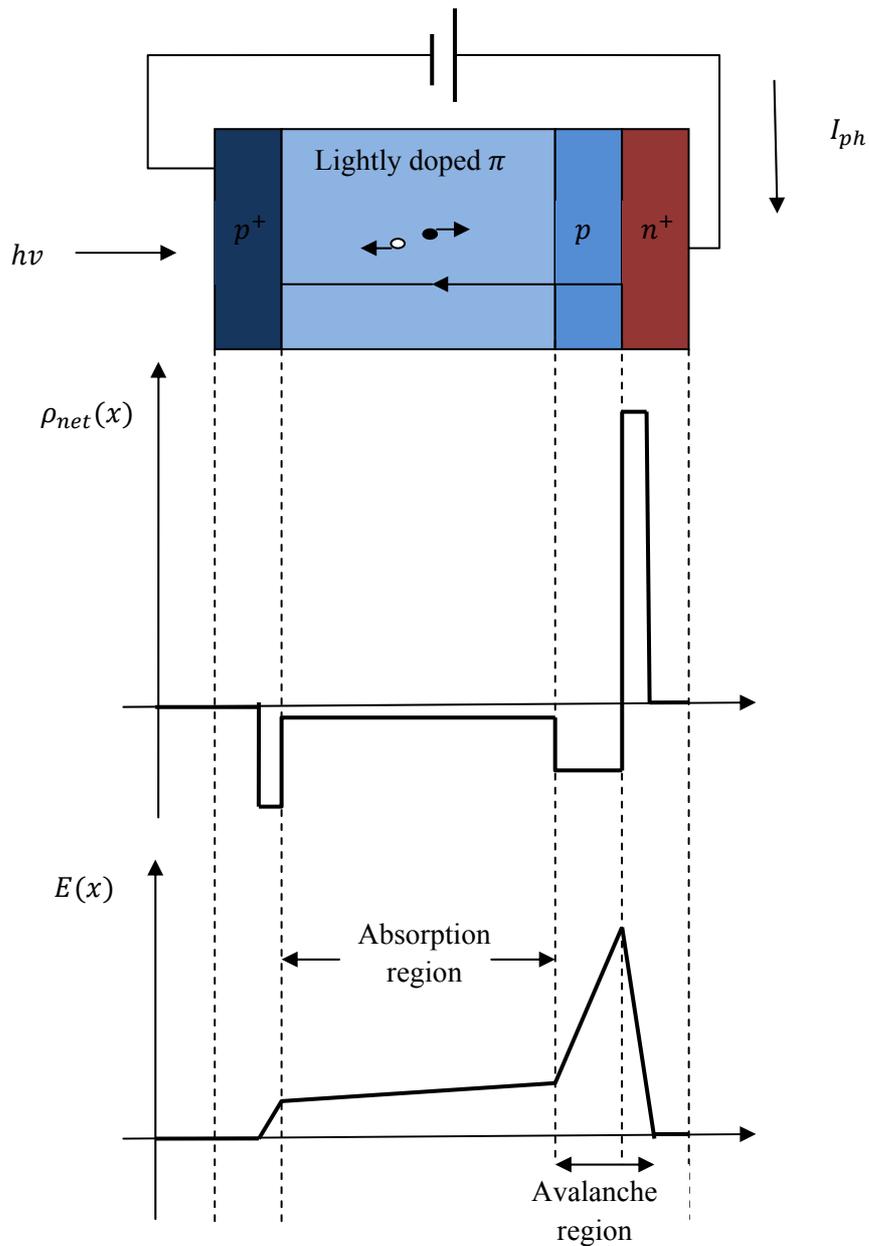


Figure 2.4 APD structure, net space charge density and field distribution

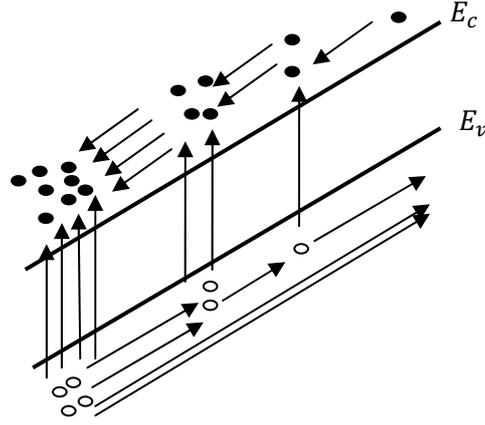


Figure 2.5 Multiplication of carriers by impact ionization in the avalanche region $\alpha_n \gg \alpha_p$

2.3.2 Impact Ionization

Impact ionization (avalanche multiplication) is the most important mechanism in junction breakdown. The ionization coefficients for electrons and holes correspond to the probability that a carrier will cause an ionization event in a unit length. For a given temperature the ionization coefficients are exponentially dependent on the electric field and are given by [4] and [3]. In [4] the ionization coefficients are given by

$$\alpha_{n,p}(\mathcal{E}) = \frac{q\mathcal{E}}{E_I} e^{(-\frac{\mathcal{E}I}{\mathcal{E}})} \quad (2.7)$$

where \mathcal{E} is the electric field, \mathcal{E}_I is the threshold field for carriers to overcome the decelerating effects of ionization scattering, and E_I is the high-field effective ionization threshold energy. For Si, the value of E_I is found to be 3.6eV for electrons and 5.0eV for holes.

In [3], it pointed out that the location of avalanche breakdown is a major concern in the design of planar APDs since breakdown location impacts such issues as device reliability, dark current-voltage (I-V) characteristics, leakage current, gain, noise and yield. It also gave a similar equation to previous one

$$\alpha_{n,p}(\mathcal{E}) = a_{n,p} e^{[-(\frac{b_{n,p}}{E^*})^{c_{n,p}}]} \quad (2.8)$$

$$E^* = \frac{|\vec{E} \cdot \vec{J}_{n,p}|}{|\vec{J}_{n,p}|} \quad (2.9)$$

where J_e, J_h are electron and hole current densities, \vec{E} is the electric field, E^* is the electric field in the direction of the current density, and $a_{n,p}$, $b_{n,p}$ and $c_{n,p}$ are experimentally determined constants. These equations also show that the generation rate due to impact ionization is exponentially dependent on the flux of carriers entering the avalanche region. Furthermore, the ionization coefficients decrease with bandgap energy.

An important parameter for describing avalanche photodiode performance is the ionization ratio $k = \frac{\alpha_p}{\alpha_n}$. For minimum noise, a large difference in ionization coefficients for electrons (α_n) and holes (α_p) is crucial. For Si, electrons are much more likely to be impact ionized than holes and the value for k between 0.003 and 0.01 can be obtained, while for Ge and III-V compounds k is between 1 and 2 for high electric fields, and high quantity APDs are difficult to fabricate. Further discussion of impact ionization modelling will be taken in chapter 3.

2.3.3 Multiplication Gain

Avalanche multiplication is a result of multiple impact ionizations, and the expressions for multiplication gains for electrons and holes in a total avalanche gain region of length L are

$$M_n = \frac{1}{1 - \int_0^L \alpha_n \exp[-\int_x^L (\alpha_n - \alpha_p) dx'] dx} \quad (2.10)$$

$$M_p = \frac{1}{1 - \int_0^L \alpha_p \exp[-\int_x^L (\alpha_n - \alpha_p) dx'] dx} \quad (2.11)$$

For $\alpha_n \gg \alpha_p$, $M_e = \exp \int_0^L \alpha_e dx$, and for $\alpha_n = \alpha_p = \alpha$, $M = \frac{1}{1 - \int_0^L \alpha dx}$. The I-V curve of an avalanche photodiode is shown in Figure 2.6. At low reverse bias, dark current is not amplified but at high reverse bias, both dark current and photocurrent are amplified.

The quantum efficiency of photons generating the primary photocurrent J_{ph} depends on the depletion width [4]

$$\eta = (1 - R) \left[1 - \frac{\exp(-\alpha W_D)}{1 + \alpha L_p} \right] \quad (2.12)$$

$$L_p = \sqrt{D_p \tau_p} \quad (2.13)$$

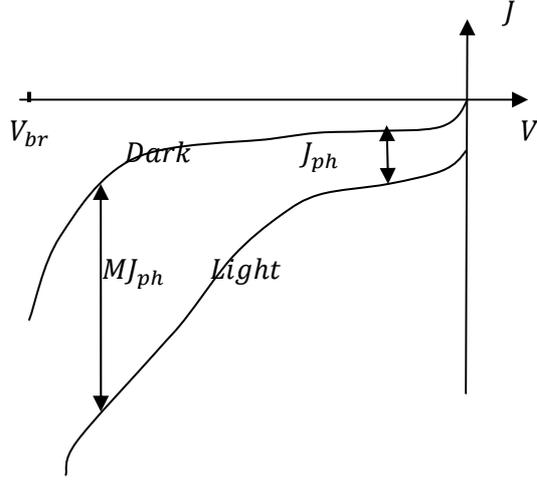


Figure 2.6 Reverse I-V curve of an APD under dark and light [5]

where α is the absorption coefficient, W_D is the depletion-layer width, L_p is minority hole diffusion length, D_p is the diffusion coefficient for holes, τ_p is the lifetime of excess carriers.

