Rochester Institute of Technology
RIT Digital Institutional Repository

Theses

2012

A Silicon p-i-n detector for a hybrid CMOS imaging system

Christopher Shea

Follow this and additional works at: https://repository.rit.edu/theses

Recommended Citation
Shea, Christopher, "A Silicon p-i-n detector for a hybrid CMOS imaging system" (2012). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact repository@rit.edu.

A Silicon *p-i-n* Detector for a Hybrid CMOS Imaging System

Christopher G. Shea

A Thesis Submitted
In Partial Fulfillment
of the Requirements of the Degree of
Master of Science
in Microelectronic Engineering

Approved by:

Professor ____
Dr. Karl D. Hirschman (Thesis Advisor)

Professor ____
Dr. Alan Raisanen (Thesis Committee Member)

Professor ____
Dr. Donald Figer (Thesis Committee Member)

Professor ____
Dr. Michael Jackson (Thesis Committee Member)

Professor ____
Dr. Robert Pearson (Program Director)

Professor ____
Dr. Sohail Dianat (Deparment Head)

DEPARTMENT OF ELECTRICAL AND MICROELECTRONIC ENGINEERING
COLLEGE OF ENGINEERING
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK
FEBRUARY 2012

IPR2025-00845
Apple-Sony EX1026 Page 2

Acknowledgment

I would like to thank Dr. Karl D. Hirschman for his steadfast support and continual guidance throughout my collegiate experience. After first taking me on as a student researcher, and then continuing as an advisor on both my senior design and thesis projects, I am eternally grateful to Dr. Hirschman for his support of my academic career. I would like to thank Dr. Don Figer for entrusting me with the completion of p-*i*-n image sensor project. Al Raisanen, Brian Ashe and Nancy Stoffel are all greatly appreciated for their support and advice throughout the project. I am grateful to have been surrounded by a bright class of graduate students to provide stimulating discussion and companionship throughout the long nights in the cleanroom. I would also like to specifically thank Jon-Paul DesOrmeaux for his continued moral, emotional, physical, social, and psychological support as only a true friend can provide, and especially my mother for none of this would be possible without her eternal love and support.

Support for this project is provided by the National Aeronautics and Space Administration (NASA) under Grant NNX07AG99G, issued through the Astronomy and Physics Research and Analysis (APRA) Program of the Science Mission Directorate to the Rochester Image Detector Laboratory (RIDL) at the Rochester Institute of

IPR2025-00845
Apple-Sony EX1026 Page 3

Technology (RIT). All simulations were performed in the Very Large Scale Integration (VLSI) Laboratory and all device fabrication was completed in the Semiconductor & Microsystems Fabrication Laboratory (SMFL), both at RIT. All photo-masks for the project were written in house at the SMFL, thanks to Mr. Thomas Grimsley.

Deposition of indium as the bump bond metal was performed at Jet Process Corporation (JPC) in North Haven, CT using the jet deposition process. Lift-off and flip-chip bonding were performed at Smart System Technology & Commercialization Center (STC) in Canandaigua, NY, as well as final packaging and wire-bonding.

IPR2025-00845
Apple-Sony EX1026 Page 4

Abstract

A fully depleted silicon p-*i*-n image sensor for a very low noise hybrid CMOS imaging system was simulated, fabricated, and electrically characterized. The image sensor was then bonded to the foundry fabricated CMOS circuitry to create the imaging system. SILVACO Atlas was used to simulate the steady state electrical operation of the device as well as the optical response. Revisions were made to an existing mask set to allow the use of both contact and projection lithography in the fabrication process. Significant process improvements were introduced to eliminate needless complexity and reduce leakage current from the previously reported 1.5×10^{-6} A/cm^2 below the goal of 2.2×10^{-9} A/cm^2. Following fabrication of the image sensors, electrical testing was performed to verify diode quality from leakage and lifetime measurements. A lift-off process was developed for thick metal layers used in the bump-bond hybridization process. Daisy-chain test parts were created to characterize the mechanical and electrical connections formed in the hybridization process. Fabricated p-*i*-n photodiode arrays were diced and hybridized to read-out integrated circuits using a flip-chip bump bond process with indium interconnects. Testing of hybridized devices is currently ongoing.

IPR2025-00845
Apple-Sony EX1026 Page 5

Table of Contents

Title Page i
Acknowledgment ii
Abstract iv
Table of Contents v
List of Tables.................... viii
List of Figures ix
1. Introduction....................1
1.1. A Review of Photodetectors....................2
1.2. A Historical Perspective....................3
1.3. Detector Materials....................4
1.4. Image Sensor Architectures....................6
1.4.1 CCD and CMOS Pixel Architectures 7
1.4.2 System Architecture: Hybrid vs. Monolithic....................9
1.4.3 Hybrid Image Detectors....................12
1.5. APRA Imaging System 13
1.5.1 APRA Hybrid Imaging System 14
1.5.2 MOSIS Read-out Integrated Circuit (ROIC)....................16
1.5.3 Silicon p-i-n Photodiode Array....................17
2. *p-i-n* Photodiode Operation....................20
2.1. Junction Electrostatics....................20
2.1.1 Current-Voltage Characteristics 22
2.1.2 Minority Carrier Lifetime....................26
2.1.3 Open-Circuit Voltage Decay (OCVD) Method....................27
2.1.4 Reverse Recovery Method....................28
2.2. Absorption and Photogeneration....................29
2.3. Illuminated Operation 33
2.3.1 Charge Collection....................34
2.3.2 Quantum Efficiency....................35
2.3.3 Point Spread Function 35
3. SILVACO Atlas TCAD Simulations 37
3.1. Numerical Poisson Solver Theory....................37
3.1.1 Electrostatic Equations 38

