Brigham Young University
BYU ScholarsArchive

Theses and Dissertations

2011-04-29

Advancements in the Solid-state Impact-ionization Multiplier (SIM) Through Theory, Simulation and Design

Michael S. Johnson
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Electrical and Computer Engineering Commons

BYU ScholarsArchive Citation

Johnson, Michael S., "Advancements in the Solid-state Impact-ionization Multiplier (SIM) Through Theory, Simulation and Design" (2011). *Theses and Dissertations*. 2636.
https://scholarsarchive.byu.edu/etd/2636

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Advancements in the Solid-state Impact-ionization Multiplier (SIM) Through Simulation, Fabrication, and Characterization

Michael S. Johnson

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Aaron R. Hawkins, Chair
Brian A. Mazzeo
Gregory P. Nordin
Stephen M. Schultz
Richard H. Selfridge

Department of Electrical and Computer Engineering

Brigham Young University

June 2011

Copyright © 2011 Michael S. Johnson

All Rights Reserved

ABSTRACT

Advancements in the Solid-state Impact-ionization Multiplier (SIM)
Through Simulation, Fabrication, and Characterization

Michael S. Johnson
Department of Electrical and Computer Engineering
Doctor of Philosophy

This dissertation outlines the study and development of a Solid-state Impact-ionization Multiplier (SIM). The SIM is a stand-alone current amplifier designed with optical detection systems in mind. The SIM amplifies signals utilizing impact ionization as a source of gain. The SIM is fabricated on silicon in order to take advantage of its favorable impact ionization coefficients. Utilizing silicon in impact ionization based gain devices makes low noise and high gains attainable.

Because it is a stand-alone device, it can be wired to an arbitrary current source making it capable of receiving an input from photodiodes of any material. This makes it possible to amplify a signal from a photodiode that has been optimized for a given wavelength. In this way, the SIM attempts to separate the absorption and multiplication portions in modern day optical detection/amplification devices such as in Avalanche Photodiodes (APDs). This flexibility allows it to be utilized in many different systems.

The SIM has gone through several iterations in the last few years. Each change has been with the purpose of increasing gain, frequency response or yield. The progression of the device has come at the hand of much thought, theory, simulation, fabrication, and testing.

One of the challenges encountered in its development has been gain controllability due to poor carrier confinement and premature breakdown. Increased gain control was developed through simulation and fabrication of a confining oxide layer. Yield and difficulties in consistent fabrication were also addressed by altering the input metallization and doping processes. The frequency response of the device has been the largest challenge in device development. Issues such as space charge, floating node voltage, edge effects and low signal amplification have caused limitations. Successes and attempts at overcoming these, and other, challenges is the basis of this dissertation of work.

Keywords: Michael S. Johnson, impact ionization, solid-state, multiplication gain, frequency response, avalanche gain, breakdown, amplifier, SIM

ACKNOWLEDGMENTS

I would never have considered a Ph.D. without Dr. Aaron R. Hawkins. I will be forever grateful to him for presenting me with this opportunity and believing in me throughout the process. Nearly everything I know about solid-state physics and device physics I learned from him. He has been irreplaceable as a mentor in my life and development as an engineer.

I also would like to express my appreciation for my committee: Dr. Richard H. Selfridge, Dr. Stephen M. Shultz, Dr. Gregory P. Nordin, and Dr. Brian A. Mazzeo. They have all aided in my education and progress. I am grateful for their flexibility and willingness to help me even on short notice.

I would also like to thank those who came before me on this project, Dr. Joshua L. Beutler and Dr. Hong Wei Lee. I spent many years alongside Josh. His ideas and understanding of fabrication processes were more than helpful. I also thank my other colleagues during the graduate program who have always been there to listen to my ideas and offer suggestions.

Last of all, I would like to thank my wife and kids. Kristen has always believed in me. She was there to lift me up during the failures and cheer me on during the successes. Their sacrifices have come without complaint as I pursued further education.

I would also like to recognize our funding sources over the years in the Raytheon Ideas Program and the National Science Foundation who made this research possible.

TABLE OF CONTENTS

LIST OF TABLES .... **XI**

LIST OF FIGURES .... **XII**

1 INTRODUCTION **1**

1.1 Introduction 1

1.2 Contributions 2

1.3 Conclusion 4

2 BACKGROUND **5**

2.1 Introduction 5

2.2 The Photodiode 5

2.2.1 PN Junction Photodiode 7

2.2.2 PIN Photodiode 9

2.2.3 Schottky Photodiode 13

2.2.4 Avalanche Photodiode 14

2.2.5 SAM APD 17

2.2.6 Single Photon APD 19

2.3 Photomultiplier Tubes 22

2.4 Transistor Based Amplifiers 24

2.5 Limitations of Present Day Detectors and Amplifiers 24

2.6 The Solid-state Impact-ionization Multiplier (SIM) 29

2.7 Conclusions and Overview 29

3 THEORY OF IMPACT IONIZATION **31**

3.1 Introduction 31

3.2 Electron Multiplication via Impact Ionization 32

3.3 Ionization Coefficients 38

3.4 Impact Ionization Gain Equations 41

3.5 Measuring Impact Ionization Coefficients 45

3.6 Impact Ionization Coefficient Ratio 49

3.6.1 Response Time 49

3.6.2 Bandwidth 51

3.7 Noise 53

3.8 Impact Ionization in Silicon 55

3.9 Conclusions 57

4 SIMULATION, MODELING AND DATA ANALYSIS 59

4.1 Introduction 59

4.2 Framework for Device Simulation 61

4.3 Development and Choice of Simulation Models 65

4.4 Device Simulation in Silvaco 68

4.4.1 Gummel Method 71

4.4.2 Newton Method 72

4.4.3 Block Method 73

4.4.4 Foundatioin for Simulation Setup in Silvaco ATLAS 73

4.4.5 Silvaco Simulation Example 77

4.5 Conclusions 85

5 CONCEPT OF THE SIM 87

5.1 Introduction 87

5.2 Advantages of the SIM 88

5.3 Basic Design ... 91

5.4 Stages of Operation ... 93

5.4.1 Pre-breakdown Stage ... 93

5.4.2 Transitional Stage ... 96

5.4.3 Breakdown Stage ... 98

5.5 Brief SIM Evolution ... 101

5.5.1 Surface SIM (Schottky) ... 102

5.5.2 Vertical SIM (Schottky) ... 103

5.5.3 Ohmic Contact SIM ... 104

5.5.4 Buried Oxide SIM ... 104

5.5.5 Future SIM for Optimal Operation ... 105

5.6 Conclusions ... 106

6 OHMIC CONTACT SIM ... 107

6.1 Introduction ... 107

6.2 Ohmic Contact Injection Node ... 107

6.3 Metal-Semiconductor Barrier ... 109

6.4 Simulation and Design ... 115

6.5 Fabrication ... 118

6.6 Current and Voltage DC Characteristics ... 119

6.7 Frequency Response ... 121

6.8 Gain Controllability ... 123

6.9 Conclusions ... 125

7 BURIED OXIDE SIM ... 127

7.1 Introduction ... 127

7.2 Impact Ionization Efficiency ... 127

7.3 Buried Oxide to Increase Ionization Efficiency 134
7.4 Fabrication 138
7.5 Gain Measurements 140
7.6 Conclusions 142
8 FREQUENCY RESPONSE AND GAIN 145
8.1 Introduction 145
8.2 Metal-Semiconductor Barrier Revisited 146
8.3 Space Charge 146
8.4 Electric Field Distribution 152
8.4.1 Edge Effects 153
8.5 Effective Resistance from Floating Voltage 160
8.5.1 Measured Floating Voltage 163
8.5.2 DC Offset Current 165
8.5.3 Four Electrode SIM 167
8.6 Conclusions 172
9 OPTIMAL SIM DESIGN 173
9.1 Introduction 173
9.2 Theoretical Limits to Optimal Design 176
9.3 Fabrication Considerations 178
9.4 SIM vs. BJT vs. MOSFET 179
9.5 Conclusions 180
10 CONCLUSION 183
10.1 Introduction 183
10.2 Future Work 184
10.3 Conclusions 185

REFERENCES **187**

A. SILVACO CODE AND SIMULATIONS **199**

A.1 Silvaco Simulation of SIM in Rectangular Coordinates 199

A.2 Mixedmode Simulation 201

A.3 3D Cylindrical Simulations in Silvaco 202

A.4 InP/InGaAs SAM APD Example from Chapter 3 207

B. FABRICATION OF BURIED OXIDE SIM **211**

B.1 SIM Fabrication Procedure for a Wet Etch Buried Oxide SIM 211

B.1.1 KOH Etch Mask, and KOH Etching 211

B.1.2 Oxide Growth and Diffusion Mask Planarization 213

B.1.3 Diffusion 214

B.1.4 Metallization 215

B.2 SIM Fabrication Procedure for a Dry Etch Buried Oxide SIM 216

B.2.1 RIE, ICP Etching 216

B.2.2 Oxide Growth and Diffusion Mask Planarization 217

B.2.3 Diffusion 218

B.2.4 Metallization 219

B.2.5 Process Charts 220

LIST OF TABLES

Table 2.1: Table showing the different types of transistor based amplifiers. 25

Table 4.1: Lists the basic ATLAS commands and a brief description. 76

Table 4.2: Outlines a few of the models used in Silvaco simulations. 83

Table 6.1: Comparison of frequency limit of the Schottky contact SIM to that of the ohmic contact SIM. 121

LIST OF FIGURES

Figure 2.1: Band and slab diagram for a PN junction photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. 8

Figure 2.2: Band and slab diagram for a PIN photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. Notice the size of the absorption region when compared to that of the PN junction photodiode. 10

Figure 2.3: a) Cross section of a typical PIN photodiode showing a photon being absorbed in the N+ region as opposed to the desired intrinsic region. b) Band diagrams of a PIN photodiode illustrating the same effect as (a). Notice in (b) that the generation of the electron-hole pair occurs outside the region where an electric field is present (bent bands). Generated carries outside of the depletion region typically recombine before they are observed. 12

Figure 2.4: Band and slab diagram for a Schottky photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. Notice that the photons must be absorbed very close to the surface of the semiconductor layer. 13

Figure 2.5: Band and slab diagram for an avalanche photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. Notice the strong bending of the bands due to a very high bias applied. This high field is necessary for avalanche gain...15

Figure 2.6: Cross section of a typical APD. Illustrated is the doping profile as well as the path of optical illumination. For the device to operated properly the photons must be absorbed in the intrinsic layer and the device must be biased sufficiently high so as to produce impact ionization events from the generated carriers. 16

Figure 2.7: InGaAs-InP SAM APD structure. a) Shows the cross section of the device. b) Illustrates the band diagram for the structure shown in (a). Take note of the corresponding distances through the device. Also, note the heterojunction in the

band diagram. This device attempts to take advantage of the desirable absorption properties of InGaAs and the more favorable multiplication properties of InP. 18

Figure 2.8: Current versus voltage curves illustrating the over-biasing of a SPAD compared to a regular APD. Notice that the current before breakdown for a SPAD is also lower than that of an APD in order to reduce premature breakdown at such high bias levels. It should also be noted that the breakdown of the SPAD doesn't occur until a 'trigger event' initiates the breakdown burst. 20

Figure 2.9: Illustrates a typical PMT structure. A photon enters the device and strikes the cathode emitting an electron. The arrows illustrate the path and magnitude of the signal as it is amplified by the dynodes and finally collected by the anode. 23

Figure 2.10: a) BJT common amplifier configurations b) MOSFET common amplifier configurations. These illustrate the simplicity of some types of amplifier circuits based on transistor technology. 25

Figure 3.1: a) Slab diagram illustrating the process of impact ionization of electrons in a high biased pn junction. b) Corresponding band diagram to (a). The diagram illustrates the transit of carriers through the electric field as they participate in impact ionization events. 33

Figure 3.2: Simple parabolic estimation for the conduction and valence bands for a semiconductor material. Simplifications of this nature are often made in order to analyze complicated semiconductor band structures. 34

Figure 3.3: Simplified version of the band structure of silicon. 35

Figure 3.4: Detailed band structure for silicon. This more complex representation shows wavevector dependence. Going from left to right along the horizontal, the wavevector is tracing out a particular one-dimensional path through the three-dimensional momentum space. The vertical axis represents energy. 36

Figure 3.5: Slab diagram showing the current density through a multiplication region in a highly biased semiconductor. This diagram helps set the boundary conditions and framework for the derivation of the gain equations that describe impact ionization.........42

Figure 3.6: Pure electron injection via optical absorption in a PIN photodiode. Biasing to achieve impact ionization is also shown in order to accurately measure ionization coefficients under different fields. 47

Figure 3.7: Slab diagrams illustrating impact ionization events in different materials. a) This material has a low ratio of α to β yielding a clean gain pattern across the bulk. b) The electrons and holes are equally likely to produce ionization events in this material introducing a long settling time. 50

Figure 3.8: Multiplication gain with respect to multiplication length for pure electron injection. Notice change in gain versus length for different ionization coefficient ratios....52

Figure 3.9: Upper limit frequency due to the ratio of α to β. Data was taken for a constant gain of 50 and plotted. Notice a system with k=0 can never be limited in frequency by impact ionization....53

Figure 3.10: Excess noise factor versus multiplication gain for various semiconductor materials and compounds. Notice also, that an estimated ratio value is given for each material. There are two values given for silicon illustrating the dependency on multiplication lengths to the ionization coefficients. Silicon has the most favorable excess noise factor of known semiconductors....56

Figure 4.1: Flowchart showing the steps to implement a device utilizing the process and device simulations to develop a desired result....60

Figure 4.2: Illustrates the different types of models and methods used in device simulation. It shows the tradeoff between ease and speed of use and the accuracy of the solution....63

Figure 4.3: The right hand side of the arrows shows the types of commands that are desired in a process simulation setup. The left hand side shows the corresponding commands used in the Silvaco platform to properly describe the device and the modeling....74

Figure 4.4: Flowchart illustrating the main areas used to analyze data in the Silvaco platform. The devices simulator (ATLAS) can take inputs from either Deckbuild or the process simulator (ATHENA). The output data can be analyzed real-time or through Tonyplot....75

Figure 4.5: An InP/InGaAs SAM APD. This device will be simulated in the following example....77

Figure 4.6: Mesh for the InP/InGaAs SAM APD. The mesh was produced in SILVACO. Proper mesh setup is necessary for accurate solutions. Notice that the size of the mesh decreases at the material and doping junctions indicated by the black solid lines. This is necessary to obtain accurate results....79

Figure 4.7: Regions and materials have been assigned in this Silvaco simulation....81

Figure 4.8: Cross section of the SAM APD as produced by Silvaco. The doping profile is shown and indicated. Note that only the concentration can be seen in the figure and not the doping type. The code must be available to see the doping type....82

Figure 4.9: Silvaco output file showing the photogeneration rate within the device. Notice that because the wavelength of light was specified correctly for the

materials the optical signal is not absorbed by the InP which is the multiplication material for this APD. 84

Figure 4.10: Current versus voltage curve for the SAM APD. Green curve is for no light incident on the APD. The red is for 1 W/cm^2. Silvaco is capable of showing an array of different outputs in this manner....84

Figure 4.11: Silvaco Structure file showing a cutline and the band diagrams along the associated line. The structure file is on the right illustrating the location of the cutline and the corresponding band diagrams are on the left....85

Figure 5.1: Shows the capability of the SIM to be wired to an arbitrary current source. a) The SIM amplifying a signal from an silicon photodiode. b) The SIM amplifying a signal from a InGaAs photodiode 89

Figure 5.2: Excess noise versus gain for various ratios of ionization coefficients. Notice as the gain goes up (this indicates an increase in electric field strength) the excess noise increases as well. This indicates that for devices limited by excess noise factor, lower gains are better....90

Figure 5.3: Two devices wired together in a cascading fashion. With the output of one device connected to the input of the next device the system gain can be large while keeping the gain per device manageable. In this example the devices are biased up to produce a gain of 10 yielding a system gain of 100. 91

Figure 5.4: Illustrates a cross section of the basic design of the SIM. The SIM is fabricated on a P+ substrate with a P- epitaxial layer. The output node has an N+ well. This version of the SIM has a Schottky contact at the injection node. 92

Figure 5.5: Surface band diagram over the top of a SIM cross section. This diagram is for an unbiased SIM. Bending in the bands is due to the built in potential at the N+ to P- junction....93

Figure 5.6: Represents a typical current versus voltage curve for the SIM. Labeled are the different stages of operation as the output voltage changes....94

Figure 5.7: Slab diagram showing the pre-breakdown stage of the SIM. At this stage the bias isn’t large enough to extend the depletion region to the input node. 95

Figure 5.8: Surface band diagram over the top of a SIM cross section. This illustrates the bending of the bands for a biased SIM where the bias isn’t yet sufficient to completely depleted the epitaxial layer between the input and output nodes. 95

Figure 5.9: Surface band diagram over the top of a SIM cross section. This illustrates the bending of the bands for a biased SIM where the bias has caused complete depletion of the epitaxial layer between the input and output nodes. The bias still isn’t sufficiently strong to initiate impact ionization events. 97

Figure 5.10: Slab diagram showing the transitional stage of the SIM. At this stage the bias is large enough to extend the depletion region to the input node but not large enough for impact ionization to occur. 97

Figure 5.11: Surface band diagram over the top of a SIM cross section. This illustrates the bending of the bands for a biased SIM where the bias has caused complete depletion of the epitaxial layer between the input and output nodes and has become sufficiently strong to initiate impact ionization events..99

Figure 5.12: Slab diagram showing the breakdown stage of the SIM. At this stage the bias is large enough to extend the depletion region to the input node and also large enough for impact ionization to occur. Notice the path of the generated holes is toward the grounded substrate. 99

Figure 5.13: Inverted surface band diagram. This shows the inverse of the valence and conduction bands. This is the potential seen by generated holes in the depletion region. As indicated by the potential drop, the holes travel to the grounded P+ substrate instead of traveling back toward the input node. 100

Figure 5.14: Top view of a surface SIM. This top view shows the doped wells in the epitaxial layer. Instead of the generated holes traveling through the epitaxial layer to the substrate they travel along the surface to grounded P+ wells. In this SIM all of the carrier transport happens at the surface as opposed to through the bulk. 102

Figure 5.15: Cross-section and operation of the Schottky contact SIM. 103

Figure 5.16: Cross-section and operation of the Ohmic contact (or pn junction) SIM. This version SIM is discussed in full detail in Chapter 7. 104

Figure 5.17: Cross-section and operation of the buried oxide SIM. This version of the SIM is discussed in full detail in Chapter 8. 105

Figure 6.1: Surface band diagram for the SIM during various stages of bias at the output node. a) No bias applied to the output node. b) Bias applied at output node isn't sufficient to fully deplete the semiconductor between the input and output. c) Bias applied at output node fully depletes semiconductor between input and output but isn't sufficient to initiate impact ionization events. d) Bias is large enough for impact ionization to occur. 108

Figure 6.2: Band diagrams illustrating the metal-semiconductor interface at the injection node for (a) the ohmic contact SIM and (b) the Schottky contact SIM. The bands are shown for the biased and unbiased cases for both versions of the SIM....................110

Figure 6.3: Band diagram for the metal-semiconductor contact in both the Schottky and ohmic contact SIMs. (a) The Schottky contact SIM. It shows a lowering of the barrier height at the metal-semiconductor interface for a change in voltage at the output node. (b) The ohmic contact SIM. It shows no change in the barrier height

at the metal-semiconductor interface but also illustrating the barrier lowering at the pn junction....112

Figure 6.4: Electric field profile between the input contact and the output N+ well. a) Corresponds to the derivative of the potential for V_1 of Figure 6.3a. b) Corresponds to the derivative of the potential for V_2 of Figure 6.3a....113

Figure 6.5: Current versus voltage curve illustrating the resistive effects of an ohmic contact and a Schottky contact....115

Figure 6.6: Current versus voltage curves for the ohmic contact SIM produced by Silvaco....117

Figure 6.7: Cross-section of the ohmic contact SIM as simulated in Silvaco. Upper-left hand shows the doping concentrations for SIM. Upper-right hand shows the potential curves for the biased SIM. Lower-left hand is the electron current density for the biased SIM. Lower-right hand corner is the hole current density for the SIM....117

Figure 6.8: a) Output current versus voltage curve for modeled and fabricated pn junction injection SIMs. The solid black and gray curves represent simulated injection currents of 50nA and 500nA respectively. These can be compared to the measured results for 50nA and 500nA injection tests represented by the diamonds and triangles respectively. The leakage current curve for the fabricated devices is also shown and is represented by the Xs. The voltage shown in the graph is applied to the output node. The fabricated device had a width of 10μm between the n-wells of the injection and output nodes. b) Gain (M) versus voltage curve for the same modeled and fabricated pn junction injection SIM devices from (a)....120

Figure 6.9: Testing setup for extracting the frequency response in the SIM....122

Figure 6.10: Illustrates the measured 3dB bandwidth for the fabricated SIM, represented by the diamond points, over a range of current levels. The theoretical response predicted by Eq. 5 is shown as a solid line, with 20 pF used for the node capacitance. A bias of approximately 40 volts was used....123

Figure 6.11: a) Cross section of the SIM showing the geometries and doping types. Approximate doping levels include: N+ type=1e18, P-type=1e15, P+ type=1e18. b) Illustrates field profiles for the horizontal cutline shown in (a) when V_{SIM} is at 50 volts and the injected current is 1 micro-Amp. c) Illustrates the field profiles for the vertical cutline in (a) for the same conditions as (b). Electric field lines were extracted from ATLAS simulations for the structure. 124

Figure 7.1: Cross section of the buried oxide SIM showing the magnitude of the electric field, as calculated in ATLAS, with V_{sim}=50V and $I_{injection}$ =10μA. The oxide depth (d_{ox}) and the well depth (D_{well}) are labeled (in this illustration d_{ox}=D_{well}). 129

Figure 7.2: Cross section of the SIM showing the magnitude of the current density, as calculated in ATLAS, with V_{sim}=50V and $I_{injection}$=10µA. The holes produced by impact ionization are shown moving diagonally down to the substrate. 129

Figure 7.3: Theoretical gain vs. voltage plot as produced by Equations 7.1 - 7.6 showing how the gain curves rise very abruptly with low electron injection efficiency. This compares the theoretical cases of 1% and 100% efficiency for electron injection into the high field region. 131

Figure 7.4: Theoretical gain vs. voltage plot as produced by Equations 7.1 - 7.6 showing the derivative (dM/dV) for Figure 7.3. 131

Figure 7.5: Simulations showing the current versus voltage curves for varying well depths at the output node. This illustrates the electric field distribution is dependent upon the doping profile and well depths and is effecting the ionization efficiency as electrons pass from input to output without passing through the optimal field region. 133

Figure 7.6: Simulation of a buried oxide SIM showing the electric field profile. Notice that the electric fields are nearly unchanged when compared to Figure 7.1. This simulation has V_{out}=50V $I_{injection}$=10μA where d_{ox}=D_{well}. 135

Figure 7.7: SIM cross-section showing the total current density and V_{sim}=85V and $I_{injection}$=10μA. It was necessary to increased V_{out} to obtain current levels comparable to the buried oxide SIM of Figure 7.2. This is due to the low impact ionization efficiency of the configuration of Figure 7.2. The current density shown moving diagonally down to the substrate are the holes produced by impact ionization. 135

Figure 7.8: Comparison of simulations in Silvaco of an ohmic contact SIM and a buried oxide SIM with an ohmic contact injection point. They are labeled as *no oxide* and *buried oxide* respectively. Notice the gradual breakdown of the buried oxide case when compared to the no oxide. This is indicative of increased impact ionization efficiency. 136

Figure 7.9: Each dot in the graph represents a separate simulation in Silvaco where the only difference from one simulation to the next was the oxide depth (shown on the x-axis). Each value was obtained by taking the derivative of gain (M) with respect to voltage at a gain of M=50. The well depth at the output for all the simulations was held at a constant 1μm. Notice that 1μm oxide depth is the low dM/dV point. 137

Figure 7.10: Fabrication of the second generation Buried Oxide SIM begins by growing a) thermal oxide to act as a KOH etch mask. b) The thermal oxide is then patterned to form the etch mask and c) KOH etched to form pedestals. d) The etch mask is then removed and a new thermal oxide is grown over the entire wafer. e) SU8 is then placed over the wafer to protect the field oxide during polishing and

f) the oxide on the top of the pedestal is removed via Chemical Mechanical Polishing (CMP). After cleaning off the SU8, g) spin on glass is spun on over the entire wafer. h) The exposed silicon pedestal tops are then doped with phosphorus spin on glass with the thermal oxide acting as a diffusion barrier in the diffusion process. After diffusion, the excess spin on glass is removed. i) Lastly the spin on glass is removed and metal is patterned onto the doped pedestals. 139

Figure 7.11: Scanning electron microscope (SEM) image of a cross-section of the buried oxide SIM. Notice the characteristic etch angle due to wet etching silicon. Labeled is the theoretical well dept on one of the pedestals. 140

Figure 7.12: Illustration showing the testing setup of the SIM devices. A cross-section of the SIM is shown with the photodiode and the parameter analyzer. 141

Figure 7.13: Comparison of the measured and simulated IV curves of the SIM. The buried oxide case has an injected current of $I_{injection}$=500nA. For the planar (no oxide) case $I_{jection}$=50nA. 142

Figure 8.1: Top views of SIM devices showing the doping and metallization. a) Shows the traditional Schottky contact SIM. b) Shows an ohmic contact SIM with elongated wells and metallization at the input and output contact points. This configuration increases the cross-sectional area of the depletion region thus reducing the space charge resistance (R_{sc}). 149

Figure 8.2: Theoretical SIM with increased doping and metallization in order to increase the cross-sectional area of the depletion region between the input and output during operation. 150

Figure 8.3: Images taken through a microscope of the top view of the metallization of the SIM. a) Shows the SIM with a older version of the contacts. b) Shows the elongated doping and metallization of a SIM designed to reduce space charge resistance. 150

Figure 8.4: $R_{barrier} + R_{sc}$ resistance versus input current between the metal semiconductor and electron collector. The theoretical curve is calculated using Equation 8.10 and Equation 6.9 assuming a SIM device made using a *p*-type epitaxial layer, Schottky injection contact, and a spacing d equal to 4μm. Measured values correspond to fabricated SIM devices with those parameters. 151

Figure 8.5: a) Highly doped n type well in a p type semiconductor. The well can be broken down into (b) the planar bottom of the doping distribution, (c) the cylindrical sides of the distribution and (d) the spherical corners. 154

Figure 8.6: Planar cut of Figure 8.5a. This figure labels the different curvatures of radius for the junction surface. 155

Figure 8.7: Theoretical current versus voltage curves illustrating the breakdown voltages for different types of junctions. The spherical region of a junction will breakdown sooner because the electric field is higher there. The geometry and abruptness of a junction are very important when analyzing the breakdown characteristics and the electric field distribution. 155

Figure 8.8: Shows the breakdown voltages for different doping levels of a one-sided abrupt junction in a silicon substrate. The different breakdown levels are for the different radii of curvature (r_j) as defined by Figure 8.6. 156

Figure 8.9: Cut-plane of a 3D simulation in Silvaco. This cut-plane is a top view of an ohmic contact SIM at about 1µm into the epitaxial layer (the depth of the N+ well at the input and output). Electric field profile is shown at this depth. The field profile reaffirms that the spherical portion of the junction has a higher field than the cylindrical and planar portions....157

Figure 8.10: Circular SIM design with the same distance from the input well to the output well. The doping and other fabrication consideration remain the same from this design to previous version of the SIM. This designs purpose was twofold. First, it decreases the space charge resistance by increasing the depletion region cross-sectional area. Secondly, it removes any spherical portion in the doping profile making a much more uniform electric field distribution. 158

Figure 8.11: Gain versus voltage curves as done in 3D Silvaco for buried oxide rectangular and circular SIM configurations. Notice the rate of change of gain with respect to voltage. The circular geometry maintains a much more gradual and even gain profile than the rectangular suggesting a higher impact ionization efficiency and more even electric field distribution with respect to carrier path....159

Figure 8.12: Cut-plane of a 3D simulation in Silvaco. This cut-plane is a top view of a circular ohmic contact SIM at about 1µm into the epitaxial layer (the depth of the N+ well at the input and output). Electric field profile is shown at this depth. Because the doping profile is circular there are no spherical portions of the junction. This helps maintain an even electric field distribution across the complete area of the depletion region. 159

Figure 8.13: Silvaco Simulation showing the shifts in the electron quasi Fermi level for injected current levels ranging from 1pA to 1µA. This data was taken near via cutline in Silvaco near the input node of a SIM. The mesh coordinates of the x-axis are in microns going from 0 (the input pn junction) toward the output pn junction of the device. 161

Figure 8.14: Resistance versus current curve. It shows the derivative of the voltage versus current from Figure 8.15. Also shown is the theoretical diode resistive limit of kT/I. Notice that the measured data is nearly two orders of magnitude higher than kT/I....164

Figure 8.15: Voltage versus current curve showing the shifting at the input voltage for a range of input current levels. 164

Figure 8.16: Voltage shifts at the input node for corresponding injection current levels. The relationship is linear showing a resistive nature. Voltage and current shift one correspond to a 1nA current pulse with no DC offset current. Voltage and current shift two correspond to a 1nA pulse with a 10nA DC current offset. Notice the changes in voltage from case one to two given the same pulse. 166

Figure 8.17: SIM with fourth stabilizing electrode between the input and output. The operation of the SIM remains unchanged but the addition of the electrode controls electric field shifts in the SIM reducing the effects of the floating voltage.....................169

Figure 8.18: Input voltage versus input current for various voltages on the stabilizing electrode. Notice the large voltage shifts for the floating and +10V cases when compared to the 0V and -10V cases. This shows a strong indication that the stabilizing electrode helps immensely to reduce the shifts in the input voltage for shifts in the input current. 170

Figure 8.19: Shows the resistance versus current for due to the floating voltage at the injection node. Shown are the resistance values for various different voltages on the stabilizing electrode. These values are calculated by taking the derivative of Figure 8.18. 171

Figure 9.1: Illustrates the optimal SIM design. The design includes the ohmic contact injection node to eliminated barrier resistance at the input. It also has the buried oxide for increased ionization efficiency. It is circular in nature to reduce edge effects and space charge resistance. It has a fourth electrode in order to reduce the effects of the floating voltage at the input node...174

Figure 9.2: Illustrates a cross section of the optimal SIM shown in Figure 9.1. The figure shows the operation in the breakdown stage of the SIM. The operation stays consistent with previous versions of the SIM. 175

Figure 9.3: Top view of the optimal SIM illustrating the metallization. This design contains all of the optimization features discussed in Chapter 8. 178

Figure A.1: Examples of 3D simulation done in Silvaco's 3D ATLAS environment. The top row on the left is a rectangular SIM. The top on the right is a circular SIM. The bottom row shows a cut plane of the SIM s on the top row. The cut plane shows the electric filed magnitude at a depth equal to the input and output well depths. 206

Figure B.1: SIM Oxide Growth Chart #1: This chart shows the necessary step gas flows, step temperatures, and step durations to grow roughly 1200 angstroms of thermal oxide to act as a KOH etch mask in subsequent steps. The low temperature provides the necessary oxide grow while minimizing the amount of diffusion within the EPI layer. 221

Figure B.2: SIM Oxide Growth Chart #2: This chart shows the necessary step gas flows, step temperatures, and step durations to grow 5000 angstroms of thermal oxide. This oxide acts as an insulator and passivation layer for the SIM. 222

Figure B.3: SIM Diffusion Chart #1 for P-8545 Spin on Dopant (SOD): This chart shows the necessary step gas flows, step temperatures, and step durations to diffuse phosphorus into the SIM pedestals to a depth of roughly 1.3 um. It is important that sufficient oxygen is always flowing into the furnace during the diffusion. Oxygen allows the phosphorus in the SOD to diffuse by turning the phosphoric acid (H3PO4) within the mixture into Phosphorus Pentoxide (P2O5). It is important to completely oxidize all the H3PO4 within the SOD otherwise it will form a bond with the silicon that cannot be removed unless it is first oxidized. Following diffusion, it is important to immediately dip the wafers in Buffered Oxide Etch (BOE) to completely remove the SOD before the P2O5 reacts with the moisture in the air to form H3PO4...223

1 INTRODUCTION

1.1 Introduction

The Solid-state Impact-ionization Multiplier (SIM) has seen many changes and improvements during my time on the project. These changes have come in attempts to produce a novel, quality amplifier. A good amplifier has qualities that include a robust design, the capability of achieving necessary gain levels, operating with sufficiently low noise and at high frequencies. When I started on the SIM project the device suffered from several deficiencies in each of these qualities. The device had several difficult fabrication steps which produced inconsistencies in fabrication and testing which was a drawback on its robustness. Some of these features lead to frequent failure in analysis and testing. Also, the structure of the SIM made achieving high gains difficult and the breakdown difficult to control. The output signals at higher gains were very noisy and the upper limit frequency was often in the tens of kilohertz making it slow when compared to the needs of many systems [1].

Many changes have been explored make the SIM a quality amplifier and overcome these deficiencies. A new design was introduced in order to improve the fabrication and output response by making it more consistent and straight forward. The design was further altered to facilitate proper carrier confinement to improve the gain and reduce the noise at high gain levels. Significant research has produced design alterations to increase the frequency response and show viable solutions to the lower speeds the device has experienced in the past. This dissertation

outlines the background behind the operation of an amplifier such as the SIM. It then discusses the structure and operation of the SIM outlining the design alterations that have made in order to produce a quality amplifier.

1.2 Contributions

The development of the SIM was a collaborative effort and couldn't have been realized without the contributions of many individuals. My unique ideas, insights and contributions were necessary in the development of the SIM. Some of my unique contributions to this project include:

1. Unique ideas in theory and design of the pn junction and buried oxide SIMs which improved fabrication consistency and gain characteristics respectively. Also, the theory and design of the elongated electrode and stabilizing electrode SIM designs which reduce space charge and dampen floating voltage resistances respectively.
2. Development of code for simulations of the SIM in ATLAS 2D, ATLA 3D, ATHENA, and MixedMode through the Deckbuild and Tonyplot user interface.
    a. Proposal of the addition of an insulated layer for carrier confinement based on simulation of the SIM.
    b. Design and verification of proper operation of the pn junction SIM, buried oxide SIM, elongated electrode SIM, stabilizing electrode SIM. This includes the structure (doping, electric field distribution, carrier path, etc.) and DC/AC characteristics of these designs. Without this verification these designs would not have been pursued.

c. Modeling of the ionization efficiency theories and the proposal of insulator depth to doping diffusion depth. This model was necessary in verifying the effectiveness of the buried oxide.

d. Simulations verifying the floating voltage at the input due to shifts in the bands through the depletion region and the proposed solution of a DC offset current and stabilizing electrode to dampen these shifts. Without this knowledge the DC offset current and stabilizing electrode would not have been realized.

3. Development of code for modeling and optimization of electric field distribution and voltage shifts using MATLAB. These models defined the ionization efficiency theory behind poor gain characteristics.
4. Contributions to fabrication recipe and mask development for the pn junction SIM, buried oxide SIM, elongated electrodes and stabilizing SIM designs based on theories and ideas developed through simulation and modeling. These have included doping levels, oxide depths, electrode lengths, device sizes, etc.
5. Development of the packaging process for the SIM and the actual packaging of the SIM in TO-5 cans through dicing, wirebonding and other techniques. Also, development of printed circuit board for DC/AC testing of packaged SIM devices. Proper packaging and testing were necessary to verify device operation.
6. Alterations to the testing setup and process. Acquisition of data verifying operation of the various SIM designs especially the stabilizing electrode SIM design. Demonstration of the dampening effects on the floating voltage of the stabilizing electrode and DC offset current through testing. Also, my

development of Excel macros has made the acquisition of tested data and its analysis more streamlined making further advancement of the SIM possible.

The development of the SIM has provided me with the unique opportunity to participate in all of the steps of device realization including: ideas/theory, design, simulation/modeling, fabrication, packaging and testing. My most influential contributions came through the great amount of time spent on the development of the SIM through simulation. This provided unique opportunities to explore ideas in device operation and physics. Any "what if" scenario could be explored and analyzed through simulation. These opportunities helped in the development of the final versions of the SIM as well as my understanding of solid-state physics. Without the simulations and models I developed the pn junction SIM and buried oxide SIM designs would now have been pursued and realized.

1.3 Conclusion

My time on the SIM project has seen its improvement in many areas. The device has become much more robust and the fabrication process has become more consistent. The gain characteristics have been greatly improved through the development of the buried oxide channel. Strides in the frequency response have been made through the elongation of the electrodes to reduce space charge and the operation of the device with a DC offset current. These advancements to the SIM have come through significant contributions on my part through ideas, theory, design, simulation, modeling, fabrication, packaging and testing.

2 BACKGROUND

2.1 Introduction

Signal detection and amplification are necessary in nearly all modern day electrical systems. This is especially true for optical systems. There are many ways that a signal can be detected and amplified. Some of these involve methods that keep the signal in its original form. Other amplification systems will change the detected signal from one form to another in the amplification process [2]. This is true for optical detection systems that convert the signal from an optical form to an electrical form. Regardless of the method of amplification, it is always beneficial that this process not add excess noise to the system [3].

This chapter will discuss typical detectors and amplifiers used in optical systems. It will outline some of the benefits and limitations of these detectors and amplifiers and set the basic groundwork for the motivation behind the research of the SIM.

2.2 The Photodiode

Probably the most common light detector in modern day systems is the photodiode [4]. The typical photodiode is simple in nature and very effective in purpose. Optical power converted to electrical power was observed during the basic development of the PN junction [5]. It was discovered that the rectifying nature of the PN junction yielded photonic applications. This technology was refined throughout the 1950s and 1960s [6], [7]. During this time junction

technology produced devices such as the solar cell [8] and light emitting diode [9]. Variations and application of the photodiode have been researched over the last several decades and continue in the present day. Photodiodes are absolutely necessary in nearly all optical systems used today.

A properly functioning photodiode generates electron-hole pairs when illuminated by light [10]. There are several types of photodiodes. Each has a purpose in industry and science. Some of the most common photodiode structures include the p-n photodiode, PIN photodiode [11], Schottky photodiode [12], and the avalanche photodiode [13].

To understand the different varieties of photodiodes it is necessary to obtain a basic understanding of how a photodiode operates and some knowledge of the terminology used in discussing optoelectronic devices. When a photon with sufficient energy enters a photodiode an electron-hole pair is generated [14]. The energy necessary for a photon to excite an electron from the valence band to the conduction band is described by the following equation:

$$h\nu \geq E_g . \qquad \textbf{2.1}$$

In Equation 2.1 [4], ν is the frequency of the light, h represents Plank's constant and E_g is the bandgap energy for the photodiode material. How readily a device produces electron-hole pairs for different frequencies and wavelengths of light is the device's quantum efficiency [15]. More simply put, the quantum efficiency (η) of a device can be defined as the ratio of photons shined incident on the device to the number electron-hole pairs produced, as in:

$$\eta = \frac{\#\ electron\text{-}hole\ pairs}{\#\ photons} . \qquad \textbf{2.2}$$

Given Equation 2.2, and the dependency of electron-hole generation on the energy of the photon, the quantum efficiency is wavelength dependent [16]. There are many factors that influence the quantum efficiency of a device. Special care must be taken into consideration

during the fabrication of devices to increase the optical input to electrical output ratio as much as possible.

Another (very similar) method of determining how effective a photodiode is at converting an optical signal to an electrical signal is to calculate its responsivity. The responsivity of a device is very similar to the quantum efficiency but in a more broad sense. The responsivity is simply the ratio of current output by a photodiode to the optical power incident on the photodiode, as in:

$$R = \frac{current}{optical\ power}. \qquad \textbf{2.3}$$

As stated earlier, there are several different types of photodiodes. Each has its purpose and niche in industry and research. Some of the most common photodiodes used will now be discussed in more detail.

2.2.1 PN Junction Photodiode

The photodiode, in its simplest form, consists of nothing more than a p-n junction as shown in Figure 2.1. When the junction is reverse biased an absorption/depletion region is formed [14]. When photons are absorbed in this region the optical signal is converted into electron-hole pairs. The electrons drift towards the N+ side of the diode while the holes drift towards the P+ side. This is due to the electric field present in the depletion region. Because this region is depleted of carriers, the recombination rate in this region for electrons and holes is relatively small. In this way the electrical signal produced is able to be collected and observed efficiently. When photons are absorbed in a non-depleted section of the device the electron-hole pairs are unable to drift due to the lack of an electric field. Also, the presence of majority carriers in the non-depleted semiconductor increases the recombination rate. Due to the lack of

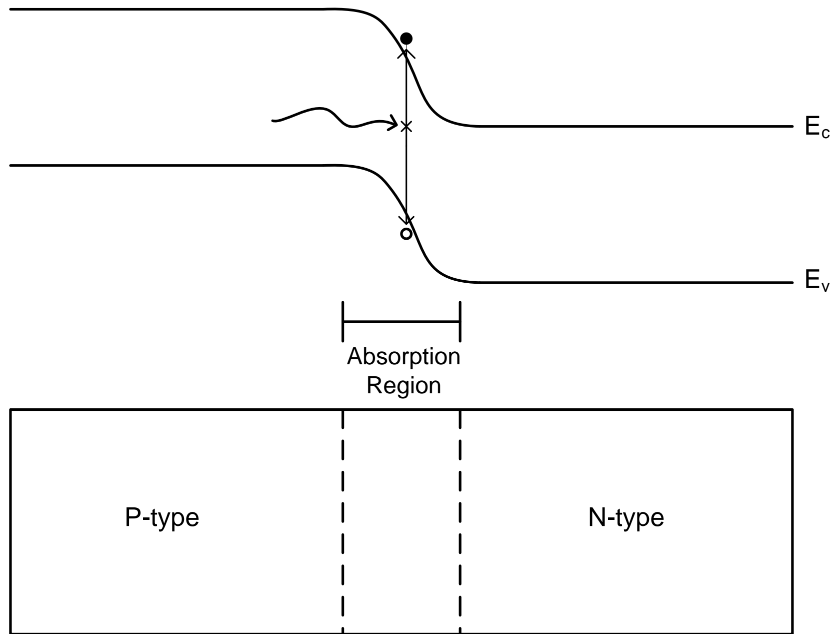


Figure 2.1: Band and slab diagram for a PN junction photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region.

an electric field and the high recombination rate these produced carriers recombine quickly without being collected. Their electrical signal is not observed and, thus, the quantum efficiency of the device decreases. For this reason, it is desired that the optical signal be absorbed in the depletion region of the device.

A proper functioning device is a very consistent and constant current generator when illuminated. The current generated by optical absorption in the depletion region can be described by

$$i_{\lambda} = \frac{\eta I_0 A e \lambda_0}{hc}, \qquad \textbf{2.4}$$

where η is the quantum efficiency of the device, I_o is the light irradiance on an absorption region of area A, λ_o is the optical wavelength, with h and c representing Planks constant and the speed of light in a vacuum respectively.

By examining Equation 2.4 [14], it can be noted that the generated current is dependent upon the area of the absorption region. In a simple p-n junction photodiode the absorption region is relatively small. This reduces the responsivity of the device as portions of the optical signal incident on the device either pass through the depletion region without being absorbed or are absorbed before entering the depletion region [17]. This problem can be decreased by increasing the size of the depletion region to allow a larger area for proper absorption of the optical signal. For this reason the PIN photodiode replaces the simple p-n junction photodiode in most practical systems today.

2.2.2 PIN Photodiode

The PIN photodiode is simply a p-n photodiode with a near intrinsic semiconductor layer in between the P-type and N-type layers, thus, the PIN name (P-type – Intrinsic – N-type). The basic structure of a PIN photodiode can be seen in Figure 2.2. The introduction of the intrinsic layer gives several advantages. It increases the size of the depletion layer which in turn increases the size of the optical absorption region increasing the responsivity of the device. Because the middle layer is nearly intrinsic only a small voltage is necessary to extend the depletion/absorption region through the entirety of this layer. Another major advantage to the increased size of the depletion region is the decreased junction capacitance [18]. Junction capacitance (C_j) is described by the following equation:

$$C_j = \frac{A\varepsilon\varepsilon_0}{w}, \qquad \textbf{2.5}$$

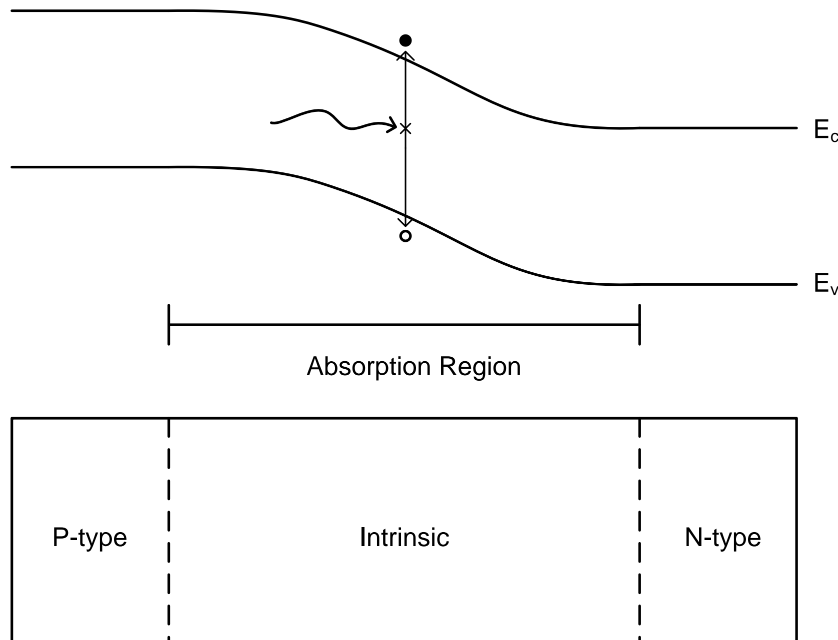


Figure 2.2: Band and slab diagram for a PIN photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. Notice the size of the absorption region when compared to that of the PN junction photodiode.

where *w* is the width of the depletion region. Notice that the junction capacitance is inversely proportional to the width (*w*) of the depletion region. This shows the relationship between the size of the depletion region and the frequency response (f_{3db}) of the photodiode. This relationship can be described in the following way

$$f_{3dB} \approx \frac{1}{2\pi R_{tot} C_j}. \quad \textbf{2.6}$$

R_{tot} represents the total resistance (series, load, etc.) experienced by the electrical signal. Taking both Equation 2.5 and Equation 2.6 into consideration it becomes apparent that an increase in the intrinsic layer width results in a corresponding increase in the frequency response of the device.

The relationship between depletion width and frequency response, however, has a trade-off. The carrier transit time through the depletion region is also increased with an increased width by the relationship

$$\tau_{drift} = \frac{w}{v_{sat}}. \qquad \textbf{2.7}$$

The carriers that drift through the electric field of the depletion region will reach a saturation velocity (v_{sat}) for that material. The width of that region determines the transit time of the carriers through the device after absorption [19]. Thus, electrical signals generated in the depletion region through optical pulses will be observed at the output of the device over a time period of τ_{drift}.

The time constant associated with the junction capacitance and internal resistance can be described as τ_{RC}. Thus, the total response time of the electrical signal produced by optical absorption is the combination of Equation 2.7 and τ_{RC}, as in:

$$\tau^2 = \tau_{drift}^2 + \tau_{RC}^2 . \qquad \textbf{2.8}$$

It now becomes apparent that the total carrier response time is a combination of these two time constants. With this in mind, the frequency response of a PIN photodiode can be improved by optimizing the width of the intrinsic layer.

Another, typically less critical, factor that affects the speed of a photodiode is the rate of diffusion for carriers absorbed outside the depletion region. This time delay can become significant if proper care isn't taken into the device geometries when considering the wavelength of the incident light. This problem typically occurs if the optical penetration depth of the incident light isn't sufficient to reach the depletion region and the photons are absorbed too close to the surface. Figure 2.3 illustrates this issue.

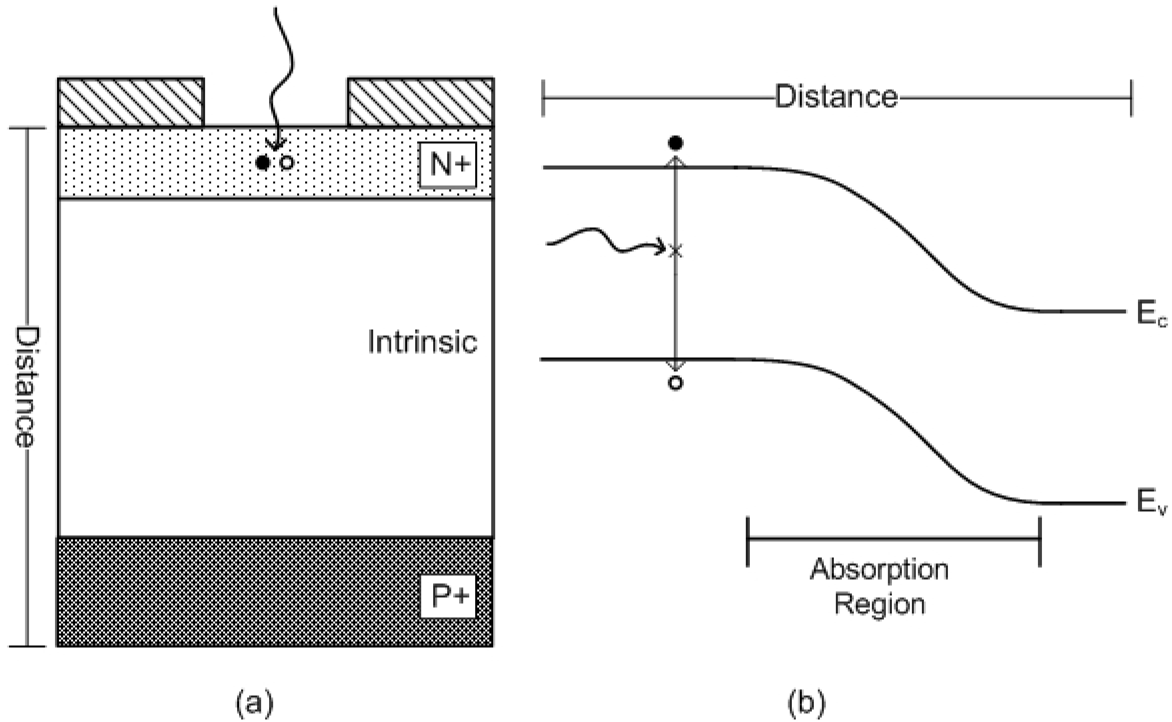


Figure 2.3: a) Cross section of a typical PIN photodiode showing a photon being absorbed in the N+ region as opposed to the desired intrinsic region. b) Band diagrams of a PIN photodiode illustrating the same effect as (a). Notice in (b) that the generation of the electron-hole pair occurs outside the region where an electric field is present (bent bands). Generated carriers outside of the depletion region typically recombine before they are observed.

To understand this concept it becomes necessary to take into consideration the diffusion length of carriers in a semiconductor material [20]. The diffusion length (*L*) is the average distance traveled by a minority carrier in a semiconductor from the point at which it is generated to the point at which it recombines as described by

$$L = \sqrt{D\tau_{carrier}}\,, \qquad \textbf{2.9}$$

where *D* is the diffusion constant for the material and $\tau_{carrier}$ is the carrier lifetime. When carriers are generated in non-depleted semiconductor, and are within a diffusion length of depleted semiconductor, it becomes probable that the carriers will diffuse into the depletion region where

they can be collected. This process creates a time delay that must be taken into consideration for systems where a significant portion of the optical signal is being absorbed outside of the depletion region. This can become quite problematic for systems that are targeting short wavelength (ultraviolet) signals that are absorbed very close to the incident surface. For this reason the Schottky photodiode is often used in systems where absorption is occurring too close to the surface.

2.2.3 Schottky Photodiode

The Schottky photodiode utilizes a Schottky contact as the 'rectifier' portion of the photodiode to create an absorption volume instead of a pn junction as seen in Figure 2.4 [21].

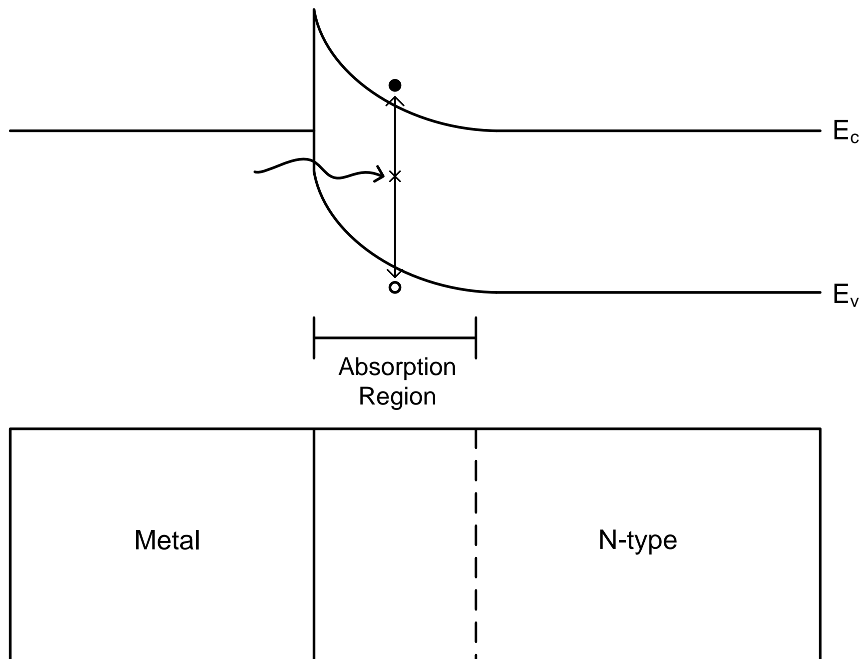


Figure 2.4: Band and slab diagram for a Schottky photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. Notice that the photons must be absorbed very close to the surface of the semiconductor layer.

This type of photodiode is much less common for typical detection systems. As stated earlier, Schottky photodiodes are usually only utilized in systems targeting ultraviolet sources due to the small size of the absorption region and the relatively small distance from the surface to the junction [22]. When a typical pn junction/PIN photodiode is used for such short wavelengths the vast majority of the incident optical signal is absorbed in the highly doped region at the surface and not the depletion region. In these systems, diffusion lengths and times become the major speed delay of the device. It also significantly reduces the quantum efficiency due to carriers that recombine before being able to diffuse into the depletion region. By using a Schottky photodiode with an anti-reflective coating both the quantum efficiency and speed delay can be significantly increased for systems targeting these wavelengths.

2.2.4 Avalanche Photodiode

The avalanche photodiode (APD) is the first device discussed in this chapter that has an internal gain mechanism [23]. For this reason the APD is often used in low light systems. The built-in multiplication process makes it possible to produce appreciable current levels in the external circuitry of a system at optical levels that would otherwise be too small.

In its simplest form, an APD can be very similar to that of a pn photodiode or, more typically, a PIN photodiode as see in Figure 2.5. The APD is quite robust and the design can be altered to fit various needs such as single photon detection, low noise output, high frequency response, etc. The most typical and basic APD structure used in optical detection systems can be seen in Figure 2.6.