The dark current in a p-n junction is composed of the standard diffusion current, generation current in the depletion region, and the current generated at the perimeter near the guard-ring due to higher field and higher number of defects at the semiconductor surface. The perimeter component gets multiplied by a different factor than the internal primary photocurrent.

In a practical device, the maximum achievable DC multiplication at high light intensities is limited by the series resistance and space charge effect, which can be combined into a series resistance R . The multiplication gain can be estimated by an empirical equation [4]

$$M_{ph} = \frac{I - I_{MD}}{I_P - I_D} = \frac{1}{1 - \left(\frac{V_R - IR}{V_B}\right)^n} \quad (2.14)$$

Where I is the total multiplied current, I_P is the unmultiplied primary photocurrent, and I_D and I_{MD} are respectively the primary and multiplied dark currents. V_R is the reverse-bias voltage, V_B is the breakdown voltage, and the exponent n is a constant depending on the semiconductor material, doping profile and radiation wavelength. For high light intensity ($I_P \gg I_D$) and $IR \ll V_B$, the maximum value of the photo-multiplication is given by

$$M_{ph}|_{max} \approx \frac{I}{I_P} = \frac{1}{1 - \left(\frac{V_R - IR}{V_B}\right)^n} \Big|_{V_R \rightarrow V_B} \approx \frac{V_B}{nIR_s} = \sqrt{\frac{V_B}{nI_P R}} \quad (2.15)$$

$$I_P = AJ_{ph} \quad (2.16)$$

When the photocurrent is smaller than the dark current, the maximum multiplication is limited by the dark current and is given by $M_{ph}|_{max} = \sqrt{\frac{V_B}{nI_{DR}}}$. Thus, it is important that the dark current is made as low as possible so that it will not limit by $M_{ph}|_{max}$.

2.3.4 Excess Noise

Since avalanche multiplication is statistical in nature, and this random process does not lead to the same multiplication of EHPs at a given distance in the depletion region for every injected photon. This leads to the major drawback of an APD: excess noise, which describes the factor by which the avalanche process increases the noise over that of a perfect noiseless gain mechanism. The noise attributed to the variation in gain is measured by the excess noise factor F , which is the ratio of the mean square of the gain to the square of the mean gain [11]

$$F = \frac{\langle M^2 \rangle}{\langle M \rangle^2} = M \left[1 - (1 - k) \left(\frac{M-1}{M} \right)^2 \right] \approx kM + 2(1 - k) \quad (2.17)$$

Where $k = \frac{\alpha_p}{\alpha_n}$ for electron injection/multiplication and for hole injection this expression still applies if k is replaced by $k' = \frac{\alpha_n}{\alpha_p}$. From this equation, k/k' should be minimised to reduce noise. Qualitatively, $k = 1$ means equal ionization by electrons and holes, and ionized holes can also produce EHPs, and such positive feedback only reinforces the fluctuation noise.

2.4 Single-Photon Avalanche Photodiodes (SPADs)

2.4.1 Basic SPAD Operation

Photon counting is the technique of choice in accurate measurements of weak optical signals and fast light pulses, in the nanosecond and picoseconds range [12]. Among commercially available Photodetectors two types of device could obtain single photon sensitivity: photomultiplier tubes (PMTs) and SPADs. [7] gave a rough history of the development of Geiger-mode avalanche photodiodes. The first PMT was invented in the RCA laboratories and became a commercial product in 1936. Further innovations have led to high sophisticated

devices available nowadays. However, PMTs have two severe problems: They are very sensitive to magnetic fields and their price is high because the complicated mechanical structure inside the vacuum container is mostly handmade. These forced the search for an alternative device to PMTs. PIN photodiodes were very successful used in most big experiments in high-energy physics, but it has not internal gain and the requirement for charge sensitive amplifier make the Signal to Noise Ratio (SNR) suffer. APDs have internal gain which improves the SNR but the gain is not enough and still some 20 photons are needed for a detectable light signal. Furthermore, the sub-Geiger mode APDs lead to poor timing resolution due to the noisy avalanche buildup, and the setup is very susceptible to interference requiring cumbersome shielding [8]. From the year 2000, the Geiger-mode APDs (SPADs) have been developed. SPADs are basically APDs biased above the breakdown voltage in the so called Geiger-mode, where the devices remain at almost zero current state for hundreds of microseconds to milliseconds, until the first minority carrier is generated and triggers the avalanche process. A high current pulse in the milliamperere range, being generated from a single carrier/incident photon, is delivered to the external circuit.

2.4.2 SPAD Structures

Early pioneering work in the development of solid state single photon detectors was carried out in the 1960s by McIntyre (reach-through structure) in the RCA laboratories [13] and by Haitz (planar thin structure) in the Shockley research laboratory [14], shown in Figure 2.7.

The reach-through structure was improved and optimised for low noise with the SLIK™ structure [15] and integrated and commercialised in the SPCM-AQR series by Perkin-Elmer Optoelectronics [16]. Since 1970s, a large variety of thin SPADs fabricated with dedicated planar CMOS processes have been proposed, such as n^+/p substrate with deep diffused n^- guard ring [17], n^+/p^+ enrichment in a p-substrate (virtual guard ring) [18], and patterned double epitaxial (DJ-SPAD)[19], shown in Figure 2.8. Table 2.2 summaries the advantages and disadvantages of both of the thin planar and thick reach-through structures.

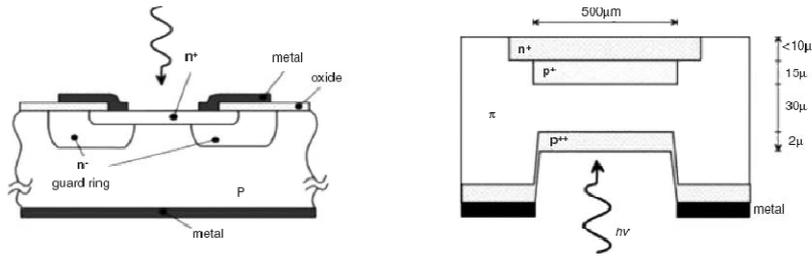


Figure 2.7 Hartz's planar structure (left) and McIntyre's reach-through structure (right)

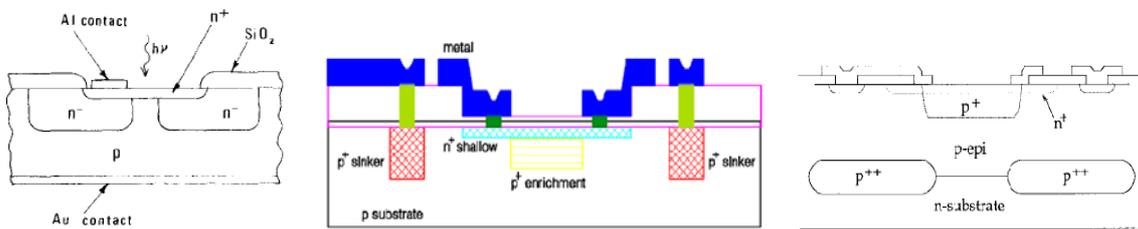


Figure 2.8 n^+/p substrate with deep diffused n^- guard ring (left), n^+/p^+ enrichment in a p -substrate virtual guard-ring (middle) and DJ-SPAD (right)

SPADs	Advantages	Disadvantages
Reach-Through	High photon detection efficiency in the visible and IR region; Large active area	High bias voltage (250-500V); Dedicated fabrication process preventing monolithic integration
Planar	Ultrafast timing resolution; High fabrication yield and low cost, low power consumption; Possibly production for monolithic array for a photon counting system	Low photon detection efficiency at larger wavelengths due to thin depletion layer

Table 2.2 Pros and cons of thin and thick SPADs

An APD in a standard 2 μm Bipolar Complementary Metal-Oxide-Semiconductor (BiCMOS) process has been reported by Bible *et al* [20], shown in Figure 2.9. The device is a p^+/n guard ring diode with an n well forming the n region and source/drain diffusions forming the p^+ region, lighter doped p base material then forms the guard ring.