3.1.2 Continuity Equations 38
3.1.3 Drift-Diffusion Transport Equations 39
3.2. Atlas Simulation Organization 39
3.2.1 Device Structure Specification 41
3.2.2 Material Models..44
3.2.3 Numerical Methods 48
3.2.4 Solutions Specification 50
3.3. Simulation Results..52
3.3.1 Current Voltage Characteristics..52
3.3.2 Quantum Efficiency..56
3.3.3 Point Spread Function 64
4. Layout and Fabrication..70
4.1. Existing Layout 70
4.2. Mask Design..71
4.2.1 Revised Test Die..72
4.3. Process Flow 73
4.3.1 Front–to–Backside Alignment..75
4.3.2 Contact Cut Etch..78
4.3.3 Aluminum Dry Etch 80
5. Electrical Testing..84
5.1. Previous Results 84
5.2. Current-Voltage Characteristics 86
5.2.1 On-State Ideality..87
5.2.2 Off-State Leakage..89
5.3. Junction Lifetime Measurements 93
5.4. Spectral Response Measurements 95
6. Hybridization..96
6.1. Lift-Off Process Development 97
6.1.1 Previous Development Work..99
6.1.2 Initial Investigation..102
6.1.3 SC 1827 on LOR-30A Development..104
6.1.4 Indium Deposition 106
6.1.5 SC 1827 on Double-Coat LOR-30A..108
6.2. ROIC Passivation Cut 110
6.3. Bump-Bond Daisy Chain Test Parts 113
6.3.1 Physical Testing..115

IPR2025-00845
Apple-Sony EX1026 Page 7

6.3.2 Electrical Testing 118
6.4. Final Hybridized Devices 119
7. Final Remarks 121
Appendix I Typical Atlas Input File 123
Appendix II Process Flow 125
References 129

List of Tables

TABLE 1 SUMMARY COMPARISON OF CCD AND CMOS TECHNOLOGIES [8] 8

TABLE 2 IMAGING SYSTEM PERFORMANCE GOALS [14] 14

TABLE 3 COMPARISON OF REVERSE BIAS LEAKAGE GOALS, SIMULATIONS AND EXPERIMENTAL RESULTS 92

List of Figures

Fig. 1.1 Data showing a) number of pixels and b) pixel area for imaging sensors reported in IEEE publications [3] 4

Fig. 1.2 Optical absorption coefficients for various semiconductor materials [4] 5

Fig. 1.3 Schematic representation of CCD and CMOS image sensors highlighting the differences of operation taken from [7] 8

Fig. 1.4 Typical monolithic CMOS image sensor [11] 10

Fig. 1.5 Schematic representation of front-side illumination (FSI) vs back-side illumination (BSI) [12] 11

Fig. 1.6 Cross-section of hybrid imager used in ATLAS detector [13] 13

Fig. 1.8 Illustration of Hybridized Image Detector cross-section 15

Fig. 1.9 Top Down Illustration of Hybridized Image Detector 16

Fig. 1.10 Optical Micrograph of ROIC die 17

Fig. 1.11 Pixel Design and Layout 17

Fig. 1.12 Layout of 256x128 p-i-n photodiode array with guard ring 18

Fig. 2.1 Current-voltage characteristics of a practical Si diode showing ideal and experimental behavior in forward and reverse modes of operation [4] 22

Fig. 2.2 Lifetime (τ_p) and Diffusion length (L_p) of Holes in n-type Si as a function of Donor density [17] 25

Fig. 2.3 Current and voltage transients observed in the methods used for carrier 27

Fig. 2.4 Complex refractive index vs. wavelength for silicon 29

Fig. 2.5 Reflectance and Transmittance at the air-silicon interface 30

IPR2025-00845
Apple-Sony EX1026 Page 10

Fig. 2.6 Absorption coefficient and penetration depth vs. wavelength for silicon 31

Fig. 3.1 Atlas Input Code Organizational Structure 39

Fig. 3.2 Atlas example mesh definition code 41

Fig. 3.3 Atlas simulation code defining electrode name as positions and doping profiles 42

Fig. 3.4 Atlas code defining material and interface properties and specifying numerical models 43

Fig. 3.5 Flow diagram for Newton method [24]....47

Fig. 3.6 Atlas specification of solution method and parameters....48

Fig. 3.7 Atlas solution specification code sample 49

Fig. 3.8 Simulated *p-i-n* diode forward bias current density 53

Fig. 3.9 Simulated *p-i-n* diode reverse bias leakage current vs. applied bias 54

Fig. 3.10 Simulated reverse bias current density at 50 V vs. minority carrier lifetime 55

Fig. 3.11 Typical quantum efficiency curve from SILVACO Atlas 56

Fig. 3.12 QE curves for surface grid resolution from 0.001 to 0.1 μm....57

Fig. 3.13 Simulated QE curves for surface dopant concentrations ranging from $1x10^{17}$ cm^{-3} to $1x10^{21}$ cm^{-3} 58

Fig. 3.14 Simulated QE curves for $n^+ - n^-$ homo-junction depths of 0.3 – 1.0 μm 59

Fig. 3.15 Dopant and electron concentration profiles on *n*-type light absorbing side 60

Fig. 3.16 Simulated QE curves for operating biases from 1 to 100 V....61

Fig. 3.17 QE curves for detector thicknesses from 10 to 300 μm 62

Fig. 3.18 Point spread function simulations showing current through pixels as a function of the source illumination position 63

Fig. 3.19 Current through pixels overlaid with maximums centered 64

IPR2025-00845
Apple-Sony EX1026 Page 11

Fig. 3.20 Simulated current vs. illumination position for 50, 75, 100, 125 & 150 V 65

Fig. 3.21 FWHM of Simulated PSF for 100 – 1200 nm and 2 – 6 kV/cm at 300 K 66

Fig. 3.22 Full width half max of point spread function vs. wavelength at 2 kV/cm for 200 K and 300 K........ 67

Fig. 3.23 Simulated hole concentration distributions for centered point source illumination at 50 V reverse bias and 300 K for various wavelengths in 5 pixel structure 68

Fig. 4.1 CAD rendering of contact lithography mask layout of (a) entire wafer and (b) close-up of test structures 70

Fig. 4.2 CAD renderings of **(a)** projection die and **(b)** contact mask layout 71

Fig. 4.3 CAD renderings of **(a)** Process characterization test die and **(b)** Pixel characterization test die 72

Fig. 4.4 Summary of key steps in revised process flow 73

Fig. 4.5 Comparison of current and newly developed front-to-back side alignment process........ 75

Fig. 4.6 Optical Micrographs of Metal layer pattern after **(a)** contact lithography **(b)** projection lithography 77

Fig. 4.7 An optical micrograph of the array after contact cut etch where the failure mechanism can be observed 78

Fig. 4.8 Optical mirograph of features in LOR layer exhibiting blow-out and scumming........ 79

Fig. 4.9 Patterned aluminum layer on bump bond side of device showing thin grid lines due to undercutting during wet etch process........ 80