The APD receives optical signals in much the same way that a PIN photodiode does. The optical signal is absorbed in the depletion region creating an electron-hole pair. The electrons and holes drift through the electric field with the electrons drifting towards the N+ region and the

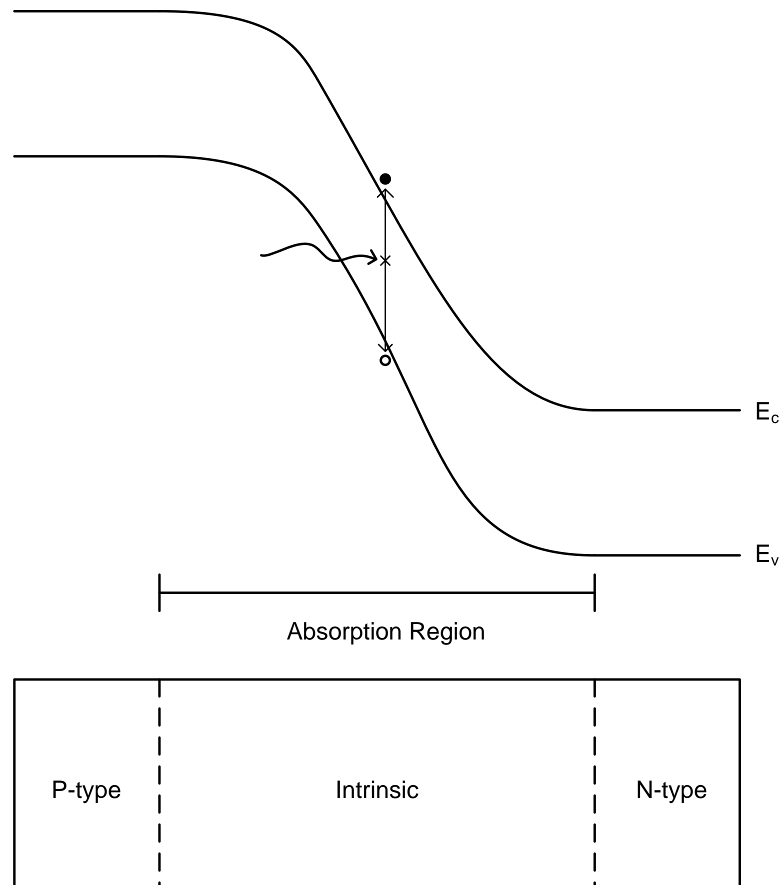


Figure 2.5: Band and slab diagram for an avalanche photodiode. The band diagrams illustrate the absorption of a photon exciting an electron from the valence band to the conduction band. The slab diagram illustrates the doping profile as well as the size of the depletion or absorptions region. Notice the strong bending of the bands due to a very high bias applied. This high field is necessary for avalanche gain.

holes drifting to the P+ region of the device. The difference between a PIN photodiode and the APD is in the electric field distribution and biasing. An APD is biased such that the electron-hole pairs are subject to a very high electric field. As electrons travel through this electric field, their drift speed depends on the electric field strength. When the electric field is increased sufficiently large they become more likely to collide with the crystal lattice. When this level of

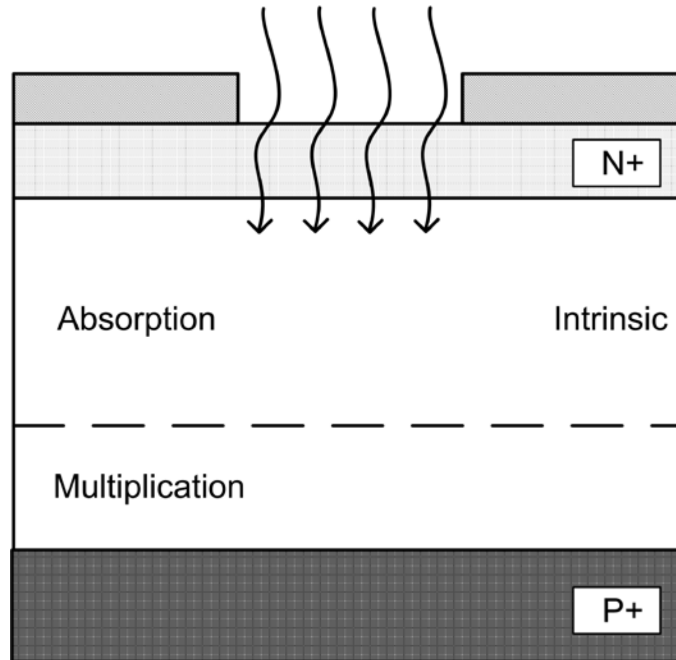


Figure 2.6: Cross section of a typical APD. Illustrated is the doping profile as well as the path of optical illumination. For the device to operated properly the photons must be absorbed in the intrinsic layer and the device must be biased sufficiently high so as to produce impact ionization events from the generated carriers.

electric field strength is reached the electrons drift speed becomes saturated at an average speed (about 10^7cm/s in silicon) known as the saturation velocity [24]. Increasing the electric fields strength beyond this point will cause electrons to gain enough energy between collisions so that when they impact the crystal lattice new electron-hole pairs are generated. These newly generated electrons can obtain sufficient energy to ionize other carriers and the process continues like a chain reaction that amplifies the original signal.

Like the PIN photodiode, the main factors that determine the response speed of an APD are the junction capacitance (τ_{RC}) and the transit time (τ_{drift}). The junction capacitance is inversely proportional to the width of the depletion layer in the absorption region. The larger the depletion layer the smaller the capacitance. As stated earlier, a larger depletion layer gives the added benefit of increased quantum efficiency in the device. But, like the PIN photodiode, there

is the tradeoff with carrier transit time through the device. The transit time needs special consideration in the APDs because of the collisions that take place during the avalanche breakdown process.

Material considerations become a big factor in the design of APDs due to the effects of the ionization coefficients in the amplification process. This topic will be discussed in further detail in a later chapter on the theory of impact ionization. At this point it is sufficient to say that semiconductor materials experience ionization event at different rates for holes and electrons. Furthermore, the optimal absorption material and multiplication material are rarely the same.

Another consideration needs to be given to APD transit time that is not present in typical photodiodes. A longer time may be required to traverse the gain region in an APD than is required to traverse a typical absorption region. At high gains extra time is required for the multiplication process as collisions are readily taking place. This time becomes more apparent and problematic in devices that are biased for gains in the hundreds to thousands range. This delay is known as the multiplication time.

Many other factors must be taken into consideration when discussing devices that utilize impact ionization as the gain mechanism. Some of these consideration include ionization coefficients, biasing, dark current, noise etc. These factors will be discussed more readily in the following chapter.

2.2.5 SAM APD

The semiconductor material of an APD (or any photodiode for that matter) determines the optical wavelengths that will be absorbed. The material will also determine the rate at which impact ionization events occur given electric field strength. With these two facts in mind it becomes apparent that the optimal material for absorption may not (and often is not) the optimal

material for multiplication. To combat this problem, time and effort has gone into the development of an APD with different absorption and multiplication materials. The Separate Absorption Multiplication APD, as shown in Figure 2.7, is such a device [25]. These devices utilize heterostructures and somewhat elaborate fabrication techniques to match the lattices of one material to that of another [26].

Figure 2.7a shows a basic InGaAs-InP SAM APD structure [27]. This type of APD is used in many current optical systems. The band diagram is shown in Figure 2.7b. The techniques used to reduce the barrier from the heterojuction are quite difficult and drive up the cost of such devices.

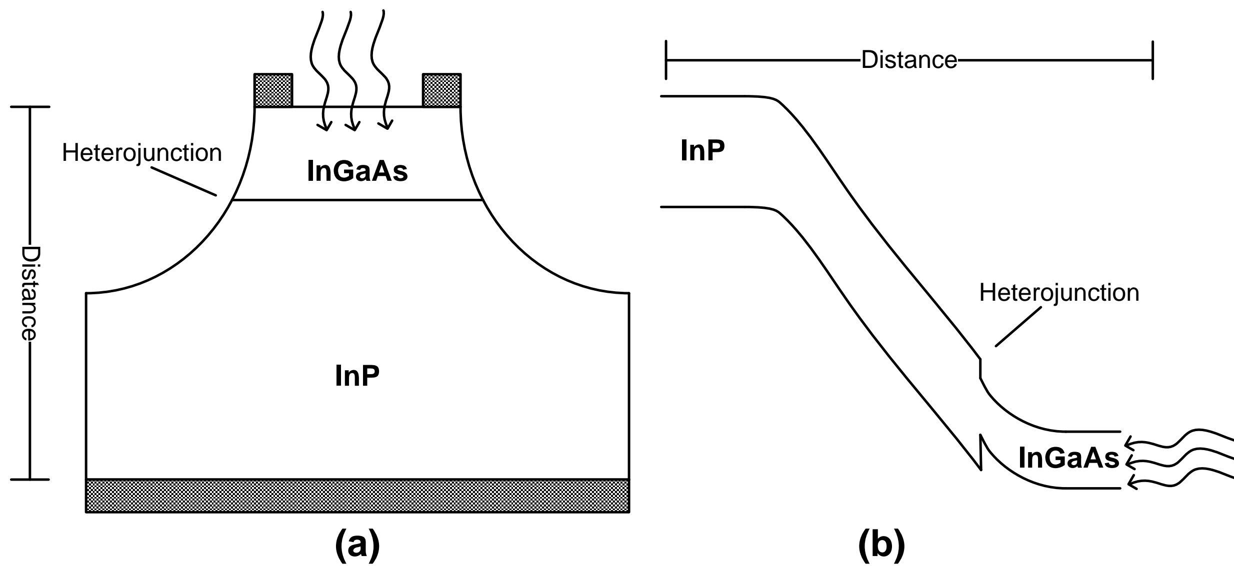


Figure 2.7: InGaAs-InP SAM APD structure. a) Shows the cross section of the device. b) Illustrates the band diagram for the structure shown in (a). Take note of the corresponding distances through the device. Also, note the heterojunction in the band diagram. This device attempts to take advantage of the desirable absorption properties of InGaAs and the more favorable multiplication properties of InP.

2.2.6 Single Photon APD

To be able to detect single photons in optical systems is a difficult problem [28]. A single photon incident on a semiconductor device will (at most) create one electron-hole pair. To be able to amplify the detection of a single photon is a difficult problem and takes very precise fabrication, biasing and noise threshold control. Photomultiplier tubes [29] (these will be discussed in the next section of this chapter) and APDs are both successfully used in detecting single photon events to some degree. But, the lack of resolution when amplifying a singular event makes it difficult, if not impossible, to monitor the event rate in these devices at appreciable frequencies.

Very high gains are obviously necessary when amplifying a single photon. The electron-hole pair generated through the absorption of this photon must create a perturbation sufficiently large to escape the noise threshold of the device in steady state. Because of these sensitive detections, an APD cannot simply be set up in a conventional manner. For single photon detection gains in the thousands are often necessary. Also, the dark currents and noise levels must be sufficiently small in steady state so as to not drown out any incoming signal. In order to achieve these results a device must be designed to handle over-biasing while maintaining a non-breakdown state of effectively zero current for relatively large time periods (milliseconds or more) while awaiting an incoming signal. As shown in Figure 2.8, when a trigger even occurs there is a corresponding jump in the current. The current versus voltage of a theoretical Single Photon Avalanche Diodes (SPADs) is compared to a typical APD response [30].

APDs capable of amplifying single photon signals are known as (SPADs) or Geiger-mode APDs (GPADs) [31], [32]. In a Geiger counter an inert gas is placed in between two highly biased electrodes [33]. When a radiation is introduced in this gas it creates a conductive

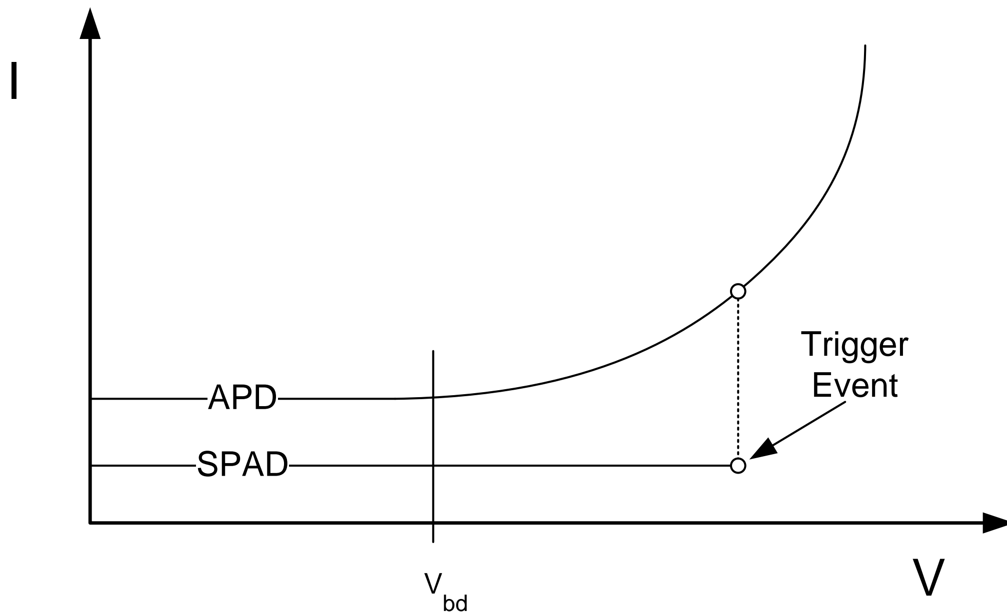


Figure 2.8: Current versus voltage curves illustrating the over-biasing of a SPAD compared to a regular APD. Notice that the current before breakdown for a SPAD is also lower than that of an APD in order to reduce premature breakdown at such high bias levels. It should also be noted that the breakdown of the SPAD doesn't occur until a 'trigger event' initiates the breakdown burst.

region. This conductive region rapidly cascades creating a current pulse detected by the counter. SPADs or GPADs operate in somewhat the same way. These diodes operate in a state of over bias. If a photon enters the SPAD in this state, it acts as a trigger event setting off a series of ionization events. These ionization events cascade causing a detectable breakdown current.

In describing this effect, it becomes apparent that (like a Geiger counter) single events are all but impossible to detect. The detected breakdown current is identical for 1 photon or 100 photons. In both cases a burst of current is detected indicating an ionization event occurred. The problem is that the current burst duration exceeds the necessary time to be able to resolve single ionization events. Often bursts must be stopped by a quenching process. This means that the

applied bias is temporarily lowered to stop the breakdown from continuing. During the quenching process any photons incident on the device will not be detected further exacerbating the resolution issue. Also, circuitry used to quench can introduce parasitic effects into the overall detection system [32].

Furthermore, operating so close to the edge of breakdown gives rise to false positives [34]. In these cases a carriers generated through thermal activity or other common means can cause a breakdown events; this burst is indistinguishable from one caused by photon absorption. Afterpulsing is also a frequent cause for false current bursts [35]. Afterpulsing occurs when carriers find themselves trapped in the lattice or shallow states in the energy bands after a burst has been quenched. The carriers are somewhat easily freed from these states when the bias is reapplied after the quench, thus, causing a new burst. Careful consideration to temperature and biasing must be taken to prevent this from occurring. SPADs are often cooled during operation and need to be carefully setup to prevent false positives bursts.

Along with the problem of low resolution, false positives, quenching and afterpulsing is that of quantum efficiency. Like all optical detectors, not every photon incident on the device will produce an electron-hole pair due to surface reflections, recombination, etc. Furthermore, even when an electron-hole pair is generated there is no guarantee that it will avalanche into a current burst. We are starting to see that many factors affect the operation of SPADs and GPADs. Devices designed to operate this close to the edge of instability often experience similar problems. Because of all of these bandwidth issues, even the most well designed single photon counting systems only operate in the megahertz range. Single photon counting technology in solid state devices still has a ways to go before it will be capable of being implemented into modern telecommunication systems.

2.3 Photomultiplier Tubes

Photomultiplier Tubes (PMTs) are often considered older technology [36]. This is somewhat understandable since their operation is based on vacuum tubes [37]. Despite this fact, they are still used in highly sensitive optical detection systems [29]. They operate through the photoelectric effect. Figure 2.9 illustrates how they operate. A cathode absorbs a photon and emits an electron. Other electrodes, referred to as dynodes, are successively arranged in sequence. Each dynode is biased higher and higher as the sequence progresses through the PMT. The voltage drop from the first dynode to the last dynode can be very large, as much as a few thousand volts. As the original electron is emitted from the cathode it travels from dynode to dynode causing more and more electrons to be emitted. This cascading effect can produce gains (M) in the range of millions depending on the number of dynodes, as seen by the following equation:

$$M = b^n \cdot \left(\frac{V}{n+1} \right)^{kn} . \qquad \textbf{2.10}$$

In Equation2.10, k is a function of the electrode material, b is an environmental constant and n is the number of dynodes in the system. This avalanche effect is somewhat similar to that of an APD as described earlier in this chapter. PMTs are usually quite sensitive, even capable of single photons, and the gains quite large reaching the thousands [38]. Because of this they have been known to burn out if too much light is allowed to be absorbed initially by the cathode. This sensitivity makes it widely used today in very low light optical detection systems.

The response time of a typical PMT is in the range of 0.1-10ns [39]. This illustrates that there is a significant tradeoff between the response time of the device and the gain. Notice, in Equation 2.10, the gain is dependent on the number of dynodes n. There is also a inverse

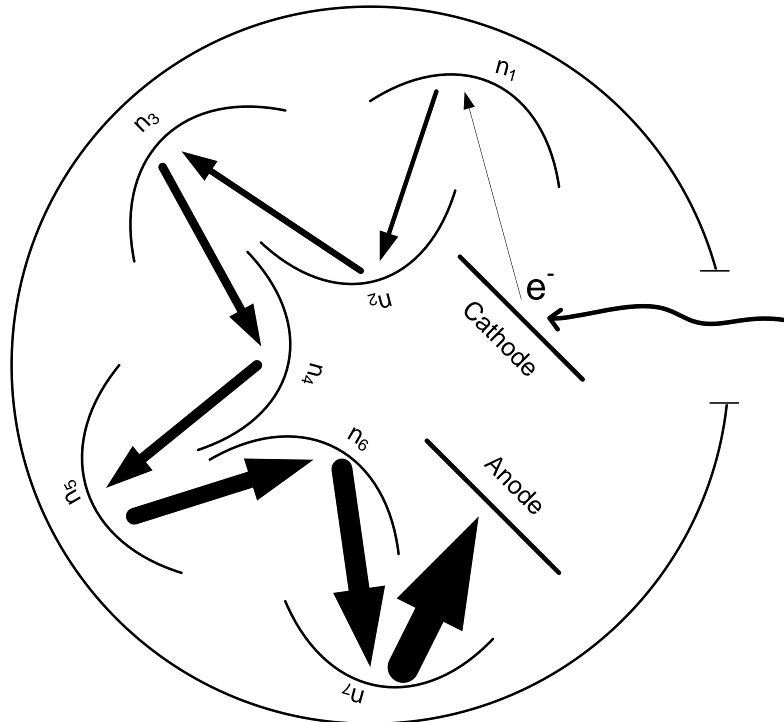


Figure 2.9: Illustrates a typical PMT structure. A photon enters the device and strikes the cathode emitting an electron. The arrows illustrate the path and magnitude of the signal as it is amplified by the dynodes and finally collected by the anode.

correlation between the number of dynodes and the speed of the device [37]. The carrier transit time from cathode to anode increases with each additional dynode stage. If high gains are needed, but faster speeds are desired, the number of dynodes can be decreased and the speed somewhat maintained by increasing the voltage drop from one dynode to the next. This allows higher gains to be achieved and keep the number of dynodes lower. This also introduces tradeoffs in that there is an increase in the emission time of secondary electrons at each dynode. Careful design for the use of a particular PMT must be taken into consideration when setting up the number of dynodes and the voltage drops between the dynodes to achieve the desired speeds and gains.

2.4 Transistor Based Amplifiers

There are many types of signal amplifiers based on transistors. Transistors make up the foundation of active circuit elements in current technology [40]. A typical use of a transistor is that of a switch. A transistor can be used as a switch because voltages and currents applied to one terminal of the device affect the currents and voltages at different terminals. The two main types of transistors are those based on bipolar junction transistor (BJT) technologies and field effect transistor (FET) technologies.

Transistors can also be used in circuits to create amplifiers. These transistor-based amplifiers can be quite simple or very complex. In their simplest form, the amplifier could be nothing more than a single BJT or FET set up properly to amplify an input voltage, as seen in Figure 2.10. More complicated forms are capable of performing many sophisticated electrical and mathematical operations and conversions. Often, these types of amplifiers are used to change the input signal to a different form at the output (from current to voltage or vice versa), as seen in Table 2.1. Because of their diversity and multiple uses, amplifiers based on transistors are the most common amplifier in current electrical systems.

Whole books have been written on transistor based amplifiers [41]. It is sufficient to say at this point that they come in many different varieties, complete many different purposes and are of absolute importance to modern circuits [42].

2.5 Limitations of Present Day Detectors and Amplifiers

Detectors and amplifiers are really the foundation for optical systems. Modern day systems implement different types of detectors and amplifiers as seen in this chapter. Some are

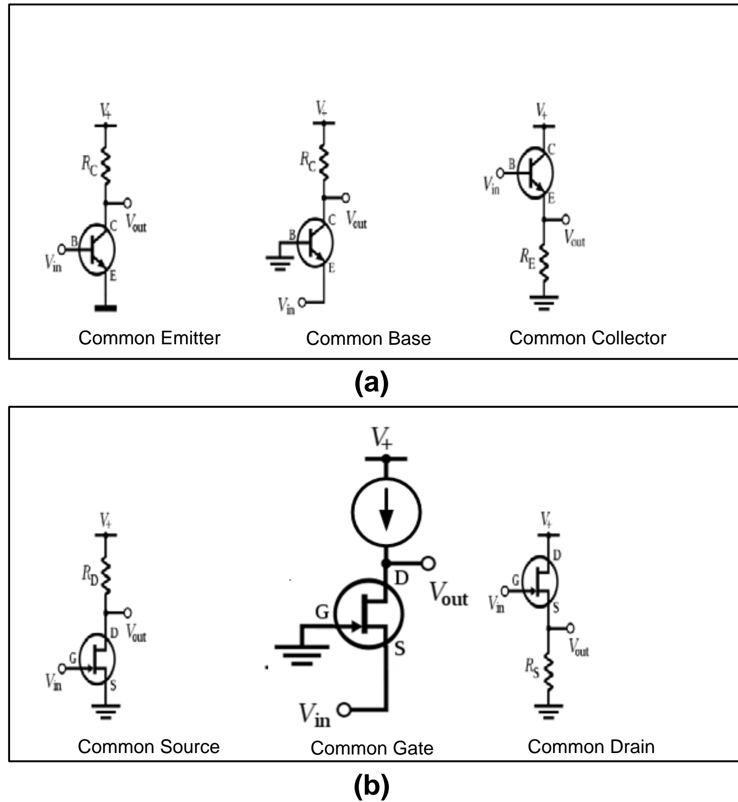


Figure 2.10: a) BJT common amplifier configurations b) MOSFET common amplifier configurations. These illustrate the simplicity of some types of amplifier circuits based on transistor technology.

detectors and amplifiers combined in one device like the APD, SPAD and PMT. Also discussed were other devices that are only detectors and lack an internal amplification mechanism such as the pn junction, Schottky, and PIN photodiodes. Furthermore, there are signal amplifiers that lack any sort of optical detection mechanism such as the transistor-based amplifiers. All of these

Table 2.1: Table showing the different types of transistor based amplifiers.

Amplifier Type	Input	Output
Current Amplifier	Current	Current
Transimpedance Amplifier	Current	Voltage
Transconductance Amplifier	Voltage	Current
Voltage Amplifier	Voltage	Voltage

devices have roles in optical detection systems but also have limits to their capabilities in these systems. Even with all of the time and money spent on developing detectors and amplifiers many of these devices are very good at one thing but lack at another. Some are sensitive but bandwidth limited, others robust yet one dimensional in usefulness. To be able to combine the pros of each type of devices would be very useful. Better yet, would be the ability to break down devices into their respective advantages and mix and match these to create an optical system. The remainder of this section will discuss some of the fundamental limitations to the different types of devices previously discussed in this chapter.

The photodiode is necessary for telecommunication optical detection systems. The photodiode comes in many different forms and varieties. They perform well and have little room for improvement in most applications. They are small, easily integrated, and robust. The PIN photodiode is rarely the bandwidth limiting factor in an optical detection system. They are commonly combined with a Transimpedance Amplifier (TIA) in optical detection systems. This combination is effective in many systems but their application into lowlight systems is the main, glaring drawback. This is often because the post-detector amplifiers (TIAs) have noise thresholds above that of the output signal of the photodiode. Thus, the TIA sets the noise floor for the system. This noise will drown out sensitive signals and reduces the effectiveness of the photodiode's capabilities.

This problem has been somewhat remedied by the introduction of the APD. The APD, as discussed earlier, has an internal gain mechanism based on impact ionization that makes low-noise signal amplification possible. The ability to amplify the detected signal via impact ionization greatly increases the sensitivity of the APD over the PIN in low-light systems. APDs, however, are limited by materials [43]. A PIN photodiode can often be fabricated from the

optimal semiconductor material for a desired wavelength. An APD can also be optimized for detection of a desired wavelength. At this point, one might ask: "*then what's the problem?*" The problem is that in a typical APD, the detector material and the multiplication material are the same. Rarely is the same material efficient at both detection of a desired wavelength and effective in the multiplication process. The reasons behind this will be discussed more thoroughly in the following chapter. It is sufficient to say at this point that an ordinary APD cannot be optimized for both detection and amplification. Furthermore, using the same material for both is often not desirable.

This has become obvious in modern day optical communication systems. In most high speed fiber optic networks the optimal wavelength range for low loss, high efficient systems, is 1300-1550nm [44]. A typical detector material for this wavelength range is InGaAs. This material exhibits very high quantum efficiencies in this range but, due to its material properties, InGaAs is poorly fit for impact ionization based gain. To overcome this, SAM APDs have been fabricated in attempts to effectively separate the absorption and multiplication regions to different materials. As was stated earlier, more desirable multiplication materials are grown epitaxially on InGaAs. This would be a perfect solution if it weren't for the fact that not all materials have compatible crystalline structures. Incompatible band gaps can't be lattice-matched. Without properly lattice matched materials the barrier between the two introduces resistive effects that hinder the bandwidth of the device greatly even when growth is possible. When two materials with mismatched lattice constants are combined it introduces strain in the lattice at the barrier. It also introduces crystalline defects.

For example, silicon is the optimal semiconductor material for clean impact ionization. The reasons for this will be discussed in the following chapter. Unfortunately, silicon cannot be

epitaxially grown onto InGaAs or any other III-V materials because of the large differences in the lattice constants of silicon and these materials [45]. If silicon were capable of being combined with these materials the drawbacks associated with it would more than negate any positive contributions the combination would provide.

There is constant effort to find compatible absorption and multiplication materials that have desirable properties. In typical systems, InP is lattice matched to InGaAs. But InP is far from ideal when it comes to impact ionization properties [46]. Much research is still being done to combine materials for optimal gain and detection. A lot of progress has been made in the development of APDs over the years. New techniques have increased their sensitivity such as in SPADs. But SPADs suffer from the same problem as a typical APD in their poor ability to efficiently absorb and multiply a signal. PMT's have also been discussed but, due to their bulky nature, dated technology and bandwidth, they aren't practical in modern day telecommunication systems.

What seems to be absent is a device that can be connected to an independent photodiode to amplify the signal. This is often done with transistor based amplifiers. The output of a photodiode can be fed into the input of a transistor based amplifier and amplified. The problem with this is that transistor based amplifiers are noisy. Because of this, their noise threshold levels consume the photodiode's output single in most practical systems. The gain mechanism in transistor based amplifiers is dependent on supplemental circuitry which can make them bulky and difficult to implement.

Ideally, there would be a device that could be wired up to an arbitrary photodiode and amplify the signal using a clean gain mechanism such as avalanche breakdown. The Solid-state Impact-ionization Multiplier (SIM) has been developed in hopes to be that device.

2.6 The Solid-state Impact-ionization Multiplier (SIM)

The SIM is a stand-alone device designed to amplify a current signal utilizing impact ionization based gain [47], [48], [1], [49], [50]. It is a silicon device which takes advantage of low cost substrates with very well known fabrication techniques and technology. Silicon also has the most desirable properties for impact ionization based gain among semiconductor materials as will be discussed shortly.

A device that can take an arbitrary current source as an input signal is very useful in many modern day systems. This is especially true in optical communication systems. It makes it possible for the optimal detector material to be used in any system and still be able to take advantage of low noise gains produced through impact ionization. In theory, the SIM operates by taking the output current from a photodiode as an input. This signal is then amplified through impact ionization in the SIM. The SIM doesn't distinguish between the different current sources and can easily be integrated into systems optimized for different wavelengths.

This dissertation is focused on the SIM, its development, and its evolutions. The development and characterization of the SIM has continued to the present day.

2.7 Conclusions and Overview

This dissertation's purpose it to delve into how the SIM operates as well as its evolutions. This chapter outlined many present day optical detectors and signal amplifiers. The convenience of a device such as the SIM was outlined as well. Chapter 2 covers the theory of impact ionization. Chapter 3 deals with modern day simulation and modeling techniques used when evaluating fabrication and operation of semiconductor devices. Chapter 4 outlines the concept of the SIM and its fundamental operation. Chapter 5 talks about one of the first major alterations to

the SIM in an attempt to increase the speed and consistency in fabrication of the SIM. Chapter 6 deals with solutions to insufficient gains in the SIM. It discusses the importance of proper electric field distribution as well as carrier confinement. Chapter 7 discusses the need for high frequencies and high gains in the SIM in order for it to be commercially viable. It also discusses some of the factors that have limited the frequency response and the gain. Chapter 8 outlines the most recent alterations and the theoretical optical SIM configuration. Chapter 9 concludes the dissertation outlining some of the future work and applications that could be applied to the SIM.

3 THEORY OF IMPACT IONIZATION

3.1 Introduction

When talking about devices that utilize avalanche breakdown as a gain mechanism it becomes necessary to take a closer look at the theory of impact ionization. In most modern day semiconductor devices, impact ionization is an undesired process [51]. When impact ionization begins to occur in a semiconductor device it is usually referred to as having reached a state of 'breakdown.' Breakdown is often undesirable in a solid state device because normal operation ceases at this point. There are, however, devices that take advantage of the point where impact ionization occurs and experiences 'breakdown'. Some of the most common of these types of devices are the avalanche photodiode (APD) and the IMPact ionization Avalanche Transit-Time (IMPATT) diode [52]. In these types of devices, the avalanche *breakdown* is usually referred to with the less abrasive names of avalanche *multiplication* or avalanche *gain*.

Impact ionization has been observed for decades in solid state devices. Over the years this process has become a source of high- speed, low-noise gains [53]. It has lead to very refined devices, such as the previously mentioned APD, that can take very low optical signals at the input and produce appreciable electrical signals at the output. A lot of time and effort has gone into the solid-state physics of avalanche multiplication based devices. Given the correct conditions, impact ionization will occur in all known semiconductors. This chapter gives a general overview of the impact ionization mechanism. Even though the material and equations

in this chapter are very broad reaching, the nature of impact ionization makes them applicable to a wide range of semiconductor materials.

3.2 Electron Multiplication via Impact Ionization

In semiconductor devices, impact ionization occurs when an electron is removed from the valence band to the conduction band by means of energy transfer from another electron. When an electron enters an electric field in a semiconductor it will begin to drift through that semiconductor in the opposite direction of the electric field. As it travels through the electric field, its drift velocity depends on the strength of the field. Eventually, the average drift velocity of electrons present in the field reaches a point of saturation known as the saturation velocity. Increasing the strength of the electric field beyond this point will not increase the average drift velocity of the electrons in the material, due to the collisions and interactions between the free electrons and the lattice, but it will increase the average energy of these free electrons. In the presence of these large electric fields, the energy of electrons that interact with the crystal lattice can be larger than the threshold energy. When this happens the electrons transfer their energy through impact ionization events.

As stated earlier, impact ionization events occur when an energized electron collides with the lattice bringing an electron from the valence band to the conduction band [4]. This process is illustrated in Figure 3.1. As seen in the figure, the newly ionized electron is now free to drift through the electric field in the same way as the original electron. If these electrons gain sufficient energy they, also, will produce ionization events. This process rapidly increases as the device enters avalanche multiplication. Figure 3.1 illustrates just two ionization events by the free carriers in the high field region.

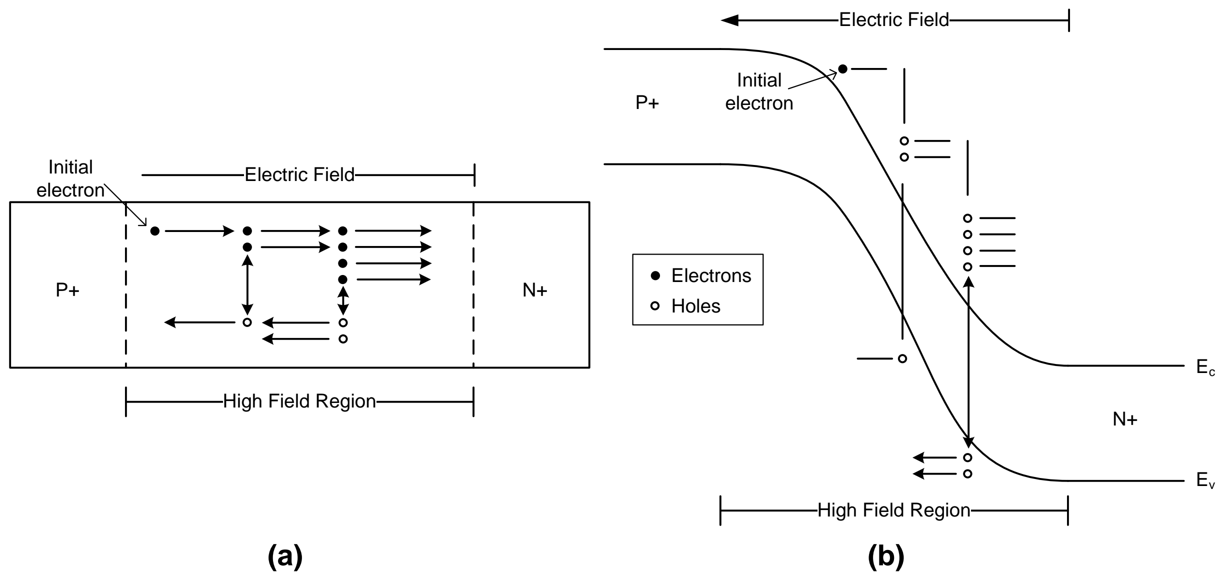


Figure 3.1: a) Slab diagram illustrating the process of impact ionization of electrons in a high biased pn junction. b) Corresponding band diagram to (a). The diagram illustrates the transit of carriers through the electric field as they participate in impact ionization events.

Clearly, not all of the free electrons within the semiconductor material experience impact ionization events at the same rate as they drift through the electric field. Some of the factors that influence this rate and how this rate affects the avalanche process will be discussed throughout the remainder of this chapter.

The bandgap energy is the difference in energy from the top of the valence band to the bottom of the conduction band in a material. The energy states in the bandgap are forbidden and can't be occupied by carriers. With this in mind, for an electron to go from the valence band to the conduction band it must absorb at least the equivalent of the bandgap energy. As seen in the previous chapter, this energy can be absorbed via a photon. The energy of the photon necessary for this to occur was outlined in Equation 2.1. Sufficient energy can also be transferred between an energetic/excited electron in the conduction band to an electron in the valence band via collisions with the lattice. In an ideal situation, the minimum energy necessary to excite an

electron from the valence band to the conduction band is the energy difference between the two bands. It takes a perfectly elastic collision, accompanied by perfect transfer of energy, for a carrier with the minimum energy (E_g) to successfully ionize another carrier. It should be sufficient to say that the transfer of energy isn't ideal. Because perfect transfer of energy from one electron to another isn't realistic, more complete models have been developed estimating the energy necessary for impact ionization. Through these models it can be shown that an energy greater than that of the bandgap is necessary. One of the most widely accepted models was presented by Wolff and will be outlined in the remainder of this section [54].

Wolff based his analysis of energy transfer in ionization events on a parabolic band structure as in Figure 3.2. He used this simple structure, along with conservation of energy and momentum, to come to his conclusion. The equation defining the E-k relationship for the band structure shown in Figure 3.2 is

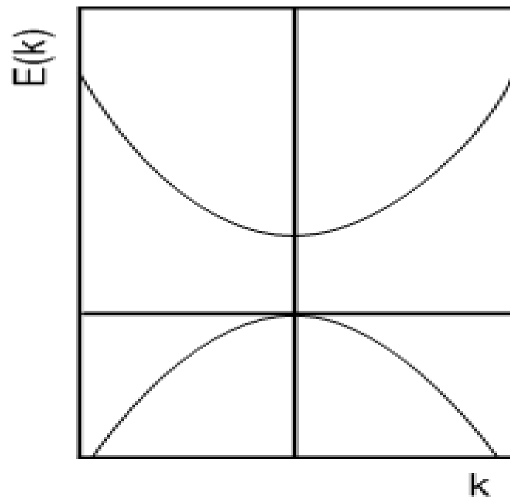


Figure 3.2: Simple parabolic estimation for the conduction and valence bands for a semiconductor material. Simplifications of this nature are often made in order to analyze complicated semiconductor band structures.

$$E(k) = \frac{\hbar^2 k^2}{2m^*}, \quad \textbf{3.1}$$

where $\hbar$ is Plank's constant, k represents distance in k-space and m^* is the carrier effective mass. This is somewhat of a stretch into ideality given the complex band structure of semiconductor devices. More complicated, yet still very simplified, version of band structures are also used to analyze other phenomenon. One example of this is in Figure 3.3, which shows a simplified version of the band structure of silicon.

Compare Figure 3.3 with Figure 3.4 [55]. They both represent the band structure of silicon with one giving more detail. Some of the intricacies of these band diagrams will be discussed in further detail later on. But it should be noted that in both Figure 3.3 and Figure 3.4 that the bands near a minimum can be approximated by a quadratic. It is also important at this juncture to point out that the band structure of a material determines the effective mass of the carriers in that material [56]. Given the ideal band structure produced by Equation 3.1, the effective mass is constant and the same as the real mass as described by

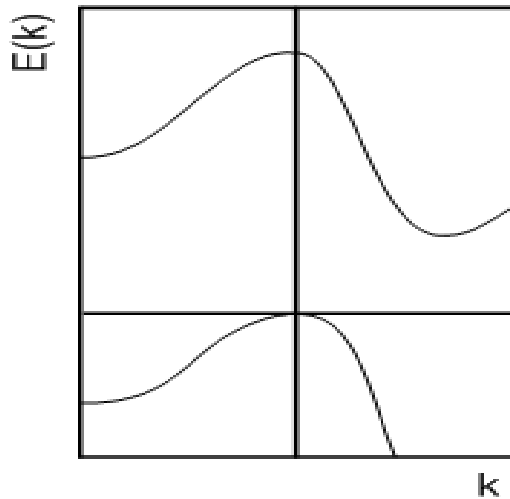


Figure 3.3: Simplified version of the band structure of silicon.

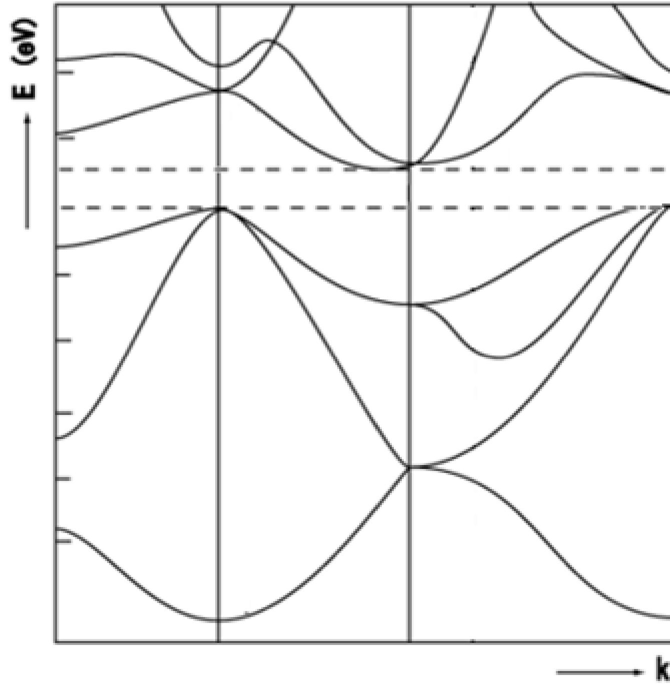


Figure 3.4: Detailed band structure for silicon. This more complex representation shows wavevector dependence. Going from left to right along the horizontal, the wavevector is tracing out a particular one-dimensional path through the three-dimensional momentum space. The vertical axis represents energy.

$$m^* = \hbar^2 \left(\frac{\partial^2 E(k)}{\partial k^2} \right)^{-1} . \quad \textbf{3.2}$$

Armed with this knowledge, we delve into Wolff's analysis of energy transfer in impact ionization events [54].

Considering the parabolic bands shown in Figure 3.2, the conduction band has an effective mass of m_e and a valence band with effective mass m_h. An electron introduced into an electric field in this semiconductor material will travel through the field with an initial velocity of v_i. As it drifts through the field it has a kinetic energy of $1/2\, m_e^* v_i^2$. The momentum would then be $m_e^* v_i$. If that initial carrier were to participate in an ionization event there would be three carriers (the initial electron and the new electron-hole pair). The electrons continue to drift in the

same direction as the initial electron and the hole drifts in the opposite direction (the same direction as the electric field). If we make the assumption that the ionization event occurred after a perfectly elastic collision involving the initial electron, we have conservation of energy in

$$\frac{1}{2}m_e v_i^2 = E_g + \left(\frac{1}{2}m_e v_e^2 \cdot 2 + \frac{1}{2} \cdot m_h v_h^2\right), \qquad \textbf{3.3}$$

where the left hand of the equation represents the energy of the initial electron and the right hand represents the energy of the three carriers after the ionization event. We also have conservation of momentum in

$$P_{i,e} = m_e v_i = m_e v_e \cdot 2 + m_h v_h \,. \qquad \textbf{3.4}$$

In this simplified case, we can assume that $m_e = m_h$. We also assume that the velocity the ionized electron-hole pair is equal and defined as $v_e = v_h = v_f$, With this assumption, Equation 3.3 and Equation 3.4 become:

$$\frac{1}{2}m_e v_i^2 = E_g + \frac{3}{2}m_e v_f^2, \qquad \textbf{3.5}$$

and

$$m_e v_i = 3 m_e v_f \,, \qquad \textbf{3.6}$$

respectively. If we now combine Equation 3.6 and Equation 3.5 for energy, we are left with

$$E_i = \frac{1}{2}m_e v_i^2 = 1.5 E_g \,. \qquad \textbf{3.7}$$

By following Wolff's method, the energy needed for an initial electron to cause an ionization event to occur in these ideal circumstances is 3/2 the band gap energy.

This estimation is fairly accurate for some semiconductor materials but the assumptions become foggy for materials with complicated band structures or those with an indirect bandgap such as silicon. Notice in Figure 3.3 and Figure 3.4 that the low energy point in the conduction

band and the high energy point in the valence band do not line up in momentum (k) space [57]. This is the definition of an indirect bandgap. Because of this, for an electron to be excited from the valence band to the conduction band, there must not only be an absorption of energy but a shift in momentum. Because of this, electrons in silicon require a much higher energy than $\frac{3}{2}E_g$ to initiate an ionization event. Wolff shows this to be true reporting a measured energy of 2.6eV for an electron to initiate impact ionization and 5eV for hole initiated impact ionization in silicon [54].

3.3 Ionization Coefficients

Impact ionization is probabilistic in nature. Due to it being highly engrained in probability theory, there are many ways to look at the nature of impact ionization. It is necessary to understand the rate at which carriers will produce ionization events in a semiconductor material. This rate is known as the ionization coefficient for a given material. This section will attempt to give a brief overview of common method to derive one of the most fundamental parts of impact ionization theory, the ionization coefficient.

Shockley first suggested that impact ionization is due mainly to “lucky electrons.” A ‘lucky electron,’ according to Shockley, is an electron that suffers no collisions long enough to obtain energy exceeding that of the threshold energy [58]. The threshold energy is the minimum energy necessary to bring an electron from the valence band to the conduction band. In semiconductor materials the threshold energy is equivalent to the energy of the bandgap of that material, as shown earlier.

In other words, Shockley theorized that when an electron enters an electric field in a semiconductor it will begin to drift through that semiconductor in the opposite direction of the

electric field. As they drift, they lose energy through collisions with phonons, impurities, or defects in the lattice structure. Carriers that interact or collide with the lattice before exceeding the threshold energy simply transfer that energy to the lattice in the form of lattice vibrations (phonons) instead of transferring that energy in the form on an ionization event. Clearly, not all of the free electrons within the semiconductor material impact ionize as they drift through the electric field. Thus, only those 'lucky' electrons that avoid collisions long enough to obtain sufficient energy for impact ionization transfer energy in this way.

A very different approach to the derivation of the ionization coefficients was taken by Wolff. He suggested that electrons travelling through an electric field in a semiconductor undergo many collisions as they gain energy. The energy lost per collision is small in relation to the total energy of the electron. Thus, the electron distribution develops an effective temperature. When the energy from this electron temperature surpasses the threshold energy ionization events occur. Increased electric field corresponds to an increased electron temperature and thus an increased ionization rate.

After Shockley and Wolff, many other theories and methods have been used in an attempt to derive the ionization coefficients for a given material [59]. Most of which have been some sort of combination between the two methods. These somewhat intuitive methods have helped bring the theory of impact ionization a long way in a relatively short time. With that in mind, modern day computational abilities have changed techniques used in today's research. There has been a shift toward numerical methods for such derivations.

The most widely used of these methods are based upon the Monte Carlo technique [60], [61], [62]. The Monte Carlo technique is not specific to any theory but instead solves widely accepted equations using computational algorithms which use random numbers to converge on a

result. This method is only as good as the number of degrees of freedom allowed in the chosen equations. This method is also very computationally taxing. Some simulations could take days, weeks, or even years to solve if left alone. Typical use of this method in the calculation of ionization coefficients involves solving Boltzmann's equation directly [63]. Things such as carrier path and phonon scattering are selected randomly while taking into consideration things such as field strength, geometries, and lattice makeup. The result is that the drift velocity, average carrier energy and other necessary variables can be obtained.

With these methods in mind, it should be noted that the most reliable technique in obtaining ionization coefficients is by measuring them experimentally. Obtaining the coefficients in this way will be quite accurate given constant variables. The method for measuring the ionization coefficients is outlined later in this chapter.

The problem with this method is when you take into consideration that coefficients are based upon the many factors which include, but are not limited to: electric field intensity, field distribution, lattice structure, and doping. The altering of any of these variables will produce different ionization coefficients. This variance can make it difficult to dynamically design devices that are based upon impact ionization as their gain mechanism. For this reason numerical methods are becoming more widely used in device design. But, once an acceptable design has been given, it remains necessary to measure the ionization properties of a fabricated device.

The first step in obtaining impact ionization coefficients analytically is by understanding the impact ionization gain equations. These equations make it possible to extract ionization coefficients given proper system set up. The next sections analyze the gain equations and the measuring of the ionization coefficients.

3.4 Impact Ionization Gain Equations

It is necessary to emphasize at this point, even though it has been mentioned previously, that holes also participate in impact ionization events in semiconductor materials. Like the electrons, this occurs when a hole obtains sufficient energy to overcome the threshold energy in that material. It is also important to note, as has been alluded to earlier, that the threshold energy is nearly always different for holes and electrons in a given semiconductor material. Because of this, and other factors, the rate at which holes and electrons ionize secondary pairs is also different. As discussed earlier, this rate is known as the ionization coefficient. The electron and hole ionization coefficients are represented by α and β respectively [64]. These coefficients are dependent upon many factors including the semiconductor material, the electric field strength and the electric field distribution.

The units for the electron and hole ionization coefficients are in inverse meters, typically represented as cm^{-1}. This is because α and β represent the reciprocal of the average distance an electron or hole must travel before experiencing an ionization event. The fact that holes and electrons ionize at different rates becomes very important in practical devices where it is nearly always desirous that the ratio of ionization coefficients (k) for electrons and holes be either very large or very small. In other words, it is desirous that either the electrons or holes contribute to the majority of ionization events [65].

From semiconductor to semiconductor the fundamental equations that describe the gain characteristics are nearly identical under normal avalanche multiplication. The main variables that change from material to material are the threshold energy (band gap) and the ionization coefficients. With this in mind, we now delve into a general explanation of the gain characteristics of semiconductors operating in avalanche breakdown.

To obtain equations that describe impact ionization, we start with the simplified setup shown in Figure 3.5. Notice that Figure 3.5 describes nothing more than a slab of highly biased doped semiconductor material. Note that the slab has a width *W*. The high bias applied to the slab creates a depletion region and an accompanied electric field. Holes introduced into the electric field naturally drift in the direction of the field and electrons drift in the opposite direction. As we have seen previously in this chapter, and in the previous chapter, this setup is common to achieve impact ionization and resembles that of the gain region in an APD. This model also assumes that there are no carriers being generated in the slab via optical absorption or generation other than through impact ionization. Carriers can only enter the slab at the boundaries. This is very similar to the operation of a typical APD where the absorption and multiplication regions are sufficiently isolated.

Now, given this model, we can describe the rate of change of electron current density and rate of change of hole current density within the slab at a position *x* as

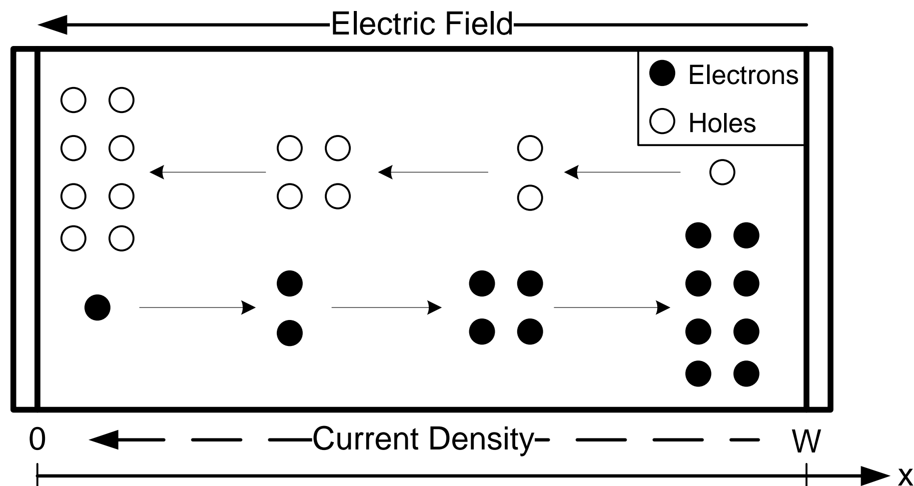


Figure 3.5: Slab diagram showing the current density through a multiplication region in a highly biased semiconductor. This diagram helps set the boundary conditions and framework for the derivation of the gain equations that describe impact ionization.

$$\frac{dJ_n}{dx} = \alpha(x)J_n(x) + \beta(x)J_p(x)\,, \qquad 3.8$$

and

$$-\frac{dJ_p}{dx} = \alpha(x)J_n(x) + \beta(x)J_p(x)\,, \qquad 3.9$$

respectively [66]. As was stated earlier the carriers entering the system must enter at the boundaries of the slab. We will further assume that all carriers entering the system are electrons. Thus, the only holes that will find themselves in the system have been generated via impact ionization events. With this assumption we have the boundary condition $J_p(W) = 0$. This assumption also leads to the current density at *W* to be

$$J_n(W) = J_n(x) + J_p(x)\,. \qquad 3.10$$

If we now combine Equation 3.8 and Equation 3.10 we get

$$\frac{dJ_n(x)}{dx} - \{\alpha(x) - \beta(x)\}J_n(x) = \beta(x)J_n(W)\,. \qquad 3.11$$

Notice that this takes the form of an ordinary differential equation. It can now be solved using the integrating factor $\exp[-\int_0^x (\alpha(x) - \beta(x))dx'] = \exp[-\varphi(x)]$. With this method, and by integrating across the entire length of the intrinsic region (0 to *W*) Equation 3.11 yields:

$$J_n(x) = \frac{\int_0^x \beta(x)J_n(W)\cdot\exp[-\int_0^x \{\alpha(x') - \beta(x')\}dx']dx + J_n(0)}{\exp[-\int_0^x \{\alpha(x') - \beta(x')\}dx']}\,. \qquad 3.12$$

Current gain for electron injection can be defined as

$$M_n = \frac{J_n(W)}{J_n(0)}\,. \qquad 3.13$$

Applying Equation 3.13 to Equation 3.12 yields

$$M_n = \frac{J_n(W)}{J_n(0)} = \frac{1}{\exp\left[-\int_0^W \{\alpha(x) - \beta(x)\}dx\right] - \int_0^W \beta(x)\exp\left[-\int_0^x \{\alpha(x') - \beta(x')\}dx\right]dx} \quad 3.14$$

The second term in the denominator can be written as:

$$-\int_0^W \beta(x)\exp\left[-\int_0^x \{\alpha(x') - \beta(x')\}dx'\right]dx = -\int_0^W \left[\underbrace{\exp\left(-\int_0^x \alpha(x')dx'\right)}_{A}\underbrace{\beta(x)\exp\left(\int_0^x \beta(x')dx'\right)}_{dB}\right]dx \quad 3.15$$

By integrating by parts $\left(\int AdB = AB - \int BdA\right)$ on the second term in the denominator things are simplified. This is setup as:

$$\begin{aligned} AB - \int BdA &= \underbrace{\exp\left[\int_0^x \alpha(x')dx'\right]}_{A}\underbrace{\exp\left[\int_0^x \beta(x')dx'\right]}_{B} - \int_0^W \underbrace{\exp\left[\int_0^x \beta(x')dx'\right]}_{B}\underbrace{\alpha(x)\exp\left[-\int_0^x \alpha(x')dx'\right]}_{dA}dx \\ &= \exp\left[-\int_0^W \{\alpha(x) - \beta(x)\}dx\right] - 1 + \int_0^W \alpha(x)\exp\left[-\int_0^x \{\alpha(x') - \beta(x')\}dx'\right]dx \end{aligned} \quad 3.16$$

Substituting this in for the second term in the denominator of Equation 3.14 yielding this common form of the equation for electron injection gain:

$$M_n = \frac{1}{1 - \int_0^W \alpha(x)\exp\left[-\int_0^x (\alpha(x') - \beta(x')dx'\right]dx} . \quad 3.17$$

The solution for hole injection can be obtained in a similar fashion and is shown here.

$$M_p = \frac{1}{1 - \int_0^W \beta(x)\exp\left[\int_x^L (\alpha(x') - \beta(x')dx'\right]dx} . \quad 3.18$$

If we assume a uniform electric filed across the gain region then Equations 3.17 and 3.18 are simplified to

$$M_n = \frac{1}{1+\frac{\alpha}{\alpha-\beta}\left[\exp\left(-x(\alpha-\beta)\right)\right]_0^W} = \frac{(\alpha-\beta)\exp\left(W(\alpha-\beta)\right)}{\alpha-\beta\exp\left(W(\alpha-\beta)\right)}, \quad \mathbf{3.19}$$

and

$$M_p = \frac{1}{1+\frac{\beta}{\beta-\alpha}\left[\exp\left(-x(\beta-\alpha)\right)\right]_0^W} = \frac{(\beta-\alpha)\exp\left(W(\beta-\alpha)\right)}{\beta-\alpha\exp\left(W(\beta-\alpha)\right)}, \quad \mathbf{3.20}$$

respectively.

Also, we will see shortly in Sections 3.5 through 3.8 that the impact ionization coefficients for holes and electrons vary from semiconductor to semiconductor. In all materials the coefficient for electrons is greater than that of the holes or vice versa. If we assume that $\alpha >> \beta$ for a given material at a given electric field strength then we get the simplified version of

$$M_n \approx \frac{\exp\left(W\alpha\right)}{1-\frac{\beta}{\alpha}\exp\left(W\alpha\right)}, \quad \mathbf{3.21}$$

and

$$M_p \approx 0. \quad \mathbf{3.22}$$

We will see in the following sections that the above case is very desirable in devices that utilize impact ionization for gain.

3.5 Measuring Impact Ionization Coefficients

As stated earlier, the best and most accurate method of obtaining ionization coefficients for a given material is by measuring them experimentally. Theoretically, the process to measure the ionization coefficients for the holes and electrons is quite simple [66]. Measure the electron current through a device without gain. Then, measure the electron current through the same

device with proper biasing to produce impact ionization gain. Divide the measured gained electron current by the measured non-gained electron current to obtain M_n. Repeat this process for hole current to obtain M_p. Apply M_n and M_p to Equations 3.19 and 3.20 to calculate α and β for the given semiconductor material.

In practice, it is a more difficult process than described in the previous paragraph. The concept is the same but the steps are more difficult to realize. The first difficulty is the step of obtaining a measured amount of current through a device without gain. As stated previously, this must be done for electron and hole current. The current must be from pure current source such that original signal consists of only one type of carrier. Without this constraint it couldn't be known which carriers are producing the gain once the device is biased for avalanche breakdown. Furthermore, the amount of current through the device prior to breakdown must be very accurately measured. Typically, these two steps (pure carrier source and accurate measurement) are performed using optical injection for the current and a PIN photodiode as the device, as seen in Figure 3.6. The type of carrier injected into the active region of the device is determined by the doping and the optical source. If the optical source is incident on the P+ doping in PIN photodiode, electron-hole pairs are created as expected. The holes are drawn to the anode while the electrons are drawn into the devices depletion layer. This is the opposite if the optical source is incident on the N+ doping side of the PIN. In this case the holes are drawn into the depletion region and the electrons are drawn to the cathode. Each of these cases is capable of producing a pure source of electron or holes respectively.

Even this process is more complicated than it sounds. Careful doping and selection of the optical source must be taken to ensure the electron-hole pairs are not being generated within the depletion region. This means that the photons must be absorbed in either the P+ or N+ regions

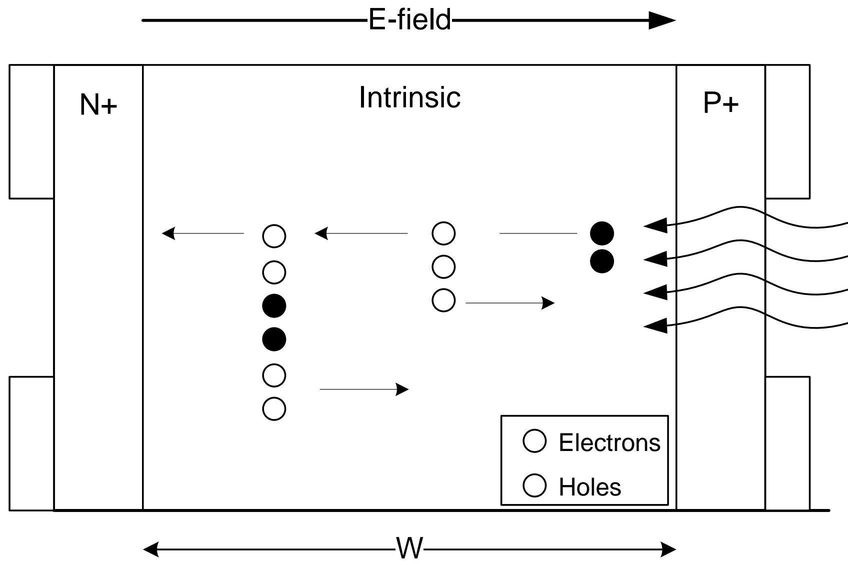


Figure 3.6: Pure electron injection via optical absorption in a PIN photodiode. Biasing to achieve impact ionization is also shown in order to accurately measure ionization coefficients under different fields.

and not in the intrinsic layer. This is to ensure that the carriers in the electric field region of the device are purely one type and not the other. This is done by ensuring the depth of the doping layer is sufficiently large when compared to the optical absorption length for the wavelength/frequency of the optical source. The doping must be quite accurate to ensure the measurement of ionization coefficients is accurate as well.