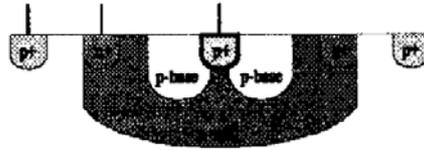


Figure 2.9 BiCMOS SPAD [20]

A low noise APD fabricated in 0.8 μm CMOS technology has been reported by Rochas *et al* [6]. To create a guard ring without additional processing steps, advantage is taken of the lateral diffusion of two n wells designed at a small distance d . The process begins with the oxidation of the p-type silicon wafer and the deposition of silicon nitride. The n-tub mask is then processed, nitride is etched and n-tub implantation is done. Two n-doped regions are separated by a gap d . A thick oxide is then grown by exposing the wafer to pure oxygen at high temperature. During this phase, the donor ions diffuse vertically deeper in the substrate as well as laterally. This is followed by the nitride etching and p-well self aligned implantation. Donor and acceptor ions diffuse deeper during the subsequent high temperature drive-in diffusion. At this step, a guard ring structure with a low doped p-region is available at the centre of the gap. During the following steps of process, contacts for p-substrate and n-well are fabricated as well as the photosensitive p^+ active region. The process steps are illustrated in Figure 2.10.

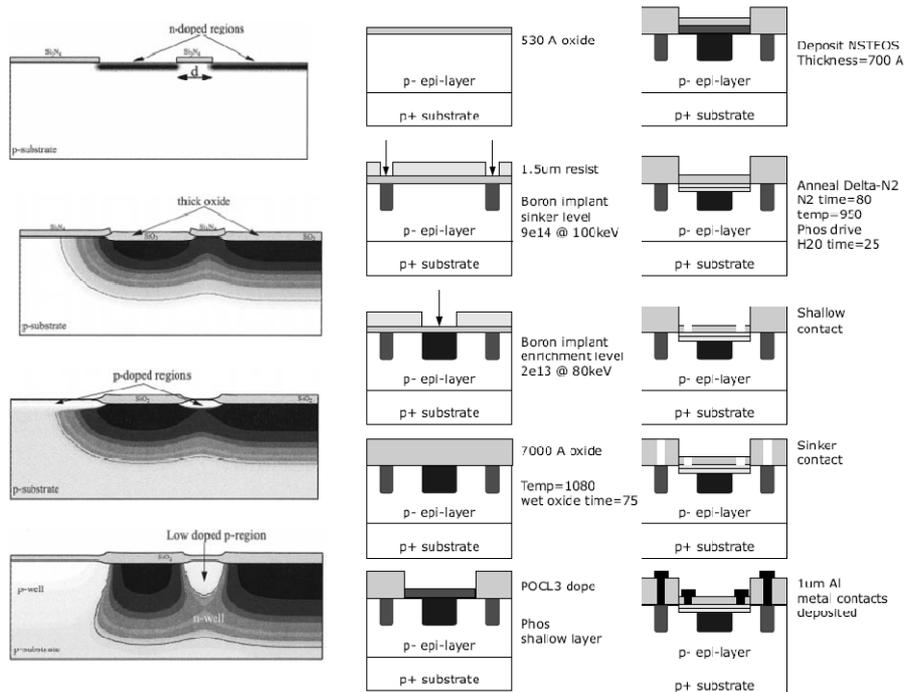


Figure 2.10 Processes for Rochas SPAD [6] and Monloney SPAD [21]

Another APD structure has been reported by Moloney *et al* [21], which allows for the use of SOI technology along with allowing optical and electrical isolation of pixels in an array. The main processing steps of fabricating the diodes are shown in Figure 2.11. A 530Å oxide is thermally grown and boron is subsequently implanted at a dose of $9e14$ atoms/cm² with an energy of 100 keV to form the p^+ sinker (anode). Boron is then again implanted at a dose of $2e23$ atoms/cm² with an energy of 80 keV to form the active area of the device, which is called the p enrichment layer. The pre-implanted thermally grown oxide is important to eliminate the surface damage caused by the implantation step. The enrichment diffusion should have the proper doping profile in order to give the desired breakdown voltage. After the implantation, an annealing step at 1100°C for 120min is performed in nitrogen followed by a diffusion step at 1100°C for 60min in a wet oxygen environment to grow a 7000 Å oxide. This step is followed by the opening of a window, which defines the n^+ area, which is called the shallow region. The shallow layer must overlap the enrichment in order to have a lower electric field outside the active area and provide a virtual guard ring. Then a thermally grown pre-implant oxide is produced. Phosphorous is then implanted either through a sacrificial oxide, or it can be POCL doped after the etch to bare silicon. Thermal growth of a 700 Å oxide follows, after which an annealing step assures repair of the crystal damage caused

during the implantation step. Contact windows are then opened with different masks being used for p^+ sinker and n^+ contacts. Finally, a 1 μm aluminium layer (1% Si) is deposited and developed. The implantation and annealing conditions are chosen to give the proper distribution of boron in order to achieve a low ($< 30\text{ V}$) breakdown voltage. The main oxidation and annealing steps were taken from a 1.5 μm CMOS process flow. The devices have a circular symmetry, in order to eliminate sharp angles that may cause edge breakdown. The substrate material is low dislocation, Czochralski $\langle 100 \rangle$ orientation silicon wafers with an initial epitaxial layer resistivity of 10–16 $\Omega\text{-cm}$ on a highly doped p type substrate. Standard clean room photolithography techniques were used to process the wafers. Five masks have been used for the fabrication of the SPAD, this could however be reduced to four by using the same contacts for the anode and cathode.

Recently, [22] has reported the design of a SPAD fabricated in 0.18 μm IBM CMOS that implements a planar multiplication region with Shallow Trench Isolation (STI) to as guard ring to prevent the edge early breakdown, shown in Figure 2.11.

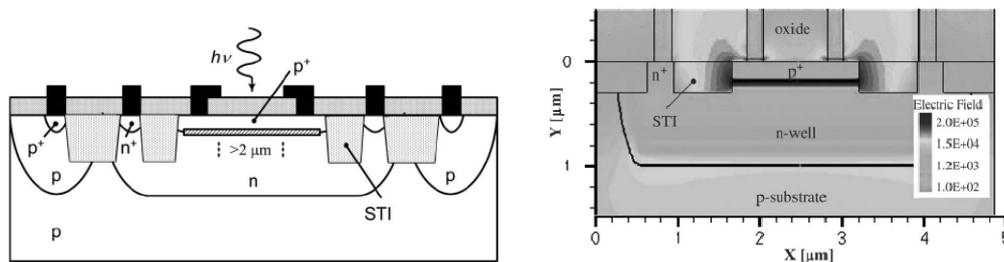


Figure 2.11 STI 0.18 μm CMOS SPAD (left) and the ISE-TCAD device simulation (right) [22]

2.4.3 SPAD Performance Parameters

2.4.3.1 Detection Probability

The detection probability of a SPAD is defined as the probability that an incident photon will produce a detection event, which is the multiplication of quantum efficiency and avalanche probability. The avalanche probability is the probability that the photogenerated electron or hole will initiate an avalanche that does not terminate prematurely. As previously discussed, in silicon, electrons have higher impact ionization coefficient than holes, so the probability of detection is higher for electron-initiated avalanches than for hole-initiated avalanches.

Furthermore, the more the device is biased above the breakdown voltage, the higher the avalanche probability and detection probability.

2.4.3.2 Dark Count Rate

Dark count rate is defined as the number of dark current induced events per second. In a SPAD, a single thermally generated electron or hole can initiate an avalanche, leading to an electrical pulse (false count) that is indistinguishable from a photon absorption.

Whenever the thermal equilibrium condition of a semiconductor system is disturbed ($pn \neq n_i^2$), processes exist to restore the system to equilibrium ($pn = n_i^2$). These processes are recombination when $pn > n_i^2$ and thermal generation when $pn < n_i^2$ [4]. Since under reverse bias the depletion region widens, and there can be significant opportunities for both generation and recombination. However, under reverse bias, n and p are both small (miniscule) in most of the depletion region $pn < n_i^2$, so that thermal generation dominates, and recombination in the depletion region is negligible. Also, under reverse bias, only a small number of carriers are available to diffuse across the junction (once in the junction they drift to the other side). With the increase of reverse bias, band to band tunnelling and carrier multiplication take place and the reverse current increases. As a result, thermal generation and tunnelling currents in the depletion region far exceed the minority carrier diffusion current. The following analysis will be concentrate on these two reverse current mechanisms.