Fig. 4.10 Resolution features etched into aluminum layer using suggested recipe. Note significant undercut of resist features 81

Fig. 4.11 Patterned aluminum pixel contacts dry etched with optimized recipe 82

Fig. 5.1 Representative current-voltage characteristics of *p-i-n* diode fabricated by Kolb with ideality fits [26]........ 84

Fig. 5.2 Reverse bias leakage current densities for several devices fabricated by Kolb 85

IPR2025-00845
Apple-Sony EX1026 Page 12

Fig. 5.3 Measured forward bias I-V characteristics of rev3 *p-i-n* diode with exponential fits and resulting ideality factors 86

Fig. 5.4 Forward bias I-V characteristics of 25X p-*i*-n diode measured from 255 K to 300 K showing temperature dependence 87

Fig. 5.5 Overlay of reverse bias leakage currents from best wafer of the final two processing runs 88

Fig. 5.6 Wafer map of leakage current uniformity for final two processing run. 89

Fig. 5.7 Measure reverse bias leakage current vs. applied potential from 260 to 350 K 90

Fig. 5.8 Reverse bias leakage current density at 50 V from 250 to 350 K with exponential extrapolation down to 200 K 91

Fig. 5.9 Diode junction voltage decay with exponential fit 93

Fig. 5.10 Short circuit spectral response of *p-i-n* diode 94

Fig. 6.1 Schematic representation of lift-off and flip chip bump bonding technique used for hybridization process 96

Fig. 6.2 Scanning electron micrograph of bump bond opening in bi-layer resist stack of HPR504 on LOR-30A [26] 99

Fig. 6.3 SEM cross-section of bump-bonds after deposition of **(a)** gold-tin **(b)** indium 100

Fig. 6.4 SEM of bump-bond array after lift-off (a) gold-tin(b) indium 101

Fig. 6.5 SEM cross-section of resist profile after development for various processes 102

Fig. 6.6 SEM cross-section and correlating optical micrograph of bump-bond 104

Fig. 6.7 SEM of new thick bi-layer LOR process after deposition of 3 μm of indium 105

Fig. 6.8 SEM of several 3 μm tall indium bumps after lift-off of excess material 106

Fig. 6.9 SEM of bi-layer resist process with two LOR layers 107

Fig. 6.10 SEM of thick bi-layer resist process after 4.6 μm of indium deposition 108

Fig. 6.11 Optical micrograph of etched and un-etched bump bond pads 111

IPR2025-00845
Apple-Sony EX1026 Page 13

Fig. 6.12 SEM cross-section of ROIC die through etched bond pad....112

Fig. 6.13 Screen capture of bump-bond daisy chain ROIC test part VLSI layout....114

Fig. 6.14 Infra-red optical micrograph of the corner of the bump bond array....116

Fig. 6.15 Optical micrograph of 5 μm thick indium bumps after bonding and separation....116

Fig. 6.16 SEM of 1 μm indium bump after bonding and separation....117

Fig. 6.17 Measured resistance of columns of 84 bump bonds....118

Fig. 6.18 Image of a hybrid *p-i-n* CMOS imaging system wire-bonded in DIP....119

IPR2025-00845
Apple-Sony EX1026 Page 14

Chapter 1

Introduction

The 2009 Nobel Prize in Physics was awarded to Willard S. Boyle and George E. Smith from Bell Labs for their 1969 invention of the charge-coupled device (CCD), a solid state image sensor that has led to many scientific discoveries and consumer applications. Image sensor technology has evolved dramatically throughout the years, introducing new technologies such as active pixels, complimentary metal-oxide-semiconductor (CMOS) image sensors and more recently the hybrid detector [1-8].

An imaging system contains several key components, including a photodetector, read-out circuitry, and a collection of optical elements. The photodetector is a transducer designed to convert an optical signal into an electrical signal. The material must have desirable electrical properties that can be modified by an optical signal. Semiconductors are perfectly suited to this task due to the profound effect electromagnetic radiation has on the material. The read-out circuitry accesses many different elements of a large array and conditions the signal for output. The optical system is responsible for collecting the light in the desired field of view, and focusing this light onto the image sensor, or focal plane.

IPR2025-00845
Apple-Sony EX1026 Page 15

Image sensors have become an invaluable tool to astronomical research enabling the discovery of new phenomena and the confirmation of models. Imaging systems operating in the harsh environments of outer space must be able to perform at cryogenic temperatures and withstand high energy radiation. This thesis continued the development of a fully depleted silicon p-*i*-n image sensor and hybridization process for a hybrid CMOS focal plane array enabling future NASA space missions.

1.1. A Review of Photodetectors

There are several types of semiconductor based photodetectors including photoconductors, photodiodes, charge-coupled devices, and phototransistors. The devices are classified by their structure and the principles upon which they operate. In a photoconductor a single piece of homogeneous semiconducting material is contacted by ohmic connections and the resistivity is modulated by the optical signal. Photodiodes consist of a metallurgical junction in a semiconducting material that is reverse biased to sweep out photo-generated carriers to the collection terminals. Charge-coupled devices are a special type of detector where an array of gated capacitors is incorporated to collect photogenerated carriers. These carriers are then transferred to the edge of the device and read out as the charge is converted to a voltage. The phototransistor is a device that uses

IPR2025-00845
Apple-Sony EX1026 Page 16

an optical signal to modulate gain of the transistor. Each device has advantages and disadvantages associated with the design and are suited to different applications.

A silicon p-*i*-n photodiode is similar to the ubiquitous p-n diode, except a nearly intrinsic region exists between the two highly doped terminals. In the reverse bias mode of operation, the entire intrinsic region is typically depleted providing a large volume for the photogeneration of carriers. One of the main advantages of the p-*i*-n photodiode is the thickness of the depleted intrinsic region can be tailored to optimize absorption at a desired wavelength. The absence of impurities in the depleted region and the use of high quality float zone silicon allow devices with extremely low leakage levels to be realized.

1.2. A Historical Perspective

Image sensors have developed very rapidly over the past five decades, helped enormously by the vast amount of research performed by the semiconductor industry. Using processes developed for integrated circuit manufacturing the image sensor industry has been able to shrink pixel size and increase total pixel counts at rates comparable to transistors in the IC industry. Luppino and Burke even created their own version of 'Moore's Law' stating pixel count and density would double every 2.5 years [2]. A graphical illustration of the trends can be seen in Fig. 1.1 a) and b) showing the pixel count and area for image sensors reported in IEEE publications [3].