Another consideration is the consistency in the electric field profile through the depletion region. This field must remain constant and thus requires a very intrinsic layer. Furthermore, the extension of the electric field into the P+ and N+ regions must be minimal through all bias levels in the current measurement process. This requires a sufficiently abrupt junction between the intrinsic layer and the P+/N+ regions. It is also necessary that the doping in the P+/N+ region be sufficiently large.

With the proper setup, values for M_n and M_p can be obtained through measurement. With these values we need a proper relationship between them via the gain equations to calculate the

ionization coefficients. A relationship between M_n and M_p can be obtained with respect to α and β through a little manipulation of Equations 3.19 and 3.20. This relationship is shown here:

$$M_n = M_p e^{(\alpha-\beta)W} \,. \qquad \textbf{3.23}$$

By substituting this relationship into Equation 3.19 we obtain

$$M_n = \frac{(\alpha-\beta)\left(\dfrac{M_n}{M_p}\right)}{\alpha - \left(\beta \dfrac{M_n}{M_p}\right)} \,. \qquad \textbf{3.24}$$

With a little more algebra, the ionization coefficients can be obtained as functions of M_n and M_p as shown in

$$\alpha = \left(\frac{1}{W}\right)\left(\frac{M_n - 1}{M_n - M_p}\right)\ln\left(\frac{M_n}{M_p}\right), \qquad \textbf{3.25}$$

and

$$\beta = \left(\frac{1}{W}\right)\left(\frac{M_p - 1}{M_p - M_n}\right)\ln\left(\frac{M_p}{M_n}\right). \qquad \textbf{3.26}$$

By applying the values obtained for M_n and M_p into Equations 3.25 and 3.26 proper measurement and calculation of the ionization coefficients can be obtained for the material used as the PIN photodiode.

Proper measurement of ionization coefficients is absolutely necessary for accurate evaluation of semiconductor materials and devices [67]. Simulation and modeling techniques rely heavily on these coefficients to obtain accurate predictions of avalanche gain based devices as will be seen in Chapter 4. These values also define a material's usefulness with respect to ionization based gain. Reasons for this will be described in Sections 3.6 and 3.7 on ionization response time and noise considerations respectively.

3.6 Impact Ionization Coefficient Ratio

We just discussed the process of obtaining accurate ionization coefficients for both electrons and holes in semiconductor materials. The ratio of these ionization coefficients becomes vital when choosing a material for devices such as the SIM or an APD. This ratio affects many factors such as response time, bandwidth and noise in a device. These factors will be discussed in the following sub-sections.

3.6.1 Response Time

The ratio of hole to electron ionization coefficients will be defined as

$$K = \frac{\beta}{\alpha}. \qquad \textbf{3.27}$$

As we learned in the previous section the ionization coefficient of a carrier (whether it be a hole or an electron) is defined as the probability of that carrier initiating an impact ionization even given a material and electric field.

Since the ionization coefficients are material specific, and are different for holes and electrons, it becomes apparent that some semiconductors function better than others as a multiplication material in impact ionization based devices. The optimal materials have coefficients where the ratio is either very large (for hole amplification) or close to zero (for electron amplification). Silicon is widely considered the optimal material for multiplication due to its very low K ratio. The benefits of a desirable ratio will now be discussed.

Impact ionization has an associated settling time or response time. This response time is directly associated to K [68]. Semiconductors with a small K will experience a shorter settling time in high electric fields as seen in Figure 3.7. Notice in Figure 3.7a, for a very small K the

electrons are significantly more likely to participate in impact ionization events than the holes. In this case the holes are not participating in impact ionization events.

They simply drift through the field region without producing other electron-hole pairs. This produces a very clean gain and fast transit time as the electrons simply drift across the field gaining up as the go. On the other hand, in Figure 3.7b, notice that the holes and electrons are equally likely to produce impact ionization events. In this case when an initial electron is introduced into the high field region it produces an electron-hole pair through impact ionization. The secondary electron continues with the original electron also producing electron-hole pairs as in Figure 3.7a. Notice, though, that the secondary hole also participates in impact ionization events producing more electron-hole pairs that continue to ionize other carriers as they travel through the electric field. This produces a 'sloshing' effect with an associated settling time. The

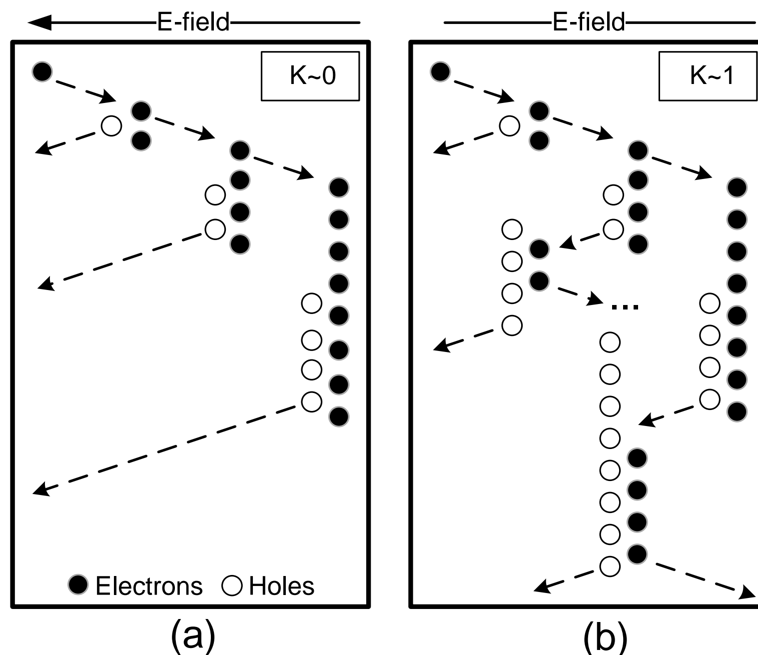


Figure 3.7: Slab diagrams illustrating impact ionization events in different materials. a) This material has a low ratio of α to β yielding a clean gain pattern across the bulk. b) The electrons and holes are equally likely to produce ionization events in this material introducing a long settling time.

response of this device is very slow and (as we will see in the following section) noisy. In theory this could go on indefinitely if the value of K were truly 1.

The ratio of α to β becomes very important when analyzing the response time of the device. It also is paramount when discussing the gain controllability and achievable gain of a device as well. Figure 3.8 [57] shows the gain with respect to the length of the gain region (measured in 'multiplication length'). Notice, for $K \approx 1$ the gain becomes very uncontrollable very quickly as a single carrier introduced into the gain region 'blows up' into theoretically infinite number of ionization events. This is very undesirable when trying to achieve clean controlled gain. As the disparity between α and β increases there is a marked change in the gain profile.

3.6.2 Bandwidth

When K is optimal, the bandwidth of the device is not affected by the ionization coefficients. Meaning if only one carrier is participating in impact ionization events the bandwidth of the device will be affected little, if any, by the response time of the gain region. The theory behind this has shown that, if the steady state gain M_o is less than $1/K$, the response time is a minimal factor [69]. As M_o gets larger than $1/K$ the limit of the gain-bandwidth due to response time becomes

$$M(\omega) = \frac{M_o}{\left(1+\omega^2 M_o^2 \tau^2\right)^{1/2}} \quad \textbf{3.28}$$

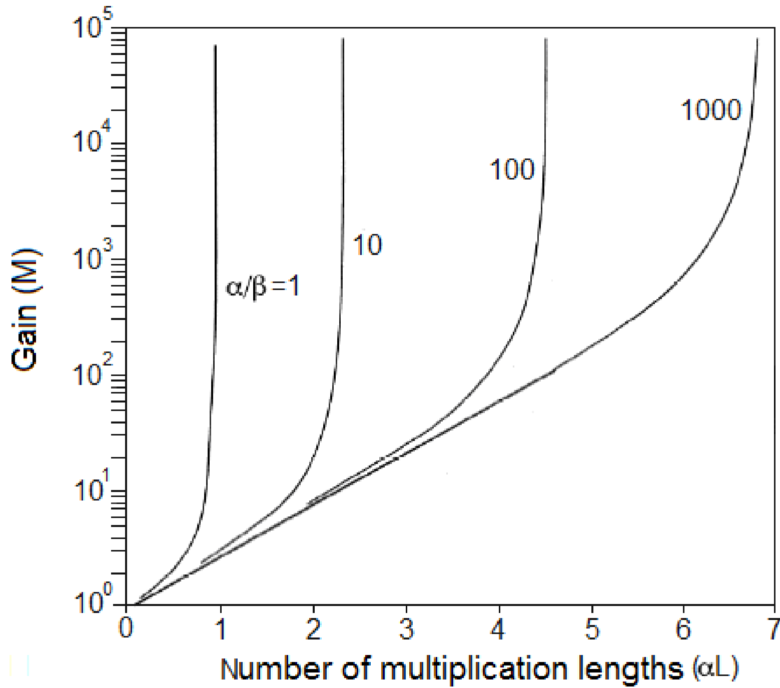


Figure 3.8: Multiplication gain with respect to multiplication length for pure electron injection. Notice change in gain versus length for different ionization coefficient ratios.

In this case τ is an effective transit time as a function of *K*. If the output response is plotted for various *K* values we get Figure 3.9 [57]. This data was taken for a system with a constant gain of 50. Notice that for optimal ratios (close to zero) the frequency response is theoretically flat for any frequency and cannot be limited by the effects of impact ionization. On the other hand, though, if the ratio is not optimal (close to one) many systems are limited in their upper limit frequency by the effects of impact ionization based gain.

As seen in this section the ratio $K = \frac{\alpha}{\beta}$ is of the upmost importance when designing avalanche breakdown devices. This ratio affects the transit time/response time of the device, the

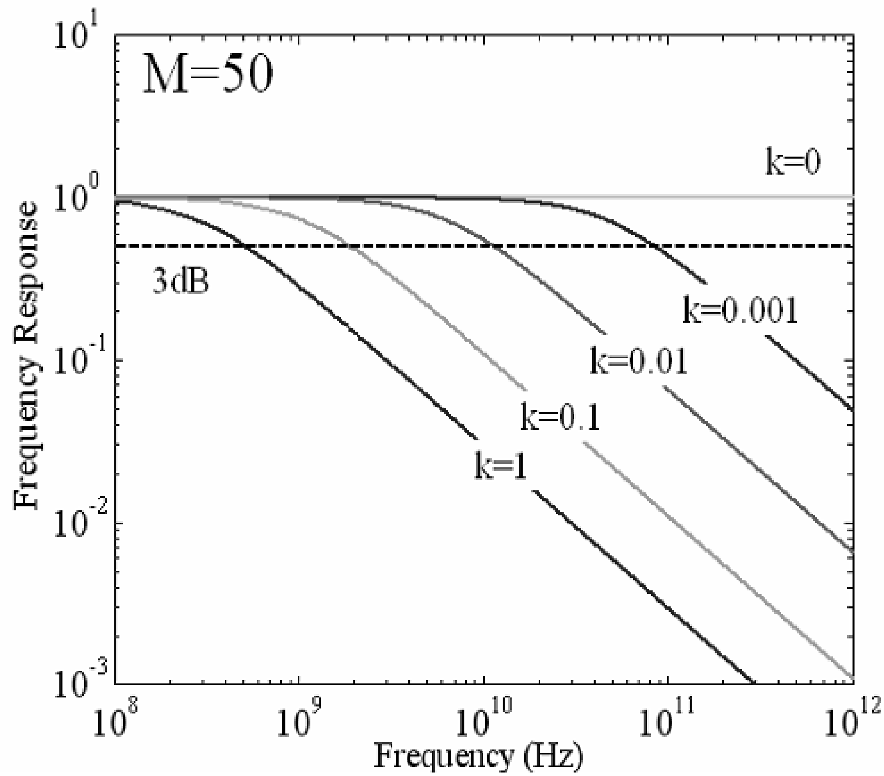


Figure 3.9: Upper limit frequency due to the ratio of α to β. Data was taken for a constant gain of 50 and plotted. Notice a system with k=0 can never be limited in frequency by impact ionization.

gain controllability, and the overall bandwidth capabilities. Also, as we will see in the next section, *K* plays a large part in the excess noise produced via impact ionization.

3.7 Noise

As alluded to in the previous section, the excess noise of an impact ionization gain based device is dependent upon $K = \frac{\alpha}{\beta}$. As was noted in Figure 3.7, when $K \approx 1$, there is a sloshing effect. This unwanted feedback, as both carriers participate in impact ionization events, not only increases the response time of the device but it introduces unwanted signal reproduction. This unwanted gain is nothing more than excess noise added to the clean desirable gain in the system.

As the ration of ionization coefficients reaches a more desirable value the number of undesirable ionization events is lowered and the random feedback introduced into the system (excess noise) is lowered as well.

Because impact ionization is very probabilistic in nature, this undesirable feedback is random. Some carriers will participate in more than the average ionization events and some less than average.

The excess noise produce by uneven participation in ionization events by carriers can be defined as

$$F = \frac{<M^2>}{<M>^2}. \quad \textbf{3.29}$$

In Equation 3.29, $<M>^2$ is the square of the average gain and $<M^2>$ is the mean square gain. What this means is that the noise is the ratio of random gain events by random carriers to the gain produced by equal ionization by all carriers. The noise is then proportional to how evenly the gain is distributed among the carriers in the high field region.

The equations that define the excess noise factor for different carriers are

$$F_e = k_{eff} M_e + \left(\frac{2-1}{M_e}\right)(1 - k_{eff}), \quad \textbf{3.30}$$

for electrons as the carrier and

$$F_h = k'_{eff} M_h - \left(\frac{2-1}{M_h}\right)(k'_{eff} - 1) \text{ [65]}, \quad \textbf{3.31}$$

for holes. In both Equation 3.30 and 3.31 M_e and M_h are the total gain produced by the electrons and holes respectively. What we observe from this is that the excess noise for both the electron and hole produced gain is based is based upon the gain or field the carrier enters and the

ionization coefficient of the carrier in the material. These are the only two controllable factors: material and field.

The effective ratio of coefficients for the material in Equations 3.30 and 3.31 can be defined as

$$k_{eff} = \frac{\int_{x_1}^{x_2} \beta M^2 dx}{\int_{x_1}^{x_2} \alpha M^2 dx} \quad [65]. \qquad \mathbf{3.32}$$

From the previous three equations we begin to see the role that the ionization coefficients have in the excess noise attributed to a device that utilized impact ionization for gain. Notice that as k_{eff} approaches unity the excess noise approaches the total gain of the device. From this it becomes apparent that for excess noise to be minimized a material with optimal ionization coefficient ratios (as close to zero as possible) are desirable. The excess noise for several different materials at a varying level of gain is shown in Figure 3.10 [70]. From this we see how important material choice is when designing multiplication devices. From this figure it becomes apparent that the material of choice for amplification is silicon.

3.8 Impact Ionization in Silicon

With all of the knowledge obtained on impact ionization in the previous sections we now turn our focus on silicon and its ionization properties [71]. The importance of silicon and its impact ionization properties are vital to this document because this is the material chosen for the SIM. Silicon was chosen as the material for the SIM because it has the most favorable impact ionization coefficient ratio of known semiconductor materials [72], [73], [74]. This was seen in several figures in the previous sections. Figure 3.10 showed that the excess noise factor for

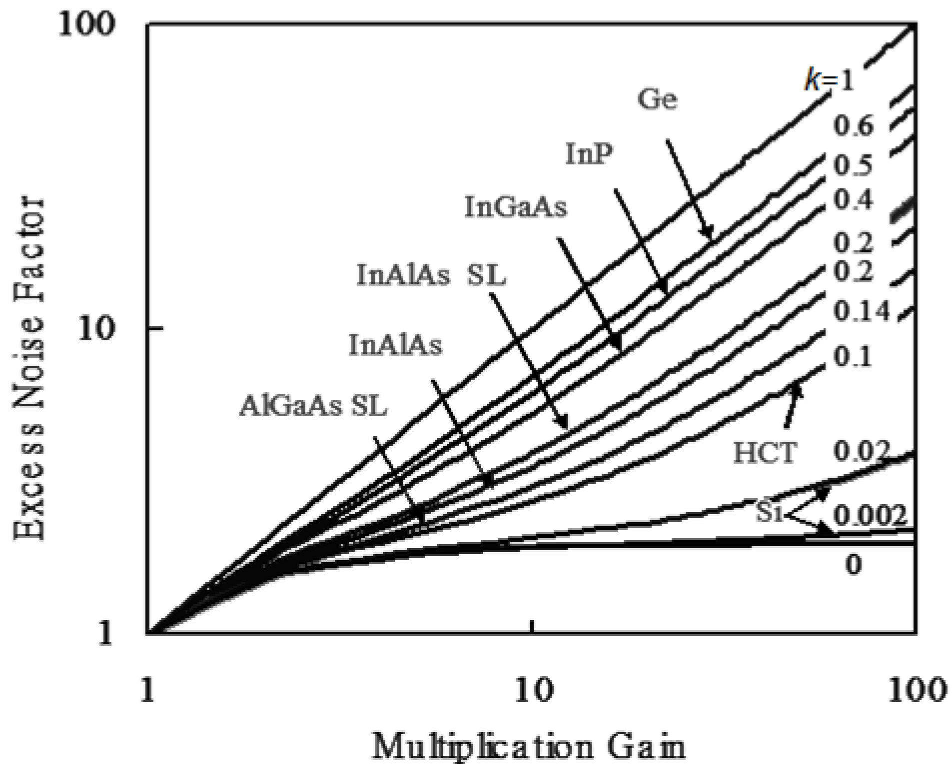


Figure 3.10: Excess noise factor versus multiplication gain for various semiconductor materials and compounds. Notice also, that an estimated ratio value is given for each material. There are two values given for silicon illustrating the dependency on multiplication lengths to the ionization coefficients. Silicon has the most favorable excess noise factor of known semiconductors.

impact ionization in various semiconductor materials. Silicon is lower than other materials used in similar devices. Figure 3.9 showed us that desirable *K* yields a more favorable bandwidth. We saw in Figure 3.7a that a material with $\beta/\alpha \approx 1$ will reduce the response time of that material when used as an amplifier [7]. As you are beginning to see, there are various reasons why silicon is the most desirable material for impact ionization based devices.

In silicon, electrons are about 20 times more probably to participate in impact ionization events than holes [5]. This puts silicon in an area of ideality that other semiconductor materials can't achieve. The reasons behind why silicon is so favorable when it comes to *K* are still

theorized [13]. Most attempts attribute it to the indirect band diagrams of silicon which require momentum shifts to excite carriers from one band to the other. These disparities could be more easily overcome by electrons than holes. Nonetheless, the desirable ratio in silicon is very well documented.

In summary, Silicon exhibits the lowest excess noise factor of any semiconductor or superlattice during impact ionization. It has the most favorable K of known semiconductor materials used in impact ionization based devices. Also, silicon fabrication technologies are by far the most researched and developed in the world. This allows for cheap fabrication and well known techniques to be readily available to those utilizing silicon for devices. Because the SIM is a stand-alone device with gains based upon the theory of impact ionization it can take full advantage of silicon's favorable multiplication characteristics. Because it is stand alone it is very diverse and can be used with photodiodes that are optimized for their absorption characteristics.

3.9 Conclusions

Impact ionization is utilized in many systems to produce signal gain. The ionization coefficients are material and field dependent values that determine how readily carriers participate in ionization events. Analysis of these coefficients and material properties yields gain equations that describe the multiplication effects due to impact ionization. One very important thing to note is that the ionization coefficients for holes and electrons are different for a given material in a given electric field. The ratio of the coefficients for a given material determines the excess noise produced and the upper limit frequency for a given material when it is utilized as an amplification device. Silicon has the most desirable ratio of ionization coefficients and for that reason it is used as the material of choice in the design and development of the SIM.

4 SIMULATION, MODELING AND DATA ANALYSIS

4.1 Introduction

This chapter is intended to give an overview of methods in semiconductor device modeling and simulation techniques. These techniques are becoming ever more important as devices are reaching the limits of semiconductor processing capabilities and crossing the line between semi-classical methods to quantum mechanical analysis of device physics. This chapter also overviews the methods and platforms used to model the SIM.

The need for more accurate and more complete simulation techniques was first brought to light through transistor development. As the development of the bipolar junction transistor (BJT) and later MOS technologies became smaller the need for process and device modeling became more important. Initially, the advancement of Technology Computer Aided Design (TCAD) software was driven by the need for process development simulation in research environments [75]. IC developments lead to concerns of device isolation along with parasitic effects which meant ever increasing process complexity. The original problem tackled by TCAD process simulation was to accurately predict active dopant distribution during processing. With real estate shrinking, dopant diffusion analysis techniques were needed to accurately model junction depths, gradients and doping profiles. More modern process simulations tackle issues dealing with elaborate heterostructures, exotic materials and chemical processes all on ever decreasing scales. Process simulators can accurately model doping, depositions, epitaxial

growths, heterostructures and many other fabrication processes as well as environmental effects during the fabrication process.

To illustrate the advantages of using a TCAD platform as opposed to trial and error in a fab, the following brief example is given. Figure 4.1 shows a flowchart of a device need or structural change going from idea to implementation utilizing TCAD. Let us assume that the device need is to analyze the effects of a varied oxide layer thickness at the gate of a MOSFET. A process simulation can be run, or the oxide thickness can be specified, in the structure definitions. Once this has occurred the device can be simulated and the data analyzed for the desired effects. If the results aren't satisfactory then the process can be repeated infinite number of times altering small effects until the desired result is achieved. Once the result is achieved the model can be fabricated, implemented and tested. Without the luxury of TCAD methods every iteration and alteration would have to be fabricated and tested individually. This can be very costly, time consuming and difficult to optimize.

The history of TCAD software development started at Stanford University with the development of the Stanford University Process Modeling (SUPREM) [61]. Newer versions of the original SUPREM have been implemented in the most widely used simulation platforms of

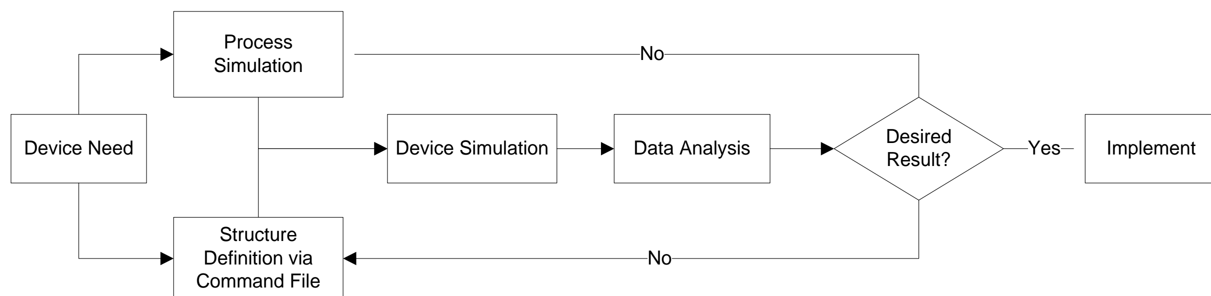


Figure 4.1: Flowchart showing the steps to implement a device utilizing the process and device simulations to develop a desired result.

today. Technology Modeling Associates, Inc. (TMA) was the first company to commercialize a version SUPREM (TMA was later acquired by Avanti and then Synopsys). Later, Silvaco produced a product named ATHENA based on SUPREM. More recently Integrated Systems Engineering (ISE) came out with their own process simulator named DIOS (ISE was later acquired by Synopsys as well). Synopsys and Silvaco are now, by far, the two leading providers of process simulation software today. Both companies also provide a line of device simulators that work in harmony with their process simulation software. These include Dessis (acquired from ISE) and Taurus, both from Synopsis and ATLAS from Silvaco. Outside of the Synopsis and Silvaco platforms, some of the more widely used simulators in research and industry include: PROMIS, PREDICT, PROSIM, ICECREM, DADOS, TITAN, MicroTec, DOPDEES, and ALAMODE. As explained earlier, these technologies offer the flexibility in industry and research to examine hypothetical scenarios rapidly at no material cost. They also provide diagnostics, analysis, and understanding while decreasing design cycle times and the time it takes a new product to reach the market. The semiconductor industry relies heavily on these TCAD platforms and owes much to their development.

4.2 Framework for Device Simulation

The methods in which a simulation platform arrives at a solution file are many. Through the years there have been numerous models and methods used to outline semiconductor device physics. Some models work better describing certain phenomenon than others. Furthermore, some models and numerical methods can be combined with each other while others might not be compatible. Because of this, there are multiple ways that a problem can be solved using simulation software.

With that in mind, we delve a little further in to some of the processes and techniques used in device simulations. One of the first things that must be understood when discussing simulation methods is the motivation behind it. The purpose is to sufficiently model the physics while minimizing the time and computational effort necessary to produce a result. This means that the simulation should produce sufficiently resolved, detailed and specific results to a problem or device structure but without excessive accuracy. Simulations are only useful if the information obtained is accurate and obtainable in a sufficiently small time frame. Figure 4.2 [59] shows the tradeoff between the ease and speed of using a particular method and the exactness and accuracy of that method. Knowing the limitations of the different methods is very important so that you know how accurate your model solution is. As stated earlier, there are some models that are excessively computationally taxing for the solution to be obtained. You would never attempt to solve Green's function [76] in the analysis of a transistor.

For the simulations involving the SIM and other typical device simulations, there are two main factors that must be solved. First, are the carrier transport equations which describe and govern how charge flows in the device. The second is the electric fields in the device. These two factors are very much married together. The dynamics of charge flow are determined by electric fields and the electric fields dynamically change as mobile carries flow. Since these two factors are dependent on one another they must be solved simultaneously in some fashion. The solving of these two factors can be extremely taxing, computationally, if approximations or assumptions aren't made. There are many ways to arrive at a solution to these factors. Some ways are very exact and others are loose approximations. There is a hand in hand trade-off between the *exactness* and computational *ease* of these techniques.

Model/Method	Use/Implementation
Compact discrete models (spice-like)	Compact discrete models (spice-like)
Drift-Diffusion Approximation	Device simulations down to ~0.5-0.1μm for varying levels of accuracy
Hydrodynamic Methods	
Boltzmann Transport Equation (Monte Carlo methods)	Device simulations down to the limits of classical theory
Quantum Corrected Hydrodynamic Methods	Analyze classical devices with correction for quantum aspects
Quantum Equations (ie: Wigner-Bolzmann) and other such Methods	Accurate up until single particle analysis
Green's Function or Direct Solution of the n-body Schrodinger Equation	Single Particle Analysis for Exact Solutions (not possible for large systems)

Ease and Speed

Exactness and Accuracy

Figure 4.2: Illustrates the different types of models and methods used in device simulation. It shows the tradeoff between ease and speed of use and the accuracy of the solution.

Years of research and development have gone into mathematics and device physics to derive the fundamental equations that are used to evaluate potential and charge on the various levels needed for semiconductor device simulation. These equations all find their origin in Maxwell's laws. Maxwell's laws [4] are shown in their differential form here:

$$\nabla \cdot E = \frac{\rho}{\varepsilon_0}, \qquad \textbf{4.1}$$

$$\nabla \cdot B = 0, \qquad \textbf{4.2}$$

$$\nabla \times E = -\frac{\partial B}{\partial t}, \quad \textbf{4.3}$$

and

$$\nabla \times B = \mu_o J + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t}. \quad \textbf{4.4}$$

From these laws, the set of fundamental equations that link electric field to carrier densities in simulations were formed. The main equations used in typical device simulation, as derived from these laws, are Poisson's Equation, the continuity equations and the carrier transport equations. These equations and their purposes in device physics and analysis will now be discussed briefly.

Poisson's Equation [77] in one of its forms is shown here:

$$\nabla^2 V = -\nabla \cdot E = -\frac{\rho}{\varepsilon_s} = \frac{q(n - p + N_A - N_D)}{\varepsilon_s}. \quad \textbf{4.5}$$

It relates variations in electrostatic potential to local charge densities. It is typically derived from Gauss' law (Equation 4.1). Both (Gauss' Law and the Poisson equation) are basically different forms of the same equation which relates the distribution of electric charge to the electric field. Poisson's equation becomes necessary when evaluating how charge density distributions affect electric fields within a depletion layer and vice versa.

The continuity equations are time dependent equations dealing with net charges. They describe the net charge of a carrier concentration by relating the generation and recombination rates and the current densities over time. One form of the continuity equations for electrons and holes are

$$\frac{\partial n}{\partial t} = G_n - U_n + \frac{1}{q}(\nabla \cdot J_n), \quad \textbf{4.6}$$

and

$$\frac{\partial p}{\partial t} = G_p - U_p - \frac{1}{q}(\nabla \cdot J_p)\,, \qquad \textbf{4.7}$$

respectively [77]. In these equations G_n and G_p are the carrier generation rates for electrons and holes respectively. While, U_n and U_p are the recombination rates. These rates are affected by things such as the type of semiconductor material, impurity concentration, defects, temperature, impact ionization and other factors such as optical exposure.

The transport equations are the equations that describe the mobility, drift and diffusion of carriers in the presence of an electric field in space as shown here:

$$J_n = q\mu_n n\mathscr{E} + qD_n \nabla n = \mu_n n \nabla E_{Fn}\,, \qquad \textbf{4.8}$$

and

$$J_p = q\mu_p n\mathscr{E} - qD_p \nabla p = \mu_p p \nabla E_{Fp}\,, \qquad \textbf{4.9}$$

for electrons and holes respectively [77]. In Equations 4.8 and 4.9 the carrier mobility for electrons and holes is represented by μ_n and μ_p, the diffusion length for the material and conditions for the same carriers are D_n and D_p and the electron and hole quasi Fermi levels are E_{Fn} and E_{Fp}.

Equations 4.5-4.9 set the foundation for semiconductor device operation. They are the framework for the device simulation platforms descried in the previous section. They define how factors such as temperature dependence, optical generation, impact ionization, static and dynamic potentials, doping levels and much more affect the device operation.

4.3 Development and Choice of Simulation Models

When TCAD models were first introduced the device characteristics were estimated using broad and simple analytical models and approximations due to their computationally demanding

nature. These assumptions and approximations were capable of sufficiently describing the devices of the time. As devices became more complicated and exponentially smaller in size these assumptions and approximations were no longer sufficient. The physics that governed the device characteristics left the classical realm and entered into the semi-classical. In some cases the quantum theories became necessary to accurately predict some characteristics. As the devices evolved into what they are today so did the numerical methods used in simulation practices.

One of the most useful equations when analyzing and modeling semiconductor devices and carrier transport is Boltzmann's Transport Equation [78], [79]. This equation is probabilistic and semi-classical in nature. It is a function that attempts to reveal the location, x, and momentum, p, of a particle in time in a semiconductor device. Its solutions are typically based, in part, on classical device physics. It is often referred to as being semi-classical because it delves into the quantum theory of Heisenberg's uncertainty principle which states that the exact position and momentum of a particle cannot both be known in a given moment. For this reason Boltzmann's Transport Equation is heavily based upon probability theory to estimate position and momentum in time given complex scattering functions. Boltzmann's Transport Equation [77] can be expressed in the simple general form shown here:

$$f(x,p,t) = \frac{\partial f}{\partial t} + v \cdot \nabla_x f + E \cdot \nabla_p f \ . \qquad \textbf{4.10}$$

The details of the function aren't necessarily important for the scope of this discussion. It should be noted that it is almost an impossibly difficult function to solve when applied to particles in semiconductor devices. This is because location and momentum exist in three dimensions each (assuming no simplification or assumptions) and time in one dimension. This gives way to seven independent variables for a computer to analyze, which in turn yields about

10^{13} unknowns. As the number of carriers being tracked through the device increases, and it becomes necessary to track theses carriers through a dynamic electric field, it becomes apparent that assumptions, to some degree, become necessary for most practical devices. Otherwise, the computational power necessary to solve systems with more than a few electrons quickly becomes overwhelming.

Various approximations and simplifications have been made in order to efficiently obtain solutions while still maintaining a model that sufficiently captures the features and intent of a device. The most frequent model used, based off of these assumptions and simplifications, is that of the drift-diffusion model [80].

The drift-diffusion model is used when analyzing devices that are on the order of a half micron or bigger in feature sizes. For many practical devices this is sufficient. When modeling the SIM, versions of the drift-diffusion model are used almost exclusively because the assumptions made to simplify Boltzmann's Transport Equation into the drift-diffusion model are applicable. These assumptions and approximations tend to lean toward classical analysis and away from quantum mechanics. In typical carrier transport this is often sufficient. Once the device/feature size gets below this point then quantum-like effects begin to play a significant role in the way the device operates. When this occurs, less simplified versions of the drift-diffusion model can often be implemented. Later, for feature sizes smaller than a half micron, many device simulation platforms often use a version of the hydrodynamic model which takes things like hot carriers and other quantum effects into more careful consideration. If the devices get even smaller and approaches feature sizes below 100nm and into the ones of nanometers then it becomes necessary to utilize particle based simulations and model like the Green's function or the Schrödinger equation [81]. The issue with more semi-classical and quantum-like models is

that they are very difficult to implement for practical systems due to their complexity. The Schrödinger equation, for example, becomes overly complex for models involving multiple electrons.

We can see from the previous paragraph that the recent technological advancements and the modern day state of the art features that device simulation capabilities are somewhat lacking. The ability to analyze new materials at ever shrinking scales is important but difficult. It is apparent that transistors in modern day technologies have active regions in the range of tens of nanometers. The drift-diffusion model is obviously not adequate in accurately predicting device operation. Simulations have moved to hydrodynamic models to handle these situations but even they are now insufficient to model the cutting edge [82]. Nevertheless, when modeling and predicting performance for larger-scale devices such as APDs, photodiodes, and the SIM, the drift-diffusion model is typically sufficient. Additional considerations need to be taken into account for impact ionization and other effects but in this device size range the assumptions used are sufficiently accurate. With the introduction of hot carriers, tunneling, impact ionization or other quantum or quantum-like effects other equations must be added to the drift-diffusion model to make it complete. Because the drift-diffusion model is the used in the Silvaco platform which, in turn, is used in simulating the SIM, its application and Silvaco will be discussed in greater detail in the next section.

4.4 Device Simulation in Silvaco

Silvaco is the platform chosen to model the SIM. Silvaco comes with a broad range of capabilities. It provides quite a set of physical models including:

- DC, AC small-signal, and full time-dependency.
- Drift-diffusion transport models.

- Energy balance and Hydrodynamic transport models.
- Lattice heating and heat sinks.
- Graded and abrupt heterojunctions.
- Optoelectronic interactions with general ray tracing.
- Amorphous and polycrystalline materials.
- General circuit environments.
- Stimulated emission and radiation
- Fermi-Dirac and Boltzmann statistics.
- Advanced mobility models.
- Heavy doping effects.
- Full acceptor and donor trap dynamics
- Ohmic, Schottky, and insulating contacts.
- SRH, radiative, Auger, and surface recombination.
- Impact ionization (local and non-local).
- Floating gates.
- Band-to-band and Fowler-Nordheim tunneling.
- Hot carrier injection.
- Quantum transport models
- Thermionic emission currents [83].

Silvaco is a professional platform capable of powerful numerical methods and a broad range of models that can be used to accurately describe many semiconductor devices.

In device simulations, many models are used. As discussed earlier, the default model used by Silvaco in device simulation is the drift-diffusion model. Also mentioned in the previous section, the drift-diffusion model is a simplification of the Boltzmann Transport Equation. The simplification is made by making certain assumptions based upon the device setup. These assumptions and simplifications may vary yielding different versions of the drift-diffusion model. Some of the typical assumptions used in the derivation are things like parabolic-like bands, steady-state conditions, one to two dimensional space, classical particle theory, discrete device dimensions, the absence of temperature gradients, low electric fields and so forth. No matter the assumptions and simplifications the drift-diffusion model is always based on Equations 4.5-4.9 (Poisson Equation, Continuity Equations, and Carrier Transport Equations) in one form or another. Other models and corrections can be added to the base model to account

for device specific phenomenon and this is usually the case. When additional models are included in a simulation other equations are added that must be solved simultaneously with Equations 4.5-4.9. Not all of the models are compatible with each other and will cause run-time errors or convergence problems when improperly combined.

The basic process to obtain a solution set is the same for the different methods used. First the device is outlined and a mesh is defined (either by a process simulation or through direct specification). The mesh is a series of nodes and quadrants that are used when solving the coupled set of equations. After the mesh and geometries of the device to be simulated are set, the materials and doping are assigned within that geometry. Then voltages and currents are applied giving rise to electric fields and current densities. The dynamics of the electric fields and current density through the device are calculated through a process of iterated solutions to a discretized version of the continuous devise. The discrete sections are defined in the geometry mesh setup and are very important to the accuracy and efficiency in obtaining a solution. An example will be given in a future section.

The system is solved by using an iterative process that makes estimates to the solution set and refines them in each iteration. Each iteration through the solving method is based upon the last iteration and is started by an initial guess at the solution set. The iterations get closer and closer to a solution of the set of equations determined in the model. All of the equations must have an agreeable solution to the device setup. If the device setup and initial guess were sufficiently well described the simulations will converge on a solution set and the data can be analyzed. There are preset convergence criteria and the iterations will continue until either the criteria are met or it becomes apparent that they will not be met (divergence).

Different methods can be used in the iterative process to analyze the system of equations used. When choosing between the different methods, the choice is usually base up the likelihood of convergence given the setup and method chosen and how quickly an accurate solution set can be obtained. Accuracy, or how well a model compares to real life, is very desirable but the trade-off between accuracy and simulation time needs to be taken into consideration. The device mesh and numerical methods chosen should be efficient and be able to handle a wide variety of structures. High accuracy requires a fine mesh that can resolve all significant features of the solution. High efficiency requires a coarse mesh that minimizes the total number of nodes. This trade-off between accuracy and numerical efficiency is frequently a problem.

In Silvaco there are three types of numerical methods that can be applied to a chosen model. These are the Gummel, Newton, and Block methods [83]. These methods are different approaches to solving the dynamic system of equations set out by the model that describe and predict the device operation.

4.4.1 Gummel Method

In the Gummel method solves the system of equations by solving them sequentially as opposed to simultaneously [84]. This is typically done by solving for Poisson's equation given the initial conditions. This is done for each mesh triangulation and the boundaries are set equal for adjoining triangulations. Using the potential obtained from this solution the continuity equations are solved followed by the current transport equations. This concludes a typical iterative step which now cycles back on itself solving for Poisson's equation using the current densities and quasi-Fermi levels obtained from the solution to the other equations. This process is repeated until the convergence restraints are met. Convergence is checked every iteration by algebraically positioning the terms on the same side of all the equations and substituting the

iterative values in as the variables. Convergence is met when the residuals are smaller than the set tolerance. The rate of obtaining a converging solution in the Gummel method is based upon the coupling between the equations. If there is little coupling the convergence is reached much quicker. Because this is often not the case the Gummel method usually converges relatively slowly.

4.4.2 Newton Method

The Newton method solves the equations simultaneously as opposed to sequentially [85]. Thus, each iteration in this method solves a linearized version of the entire system of equations. Because of the complexity of this problem (solving a system of equations as opposed to a single equation) each iteration takes significantly longer than the Gummel method. On the other hand the solution is often converged upon much more quickly given sufficiently accurate initial guesses to the system. As a result of this, the Newton method becomes much more dependent upon accurate initial guessing techniques than the Gummel.

The Newton method is the preferred method for systems where an accurate initial guess is possible. It is also preferred when the biasing of the system is larger. As the system biasing increases the coupling between the electric fields and the current densities increases. With an increased relationship between current density and electric field the Gummel method becomes quite lengthy in obtaining convergence. For most systems the Newton method is the preferred method and as such is the default numerical method technique for Silvaco simulations of the SIM.

4.4.3 Block Method

The Block method can, to some degree, be thought of as a combination of the Gummel and Newton methods [86]. The Block methods used by Silvaco go through iterations that solve subgroups of equations simultaneously while stepping through these subgroups sequentially. This makes it possible to incorporate more models in the simulation. These models can be approached separately and then combined. Often, this means Newton's method is used to calculate and update the potential and carrier concentrations. Then the Gummel method is used to analyze additional models that are incorporated into the system (i.e.: lattice heating, tunneling etc.) Note: The Newton and Gummel methods are both capable of handling additional models as well. But not all models and methods are compatible.

4.4.4 Foundatioin for Simulation Setup in Silvaco ATLAS

As mentioned in the previous section, the majority of analyses of the physics behind the operation of the SIM have been done using the Silvaco platform. As stated earlier, Silvaco is based off of the SUPREM platform developed at Stanford University. Silvaco is made up of a group of applications that are made to operate together in harmony. Figure 4.3 illustrates the main applications in the Silvaco platform and how they interact [83].

ATHENA is the process simulator. The structure files produced by ATHENA can be directly input into ATLAS for device simulations. If desired, Deckbuild can be used to create a command file independent of a process simulator which can be used by ATLAS for device simulation. In this case, the structure and material definitions are specified in the Deckbuild runtime environment (more will be explained about this shortly). Once a device has been designed using either ATHENA or Deckbuild, Atlas is used for the device simulation and

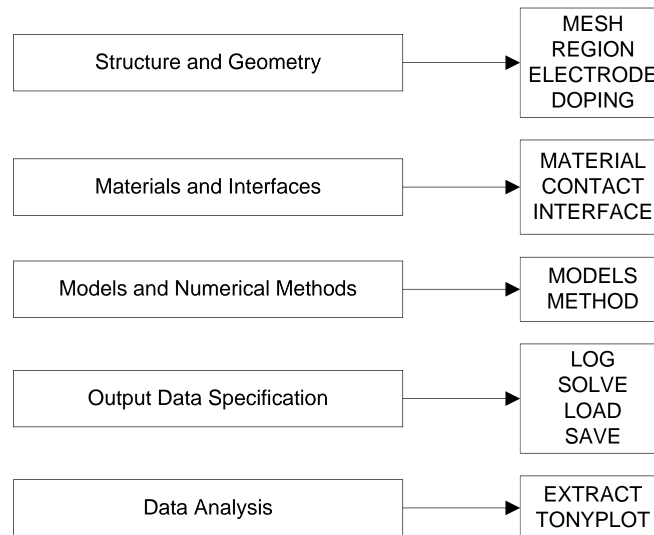


Figure 4.3: The right hand side of the arrows shows the types of commands that are desired in a process simulation setup. The left hand side shows the corresponding commands used in the Silvaco platform to properly describe the device and the modeling.

integration into theoretical systems. During the device simulation, numerous effect are being monitored and recorded as voltages and currents are applied to the device and specified electrodes. ATLAS keeps track of parameters such as mobility, carrier concentration, electric field, recombination, generation, energy levels and much more. Some of this data can be observed real-time via the runtime output. All of the data is output into a series of log and structure files that can be analyzed and observed in the Tonyplot user interface [87]. Once a device has been deemed operable it can be *packaged* and exported to Silvaco's spice (discrete device level) simulator interface and inserted into a circuit system. This interface, known as Mixedmode, allows for DC, AC and transient responses to be done on circuit systems with the *packaged* device integrated seamlessly inside.

Athena, the process simulator, can be used to accurately model fabrication steps [88]. Through proper setup of the simulation diffusion lengths, an oxide growth, depositions, etches and much more can be simulated with extensive control of the theoretical environment. The

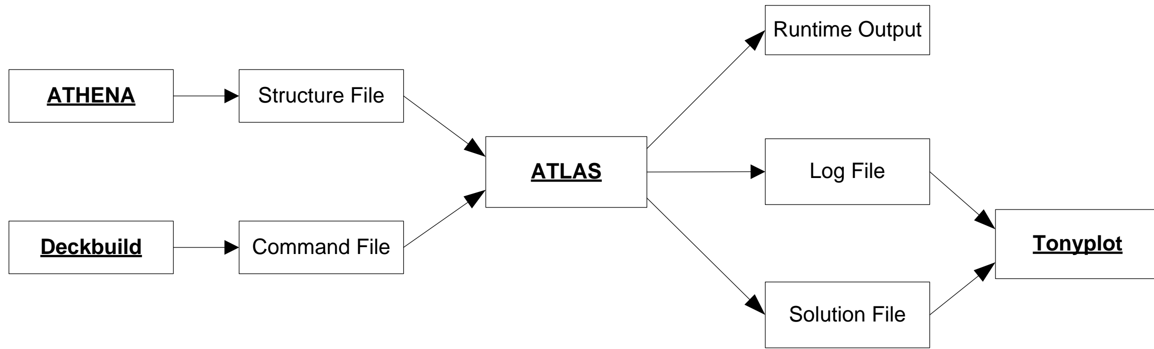


Figure 4.4: Flowchart illustrating the main areas used to analyze data in the Silvaco platform. The devices simulator (ATLAS) can take inputs from either Deckbuild or the process simulator (ATHENA). The output data can be analyzed real-time or through Tonyplot.

setup for a process simulation in Athena is generated in Deckbuild. In Deckbuild commands are outlined to specify the geometries, resolution, and physical factors that a semiconductor material encounters in the fabrication process. Control over semiconductor materials, temperatures, timelines, chemical exposures, impurities, etc., are taken into consideration during these *fabrication* sessions.

As stated earlier, Deckbuild is used to create a command file for Athena process simulations or Atlas (for device simulations without doing a process simulation previously). At this time we will discuss the abilities of Deckbuild in designing semiconductor devices for simulation in Atlas.

In the Deckbuild runtime interface the structure and geometry, materials and interfaces, models and numerical methods, and output data specification are outlined. The ATLAS commands for each of these can be seen in Figure 4.2. These parameters can be used to specify device and material characteristics such as carrier mobility, recombination, generation and other effects. After the device has been setup and the models chosen voltages and currents can be applied to the metal contacts for single device DC, AC and transient testing. During the running

process the current and voltage can be monitored in Deckbuild as well as the convergence of the overall model. The results are output to Tonyplot in the form of a data curves as well as a cross sections of the device for further analysis and manipulation. The results can also be extracted and exported for analysis in other mediums.

We will learn more about the specifics of the SIM in Chapter 4.5. In the meantime, we will discuss some of the nuances of the analysis of the SIM using Silvaco and why it was necessary to arrive at present versions of the SIM.

Table 4.1: Lists the basic ATLAS commands and a brief description.

Command	Description
MESH	Sets the resolution for the discretization of the device during simulation.
REGION	Specifies the different material region in the devices geometry.
ELECTRODE	Specifies locations where voltages and currents can be applied.
DOPING	Sets the doping levels, type, junction depth, etc.
MATERIAL	Used to assign a material type to each REGION
MODELS	Determines which models will be used to evaluate the carrier densities, carrier transport, and electric fields in the system.
CONTACT	Define the type of metal/work function for each specified ELECTRODE.
BEAM	Define and specify a light source for simulations with an optical source.
METHOD	Specify the numerical methods used in evaluating the chosen models.
LOG	Output command for applied conditions. Produces a .log file which can be evaluated using TONYPLOT.
SOLVE	Used to apply/change the voltage or current at a specified ELECTRODE.
LOAD	Import a file from ATHENA.
SAVE	Save a design to be exported to MIXEDMODE or TONYPLOT.
EXTRACT	Specify specific parameters (aside from the default) that need to be evaluated during runtime.
TONYPLOT	Open TONYPLOT and evaluate a saved file.

Being able to analyze variations to currents, voltages, geometries, doping levels, temperatures and other factors was absolutely necessary during the development of the SIM. Analyzing the effects of these changes through traditional fabrication and testing techniques would prove costly and time consuming. Having complete control over the devices structure with predictive results through Silvaco made testing the latest theories and queries possible.

The process to set up a simulation using Deckbuild is rather straight forward for most basic devices. The first thing that needs to be done is to understand the statements that are used to define a device's geometry and the parameters that simulate the fields and current densities through the device. A more in-depth look at some of the ATLAS statements can be seen in Table 4.1 [83].

4.4.5 Silvaco Simulation Example

With this information it would be useful to see an example. The following section will walk through the setup and evaluation of a simulation in Silvaco. (The full code for this simulation can be found in Appendix A). The example will be the simulation of a SAM APD.

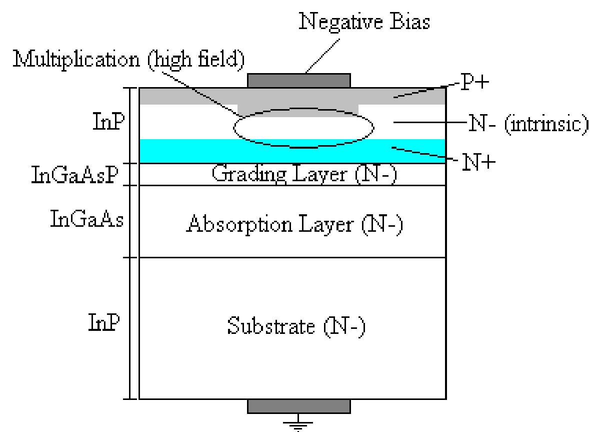


Figure 4.5: An InP/InGaAs SAM APD. This device will be simulated in the following example.

The basic structure can be seen in Figure 4.5. This structure is similar to that discussed in Section 2.2.5 and shown in Figure 2.7. Using an SAM APD as an example was chosen because it a wide range of Silvaco's capabilities, including: heterostructures, optical signals, gradient layers, voltages, etc. Silvaco is capable of much more than will be shown in this example but this will give a foundation for the capabilities that Silvaco has when modeling devices like the SIM.

The first thing that needs to be done when setting up a simulation is to specify a MESH. The mesh can either be setup in cylindrical coordinates or in rectangular coordinates. The following is an example of a mesh setup in rectangular coordinates. Once the type of mesh is specified then the locations and spacing are setup. For a rectangular mesh this is done by specifying the location of the changes in the mesh spacing starting at the smallest x coordinate in the design (this is usually at x=0 but it can be a negative value if desired). After specifying the mesh in the x direction, the y direction is specified in the same way (starting from smallest and going to largest coordinate). The x.mesh and y.mesh commands are used for this process. The mesh used in this example is shown here:

```
x.mesh loc=0.0      spac=0.5
x.mesh loc=15.5     spac=0.5
x.mesh loc=16.0     spac=0.5
x.mesh loc=20.5     spac=0.5
x.mesh loc=19.0     spac=0.5
x.mesh loc=25       spac=0.5

y.mesh loc=-.5      spac=0.2
y.mesh loc=-.25     spac=0.2
y.mesh loc=0        spac=0.1
y.mesh loc=0.75     spac=0.05
y.mesh loc=1.0      spac=0.05
y.mesh loc=1.15     spac=0.05
y.mesh loc=2.20     spac=0.1
y.mesh loc=2.25     spac=0.1
y.mesh loc=2.3      spac=0.1
y.mesh loc=3.25     spac=0.1
y.mesh loc=5.75     spac=0.25
y.mesh loc=5.8      spac=0.25
```

Notice that for each command (x.mesh or y.mesh) there is a corresponding location (loc) and spacing (spac) command. These set boundaries for different special resolutions in the

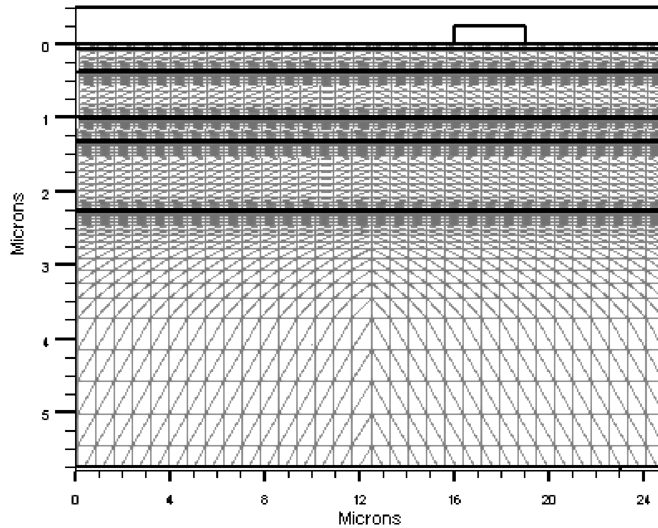


Figure 4.6: Mesh for the InP/InGaAs SAM APD. The mesh was produced in SILVACO. Proper mesh setup is necessary for accurate solutions. Notice that the size of the mesh decreases at the material and doping junctions indicated by the black solid lines. This is necessary to obtain accurate results.

simulation. For every change in region, material, doping, electrode etc., in the x direction, there should be a restructure of the mesh with a new location command (this is true even if the spacing remains the same during the change). The resulting mesh can be seen in Figure 4.6. The location and spacing units are in microns. The setup of the MESH is probably the most important part of the simulation when it comes to runtime efficiency, resolution of results, and convergence issues. If the mesh is too narrow then the simulation could take days to converge. If it is too broad then the results will be either faulty due to the loss of features or insufficiently resolved. It takes quite a bit of time to get used to setting up meshes where the simulations run efficiently and the results maintain their integrity.

Once the mesh is setup, every portion of the mesh must be assigned to a REGION. The regions are simply geometric assignments for different materials. The MATERIAL command is used in conjunction with the region command and can be seen by this section of code:

```
region num=1 material=InP        x.min=0 x.max=25 y.min=2.25 y.max=5.75
region num=2 material=InGaAs  x.min=0 x.max=25 y.max=2.25 y.min=1.15 x.comp=.47
region num=3 material=InGaAsP x.min=0 x.max=25 y.max=1.15 x.comp=.47 y.comp=1 grad.34=.15
region num=4 material=InP        x.min=0 x.max=25 y.min=0    y.max=1
region num=5 material=Air        x.min=0 x.max=25 y.min=-.5  y.max=0
region num=6 material=Air        x.min=0 x.max=25 y.min=5.75 y.max=5.8
```

It should also be noted that every section of the mesh must have a corresponding region and material assigned to it. There are a wide variety of materials that can be chosen as seen in the code above. The *x.min*, *x.max*, *y.min* and *y.max* commands specify the location of the regions/materials and the units stay consistent (microns) with those used when defining the mesh. Notice in the above code that region number 2 and 3 have been assigned composition numbers (*x.comp* and *y.comp*) based upon the following equation

$$E_g = 1.35 + x(0.642 + 0.758x) + y(0.101y - 1.101) - xy(0.28x - 0.109y + 0.159)\,. \qquad \textbf{4.11}$$

Where 1.35eV is the energy gap of pure InP and the x and y composition factors alter the material in the following manner $In_{1-x}Ga_xAs_yP_{1-y}$. If the x and y composition both are zero then the material is simply InP. By grading (*grad.34=.15*) the x composition from zero to ~1/2 and the y composition from zero to one a seamless transition is made from InP to InGaAs in the InGaAsP. This grading will smooth discontinuities in the heterojunctions between the materials and eliminate anomalous barriers. The region setup per Silvaco can be seen in Figure 4.7

Once the regions and materials are specified ELECTRODEs are assigned to different locations. The electrodes geometries must be in contact with a suitable material. The electrodes are used to specify where voltages and currents can be applied to the device. It is effectively like adding a contact of some sort to the semiconductor material. The code is shown here:

```
elec num=1   name=Anode    x.min=16       x.max=19.0             y.min=-0.25             y.max=0.0
elec num=2   name=Cathode  x.min=0.0      x.max=23               y.min=5.75              y.max=5.8
```

Each electrode must be named and numbered for future application of voltages and other material/contact information.