In indirect-bandgap semiconductors such as Si and Ge, the dominant transition are indirect recombination or generation via bulk traps, of density N_t and energy E_t present within the bandgap [4]. As a result, the thermal generation is dominated by trap-assisted thermal generation, though band-to-band thermal generation and tunnelling assisted thermal generation could also occur. In [8], firstly, the net generation rate G is derived in steady state

$$G = N_t \cdot \frac{\langle e_n \rangle \langle e_p \rangle - np \langle c_n \rangle \langle c_p \rangle}{\langle e_n \rangle + \langle e_p \rangle + n \langle c_n \rangle + p \langle c_p \rangle} \quad (2.18)$$

where N_t is the density of traps, $\langle e_n \rangle$ and $\langle e_p \rangle$ are average rate of electron and hole to escape from a trap respectively, and $\langle c_n \rangle$ and $\langle c_p \rangle$ are average rate of electron and hole to be captured by a trap. In thermal equilibrium, $\langle c_n \rangle$ and $\langle e_n \rangle$, $\langle c_p \rangle$ and $\langle e_p \rangle$ are linked by

$$\langle e_n \rangle = n \langle c_n \rangle \exp \left[\frac{E_t - F_n}{kT} \right] = n_i \langle c_n \rangle \exp \left[\frac{E_t - E_i}{kT} \right] \quad (2.19)$$

$$\langle e_p \rangle = p \langle c_p \rangle \exp \left[\frac{F_p - E_t}{kT} \right] = n_i \langle c_p \rangle \exp \left[\frac{E_t - E_i}{kT} \right] \quad (2.20)$$

with F_n the quasi-Fermi level for electrons, E_i the intrinsic Fermi level, E_t the energy level of the trap and n and p is give by

$$n = n_i \exp \left[\frac{F_n - E_i}{kT} \right] \quad (2.21)$$

$$p = n_i \exp \left[\frac{E_i - F_p}{kT} \right] \quad (2.22)$$

The SPAD dark count rate caused by thermal generation is obtained by integrate over the depletion region the product of net generation rate and avalanche probability.

The second step is to including the field dependency of the generation rate. To include tunnelling, Hurkx introduced field enhancement factors in the standard SRH model (Hurkx Tunneling SRH model in Sentaurus Device simulation). In this model, the escape rates are modified and multiplied by $1 + \Gamma_{n,p}$, $\Gamma_{n,p}$ is trap-assisted tunnelling factor.

The field enhanced trap assisted tunnelling (TAT) SRH generation rate is give by

$$G_{SRH(Hurkx(tunneling))} = \frac{n_i}{\frac{\tau_n}{1+\Gamma_n} \exp \left[\frac{E_i - E_t}{kT} \right] + \frac{\tau_p}{1+\Gamma_p} \exp \left[\frac{E_t - E_i}{kT} \right]} \quad (2.23)$$

where τ_n and τ_p are the lifetimes of electrons and holes respectively.

For SPAD of surface area S , the resulting dark count rate $DRC_{SRH(Hurkx(tunneling))}$ is give by

$$DRC_{SRH(Hurkx(tunneling))} = S \int_{z_0}^{z_w} P_p(z) G_{SRH(Hurkx(tunneling))} dz \quad (2.24)$$

The dark count rate caused by thermal generation in the depletion region is dependent on the fabrication process. Trap concentration, energy levels of the different traps in the forbidden band, ability for the traps to catch and release carriers all influence the final dark count rate. It is also dependent on the surface area, temperature, and excess bias voltage ($P_p(z)$ and depletion width). Traps located in the high field region contribute most to DCR.

Band-to-band tunnelling should be combined with SRH trap-assisted thermal generation at high field. Hurkx model (band-to-band tunnelling model used in Sentaurus Device simulation) is commonly used for the tunnelling generation rate in silicon devices. Then the total DRC is the sum of two contributions: $SRH(Hurkx(tunneling))$ and $Band2band(Hurkx)$.

2.4.3.3 Afterpulsing

The carriers travel through the depletion region during a Geiger-pulse may be caught for a finite time by trapping centres and released after the recharge of the SPAD, which may trigger an avalanche breakdown and induce an unwanted pulse called an afterpulse.

2.4.3.4 Timing Jitter

There is a statistical variation of the time interval between the pulse arrive and the resulting electrical signal from the SPAD, which is called timing jitter. There are four sources of timing jitter for a SPAD [2]. First, the photon detection probability is less than unity and the transmitted optical pulse has finite time duration. The SPAD can detect a photon at the leading edge, the middle, and the trailing edge of the pulse. Second, the photogenerated electron requires finite time to drift from the generation point to the high field region where avalanche occurs. Depending on how deep into the detector the photo propagates before it is absorbed, the photogenerated electron may have a short or longer drift delay. The third contribution to timing jitter arises from statistical variation in the time it takes the avalanche current to grow to its resistance-limited value and the fourth contribution is the thermal noise due to SPAD resistance.

2.4.4 SPAD Design Considerations

In [7] and [8], choices of parameters are concluded in the design of the device:

- Semiconductor material: Influence the photodetection efficiency and the range of wavelength to be detected
- p-silicon on n-substrate: highest detection efficiency for blue light
- n-silicon on p-substrate: highest detection efficiency for green light
- Layer thickness: range of wavelengths, optical crosstalk
- Doping concentrations: operating voltage and its range
- Impurities and crystal defects: dark count rate, afterpulses
- Area: gain, fill factor, dynamic range, recovery time

In particular, some of the technological parameters are of interest. A decrease of junction depth is beneficial for the collection of photons with shorter penetration depths (UV/blue photons). For p^+n junction, if n-well donor concentration is kept constant, a shift of the depletion region towards the surface will decrease the efficiency for red/IR photon collection, but is better for collection of UV photons. A lower n-well active donor concentration allows a better collection of IR photons, but the opposite for UV photons. Thus, a large spectral range of detection is obtained with a junction depth as close as possible to the surface and a relatively low n-well doping level. There is also a trade-off between timing resolution and IR detection efficiency: a small confined volume of multiplication depletion region will result excellent timing resolution. The tunnelling contribution is a physical limitation, and can be neglected at low n-well doping but becomes dominant with high n-well doping. A low n-well doping helps to keep both thermal and tunnelling generations at a low level.

2.5 Conclusion

Many of the fundamental physical theory and concepts associated with APDs and SPADs have been introduced, from the operation of the devices to the different characteristics and performance parameters that can be applied to the two types of devices. Different device structures were reviewed as well as some of the novel fabrication process, in order to get a comprehensive feeling of the process evolution and design trend. It was essential that these basic concepts were well understood before the issue of high DCR/dark current could be addressed through TCAD simulations and process optimization. The factors affecting DCR were studied, as well as the physical models built to describe those factors.

The chapter ended with some general design considerations with an emphasis on how the modification of parameters would affect the device performance.

Chapter 3

The TCAD Simulations

3.1 Introduction

Technology Computer-Aided Design (TCAD) today becomes an extremely important research and development activity in the semiconductor industry. A modern integrated circuit cannot be developed without the massive use of computer aided design (CAD) in any step of the complex flow from the idea to the final product. The use of TCAD makes it possible to explore technologies and design concepts virtually: Firstly, it models the complex flow of semiconductor device fabrication steps and generates detailed geometric information and doping profile distribution of the device. Secondly, it uses the information of the first step to predict the device characteristics, which could be extracted to device circuit models to be implemented in any circuit simulator. A complete TCAD simulation involves the following steps:

- Device generation either using process simulator to virtually fabricate the device or using device structure editor to draw the device
- Mesh creation suitable for device simulation
- Device simulation that solves the mathematical equations with specific physical models to describe the device characteristics and behaviour
- Visualisation of all the simulation results by generating plots and diagrams

In this chapter each of these TCAD simulation steps are described, with the TCAD software suite Sentaurus™ by Synopsys (former ISE AG) including the main simulators DIOS, Sentaurus Process (process simulators) and Sentaurus Device (device simulator). Since the simulation flow is just one of the multiple possible implementations, some parts cannot be generalized and applied to other TCAD installations such as TSUPREM4 and MEDICI.

3.2 Device Generation

The first step of TCAD simulation flow is to generate a device structure that is suitable for device simulation, which means that the device structure should be described by its boundary and materials. There are two methods to perform this step: device structure edition and process simulation. The former approach uses a device editor such as Mdraw (ISE-TCAD) or Sentaurus Structure Editor (Synopsys) to draw the device either by script command or mouse. The second method uses process simulators such as Dios (ISE-TCAD) or Sentaurus Process (Synopsys) to simulate the full device fabrication flow in certain level of detail.

3.2.1 Device Structure Editor

The generation and edition of a device using a device structure editor is to draw the structure of the device, either interactively using a graphical user interface (GUI) or in batch mode using scripts. Several basic geometrical elements are available like rectangles and lines in 2D and cuboids and cylinder in 3D. Each of these geometrical elements is defined in terms of materials (silicon, poly-silicon, oxide, aluminium, etc). Doping concentrations are defined using analytical functions. In Mdraw (ISE-TCAD) an embedded TCL interpreter and a script can control the program. The embedded interpreter makes it easy to generate parameterized structures such as varying gate length. Another new Sentaurus Structure Editor (Synopsys) even provides more functions such as externally generating doping profiles, defining local refinements and inclusion of external submeshes. There is also a process emulator Procem with Sentaurus Structure Editor, which can provide scripting functions to emulate TCAD process steps.