IPR2025-00845
Apple-Sony EX1026 Page 17

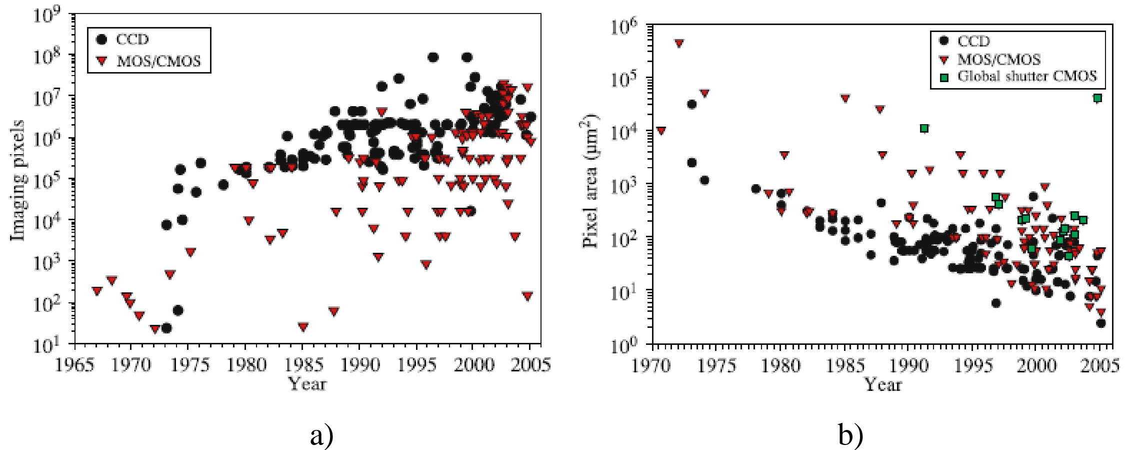

Fig. 1.1 Data showing a) number of pixels and b) pixel area for imaging sensors reported in IEEE publications [3]

Like Moore's Law, this trend will eventually be faced by fundamental limitations restricting further growth. The number of pixels in a device is limited by the size of each pixel and the substrates used for manufacturing. Currently, most high grade scientific imagers are produced on 150 mm substrates, leaving the industry room for growth up to the current IC industry standard of 300 mm. The large areas required by the detectors however make defect densities a primary concern. The size of individual pixels is limited by the volume of detecting material required to produce a suitable signal for the sensing circuitry. The ability to prevent bleeding of a signal into adjacent pixels also limits the reduction of pixel area and becomes more difficult with increasing detector thickness.

1.3. Detector Materials

There are many different semiconducting materials that can be used as photodetectors. The structure of the material defines the fundamental properties of what

IPR2025-00845
Apple-Sony EX1026 Page 18

types of radiation a material absorbs and how strongly it does so as shown by the absorption coefficient. An excellent illustration of the optical properties of various semiconductor materials shown in Fig. 1.2 is taken from Sze.

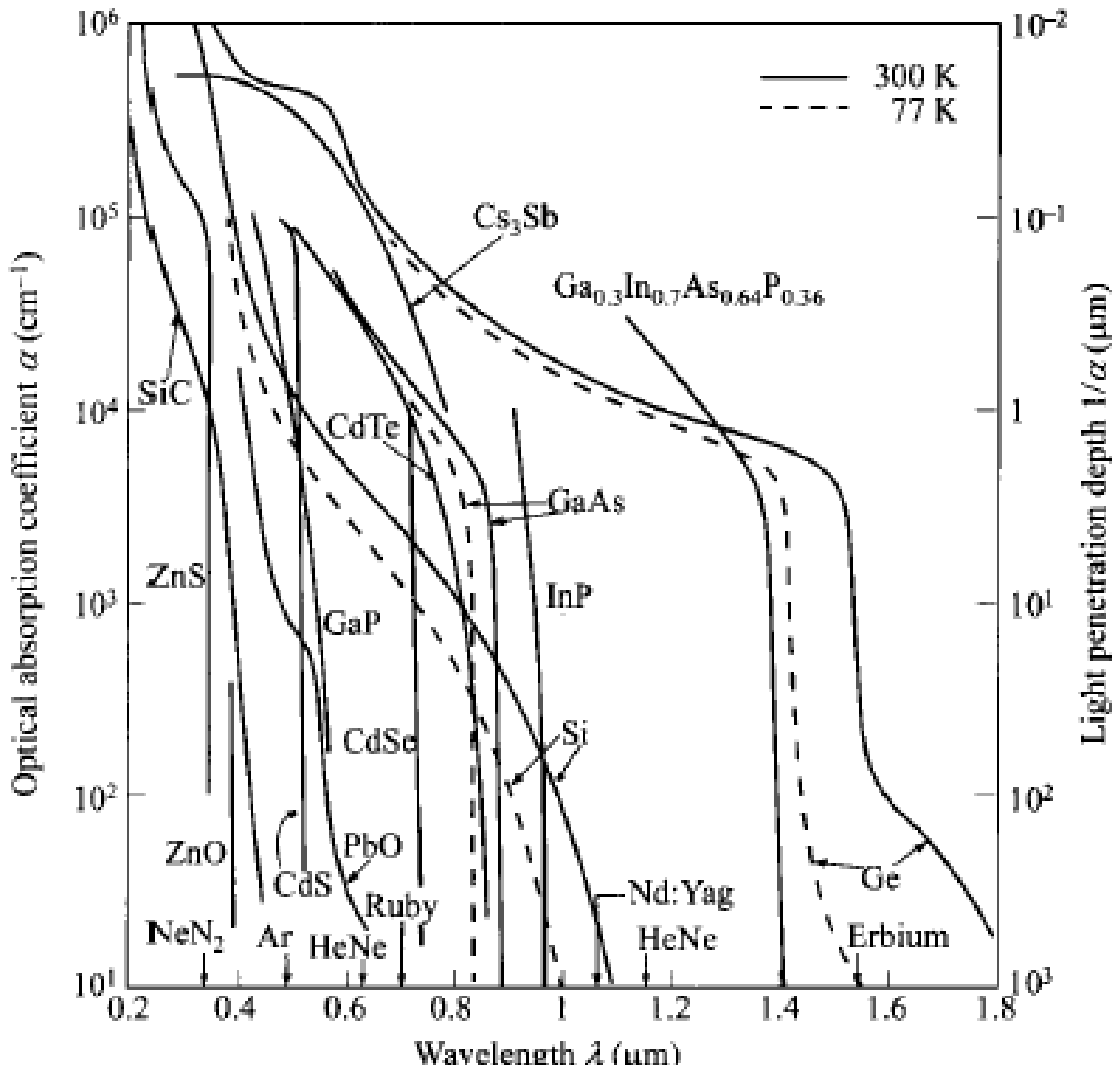

Fig. 1.2 Optical absorption coefficients for various semiconductor materials [4]