DOPING is the next thing to be specified. There are many different ways to specify the doping in Silvaco. The junction can be specified to be very abrupt or a gradient. The doping concentration in the well can be uniform, Gaussian or graded in nature. This is an example of how the doping was specified for the regions of the SAM APD:

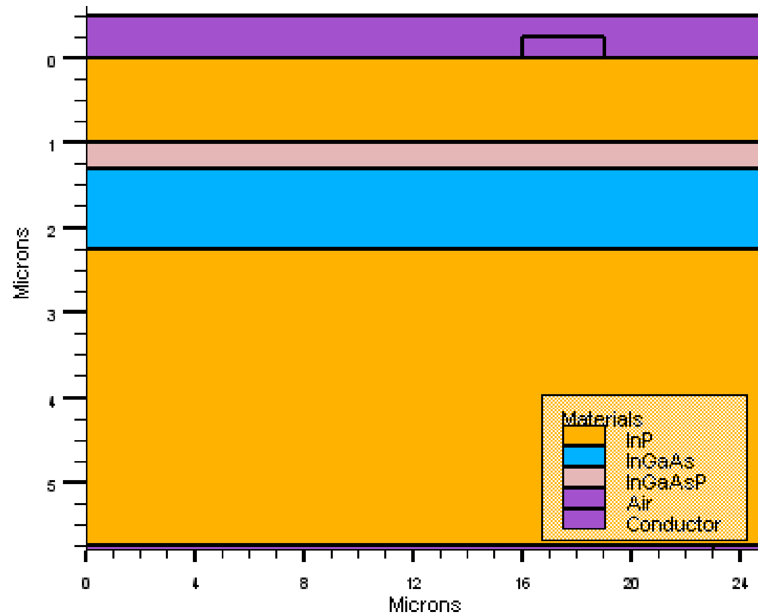


Figure 4.7: Regions and materials have been assigned in this Silvaco simulation.

```
#InGaAs
doping    uniform reg=2          n.type conc=5e15

#InP (Avalanche Region (n- Well) doping)
doping    uniform reg=4          n.type conc=3e15
doping    uniform reg=4          n.type conc=1e17  x.left=0 x.right=23 y.min=0.75 y.max=1

# (P+ layer) doping
doping    uniform reg=4          p.type conc=1e18  x.left=0 x.right=20 y.min=0.0  y.max=0.25

# (P- Well) doping
doping    uniform reg=4          p.type conc=5e15 x.left=15.5 x.right=20.5 y.min=0.0 y.max=0.5

#Substrate (N-) doping
doping    uniform reg=1          n.type conc=1e16
doping    uniform reg=1          n.type conc=8e16 x.left=0 x.right=23 y.min=3.25 y.max=5.75

#InGaAsP
doping    uniform reg=3          n.type conc=5e15
```

Note that the pound sign (#) is for comments within the code. When an entire region is to be doped the same, there is no need for the *x* and *y* specifications. In this case all that is needed is

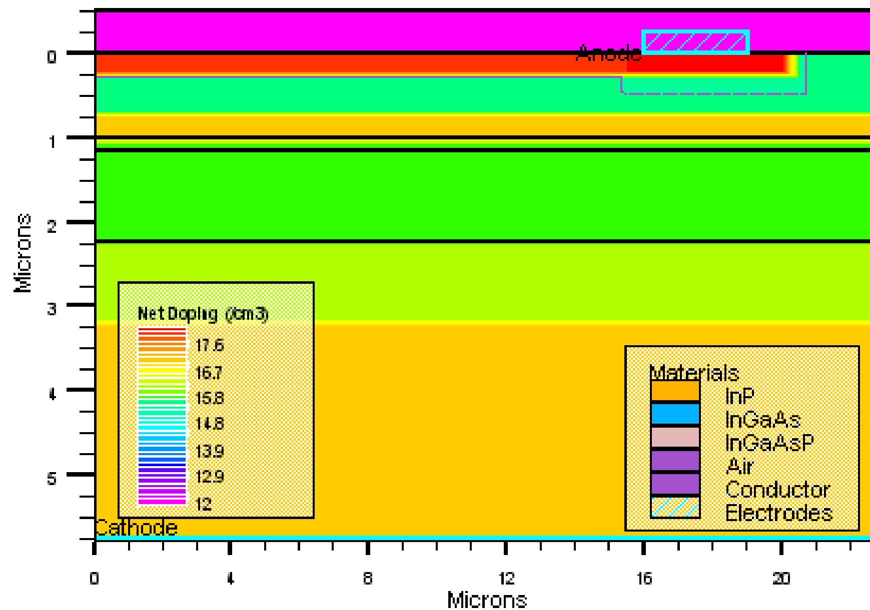


Figure 4.8: Cross section of the SAM APD as produced by Silvaco. The doping profile is shown and indicated. Note that only the concentration can be seen in the figure and not the doping type. The code must be available to see the doping type.

the doping type and concentration. This is seen for regions 1, 2 and 3. Figure 4.8 shows the doping profile for the APD produced by Silvaco. After the doping all that is left for the structure setup is to specify the electrode material. In this case aluminum was used as shown in the code:

```
contact     name = Anode        Aluminum
contact     name = Cathode      Aluminum
```

Because this is an APD there needs to be a light source for simulation. In this case the light source was specified using the BEAM command as follows:

```
beam    num=1 x.origin=5 y.origin=-1.0 angle=90.0 wavelength=1.55 max.window=5
```

The origin of the light source is specified as well as the angle and wavelength (in microns). To make sure you have the location of the beam in the right spot and the wavelength correctly specified, Silvaco allows you to see the photogeneration rate in the semiconductor. This is shown in Figure 4.9.

Table 4.2: Outlines a few of the models used in Silvaco simulations.

Model	Syntax	Notes
Shockley-Read-Hall	SRH	Uses fixed minority carrier lifetimes. Should be used in most simulations.
Concentration Dependent	CONSRH	Uses concentration dependent lifetimes. Recommended for Si.
Auger	AUGER	Direct transition of three carriers. Important at high current densities.
Selberherr's Model	IMPACT SELB	Recommended for most cases. Especially those involving impact ionization
Klaassan Band to Band	BBT.KL	Included direct and indirect transitions

Now that the structure and all of the peripherals are setup the models and methods must be specified. The models used in Silvaco are based upon variations of the drift-diffusion model outlined in the previous section. When specifying different models in Silvaco it is essentially telling Silvaco to focus on certain parts of the model or add additional equations to the base model equations to enhance the solution set. In this case the following models are specified in the code:

```
models srh consrh auger impact selb bbt.kl
method Newton
```

Table 4.2 shows the models and some brief notes about the model. In the above code the Newton method was specified. More can be learned about this by reading Section 4.4.2.

From this point the only thing left to do is apply voltage to the specified electrodes and specify the intensity of the beam. This is done by using the following code:

```
solve   b1=1e-5
solve   b1=0.001
solve   b1=0.01
solve   b1=0.1
solve   b1=0.5
solve   b1=1.0

solve   vAnode=0
solve   vAnode=-.1
solve   vAnode=-1    vstep=-1   name=Anode  vfinal=-20
```

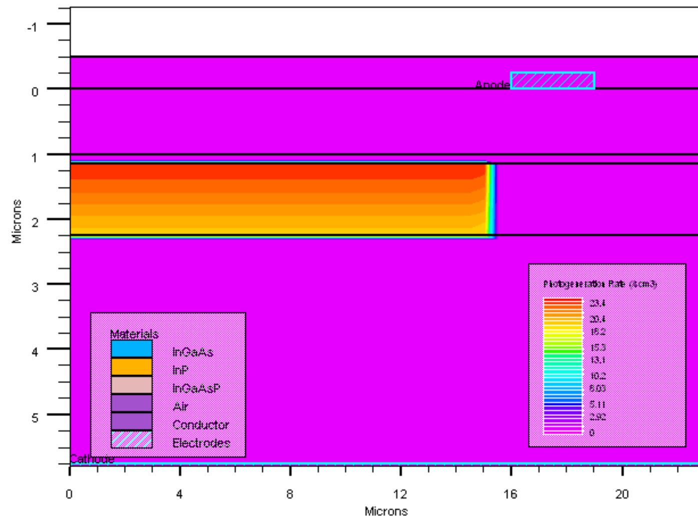


Figure 4.9: Silvaco output file showing the photogeneration rate within the device. Notice that because the wavelength of light was specified correctly for the materials the optical signal is not absorbed by the InP which is the multiplication material for this APD.

The units of the beam power are in Watts/cm and the units of the voltage are in Volts. With this code the output data can be observed in Tonyplot. Figure 4.6 through Figure 4.9 give

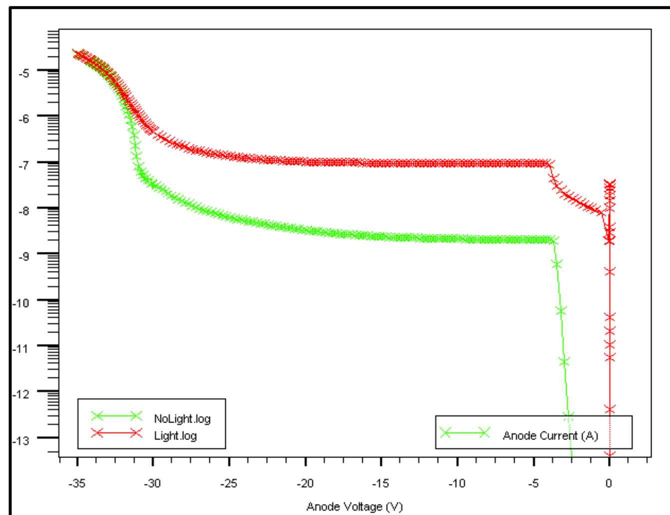


Figure 4.10: Current versus voltage curve for the SAM APD. Green curve is for no light incident on the APD. The red is for 1 W/cm^2. Silvaco is capable of showing an array of different outputs in this manner.

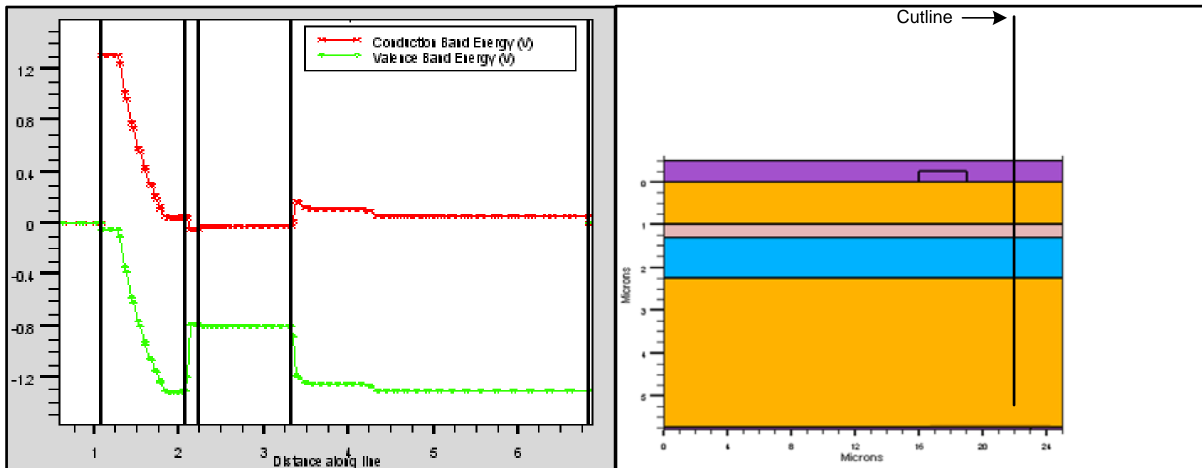


Figure 4.11: Silvaco Structure file showing a cutline and the band diagrams along the associated line. The structure file is on the right illustrating the location of the cutline and the corresponding band diagrams are on the left.

examples of a few of the cross-sectional structure diagrams viewable in Tonyplot. You can also view output current versus voltage curves in Figure 4.10 or the band diagrams from Silvaco for the device in Figure 4.11.

As this example has shown, Silvaco has many capabilities and is very useful when trying to analyze a device and its operation before fabrication. Many of the parameters we specified above could easily be tweaked and the results viewed for comparison. This allows a lot of flexibility when analyzing the '*what if*' scenarios often encountered in device physics.

4.5 Conclusions

Simulation and modeling techniques are vital in modern day research of device physics. This is no different in the research and development of the SIM. Extensive simulations of the SIM have been done using the Silvaco platform. Silvaco code for simulations of the SIM can be

found in Appendix 0. These simulations have helped verify designs and explore various scenarios otherwise not achievable given financial constraints.

5 CONCEPT OF THE SIM

5.1 Introduction

The Solid-state Impact-ionization Multiplier (SIM) has been mentioned briefly in the previous chapters of this dissertation. As was stated earlier, the SIM was designed in an effort to fill a need in the semiconductor device world. The amplification process within the SIM is identical to that of an avalanche photodiode in that it is based on impact ionization events. Unlike an APD, however, the signal enters the amplification region of the SIM electrically from an arbitrary current source instead of an optical source [23]. Because of this, and the desire to utilize the SIM in optical systems, it was designed to be used in tandem with a photodiode as the current source. As was discussed in Chapter 1, one of the main difficulties in APD design is the material choices for the optical absorption region and the signal multiplication region. In a standard APD these materials are one and the same. Much research has gone into efforts to separate the material used for each of these operations whether by epitaxial growth techniques or wafer bonding [89], [43]. The SIM was designed to be a stand-alone amplifier that utilizes impact ionization based gain (avalanche breakdown) to amplify a signal from an arbitrary current source enabling complete separation of these two operations. Because of the ability to separate the absorption and amplification of a system, the optimal materials for each can be used. The basic design, operation and aspects of the SIM will be discussed in further detail throughout the remainder of this chapter.

5.2 Advantages of the SIM

The purpose of developing any device is so there is some sort of advantage given by that device in a system. The SIM is no exception to this. There are modern day systems that use impact ionization as the mechanism to generate low-noise current gains. The SIM has potential to add advantages to these systems through its unique attributes. As stated in the introduction to this chapter, in operation, the signal source for an APD comes from electron-hole pairs generated within its depletion region. Once the signal is detected, it is amplified in the multiplication region of the device. Ideally, the optimal semiconductor material for detection would also be the optimal material for amplification but this is rarely the case. It becomes apparent that having the depletion/detection region and the multiplication region in the same device makes it is difficult to utilize the most ideal semiconductors for both operations. Separate Avalanche and Multiplication (SAM) APDs and wafer-fused APDs were developed to address this issue by combining different semiconductors using heterojunctions [89]. But these devices require difficult and often elaborate fabrication techniques making it expensive and sometimes impractical for some systems.

It becomes obvious that this issue was one of the main driving forces for the development of the SIM. The SIM produces impact ionization based current gain receiving its signal via connection to an arbitrary current source (not through a semiconductor junction). Because of this, when the SIM is utilized in optical detection systems it introduces a natural separation of the detection and amplification of the signal in the system. The system detector can be a photodiode tailor made for optimizing the absorption of the input signal's optical wavelength. Likewise, the material used for the amplification can be optimized by using a semiconductor capable of low noise high bandwidth impact ionization based amplification. The SIM's capability to handle an

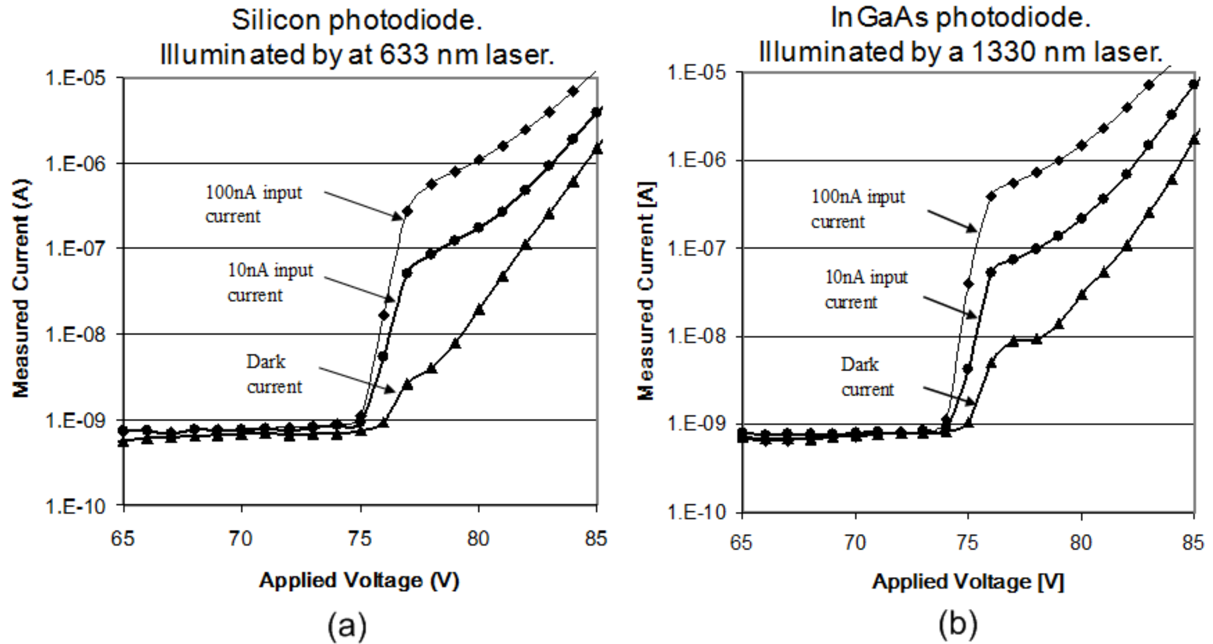


Figure 5.1: Shows the capability of the SIM to be wired to an arbitrary current source. a) The SIM amplifying a signal from an silicon photodiode. b) The SIM amplifying a signal from a InGaAs photodiode

input signal from an arbitrary current source is illustrated in Figure 5.1 [47]. Notice that the current versus voltage curves for the same SIM wired up to two different photodiodes produces nearly identical outputs.

Another advantage of being a stand-alone device is that it can be cascaded for large gains. That is, the output of one SIM can become the input of another, as seen in Figure 5.3. Cascading in this way provides further advantages with regards to signal to noise ratios and makes it possible to achieve very high gains at lower fields and applied voltages. The signal to noise ratio is a direct result of the excess noise factor. As seen in Figure 5.2 [65], when the field is increased in impact ionization based devices, the excess noise factor is also increased. Achieving high gains without extreme fields is very desirable but this is difficult since the gain of a device is directly proportion to the electric field within the device [90]. One device biased to achieve a gain of one hundred might produce a significant amount of noise at those electric

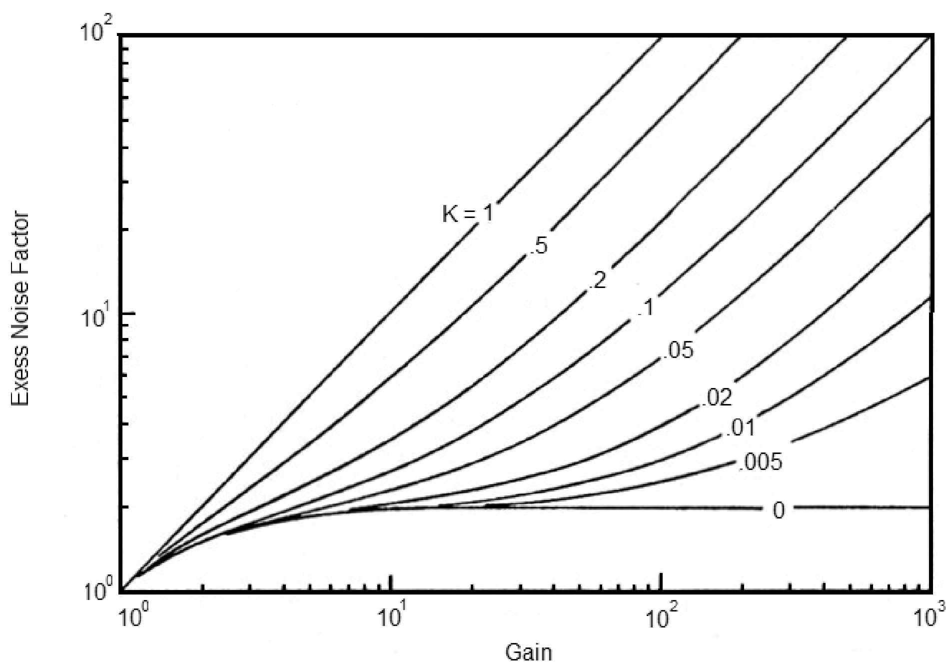


Figure 5.2: Excess noise versus gain for various ratios of ionization coefficients. Notice as the gain goes up (this indicates an increase in electric field strength) the excess noise increases as well. This indicates that for devices limited by excess noise factor, lower gains are better.

field levels. But, if both devices in Figure 5.3 were biased to achieve a gain of ten it produces a system gain of one hundred at smaller electric fields. This greatly reduces the voltages necessary to achieve appreciable gains. There are obvious tradeoffs to the number of devices cascaded together and the speed, gain, and noise output of a system. These tradeoffs won't be discussed in great detail at this time but it is sufficient to say that having this as a feature in an amplification device greatly increases its usefulness and flexibility.

Also, when discussing noise considerations, and a device of the nature of the SIM, it could be asked "What is the advantage of a photodiode/SIM tandem over a photodiode/transimpedance amplifier (TIA) combination?" One of the answers is that the SIM can operate below the noise floor of a typical TIA used to convert the current signal into readable voltage levels, thus, increasing sensitivity [91]. A current source can feed a signal into the SIM

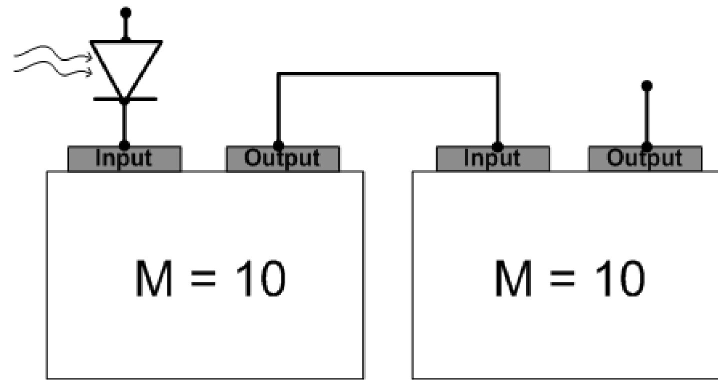


Figure 5.3: Two devices wired together in a cascading fashion. With the output of one device connected to the input of the next device the system gain can be large while keeping the gain per device manageable. In this example the devices are biased up to produce a gain of 10 yielding a system gain of 100.

where it is amplified through a desirable avalanche process and can then, if desired, be fed into a TIA for voltage readout.

Probably the biggest advantage of all is that the SIM is fabricated from silicon. The material of choice for amplification in semiconductor devices is silicon. Silicon is the optimal semiconductor material for impact ionization based amplification for a variety of reasons including: cost, ease of fabrication, desirable ionization coefficients, robustness, among others (as was stated in the chapter on the theory of impact ionization).

5.3 Basic Design

The device cross-section illustrated in Figure 5.4 represents a version of the SIM. Several variations of the SIM have been introduced since the original version. These variations have maintained the same operation principal despite their changes. A brief look at these variations will be outlined later in this chapter and a more in-depth look at some of these versions will be seen in later chapters.

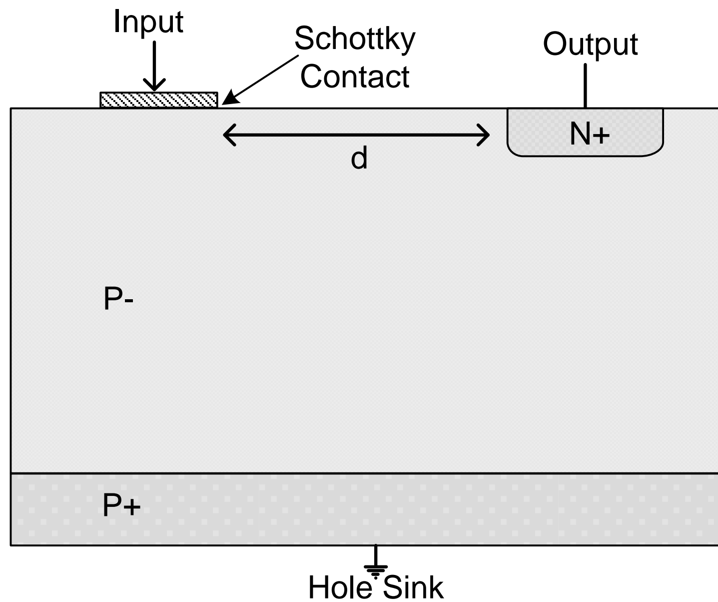


Figure 5.4: Illustrates a cross section of the basic design of the SIM. The SIM is fabricated on a P+ substrate with a P- epitaxial layer. The output node has an N+ well. This version of the SIM has a Schottky contact at the injection node.

Figure 5.4 illustrates a version of the SIM fabricated on a P+ substrate with a P- epitaxial layer. The SIM is a three electrode device (input, output and ground). These contact points are outlined as the current source (input node), bias node (output node) and the hole sink (ground). There is also an N+ well at the output electrode necessary for correct operation and bias which will be outlined in the following section. The distance between the input and the output is labeled *d* in the figure as well. This distance is important when describing the operation and characteristics of the SIM as well as the bias necessary for operation. Figure 5.5 shows the same cross-section from Figure 5.4, with a surface band diagram overlaid above it. The surface band diagram illustrates the SIM before any currents or voltages are applied. The bending of the bands represents the built in potential at the semiconductor junctions.

5.4 Stages of Operation

As expected, when different voltages and currents are applied to the SIM its operation changes. To help explain the operation of the SIM, the different operation will be explained using the structure found in Figure 5.5 and current versus voltage curve shown in Figure 5.6. The different stages outlined in Figure 5.6 will now be discussed in further detail.

5.4.1 Pre-breakdown Stage

The SIM operates by biasing the output with a positive voltage and grounding the substrate. The input node is a floating voltage point where a current source is connected as the input signal. The application of this voltage leads to the reverse bias of the junction between the N+ well at the output and the P- epitaxial layer. As the voltage at the output node is

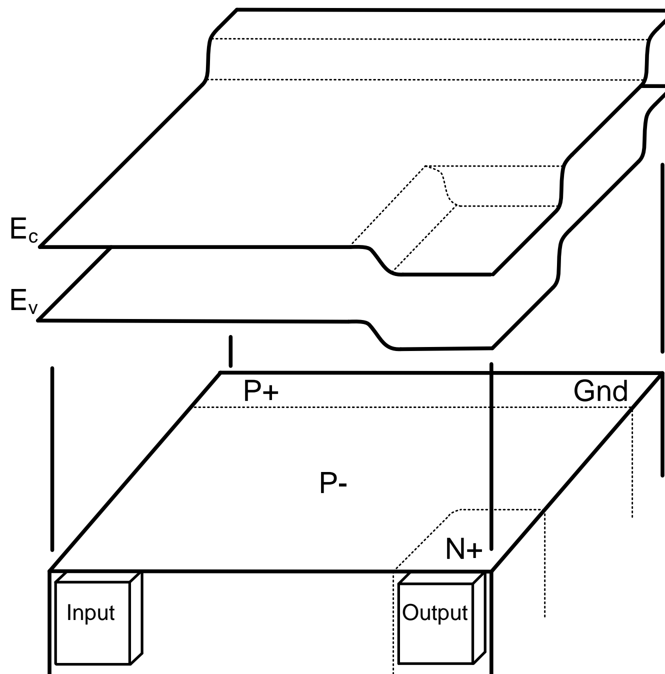


Figure 5.5: Surface band diagram over the top of a SIM cross section. This diagram is for an unbiased SIM. Bending in the bands is due to the built in potential at the N+ to P- junction.

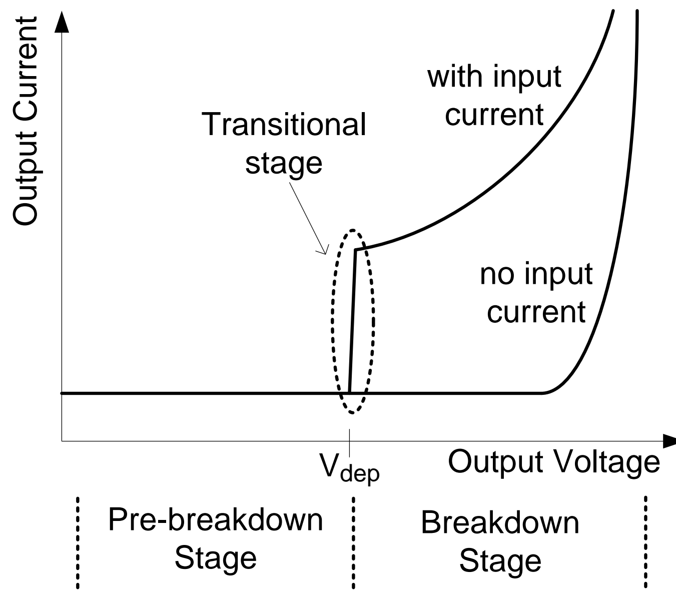


Figure 5.6: Represents a typical current versus voltage curve for the SIM. Labeled are the different stages of operation as the output voltage changes.

increased there is an increase in the bending of the bands and the depletion layer at the PN junction begins to extend itself outward into the epitaxial layer in the vertical and horizontal direction as seen in Figure 5.8.

At lower biases the electric field isn't sufficiently large to produce impact ionization events. By monitoring the current entering and leaving each of the three nodes we begin to get a feel for the way the device is operating during the different stages. During this pre-breakdown stage, carriers injected at the input node simply recombine in the bulk and the output current is zero. Holes enter the device from the ground to replace those recombining in the bulk and we get a net current flow from the injection node to the substrate satisfying Kirchhoff's current law. This is illustrated in the slab diagram in Figure 5.7. It can also be seen in the current versus voltage curves in Figure 5.6. Notice that the output current during the pre-breakdown voltages is

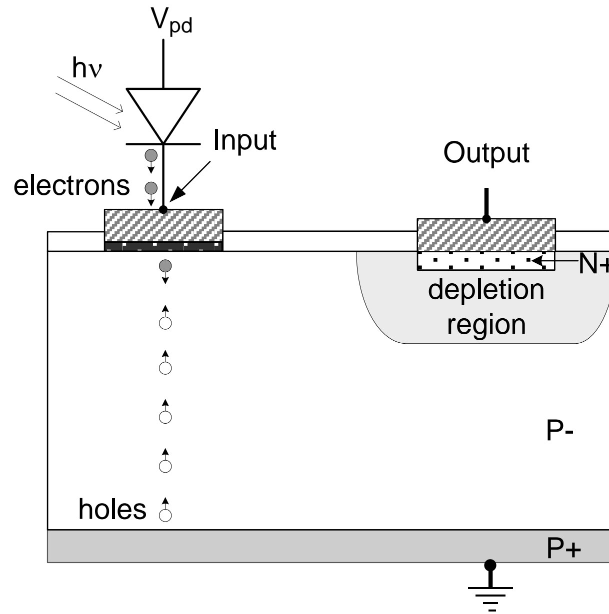


Figure 5.7: Slab diagram showing the pre-breakdown stage of the SIM. At this stage the bias isn't large enough to extend the depletion region to the input node.

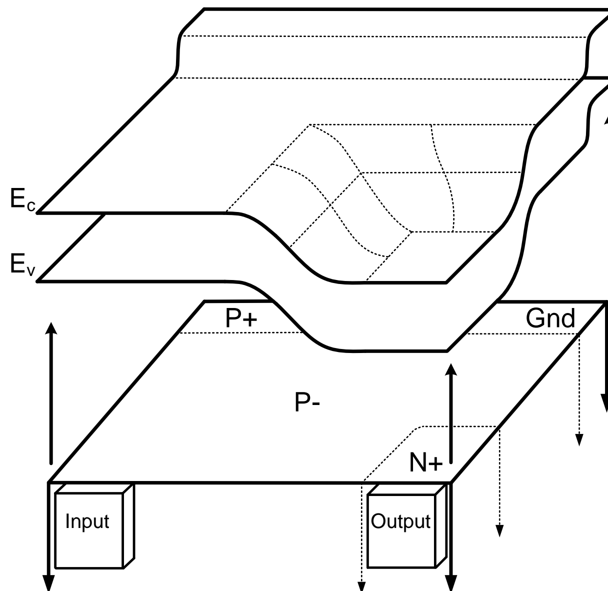


Figure 5.8: Surface band diagram over the top of a SIM cross section. This illustrates the bending of the bands for a biased SIM where the bias isn't yet sufficient to completely depleted the epitaxial layer between the input and output nodes.

low and doesn't represent the input signal. Depending on the doping and geometries (length of *d*), the pre-breakdowns stage may take a very large amount of voltage at the output node to be

overcome. The breakdown voltage for a one-sided junction (which is a very good approximation for the SIM) is shown here:

$$V_{BD} = \frac{\varepsilon_s \mathscr{E}_m}{2qN} \text{ [4].} \quad \textbf{5.1}$$

Here, N represents the doping of the P- epitaxial layer of the one sided junction. The size of the depletion region must extend within a depletion length of the injection node for current to begin flowing from the input to the output. The depletion length into a semiconductor can be calculated given the reverse bias voltage and the doping profile (N+ well and P- epitaxial layer) by the following equation:

$$d = \sqrt{\frac{2\varepsilon_0 \varepsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)(V_{bi} - V_{sim})} \text{ [4].} \quad \textbf{5.2}$$

It should be noted that this is assuming an abrupt-like junction between the N+ well and the P- epitaxial layer and is dependent on the material, doping level and bias.

5.4.2 Transitional Stage

During the transitional stage of the SIM's operation, current begins to flow from the input injection node to the output node. This occurs with the depletion region within the device comes within a diffusion length of the injection point. Figure 5.9 illustrates the surface band structure for a voltage at the output node to create an electric field strong enough to completely deplete the carriers between the input and output. When this occurs there is a noticeable jump in the current flow at the output node as seen in Figure 5.6. There is also a corresponding drop in the current at the hole sink which is not illustrated in the figure. In this voltage range the current transitions from the hole sink to the output node in whole or at least in part as indicated by the slab diagram in Figure 5.10. Keeping Kirchhoff's current law in mind we see that the current at the input must

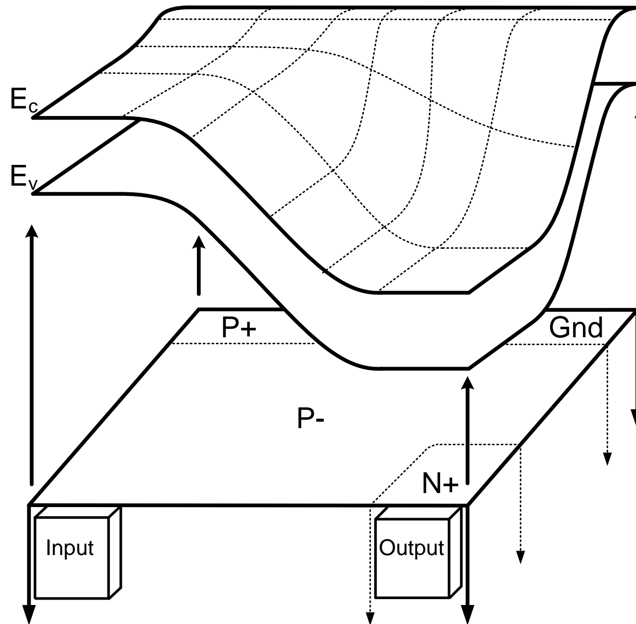


Figure 5.9: Surface band diagram over the top of a SIM cross section. This illustrates the bending of the bands for a biased SIM where the bias has caused complete depletion of the epitaxial layer between the input and output nodes. The bias still isn't sufficiently strong to initiate impact ionization events.

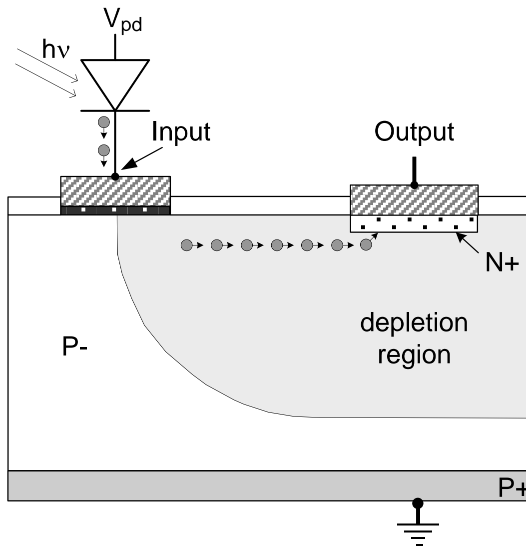


Figure 5.10: Slab diagram showing the transitional stage of the SIM. At this stage the bias is large enough to extend the depletion region to the input node but not large enough for impact ionization to occur.

be equal and opposite of the combined currents at the output and the grounded substrate,

$$I_{\text{input}} + I_{\text{output}} + I_{\text{hole sink}} = 0\,. \qquad \textbf{5.3}$$

After the current flow has transitioned from 'input to substrate' to 'input to output' continued bias is applied until breakdown is reached.

5.4.3 Breakdown Stage

After reaching the transitional stage, additional voltage is often necessary before breakdown begins to occur for the input signal to be amplified. Again, this depends upon the geometries and doping of the device as was shown in Equations 5.1 and 5.2.

The surface band diagram for the breakdown stage is seen in Figure 5.11. Once the maximum electric field in the depletion region is high enough for breakdown to occur there is gain in the output signal as seen in Figure 5.6. This is because the injected electrons travel through the depletion region and enter fields sufficiently large for impact ionization to occur. These carriers gain sufficient energy to ionize other carrier during the generation of electron-hole pairs. During this stage in the operation, we start to notice an increase in the substrate current equal to that of the gained signal at the output, as can be seen in the slab diagram of Figure 5.12. In other words, the substrate (hole sink) current is equal and opposite the magnitude of the output current minus the original injected signal as in

$$\left|I_{\text{hole sink}}\right| = \left|I_{\text{output}} - I_{\text{input}}\right|. \qquad \textbf{5.4}$$

From Equation 5.4 it becomes obvious that the increase in current at the hole sink is due to the electron-hole pairs that are created during impact ionization. The electrons from these generated pairs join the originally injected electrons as part of the output signal and the holes go to the grounded substrate (hole sink). Having a potential drop from the output to the substrate to *bleed* off the generated holes is a necessary feature for proper operation. This feature is illustrated in Figure 5.13. Without the hole sink, the holes would not be directed away from the injection

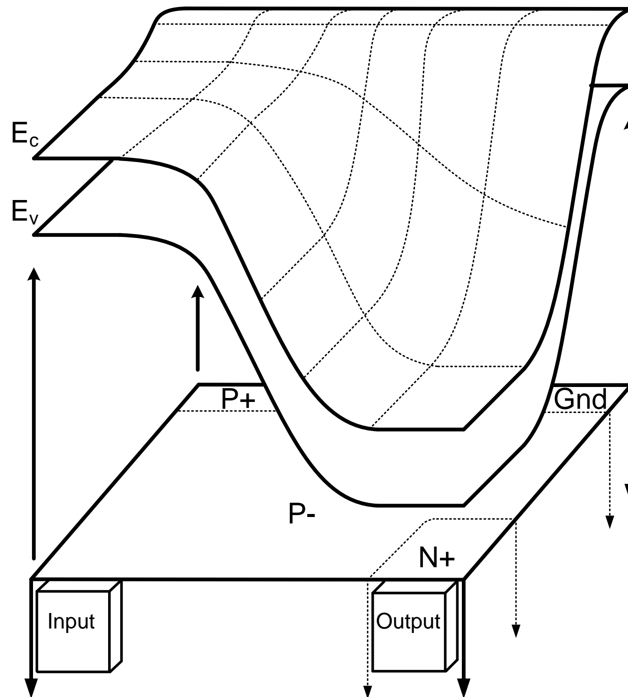


Figure 5.11: Surface band diagram over the top of a SIM cross section. This illustrates the bending of the bands for a biased SIM where the bias has caused complete depletion of the epitaxial layer between the input and output nodes and has become sufficiently strong to initiate impact ionization events.

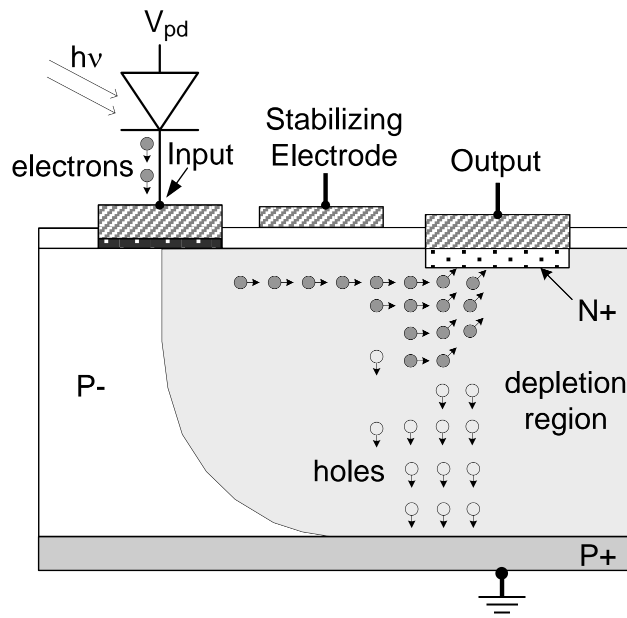


Figure 5.12: Slab diagram showing the breakdown stage of the SIM. At this stage the bias is large enough to extend the depletion region to the input node and also large enough for impact ionization to occur. Notice the path of the generated holes is toward the grounded substrate.

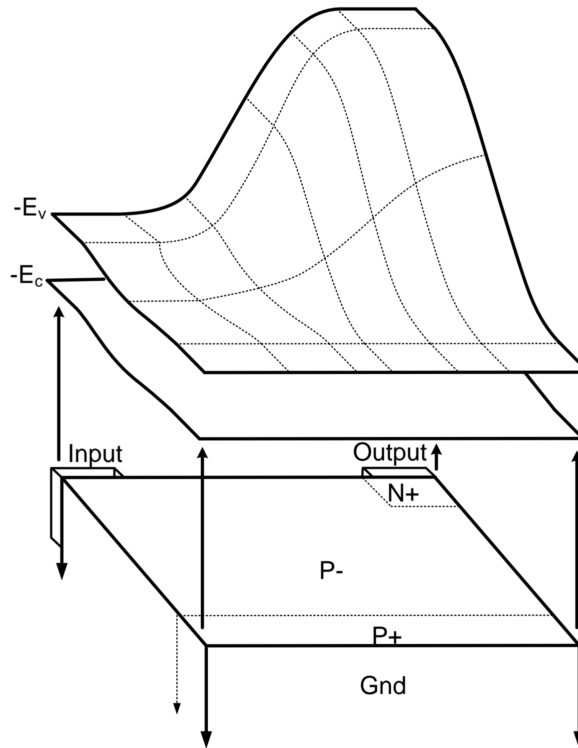


Figure 5.13: Inverted surface band diagram. This shows the inverse of the valence and conduction bands. This is the potential seen by generated holes in the depletion region. As indicated by the potential drop, the holes travel to the grounded P+ substrate instead of traveling back toward the input node.

node. Thus, they would drift back to the injection point where they would recombine with the input signal or other ionized electrons, effectively killing the gain.

As mentioned previously, the voltage at the input node floats. One other noteworthy point with regards to this is that once the depletion region fills the entire channel the voltage drop between the output node and input node is locked down by the voltage drop across the depletion region. This *locked down* voltage may be misleading. This is because the voltage drop across the depletion region is a function of the carriers present in that region. If the current at the injection point were constant then the voltage drop also is constant and the floating voltage never changes. But due to the electric field lowering through changes in carrier concentration in a

depleted region the voltage at the injection point becomes a direct function of the input current and the gain. This creates a resistive effect as we try to rapidly change the current levels at the injection node. The topic of a resistance due to the floating voltage will be discussed in detail later. At this point it is sufficient to say that high gains can effectively be achieved through proper biasing of the SIM.

5.5 Brief SIM Evolution

All solid-state devices go through alterations and changes as they develop and evolve. The SIM is no exception to this. As the research and development of the SIM has gone forward, alterations and changes have been made with many factors in mind. Consistency in fabrication has been one of those considerations. The SIM has been developed and fabricated in the cleanroom at Brigham Young University and, as such, it has become necessary to alter the design, at times, to accommodate the strengths of our fabrication capabilities. Focuses on consistency and yield in fabrication and design has produced consistency in testing and an increased robustness and durability to the SIM.

Performance is another main factor when examining alterations made to the SIM over the years. All devices have performance flaws. Researchers are constantly trying to come up with innovations that improve device capabilities and performance. Some of the main factors that are taken into consideration in the SIM are the gain and frequency response. Several alterations have been made with these two factors in mind. The remainder of this dissertation will deal with these two factors and other versions of the SIM. A brief outline of the most notable versions of the SIM will be given throughout the rest of this chapter. These outlines will be expounded on in later chapters.

5.5.1 Surface SIM (Schottky)

The surface SIM was one of the original versions of the SIM [47]. In this version the hole collectors are a series of doped wells at the surface as opposed to being the P+ substrate as shown previously in this chapter. Figure 5.14 illustrates the relative structure. Keep in mind that Figure 5.14 is showing a top view of the SIM as opposed to a side cross-sectional cut as shown previously. The operation principal of the surface SIM is much the same as the previously discussed version. An electron signal is injected at the input node and a voltage is applied to the output node. The voltage at the output is increased until the fields reach sufficient strength that the electron-hole pairs are generated through impact ionization. As always, the electrons are collected at the output node but in this case the holes are directed to the wells at the surface. This structure required more fabrication steps and lacked the consistency and durability of that of the vertical SIM structures. The vertical SIM was designed to make the fabrication more consistent and reduce the number of devices being burned out during the testing process. It also

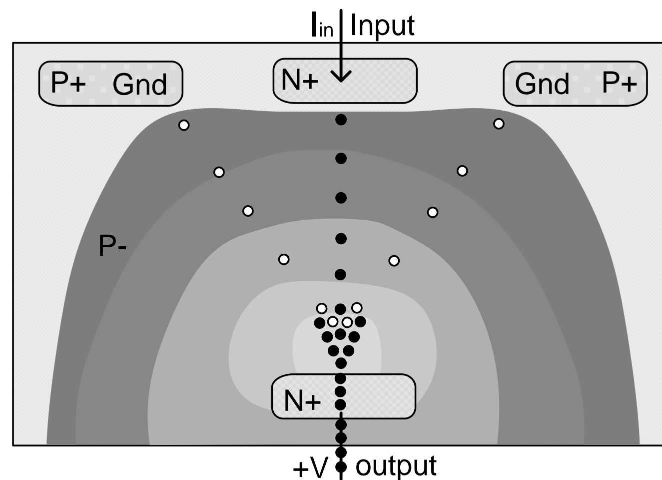


Figure 5.14: Top view of a surface SIM. This top view shows the doped wells in the epitaxial layer. Instead of the generated holes traveling through the epitaxial layer to the substrate they travel along the surface to grounded P+ wells. In this SIM all of the carrier transport happens at the surface as opposed to through the bulk.

made the electric fields more consistent so they somewhat resemble more traditional avalanche based amplifiers.

5.5.2 Vertical SIM (Schottky)

Due to the surface SIM suffering from several problems associated with high amounts of current traveling along the surface of the semiconductor, a new type of geometry was explored [92], [48]. This version of the SIM is seen in Figure 5.15. Having the holes travel through the bulk of the device is an obvious fix to that problem and at the same time it increases fabrication consistency and reduces doping steps. The operation of this version of the SIM was discussed earlier in this chapter. The metallization of the injection node did remain consistent from the surface SIM to the vertical SIM. This feature (the Schottky contact) becomes one of the more difficult points of the fabrication process and brings up questions about the metal-semiconductor barrier resistance.

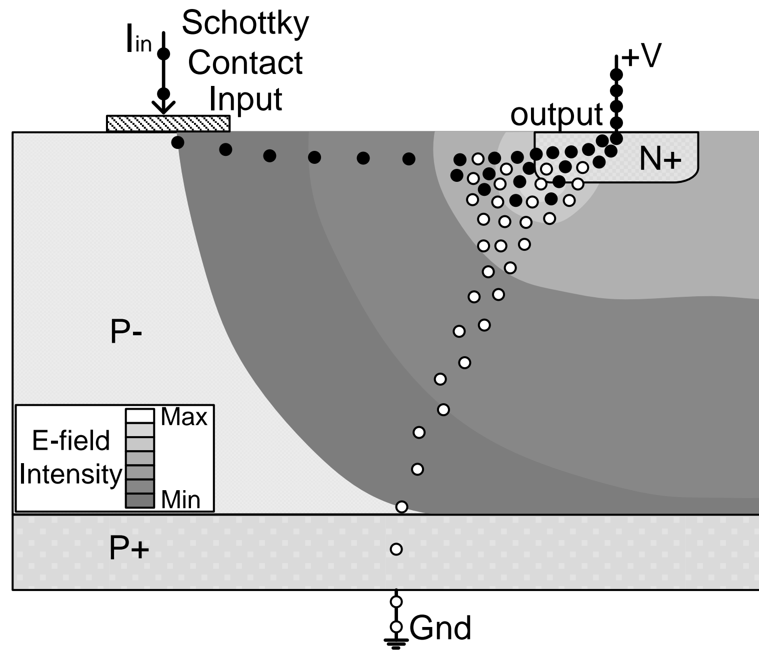


Figure 5.15: Cross-section and operation of the Schottky contact SIM.

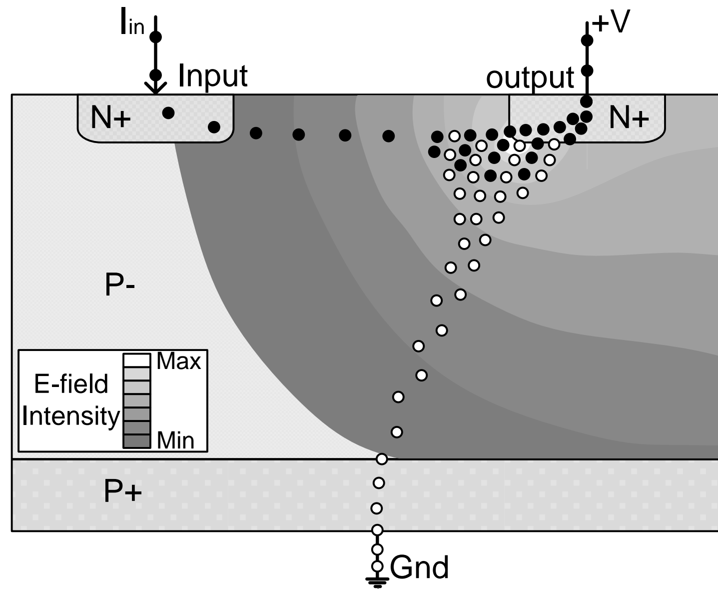


Figure 5.16: Cross-section and operation of the Ohmic contact (or pn junction) SIM. This version SIM is discussed in full detail in Chapter 7.

5.5.3 Ohmic Contact SIM

The Ohmic contact SIM seen in Figure 5.16 was designed with the purpose of overcoming the barrier resistance introduced by the Schottky contact at the injection node [49]. The discussion of this version of the SIM comprises chapter 6 of this document. At this point it is sufficient to point out that the general operation and concept of the SIM were maintained and the difficult step of metallization was simplified with this version.

5.5.4 Buried Oxide SIM

One of the main issues with previous versions of the SIM was gain controllability. The buried oxide SIM seen in Figure 5.17 was designed to increase the ionization efficiency and, thus, the gain controllability of the SIM [50]. This version of the SIM is discussed in length in chapter 7. As the previous versions of the SIM, this version maintains the same concepts of gain

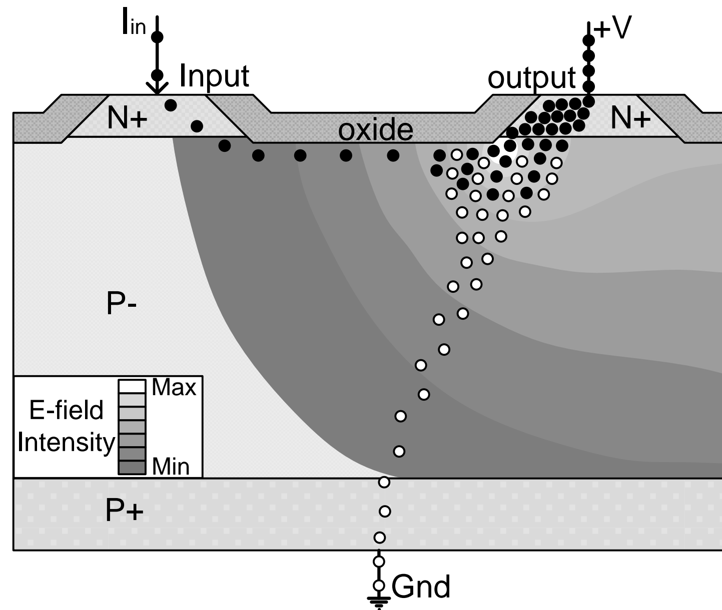


Figure 5.17: Cross-section and operation of the buried oxide SIM. This version of the SIM is discussed in full detail in Chapter 8.

operation. As will be discussed later, the introduction of pedestals confined the carries through the optimal electric field profile for more efficient gains. The introduction of this step did somewhat complicate the fabrication process but this step was necessary to be able to produce devices capable of achieving appreciable gains on a consistent basis.

5.5.5 Future SIM for Optimal Operation

The most recent alterations of the SIM have been done with hopes of increasing the frequency response. Chapter 8 will discuss the challenges faced in achieving gain and high frequencies. Chapter 9 will discuss the alterations necessary to achieve the optimal performance and operation in the SIM. As usual, these alterations leave the main points of device operation consistent with previous versions of the SIM.

5.6 Conclusions

This chapter reiterated the need that a device such as the SIM fulfills. It also touched on the advantages the SIM has over other devices used in modern day systems. After this, the basic operation of the SIM was outlined. This basic operation is consistent through all versions of the SIM. The variations of the SIM were briefly described in anticipation for their discussion in later chapters.

6 OHMIC CONTACT SIM

6.1 Introduction

In Section 5.5.3 of the previous chapter the ohmic contact SIM was briefly introduced. In the operation of previous versions of the SIM, current from an external source is injected through a Schottky contact. This chapter introduces a SIM design that utilizes an ohmic contact with a pn junction injection point [49]. One of the main benefits to the ohmic contact SIM is that the new design makes the devices more consistent to fabricate. Schottky contact characteristics can be highly dependent on semiconductor surface and metal deposition conditions which make consistency difficult [93], [94]. The Schottky contact SIM also suffered from frequency response limitations at low input currents due to thermionic emission over the Schottky barrier. This chapter will also touch on the frequency response limits for the pn junction injection device which were investigated in the hopes of improving the SIM's performance with low input currents [95].

6.2 Ohmic Contact Injection Node

The SIM operational stages (as discussed in Section 5.4) for the ohmic contact version can be seen in Figure 6.1. By comparing Figure 6.1a with Figure 5.5 you can see the main structural differences between the ohmic contact version of the SIM and the previous Schottky contact

SIM. Notice that the two structures are very similar with the main difference being at the injection node.

As you can see in Figure 6.1, the operation of the SIM is consistent with the previous version of the SIM. For comparison purposes, this figure can be paralleled to the surface band

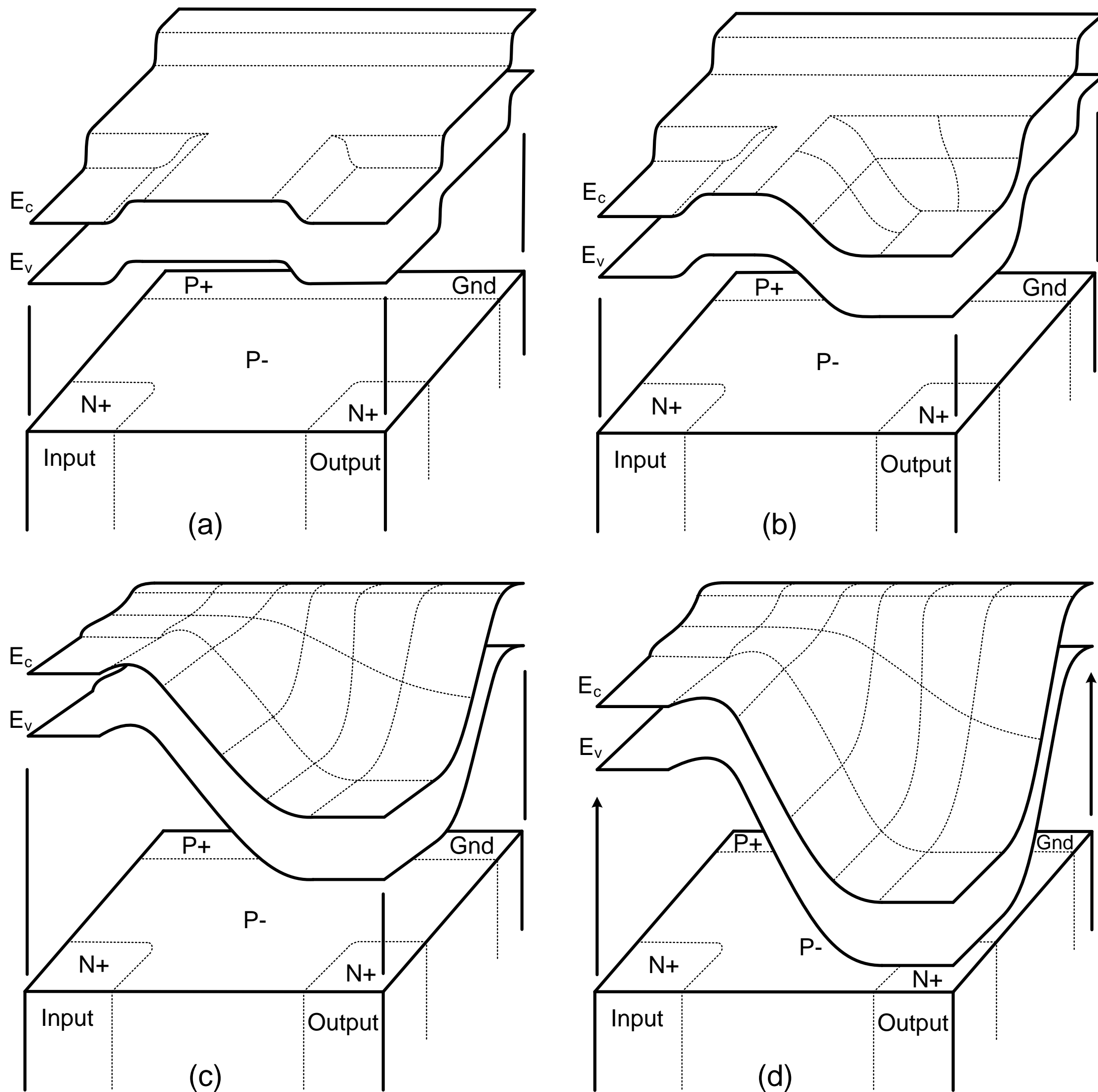


Figure 6.1: Surface band diagram for the SIM during various stages of bias at the output node. a) No bias applied to the output node. b) Bias applied at output node isn't sufficient to fully deplete the semiconductor between the input and output. c) Bias applied at output node fully depletes semiconductor between input and output but isn't sufficient to initiate impact ionization events. d) Bias is large enough for impact ionization to occur.

diagram figures of the previous chapter. The doping profile has changed and this has affected the overall band diagram in some ways but the concept remains consistent. Figure 6.1a illustrates the unbiased surface band diagram. When a bias is applied to the output we see the bands bending as shown in Figure 6.1b. If sufficient bias is applied to the output node the semiconductor between the input and the output (and between the output and the substrate) will become completely depleted as indicated by Figure 6.1c. At this point, if the bias is increased, there is an increase in the rate of change of the band energy from the input to the output. This can be seen in Figure 6.1d. This rate of change facilitates impact ionization events. By further increasing the bias the gain will increase. This process is identical to that explained in the previous chapter in Sections 5.4.1-5.4.3.