The advantage with this approach is that the fabrication process is not needed, thus new device concepts can be tested. It is also possible to include Secondary Ion Mass Spectrometry (SIMS) concentration depth profiles in this approach and further improve the model and data from measurements. But the disadvantage is that the generated concept structure may not be possible or very hard to fabricate with particular process.

3.2.2 Process Simulation

Process simulation models every process step of a silicon wafer from the start of processing until the electrical test. The level of detail of input parameters to the process simulator

heavily depends on complexity of the physical model implemented in the process simulator. The developments of models for each process steps are separate fields of research. The process simulation flow consists of a sequence of single commands for the process simulator, such as the type of process steps (diffusion, etching, implantation, etc) and process conditions in terms of parameters (time, temperature, pressure, etc). These parameters are called input parameters. A CMOS example of the syntax of such a process flow is illustrated in Appendix I. Advanced calibration of the models for a certain range of input parameters such as temperature is also available, and this leads to the second parameters called model parameters, which describe the underlying physics of the process simulator model. The TCAD software takes the appropriate process models and adopts them to fit the constraints of the process simulator. The mesh for solving the partial-differential equations typical for the physical and chemical processes occurring during the processing is normally of unstructured type, to model the steep gradients of the doping distributions with good accuracy, but with a low number of mesh points where physical fields (doping concentration, point defect concentration, etc) are not varying that much. Detailed description of process simulators such as DIOS and Sentaurus Process can be found in [23] [24].

There are several types of applications of process simulation, and the one used in this project is illustrated in Figure 3.1. In this application, a previously installed process flow is simulated to generate the electrical parameters of the device by a process simulation followed by a device simulation. The objective of this type of application is to optimize the process in terms of device parametric performance. Furthermore, in terms of process complexity versus cost, simulations are carried out to optimize an already implemented process further. In addition the statistical sensitivity of device parameters such as breakdown voltage, dark current or threshold voltage on certain process flow parameters like implantation energy or dose of a particular implant step could be calculated.

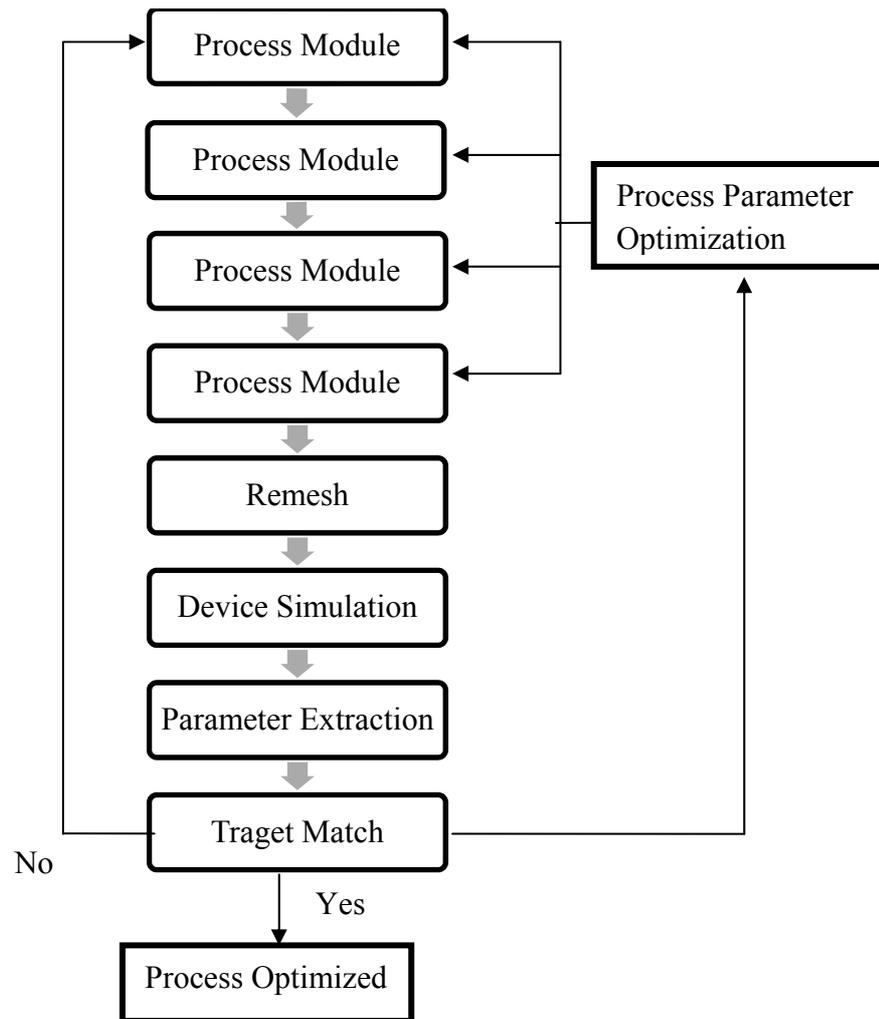


Figure 3.1 A common process simulation strategy

3.3 Refinement for Device Simulation

In order to perform a device simulation, which is based on finite element method (FEM), a specific suitable mesh is needed. The requirements on different grids of process and device simulations are very different. The process simulation grid has to follow steep gradients in the doping profile and must resolve internal interfaces such as Si/SiO₂ with accuracy sufficient to model segregation and dopant transportation across interfaces. And it must be able to adapt to changes in the boundary during process steps such as etching or oxidation. On the other hand, the device simulation grid must resolve mainly physical quantities during the device simulation, such as carrier concentrations (not doping concentration) and electrostatic potential. Therefore, the grid obtained from the process simulation is not suitable

to get accurate results for device simulations, and a remesh/refinement of the structure is necessary. The boundaries where currents occur during the device operation have to be represented appropriately by the mesh. A proper grid for carrier transport simulation should follow the current flows during device operation, such as p-n junctions of a diode, channel in a MOSFET device. A general and automatic approach is to use a two-step strategy to generate a suitable device simulation grid. First, a coarse grid is used with the device simulator to obtain a coarse representation of the physical fields in the device, which acts the refinement criteria for a second iteration of the grid remesh. This method can lead to faster device simulation in big 2D grids or 3D grids because of improved convergence.

3.4 Device Simulation

A device simulation provides information about the inner conditions and the terminal characteristics of the device and predicts the behaviour of the device. The main inputs for device simulation are:

- Doping concentration of different doping species on a mesh (.dat file in DIOS or .tdr file in Sentaurus Process)
- Structure boundary information of the region which is evaluated with device simulation, including material types and detailed surface/interface shapes (.grd file in DIOS or .tdr file in Sentaurus Process)
- Contacts information define in process simulation file

A Sentaurus Device Syntax example is in Appendix II.

The numerical approach of device simulation could be applied in both 2D and 3D, but the memory consumption for a 3D simulation is much larger compared to a 2D simulation. A reasonable estimation is that the memory consumption is around 300MB~500MB for 2D simulation compared with 1GB~2GB for a 3D simulation

3.4.1 Carrier Transport Models

Physical phenomena in semiconductor devices are very complicated and, depending on applications, are described by partial differential equations of different level of complexity.

Coefficients and boundary conditions of equations such as mobility, generation-recombination rate, material-dependent parameters, interface and contact boundary conditions, can be very complicated and depending on microscopic physics, the structure of the device, and the applied bias [23]. The TCAD device simulation solves these partial differential equations describing the electrical potential distribution and carrier transportation in a doped semiconductor material. A group of carrier transport equations can be combined to form a simulation model depending on the device under investigation and the modelling accuracy required. In Sentaurus Device, three different simulation models can be selected:

- Drift-Diffusion (DD): Isothermal simulation, described by basic semiconductor equations. Suitable for low-power density devices with long active regions
- Thermodynamic (TD): Accounts for self-heating. Suitable for devices with low thermal exchanges, particularly, high-power density devices with long active regions
- Hydrodynamic (HD): Accounts for energy transport of the carriers. Suitable for devices with small active regions

The basic physics for a carrier transport are covered by the Poisson equation and the continuity equations for electrons and holes:

$$\nabla \cdot \varepsilon \nabla \phi = -q(p - n + N_D - N_A) - \rho_{trap} \quad (3.1)$$

$$\nabla \cdot \vec{J}_n = qR_{net} + q \frac{\delta n}{\delta t} \quad (3.2)$$

$$-\nabla \cdot \vec{J}_p = qR_{net} + q \frac{\delta p}{\delta t} \quad (3.3)$$

where ε is the electrical permittivity, q is the elementary electronic charge, n and p are the electron and hole densities, N_D and N_A are the concentrations of active (ionized) donors and acceptors, ρ_{trap} is the charge density contribute by traps and fixed charges, R_{net} is the net electron-hole recombination rate, \vec{J}_n and \vec{J}_p are the electron and hole current densities respectively. The different carrier transport models describe the current densities in different forms. However, the Poisson and the continuity equations have to be solved for all of them.