Silicon can be seen to absorb light at a reasonable rate (10^2 – 10^4 cm^{-1}) over a relatively wide range of photon wavelengths compared to other materials making, it an

IPR2025-00845
Apple-Sony EX1026 Page 19

attractive material. Although germanium has an even higher absorption over a wider range of wavelengths, it has several drawbacks that limit its applications. The rarity of the material makes it considerably more expensive, and a smaller band gap creates much higher thermally generated noise. All work in this project was completed on ultra-high purity float zone silicon substrates. Float zone substrates are produced by passing an RF induction coil over a poly-Si rod, melting the material and inducing a single crystal growth. The absence of a crucible or any direct contact with melted Si, as found in traditional Czochralski grown substrates, reduces the introduction of carbon and oxygen impurities. The lower oxygen levels eliminate the need to form an oxygen depleted (denuded) zone resulting in consistent material properties throughout the thickness of the wafer.

1.4. Image Sensor Architectures

Image sensors can be classified based on the architecture at either the system level or the pixel level. The two types of system level architecture are monolithic and hybrid. A monolithically integrated image sensor contains both the active photodetector region and all circuitry necessary to convert the charge to a signal conditioned for output on a single substrate. In hybrid image detectors the active sensing region and readout circuitry are fabricated on separate substrates and the hybridization process makes the electrical

IPR2025-00845
Apple-Sony EX1026 Page 20

connections between the two substrates. Hybrid detectors allow the substrate and fabrication processes for the image sensor and read out circuitry to be optimized separately, enabling ultra-low noise imaging systems that are highly sensitive and radiation hardened. The two types of pixel architecture are CCD and CMOS. The CCD is a more mature technology and the typical choice for most high end applications, although CMOS devices are steadily gaining market share.

1.4.1 CCD and CMOS Pixel Architectures

Charge-transfer devices were originally conceived as shift register devices [5], but their applications to image detection were readily apparent [6]. The main difference between the two architectures is the method of addressing the collected charge as shown in Fig. 1.3. In CCD's, overlapping gate structures serve to shift the charge to the edge of the array where it is read out, however in CMOS devices each pixel contains at least one transistor as an access device. Complex devices include additional transistors to amplify the signal and reduce noise. The use of CMOS processing also greatly eases process integration challenges of simultaneously fabricating logic circuitry in the periphery.

IPR2025-00845
Apple-Sony EX1026 Page 21

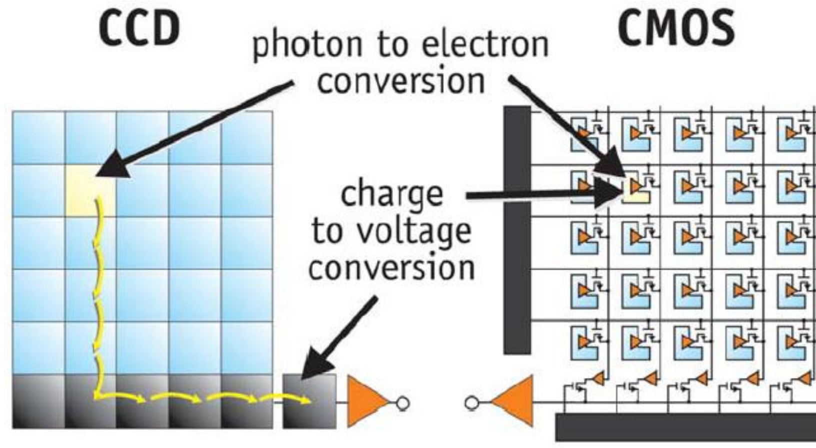

Fig. 1.3 Schematic representation of CCD and CMOS image sensors highlighting the differences of operation taken from [7]

Both architectures have their advantages and disadvantages. A summary of key aspects of the designs is provided in Table 1 [8]. Historically, CCD's are known for delivering a higher quality sensor although at a much higher price. CMOS sensors have benefited greatly from the advanced processing techniques developed by the memory and logic sectors making them extremely cost effective.

TABLE 1 SUMMARY COMPARISON OF CCD AND CMOS TECHNOLOGIES [8]

CCD Technology	CMOS Technology
Highly optimized for optical detection, special fabrication requirements	Benefits from advances in manufacture of high-volume digital products
Very high signal-to-noise	Noise typically higher
Low photoresponse nonuniformity (PRNU), low fixed-pattern noise (FPN)	High PRNU, high FPN, improved by gain and offset correction
Low dark current	Dark current typically higher
High power dissipation	Low power consumption
Complex driver electronics, no on-chip logic and digitization	Single power supply operation, digital output
Serial readout, no windowing capability	Random addressing capability

IPR2025-00845
Apple-Sony EX1026 Page 22

Passive and Active CMOS Pixels

The type of pixel circuitry implemented in a CMOS sensor is an important aspect of system performance and can be used to categorize the device. The first designs were passive sensors that utilize a single transistor per pixel as an access device. Active sensors incorporate per-pixel amplification through the use of source-followers to greatly increase pixel performance [9]. Additional circuitry can further enhance device performance by reducing undesirable effects due to device variation and noise, however the additional transistors occupy precious real estate within the pixel [10]. The percentage of pixel area available for sensing incoming light is known as the fill factor. The two most prominent techniques used to mitigate the effects of reduced fill factors are discussed in the following sections.

1.4.2 System Architecture: Hybrid vs. Monolithic

The first image sensors were monolithically integrated CCD's based upon the charge transfer device conceived by Boyle and Smith. In a monolithic device, both the active sensing region and the read out circuitry are fabricated within the same substrate. This can require greater process complexity due to the vastly different doping requirements of the two regions. A top down micrograph of a typical monolithic CMOS image sensor is shown in Fig. 1.4 [11]. The sensor array is in the center and is

IPR2025-00845
Apple-Sony EX1026 Page 23

surrounded by the logic circuitry used to select the rows and columns and condition the signal for output.