As previously seen, the Schottky contact SIM had the metal-semiconductor contact at the injection node which was made directly between the epitaxial layer and the deposition metal. This formed the Schottky contact at the p-type epitaxial layer. One of the primary factors considered in altering the structure of the SIM was the fabrication process. The new ohmic contact design makes fabrication much more consistent and reduces the number of critical steps in this process. This ease of fabrication expedites the transition to testing.

6.3 Metal-Semiconductor Barrier

One of the main changes to the band diagrams between the Schottky contact SIM and the ohmic contact isn't seen in Figure 6.1. This is because the surface band diagrams shown in Figure 6.1 are solely for the bands within the semiconductor and don't take into consideration the metal-semiconductor contact. These differences are illustrated in the band diagrams of Figure 6.2. Notice that in Figure 6.2b (ohmic contact) the electrons at the interface tunnel through the

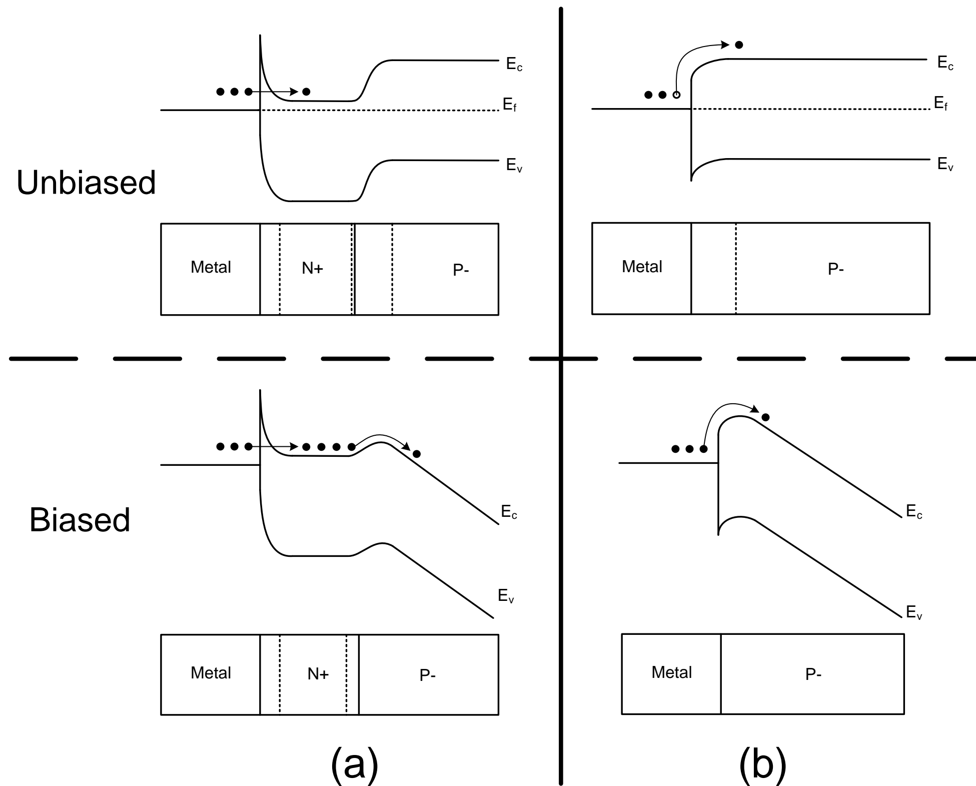


Figure 6.2: Band diagrams illustrating the metal-semiconductor interface at the injection node for (a) the ohmic contact SIM and (b) the Schottky contact SIM. The bands are shown for the biased and unbiased cases for both versions of the SIM.

barrier as opposed to Figure 6.2a (Schottky contact) where the electrons must obtain sufficient energy to overcome the barrier via thermionic emission. This difference is key when comparing the resistive effects of both barriers.

The understanding of this barrier interface is necessary in order to see the motivation behind the move from a Schottky to an ohmic contact. In order to better grasp the barrier issues, a derivation of the barrier resistance will now be shown. It begins by taking the basic Schottky SIM structure from Chapter 4 to calculate the barrier resistance ($R_{barrier}$). This can then be compared analytically to the resistance from an ohmic contact.

As was mentioned previously, the barrier resistance is due to thermionic emission over the metal-semiconductor barrier as seen in Figure 6.2a. In order to extract a resistance from this barrier it becomes necessary to relate the current and voltage through the barrier. It is necessary at this juncture to point out that the height of the barrier isn't constant for different voltages and currents. As the voltage applied to the output changes there is a corresponding change in the height of the barrier at the input (considering the SIM is operating in either the transitional or breakdown stages and the channel is completely depleted). Figure 6.3a shows this barrier lowering for the Schottky contact SIM. Notice in Figure 6.3b (the ohmic contact) that the barrier at the metal-semiconductor interface isn't affected by the voltage at the output node. There is, though, a lowering of the barrier at the N+ to P- junction that will be taken into consideration at the appropriate time. At this time, the focus will be placed on the Schottky contact barrier.

From the figure, there is a drop in the barrier height ($\Delta\Theta$) for a corresponding increase in the voltage at the output (ΔV). The current through the device is described by the following

$$I = I_o e^{\Delta\Theta / kT} + I_d \text{ [4]}. \qquad \textbf{6.1}$$

In Equation 6.1, the first term corresponds to the current through the barrier and the second term is the generated current within the semiconductor. Thus, I_o is the current through the device as if there were no barrier lowering, $\Delta\Theta$ is the amount of barrier lowering due to ΔV, k is Boltzmann's constant, T is temperature and I_d is the dark current within the semiconductor. To obtain the rate of change in current with respect to the rate of change in barrier height the derivative is taken as in

$$\frac{dI}{d\Delta\Theta} = \frac{(I - I_d)}{kT}. \qquad \textbf{6.2}$$

Resistance is dI/dV or in this case

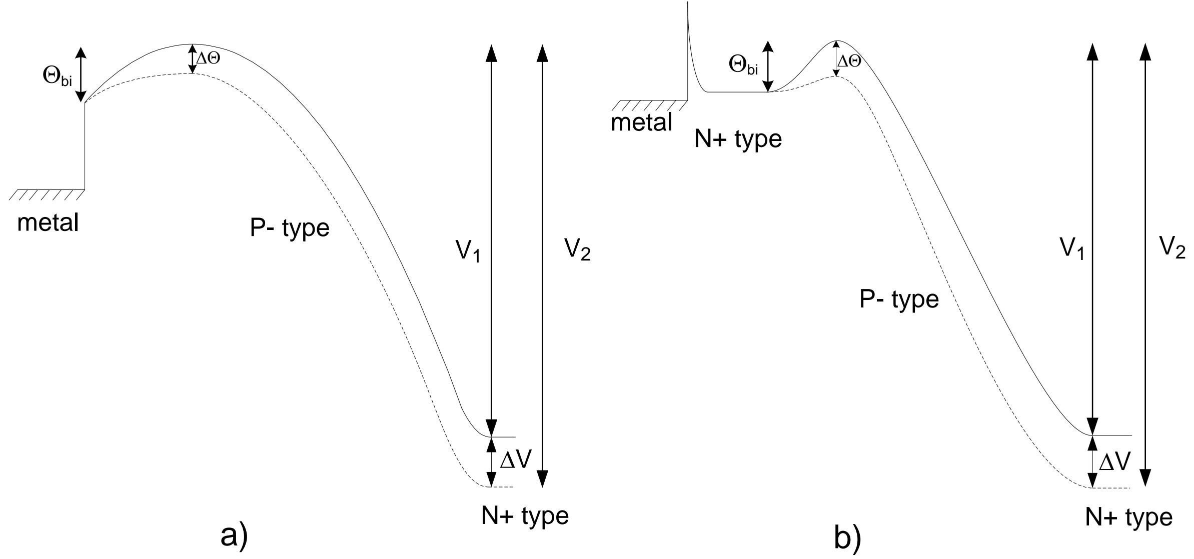


Figure 6.3: Band diagram for the metal-semiconductor contact in both the Schottky and ohmic contact SIMs. (a) The Schottky contact SIM. It shows a lowering of the barrier height at the metal-semiconductor interface for a change in voltage at the output node. (b) The ohmic contact SIM. It shows no change in the barrier height at the metal-semiconductor interface but also illustrating the barrier lowering at the pn junction.

$$\frac{dI}{dV} = \frac{-dI}{d\Delta\Theta}\frac{d\Theta}{dV}. \qquad \textbf{6.3}$$

The $d\Theta/dV$ is obtained by examining the electric fields between the input and the output nodes when the channel is completely depleted. Figure 6.4 shows the electric field versus position for both V_1 (Figure 6.4a) and V_2 (Figure 6.4b) of Figure 6.3a. These electric field profiles are obtained by taking the derivative of the potential with respect to distance across the channel from the input electrode to the output N+ well. As indicated by Figure 6.4a and Figure 6.4b the area under the electric field curve between 0 and W is equal to the height of the barrier for V_1 and V_2 respectively. In both cases the barrier height can be written as

$$\Theta = \frac{qN_A}{\varepsilon_s}\frac{W^2}{2}. \qquad \textbf{6.4}$$

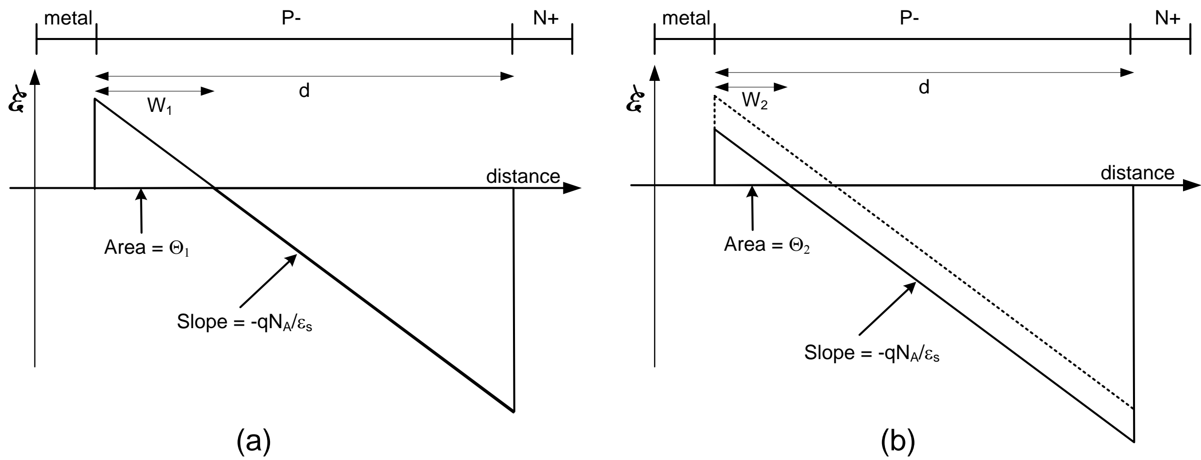


Figure 6.4: Electric field profile between the input contact and the output N+ well. a) Corresponds to the derivative of the potential for V_1 of Figure 6.3a. b) Corresponds to the derivative of the potential for V_2 of Figure 6.3a.

In Equation 6.4, q is the charge of an electron, N_A is the epitaxial doping level and, ε_s is the permittivity of silicon. Notice that for an increase in the output voltage there is a corresponding decrease in W. This change is exaggerated in the figures for illustrative purposes. But the relationship between the voltage at the output and the distance W is

$$V = \frac{qN_A}{\varepsilon_s}\frac{(d-W)^2}{2}. \qquad \textbf{6.5}$$

Combining Equations 6.4 and 6.5 yields the derivative of the barrier height with respect to the output voltage as in

$$\frac{d\Theta}{dV} = \frac{-W}{d-W}. \qquad \textbf{6.6}$$

The relationship between W and the current through the device can be obtained from Equations 6.1 and 6.4 to get

$$W = \sqrt{\frac{2\varepsilon_s}{qN_A}} \cdot \sqrt{\Theta} = \sqrt{\frac{2\varepsilon_s}{qN_A}} \cdot \sqrt{\Theta_{bi} - kT \ln\left(\frac{I - I_d}{I_o}\right)}. \qquad \textbf{6.7}$$

Using the relationship for *W* in Equation 6.7 in Equation 6.6 yields

$$\frac{d\Theta}{dV} = \frac{1}{\left(\frac{d}{\sqrt{\frac{2\varepsilon_s}{qN_A}} \cdot \sqrt{\Theta_{bi} - kT \ln\left(\frac{I - I_d}{I_o} \right)}} \right) - 1}. \qquad \textbf{6.8}$$

From this we can get the barrier resistance by applying Equations 6.2, 6.6 and 6.8 to Equation 6.3 yielding:

$$R_{barrier} = \left(\frac{dI}{dV} \right)^{-1} = \frac{kT}{I - I_d} \frac{d - W}{W} = \frac{kT}{I - I_d} \left(\left(\frac{d}{\sqrt{\frac{2\varepsilon_s}{qN_A}} \cdot \sqrt{\Theta_{bi} - kT \ln\left(\frac{I - I_d}{I_o} \right)}} \right) - 1 \right). \qquad \textbf{6.9}$$

Theoretical values for current versus voltage of a Schottky contact and an ohmic contact can be seen in Figure 6.5. The nature of a Schottky contact is that of a rectifier while the ohmic contact appears linear like a resistor. The slope of the ohmic contact curve depends on the doping of the N+ region among other factors. The high N+ doping at the junction creates a thin barrier that can be tunneled through and the tunneling relationship is linear with current like a resistor. This corresponds to an ohmic behavior as seen in the current versus voltage curves of Figure 6.5. This linear, resistive, pattern yields the name *ohmic contact*. The current versus voltage characteristics are diode-like for the Schottky contact. The slope of these curves yields the resistance of the contact. It becomes apparent from this that at very low currents the slope of the Schottky contact produces a significantly higher resistive effect than that of the ohmic contact due to the rate of change of current with respect to voltage near the origin.

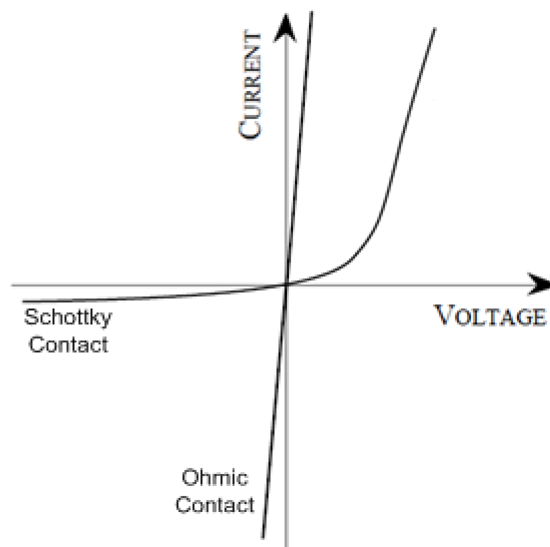


Figure 6.5: Current versus voltage curve illustrating the resistive effects of an ohmic contact and a Schottky contact.

6.4 Simulation and Design

As we learned in Chapter 4, simulation and modeling techniques are valuable when trying to search through possible solutions to problems, improve device operation, or test device theories. Silvaco is often used to simulate device operations for the SIM. As we just learned, one of the issues experienced with the Schottky contact SIM was the barrier resistance caused by the Schottky contact that can be seen in Figure 6.2 and its corresponding resistive effects from the current versus voltage curves in Figure 6.5.

The initial investigation into changing the injection node structure was to overcome the barrier issue at the metal-semiconductor junction. Initial attempts (in simulation) to overcome this were to create a P+ well under the injection node with sufficiently high doping so that, when properly biased, the injected signal tunnels through the barrier. Simulations showed little to no

change in the resistance at achievable doping levels with this method because the barrier was not sufficiently narrow for tunneling to occur at the voltages used during normal operation.

The next step, in attempts to decrease barrier resistance, was to create an ohmic contact at the injection node by creating a highly doped N+ well under the injection node. This hadn't been attempted before due to the introduction of a pn junction in the semiconductor bulk. Fears that this junction would affect proper depletion of the channel, among other concerns, discouraged this line of thinking. If the channel doesn't properly deplete or the field profile isn't developed correctly, gain may become unachievable. Fortunately, the introduction of a pn junction at the injection node proved not to be an issue in simulation and tests. When looking at these results, it was difficult to be conclusive since tunneling effects and other quantum effects are often difficult to simulate for reasons discussed in Chapter 4.

The simulation of the Ohmic contact SIM was performed in the ATLAS platform of Silvaco. Selberherr's model for impact ionization was used to model the ionization events and the Shockley-Read-Hall recombination model for carrier recombination events. Careful consideration of the mesh setup at the N+ to P- interface was taken in order to simulate an abrupt-like junction. The basic structure in Silvaco can be seen in Figure 6.7.

Typical voltages were applied to the output node to test the DC characteristics of the structure. The results can be seen in Figure 6.6. Notice from the curves that the design is able to produce appreciable gains at achievable voltage levels for low input currents. See Appendix A.1 for the Silvaco ATLAS code used to produce this simulation.

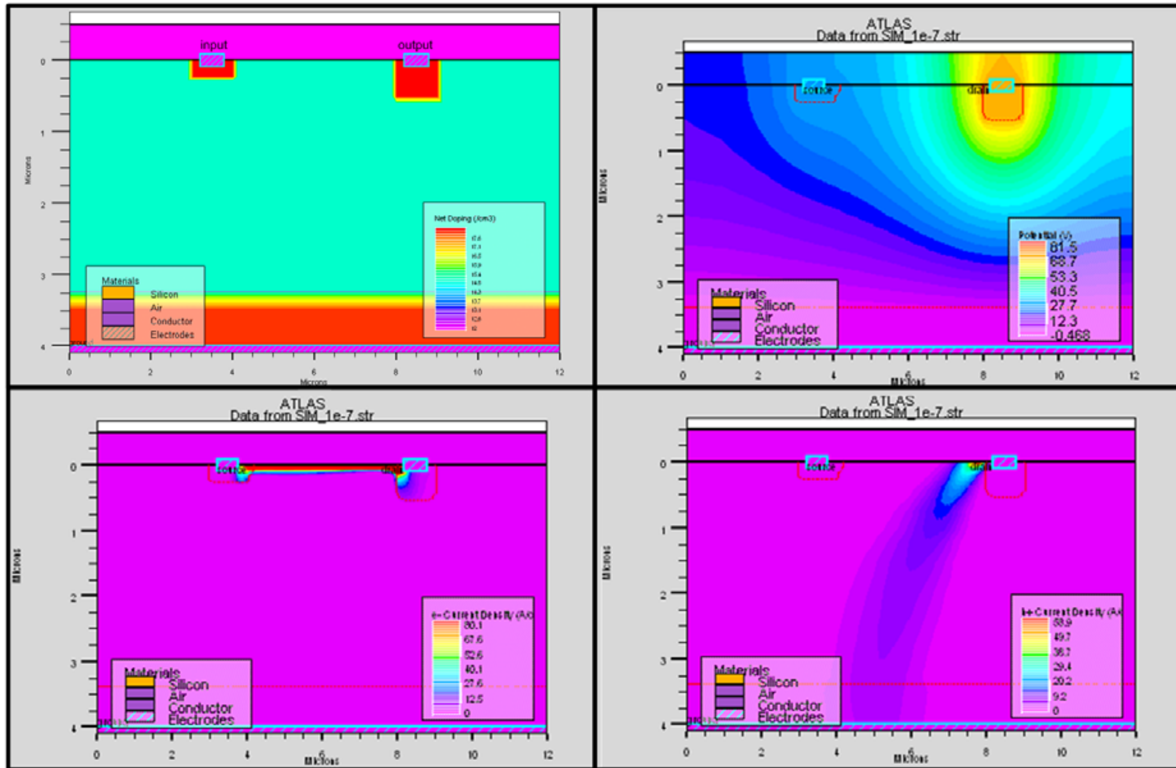


Figure 6.7: Cross-section of the ohmic contact SIM as simulated in Silvaco. Upper-left hand shows the doping concentrations for SIM. Upper-right hand shows the potential curves for the biased SIM. Lower-left hand is the electron current density for the biased SIM. Lower-right hand corner is the hole current density for the SIM.

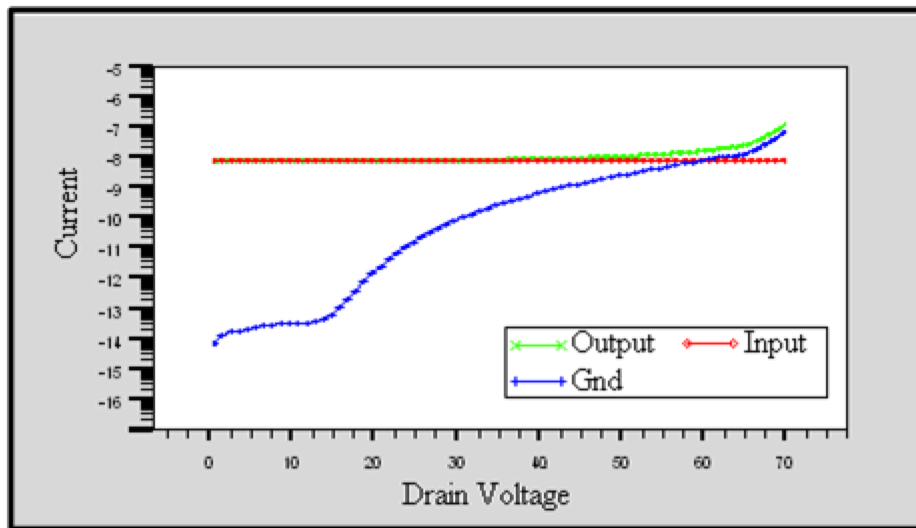


Figure 6.6: Current versus voltage curves for the ohmic contact SIM produced by Silvaco.

These simulation results indicate that the introduction of an N+ well at the injection node does not affect the expected operation of the SIM. They also indicated that these changes, at worst, leave the resistances and frequency response of the device unchanged. There were also promising signs that pointed to a decrease in resistance and an overall increase in the frequency capabilities of the SIM. Also, because an N+ doping step was already included in the fabrication process, it is fairly seamless to integrate into the current process. Armed with this information, steps for the design and fabrication of the ohmic contact SIM were taken.

6.5 Fabrication

As alluded to earlier, one of the biggest advantages to the new design is that it makes the devices more consistent to fabricate. Metallization can be quite difficult and the barrier is highly dependent on semiconductor surface and metal deposition conditions. This is especially true when designing well characterized Schottky contacts. Because of this, Schottky contact SIMs often suffered from inconsistent testing results.

To ensure that ohmic contact SIMs produce the desired current gains and contact resistance, the devices were fabricated in silicon after the models developed in Silvaco's ATLAS. As usual, the fabrication was performed using silicon epitaxial wafers doped to approximately 3Ω-cm in the 10μm p-type epitaxial layer and 0.01Ω-cm in the p-type substrate. First, a 30nm layer of silicon dioxide was thermally grown on the silicon surface in a tube furnace. Wells of varying widths (30–360μm) and spacing (3–9μm) were then patterned with photoresist on the oxide and then the wafers implanted with a heavy phosphorous dose. Notice that this step produced both the input and output well simultaneously. Wafers were then

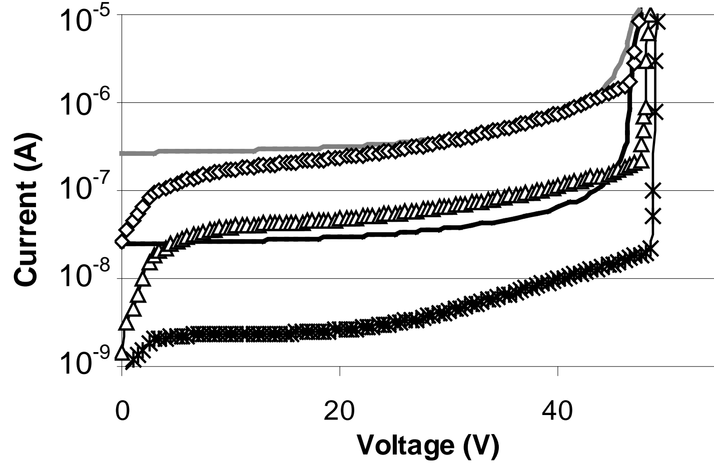
activated at 1050 C in a forming gas environment for 30 seconds. Front and backside aluminum thermal evaporation served to create a substrate contact and individual leads to devices. Wafers were then annealed in a forming gas environment for 5 min at 450 C to produce low-resistance ohmic contacts to heavily doped regions. More details on the fabrication of SIM devices can be found in Appendix 0.

6.6 Current and Voltage DC Characteristics

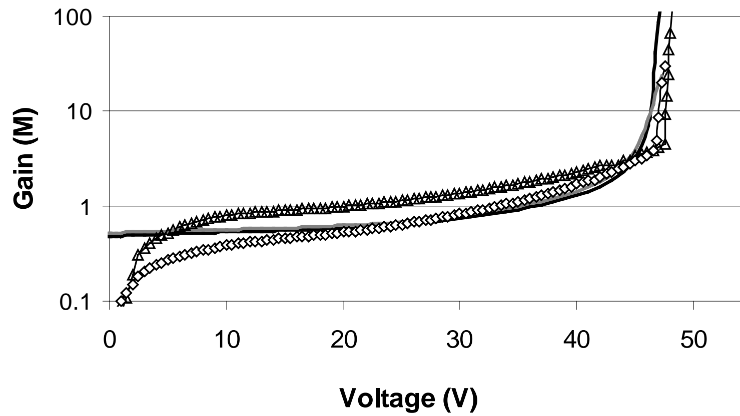
Fabricated devices were packaged in TO-5 cans and setup for testing. The current versus voltage relationship and current gain were measured using a silicon photodiode (Perkin Elmer VTP1012) as a current source. Gain (*M*) is calculated using

$$M = \frac{I_{output} - I_{dark}}{I_{input}}. \quad \textbf{6.10}$$

The photodiode was wired to the injection node of the SIM and reversed biased. A laser emitting 870 nm light (Fiberlink XA-1000A-1) was used to illuminate the photodiode, injecting electrons into the SIM. The laser was modulated with an Agilent 33250A waveform generator. Source-meter units (Keithley 2400 and 2410) biased the SIM and photodiode while monitoring corresponding input and output currents. Figure 6.8 shows measurements for different input currents along with ATLAS simulations for equivalent devices. At low voltages applied to the output electrode, the electric field is too weak for impact ionization to occur. As the bias is increased so that the depletion region approaches the n-well beneath the injection point, electrons are injected into the depletion region where they drift toward the positively biased output electrode (without impact ionization gain). With continued increase of the bias, impact ionization begins to occur as the electric field becomes stronger. With increased bias, breakdown occurs in the high vertical field between the output electrode and the substrate and in



(a)



(b)

Figure 6.8: a) Output current versus voltage curve for modeled and fabricated pn junction injection SIMs. The solid black and gray curves represent simulated injection currents of 50nA and 500nA respectively. These can be compared to the measured results for 50nA and 500nA injection tests represented by the diamonds and triangles respectively. The leakage current curve for the fabricated devices is also shown and is represented by the Xs. The voltage shown in the graph is applied to the output node. The fabricated device had a width of 10µm between the n-wells of the injection and output nodes. b) Gain (M) versus voltage curve for the same modeled and fabricated pn junction injection SIM devices from (a).

the horizontal field between the injection point and the output. When this happens, there is a marked jump in the output current at around 45 V for the device shown in Figure 6.8. Figure 6.8b shows the gain versus voltage curve for the same device indicated in Figure 6.8a.

It should be noted that the floating voltage at the injection node is not considered a direct function of V_{out}. It is directly related to the current through the depletion region and, thus,

depends on the gain which is a function of the output voltage. For the gain levels in Figure 6.8, the change in floating voltage (ΔV) over this range of output voltage is: $\Delta V = 0.64$ Volts. Also, the data from Figure 6.8 was obtained from a device with a width of 10μm between the n-wells of the injection and output nodes. Devices of varying spacing have been tested as well. Varied spacing causes the devices to deplete at slightly different voltages but once breakdown voltages are reached no noticeable effect on the gain or frequency response of the device has been shown.

6.7 Frequency Response

For frequency response measurements, the laser source described above was driven by sinusoidal signals of different frequencies using a waveform generator (Agilent 33250A) as seen in Figure 6.9. The SIM's output signal was connected through a bias tee (Picosecond Pulse Labs 5530A) to a transimpedance amplifier (Femto HCA-100M-50K-C) converting the ac output into a voltage signal. This signal was analyzed and measured on an oscilloscope (Tektronix TDS 340A) [1]. The SIM was operated at an approximate gain of 3 as obtained by Equation 6.10 for the measurements shown in Figure 6.10. This figure has the 3dB frequencies versus various input currents. These results are very comparable to previous Schottky contact SIMs as can be seen in Table 6.1.

Table 6.1: Comparison of frequency limit of the Schottky contact SIM to that of the ohmic contact SIM.

Current (μA)	3dB Frequency (kHz) PN SIM	3dB Frequency (kHz) Schottky SIM
0.1	40	27
1.0	240	135
3.0	440	280
6.0	540	440

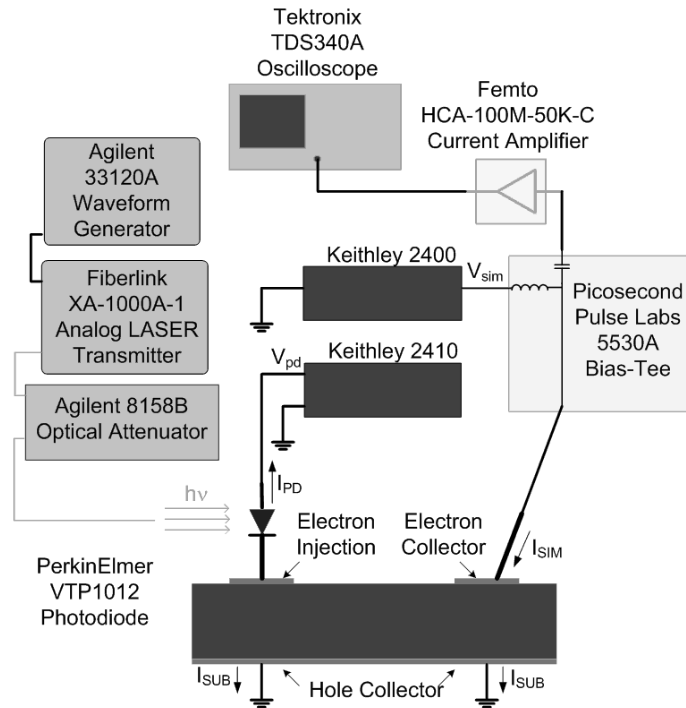


Figure 6.9: Testing setup for extracting the frequency response in the SIM.

Even though the frequency responses were comparable, meaning there was no loss of functionality with an improved fabrication design, this was not quite what was expected. With the reduction in the barrier height resistance there should have been a more noticeable increase in the 3dB frequencies for the ohmic contact SIM. One of the reasons for the damped improvement was the presence of a floating voltage at the injection node. It was at this point in the evolution of the SIM that the floating voltage at the input was suspected of causing major speed delays in the operation of the SIM.

The floating voltage exists because there is no external voltage to lock down the voltage at the injection node. Because of this, the voltage at this point becomes a direct function of the current density found in the depletion region of the device. And the current density present in

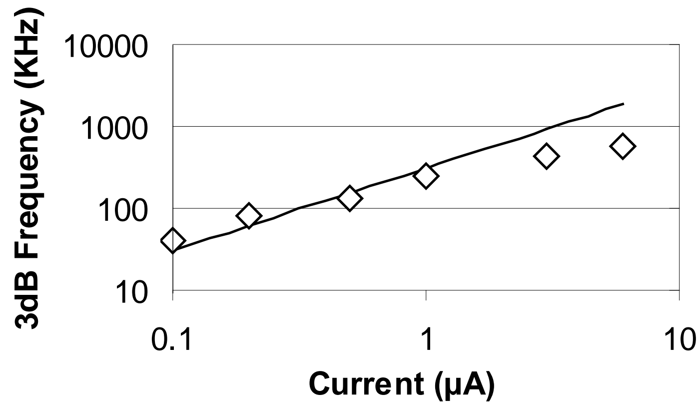
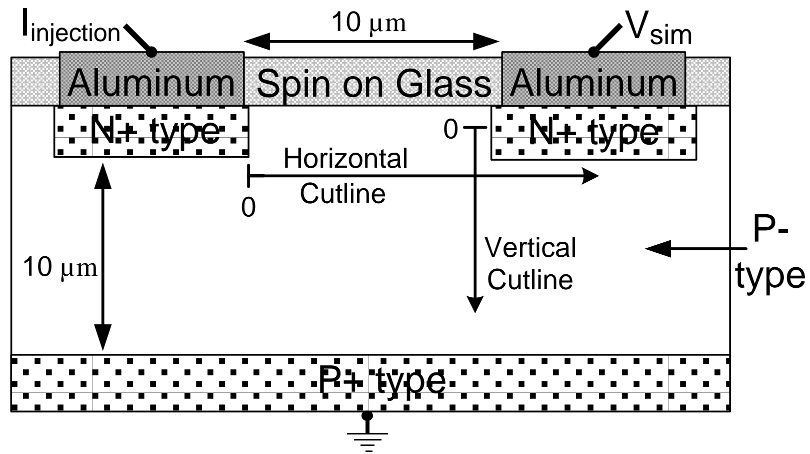


Figure 6.10: Illustrates the measured 3dB bandwidth for the fabricated SIM, represented by the diamond points, over a range of current levels. The theoretical response predicted by Eq. 5 is shown as a solid line, with 20 pF used for the node capacitance. A bias of approximately 40 volts was used.

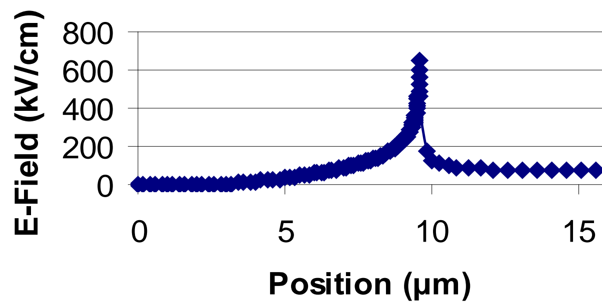
the depletion region is largely determined by the injected current levels. This means that the injection point voltage is a function of the injection point current yielding a resistive effect. This resistive effect was previously noted but written off to be insignificant compared to the barrier resistance. At this point though, it became obvious that the resistive effect played a larger role in the operation of the SIM than previous thought. The floating voltage resistance will be discussed in full in Chapter 8 in Section 8.5 of that chapter.

6.8 Gain Controllability

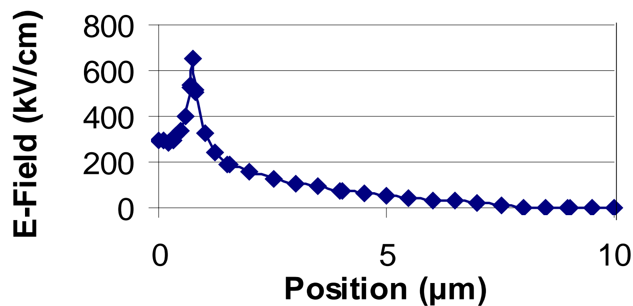
Well designed silicon APDs have gain curves that increase more gradually with voltage which is desirable because the gain settings are more controllable. The reason for the more abrupt breakdown in this SIM lies in the electric field distribution. Figure 6.11a shows the device geometry, while Figure 6.11b and Figure 6.11c show the electric field extracted from the device simulations from Silvaco. The curves are taken along cut-lines in the horizontal and vertical directions as displayed in Figure 6.11a. The plots show that the electric field is very high near the output electrode but relatively low elsewhere. Most of the impact ionization takes



(a)



(b)



(c)

Figure 6.11: a) Cross section of the SIM showing the geometries and doping types. Approximate doping levels include: N+ type=1e18, P-type=1e15, P+ type=1e18. b) Illustrates field profiles for the horizontal cutline shown in (a) when V_{SIM} is at 50 volts and the injected current is 1 micro-Amp. c) Illustrates the field profiles for the vertical cutline in (a) for the same conditions as (b). Electric field lines were extracted from ATLAS simulations for the structure.

place in these high field regions. Because the electric field is so high, even small changes in field magnitude create large changes in impact ionization generation [96]. Consequently, small voltage changes on the SIM's output node lead to abrupt breakdown in the current versus voltage curves. Creating future SIM designs that have more controllable gains will involve engineering the electric field regions to be more equally distributed between electrodes – something closer to constant magnitudes throughout the depletion region. This can be done through doping and geometry alterations. Research and solutions to the gain controllability can be seen in the following chapter.

6.9 Conclusions

The ohmic contact SIM operates with a similar performance to the previous design based upon the Schottky contact. The frequency response is limited by the resistive effects caused by shifts in the quasi-Fermi level. These shifts are due to the varying current levels through the depletion region and result in a kT/qI resistive limiting factor. Future considerations will be given to this effect and how to reduce it, possibly by introducing a DC current offset. This could potentially increase the frequency response of the device. Other considerations to improve the speed will include space charge resistance characterization by altering the device geometry. As we will see now, changes to the ohmic contact SIM have been focused on making the gain curves more gradual in order to improve the output controllability.

7 BURIED OXIDE SIM

7.1 Introduction

As we saw in the previous chapter, the ohmic contact SIM suffered from gain controllability issues. Steps were taken to alter the SIM design in order to overcome these problems by confining the carrier path through the gain region. This chapter deals with the design changes and theory behind the gain controllability, impact ionization efficiency and the development of the buried oxide SIM design.

7.2 Impact Ionization Efficiency

Figure 6.1 illustrated the ohmic contact SIM device structure and gave insight into the surface band structure for the device in various stages of operation. Notice in the breakdown stage that the band structure yields a potential gradient for the electrons to stay along surface of the device as the drift from input to output [96]. This, accompanied with the higher concentration of defects and impurities along device surfaces, leads to the carriers being somewhat confined to the surface of the device. This wouldn't be an issue if it wasn't for a non-ideal electric field profile with respect to this carrier path. The distribution of the electric field through the depletion region and the path of the carriers determine the impact *ionization efficiency* of injected electrons. We define impact ionization efficiency as the ratio (η) of

electrons passing through the high electric field and experiencing impact ionization (n_{amp}) to the total injected electrons (n_{in}) as in:

$$\eta = \frac{n_{amp}}{n_{in}}. \qquad \textbf{7.1}$$

Using Silvaco's ATLAS, simulations were run to determine the electric field profile. This is illustrated in Figure 7.1. As discussed before, the electric fields are created due to the bias applied at the output electrode. The bias creates a vertical field between the output electrode and the substrate and a horizontal field between the injection point and the output [97]. As expected, the combination of the vertical and horizontal fields creates a maximum electric field located at the corner of the output's pn junction as seen in Figure 7.1. Also, the doping profile plays a big role in the electric field distribution. Radius of curvature of the doping changes the electric field distribution as well [5]. The location of highest radius of curvature has a higher electric field (Doping junction edge effects due to doping profiles can play a huge part in the electric field distribution. This will be discussed more fully in the following chapter. In Figure 7.1 this effect was exaggerated for illustrative purposes.). With this information the electron current density was also taken from Silvaco simulations. This is shown in Figure 7.2.

Now, taking into consideration the path of the carriers from Figure 7.2 and the location of the electric field's highest magnitude shown in Figure 7.1, it becomes apparent that the majority of the injected carriers do not pass through the optimal amplification region. This greatly reduces the number of electrons that are being amplified (n_{amp} from Equation7.1) due to impact ionization. At low impact ionization efficiency, if there is some net signal gain, a small fraction of the input electrons are experiencing very high gains while the larger portion of the input electrons are experiencing little gain. This can lead to an increased excess noise factor compared to a signal that is being uniformly amplified.

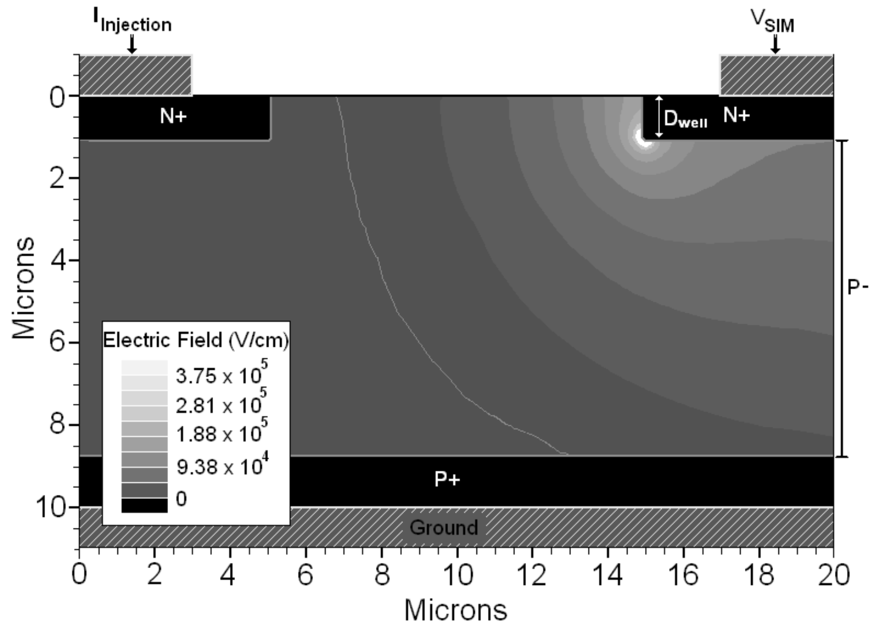


Figure 7.1: Cross section of the buried oxide SIM showing the magnitude of the electric field, as calculated in ATLAS, with V_{sim}=50V and $I_{injection}$ =10µA. The oxide depth (d_{ox}) and the well depth (D_{well}) are labeled (in this illustration d_{ox}=D_{well}).

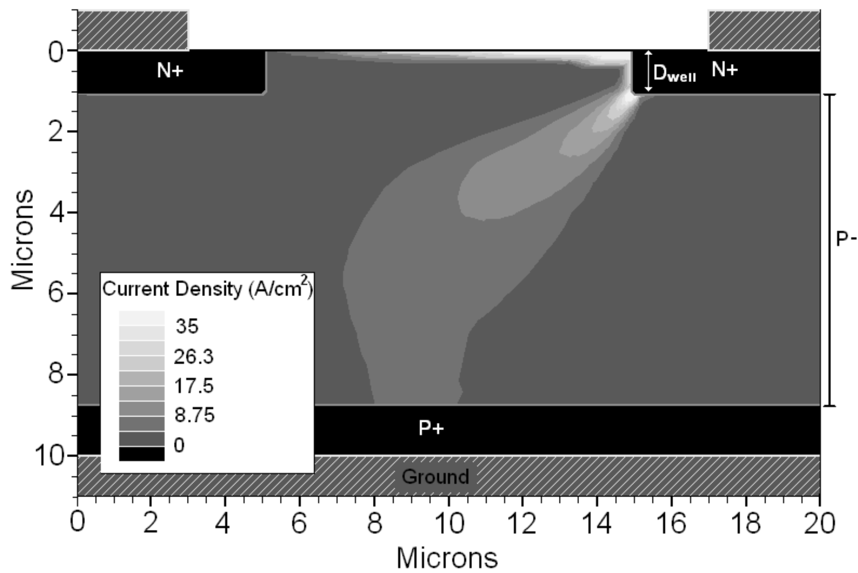


Figure 7.2: Cross section of the SIM showing the magnitude of the current density, as calculated in ATLAS, with V_{sim}=50V and $I_{injection}$=10µA. The holes produced by impact ionization are shown moving diagonally down to the substrate.

To illustrate how the effects of low impact ionization efficiency influence the gain versus voltage relationship, consider the gain equation for impact ionization initiated by electrons in silicon:

$$\frac{1}{M} = \exp\left(-\int_0^W (\alpha-\beta)dx\right) - \int_0^W \beta \exp\left(-\int_x^W (\alpha-\beta)dx'\right)dx, \qquad \textbf{7.2}$$

where α and β are the electron and hole ionization coefficients respectively [98]. When 100% of the electrons are capable of contributing toward the gain (M) then the output current, I_{out}, is defined as

$$I_{out} = I_{in}M \ , \qquad \textbf{7.3}$$

where M represents some average gain experienced by a given electron and I_{in} is the input current. In the case where only a fraction of the injected electrons are entering the high field region and effectively contributing toward the gain we have an output signal of

$$I_{out} = I_{in}\eta M + I_{in}(1-\eta)\,. \qquad \textbf{7.4}$$

The electron and hole ionization coefficients (α and β) are dependent upon the magnitude of the electric field in which the carriers are found by

$$\alpha = a_1 e^{-\frac{a_2}{E}}, \qquad \textbf{7.5}$$

and

$$\beta = b_1 e^{-\frac{b_2}{E}}. \qquad \textbf{7.6}$$

Here, a_1, a_2 and b_1, b_2 are constants which depend upon the range of the electric field’s magnitude [65].

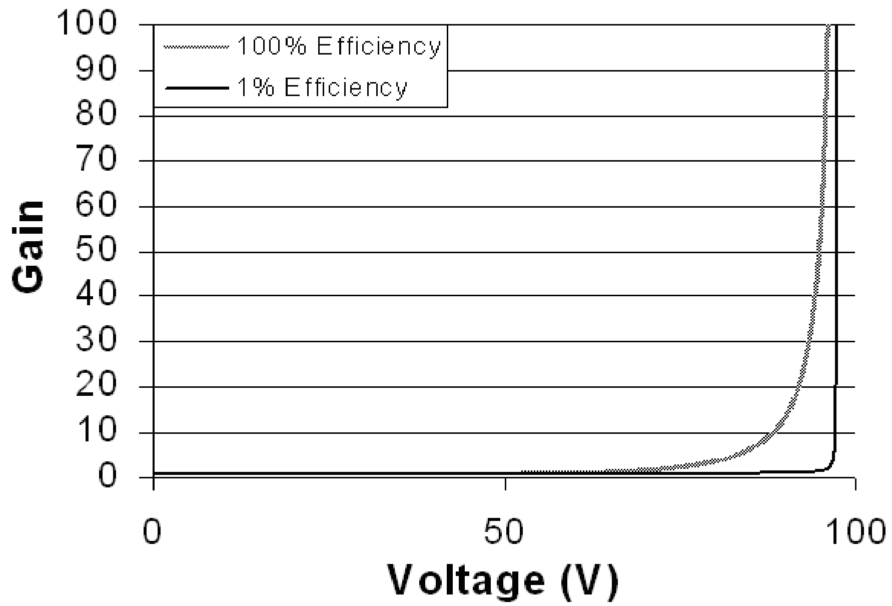


Figure 7.3: Theoretical gain vs. voltage plot as produced by Equations 7.1 - 7.6 showing how the gain curves rise very abruptly with low electron injection efficiency. This compares the theoretical cases of 1% and 100% efficiency for electron injection into the high field region.

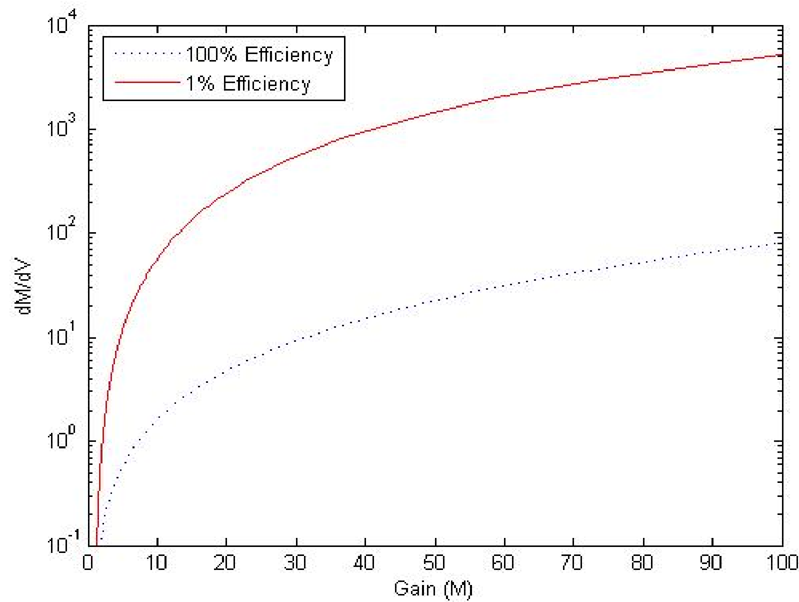


Figure 7.4: Theoretical gain vs. voltage plot as produced by Equations 7.1 - 7.6 showing the derivative (dM/dV) for Figure 7.3.

Using Equations 7.2-7.6, a plot of the relative gains for 100% efficiency and 1% efficiency can be plotted over a range of voltages. This is shown in Figure 7.3 and illustrates the importance of achieving high impact ionization efficiency. Note the rate of change of the gain (*M*) with respect to the voltage (*dM/dV*) shown in Figure 7.4. For the 1% efficiency case, this rate is much larger for a given voltage. Lower *dM/dV* is generally preferable so that a specific gain is easier to set and maintain. One case where a large *dM/dV* is preferred is for APD's operating in Geiger mode [31]. Unlike the SIM, a Geiger mode APD operates at large enough voltages to place it in an unstable regime well above the breakdown voltage. In this regime, a single photon will set off a significant avalanche process producing currents in the mA range. This large *dM/dV* is desired and necessary to produce this effect. In the SIM, a large *dM/dV* indicates that only a small fraction of the injected electrons are experiencing avalanche events while the larger portion of electrons experience little gain. For this reason, when the SIM operates with poor efficiency, *dM/dV* must be large to produce appreciable gain levels.

The fraction of electrons that enter the optimal electric field region (η in Equation 7.1) is based upon the depth of the high field region with respect to the surface of the device. This is because the fraction of electrons which contribute to the gain of the device (n_{amp}) is dependent on the number of electrons which diffuse to a depth into the device where there are significant electric fields. The vertical distance into the device that the carriers are able to diffuse is based upon:

$$\frac{d\Delta n}{dt}\Big|_x = -D_n \frac{d\Delta n}{dx}, \qquad \textbf{7.7}$$

where Δn is the change in carrier concentration and D_n is the diffusion coefficient for the material [99]. As shown by the equation, the number of carriers that diffuse a certain distance into the bulk is proportional to the concentration gradient at a given point and inversely

proportional to the distance [100]. The number of carriers that diffuse into the bulk a distance *d* changes dramatically for a constant diffusion time. Considering the length of the depletion region (~10μm) and the saturation velocity of electrons in silicon (1.05x10^{7}cm/s), it only takes approximately 100ps for injected carriers to travel the length of the depletion region in the SIM [101]. In this time there would be few carriers that diffuse to a distance of d=0.5μm into the bulk. To illustrate this concept, simulations with well depths (D_{well} from Figure 7.1) of D_{well}=0.5μm, 1.5μm, and 3.0μm were performed using ATLAS. Figure 7.5 contains the results of these simulations showing the gain (M) versus voltage curve as D_{well} varied. As shown earlier, *dM/dV* is an excellent indicator of impact ionization efficiency, with a larger *dM/dV* indicating a less efficient geometry. The gain plot for D_{well}=0.5μm is more gradual than that of D_{well}=3.0μm indicating a more efficient geometry. It might be assumed by these simulations that the problem of poor impact ionization efficiency can be resolved by simply utilizing very

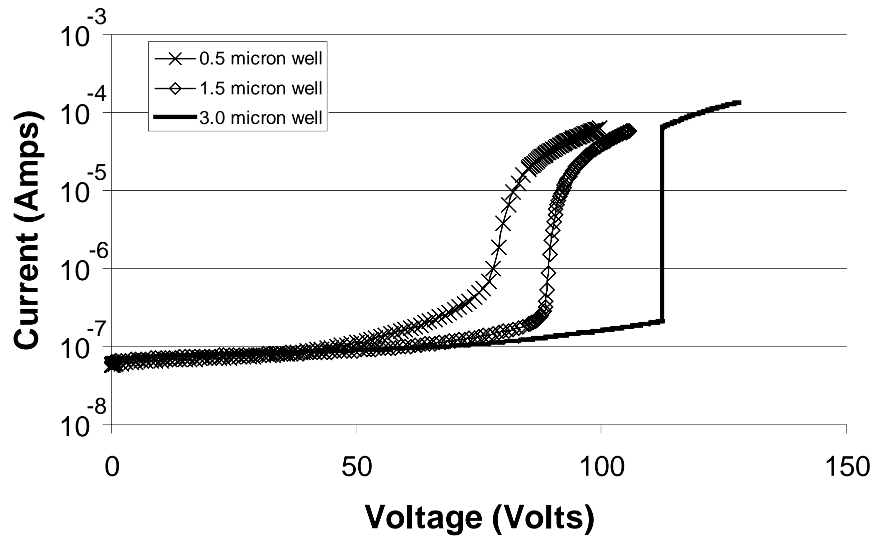


Figure 7.5: Simulations showing the current versus voltage curves for varying well depths at the output node. This illustrates the electric field distribution is dependent upon the doping profile and well depths and is effecting the ionization efficiency as electrons pass from input to output without passing through the optimal field region.

shallow wells at the output electrode. But achieving wells sufficiently shallow to offset the poor ionization efficiency can be difficult, especially when applying the high voltage to the wells required for operation in the avalanche gain regime.

7.3 Buried Oxide to Increase Ionization Efficiency

The importance of increasing impact ionization efficiency led to the development of the buried oxide SIM design. Through simulation, attempts were made to direct the carriers through the optimal gain region. This was done by inserting a region of oxide between the n+ well of the injection node and that of the output node. This structure, as produced by Silvaco, is illustrated in Figure 7.6. As previously stated, the buried oxide was introduced to direct injected carriers into the maximum field region, optimizing the impact ionization efficiency. As indicated by comparing Figure 7.6 and Figure 7.1, the introduction of oxide leaves the electric fields virtually unchanged. The maximum electric field region remains at the lower left corner of the n+ well at the output electrode regardless of the depth of the well (D_{well}) or the oxide depth (d_{ox}). With this knowledge we now compare the carrier path of the SIM with the oxide introduce in the channel. This is seen in Figure 7.7 [102]. Notice that the carrier path is altered significantly due to the buried oxide. From simulation we see that the carriers continue to travel along the surface of the epitaxial layer next to the oxide. There is still diffusion into the device but the majority of the carriers stay near the oxide/semiconductor interface. These results were very promising as the majority of the carriers were now passing through the optimal electric field profile. In theory this would greatly increase the impact ionization efficiency of the structure.

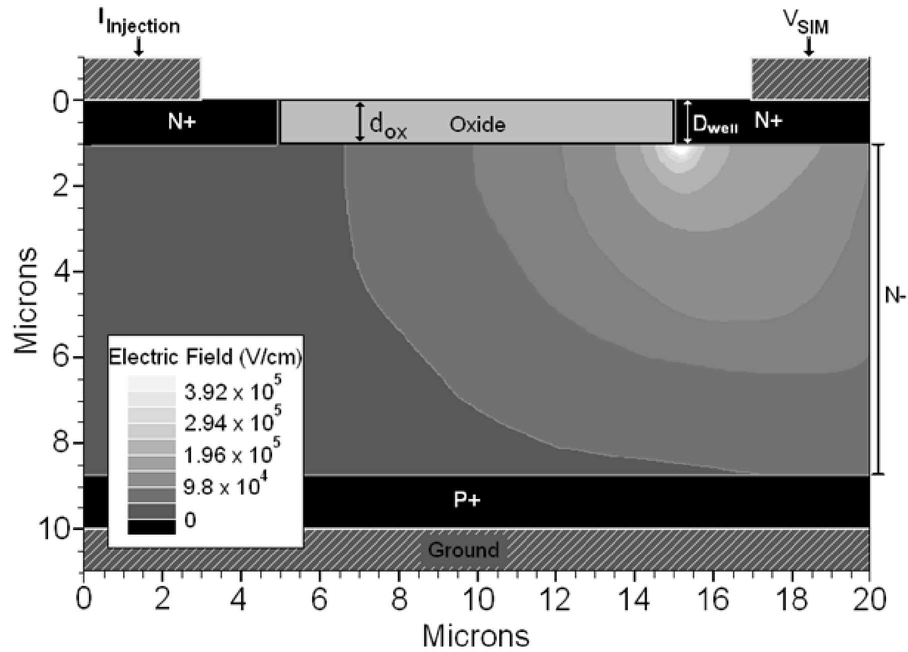


Figure 7.6: Simulation of a buried oxide SIM showing the electric field profile. Notice that the electric fields are nearly unchanged when compared to Figure 7.1. This simulation has V_{out}=50V $I_{injection}$=10µA where d_{ox}=D_{well}.