3.4.1.1 The Drift-Diffusion Model

The most widely used approach to model the current densities is the drift-diffusion model [25]. The assumptions of the simplified drift-diffusion model are:

- Full ionization: all dopants are assumed to be ionized (shallow dopants)
- Non-degenerate: the Fermi energy is assumed to be at least $3kT$ below/above the conduction/valence band edge
- Steady state: all variables are independent of time
- Constant temperature: the temperature is constant throughout the device

The drift-diffusion equations are:

$$\vec{J}_n = -nq\mu_n\nabla\Phi_n = -nq\mu_n\vec{F} + qD_n\nabla n \quad (3.4)$$

$$\vec{J}_p = -pq\mu_p\nabla\Phi_p = -pq\mu_p\vec{F} - qD_p\nabla p \quad (3.5)$$

where μ_n and μ_p are the electron and hole mobilities, q is the elementary charge, n and p describe the electron and hole densities, and Φ_n and Φ_p are the electron and hole quasi-Fermi potentials respectively. \vec{F} is the electric field vector, D_n and D_p are the diffusion constants for electrons and holes respectively. These equations indicate that the current consists of a field driven drift part and a carrier concentration driven diffusion part.

3.4.1.2 The Thermodynamic Model

The thermodynamic model extends the drift-diffusion model with a temperature dependent equation to account for electrothermal effects, under the assumption that the charge carriers are in thermal equilibrium with the lattice. The carrier temperatures and the lattice temperature are described by a single temperature T :

$$\vec{J}_n = -nq\mu_n(\nabla\Phi_n + P_n\nabla T) \quad (3.6)$$

$$\vec{J}_p = -pq\mu_p(\nabla\Phi_p + P_p\nabla T) \quad (3.7)$$

P_n and P_p are the absolute thermoelectric power of electrons and holes, and they are doping dependent coefficients. To model the thermal transport, an additional equation is added [25]:

$$c_L \frac{\delta T}{\delta t} - \nabla \cdot \kappa \nabla T = -\nabla \cdot [(P_n \nabla T + \Phi_n) \vec{J}_n + (P_p \nabla T + \Phi_p) \vec{J}_p] - (E_c + \frac{3}{2} kT) \nabla \cdot \vec{J}_n - (E_v - \frac{3}{2} kT) \nabla \cdot \vec{J}_p + q R_{net} (E_c - E_v + 3kT) \quad (3.8)$$

c_L is the lattice heat capacitance, κ is the thermal conductivity of the material, E_c and E_v are the conduction and valance band energies respectively, and k is the Boltzmann constant.

3.4.1.3 The Hydrodynamic Model

With continued scaling into the deep submicron regime, neither internal nor external characteristics of state-of-the-art semiconductor devices can be described properly using the conventional drift-diffusion transport model. In particular, the drift-diffusion approach cannot reproduce velocity overshoot and often overestimates the impact ionization generation rates. The Monte Carlo method for the solution of the Boltzmann kinetic equation is the most general approach, but because of its high computational requirements, it cannot be used for the routine simulation of devices in an industrial setting [25]. In literature this model is known as energy balance model, and the term hydrodynamic exists in context with more complex models which apply the method of moments. Detailed derivation of the energy balance model from the hydrodynamic model can be found in [26].

In the hydrodynamic model the carrier temperatures T_n and T_p are treated independently from the lattice temperature T_l . The model consists of the basic semiconductor equations, Poisson and continuity equations for electrons and holes, and the energy conservation equations for electrons, holes and lattice.

The current densities are modelled as follows [25]:

$$\vec{J}_n = q\mu_n(n\nabla E_c + kT_n\nabla n + f_n^{td}kn\nabla T_n - \frac{3}{2}nkT_n\nabla \ln m_n) \quad (3.9)$$

$$\vec{J}_p = q\mu_p(p\nabla E_c + kT_p\nabla p + f_p^{td}kp\nabla T_p - \frac{3}{2}pkT_p\nabla \ln m_p) \quad (3.10)$$

The first term takes into account the contribution due to the spatial variations of electrostatic potential, electron affinity, and the band gap. The three remaining terms take into account the contribution due to the gradient of concentration, the carrier temperature gradients, and the spatial variation of the effective masses m_n and m_p .

The coupling between carriers and lattice is done over the energy balance equations:

$$\frac{\delta W_n}{\delta t} + \nabla \cdot \vec{S}_n = \vec{J}_n \cdot \nabla E_c + \left. \frac{\delta W_n}{\delta t} \right|_{coll} \quad (3.11)$$

$$\frac{\delta W_p}{\delta t} + \nabla \cdot \vec{S}_p = \vec{J}_p \cdot \nabla E_v + \left. \frac{\delta W_p}{\delta t} \right|_{coll} \quad (3.12)$$

$$\frac{\delta W_L}{\delta t} + \nabla \cdot \vec{S}_L = \left. \frac{\delta W_L}{\delta t} \right|_{coll} \quad (3.13)$$

W_n , W_p and W_L are the energy densities while \vec{S}_n , \vec{S}_p and \vec{S}_L are the energy transport terms for electrons, holes and lattice respectively. The terms $\left. \frac{\delta W_n}{\delta t} \right|_{coll}$, $\left. \frac{\delta W_p}{\delta t} \right|_{coll}$ and $\left. \frac{\delta W_L}{\delta t} \right|_{coll}$ are the energy density changes which result from the interaction between electrons and holes with each other and with the lattice. The energy balance equations can be interpreted as follows:

“The energy density change of the carriers plus spatial energy flux of those carriers (LHS) equals to the heat generated or dissipated by the carriers while they flow along a field gradient and a collision term where the interaction between electrons and holes with each other and their coupling with the lattice are taken into account (RHS).”

In the energy balance model, the energy conservation must be met. The only way energy can leave the device is through optical emission, current flowing through the contacts and heat flux through the thermodes. The energy transport terms can be written as follows:

$$\vec{S}_n = -\frac{5r_n}{2} \left(\frac{kT_n}{q} \vec{J}_n + f_n^{hf} \hat{\kappa}_n \nabla T_n \right) \quad (3.14)$$

$$\vec{S}_p = -\frac{5r_p}{2} \left(\frac{kT_p}{q} \vec{J}_p + f_p^{hf} \hat{\kappa}_p \nabla T_p \right) \quad (3.15)$$

$$\vec{S}_L = -\kappa_L \nabla T_L \quad (3.16)$$

where

$$\hat{\kappa}_n = \frac{k^2}{q} n \mu_n T_n \quad (3.17)$$

$$\hat{\kappa}_p = \frac{k^2}{q} p \mu_p T_p \quad (3.18)$$

f_n^{hf} and f_p^{hf} are the heat flux parameters, r_n and r_p are the energy flux parameters, for electrons and holes respectively. These are Sentaurus Device parameters which can be set

according to different publication on this topic. Detailed derivation and model parameters could found in [25].

One of the important application of this model is that the simulation of breakdowns of a device. The use of the hydrodynamic transport model avoids the onset of premature breakdowns due to the local field assumed in the drift-diffusion model.

3.4.2 Mobility Models

The device simulator uses a modular approach for the description of carrier mobilities, which is influenced by several parameters such as temperature, doping, carrier-carrier scattering, high injection [25].

For high electric fields the mobility becomes field dependent. An effective field computed by the carrier temperature is used as the driving force for the mobility degradation. The driving force model is required for hydrodynamic simulation and the effective field/driving force for electrons and holes are as follows:

$$F_{hfs,n} = \sqrt{\frac{\max(w_n - w_0, 0)}{\tau_{e,n} q \mu_n}} \quad (3.19)$$

$$F_{hfs,p} = \sqrt{\frac{\max(w_p - w_0, 0)}{\tau_{e,p} q \mu_p}} \quad (3.20)$$

Where $w_{n,p} = 3kT_{n,p}/2$ is the average electron/hole thermal energies while $w_0 = 3kT_L/2$ is the equilibrium thermal energy. $\tau_{e,n}$ and $\tau_{e,p}$ are the energy relaxation times for electrons and holes respectively. Lowering the energy relaxation time is effectively increasing the driving force. The effective field/driving force is then inserted into the basic Canali model to calculate its effect on the mobility [25]:

$$\mu(F) = \frac{(\alpha+1)\mu_{low}}{\alpha + \left[1 + \left(\frac{(\alpha+1)\mu_{low} F_{hfs}}{v_{sat}}\right)^2\right]^{1/\beta}} \quad (3.21)$$

μ_{low} is the low field mobility while v_{sat} is the saturation velocity. β takes the into account the temperature dependence while other parameters are fitting parameters for different materials.