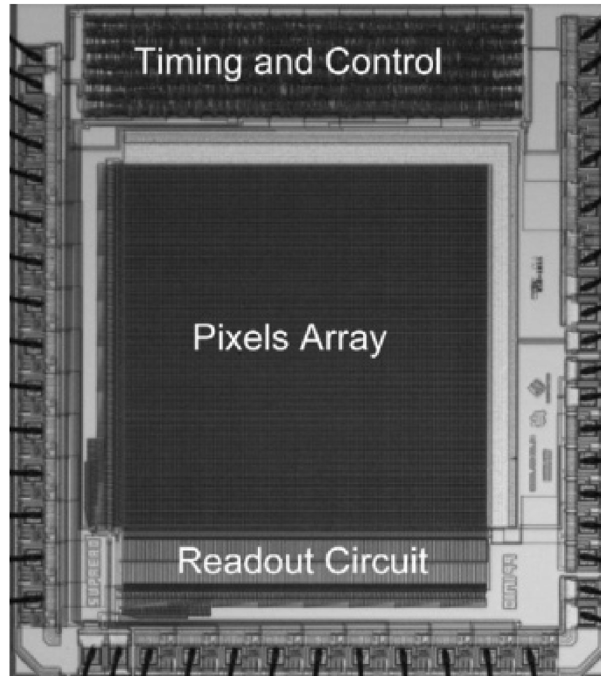

Fig. 1.4 Typical monolithic CMOS image sensor [11]

Front-side vs. Back-side Illumination

Topography and the use of opaque materials in the fabrication of image sensors leads to a loss signal before it reaches the photodetector due to absorption, reflection, and scattering. Transparent conductors such indium-tin-oxide (ITO) are often used to reduce the need for opaque metals, though imperfect and curved interfaces still cause some scattering of the light. This reduction of signal is enhanced even further in multi-transistor pixel designs with decreased fill factors, which are more common than older single transistor designs.

IPR2025-00845
Apple-Sony EX1026 Page 24

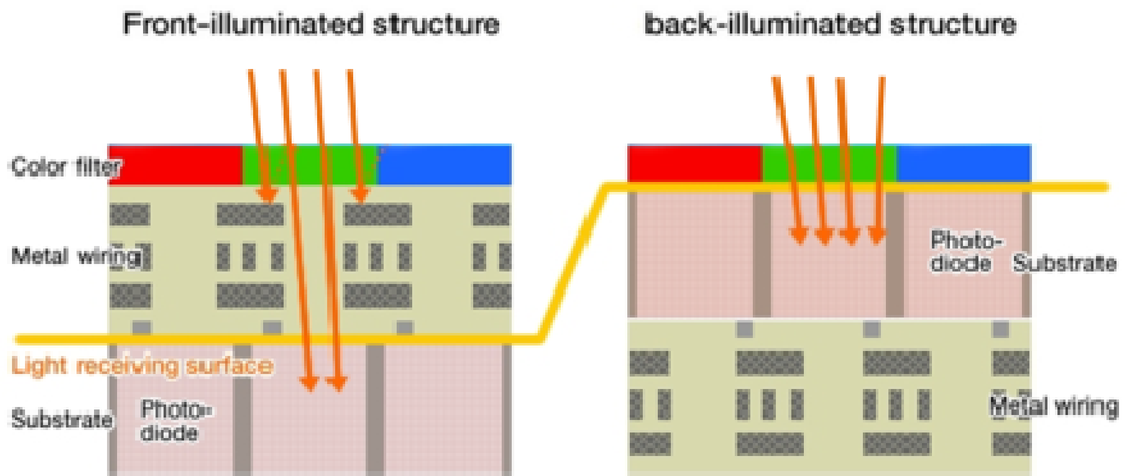

Fig. 1.5 Schematic representation of front-side illumination (FSI) vs back-side illumination (BSI) [12]

To eliminate virtually all possible sources of external light loss, one technique is to flip the substrate over after the processing of front-side is completed and illuminate the device from the back. This often requires additional processing, as most devices are fabricated in substrates too thick to allow light of sufficient intensities to penetrate into the depleted regions. The devices are can be thinned, taking special care to leave a high quality backside-surface to avoid the creation of defect states. Often specialty thinned and double side polished substrates are used for the creation of these devices. Anti-reflection layers are also used to almost completely eliminate all reflected light. A comparison of front-side illuminated (FSI) and back-side illuminated (BSI) structures is shown in Fig. 1.5.

IPR2025-00845
Apple-Sony EX1026 Page 25

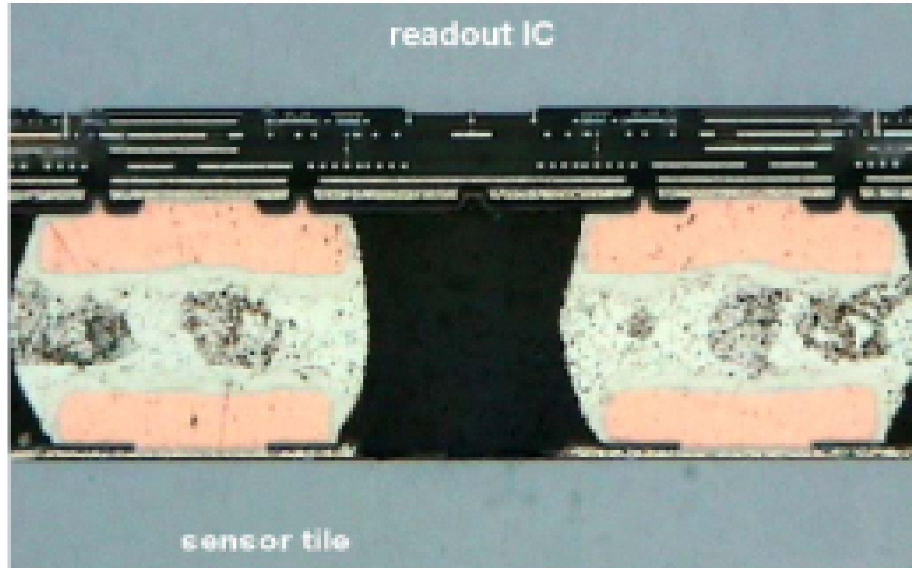

Fig. 1.6 Cross-section of hybrid imager used in ATLAS detector [13]