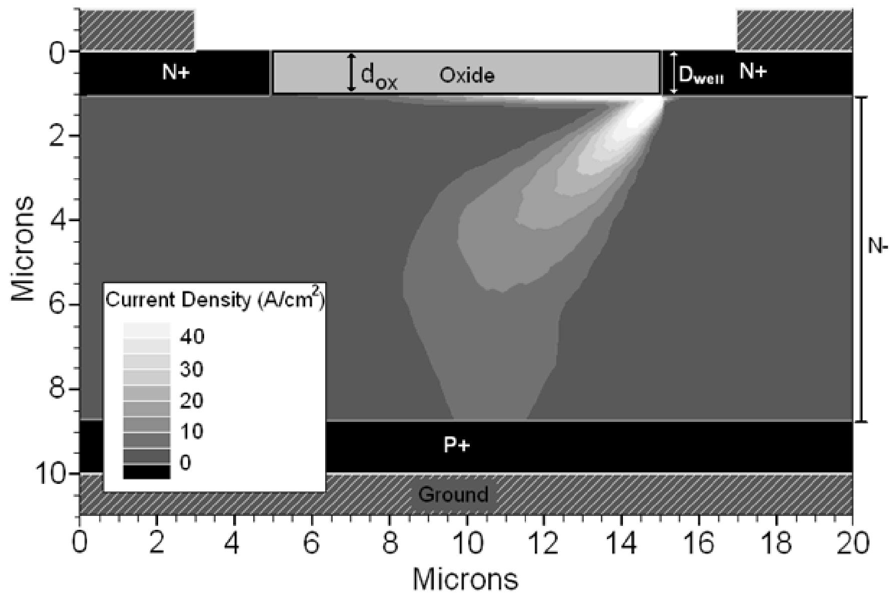


Figure 7.7: SIM cross-section showing the total current density and V_{sim}=85V and $I_{injection}$=10µA. It was necessary to increased V_{out} to obtain current levels comparable to the buried oxide SIM of Figure 7.2. This is due to the low impact ionization efficiency of the configuration of Figure 7.2. The current density shown moving diagonally down to the substrate are the holes produced by impact ionization.

In order to attempt to verify the increase in ionization efficiency due to the oxide simulations with and without this buried oxide layer were performed for comparison. Figure 7.8 illustrates the results of these simulations. The curves in the figure indicate improved impact ionization efficiency due to the introduction of buried oxide. As shown, the change in gain (M) with respect to voltage is significantly different between the cases with and without oxide. Also, notice that the voltage necessary to produce a gain of ten is approximately 60V in the buried oxide case compared to about 93V for the no oxide case due to changes in carrier path. With this information it was concluded that the introduction of oxide has made a marked improvement on the impact ionization efficiency of the SIM.

With this information simulations were run with the purpose of finding the optimal ratio between the depth of the oxide and the depth of the n+ well. To do this, various simulations for different oxide depths (d_{ox} from Figure 7.6) ranging from d_{ox}=0.0µm to d_{ox}=2.0µm were

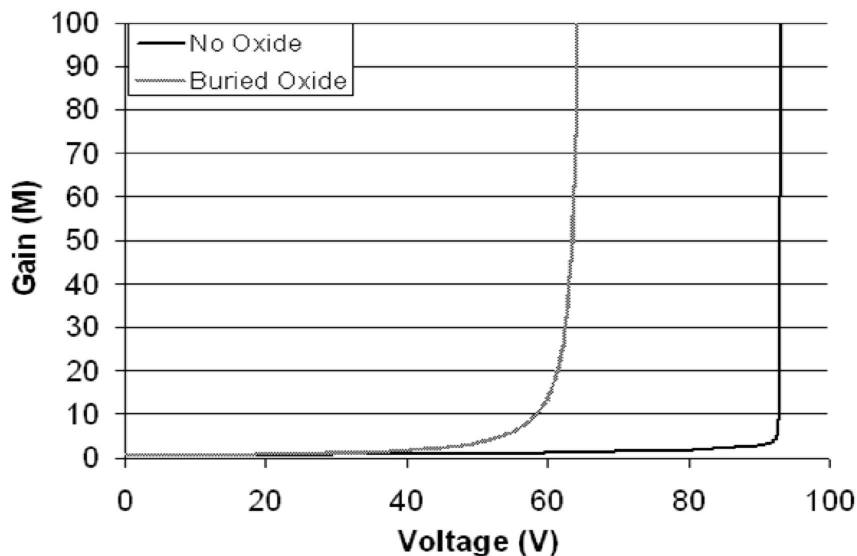


Figure 7.8: Comparison of simulations in Silvaco of an ohmic contact SIM and a buried oxide SIM with an ohmic contact injection point. They are labeled as *no oxide* and *buried oxide* respectively. Notice the gradual breakdown of the buried oxide case when compared to the no oxide. This is indicative of increased impact ionization efficiency.

performed while maintaining D_{well}=1μm constant. Using the output current-voltage curve, a gain (M) versus voltage curve was produced for each oxide depth. Then selecting a gain of M=50 for all oxide depths, the derivative of gain (M) with respect to voltage (dM/dV) was taken. These results are shown in Figure 7.9. The results indicate that dM/dV, at an achievable gain, is very dependent upon the ratio of d_{ox}/D_{well}. This, in turn, shows the variation in impact ionization efficiency for the different ratios (with lower dM/dV indicating better efficiency). It may come as no surprise that the optimal ratio of oxide depth to n+ well depth is one to one. This is because the maximum electric field is at the same depth of the n+ well and an equal oxide depth directs the majority of electrons into this field. It is also important to point out a less obvious result: for more efficient ionization, it is better to err on the side of deeper oxide than on the side of too shallow oxide when compared to the depth of the well. This is because as the magnitude of d_{ox} surpasses that of D_{well}, there is minimal diffusion away from the high field region after the carriers have passed the oxide (as long as d_{ox} isn't significantly larger than D_{well}).

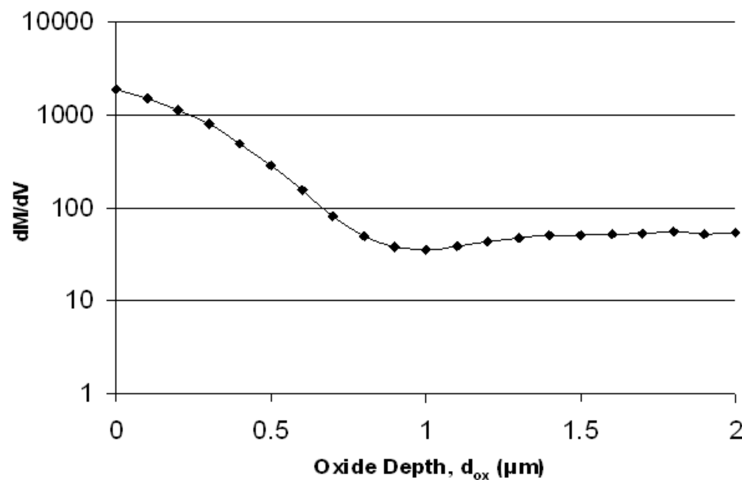


Figure 7.9: Each dot in the graph represents a separate simulation in Silvaco where the only difference from one simulation to the next was the oxide depth (shown on the x-axis). Each value was obtained by taking the derivative of gain (M) with respect to voltage at a gain of M=50. The well depth at the output for all the simulations was held at a constant 1μm. Notice that 1μm oxide depth is the low dM/dV point.

These results were very promising and indicated that a structural change to the SIM could vastly improve the gain controllability and the ionization efficiency. This makes achieving higher gains more possible and decreases the excess noise factor in the devices during high gain operation.

7.4 Fabrication

The fabrication of the buried oxide devices takes place on the same P+ substrates as previous versions of the SIM. There are several different ways to fabricate the buried oxide SIM devices and this process has evolved over time. There are two different versions of the fabrication process which are nearly identical. One uses wet etch techniques and the other utilized dry etch. This section focuses on the wet etch process. The complete fabrication process can be found in B for both the wet etch and dry etch techniques.

Figure 7.10 illustrates the fabrication process for the buried oxide SIM. These were fabricated on a p-type silicon epitaxial wafer with an EPI layer of approximately 3 Ω-cm resistivity and 10μm thick with a substrate of 0.01 Ω-cm in resistivity. 100 nm of oxide is thermally grown on the wafer. The layer of oxide is patterned with resist and the field oxide is etched in buffered HF so that features of varying widths (30–360μm) and spacing (3–9μm) remain.

These features mask of the tops of the pedestals as the wafer is etched in hot KOH to form silicon pedestals of roughly 3 um. After etching a 500 nm thick layer of thermal oxide is grown on the wafer. This passivates the surface, acts as a doping mask, and creates the buried oxide feature between the N wells. SU-8 3005 is then spun on the wafers. The minimum thickness of SU-8 3005 is about 3 um. This corresponds well to the top surface of the pedestals.

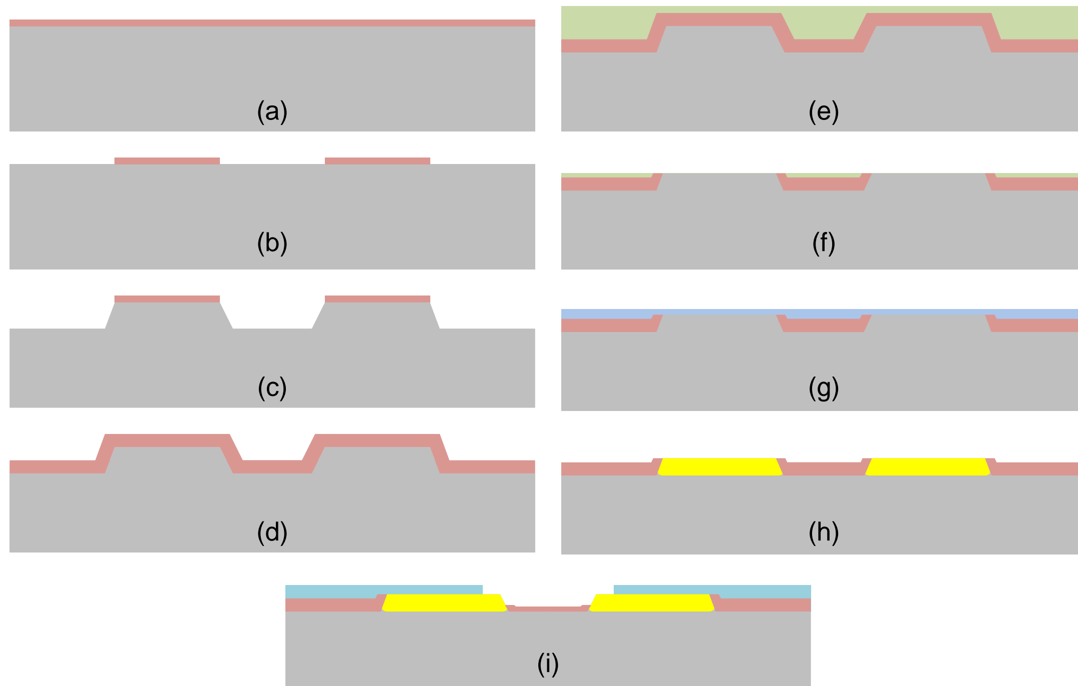


Figure 7.10: Fabrication of the second generation Buried Oxide SIM begins by growing a) thermal oxide to act as a KOH etch mask. b) The thermal oxide is then patterned to form the etch mask and c) KOH etched to form pedestals. d) The etch mask is then removed and a new thermal oxide is grown over the entire wafer. e) SU8 is then placed over the wafer to protect the field oxide during polishing and f) the oxide on the top of the pedestal is removed via Chemical Mechanical Polishing (CMP). After cleaning off the SU8, g) spin on glass is spun on over the entire wafer. h) The exposed silicon pedestal tops are then doped with phosphorus spin on glass with the thermal oxide acting as a diffusion barrier in the diffusion process. After diffusion, the excess spin on glass is removed. i) Lastly the spin on glass is removed and metal is patterned onto the doped pedestals.

The conformal quality of SU-8 provides a flat field area with minimal beading on the tops of the pedestals. Using CMP, the SU-8 beading on the pedestals tops is easily removed along with the oxide layer on the pedestal top. After polishing, the SU-8 is removed and Honeywell P-8545 phosphorus doped spin on glass is applied and diffused into the wafer. Following diffusion, the excess glass is removed in buffered HF and front and back sides of the wafer are metalized. After metallization, wafers were annealed for 30 minutes at 450° C in a forming gas environment. A detailed recipe of the second generation buried oxide SIM is found in Appendix A.

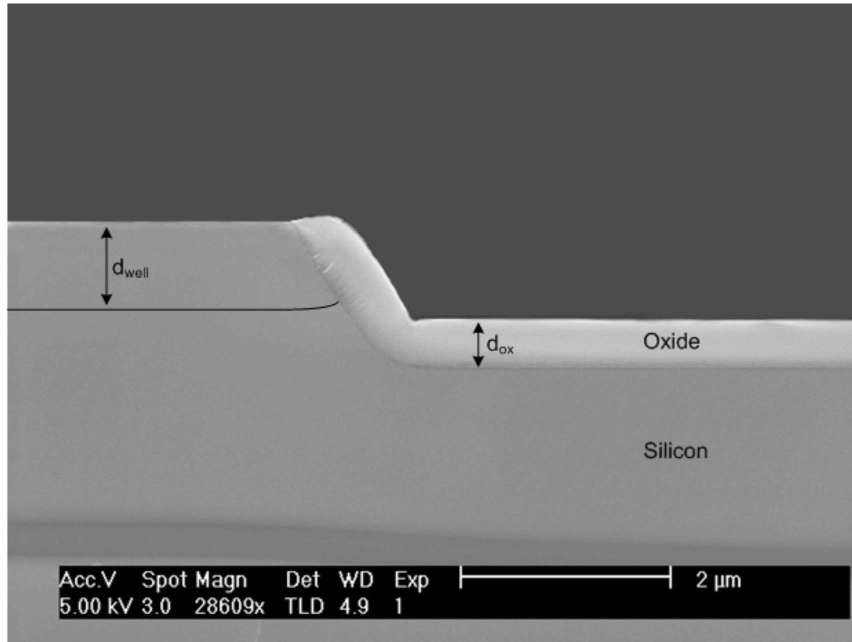


Figure 7.11: Scanning electron microscope (SEM) image of a cross-section of the buried oxide SIM. Notice the characteristic etch angle due to wet etching silicon. Labeled is the theoretical well dept on one of the pedestals.

After completing the fabrication of the buried oxide SIM, the wafer was cleaved in order to be observed using a scanning electron microscope (SEM). The SEM image in Figure 7.11 shows the buried oxide layer that extends between the input electrode and the output electrode.

7.5 Gain Measurements

After fabrication of the buried oxide SIMs, the wafers were diced up. The devices were wire bonded and packed in TO-5 cans for testing. A silicon photodiode (Perkin Elmer VTP1012) was wired to the injection node of both planar and buried oxide SIMs and illuminated by a laser emitting 870 nm light (Fiberlink XA-1000A-1). The photodiode, substrate and output were all biased using an Agilent 4156 Semiconductor Parameter Analyzer. The setup is shown in

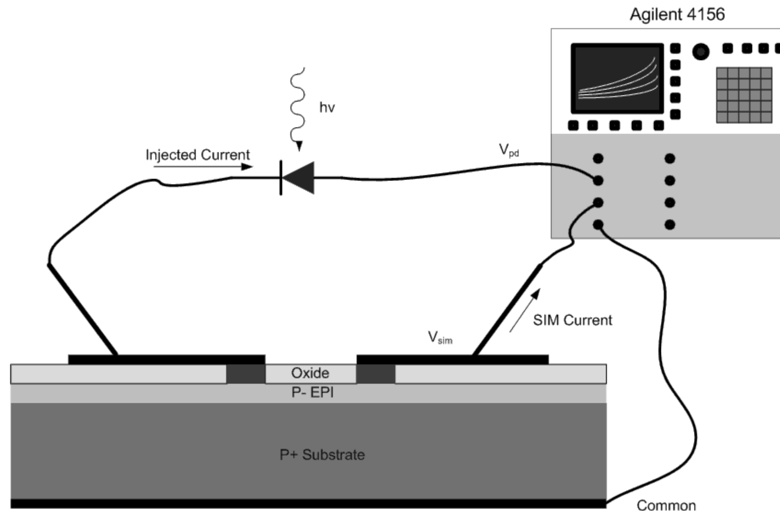


Figure 7.12: Illustration showing the testing setup of the SIM devices. A cross-section of the SIM is shown with the photodiode and the parameter analyzer.

Figure 7.12 [70] and compared to ATLAS simulations for equivalent devices. The IV curves show typical operation of both the oxide and no oxide SIM cases. In both instances, the electric fields produced by lower voltages are too weak for impact ionization to occur. As the bias increases, the depletion region gets larger and current begins to easily flow between the injection and output electrodes but without gain. Once the bias is sufficiently strong, impact ionization begins. In the planar SIM devices, there are few electrons entering the maximum electric field region [13]. Because of this, there is no noticeable gain until the amplification of these few electrons becomes very large. This occurs with only a small change in voltage resulting in a marked jump in current. On the other hand, the SIM devices with oxide show a much more gradual increase in gain. Figure 7.13 shows the measured dark current levels for both cases with and without oxide. Devices with oxide have shown a better signal to noise level and lower relative dark currents. Tested devices have stable dark current levels and show no indication of device damage or increased noise due to undesired corner or edge breakdown effects.

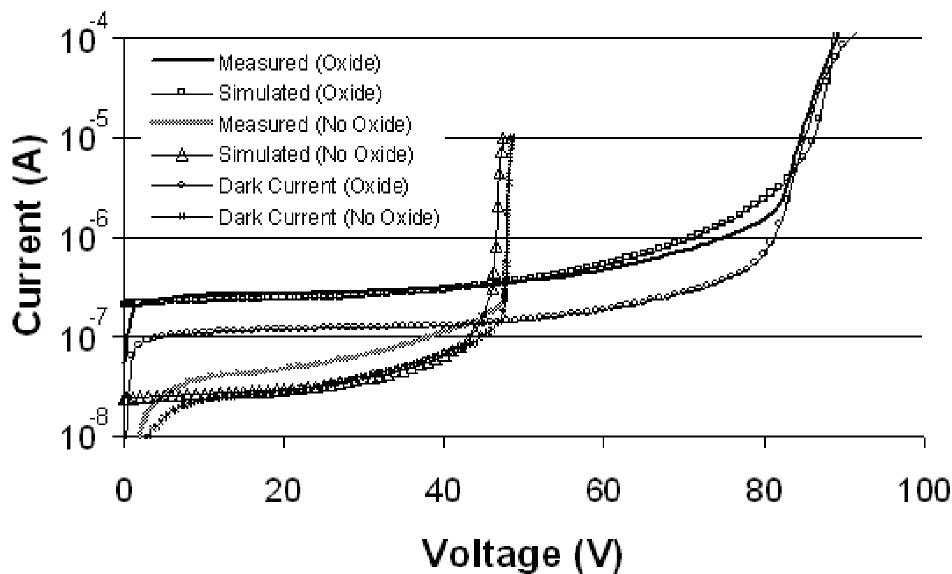


Figure 7.13: Comparison of the measured and simulated IV curves of the SIM. The buried oxide case has an injected current of $I_{injection}$=500nA. For the planar (no oxide) case $I_{jection}$=50nA.

It should be noted in Figure 7.13, that the case without oxide breaks down sooner than the case with oxide. This is because the devices used to obtain these measurements were of different sizes. In this case, the distance between the injection and output electrodes is 5µm for the planar device as opposed to 9µm for the buried oxide device.

7.6 Conclusions

Well designed impact ionization based amplifiers have gain curves that increase gradually with voltage. This attribute is desirable because it allows gain levels to be more easily achieved, controlled, and maintained. Such gain curves were not easily obtained in previous versions of the SIM because the amplification occurred along the surface as opposed to traditional impact ionization based devices that amplify vertically through the bulk of a semiconductor. Surface amplification leads to difficulties in obtaining high impact ionization efficiencies due to the path of the carriers through depletion regions. In previous versions of the

SIM, the majority of injected carriers did not pass through the optimal amplification region due to the electric field distribution. By introducing an oxide layer between the injection and output electrodes, injected electrons can be directed properly into the optimal electric field. Optimizing the ratio of oxide thickness to well depth (d_{ox} to D_{well}) allows for maximum impact ionization efficiencies and greatly improves the gain characteristics for the SIM. Fabricated devices show an improvement in the *dM/dV* of the buried oxide SIM design. For a constant gain of M=18 the buried oxide SIM was measured to have a *dM/dV*=1.4 and the planar SIM measured *dM/dV*=416. This shows the significant benefits to impact ionization efficiency by introducing an insulator to direct electron paths.

8 FREQUENCY RESPONSE AND GAIN

8.1 Introduction

Speed and gain are related to each other in amplification devices. The gain-bandwidth product is very important when determining the limits of a devices capabilities and usefulness [103], [104], [105]. To become useful, an amplifier needs to operate at a certain gains and speeds for different systems. The SIM is no exception. For the SIM to be commercially viable as an amplification device it is necessary that it operate at appreciable frequencies and produce sufficient gains.

There are many factors that limit the gain and frequency of an amplifier. Some of these factors deal with the total system resistance, others deal with the electric fields and some might deal with overcoming noise. We've discussed in previous chapters some of the alterations that have been made to the SIM during its evolution in order to overcome some of these factors. This chapter will attempt to provide a comprehensive list of the main factors that have limited the gain and frequency of the SIM. It will outline the problems and list solutions that have been taken or potential solutions to the problems. Once the limiting factors have been addressed and solutions outlined, the optimal SIM structure and operational procedure will be discussed in Chapter 8.

8.2 Metal-Semiconductor Barrier Revisited

As we saw in Chapter 6, one of the first issues addressed in the evolutionary process of the SIM was the barrier resistance at the Schottky contact injection node [106]. This was discussed in detail in Section 6.3. It was noted that the solution to the problem was the introduction of an ohmic contact at the injection node with a corresponding pn junction doping profile. In an optimal SIM device this would also be incorporated for all of the benefits listed in that chapter. A good ohmic contact is necessary for the SIM to achieve high frequency responses at appreciable gain levels [107].

8.3 Space Charge

One topic that hasn't been covered much in the previous chapters, but has been on the minds of SIM research group members, is the effects of space charge in device operation. Space charge resistance is derived by starting with the one dimensional form of Poisson's equation [108] seen here:

$$\frac{d\xi}{dx} = \frac{\rho}{\varepsilon_s} . \qquad \textbf{8.1}$$

Poission's equation relates the electric field distribution with the current density. It explains how the presence of charges changes the electric field and cause a resistive effect to other changes in charge density. The first step is to put charge density into a form we are more familiar with in semiconductor physics, as in:

$$\rho = q(n - p + N_A - N_D) . \qquad \textbf{8.2}$$

In this equation, n and p are the electron and hole concentrations respectively and N_A and N_D are the doping levels of the semiconductor slab in question. When evaluating the SIM the

semiconductor being evaluated is the p- epitaxial layer. The doping levels in depleted semiconductor are negligible when compared to the injected electron concentration. Also, because the SIM operates by efficiently collecting holes at the substrate the electrons are the only carrier in question. With this in mind, Equation 8.2 reduces to

$$\rho = qn , \qquad \textbf{8.3.}$$

yielding this form of Equation 8.1

$$\frac{d\xi}{dx} = \frac{qn}{\varepsilon_s} . \qquad \textbf{8.4}$$

The next step is to integrate from the input to the output (across the entire length of the depletion region) obtaining the total charge density in this region as in

$$\int d\xi = \int_0^{W_d} \frac{qn}{\varepsilon_s} dx , \qquad \textbf{8.5}$$

where W_d is the depletion width (distance from input to output). This step produces the electric field due to the space charge resulting in

$$\xi = \frac{q(n + N_A)W_d}{\varepsilon_s} . \qquad \textbf{8.6}$$

Integrating one more time yields the potential due to space charge. As in

$$\int \xi = \int \frac{qnW_d}{\varepsilon_s} dx \rightarrow -V = \frac{qnW_d^2}{\varepsilon_s} . \qquad \textbf{8.7}$$

The relation between electron concentration and current is shown here

$$n = \frac{I}{Av_s q}, \qquad \textbf{8.8}$$

where A is the cross-sectional area of the depletion region, I is current and v_s is the saturation velocity in the material (silicon). Combining this equation with Equation 8.7, yields

$$V = \frac{IW_d^{\ 2}}{2Av_s\varepsilon_s}. \qquad \textbf{8.9}$$

Now, using ohms law (V=IR) we extract space charge resistance to be

$$R_{sc} = \frac{W_d^{\ 2}}{2Av_s\varepsilon_s}. \qquad \textbf{8.10}$$

With this relationship, we see that the space charge resistance seen by an injected signal is proportional to the square of the distance from input to output and inversely proportional to the cross-sectional area of the depletion region. This means the best way to reduce the effects of space charge is to change the geometries of the device. Given our fabrication limits in the cleanroom, reducing the distance from the input to the output (W_d) isn't a viable option. Design alterations were made to change the cross-sectional area of the channel.

In order to increase the cross-sectional area of the depletion region (A in Equation 8.10) there needs to be an elongation of the wells and metallization in the SIM fabrication design. Figure 8.1a shows a top view of the Schottky contact SIM. Figure 8.1b illustrates the top view of the doping and metallization of an ohmic contact SIM with elongated doping profile and metallization. This configuration can greatly increases the cross-sectional area (A) of the SIM channel as indicated by Figure 8.2 [70]. By increasing this area there is a corresponding decrease in the effects of space charge resistance. To overcome the upper limit frequency due to space charge it becomes necessary to change the geometries to reduce the resistance values.

These geometric changes were put into practice and images taken through a microscope of the two versions of the SIM seen in Figure 8.1 are shown in Figure 8.3. The metallization for an original version of the SIM and the model altered to reduce space charge resistance.

The cross sectional area in the SIM device is estimated by the width of the depletion region in the direction perpendicular to the path between the metal-semiconductor contact and

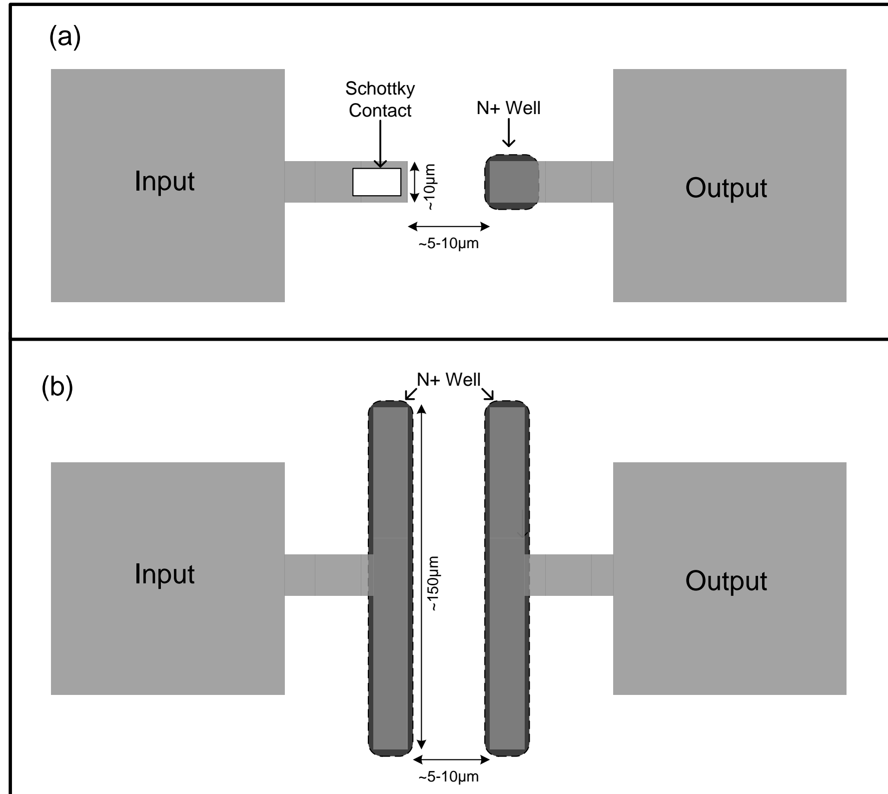


Figure 8.1: Top views of SIM devices showing the doping and metallization. a) Shows the traditional Schottky contact SIM. b) Shows an ohmic contact SIM with elongated wells and metallization at the input and output contact points. This configuration increases the cross-sectional area of the depletion region thus reducing the space charge resistance (R_{sc}).

the electron collector, multiplied by the depth of the depletion region. With that being said, we can get an estimated value of the space charge resistance for different geometries. The magnitude of the resistance term R_{sc} is verified by matching measured values with theoretical equations. Substitution of actual device parameters into Equation 8.10 yields a theoretical representation R_{sc}.

With a value for R_{sc}, it can be taken a step further by taking into consideration the barrier resistance of the Schottky contact SIM to get an idea of where frequency responses are being hit. This is done by taking into consideration Equation 6.9 to get a value for $R_{barrier}$ as a function of input current. Device parameters are also taken into consideration. Figure 8.4 shows the results

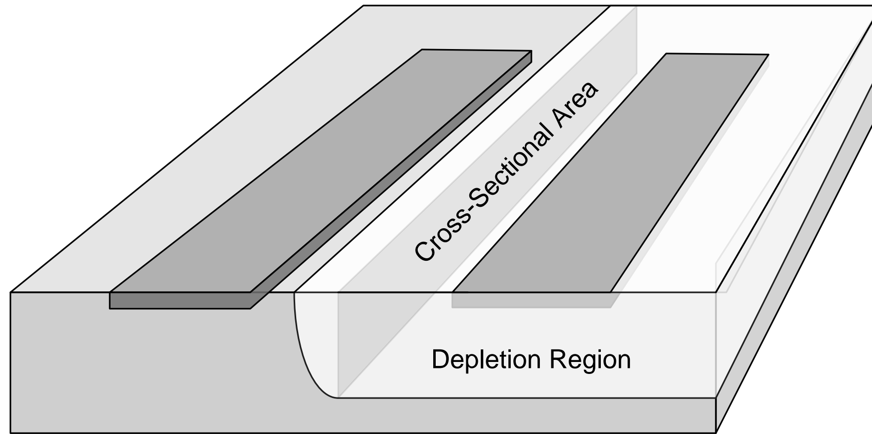
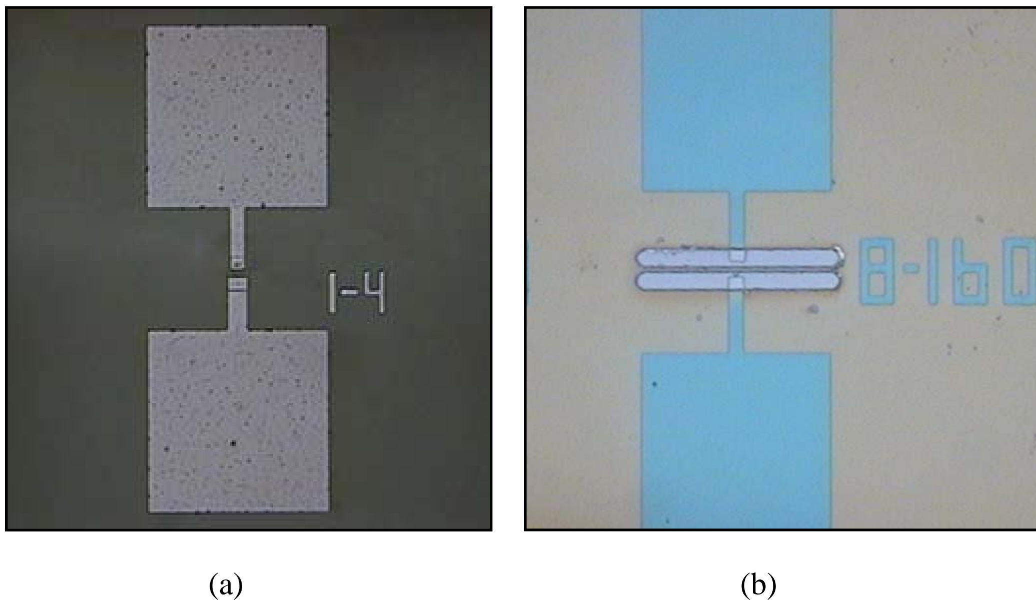


Figure 8.2: Theoretical SIM with increased doping and metallization in order to increase the cross-sectional area of the depletion region between the input and output during operation.

of this setup as a function of the input current. The values were obtained for d = 4 um, P type doping at $3x10^{15}$ cm^{-3}, and the metal-semiconductor barrier at 0.45eV [109] (nickel silicide on P type silicon). A dark current (I_d) of approximately 1nA was used to represent the real device in. When examining the curve shown in Figure 8.4[1], this dark current term dominates the



(a) (b)

Figure 8.3: Images taken through a microscope of the top view of the metallization of the SIM. a) Shows the SIM with a older version of the contacts. b) Shows the elongated doping and metallization of a SIM designed to reduce space charge resistance.

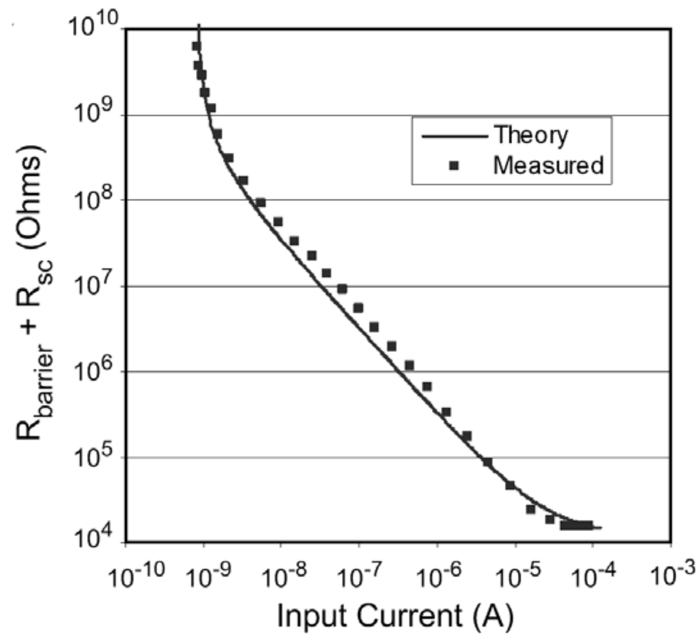


Figure 8.4: $R_{\text{barrier}} + R_{\text{sc}}$ resistance versus input current between the metal semiconductor and electron collector. The theoretical curve is calculated using Equation 8.10 and Equation 6.9 assuming a SIM device made using a *p*-type epitaxial layer, Schottky injection contact, and a spacing d equal to 4µm. Measured values correspond to fabricated SIM devices with those parameters.

resistance curve at low currents. At high currents, space-charge resistance dominates as $R_{barrier}$ drops below R_{sc}.

Measured $R_{barrier} + R_{sc}$ for the device with the same parameters are also shown in the figure to verify theory. Measurements were made using an HP/Agilent 4156 with a grounded connection to the metal-semiconductor contact while the contact to the electron collector is swept in voltage. The derivative of the measured current versus swept voltage is then used to calculate the total resistance for a given current. The calculated and measured values shown match very closely confirming that thermionic barrier emission and space charge are the dominant resistance effects in this particular SIM.

The ohmic contact design was made in part to eliminate (or at least reduce) the barrier resistance. Even with the Ohmic contact design there would still be a significant effect by the

space charge resistance at appreciable current levels. This does support the need for the ohmic contact but it also makes it obvious that a change in geometry to lower the space charge resistance is also necessary in the design of an optimal SIM device. Circular geometries have also been explored with favorable results. These designs will be discussed in detail in the next section due to another advantage that they hold.

8.4 Electric Field Distribution

For devices that are based on impact ionization, the electric field distribution is of the upmost importance to achieve consistent and controllable gains. In the most common device that utilizes impact ionization for gain, the APD, the electric field profile is established vertically through the bulk of the material. As we learned in the previous chapters, the electric field in the SIM is established in the vertical and horizontal directions due to the three terminals of the device as seen in Figure 6.11. Having three terminals greatly complicates the distribution of the electric field and turns its analysis into a non-trivial matter.

In SIM operation, the horizontal field gets locked down to an extent because of the floating voltage at the input node (this was briefly discussed in Chapters 5-6 and will be discussed at length in Section 8.5 of this chapter). The horizontal field is built up between the output node and the hole sink (grounded substrate). Because of this, it is more difficult to achieve a uniform electric field profile through the high field region. This is compounded by the fact that in the SIM the doping isn't planar across the surface and substrate of the device like in a typical APD but is a well in much the same way as a BJT or MOSFET.

Because of the differences in the electric field profile between traditional impact ionization based devices and the SIM, more care must be taken into doping depths profile when analyzing the electric field distribution.

Poor electric field distribution is something that many early versions of the SIM suffered from. The design of the buried oxide SIM was researched to handle this specific problem. The buried oxide SIM was discussed fully in Chapter 7 but some of the effects that alter the electric field distribution were not fully covered. At this time, more discussion and attention will be given to the electric field distribution and further alterations that could yield even higher impact ionization efficiencies in future SIM devices. We begin this by studying edge effects in for junctions in semiconductor devices.

8.4.1 Edge Effects

Non-uniform electric field distribution is often a result of edge effects [110]. Edge effects are manifest as changes in the electric field profile as a result of different junctions and geometries. Radius of curvature, in conductors or doped regions, plays a large role in the electric field profile through a region [111], [112]. For discussion purposes, let us assume there is a highly doped N+ well in a slab of p type semiconductor as indicated by Figure 8.5a. Notice that the doped well can be broken down into several different geometrical surfaces at the junction. Figure 8.5b illustrates the bottom of the well being approximated as a planar surface between at the pn junction. Figure 8.5c shows the area that can be approximated as a cylindrical surface at the junction. Finally, Figure 8.5d shows the spherical portion of the well. As far as electric field distribution goes, it matters a great deal what the geometry of the junction surface is and how abrupt the junction is. This can be seen by the theoretical current versus voltage curve of Figure 8.7 which illustrates the breakdown characteristics of a pn junction of the different geometry

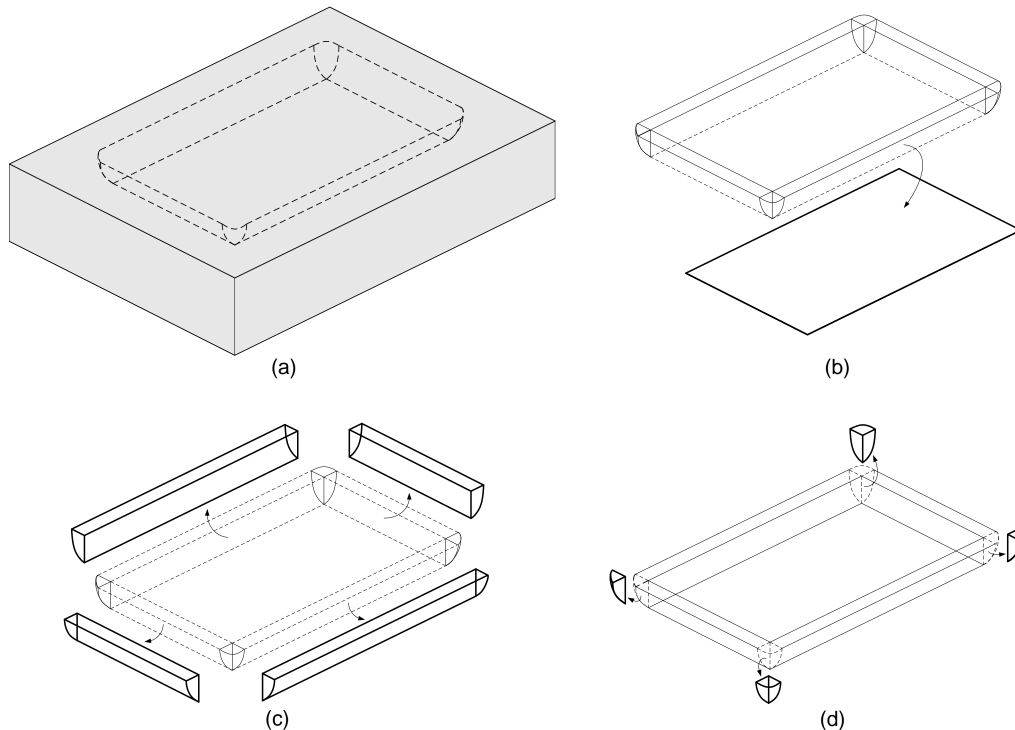


Figure 8.5: a) Highly doped n type well in a p type semiconductor. The well can be broken down into (b) the planar bottom of the doping distribution, (c) the cylindrical sides of the distribution and (d) the spherical corners.

types shown in Figure 8.5. The electric fields at the spherical portions of a junction will be higher than those at the cylindrical or planar potions [113]. Because of this, the first part of the junction to breakdown will be the spherical followed by the cylindrical and finally the planar (all other factors considered equal).

How much of difference there is between the three types of junctions depends upon the radius of curvature of the junctions. Figure 8.6 [70] labels the junction of Figure 8.5a. Quantitative data to back up this theory can be found in Figure 8.7 is shown in Figure 8.8 [70]. Figure 8.8 shows data for a one-sided, abrupt junction in silicon. The various curves are labeled for different types of junctions (spherical, cylindrical and planar) and also for varying radii of

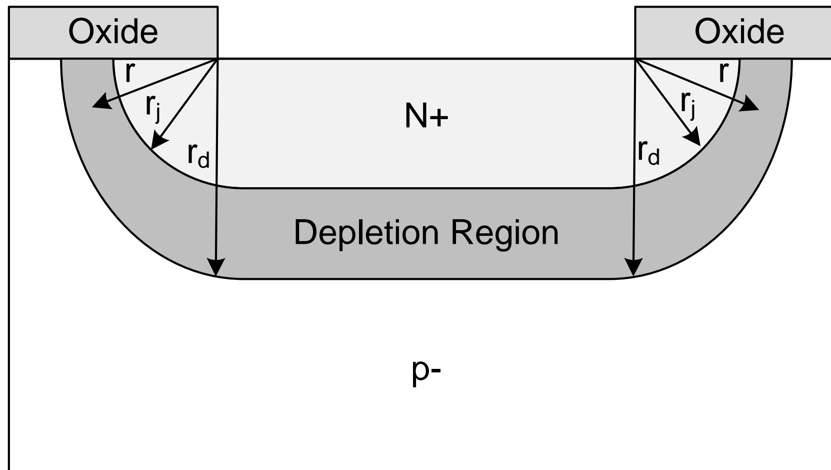


Figure 8.6: Planar cut of Figure 8.5a. This figure labels the different curvatures of radius for the junction surface.

curvature (0.1µm, 1µm, 10µm and the planar case of ∞). The data supports the theory given by Figure 8.7 and observation from Silvaco simulations.

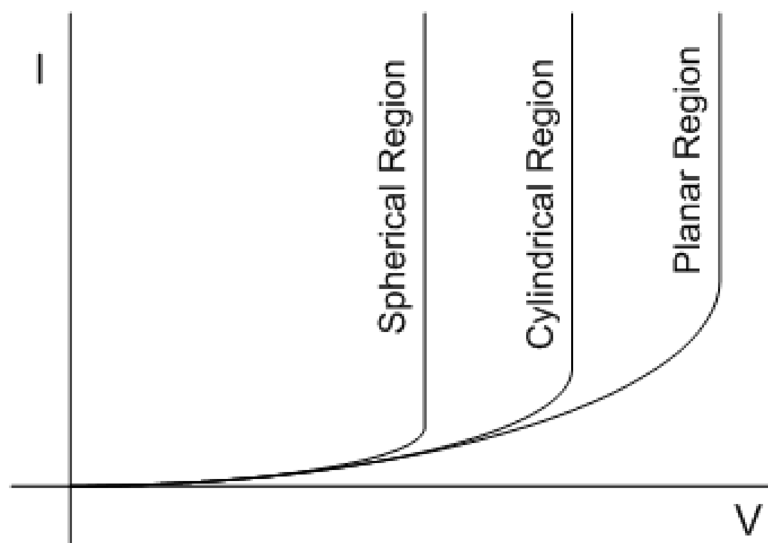


Figure 8.7: Theoretical current versus voltage curves illustrating the breakdown voltages for different types of junctions. The spherical region of a junction will breakdown sooner because the electric field is higher there. The geometry and abruptness of a junction are very important when analyzing the breakdown characteristics and the electric field distribution.

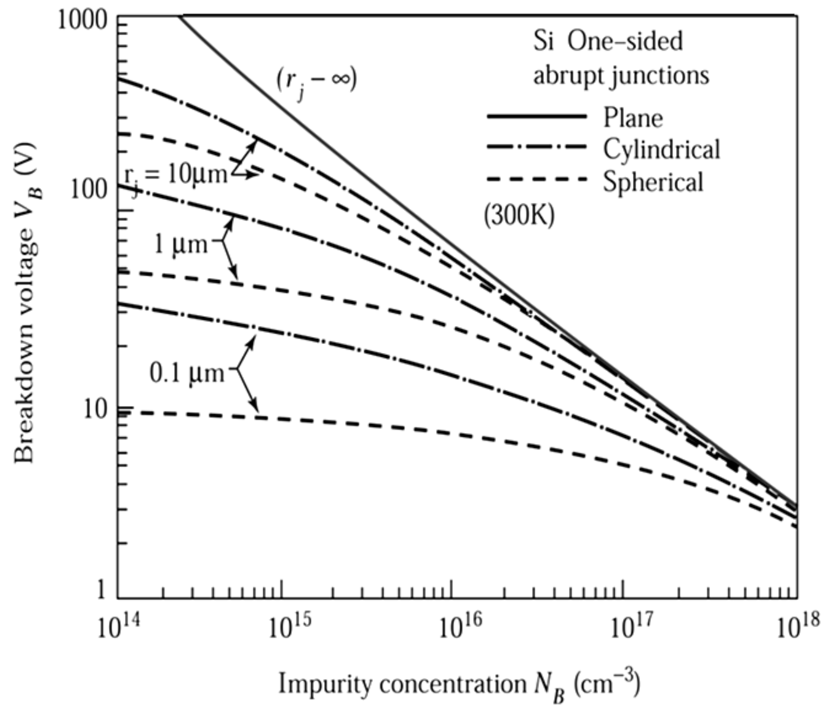


Figure 8.8: Shows the breakdown voltages for different doping levels of a one-sided abrupt junction in a silicon substrate. The different breakdown levels are for the different radii of curvature (r_j) as defined by Figure 8.6.

We can recall from the previous chapter that the buried oxide design was to aide in carrier path through the spherical and cylindrical portions of the n well so they travel through the optimal field. What was only briefly mentioned is the effect the doping geometry and edge effects have on the electric field. Further simulations done in Silvaco's 3D platform were done to monitor the electric field edge effects at the different types of junctions. Figure 8.9 shows a cut-plane from the 3D simulation from Silvaco of an elongated version of the ohmic contact SIM in rectangular coordinates. The cut-plane was taken at the bottom of the doped well and shows the electric field profile at this point. Notice that the corners of the well have a higher electric field than the rest of the doped area. This corresponds to a quicker breakdown point because of the higher field. It also affects the impact ionization efficiency because the paths the carrier take will depend on the kind of electric field profile the experience. Because of this, different carriers

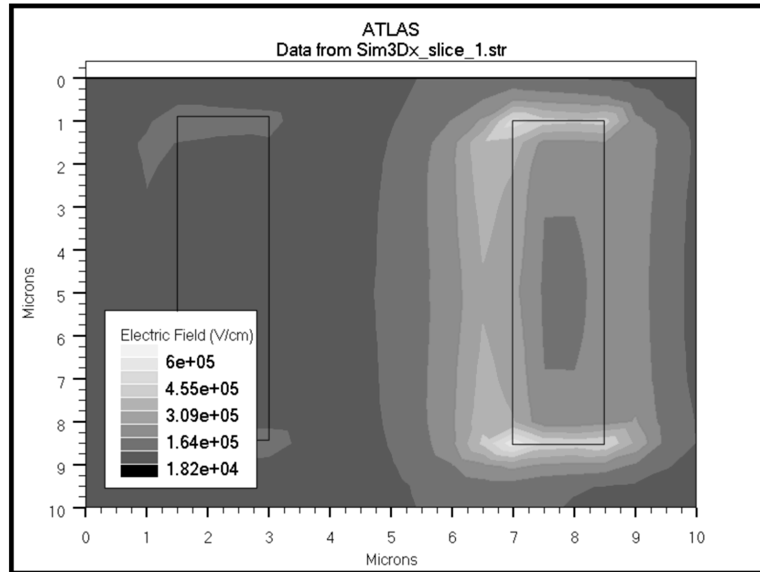


Figure 8.9: Cut-plane of a 3D simulation in Silvaco. This cut-plane is a top view of an ohmic contact SIM at about 1µm into the epitaxial layer (the depth of the N+ well at the input and output). Electric field profile is shown at this depth. The field profile reaffirms that the spherical portion of the junction has a higher field than the cylindrical and planar portions.

will participate in different amounts of ionization events. This greatly increases the excess noise of the system and the disproportionately represents the input signal as a whole.

Notice the contours in Figure 8.9. These contours illustrate regions of equal electric field strength. Ideal carrier path is in directions exactly perpendicular to these contours. This ensures that the carriers travel through the optimal gain region (highest field) as the drift from input to output. If all carriers took this path (perpendicular to the electric field contours) all carriers would experience the same field profile as they traversed the gain region. This would greatly increase the ionization efficiency of the device and make the gain profile more evenly distributed. However, the carriers traverse the device in non-ideal paths given a non-ideal electric field distribution.

The development of the circular SIM design was twofold as seen in Figure 8.10. First, it efficiently elongated the depletion region between the input and the output thus, reducing space charge resistance (see the previous section). Secondly, its design removes the spherical portion

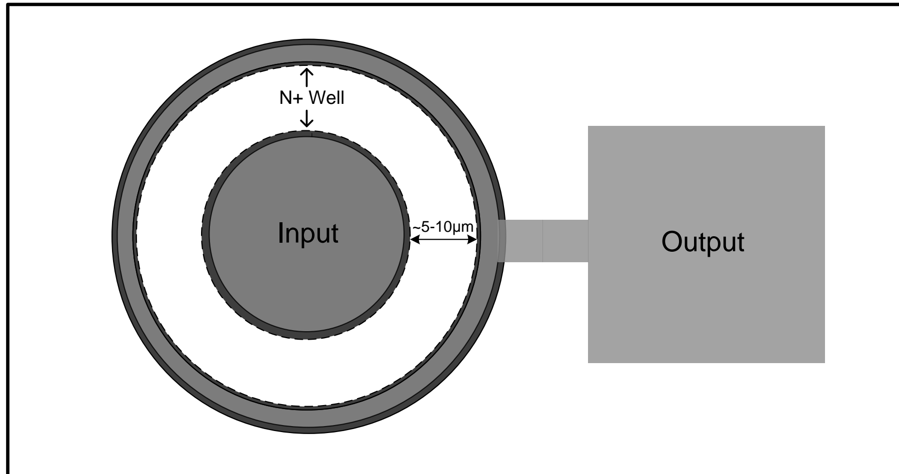


Figure 8.10: Circular SIM design with the same distance from the input well to the output well. The doping and other fabrication consideration remain the same from this design to previous version of the SIM. This designs purpose was twofold. First, it decreases the space charge resistance by increasing the depletion region cross-sectional area. Secondly, it removes any spherical portion in the doping profile making a much more uniform electric field distribution.

of the doped well making the electric field distribution more even and uniform through the gain region. This greatly reduces premature breakdown due to edge effects. Simulations for the circular SIM design shown in Figure 8.10 were performed in 3D Silvaco for comparison purposes with rectangular simulations. The resultant cut-plane for the circular design illustrating the electric field distribution is shown in Figure 8.12. This cut-plane was taken at the same depth into the epitaxial layer as that of Figure 8.9 (at the input and output node's n well). Notice from the electric field distribution that the electric field distribution is much more even throughout the distance from the input (inner ring) to the output (outer ring).

The gain versus voltage curves for the simulations represented by Figure 8.9and Figure 8.12 are shown in Figure 8.11. Notice the rate of change of gain versus voltage (*dM/dV*) for the two different structures. The circular structure is more gradual, while the rectangular structure breaks down quite abruptly. This confirms the theory that the circular structure cuts down on edge effects and increases the impact ionization efficiency of the device.

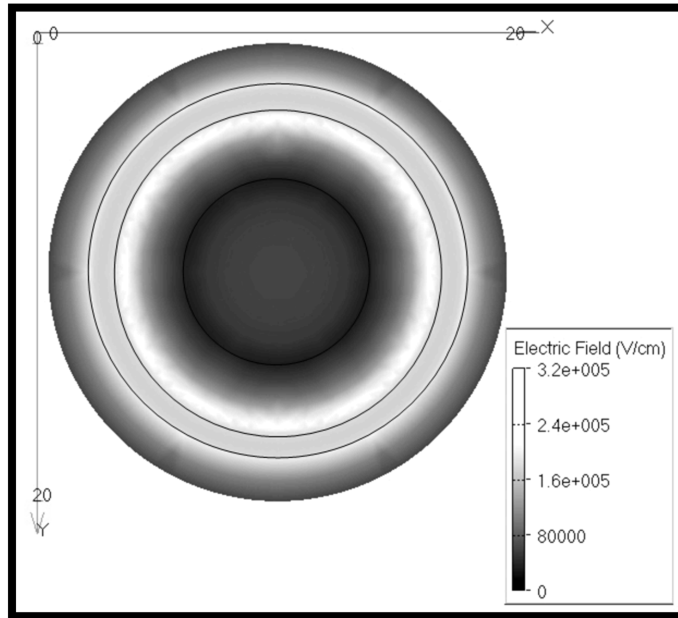


Figure 8.12: Cut-plane of a 3D simulation in Silvaco. This cut-plane is a top view of a circular ohmic contact SIM at about 1µm into the epitaxial layer (the depth of the N+ well at the input and output). Electric field profile is shown at this depth. Because the doping profile is circular there are no spherical portions of the junction. This helps maintain an even electric field distribution across the complete area of the depletion region.

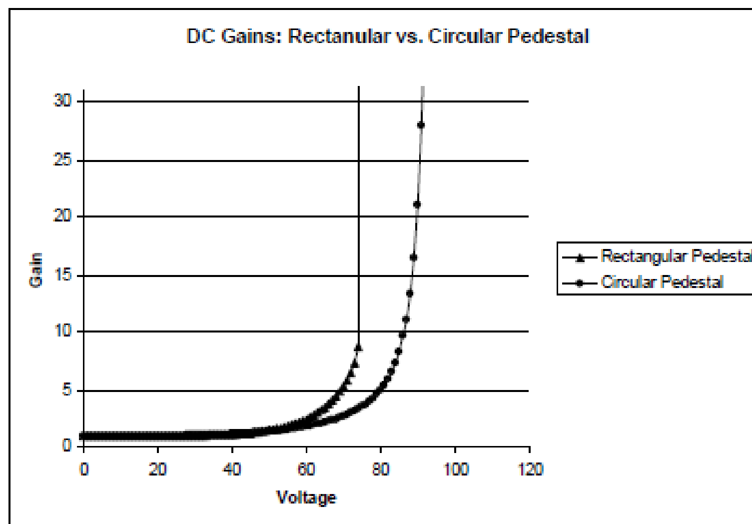


Figure 8.11: Gain versus voltage curves as done in 3D Silvaco for buried oxide rectangular and circular SIM configurations. Notice the rate of change of gain with respect to voltage. The circular geometry maintains a much more gradual and even gain profile than the rectangular suggesting a higher impact ionization efficiency and more even electric field distribution with respect to carrier path.

Because of the results seen through simulation and corresponding results observed through fabrication and testing, an optimal version of the SIM is of a circular nature in order to reduce the anomalies due to electric field edge effects. This design also facilitates the reduction in space charge resistance.

8.5 Effective Resistance from Floating Voltage

The restive effects due to the voltage shift at the input electrode were briefly discussed in previous chapters. Because there is no external voltage source to lockdown the voltage at the input node, it is free to float. In typical operation, the semiconductor between the input and the output is completely depleted. The introduction of carriers in this depletion region causes the electric field to shift. This shift in the electric field leads to a corresponding shift in the voltage at the input node. In all physical devices, a shift in voltage as a result of current is defined as a resistance.

To be more specific, the voltage shift, ΔV, at the injection point is related to the shift in the quasi Fermi energy of the depletion region. The shifts in the electron quasi Fermi level due to current level were observed in simulation and are seen in Figure 8.13. You can see from the figure that, as the input current changes, there is a corresponding shift in the electron quasi Fermi level through the depletion region. In order to gain a better understanding, a resistive value must be derived from these shifts. This is done by starting with the equation for the electron quasi Fermi level here

$$E_{Fn} = E_i + kT \ln\left(\frac{n}{n_i}\right). \qquad \textbf{8.11}$$

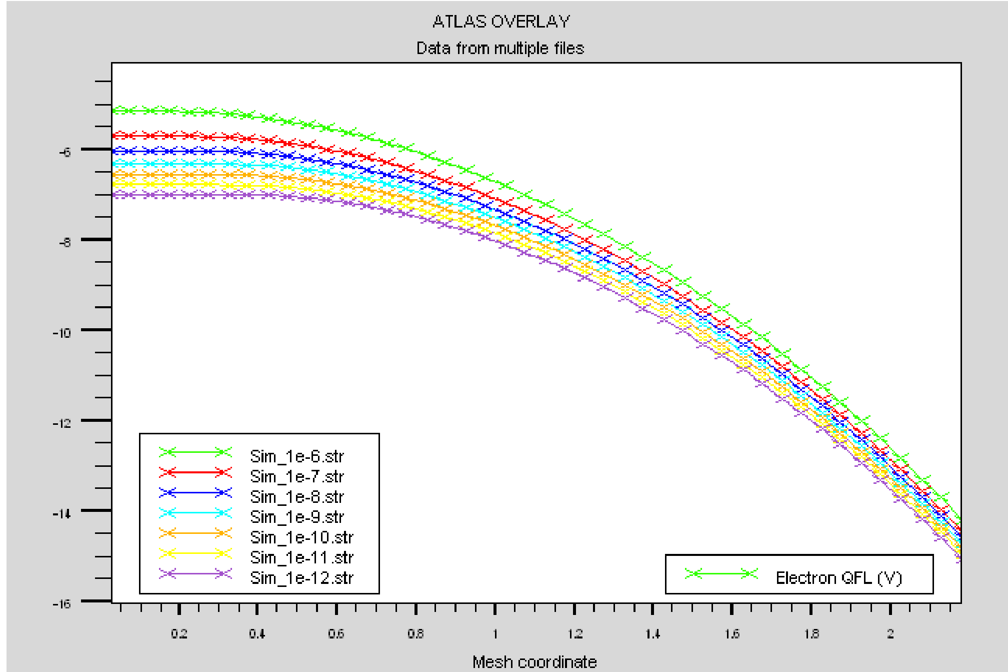


Figure 8.13: Silvaco Simulation showing the shifts in the electron quasi Fermi level for injected current levels ranging from 1pA to 1µA. This data was taken near via cutline in Silvaco near the input node of a SIM. The mesh coordinates of the x-axis are in microns going from 0 (the input pn junction) toward the output pn junction of the device.