3.4.3 Carrier Generation

Avalanche generation is a field driven effect and particularly important for simulations with high field injection. EHPs produced due to avalanche generation (impact ionization) require certain threshold field strength and the possibility of acceleration (wide space charge regions). If the width of a space charge region is greater than the mean free path between two ionizing impacts, charge multiplication occurs, which can cause electrical breakdown. The reciprocal of the mean free path is the ionization coefficient α , as discussed in Chapter 2. With these coefficients for electrons and holes, the generation rate can be expressed as:

$$G^{ii} = \alpha_n n v_n + \alpha_p p v_p \quad (3.22)$$

In the hydrodynamic transportation framework the driving force for the avalanche generation is the carrier temperature, which is translated into an effective electric field used for conventional impact ionization models.

Different impact ionization models of the threshold behaviour of impact ionization coefficients have been implemented in the Sentaurus Device framework: van overstraeten-de Man, Okuto-Crowell, Lackner, Univeristy of Bologna. In the van Overstraeten-de Man model, the impact ionization coefficient is defined by [25]:

$$\alpha(F_{ava}) = \gamma a \exp\left(-\frac{\gamma b}{F_{ava}}\right) \quad (3.23)$$

with

$$\gamma = \frac{\tanh\left(\frac{\hbar w_{op}}{2kT_0}\right)}{\tanh\left(\frac{\hbar w_{op}}{2kT}\right)} \quad (3.24)$$

The factor γ contains the optical phonon energy term $\hbar w_{op}$ and models the temperature dependence of the phonon gap with the carriers. The parameters a and b are material dependent fitting parameters. Same as discussed in Chapter 2, a higher effective field increases the impact ionization coefficient and the avalanche generation. The van Overstraeten and de Man model are applicable over the range of fields $1.75 \times 10^5 Vcm^{-1}$ to $6 \times 10^5 Vcm^{-1}$.

3.4.4 Carrier Recombination

3.4.4.1 Radiative Recombination

The radiative recombination of an electron and a hole results in the emission of a photon. There are two different recombination mechanisms:

- Spontaneous recombination of an electron in the conduction band and a hole in the valence band with the emission of a photon
- Photon stimulated recombination of an electron in the conduction band and a hole in the valence band which results in the emission of an additional photon

3.4.4.2 Non-Radiative Recombination

There are two typical non-radiative recombination mechanisms:

- Shockley-Read-Hall (SRH) recombination is the trap-level supported recombination effect of a hole and an electron. It requires energy levels deep within the band gap of the semiconductor caused by lattice defects
- For Auger recombination, the excess energy of the EHP recombination is transferred to another electron/hole in the valence/conduction band

Field enhancement reduces SRH recombination lifetimes in regions of strong electric fields. It must not be neglected if the electric field exceeds a value of approximately 3×10^5 V/cm in certain region of the device. For example, the I-V characteristics of reverse-biased p-n junctions are extremely sensitive to defect-assisted tunnelling, which causes EHP generation before band-to-band tunnelling or avalanche generation sets in. The Hurkx TAT model discussed in chapter 2 is used for device simulation.

3.5 Other Simulations

There are some other types of simulations in addition the standard DC simulation discussed in the previous device simulation section. These special simulations are: AC simulation, transient simulation, mixed-mode simulation, statistical simulation.

3.6 Conclusion

Most of the aspects of TCAD simulations were introduced, from device generation to detailed models for device simulation. It is mandatory to understand the methods of mesh generation, refinement, and the physical models to be included for the device simulation, since all these factors would affect the simulation results and accuracy. Even most of the physical models were included, the simulation solver still won't converge due to various problems, and iteration is needed. There exist similar models to describe the same physical effect, and it has to be decided which model would be more suitable for a particular device under investigation.

References

- [1] Alessandro Spinelli and Andrea L. Lacaita, “Physics and Numerical Simulations of Single Photon Avalanche Diodes”, IEEE Journal of Solid-State Circuits, Volume 40, Number 9, September 2005
- [2] Brian F. Aull, Andrew H. Loomis, Douglas J. Young, Richard M. Heinrichs, Bradley J. Felton, Peter J. Daniels, and Deborah J. Landers, “Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging”, MIT Lincoln Laboratory Journal, Volume 13, Number 2, 2002
- [3] Joe N. Haralson, II and Kevin F. Brennan, “Novel Edge Suppression Techniques for Planar Avalanche Photodiodes”, IEEE Journal of Quantum Electronics, Volume 35, Number 12, December 1999
- [4] S. M. Sze and K. K. Ng, “Physics of Semiconductor Devices”, 3rd Edition, John Wiley & Sons Inc, Nov 2006, ISBN 978-0-471-14323-9
- [5] K. K. Ng, “Complete Guide to Semiconductor Devices”, 2nd Edition, Wiley-IEEE Press, June 2002, ISBN 978-0-471-20240-0
- [6] A. Rochas, A. R. Pauchard, P-A. Besse, D. Pantic, Z. Prijic, R .S. Popovic, “Low-Noise Silicon Avalanche Photodiodes Fabricated in Conventional CMOS Technologies”, IEEE Transactions on Electron Devices, Volume 49, Number 3, March 2002
- [7] D. Renker, “Geiger-Mode Avalanche Photodiodes, History, Properties and Problems”, Nuclear Instruments and Methods in Physics Research, Section A, 567, June 2006
- [8] A. Rochas, “Single Photon Avalanche Diode in CMOS Technology”, Ph.D. Thesis, EPFL, 2003
- [9] S. O. Kasap, “Optoelectronics and Photonics – principles and Practices”, Prentice Hall, 2001

- [10] J. N. Haralson and K. F. Brennan, "Novel Edge Suppression Technique for Planar Avalanche Photodiodes", IEEE Journal of Quantum Electron, Volume 35, Number 12, December 1999
- [11] R. J. McIntyre, "Multiplication Noise in Uniform Avalanche Diodes", IEEE Transactions on Electron Devices, Volume ED-13, Number 1, January 1966
- [12] A. Lacaita, S. Cova, M. Ghioni, and F. Zappa "Single-Photon Avalanche Diode with Ultrafast Pulse Response Free from Slow Tails", IEEE Electron Device Letters, Volume 14, Number 7, July 1993
- [13] R. J. McIntyre, "Theory of Microplasma Instability in Silicon", Journal of Applied Physics, Volume 32, Issue 6, pp. 983-995, June 1961
- [14] R. H. Haitz, "Model for the Electrical Behavior of a Microplasma", Volume 35, Issue 5, pp. 1370-1376, May 1964
- [15] H. Dautet, Pierre Deschamps, Bruno Dion, Andrew D. MacGregor, Darleene MacSween, Robert J. McIntyre, Claude Trottier, and Paul P. Webb, "Photon Counting Techniques with Silicon Avalanche Photodiodes", Applied Optics, Volume 32, Number 21, July 1993
- [16] http://optoelectronics.perkinelmer.com/content/Datasheets/DTS_SPCMAQR.pdf
- [17] S. Cova, A. Longoni, and A. Andreoni, "Towards Picosecond Resolution with Single-Photon Avalanche Diodes", Review of Scientific Instruments, 52(3), March 1981
- [18] F. Zappa, M. Ghioni, S. Cova, L. Varisco, B. Sinnis, A. Morrison, and A. Mathewson, "Integrated Array of Avalanche Photodiodes for Single-Photon Counting", IEEE Proceeding of the 27th European Solid-State Device Research Conference, September 1997.
- [19] A. Spinelli, M. Ghioni, S. Cova, and L. M. Davis, "Avalanche Detector with Ultraclean Response for Time-Resolved Photon Counting", IEEE Journal of Quantum Electronics, Volume 34, Number 5, May 1998