1.4.3 Hybrid Image Detectors

The development of advanced packaging techniques has enabled the creation of hybrid image detectors, where the sensor and readout circuit are fabricated on separate substrates and then connections are made between the two die. The large matrix of densely spaced interconnections between the substrates required for the pixel arrays makes traditional connections technologies such as wire-bonding and screen-printing incompatible. A lithographic pattern transfer combined with electroplating and lift-off processes have enabled large arrays of bumps to be formed with at a very fine pitch. A cross section of the hybrid sensor used in the ATLAS detector is shown in Fig. 1.6. Hybrid detectors are inherently back-side illuminated as the front-sides of the two chips are mated to each other. Wire-bond connections are made to the backside of the photo-

IPR2025-00845
Apple-Sony EX1026 Page 26

diode to set the operating potential and the periphery of the read-out integrated circuit (ROIC) for communication.

Another large advantage of hybrid detectors is the ability use different substrates and fabrication sequences for the sensor and read out circuitry. The read-out circuitry requires high doping levels for radiation hardness, to prevent latch-up and maintain proper device operation. The sensor requires low doping levels to reduce leakage currents and increase the width of the space charge region.

1.5. APRA Imaging System

In 2006, the Rochester Imaging Detector Laboratory (RIDL) at the Rochester Institute of Technology (RIT) was awarded a grant from the National Aeronautic and Space Administration (NASA) for "A Very Low Noise CMOS Detector" under the Astronomy and Physics Research and Analysis (APRA) Program of the Science Mission Directorate [14]. The stated goal of the project from the grant proposal was "to design, fabricate, and measure the noise of a novel hybrid CMOS detector with ΣΔ (sigma-delta) pixel design at cryogenic temperatures." The initial plan called for the read out circuit to be fabricated through MOSIS, a low-cost prototyping and small-volume production service, and the detectors to be fabricated in the Semiconductor and Microsystems Fabrication Laboratory (SMFL) at RIT.

IPR2025-00845
Apple-Sony EX1026 Page 27

1.5.1 APRA Hybrid Imaging System

The APRA Hybrid imaging system was designed as a proof-of-concept device for astronomical purposes and was expected to meet the performance goals specified Table 2 [14]. A dark current of less than 1 e-/pixel/sec was desired at an operating temperature of 200 K from an 8-12 um pixel. To meet the performance goals, a hybrid architecture was designed with a silicon CMOS ROIC mated to a silicon p-*i*-n photodiode detector array using indium bump bonds.

TABLE 2 IMAGING SYSTEM PERFORMANCE GOALS [14]

Parameter	Value
Format	256x128
Pixel Size	8-12 μm
Read Noise	<1 e^- RMS
Dark Current (@200 K)	<0.1 e^-/s/pixel
QE[b]	>85%
Latent Image (after full well)	<1 e^-
Charge Rate Capacity	>10^8 e^-/s
Dynamic Range	>10^6
Operating Temperature	20 K – 300 K
Fill Factor	100%
Susceptibility to Radiation Damage	immeasureable[b]
Susceptibility to Radiation Transients	immeasureable
Maximum Frame Rate	1000 fps
Power (@30 fps)	<1 nW/pixel
Power (@0.1 fps)	<100 pW/pixel
Technology Readiness Level[c]	4

The ROIC design was modified to allow bump bond contacts and fabricated through a foundry service. The diodes were fabricated at the RIT SMFL and the hybridization process was developed in cooperation with the Smart System Technology

& Commercialization Center (STC). A modest array size was chosen to reduce the negative impacts of poor reliability from immature processes on the proof-of-concept design. The pixel size was originally intended to be between 8-12 μm but was ultimately relaxed slightly to 15 μm due to lithographic constraints.

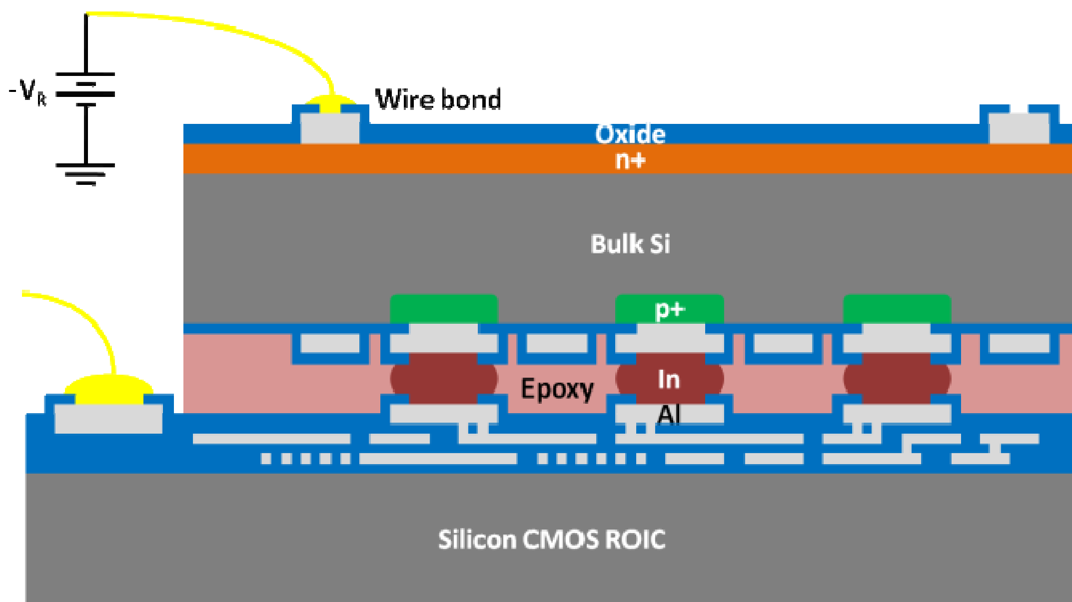

Fig. 1.8 Illustration of Hybridized Image Detector cross-section

Cross-section and top-down illustrations of the detector are shown in Fig. 1.8 and Fig. 1.9. Back-side illumination ensures a 100 % fill factor and the large depleted intrinsic region enables high quantum efficiencies throughout a wide range of wavelengths. The hybridization process uses techniques developed by the IC industry for flip chip packaging to create bump bond interconnects between the two chips. The gap is backfilled with epoxy to provide structural support and protection from environmental effects.