E_i is the intrinsic Fermi level, k is Boltzmann's constant, T is temperature, n is the free electron concentration and n_i is the intrinsic electron concentration. By making a substitution for the relationship between free electron concentration and current (Equation 8.8) we get

$$E_{Fn} = E_i + kT\ln\left(\frac{n}{n_i}\right)\Bigg|_{n=\frac{I}{Avq}} \rightarrow E_{Fn} = E_i + kT\ln\left(\frac{I}{n_i Avq}\right). \quad \textbf{8.12}$$

We now take the difference between E_{Fn1} and E_{Fn2} such as in

$$E_{Fn1} - E_{Fn2} = \left[E_i + kT\ln\left(\frac{I_1}{n_i Avq}\right)\right] - \left[E_i + kT\ln\left(\frac{I_2}{n_i Avq}\right)\right]. \quad \textbf{8.13}$$

Which can be reduced to

$$E_{Fn1} - E_{Fn2} = kT\ln\left(\frac{I_1}{n_i Avq}\right) - kT\ln\left(\frac{I_2}{n_i Avq}\right). \quad \textbf{8.14}$$

Using the following logarithmic relationship $\left[\ln(A)-\ln(B)\right]=\ln\left(\frac{A}{B}\right)$, Equation 8.14 is reduced to

$$E_{Fn1}-E_{Fn2}=kT\ln\left(\frac{\left(\frac{I_1}{n_i A\nu q}\right)}{\left(\frac{I_2}{n_i A\nu q}\right)}\right). \qquad \textbf{8.15}$$

Algebraic simplification yields

$$\Delta E_{Fn}=E_{Fn1}-E_{Fn2}=kT\ln\left(\frac{I_1}{I_2}\right). \qquad \textbf{8.16}$$

The relationship between shifts in the quasi Fermi level and shifts in the voltage can be defined as by

$$\Delta V=\Delta E_F/q\,. \qquad \textbf{8.17}$$

Applying this equation to Equation 8.16 produces

$$\Delta V=\frac{kT}{q}\ln\left(\frac{I_1}{I_2}\right). \qquad \textbf{8.18}$$

This a relationship between shifts in the voltage caused by changes in the input current.

This shift in voltage at the injection node is directly related to the frequency response for the SIM. As you can see the voltage varies with input current. This results in an effective resistance existing at this node. This resistance can be calculated by taking a derivative of the relationship in 8.18 yielding

$$R_{\Delta}=\frac{dV}{dI}=\frac{kT}{qI}\,. \qquad \textbf{8.19}$$

When combined with the capacitance on the node (which is a combination of the depletion layer capacitance and the capacitance of the current source connected to the SIM), a charging time constant, τ, can be computed:

$$\tau = R_{\Delta} C_d = \frac{C_d kT}{qI}, \qquad 8.20$$

where C_d is the total capacitance at the injection. This time constant represents the time delay for a change in the SIM's output current due to a change in its input current. Because of this time constant, frequency response is directly related to the injection current level when R_{Δ} is greater than the contact resistance for the SIM. Furthermore, this is a theoretical limit not taking into consideration the effects of gain within the depletion region. The addition of carriers via gain could increase the resistive effects and push them even further above $kT/(qI)$.

8.5.1 Measured Floating Voltage

In order to obtain a better understanding of the resistive effects of the floating voltage devices were tested in such a way that the voltage at the input could be monitored for shifts over a range of current inputs. This was done using the Agilent 4156 Semiconductor Parameter Analyzer. The devices were hooked up such that a constant voltage could be applied to the output with the substrate grounded. The input current was the swept from a low current level to a high current while monitoring the voltage at the input. The data is shown in Figure 8.15. Over a current range of a little more than two orders of magnitude there is a current

Now, taking the derivative of the relationship between the voltage at the input and the injected current from Figure 8.15 yields a resistance due to this shift. Figure 8.14 shows resistance versus current curves obtained in this method. The theoretical limit of kT/I is also shown in Figure 8.14 for reference purposes. Notice that the measured resistance values are

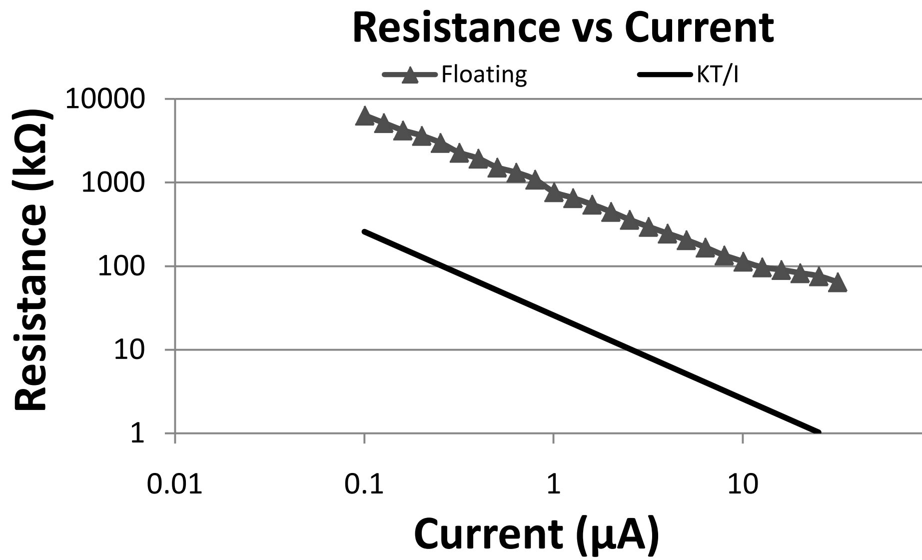


Figure 8.14: Resistance versus current curve. It shows the derivative of the voltage versus current from Figure 8.15. Also shown is the theoretical diode resistive limit of kT/I. Notice that the measured data is nearly two orders of magnitude higher than kT/I.

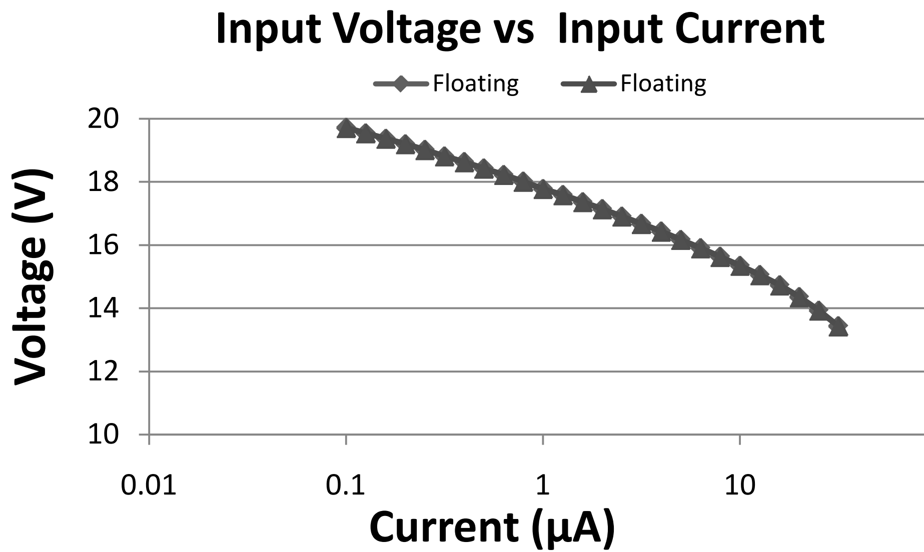


Figure 8.15: Voltage versus current curve showing the shifting at the input voltage for a range of input current levels.

nearly two orders of magnitude higher than the theoretical limits. Lowering the effects of the floating voltage brings the measured resistance down to levels comparable to the theoretical limit and increases the upper limit frequency of the device.

8.5.2 DC Offset Current

The floating voltage is more pronounced for larger shifts in the input current. This can be seen by evaluating Equation 8.18. For a change from I_1 to I_2 there is a corresponding shift in the voltage. This shift is larger for larger differences between I_1 and I_2. To help grasp this understanding more thoroughly, an example is given.

Figure 8.16 represents the voltage (floating) on the input of the SIM for various different current levels as produced by Silvaco. If we take into consideration a current pulse, ΔI, at the input there is a corresponding voltage shift ΔV. The magnitude of the voltage shift depends on the magnitude of the current shift. Assuming no injected current before a current pulse then the initial current level is equal to the dark current of the SIM (10pA) and the current pulse has a magnitude of 10nA. This means the current changes from 10pA to 1nA or a change of two orders of magnitude. The corresponding voltage shift is seen in Figure 8.16. But, if the same pulse is injected with a DC current offset of 10nA then the shift in current goes from 10nA to 11nA. This is a shift of a fraction of the original signal. The voltage shifts in both cases correspond to a pulse of 1nA but the addition of the DC offset current yields a significantly smaller voltage shift. This smaller voltage shift corresponds to a smaller R_Δ which improves the response time of the system greatly

Having a DC and AC current signal combined is not a new concept and is used in practice today [114]. The operation of many devices such as diodes and Bipolar Junction

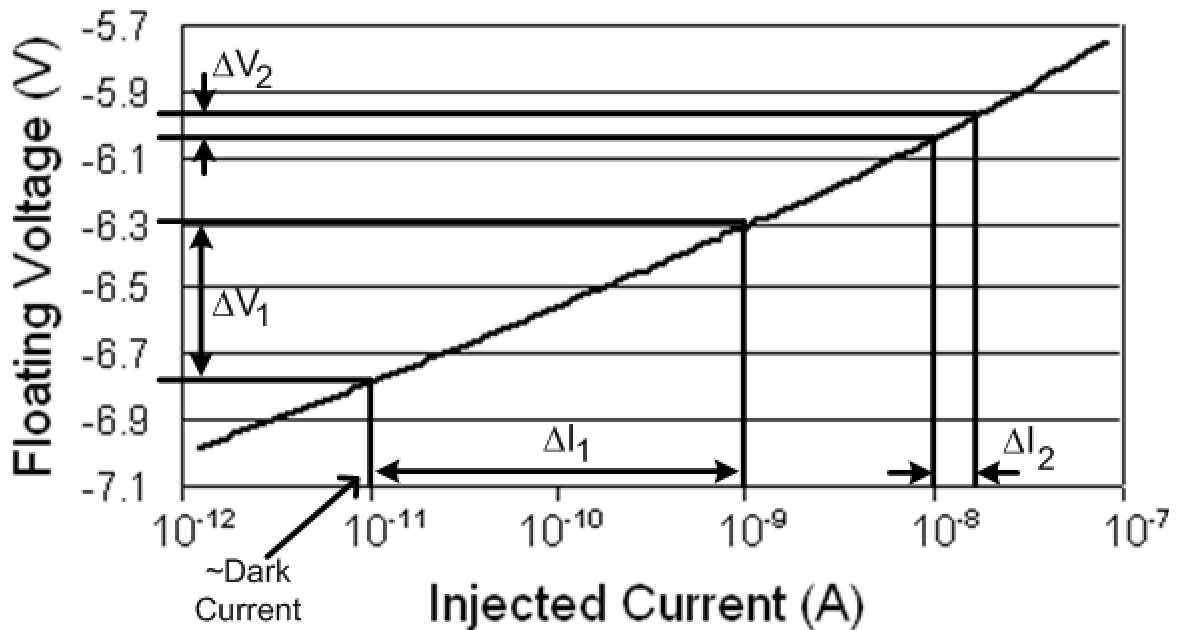


Figure 8.16: Voltage shifts at the input node for corresponding injection current levels. The relationship is linear showing a resistive nature. Voltage and current shift one correspond to a 1nA current pulse with no DC offset current. Voltage and current shift two correspond to a 1nA pulse with a 10nA DC current offset. Notice the changes in voltage from case one to two given the same pulse.

Transistors (BJT) utilize large DC current offsets to reduce these types of internal electric field shifts. A BJT doesn't have a voltage that can float due to shifts in the current but, unlike the SIM, the depletion region between the base and the collector shifts in size due to the introduction of carriers. This shift causes a time delay in the operation much in the same way the voltage shifts in the SIM introduces resistive effects. The introduction of quiescent current levels reduces these resistances for small AC signals [115]. To fully utilize the SIM, a DC offset current should be applied to reduce this resistive effect. Optimal SIM operation involves the necessary circuitry and system setup to include this DC offset.

8.5.3 Four Electrode SIM

Unwanted shifts in the voltages and electric fields of a device aren't something unique to the SIM. MOSFETs and BJTs both suffer from problems due to shifting quasi Fermi levels. In a MOSFET, for example, all of the electrodes (drain and source) are tied down by external voltages, unlike the SIM. This means that they can't shift due to changes in the carrier concentration in the depletion region. Even though the voltages can't shift, the influx of carriers changes the quasi Fermi level in the depletion region. These shifts in the quasi Fermi level correspond to changes in the barrier height at the injection node. At this point the changes in the barrier height are also affected by the output voltage. This is a short channel or high bias effect. It occurs when the depletion region at the drain reaches the depletion region at the source. When this occurs the voltage applied at the drain begins to control the current through the source by lowering the barrier at the source side pn junction. This effect is often controlled by increasing the doping in the channel to prevent the depletion region from reaching the source. A similar change in doping was considered in the SIM and simulation indicated that it would help in reducing the effects of the floating voltage. Initial fabrication and testing didn't show the same results for an unknown reason. The study of other solutions and theories have hindered further exploration of the possibility.

Like the MOSFET, the BJT has electrodes that are not capable of shifting in voltage. But the introduction of carriers causes shifts in the size of the depletion region. This isn't a problem in the SIM because the size of the depletion region is basically locked down once it fully depletes the epitaxial layer between the input and output. In the BJT, however, the depletion region between the base and the collector shifts in size due to carriers flow. As carriers diffuse through the base doping and enter the collector doped region they shift the quasi Fermi level

causing a reduction in the electric field and a corresponding change in the size of the depletion region. This shifting in the size of the depletion region creates a time delay in the operation. It is often overcome through doping changes and heterostructures. What happens in BJT is quite similar to what occurs in the SIM. The difference being that, in a BJT, the voltage is static but the depletion width varies and in the SIM the depletion width is static and the voltage varies.

It is becoming apparent that maintaining static electric fields for high concentrations of carriers is desirable in various different devices. The significance of the delay caused by the floating voltage has lead to many ideas on how to overcome or reduce its effects. One of the more promising ideas is the addition of a fourth electrode. The fourth electrode acts as a stabilizing electrode helping the input injection node maintain a more constant voltage as the output voltage as a result of input current levels shifts. The stabilizing electrode is insulated from the silicon as shown in Figure 7.17.

This stabilizing electrode resembles that of a MOSFET's gate. But it functions in a very different way. The gate of a MOSFET is used to create an inversion layer between the source and drain. This is not desirable in the SIM. An inversion layer would make gain impossible as carriers would simply travel from input to output without participating in impact ionization events. For this reason, in an n-p-n type MOSFET it is desired that the gate voltage be positive. This is the opposite in the case of the SIM.

Recall from Equation 8.18 that the changes in the voltage are dependent upon the changes in current. Figure 8.18 shows these changes in voltage for different currents at the input and for different voltages applied to the stabilizing electrode. Notice that for negative or zero bias placed on the stabilizing electrode that the shifts in voltage are greatly reduced over the same set of currents when compared to when it is positive or floating. This strongly indicates

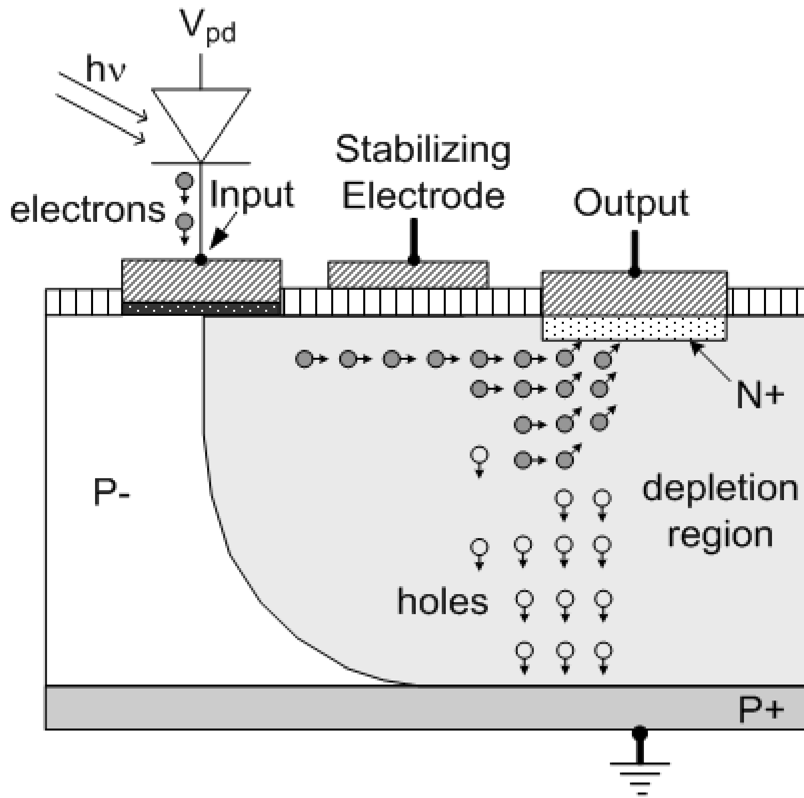


Figure 8.17: SIM with fourth stabilizing electrode between the input and output. The operation of the SIM remains unchanged but the addition of the electrode controls electric field shifts in the SIM reducing the effects of the floating voltage.

that the electrode is maintaining a more consistent quasi Fermi level and electric field near the input electrode as the currents change.

By taking the values obtained through testing and comparing them with theory and Equation 8.18, it becomes apparent that there is some factor, γ, which is dependent upon the stabilizing electrode. This factor represents the influence shifts in the current have to shifts in the voltage. The factor changes Equation 8.18 to look like this

$$\Delta V = \gamma \frac{kT}{q} \ln\left(\frac{I_1}{I_2}\right). \qquad \textbf{8.21}$$

In the case where γ=1, Equation 8.21 is equal to Equation 8.18. This is the case where the stabilizing electrode is left floating (as seen in Figure 8.14) and the SIM operates independent of

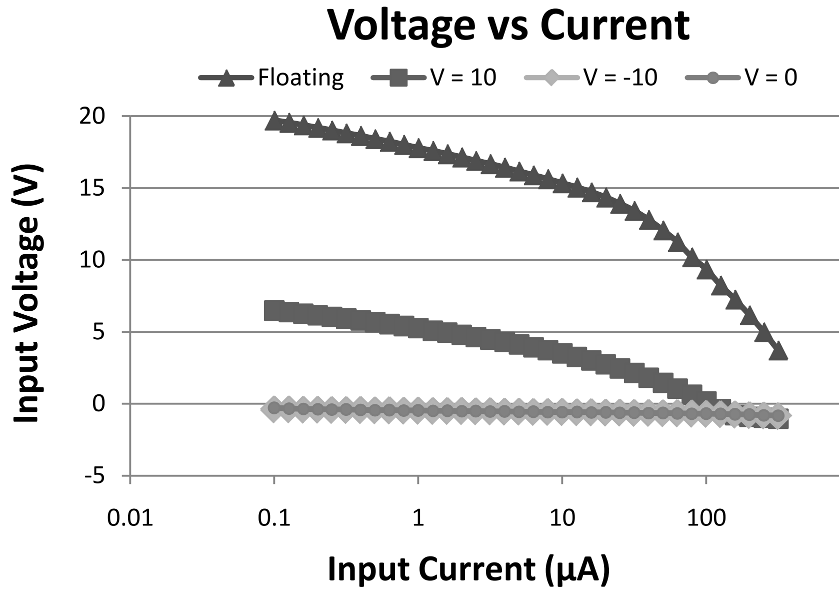


Figure 8.18: Input voltage versus input current for various voltages on the stabilizing electrode. Notice the large voltage shifts for the floating and +10V cases when compared to the 0V and -10V cases. This shows a strong indication that the stabilizing electrode helps immensely to reduce the shifts in the input voltage for shifts in the input current.

its influence. This is the worst case scenario and is the point where voltage shifts are the most dependent (maximized) on changes in the input current. As voltages are applied to the stabilizing electrode the value of γ decreases which loosens the dependence of the ΔV on current shifts (difference in I_1 and I_2) and stabilizes the input floating voltage. This, in turn, reduces the resistance due to the floating voltage from Equation 8.19.

The γ factor is dependent upon many different things and is quite complicated. This is because the influence of the stabilizing electrode depends on factors such as oxide depth, location of the electrode, voltage level on the electrode itself, voltage at the output electrode, channel length, carrier concentration, doping, etc. Due to the complexity of the factor, values for various cases were extracted numerically in order to gain a better understanding of the influence of the stabilizing electrode. This was done by taking the average voltage shift (for the floating voltage case) over the range of currents shown in Figure 8.18 and comparing it to the voltage

shifts when a voltage has been applied to the stabilizing electrode. Taking the ratio of the floating voltage values (as a normalization) to the applied voltage values yields the γ factor.

Using this method, the floating voltage case obviously yielded γ=1. The γ factor is nearly unchanged for a positive voltage on the stabilizing node (V_s=10) yielding γ=0.68. If the stabilizing voltage is tied to ground (V_s=0) the shifts in floating voltage are significantly reduced with an extracted value of $\gamma \approx 0.06$. It is slightly lower for V_s=-10, at $\gamma \approx 0.03$. Figure 8.19 shows the resistance versus current curves taking into consideration the γ factor due to the stabilizing voltage values. Notice the difference in resistance caused by proper bias applied to the stabilizing electrode. As we just discussed, the floating voltage and the (V_s=10) are very close in resistance while the grounded electrode and the (V_s=-10) are very near the theoretical limit of kT/I.

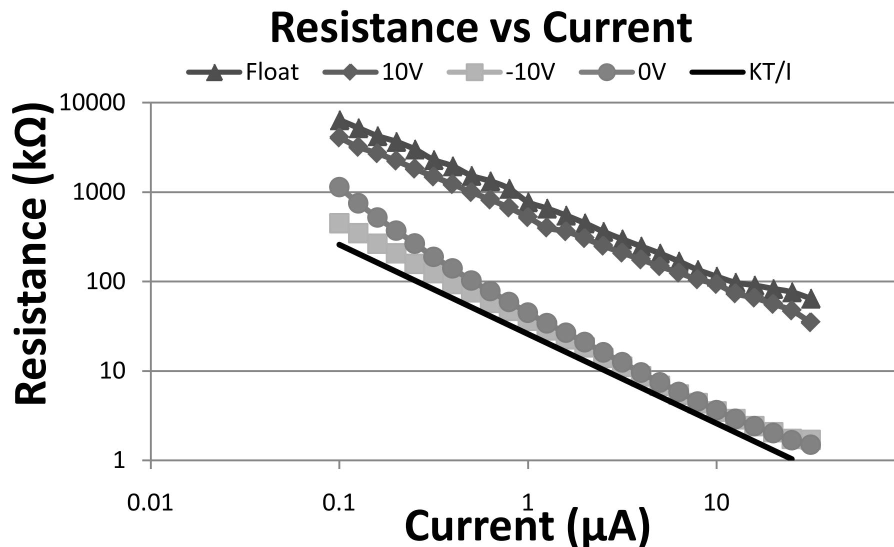


Figure 8.19: Shows the resistance versus current for due to the floating voltage at the injection node. Shown are the resistance values for various different voltages on the stabilizing electrode. These values are calculated by taking the derivative of Figure 8.18.

This information gives strong evidence that the introduction of a stabilizing electrode in between the input and output electrodes can greatly reduce the fluctuations in the floating voltage. By reducing these shifts in the floating voltage the resistance can also be greatly reduced. An optimal SIM configuration contains a stabilizing electrode for this reason.

8.6 Conclusions

There are many factors that affect the gain and frequency response of a device. Versions of the SIM have suffered from many different factors that have hindered its gain or frequency response. Some of the limitations that have been overcome already are the barrier resistance and low impact ionization efficiency. Solutions to other factors, such as space charge and the resistive effect attributed to the floating voltage, are still being explored. The capability of overcoming these factors seems to be possible with a few alterations. By incorporating all of the different solutions an optimal SIM design can be realized.

9 OPTIMAL SIM DESIGN

9.1 Introduction

With the information gleaned from the previous chapter, there is an optimal design that has yet to be realized. This *optimal* version of the SIM includes the designs and alterations discussed previously to maximize both gain and frequency response as much as possible. Figure 9.1 shows this design. The design includes an ohmic contact at the injection electrode instead of a Schottky contact. This is indicated by the N+ wells at both the input and output nodes in the figure. The ohmic contact design cuts down on the barrier resistance and it gives the added benefit of consistency in fabrication. This was discussed in Chapters 5 and 7. This version of the SIM is also fabricated with pedestals, as in the buried oxide design of Chapter 6. The buried oxide allows for increased ionization efficiency by controlling carrier path through the optimal electric field. The design also benefits from elongated input and output wells and metal contacts. As we learned in Section 8.3, the elongation of the input and output provides a geometry change that increases the cross-sectional area of the depletion region. This proportionally cuts down on the space charge resistance during operation. This elongation feature comes in the form of the circular design as discussed in Section 8.4.1. The circular design further enhances the ionization efficiency by reducing the edge effects of spherical doping profiles. This, in turn, reduces premature breakdown and allows all carriers to encounter a similar electric field profile as they traverse from input to output. The device also benefits from a fourth *stabilizing* electrode

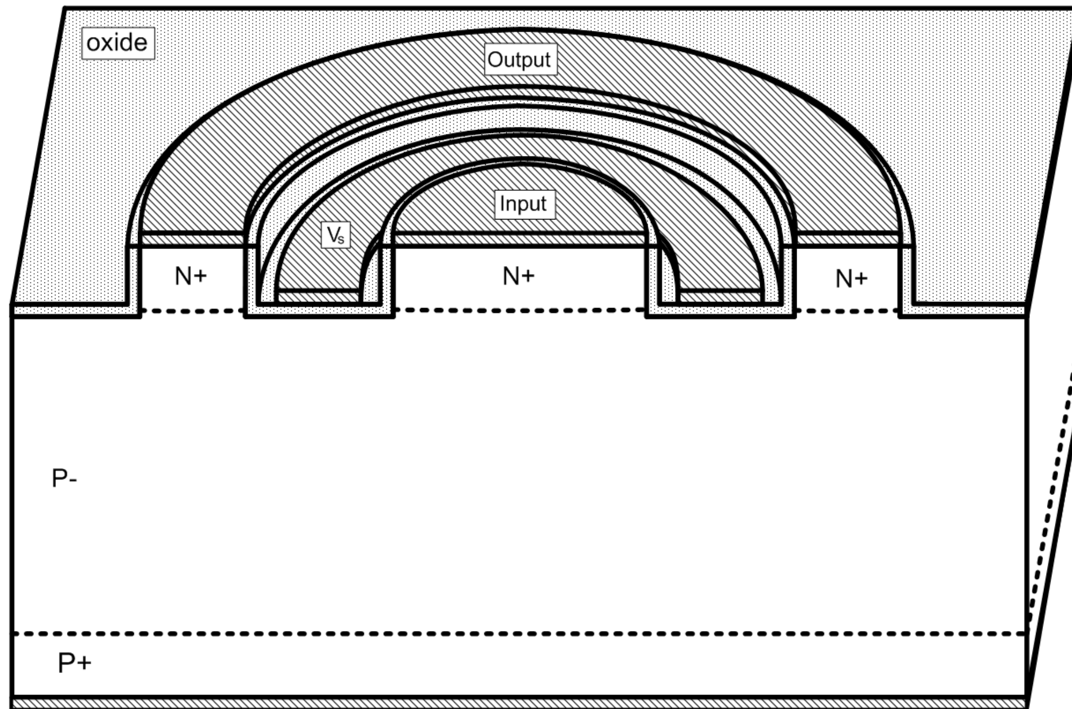


Figure 9.1: Illustrates the optimal SIM design. The design includes the ohmic contact injection node to eliminated barrier resistance at the input. It also has the buried oxide for increased ionization efficiency. It is circular in nature to reduce edge effects and space charge resistance. It has a fourth electrode in order to reduce the effects of the floating voltage at the input node.

between the input and output electrodes (labeled V_s in the figure). This insulated electrode reduces the effects of the floating voltage as discussed in Section 8.5.3. Also, during the operation of this version of the SIM, it is beneficial to include a DC current offset to reduce the effects of the floating voltage at the input node. The benefits of a DC current offset were discussed in Section 8.5.2. These benefits include a reduction in voltage shifts at the input node by maintaining a more consistent carrier distribution in the depletion region during operation.

The operation of this optimal SIM design is consistent with previous versions of the SIM as shown in Figure 9.2. Electrons injected at the inner circular electrode (input) travel outward toward the outer circular ring (output). As they enter electric field sufficiently high for ionization events, electron hole pairs are generated. The generated electrons continue toward the

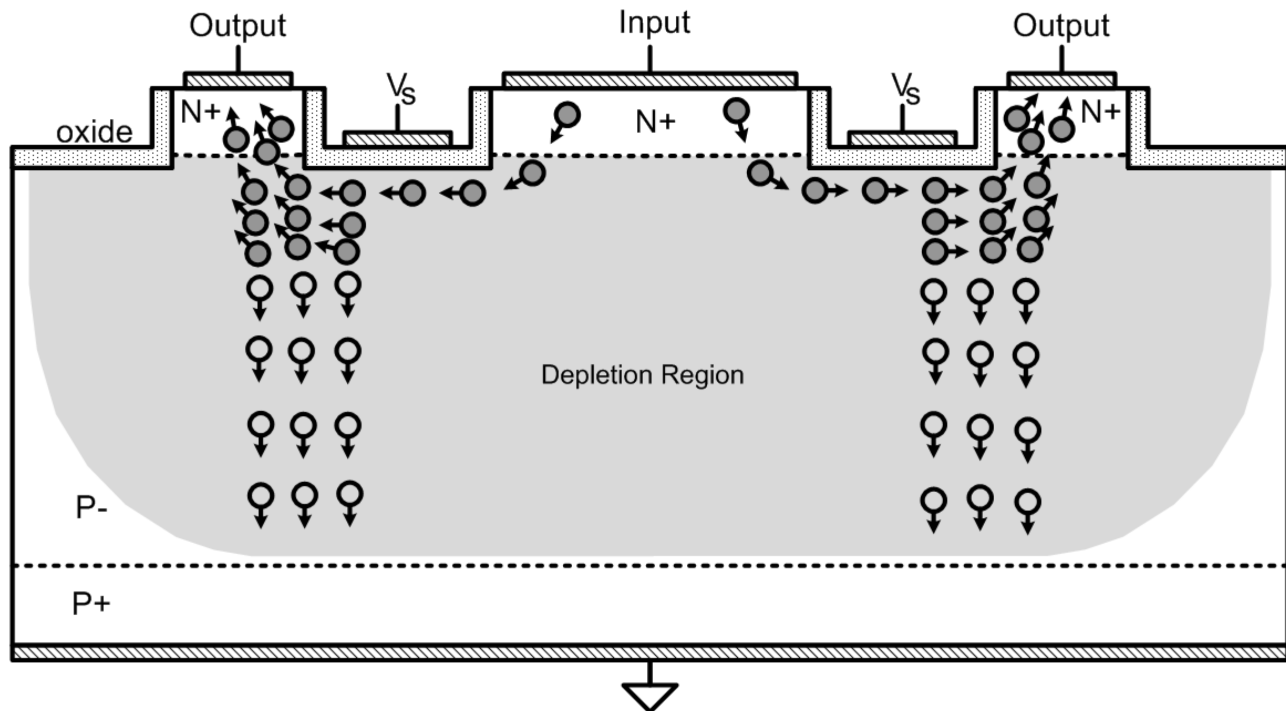


Figure 9.2: Illustrates a cross section of the optimal SIM shown in Figure 9.1. The figure shows the operation in the breakdown stage of the SIM. The operation stays consistent with previous versions of the SIM.

output electrode and the holes are drawn toward the substrate. Also, it becomes very evident from this cross-section that the electrons are more spread out as they travel from input to output in the circular setup reducing the effects of space charge resistance.

Unfortunately, due to constraints on time and resources only combinations of the above features have been tested together. An optimal SIM device with all of the features combined has not been realized. This has made the analysis of the different parts of the design difficult due to the limiting factors of the non-optimized features. When the circular design was tested to see the reduction in space charge the fourth electrode had yet to be added making the floating voltage the frequency limit. Later, when the fourth electrode was added, the mask set used made it impossible to fabricate with elongated features making space charge the frequency limiter.

Because of this, only theoretical limits to the complete optimal SIM can be given. This is done in the following section.

9.2 Theoretical Limits to Optimal Design

With the optimal SIM design laid out in the previous section, it becomes necessary to further discuss the theory behind its operation. By doing this, a better understanding of the operation of this version of the SIM can be obtained. This includes theoretical resistances and frequency limits. It is important at this point to note that several of the optimization features have been used together in the past. But these devices never used the optimization features all together in one device. Ohmic contact devices with buried oxide and elongated electrodes have been tested with the DC offset. But at that time the usefulness of the stabilizing electrode were not well knows. Because they lacked the stabilizing electrode they were severely hampered by the floating voltage effects. Later, the stabilizing electrode was added using a mask set previously designed but this setup didn't include the fabrication of the elongated electrodes. Thus, an ohmic contact, buried oxide, SIM with a stabilizing electrode was fabricated and tested using DC current offset techniques, but they were limited by the space charge of the device due to the lack of elongated electrodes. This section will analyze the theoretical upper limit frequency response assuming all of the features are combined together in an optimal SIM design.

First, the benefits of elongated electrodes with respect to the space charge resistance (R_{sc}), as discussed in Section 8.3, are analyzed. This is done by taking into consideration Equation 8.10 and the geometry of different SIM structures. The only difference between previous versions of the SIM and the elongated version of the SIM (with respect to Equation 8.10) is the cross-sectional area (A) of the depletion region. Depletion width of a typical device is about

5μm. Previous version of the SIM had an electrode length of approximately 5μm and a depletion depth of 10μm. With this information, and the saturation velocity and permittivity of Silicon, a value of $R_{sc}\approx 24k\Omega$ is obtained. Now taking into consideration the elongated electrode length of 360μm, with all other factors held constant, a value of $R_{sc}\approx 330\Omega$ is calculated. Assuming a system capacitance of C_d=10pF, the addition of the elongated electrodes increases the upper limit 3dB frequency (due to space charge) from approximately 670kHz to 48MHz.

Next, the reduction of floating voltage effects by the introduction of the stabilizing electrode is considered. The resistance values obtained analytically for the floating voltage resistance from Section 8.5.3 are used for this analysis. For a 1μA input current, without the stabilizing electrode, a resistance of $R\approx 770k\Omega$. With the stabilizing electrode the resistance is approximately $R\approx 26k\Omega$. Using the same capacitance as before (C_d=10pF), the resistance produces an upper limit frequency (due to the floating voltage) of only about 25kHz without the stabilizing electrode. With the addition of the stabilizing electrode the limit improves to approximately 600kHz.

Finally, the efforts to reduce the floating voltage effects are taken a step further by including a DC offset current. If a DC offset current of 10μA as explained in Section 8.5.2 and the γ factor are taken into consideration (i.e. Equation 8.21) the theoretical upper limit frequency (due to the floating voltage) improves to 2.5MHz without the stabilizing electrode and 80MHz with the stabilizing electrode.

Using the information from the above three paragraphs the theoretical frequency limits of the optimal SIM can be extracted. Again, this is done using the same system capacitance as used before. With the resistance values for the elongated electrodes, stabilizing electrode and DC

offset being used in tandem the upper limit frequency is approximately 30MHz. This is a marked improvement from measured results for combinations of the above improvements.

9.3 Fabrication Considerations

The fabrication of the optimal SIM design would require a new mask set that takes all of the features into consideration. A theoretical top-view of this version of the SIM showing the metallization can be seen in Figure 9.3. There are a few things that need to be taken into consideration during the fabrication and testing of the device. First, notice that there is an output pad that would be used as a point of contact when biasing the output electrode. There isn't a corresponding pad for the input or the stabilizing electrode. This may require wirebonding techniques to access these electrodes.

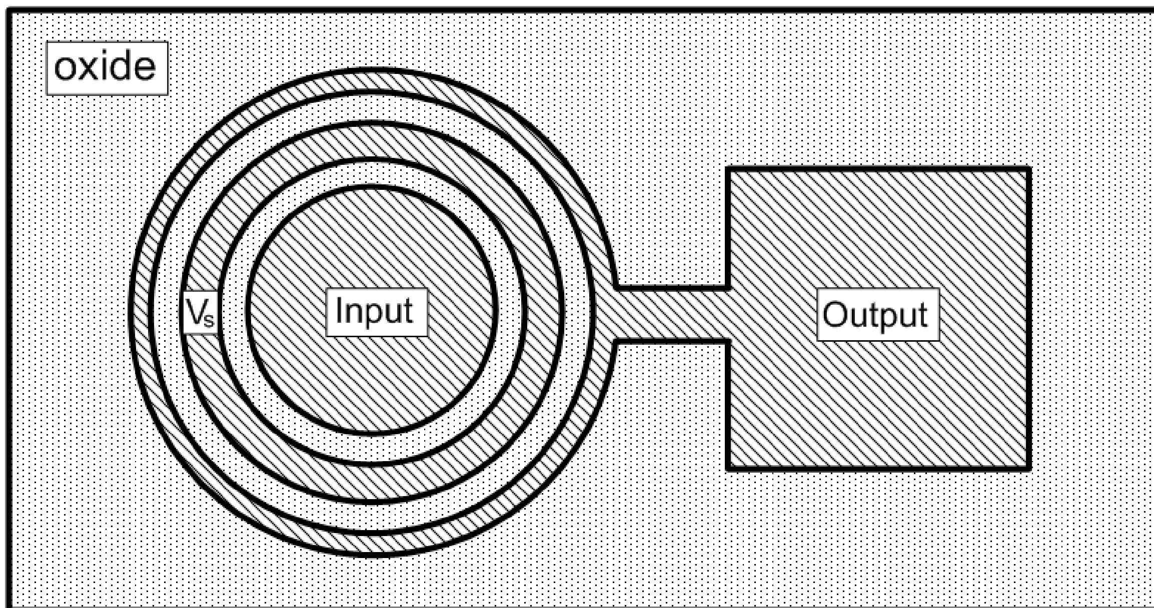


Figure 9.3: Top view of the optimal SIM illustrating the metallization. This design contains all of the optimization features discussed in Chapter 8.

Second, because the device is circular in nature it rules out the possibility of using the wet etch technique for the buried oxide pedestal features. This makes the fabrication process slightly more complicated and takes away some of the flexibility of the wet etch option.

Third, the aligning is important to keep equal distance between of the stabilizing electrode between all portions of input and output. Having this electrode shifted slightly up, down, left or right could introduce effects that prevent taking full advantage of the γ factor for the stabilizing electrode.

The fabrication techniques for the optimal SIM are identical to those seen in Appendix B.2 except a proper mask for the final metallization step would need to be designed in order to include the stabilizing electrode.

9.4 SIM vs. BJT vs. MOSFET

It should be noted that the optimal SIM design resembles a MOSFET or a BJT in many ways. They can all follow a typical n-p-n doping profile and the fourth stabilizing electrode is insulated and located in similar to a MOSFET's gate. Even thought the SIM resembles a MOSFET and BJT they are very different in operation and design. These differences include their mode of operation, doping concentrations, doping profile and overall geometry.

As stated before, the BJT and SIM have similar band profiles. In a BJT, electrons leave the emitter and diffuse through a narrow base region into the depleted doping of the collector. This base region is heavily doped to reduce hole recombination in the emitter and quite narrow to reduce the number of electrons lost due to recombination in the base. This feature is nothing like what is seen in the SIM. Typical BJTs also have their carriers travel through the bulk as opposed to along the surface like the SIM [4].

The doping of a MOSFET more closely resembles the SIM. The channel has a low (intrinsic) doping profile that allows for quick depletion and (with a positive gate bias) inversion. The inversion of the channel is how a MOSFET works. Without inversion current from the source to the drain would only occur in pinch off, which is an undesirable full depletion of the channel. This undesirable state of pinch off is the desirable state for SIM operation (Meaning, the SIM operates by fully depleting the semiconductor between the input and the output.). Inversion of this region would be detrimental to the operation of the SIM because it would make amplification of the input signal through impact ionization impossible. In fact, as we saw in the previous section, a positive voltage on the stabilizing electrode is not favorable. As far as geometries go, the buried oxide geometry is something that isn't typical in MOSFET structures. Also, the circular structure to reduce edge effects is not utilized in FETs.

Finally, while impact ionization is possible in BJTs or MOSFETs it is undesirable and often can cause damage to the device. Impact ionization is not only desirable but the doping and geometries of the SIM were made to handle these high fields and breakdown conditions.

9.5 Conclusions

Different optimization features have been utilized in the fabrication of the SIM. Each of these features was designed to either increase the gain or increase the frequency response. A useful amplifier will be able to handle appreciable gains at sufficient speeds. The combination of all of the optimization features into a singular device produces the optimal SIM. This theoretical structure yields a much improved frequency response from previous versions of the SIM. This is done by reducing the overall resistance of the device during operations with the addition of elongated doping and electrodes. This elongation feature is used together with a stabilizing

electrode that greatly reduces fluctuations of the voltage at the input. The combination of these features increases the upper limit frequency of the device. While doing this, it maintains excellent impact ionization efficiency and gains by utilizing the buried oxide fabrication technique and circular design.

10 CONCLUSION

10.1 Introduction

The advancements to the SIM have taken place in order to create a novel, quality amplifier. The design was conceived in order to fill a hole in the amplification world. More specifically, to make the separation of detection and amplification of optical signals easier, cheaper and trivial. This need arises when analyzing current optical systems. Many systems utilize photodiodes in conjunction with slow, noisy transistor based amplifiers. Others utilized APDs that make optimal detection and multiplication material difficult if not impossible to combine. The introduction of a SIM makes separation of the detector and amplifier in an APD-like structure achievable.

In order to make this device viable in present day systems, an evolutionary process has taken place. Alterations to the design have been made to improve gain and frequency response as well as consistency in fabrication. These alterations have been documented in this dissertation culminating in an optimal design that would, theoretically, greatly increase the frequency response and maintain a stable gain profile. This body of work still leaves much to be accomplished. The vision of a SIM integrated in an optical detection system has not died.

10.2 Future Work

From the information outlined in the previous chapter, the most glaring future work is the realization and characterization of the optimal SIM structure. The realization of this device was just out of reach. The theory behind its improved functionality has been outlined. The simulation and modeling to support this theory has also been explored. The fabrication techniques and processes for its development are all set in place and well established. With sufficient time and resources, the optimal design of the SIM can be realized with confidence. Its realization, though, is still not the culmination of this project. The original vision of its integration into an optical detection system remains. Integration in to an on-chip optical detection system would bring closure to the SIM project. The research and design of the device has been performed throughout the years with this ultimate goal in mind.

Further characterization of the noise, gain, and bandwidth are also included in future efforts. The noise characteristics of the SIM have never been well established. It has been assumed that the noise of the device would somewhat mirror that of other impact ionization and silicon based devices. The SIM is unique in structure and purpose which makes these efforts necessary in future analysis.

Because of the flexibility of the SIM, a more thorough investigation into applications should also be looked at. The SIM is a stand-alone amplifier that can be used in nearly any system needing the amplification of a current based signal. There are many systems that meet this description. Some of these that have been suggested by past participants of the project include audio amplifiers and charge-coupled devices [70].

10.3 Conclusions

This dissertation outlined the study, research and development of the SIM. This stand-alone amplifier was designed with the intent of improving present day optical detection systems. The background behind this effort was established in Chapter 1. There, it was established that there is a need for the development of a new kind of amplifier. Such a device would be capable of utilizing impact ionization for gain and handling an arbitrary current source as the input.

In Chapter 3 the theory behind its development was established. The usefulness of impact ionization gain as a clean, fast source of amplification was developed. It was shown that silicon is the optimal material for the device.

The development of any device in modern day research takes sophisticated modeling and simulation techniques. Chapter 4 outlined the development of device simulation software platforms. It also outlined the numerical methods and processes used in semiconductor device simulations. This has been critical in the development of the SIM. The SIM has been modeled extensively over the years using the Silvaco platform.

Chapters 5, 6 and 7 discussed the operation of the SIM and some of the established alterations. These alterations have been based on research and simulation improving the overall gain and frequency response of the SIM.

Chapter 8 established issues with the gain and speed of the device. These issues come with corresponding modifications that have been proposed, researched, simulated and developed in efforts to overcome the issues dealing with the gain controllability and speed. The combination of all of the researched modifications into an optimal device was discussed in Chapter 9.

Future work remains for the development of what still has potential to be an important device in modern day optical detection systems.

REFERENCES

[1] J. L. Beutler, C. S. Clauss, M. S. Johnson, A. R. Hawkins, M. D. Jack, G. R. Chapman, and K. Kosai, "Frequency response of solid-state impact ionization multipliers," *Journal of Applied Physics,* vol. 101, pp. 023117-023117-9, 2007.

[2] J. L. Hullett and S. Moustakas, "Optimum Transimpedance Broad-Band Optical Pre-Amplifier Design," *Optical and Quantum Electronics,* vol. 13, pp. 65-69, 1981.

[3] B. M. Oliver, "Thermal and quantum noise," *Proceedings of the IEEE,* vol. 53, pp. 436-454, 1965.

[4] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices* vol. 3rd. Holbokem, New Jersey: John Wiley & Sons, Inc, 2007.

[5] J. Maserjian, "Determination of Avalanche Breakdown in pn Junctions," *J. Appl. Phys,* vol. 30, pp. 1613-1614, 1959.

[6] K. G. McKay, "Avalanche Breakdown in Silicon," *Phys. Rev.,* vol. 94, p. 877, 1954.

[7] K. G. McKay and K. B. McAfee, "Electron Multiplication in Silicon and Germanium," *Phys. Rev.,* vol. 91, p. 1079, 1953.

[8] E. J. Charlson and J. C. Lien, "Al P-Silicon Mos Photovoltaic Cell," *Journal of Applied Physics,* vol. 46, pp. 3982-3987, 1975.

[9] P. B. Page and J. R. Peters, "Light Emitting Diode Displays," *Electrical Communication,* vol. 49, pp. 80-87, 1974.

[10] S. M. Sze, *Selected solutions for semiconductor devices : physics and technology*. New York ; Chichester: John Wiley, 1985.

[11] J. C. Campbell, A. G. Dentai, G. J. Qua, J. Long, and V. G. Riggs, "Planar Ingaas Pin Photodiode with a Semi-Insulating Inp Cap Layer," *Electronics Letters,* vol. 21, pp. 447-448, 1985.

[12] T. P. Lee, C. A. Burrus, M. A. Pollack, and R. E. Nahory, "High-Speed Schottky-Barrier Photodiode in Lpe Inxga1-Xas for 1.0 Mu to 1.1 Mu Wavelength Region," *IEEE Transactions on Electron Devices,* vol. 22, pp. 1062-1062, 1975.

[13] R. J. McIntyre, "A new look at impact ionization-Part I: A theory of gain, noise, breakdown probability, and frequency response," in *Electron Devices, IEEE Transactions on* vol. 46, ed, 1999, pp. 1623-1631.

[14] J. Wilson and J. F. B. Hawkes, *Optoelectronics : an introduction*, 3rd ed. London: Prentice Hall Europe, 1998.

[15] R. W. Brown and Chamberl.Sg, "Quantum Efficiency of a Silicon Gate Charge-Coupled Optical Imaging Array," *Physica Status Solidi a-Applied Research,* vol. 20, pp. 675-685, 1973.

[16] P. K. Cheo, *Fiber optics and optoelectronics*, 2nd ed ed. Englewood Cliffs ; London: Prentice-Hall International, 1990.

[17] E. R. Weber, *Quantum efficiency in complex systems. Pt. 1, Biomolecular systems*. San Diego, Calif. ; London: Elsevier, 2010.

[18] J. E. Parrott and L. P. Leonidou, "General Theory of P-N-Junction Capacitance," *Physica Status Solidi a-Applied Research,* vol. 25, pp. 231-240, 1974.

[19] D. Jager, R. Heidemann, and R. Kassing, "Influence of Carrier Transit-Time on Small-Signal Admittance of Sclc Diodes," *Physica Status Solidi a-Applied Research,* vol. 42, pp. 657-665, 1977.

[20] J. H. Reynolds and Meulenbe.A, "Measurement of Diffusion Length in Solar Cells," *Journal of Applied Physics,* vol. 45, pp. 2582-2592, 1974.

[21] S. Hasegawa, A. Tanaka, and T. Sukegawa, "Measurement of Optical-Absorption Coefficient in the Depletion Region of Gaas Schottky-Barrier Photodiode," *Japanese Journal of Applied Physics Part 1-Regular Papers Short Notes & Review Papers,* vol. 23, pp. 1152-1152, 1984.

[22] J. Ohsawa, T. Kozawa, O. Ishiguro, and H. Itoh, "Selective detection of blue and ultraviolet light by an InGaN/GaN Schottky barrier photodiode," *Japanese Journal of Applied Physics Part 2-Letters & Express Letters,* vol. 45, pp. L614-L616, Jul 2006.

[23] J. C. Campbell, S. Demiguel, M. Feng, A. Beck, G. Xiangyi, W. Shuling, Z. Xiaoguang, L. Xiaowei, J. D. Beck, M. A. Kinch, A. Huntington, L. A. Coldren, J. Decobert, and N. Tscherptner, "Recent advances in avalanche photodiodes," *Selected Topics in Quantum Electronics, IEEE Journal of,* vol. 10, pp. 777-787, 2004.

[24] T. Ikoma and K. Hara, "Temperature-Dependence of Hole Saturation Velocity in Silicon," *Applied Physics,* vol. 3, pp. 431-432, 1974.

[25] O. H. Kwon, M. M. Hayat, J. C. Campbell, B. E. A. Saleh, and M. C. Teich, "Optimized breakdown probabilities in Al0.6Ga0.4As-GaAs heterojunction avalanche photodiodes," *IEEE Electron Device Letters,* vol. 25, pp. 599-601, Sep 2004.

[26] P. H. Zhang, V. H. Crespi, E. Chang, S. G. Louie, and M. L. Cohen, "Computational design of direct-bandgap semiconductors that lattice-match silicon," *Nature,* vol. 409, pp. 69-71, Jan 4 2001.

[27] K. Nishida, K. Taguchi, and Y. Matsumoto, "InGaAsP heterostructure avalanche photodiodes with high avalanche gain," *Appl. Phys. Lett.,* vol. 35, pp. 251-253, 1979.

[28] J. C. Blakesley, P. See, A. J. Shields, B. E. Kardynal, P. Atkinson, I. Farrer, and D. A. Ritchie, "Efficient Single Photon Detection by Quantum Dot Resonant Tunneling Diodes," *APS,* vol. 94, p. 067401, 2005.

[29] E. Hergert, "Photomultiplier tubes count photons," *Laser Focus World,* vol. 33, p. 129, 1997.

[30] S. F. Kou, C. Qian, J. Eryou, X. Wei, and Q. Ran, "Single-photon time response of reach-through APD," *Optics Communications,* vol. 281, pp. 2481-2484, May 1 2008.

[31] W. J. Kindt, N. H. Shahrjerdy, and H. W. van Zeijlv, "A silicon avalanche photodiode for single optical photon counting in the Geiger mode," *Sensors and Actuators,* vol. 60, pp. 98-102, 1997.

[32] H. Daulet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb, "Photon counting techniques with silicon avalanche photodiodes," *Applied Optics,* vol. 32, pp. 3894-3894-3900, 1993.

[33] H. Brower, "Demonstration of Photoelectric Effect with a Geiger Counter," *American Journal of Physics,* vol. 44, pp. 305-306, 1976.

[34] H. Demir, G. Erbay, K. M. Kir, K. Omurlu, F. Berk, and C. Aktolun, "Clinical validation of technetium-99m MEBI-gated single-photon emission computed tomography (SPECT) for avoiding false positive results in patients with left bundle-branch block: Comparison with stress-rest nongated SPECT," *Clinical Cardiology,* vol. 26, pp. 182-187, Apr 2003.

[35] A. Giudice, F. Zappa, P. Maccagnani, A. Gulinatti, M. Ghioni, S. Cova, and R. Biasi, "High-rate photon counting and picosecond timing with silicon-SPAD based compact detector modules," *J. Modern Optics,* vol. 54, pp. 225-225-237, 2007.

[36] E. J. Lerner, "Photomultiplier tubes stay the course," *Laser Focus World,* vol. 37, p. 151, 2001.

[37] E. J. Lerner, "Photomultiplier tubes offer high-end sensitivity," *Laser Focus World,* vol. 32, p. 87, 1996.

[38] W. Paper, "Noise in Silicon Photomultipliers and Vacuum Photomultiplier Tubes," vol. 2008, ed: Sensl, p. 20.

[39] D. Zankowsky, "How to pick the proper photomultiplier tube," *Laser Focus World,* vol. 32, p. 159, 1996.

[40] W. Shockley, "Path to Conception of Junction Transistor," *IEEE Transactions on Electron Devices,* vol. 23, pp. 597-620, 1976.

[41] S. W. Amos, *Principles of transistor circuits : introduction to the design of amplifiers, receivers and digital circuits*, 8th ed. Oxford: Butterworth-Heinemann, 1994.

[42] W. Shockley, "The Path to the Conception of the Junction Transistor," *IEEE Transactions on Electron Devices,* vol. 31, pp. 1523-1546, 1984.

[43] A. R. Hawkins, W. S. Wu, P. Abraham, K. Streubel, and J. E. Bowers, "High gain-bandwidth-product silicon heterointerface photodetector," *Applied Physics Letters,* vol. 70, pp. 303-305, Jan 20 1997.

[44] A. Yoshizawa and H. Tsuchida, "A 1550 nm single-photon detector using a thermoelectrically cooled InGaAs avalanche photodiode," *Japanese Journal of Applied Physics Part 1-Regular Papers Short Notes & Review Papers,* vol. 40, pp. 200-201, Jan 2001.

[45] C. M. Soukoulis and North Atlantic Treaty Organization. Scientific Affairs Division., *Photonic band gap materials*. Dordrecht ; London: Kluwer Academic, 1996.

[46] T. P. Pearsall, "Threshold Energies for Impact Ionization by Electrons and Holes in Inp," *Applied Physics Letters,* vol. 35, pp. 168-170, 1979.

[47] H.-W. Lee and A. R. Hawkins, "Solid-state current amplifier based on impact ionization," *Applied Physics Letters,* vol. 87, pp. 073511-073511-3, 2005.

[48] H. W. Lee, J. L. Beutler, and A. R. Hawkins, "High Gain Effects for Solid-State Impact-Ionization Multipliers," *Quantum Electronics, IEEE Journal of,* vol. 42, pp. 471-476, 2006.

[49] M. S. Johnson, J. L. Beutler, A. P. Nelson, and A. R. Hawkins, "Solid-State Impact-Ionization Multiplier With P-N Junction Injection Node," *IEEE Transactions on Electron Devices,* vol. 56, pp. 1360-1364, Jun 2009.

[50] M. S. Johnson, J. L. Beutler, A. P. Nelson, Y. Tseung, and A. R. Hawkins, "SiO2 Barriers for Increasing Gain Events in Solid-State Impact-Ionization Multipliers," *IEEE Journal of Quantum Electronics,* vol. 45, pp. 1068-1073, Sep 2009.

[51] J. S. Yuan, "Avalanche Breakdown Effects on AlGaAs/GaAs HbT Performance," *International Journal of Electronics,* vol. 74, pp. 909-916, Jun 1993.

[52] J. K. Mishra, A. K. Panda, and G. N. Dash, "An extremely low noise heterojunction IMPATT," *Electron Devices, IEEE Transactions on,* vol. 44, pp. 2143-2148, 1997.

[53] P. L. Ntake and D. R. Conn, "Frequency Multiplication by a P-I-N Diode When Driven into Avalanche Breakdown," *IEEE Transactions on Microwave Theory and Techniques,* vol. Mt23, pp. 477-485, 1975.

[54] P. A. Wolff, "Theory of Electron Multiplication in Silicon and Germanium," *Phys. Rev.,* vol. 95, p. 1415, 1954.