- [20] A. Biber, P. Seitz, and H. Jäckel, “Avalanche Photodiode Image Sensor in Standard BiCMOS Technology”, IEEE Transactions on Electron Devices, Volume 47, Number 11, November 2000
- [21] A. M. Moloney, A. P. Morrison, J. C. Jackson, A. Mathewson, J. Alderman, J. Donnelly, B. O’Neill, A. M. Kelleher, G. Healy, and P. J. Murphy, “Monolithically Integrated Avalanche Photodiode and Transimpedance Amplifier in a Hybrid Bulk/SOI CMOS process”, Electronics Letters, Volume 39, Number 4, February 2003
- [22] H. Finkelstein, M. J. Hsu, and S. C. Esener, “STI-Bounded Single-Photon Avalanche Diode in a Deep-Submicrometer CMOS Technology”, IEEE Electron Device Letters, Volume 27, Number 11, November 2006
- [23] Synopsys, Mountain View, California, DIOS
http://www.synopsys.com/products/tcad/dios_ds.html
- [24] Synopsys, Mountain View, California, Sentaurus Process
http://www.synopsys.com/products/tcad/sentaurus_procs_ds.html
- [25] Synopsys, Mountain View, California, Sentaurus Device Userguide
- [26] A. Schenk, “Semiconductor Devices – Physical Principles and Simulations”, ETH Zurich, 2004

Appendix II

Sentaurus Process Syntax

```
#-----
# 2D nMOSFET (0.18um technology)
#-----

#--- Declare initial grid (half structure) -----
line x location= 0.0      spacing= 1.0<nm>  tag=SiTop
line x location=50.0<nm> spacing=10.0<nm>
line x location= 0.5<um> spacing=50.0<nm>
line x location= 2.0<um> spacing= 0.2<um>
line x location= 4.0<um> spacing= 0.4<um>
line x location=10.0<um> spacing= 2.0<um>  tag=SiBottom

line y location=0.0      spacing=50.0<nm>  tag=Mid
line y location=0.40<um> spacing=50.0<nm>  tag=Right

#--- Silicon substrate definition -----
region silicon xlo=SiTop xhi=SiBottom ylo=Mid yhi=Right

#--- Initialize the simulation -----
init concentration=1.0e+15<cm-3> field=Phosphorus

#--- p-well, anti-punchthrough & Vt adjustment implants -----
implant Boron dose=2.0e13<cm-2> energy=200<keV> tilt=0 rotation=0
implant Boron dose=1.0e13<cm-2> energy= 80<keV> tilt=0 rotation=0
implant Boron dose=2.0e12<cm-2> energy= 25<keV> tilt=0 rotation=0

#--- p-well: RTA of channel implants -----
diffuse temperature=1050<C> time=10.0<s>

#--- Saving structure -----
struct tdr=NMOS1 ; # p-Well

#--- MGOALS settings for automatic meshing in newly generated layers -
mgoals on min.normal.size=1<nm> max.lateral.size=2.0<um> \
normal.growth.ratio=1.4 accuracy=2e-5

#--- Gate oxidation -----
diffuse temperature=850<C> time=10.0<min> O2
grid remesh
select z=Boron
layers
struct tdr=NMOS2 ; # GateOx

#--- Poly gate deposition -----
deposit poly type=anisotropic thickness=0.18<um>
#--- Poly gate pattern/etch -----
mask name=gate_mask left=-1 right=90<nm>
etch poly type=anisotropic thickness=0.2<um> mask=gate_mask
etch oxide type=anisotropic thickness=0.1<um>
struct tdr=NMOS3 ; # PolyGate

#--- Poly reoxidation -----
```

```

diffuse temperature=900<C> time=10.0<min> O2 pressure=0.5<atm> \
    mgoals.native
struct tdr=NMOS4 ; # Poly Reox

#--- LDD implantation -----
refinebox silicon min= {0.0 0.05} max= {0.1 0.12} \
    xrefine= {0.01 0.01 0.01} yrefine= {0.01 0.01 0.01} add
refinebox remesh

implant Arsenic dose=4e14<cm-2> energy=10<keV> tilt=0 rotation=0
diffuse temperature=1050<C> time=0.1<s> ; # Quick activation
struct tdr=NMOS5 ; # LDD Implant

#--- Halo implantation: Quad HALO implants -----
implant Boron dose=0.25e13<cm-2> energy=20<keV> \
    tilt=30<degree> rotation=0
implant Boron dose=0.25e13<cm-2> energy=20<keV> \
    tilt=30<degree> rotation=90<degree>
implant Boron dose=0.25e13<cm-2> energy=20<keV> \
    tilt=30<degree> rotation=180<degree>
implant Boron dose=0.25e13<cm-2> energy=20<keV> \
    tilt=30<degree> rotation=270<degree>

#--- RTA of LDD/HALO implants -----
diffuse temperature=1050<C> time=5.0<s>
struct tdr=NMOS6 ; # Halo RTA

#--- Nitride spacer -----
deposit nitride type=isotropic thickness=60<nm>
etch nitride type=anisotropic thickness=84<nm>
etch oxide type=anisotropic thickness=10<nm>
struct tdr=NMOS7 ; # Spacer

#--- N+ implantation -----
refinebox silicon min= {0.04 0.05} max= {0.18 0.4} \
    xrefine= {0.01 0.01 0.01} yrefine= {0.05 0.05 0.05} add
refinebox remesh
implant Arsenic dose=5e15<cm-2> energy=40<keV> \
    tilt=7<degree> rotation=-90<degree>

#--- N+ implantation & final RTA -----
diffuse temperature=1050<C> time=10.0<s>
struct tdr=NMOS8 ; # S/D implants

#--- Contacts -----
deposit Aluminum type=isotropic thickness=30<nm>

mask name=contacts_mask left=0.2<um> right=1.0<um>
etch Aluminum type=anisotropic thickness=0.25<um> mask=contacts_mask
etch Aluminum type=isotropic thickness=0.02<um> mask=contacts_mask

#--- Reflect -----
transform reflect left
struct smesh=NMOS ; # Final

# save final structure:
# - 1D cross sections
SetPlxList {BTotal NetActive}
WritePlx NMOS_channel.plx y=0.0 silicon

```

```
SetPlxList {AsTotal BTotal NetActive}  
WritePlx NMOS_ldd.plx y=0.1 silicon
```

```
SetPlxList {AsTotal BTotal NetActive}  
WritePlx NMOS_sd.plx y=0.35 silicon
```

Appendix II

Sentaurus Device Syntax

* HD

```
File{
  Grid      = "n1_msh.tdr"
  Plot      = "n4_des.tdr"
  Parameter = "pp4_des.par"
  Current   = "n4_des.plt"
  Output    = "n4_des.log"
}

Electrode{
  { Name="source"    Voltage=0.0 }
  { Name="drain"     Voltage=0.0 }
  { Name="gate"      Voltage=0.0 }
  { Name="substrate" Voltage=0.0 }
}

Physics{
  Hydrodynamic(eTemperature)
  eQCvanDort
  EffectiveIntrinsicDensity( OldSlotboom )
  Mobility(
    DopingDep
    eHighFieldsaturation( CarrierTempDrive )
    hHighFieldsaturation( GradQuasiFermi )
    Enormal
  )
  Recombination(
    SRH( DopingDep )
  )
}

Plot{
*--Density and Currents, etc
  eDensity hDensity
  TotalCurrent/Vector eCurrent/Vector hCurrent/Vector
  eMobility hMobility
  eVelocity hVelocity
  eQuasiFermi hQuasiFermi

*--Temperature
  eTemperature Temperature * hTemperature

*--Fields and charges
  ElectricField/Vector Potential SpaceCharge

*--Doping Profiles
  Doping DonorConcentration AcceptorConcentration

*--Generation/Recombination
```

```

SRH Band2Band * Auger
AvalancheGeneration eAvalancheGeneration hAvalancheGeneration

*--Driving forces
eGradQuasiFermi/Vector hGradQuasiFermi/Vector
eEparallel hEparallel eENormal hENormal

*--Band structure/Composition
BandGap
BandGapNarrowing
Affinity
ConductionBand ValenceBand
eQuantumPotential
}

Math {
  Extrapolate
  Iterations=20
  Notdamped =100
  RelErrControl
  ErRef(Electron)=1.e10
  ErRef(Hole)=1.e10
}

Solve {
  *- Build-up of initial solution:
  NewCurrentFile="init"
  Coupled(Iterations=100){ Poisson }
  Coupled{ Poisson Electron Hole eTemperature }

  *- Bias gate to target bias
  Quasistationary(
    InitialStep=0.01 Increment=1.35
    MinStep=1e-5 MaxStep=0.2
    Goal{ Name="gate" Voltage= 1.1 }
  ){ Coupled{ Poisson Electron Hole eTemperature } }

  *- drain voltage sweep
  NewCurrentFile=""

  Quasistationary(
    InitialStep=1e-3 Increment=1.35
    MinStep=1e-5 MaxStep=1.1
    Goal{ Name="drain" Voltage= 1.1 }
  ){ Coupled{ Poisson Electron Hole eTemperature }
    CurrentPlot(Time=(Range=(0 1) Intervals=20))
  }
}
* none
}

```