[55] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors," *Phys. Rev.,* vol. 14, p. 556, 1976.

[56] A. W. Overhauser, "Effective-Mass Variation in Silicon Inversion Layers," *Physical Review B,* vol. 18, pp. 2884-2887, 1978.

[57] H.-W. Lee, "Solid-State Impact-Ionization Multiplier," Dissertation/Thesis, 2006.

[58] S. Tam, P. K. Ko, and C. M. Hu, "Lucky-Electron Model of Channel Hot-Electron Injection in Mosfets," *IEEE Transactions on Electron Devices,* vol. 31, pp. 1116-1125, 1984.

[59] K. F. Brennan, *: The physics of Semiconductors with applications to optoelectronic devices.*: Cambridge, 1999.

[60] J. E. Hoogenboom, "Generalized Adjoint Monte-Carlo Technique," *Atomkernenergie,* vol. 25, pp. 232-232, 1975.

[61] D. Vasileska and S. M. Goodnick, "Computational electronics," *Materials Science & Engineering R-Reports,* vol. 38, pp. 181-236, Jul 15 2002.

[62] C. Jacoboni and P. Lugli, *The Monte Carlo method for semiconductor device simulation*. Wien: Springer-Verlag, 1989.

[63] J. Biel and J. Marro, "Generalization of Boltzmann-Equation," *Nuovo Cimento Della Societa Italiana Di Fisica B-General Physics Relativity Astronomy and Mathematical Physics and Methods,* vol. B 20, pp. 25-54, 1974.

[64] P. Yuan, K. A. Anselm, C. Hu, H. Nie, C. Lenox, A. L. Holmes, B. G. Streetman, J. C. Campbell, and R. J. McIntyre, "A new look at impact ionization-Part II: Gain and noise in short avalanche photodiodes," in *Electron Devices, IEEE Transactions on* vol. 46, ed, 1999, pp. 1632-1639.

[65] P. P. Webb, "Measurements of Ionization Coefficients in Silicon at Low Electric Fields," *Solid-St. Electron.,* vol. 16, p. 1189, 1973.

[66] A. R. Hawkins, "Silicon-Indium-Gallium-Arsenide Avalanche Photodetectors," PhD, University of California Santa Barbara, 1998.

[67] V. A. Kuzmin, N. N. Kryukova, A. S. Kyuregyan, T. T. Mnatsakanov, and V. B. Shuman, "Impact Ionization Coefficients of Electrons and Holes in Silicon," *Soviet Physics Semiconductors-Ussr,* vol. 9, pp. 481-482, 1975.

[68] R. J. McIntyre, "A new look at impact ionization - Part I: A theory of gain, noise, breakdown probability, and frequency response," *IEEE Transactions on Electron Devices,* vol. 46, pp. 1623-1631, Aug 1999.

[69] R. B. Emmons, "Avalanche-Photodiode Frequency Response," *J. Appl. Phys.,* vol. 38, pp. 3705-3714, 1967.

[70] J. L. Beutler, "Frequency Response and Gain Enhancement of Solid-state Impact-ionization Multipliers (SIMs)," 2010.

[71] C. D. Bulucea and Prisecar.Dc, "Particularities of Avalanche Multiplication in Silicon P-N-Junctions Due to Carrier Generation within Space-Charge Region," *Revue Roumaine De Physique,* vol. 18, pp. 961-980, 1973.

[72] D. Dascalu and N. Popescu, "Avalanche Breakdown Calculations for Abrupt Silicon P+ Nn+ (or N+ Pp+) Structures," *Revue Roumaine De Physique,* vol. 26, pp. 699-707, 1981.

[73] P. V. Akimov, I. V. Grekhov, and Y. N. Serezhkin, "Avalanche Breakdown in High-Voltage Large Area Silicon P-N-Junctions," *Radiotekhnika I Elektronika,* vol. 20, pp. 373-380, 1975.

[74] K. I. Nuttall and M. W. Nield, "Prediction of Avalanche Breakdown Voltage in Silicon Step Junctions," *International Journal of Electronics,* vol. 37, pp. 295-309, 1974.

[75] T. A. Fjeldly, T. Ytterdal, and M. Shur, *Introduction to device modeling and circuit simulation*. New York ; Chichester: Wiley, 1998.

[76] W. Nelson, P. Bokes, P. Rinke, and R. W. Godby, "Self-interaction in Green's-function theory of the hydrogen atom," *Physical Review A,* vol. 75, pp. -, Mar 2007.

[77] S. M. Sze, *Semiconductor devices : physics and technology*, 2nd ed. New York: Wiley, 2002.

[78] C. V. Pao, "Nonlinear Boltzmann-Equation in Transport-Theory," *Transactions of the American Mathematical Society,* vol. 194, pp. 167-175, 1974.

[79] V. K. Magas, L. P. Csernai, E. Molnar, A. Nyiri, and K. Tamosiunas, "Modified Boltzmann transport equation," *Nuclear Physics A,* vol. 749, pp. 202c-205c, Mar 7 2005.

[80] X. S. Xu, "A drift-diffusion model for semiconductors with temperature effects," *Proceedings of the Royal Society of Edinburgh Section a-Mathematics,* vol. 139, pp. 1101-1119, 2009.

[81] J. Bourgain, *Green's function estimates for lattice Schrödinger operators and applications*. Princeton, N.J. ; Oxford: Princeton University Press, 2005.

[82] D. L. Woolard, H. Tian, R. J. Trew, M. A. Littlejohn, and K. W. Kim, "Hydrodynamic Electron-Transport Model - Nonparabolic Corrections to the Streaming Terms," *Physical Review B,* vol. 44, pp. 11119-11132, Nov 15 1991.

[83] "ATLAS user's manual," in *Silvaco International*, ed, 2004.

[84] Y. T. Tsai and L. C. Huang, "Simulation of amorphous silicon thin-film transistor including adapted Gummel method," *International Journal of Numerical Modelling-Electronic Networks Devices and Fields,* vol. 10, pp. 3-11, Jan-Feb 1997.

[85] I. K. Argyros and S. Hilout, "A unifying theorem for Newton's method on spaces with a convergence structure," *Journal of Complexity,* vol. 27, pp. 39-54, Feb 2011.

[86] Z. B. Omar and M. Suleiman, "Solving First Order Systems of Ordinary Differential Equations Using Parallel R-Point Block Method of Variable Step Size and Order," *Chiang Mai Journal of Science,* vol. 37, pp. 1-13, Jan 2010.

[87] "Tonyplot User's Manual," in *Silvaco International*, ed, 2004.

[88] "Athena User's Manual," in *Silvaco International*, ed, 2004.

[89] A. R. Hawkins, T. E. Reynolds, D. R. England, D. I. Babic, M. J. Mondry, K. Streubel, and J. E. Bowers, "Silicon heterointerface photodetector," *Applied Physics Letters,* vol. 68, pp. 3692-3694, Jun 24 1996.

[90] P. P. Webb, R. J. McIntyre, and J. Conradi, "PROPERTIES OF AVALANCHE PHOTODIODES," *RCA (Radio Corporation of America) Review,* vol. 35, pp. 234-278, 1974.

[91] S. M. Park and C. Toumazou, "Low noise current-mode CMOS transimpedance amplifier for giga-bit optical communication," in *Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on*, 1998, pp. 293-296 vol.1.

[92] H.-W. Lee, J. L. Beutler, and A. R. Hawkins, "Surface structure silicon based impact-ionization multiplier for optical detection," *Optics Express,* vol. 13, pp. 8760-8765, 2005.

[93] E. Hökelek and G. Y. Robinson, "Aluminum/Nickel Silicide Contacts on Silicon," *Thin Solid Films,* vol. 53, pp. 135-135-140, 1978.

[94] L. Jianwen and P. Hongfang, "The Effect of Composition on the Barrier Height and Contact Resistance of TixW1-x/Si Contacts," *J. Electron. Mater.,* vol. 16, pp. 199-199-201, 1987.

[95] S. S. Cohen, P. A. Piacente, G. Gildenblat, and D. M. Brown, "Platinum silicide ohmic contacts to shallow junctions in silicon," *J. Appl. Phys.,* vol. 53, pp. 8856-8852-8862, 1982.

[96] H. Kyuwoon and D. H. Navon, "Breakdown voltage optimization of silicon p-π-v planar junction diodes," *Electron Devices, IEEE Transactions on,* vol. 31, pp. 1126-1135, 1984.

[97] R. C. Bassus, E. Falck, and W. Gerlach, "Application of the Evolution Strategy to Optimize Multistep Field Plates for High-Voltage Planar Pn-Junctions," *Archiv Fur Elektrotechnik,* vol. 75, pp. 345-349, Jul 1992.

[98] B. Eitan, D. Frohmanbentchkowsky, and J. Shappir, "Impact Ionization at Very Low Voltages in Silicon," *Journal of Applied Physics,* vol. 53, pp. 1244-1247, 1982.

[99] D. Ritter, K. Weiser, and E. Zeldov, "Steady-state photocarrier grating technique for diffusion-length measurement in semiconductors: Theory and experimental results for amorphous silicon and semi-insulating GaAs," *J. Appl. Phys.,* vol. 62, p. 4563, 1987.

[100] C. Canali, G. Majni, R. Minder, and G. Ottaviani, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *Electron Devices, IEEE Transactions on,* vol. 22, pp. 1045-1047, 1975.

[101] M. Saritas and H. D. McKell, "Comparison of minority-carrier diffusion length measurements in silicon by the photoconductive decay and surface photovoltage methods," *J. Appl. Phys.,* vol. 63, pp. 4561-4567, 1988.

[102] M. S. Johnson, R. Shuldberg, and A. R. Hawkins, "Improving Gain Efficiency in Planar Impact Ionization Devices," *ECS Transactions,* vol. 33, pp. 143-153, 2010.

[103] R. C. Becker and J. B. Beyer, "On Gain-Bandwidth Product for Distributed-Amplifiers," *IEEE Transactions on Microwave Theory and Techniques,* vol. 34, pp. 736-738, Jun 1986.

[104] T. I. Pritchard and W. R. Wilson, "Understanding Gain-Bandwidth Product of Bipolar-Transistors," *Electronic Engineering,* vol. 47, pp. 40-41, 1975.

[105] R. G. Roozbahani, "BJT-BJT, FET-BJT, and FET-FET - On improving the gain-bandwidth product of cascode configurations.," *IEEE Circuits & Devices,* vol. 20, pp. 17-22, Nov-Dec 2004.

[106] W. Monch, "Barrier heights of real Schottky contacts explained by metal-induced gap states and lateral inhomogeneities," *Journal of Vacuum Science & Technology B,* vol. 17, pp. 1867-1876, Jul-Aug 1999.

[107] M. Finetti, P. Ostoja, S. Solmi, and G. Soncini, "Aluminum-Silicon Ohmic Contact on Shallow N+-P Junctions," *Solid-State Electronics,* vol. 23, pp. 255-&, 1980.

[108] S. M. Sze, *Physics of Semiconductor Devices* vol. 2nd. New York: John Wiley and Sons, 1981.

[109] M. Morschbach, S. Diemer, M. Oehme, and E. Kasper, "Design of a Monolithic 79 GHz Rectenna with MOTT Diodes," *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems Digest of Papers,* pp. 122-122- 125, 2003.

[110] O. Militaru, L. Borrello, C. Bozzi, M. Da Rold, R. Dell'Orso, S. Dutta, A. Messineo, A. Mihul, G. Tonelli, P. G. Verdini, R. Wheadon, and Z. Xie, "Study of edge effects in the breakdown process of p(+) on n-bulk silicon diodes," *Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment,* vol. 439, pp. 262-269, Jan 11 2000.

[111] S. M. Sze and G. Gibbons, "Effect of Junction Curvature on Breakdown voltages in Semiconductors," *Solid-State Electron.,* vol. 9, pp. 831-845, 1966.

[112] Y. S. Melnikova, "Lowering of the Instability Threshold of a Homogeneous Avalanche Breakdown of P+-N Junctions in Silicon," *Soviet Physics Semiconductors-Ussr,* vol. 24, pp. 40-43, Jan 1990.

[113] W. Fulop, "Calculation of Avalanche Breakdown Voltages of Silicon p-n Junctions," *Solid-St. Electron.,* vol. 10, p. 39, 1967.

[114] N. T. Stringer, "The effect of DC offset on current-operated relays," *IEEE Transactions on Industry Applications,* vol. 34, pp. 30-34, Jan-Feb 1998.

[115] R. C. Jaeger, *Microelectronic circuit design* vol. 2nd ed. New York: McGraw-Hill Higher Education, 2004.

LIST OF PUBLICATIONS

1. Joshua L. Beutler, Carleton S. Clauss, Michael S. Johnson, and Aaron R. Hawkins, “Frequency Response of Solid-State Impact-Ionization Multipliers,” *Journal of Applied Physics* vol. 101, 023117 (2007).

2. Michael S. Johnson, Joshua L. Beutler, Alan P. Nelson, and Aaron R. Hawkins, “Solid-state Impact-ionization Multiplier with PN Junction Injection Node,” *IEEE Journal of Transactions on Electron Devices* vol.56, 1360-1363, (2009).

3. Michael S. Johnson, Joshua L. Beutler, Alan P. Nelson, and Aaron R. Hawkins, “SiO2 Barriers for increasing gain events in Solid-state Impact-ionization Multipliers,” *IEEE Journal of Quantum Electronics* vol. 45, 1068-1073, (2009).

4. Michael S. Johnson, Roy P. Shuldberg, and Aaron. R. Hawkins, "Improving Gain Efficiency in Planar Impact Ionization Devices," *ECS Transactions,* vol. 33, 143-153, (2010).

A. **SILVACO CODE AND SIMULATIONS**

A.1 Silvaco Simulation of SIM in Rectangular Coordinates

The following code was written for operation in the Deckbuild environment in Silvaco's Altas platform. The code is for a device simulation of the ohmic contact SIM discussed in Chapter 6. It simulates the device in rectangular coordinates for both DC and transient response.

```
#-------------------------------------------------------------------------------
Go atlas

mesh space.mult=1.0
#-----------
#SET UP MESH
#-----------

x.mesh loc=0        spac=1
x.mesh loc=4        spac=0.1
x.mesh loc=3.16     spac=0.05
x.mesh loc=3.84     spac=0.05
x.mesh loc=5        spac=0.1
x.mesh loc=6        spac=0.1
x.mesh loc=8        spac=0.1
x.mesh loc=9        spac=0.1
x.mesh loc=10.2     spac=.1
x.mesh loc=10.8     spac=.1
x.mesh loc=12       spac=.1

y.mesh loc=-.5      spac=0.1
y.mesh loc=0        spac=0.01
y.mesh loc=0.14     spac=0.01
y.mesh loc=.5       spac=0.1
y.mesh loc=1        spac=0.25
y.mesh loc=2        spac=0.5
y.mesh loc=3.5      spac=0.25
y.mesh loc=4.0      spac=0.5
y.mesh loc=4.1              spac=.1

#---------------
# set up Regions
#---------------

region   num=1 material=silicon   x.min=0 x.max=12        y.min=0  y.max=4
```

```
region     num=2 material=air   x.min=0    x.max=12              y.min=-0.5             y.max=0
region     num=3 material=air   x.min=0    x.max=12              y.min=4   y.max=4.1


#---------------------
# SET UP ELECTRODES
#---------------------

elec num=1   name=drain    x.min=8.2      x.max=8.8              y.min=-0.1             y.max=0.1
elec num=2   name=source   x.min=3.2    x.max=3.8              y.min=-0.1             y.max=0.1
elec num=3   name=ground   x.min=0      x.max=12              y.min=4   y.max=4.1

#--------------
# SET UP DOPING
#--------------
#Epi doping
doping     uniform reg=1           p.type conc=1.5e15

#Drain Well
doping     uniform reg=1           n.type conc=1e18 x.left=8 x.right=9 y.min=0.0 y.max=.5

#Source Well
doping     uniform reg=1           n.type conc=1e18 x.left=3.16 x.right=3.84 y.min=0.00 y.max=.14

#Substrate Doping
doping     uniform reg=1           P.type conc=1e18 x.left=0 x.right=12.0 y.min=3.5 y.max=4

#########
#Tonyplot
#########

#------------------------
# SET UP MATERIAL & MODEL
#------------------------

#material taup0=2.e-6 taun0=2.e-6
#models    conmob  fldmob  srh  auger  consrh
models conmob fldmob srh bbt.kl auger consrh KLA impact fermidirac

#----------------
# SET UP CONTACTS
#----------------

contact        name = source   current

#-------
# SOLVE
#-------

impact selb
output con.band val.band band.param
solve init

method Newton
log      outf=SIMnoLight.log

save     outfile=initial.str
Tonyplot initial.str

solve    isource=-1e-16
solve    isource=-1e-15
solve    isource=-1e-14
```

```
solve    isource=-1e-13
solve    isource=-1e-12
solve    isource=-1e-11
solve    isource=-1e-10
solve    isource=-1e-9
solve    isource=-1e-8

solve    vdrain=.001
solve    vdrain=.01
solve    vdrain=.1 vstep=0.45  name=drain vfinal=1.0
solve    vdrain=2  vstep=1  name=drain vfinal=52

save     outfile=SIMnoLight.str
tonyplot SIMnoLight.str
tonyplot SIMnoLight.log

#----------------------------------------------------------------
# SECTION 8: TRANSIENT RESPNONCE
#----------------------------------------------------------------
solve    isource=-1e-8


log      outf=PulseSIMnoLight.log
solve    isource=-1e-8 ramptime=0  tstop=1e-9  dt=1e-11
solve    isource=-1.1e-8 ramptime=0  tstop=10e-9 dt=1e-11
solve    isource=-1e-8 ramptime=0  tstop=15e-9 dt=1e-11
tonyplot PulseSIMnoLight.log
quit
#--------------------------------------------------------------------------------
```

A.2 Mixedmode Simulation

The following code imports an output file as a *packaged* device and simulates the device in a spice-like environment. This environment is Silvaco's Mixedmode. The simulation takes a device previous simulated in ATLAS and applies the previous output file to an external circuit. It is capable of DC, AC and transient tests. This code was used to simulate the integration of the buried oxide SIM into a detector circuit.

```
#--------------------------------------------------------------------------------
go atlas
#------------------------------------
# START MIXEDMODE (SPICE)
#------------------------------------
.begin

.log   outfile=Pulse_Cap500fF

#------------------------------------
# LOAD DC SOLUTION/SAVE TRAN SOLUTION
#------------------------------------
.load infile=DCSIM
```

```
.save outfile=TranSIM

#------------------------------------
# NODE SETUP AND CURRENT PULSE
#------------------------------------
vsim    1 0 55
asim    1=drain 2=source 0=ground infile=initial.str
# PULSE SETUP      pulse StartVal EndVal TimeDelay RiseTime FallTime PulseWidth Period
iin     0 2 -1e-6 pulse  -1e-6   -2e-6    10e-15      0       0      5e-3     30e-3
C1      2 0 500f

#------------------------------------
# SETUP CYLINDRICAL AND OPTIONS
#------------------------------------
.options CYLINDR

#------------------------------------
# ALTER DC SETUP (Iinjection)
#------------------------------------
#.DC    name  start  end  steps/increments
#.DC DEC iin   -1e-6 -1e-6      3

#------------------------------------
# TRANSIENT PULSE SETUP
#------------------------------------
.numeric DTMIN=1e-15
#.TRAN  StepIntervals  TotalRunTime
.tran      1e-15          10e-3

#------------------------------------
# END MIXEDMODE (SPICE)
#------------------------------------
.end

#------------------------------------
# MODELS, IMPACT
#------------------------------------
models  device=asim region=1 conmob fldmob srh bbt.kl auger consrh KLA impact fermidirac
impact  device=asim selb region=1


####################
# END SIMULATION
####################
quit
#---------------------------------------------------------------------------
```

A.3 3D Cylindrical Simulations in Silvaco

This code simulates an ohmic contact SIM in 3D. Typical simulations are only done in 2D. In this way it was possible to gauge the effects on space charge for elongated contacts. Sample outputs from this code are shown in after the code.

```
#--------------------------------------------------------------------------------
Go atlas

mesh THREE.D CYLINDRICAL
#MESH
#-----------
#SET UP MESH
#-----------

r.mesh loc=0     spac=1
r.mesh loc=10    spac=1
r.mesh loc=11    spac=1
r.mesh loc=25    spac=1
r.mesh loc=26    spac=1
r.mesh loc=30    spac=1
r.mesh loc=31    spac=1
r.mesh loc=40.0  spac=1

a.mesh loc=0            spac=120
a.mesh loc=360          spac=120

z.mesh loc=-0.2         spac=1
z.mesh loc=0.0          spac=1
z.mesh loc=0.5          spac=1
z.mesh loc=9.5          spac=1
z.mesh loc=10           spac=1
z.mesh loc=10.2         spac=1


#---------------
# Set up Regions
#---------------

region  num=1 material=silicon a.min=0 a.max=360 r.min=0 r.max=40  z.min=0  z.max=10
region  num=2 material=air     a.min=0 a.max=360 r.min=0 r.max=40  z.min=-.2 z.max=0
region  num=3 material=air     a.min=0 a.max=360 r.min=0 r.max=40  z.min=10 z.max=10.2
region  num=4 material=oxide  x.min=3 x.max=7   y.min=0    y.max=1   z.min=0 z.max=10

#---------------------
# SET UP ELECTRODES
#---------------------

elec num=1   name=source   a.min=0.0 a.max=360.0  r.min=0   r.max=10   z.min=-.2  z.max=0
elec num=2   name=drain    a.min=0.0 a.max=360.0  r.min=26   r.max=30   z.min=-.2  z.max=0
elec num=3   name=ground   a.min=0.0 a.max=360.0  r.min=0   r.max=40  z.min=10 z.max=10.2

#--------------
# SET UP DOPING
#--------------
#Epi doping
doping     uniform reg=1           p.type conc=1e15

#Source Well
doping uniform reg=1 n.type conc=1e18 a.min=0 a.max=360 r.min=0 r.max=11   z.min=0.0 z.max=0.5

#Drain Well
doping uniform reg=1 n.type conc=1e18 a.min=0 a.max=360.0 r.min=25 r.max=31 z.min=0.0 z.max=0.5

#Substrate Doping
doping uniform reg=1 p.type conc=1e18 a.min=0 a.max=360 r.min=0 r.max=40 z.min=9.5 z.max=10

#------------------------
```

```
# SET UP MATERIAL & MODEL
#------------------------

#material taup0=2.e-6 taun0=2.e-6

#----------------
# SET UP CONTACTS
#----------------

contact name=source current

#-------
# SOLVE
#-------

impact selb
#output con.band val.band band.param
solve init

method Newton
#DT.MAX=5e-12
log     outf=Sim3Dcylindrical2x.log

save     outfile=initial.str
tonyplot3d initial.str

solve   isource=0
solve   isource=-1e-20
solve   isource=-1e-19
solve   isource=-1e-18
solve   isource=-1e-17
solve   isource=-1e-16
solve   isource=-2e-16
solve   isource=-3e-16
solve   isource=-4e-16
solve   isource=-5e-16
solve   isource=-6e-16
solve   isource=-7e-16
solve   isource=-8e-16
solve   isource=-9e-16
solve   isource=-1e-15
solve   isource=-1e-14
solve   isource=-1e-13
solve   isource=-1e-12
solve   isource=-1e-11
solve   isource=-1e-10
solve   isource=-1e-9
solve   isource=-1e-8
solve   isource=-1e-7

solve    vdrain=0
solve    vdrain=.001
solve    vdrain=.01
solve    vdrain=.1
solve    vdrain=1 vstep=2  name=drain vfinal=165


save     outfile=Sim3Dcylindrical2x.str
tonyplot3d Sim3Dcylindrical2x.str
tonyplot Sim3Dcylindrical2x.log
quit
#----------------------------------------------------------------
```

```
# SECTION 8: TRANSIENT RESPONSE
#------------------------------------------------------------------
solve    isource=-1e-12

log      outf=PulseSim.log
solve    isource=-1e-12          ramptime=0  tstop=1e-10  dt=1e-15
solve    isource=-1e-11          ramptime=0  tstop=10e-10 dt=1e-15
solve    isource=-1e-12          ramptime=0  tstop=15e-10 dt=1e-15
tonyplot3d PulseSim.log
quit
#--------------------------------------------------------------------------------
```

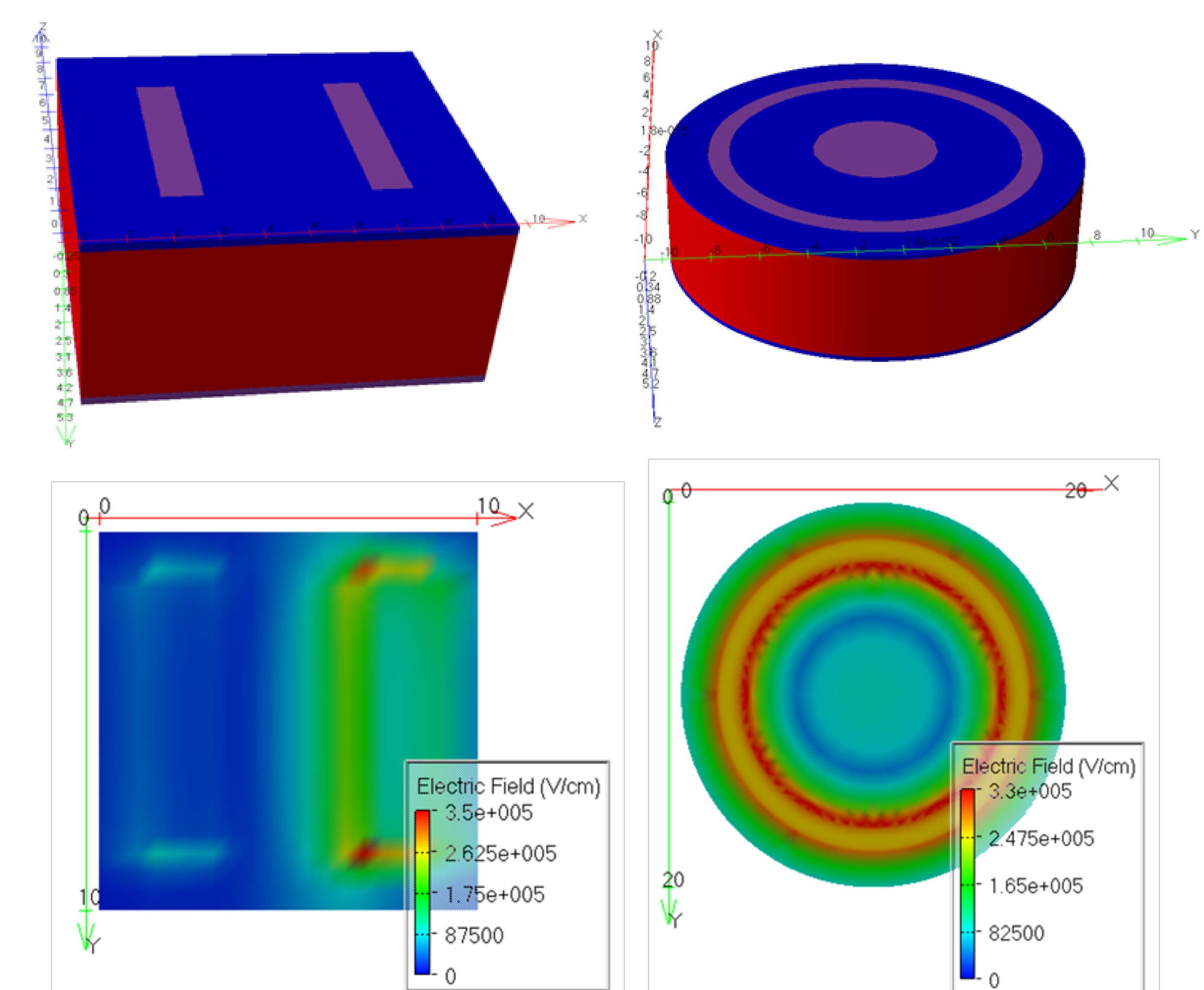


Figure A.1: Examples of 3D simulation done in Silvaco's 3D ATLAS environment. The top row on the left is a rectangular SIM. The top on the right is a circular SIM. The bottom row shows a cut plane of the SIM s on the top row. The cut plane shows the electric filed magnitude at a depth equal to the input and output well depths.

A.4 InP/InGaAs SAM APD Example from Chapter 3

This is the complete code used for the example simulation given in Section 4.4.5 of Chapter 4. The simulation is for an InP/InGaAs SAM APD. The simulation handles many different types of models and commands such as voltage, current, heterostructures, optical absorption, impact ionization etc. For this reason it was used as the example in Chapter 4.

```
#-------------------------------------------------------------------------------
Go Atlas

mesh space.mult=1.0
#-----------------------------------------
#SET UP MESH
#-----------------------------------------

x.mesh loc=0.0        spac=0.5
x.mesh loc=15.5       spac=0.5
x.mesh loc=16.0       spac=0.5
x.mesh loc=20.5       spac=0.5
x.mesh loc=19.0       spac=0.5
x.mesh loc=25         spac=0.5

y.mesh loc=-.5        spac=0.2
y.mesh loc=-.25       spac=0.2
y.mesh loc=0          spac=0.1
y.mesh loc=0.75       spac=0.05
y.mesh loc=1.0        spac=0.05
y.mesh loc=1.15       spac=0.05
y.mesh loc=2.20       spac=0.1
y.mesh loc=2.25       spac=0.1
y.mesh loc=2.3        spac=0.1
y.mesh loc=3.25       spac=0.1
y.mesh loc=5.75       spac=0.25
y.mesh loc=5.8        spac=0.25
#-----------------------------------------
# Set up Regions
#-----------------------------------------
#going from Substrate up

region   num=1 material=InP x.min=0   x.max=25 y.min=2.25       y.max=5.75
region   num=2 material=InGaAs x.min=0 x.max=25 y.max=2.25 y.min=1.15 x.comp=.47
region   num=3 material=InGaAsP        x.min=0 x.max=25           y.max=1.15  x.comp=.47      y.comp=1
grad.34=.15
region   num=4 material=InP   x.min=0  x.max=25          y.min=0  y.max=1
region   num=5 material=Air    x.min=0 x.max=25          y.min=-.5          y.max=0
region   num=6 material=Air    x.min=0 x.max=25          y.min=5.75         y.max=5.8

#-----------------------------------------
# SET UP ELECTRODES
#-----------------------------------------

elec num=1   name=Anode    x.min=16    x.max=19.0        y.min=-0.25        y.max=0.0
elec num=2   name=Cathode  x.min=0.0    x.max=23          y.min=5.75         y.max=5.8
```

```
#-----------------------------------------
# SET UP DOPING
#-----------------------------------------

#InGaAs
doping     uniform reg=2           n.type conc=5e15

#InP
#Avalanche Region(n- Well) doping
doping     uniform reg=4           n.type conc=3e15
doping     uniform reg=4           n.type conc=1e17 x.left=0 x.right=23 y.min=0.75 y.max=1

#(P+ layer) doping
doping     uniform reg=4           p.type conc=1e18 x.left=0 x.right=20 y.min=0.0 y.max=0.25

#(P- Well) doping
doping     uniform reg=4           p.type conc=5e15 x.left=15.5 x.right=20.5 y.min=0.0 y.max=0.5

#Substrate (n-) doping
doping     uniform reg=1           n.type conc=1e16
doping     uniform reg=1           n.type conc=8e16 x.left=0 x.right=23 y.min=3.25 y.max=5.75

#InGaAsP  Graded doping
doping     uniform reg=3           n.type conc=5e15

#Check structure
#-----------------------------------------
# SET UP MATERIAL & MODEL
#-----------------------------------------

Material Material=InP
MATERIAL MATERIAL=InGaAs

models srh bbt.kl auger consrh impact selb

#-----------------------------------------
# SET UP CONTACTS
#-----------------------------------------

contact       name = Anode    Aluminum
contact       name = Cathode  Aluminum

#--------------------------------------------
# Light Source
#--------------------------------------------

beam     num=1 x.origin=5 y.origin=-1.0 angle=90.0 wavelength=1.55  max.window=5

#-----------------------------------------
# SOLVE
#-----------------------------------------

impact selb
output con.band val.band band.param photogen
solve init

method Newton
log      outf=Output.log

save     outfile=initial.str
tonyplot initial.str
```

```
solve    b1=1e-7
solve    b1=1e-6
solve    b1=1e-5
solve    b1=0.001
solve    b1=0.01
solve    b1=0.05
solve    b1=0.075
solve    b1=0.1
solve    b1=0.3
solve    b1=0.5
solve    b1=0.75
solve    b1=1.0

solve    vAnode=0
solve    vAnode=-.001
solve    vAnode=-.01
solve    vAnode=-.1
solve    vAnode=-1 vstep=-1  name=Anode    vfinal=-20

save     outfile=Output.str
tonyplot Output.str
tonyplot Output.log
Quit
#--------------------------------------------------------------------------------
```

B. FABRICATION OF BURIED OXIDE SIM

B.1 SIM Fabrication Procedure for a Wet Etch Buried Oxide SIM

Use the Wacker wafers with a P+ substrate and a P- EPI layer. Make sure they are not reclaim wafers because reclaim wafers have no EPI layer left [70].

B.1.1 KOH Etch Mask, and KOH Etching

1) Create the KOH etch mask by growing 1200 Angstroms of thermal oxide on the wafers. Do it quickly using wet oxide. Make sure the bubbler has a 10:1 ratio of H2O:HCl. This is always critical. It really helps reduce contamination in the wafer. This is manifest by an improvement in leakage currents. To grow the oxide, follow Oxide Growth Chart #1.

2) After oxide growth, dehydrate wafers, apply HMDS and Spin on AZ 2020 at 3000 RPMs. Soft bake for 60 seconds at 110° C on the hotplate. Use N+ well mask and the ISCH recipe on the south aligner. Only change the exposure time in the ISCH program to 20 seconds.

3) DON'T FORGET the post exposure bake before you develop. Post exposure bake on hotplate for 60 seconds at 110° C.

4) Develop wafer in AZ 300 MIF developer. It usually takes awhile sometimes 3 minutes. It requires some eyeballing. The easiest way to tell when the resist is developed is that the small fine tuning alignment marks are developed into the resist.

5) Rinse the wafer well and hard bake on the hotplate at 110° C for 5 minutes.

6) Do an O2 descum in the Planar Etcher (PE2) for 15 seconds at 50watts. Then turn the wafers 180° in the PE2 and etch for another 15 seconds at 50 watts.

7) Remove the thermal oxide to create the KOH etch mask by dipping in Buffered Oxide Etch (BOE) until the surface of the wafer becomes mostly hydrophobic. Then over-etch for another 60 seconds. This over-etching will cause undercutting on the resist and help increase in spacing between the wells thus making devices of smaller spacing useful.

8) Rinse the wafer in De-Ionized Water (DI Water) for at least 60 seconds.

9) Remove the remaining resist in NMP on a hotplate set to 65° C. Allow the NMP about 20 minutes to fully remove the resist. Then rinse in DI Water BEFORE using Isopropyl Alcohol (IPA) or Acetone.

10) Start the KOH etching process by adding DI Water to the KOH etching bath. Fill the water to the level of the KOH in the glassware.

11) Turn on the heat to the bath by powering up the Modutek hot bath controller. The temperature should be preset to 50° C. This is done by pressing the "Timer" button followed by the "Reset" button on the Modutek controller. Any additional information on the controller and bath can be found at http://www.ee.byu.edu/cleanroom/Hot_Pot.phtml a PDF of the manual can be found at http://www.ee.byu.edu/cleanroom/Online_Manuals/Hot_Pot%20Controller.pdf

12) Make sure your KOH solution is about 75% KOH solution from the container and 25% DI Water by volume. Water is important in this whole reaction. Without sufficient amounts of water, the wafer will become pitted and not etch smoothly.

13) Allow about 60 minutes for the bath and KOH to stabilize in temperature. Etch wafers one at a time. Etch for 10 minutes and then check the etching progress with the profilometer. The target height is 3 um. Typically etch to about 3.2 um.

14) After proper pedestal height is achieved rinse the KOH solution off. It is very viscous. Rinse in DI Water for 10 minutes.

15) Etch off remaining KOH etch mask and backing oxide on wafers in BOE.

16) Rinse and get ready for oxide growth.

B.1.2 Oxide Growth and Diffusion Mask Planarization

1) Do a 3 minute, 50 watt descum in the PE2 on the backs and then on the front of the wafers.

2) Quick dip in BOE and rinse in DI Water.

3) Grow 5000 Angstroms of thermal oxide in tube furnace according to Oxide Growth Chart #2. Do not forget the 10:1 DI Water:HCl ratio in the bubbler.

4) After oxide growth, dehydrate wafers apply HMDS and spin on SU-8 3005. BE SURE TO USE THE SMALL CHUCK WHEN YOU SPIN TO KEEP SU8 OFF THE BACK OF THE WAFER. Spin speed for 5000 RPM with a ACL of 1000 for 60 seconds. Then bake for 5 minutes at 65° C then ramp to 95° C and hold for another 5 minutes. Remove wafer and cure resist by exposing wafer under the south aligner for 60 seconds.

5) Hard bake resist with 5 minutes at 65° C then ramp to 95° C and hold for 5 minutes. Then ramp to 110° C and hold for 5 minutes then to 150° C for 5 minutes THEN ramp to 200° C for 10 minutes. Then turn off hotplate an allow wafers to cool on cooling hotplate until hotplate temperature is below 90° C.

6) Remove wafers from hotplate and get ready to polish with the CMP.

7) The CMP directions are found at http://www.ee.byu.edu/cleanroom/cmp.phtml

8) Use the SU8-5 recipe in the CMP and the Eminess Ultrasol A-15 slurry. It works most excellently to polish SU8. Dilute the slurry about 10:1 or 5:1 DI Water to slurry. The 5 gallon dilute slurry bucket is labeled "SIM Group A-15 DIL." The concentrated slurry bucket is labeled "SIM Group."

9) Be sure to stir the diluted slurry while polishing to that the mixture does not settle.

10) Polish the SU8 down until the tops of the devices pedestals have the oxide polished down to bare silicon.

11) Rinse polished wafers in DI Water. Place in Nanostrip on a hot plate at 90° C for at least 1 hour.

B.1.3 Diffusion

1) Remove wafers from Nanostrip. Rinse thoroughly in DI Water. Dehydration bake the wafers in the oven at 150° C for 30 minutes. Remove wafers and do a quick BOE dip. Rinse in DI Water and place in the PE2 chamber and vacuum down. Do NOT activate the plasma. This vacuum chamber is merely to pull water out of the wafer without growing an oxide on the wafer surface.

2) Remove the Honeywell P-8548 Spin on Dopant (SOD) from the refrigerator at least 4 hours before applying the SOD to the wafers.

3) Use Pipets and the Headway spinner to apply SOD. Use a pipet to cover 80% of the wafer with SOD then spin the wafer at 500 RPM for about 2 seconds and slowly ramp to 3000 RPM. After 20 seconds, remove the wafer and place it on a hotplate at 150° C for 60 seconds.

4) Diffuse wafers in the furnace according to Diffusion Chart #1.

5) Immediately etch remaining SOD from the wafers after they come out of the furnace. It should only take about 45 seconds in BOE. Then rinse in DI Water.

6) You can check if the glass is removed by applying voltage to the N+ wells and substrate and observing the current.

B.1.4 Metallization

1) Dehydrate wafers, apply HMDS, and a thick layer of AZ 3330 photoresist. This is done by using the small chuck on the spinner so the resist does not seep onto the back of the wafer and using a max spin speed of 3000 RPM with ACL of 500. This helps the resist clear the pedestals.

2) Soft bake the wafer for 12 minutes at 90° C on a hotplate. Expose on the south aligner with metal mask on the ISCH program. Edit the program and change the exposure time to 20 seconds.

3) Develop in AZ 300 MIF developer. This takes several minutes. Verify development under the microscope and ensure that the resist has fully developed

4) Descum in the PE2 30 seconds at 50 watts and turn the wafers 180° and do another 30 second descum at 50 watts.

5) Prepping the wafer for metallization needs to be done right before placing the wafer into the sputterer.

6) Quick dip in BOE and rinse in DI water. Then dip into a 1:1 Di Water:HCl solution for 30 seconds. Rinse and dry and place directly into the sputterer.

7) Pump the sputter to 4 e-6torr. Then deposit TiW on Cathode 3 DC sputtering at 150mA current for 1000 seconds. This gives about 70nm for TiW as a barrier metal.

8) Vent the sputterer and place the wafer into the thermal evaporator and deposit 1 um of aluminum on top of the TiW.

9) Vent the evaporator and flip the wafers to deposit 300nm of aluminum on the back side.

10) Lift off the front metal in acetone. This requires a fair amount of scrubbing with a foam tip to remove the metal. Rinse in IPA and acetone and blow dry.

11) Anneal in forming gas environment at 450° C for 30 minutes. Run the forming gas on full flow on the pin gauge.

12) The wafer is finished.

B.2 SIM Fabrication Procedure for a Dry Etch Buried Oxide SIM

Use the Wacker or Montco wafers with a P+ substrate and a P- EPI layer. Make sure they are not reclaim wafers because reclaim wafers have had the EPI layer polished off.

B.2.1 RIE, ICP Etching

1) Spin SU-8 2002 onto the wafers and soft bake. Then expose the wafers with the first N+ well mask for 8 seconds. Don't forget the post-exposure bake!

2) Develop in SU-8 developer and rinse in IPA and then blow dry.

3) Descum wafers at 150 W for 2 minutes in the PE2.

4) RIE etch wafers in the Trion etcher one at a time. Each wafer requires several recipes to etch properly. Training is required!

5) Make sure the Trion chamber is clean by running the "Clean-New" recipe for six minutes on both steps.

6) Load the wafer and etch it with the "Evan-Bosch-Si3" etch. This etch take about 20 minutes and must be run twice to achieve a pedestal height of roughly 2.5 um.

7) Following the etching of one wafer, run the "Clean-New" etch as described previously to clean the chamber.

8) Load the etched wafers into the PE2 and do another descum etch at 150 W for 2 minutes. This removes any teflons that have built up on the wafer.

B.2.2 Oxide Growth and Diffusion Mask Planarization

1) Quick dip in BOE and rinse in DI Water.

2) Grow 5000 Angstroms of thermal oxide in tube furnace according to Oxide Growth Chart #2. Do not forget the 10:1 DI Water:HCl ratio in the bubbler.

3) After oxide growth, dehydrate wafers apply HMDS and spin on SU-8 3005. BE SURE TO USE THE SMALL CHUCK WHEN YOU SPIN TO KEEP SU8 OFF THE BACK OF THE WAFER. Spin speed for 5000 RPM with an ACL of 1000 for 60 seconds. Then bake for 5 minutes at 65° C then ramp to 95° C and hold for another 5 minutes. Remove wafer and cure resist by exposing wafer under the south aligner for 60 seconds.

4) Hard bake resist with 5 minutes at 65° C then ramp to 95° C and hold for 5 minutes. Then ramp to 110° C and hold for 5 minutes then to 150° C for 5 minutes THEN ramp to 200° C for 10 minutes. Then turn off hotplate an allow wafers to cool on cooling hotplate until hotplate temperature is below 90° C.

5) Remove wafers from hotplate and get ready to polish with the CMP.

6) The CMP directions are found at http://www.ee.byu.edu/cleanroom/cmp.phtml

7) Use the "SU8-5" recipe in the CMP and the Eminess Ultrasol A-15 slurry. It works most excellently to polish SU8. Dilute the slurry about 10:1 or 5:1 DI Water to slurry. The 5 gallon dilute slurry bucket is labeled "SIM Group A-15 DIL." The concentrated slurry bucket is labeled "SIM Group."

8) Be sure to stir the diluted slurry while polishing to that the mixture does not settle.

9) Polish the SU8 down until the tops of the devices pedestals have the oxide polished down to bare silicon.

10) Rinse polished wafers in DI Water. Place in Nanostrip on a hot plate at 90° C for at least 1 hour.

B.2.3 Diffusion

1) Remove wafers from Nanostrip. Rinse thoroughly in DI Water. Dehydration bake the wafers in the oven at 150° C for 30 minutes. Remove wafers and do a quick BOE dip. Rinse in DI Water and place in the PE2 chamber and vacuum down. Do NOT activate the plasma. This vacuum chamber is merely to pull water out of the wafer without growing an oxide on the wafer surface.

2) Remove the Honeywell P-8548 Spin on Dopant (SOD) from the refrigerator at least 4 hours before applying the SOD to the wafers.

3) Use Pipets and the Headway spinner to apply SOD. Use a pipet to cover 80% of the wafer with SOD then spin the wafer at 500 RPM for about 2 seconds and slowly ramp to 3000 RPM. After 20 seconds, remove the wafer and place it on a hotplate at 150° C for 60 seconds.

4) Diffuse wafers in the furnace according to Diffusion Chart #1.

5) Immediately etch remaining SOD from the wafers after they come out of the furnace. It should only take about 45 seconds in BOE. Then rinse in DI Water.

6) You can check if the glass is removed by applying voltage to the N+ wells and substrate and observing the current.

B.2.4 Metallization

1) Dehydrate wafers, apply HMDS, and a thick layer of AZ 3330 photoresist. This is done by using the small chuck on the spinner so the resist does not seep onto the back of the wafer and using a max spin speed of 3000 RPM with ACL of 500. This helps the resist clear the pedestals.

2) Soft bake the wafer for 12 minutes at 90° C on a hotplate. Expose on the south aligner with metal mask on the ISCH program. Edit the program and change the exposure time to 20 seconds.

3) Develop in AZ 300 MIF developer. This takes several minutes. Verify development under the microscope and ensure that the resist has fully developed

4) Descum in the PE2 30 seconds at 50 watts and turn the wafers 180° and do another 30 second descum at 50 watts.

5) Prepping the wafer for metallization needs to be done right before placing the wafer into the sputterer.

6) Quick dip in BOE and rinse in DI water. Then dip into a 1:1 Di Water:HCl solution for 30 seconds. Rinse and dry and place directly into the sputterer.

7) Pump the sputter to 4e-6torr. Then deposit TiW on Cathode 3 DC sputtering at 150mA current for 1000 seconds. This gives about 70nm for TiW as a barrier metal.

8) Vent the sputterer and place the wafer into the thermal evaporator and deposit 1 um of aluminum on top of the TiW.

9) Vent the evaporator and flip the wafers to deposit 300nm of aluminum on the back side.

10) Removing metal from a liftoff photoresist mask can be difficult when a planetary is used. The planetary is necessary to coat the sidewalls of the pedestals. Tape liftoff on the front of the

wafer is most effective. Take a strip of scotch tape long enough to traverse the wafer and press firmly onto the wafer, then pull off. Repeat this until the whole wafer has been done.

11) Lift off any remaining metal in acetone. Rinse in IPA and acetone and blow dry.

12) Anneal in forming gas environment at 450° C for 30 minutes. Run the forming gas on full flow on the pin gauge.

13) The wafer is finished.

B.2.5 Process Charts

The following charts are used for oxide growths and doping diffusion in the Bruce furnace [70].

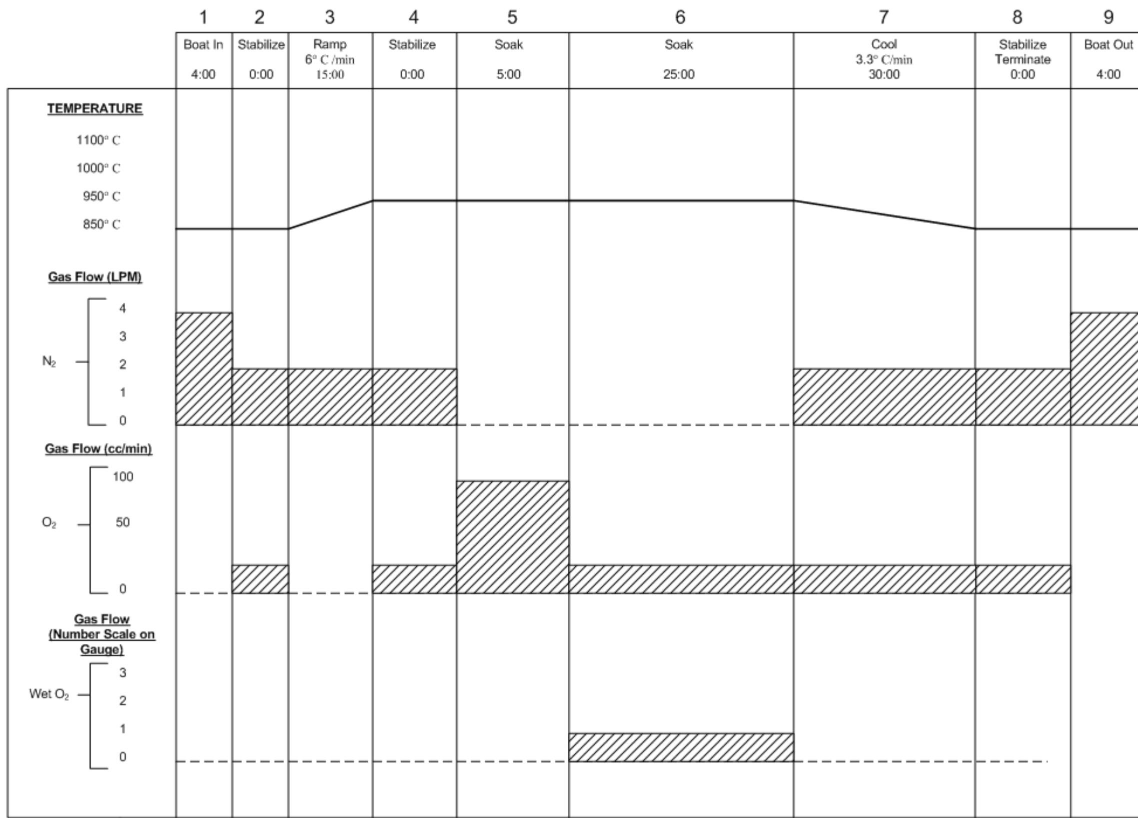


Figure B.1: SIM Oxide Growth Chart #1: This chart shows the necessary step gas flows, step temperatures, and step durations to grow roughly 1200 angstroms of thermal oxide to act as a KOH etch mask in subsequent steps. The low temperature provides the necessary oxide grow while minimizing the amount of diffusion within the EPI layer.

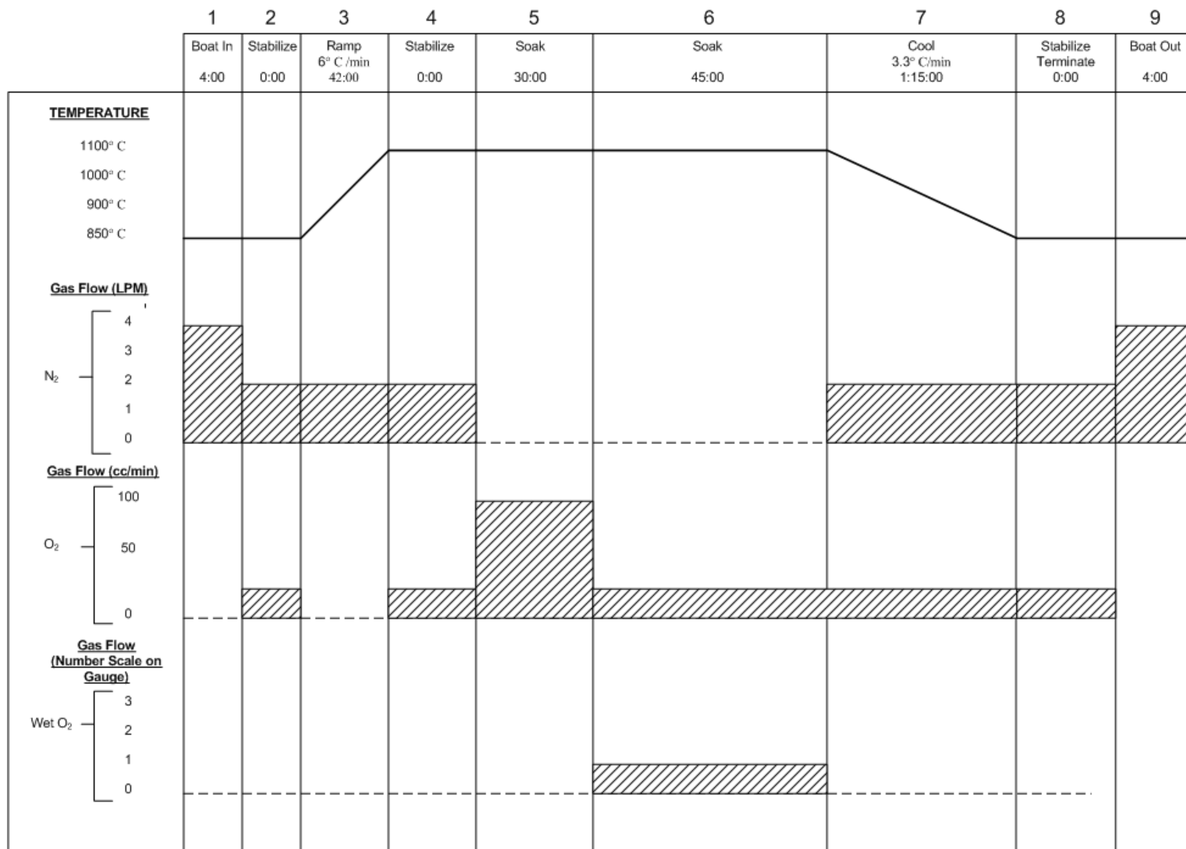


Figure B.2: SIM Oxide Growth Chart #2: This chart shows the necessary step gas flows, step temperatures, and step durations to grow 5000 angstroms of thermal oxide. This oxide acts as an insulator and passivation layer for the SIM.

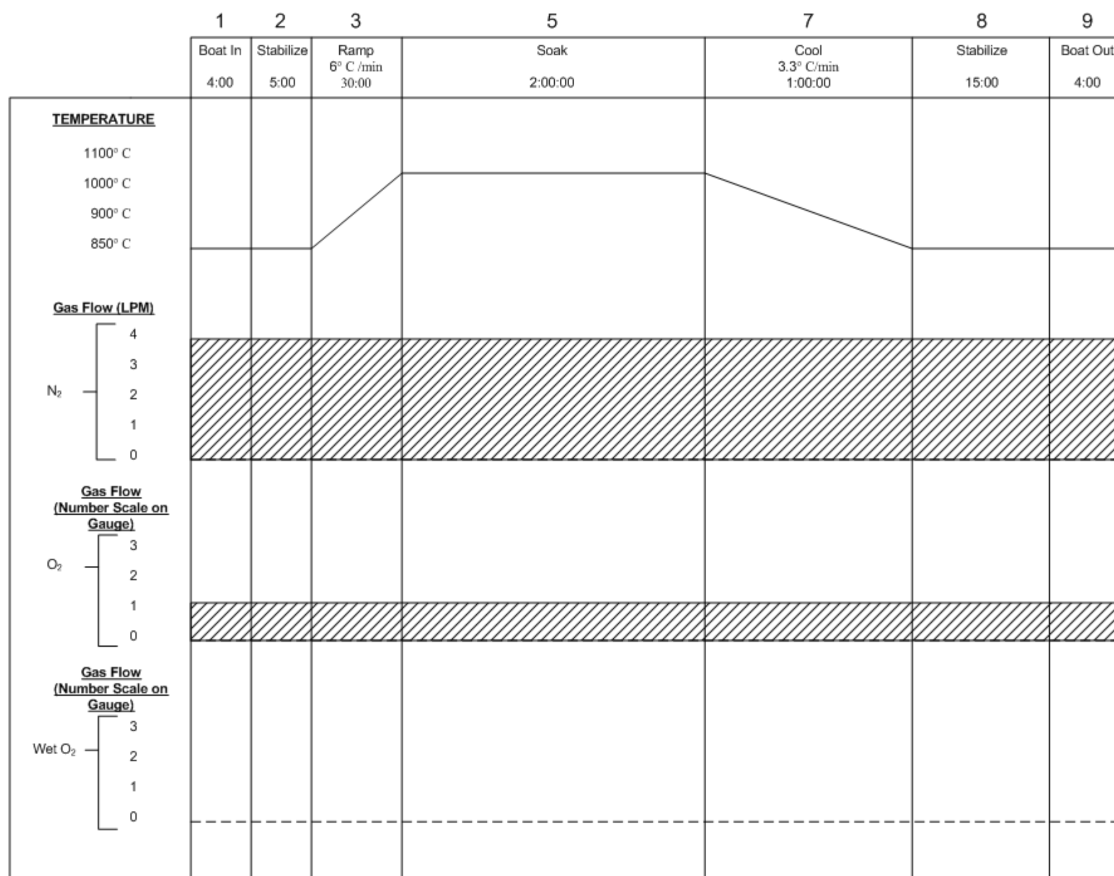


Figure B.3: SIM Diffusion Chart #1 for P-8545 Spin on Dopant (SOD): This chart shows the necessary step gas flows, step temperatures, and step durations to diffuse phosphorus into the SIM pedestals to a depth of roughly 1.3 um. It is important that sufficient oxygen is always flowing into the furnace during the diffusion. Oxygen allows the phosphorus in the SOD to diffuse by turning the phosphoric acid (H_3PO_4) within the mixture into Phosphorus Pentoxide (P_2O_5). It is important to completely oxidize all the H_3PO_4 within the SOD otherwise it will form a bond with the silicon that cannot be removed unless it is first oxidized. Following diffusion, it is important to immediately dip the wafers in Buffered Oxide Etch (BOE) to completely remove the SOD before the P_2O_5 reacts with the moisture in the air to form H_3PO_4.