IPR2025-00845
Apple-Sony EX1026 Page 29

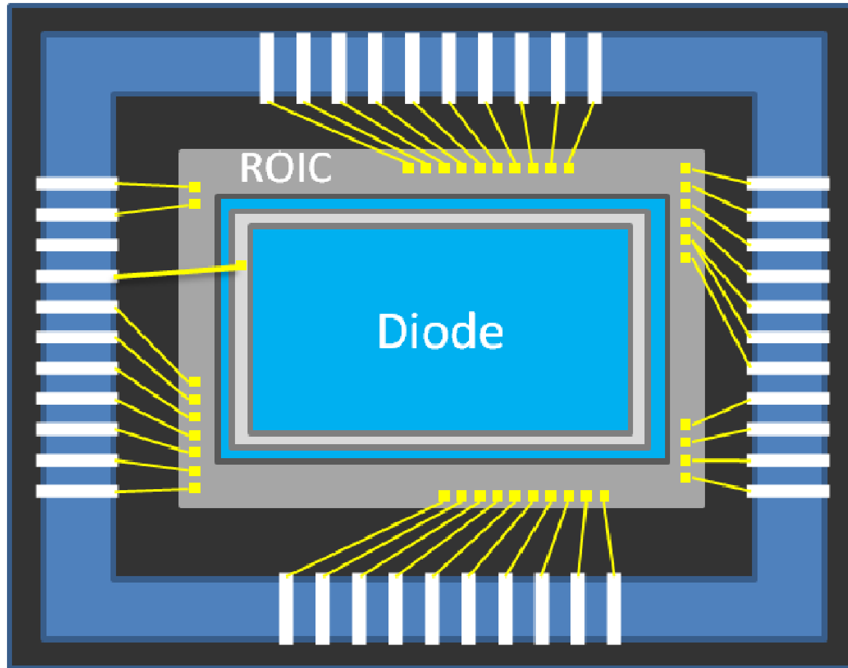

Fig. 1.9 Top Down Illustration of Hybridized Image Detector

1.5.2 MOSIS Read-out Integrated Circuit (ROIC)

The ROIC utilizes an oversampling sigma-delta (ΣΔ) analog-to-digital conversion technique to achieve an RMS read noise of < 1 e^-/pixel/sec. The circuits were designed by Dr. Zeljko Ignjatovic, an Assistant Professor of Electrical and Computer Engineering at the University of Rochester. The ROICs were manufactured at TSMC through the MOSIS foundry service using a 0.35 μm 2-poly 4-metal CMOS process. **Fig. 1.9** shows a top-down illustration of the ROIC and diode wire-bonded into the DIP package. An optical micrograph stitched together shows the full ROIC die in Fig. 1.10. The array of bump bond contacts can be seen as the shaded rectangle in the center of the die.

IPR2025-00845
Apple-Sony EX1026 Page 30

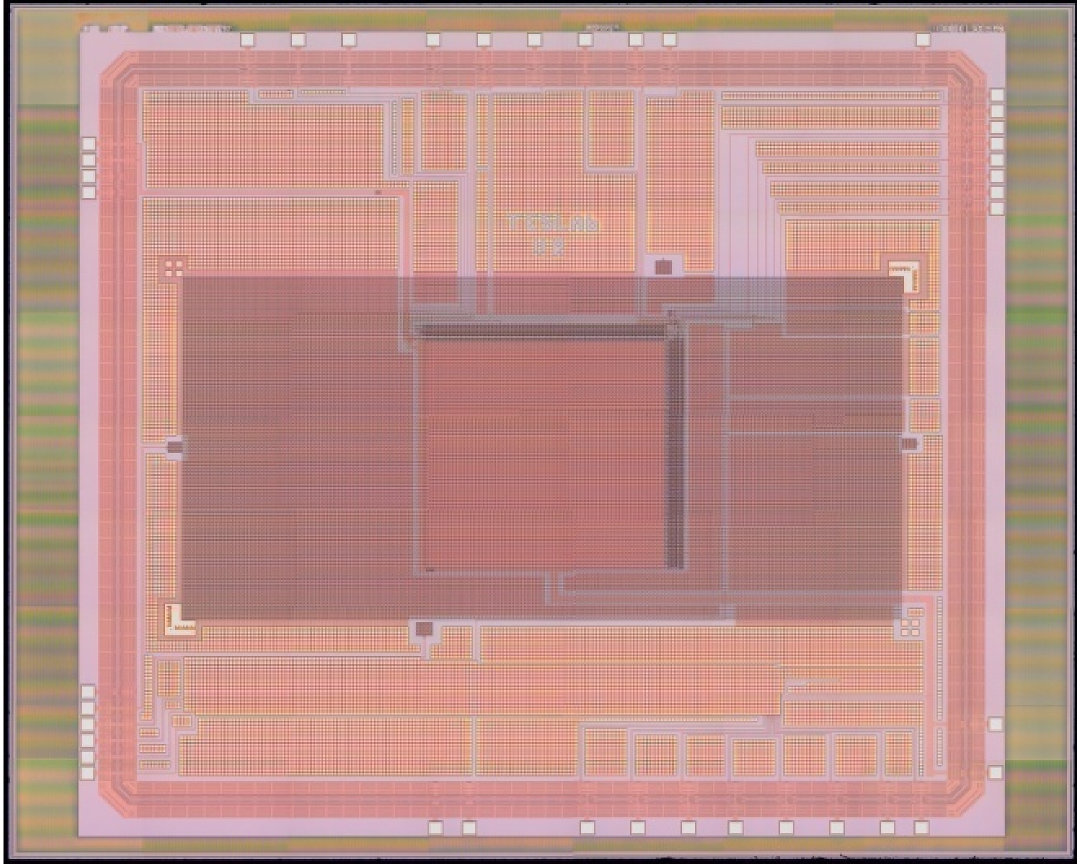

Fig. 1.10 Optical Micrograph of ROIC die

1.5.3 Silicon p-i-n Photodiode Array

The APRA imaging system consists of an array of p-*i*-n photodiodes at a pitch of 15 μm. The pixel is defined by an 11 μm p^{+}region connected to a 9 μm aluminum pad through a 7 μm contact opening. The aluminum pad is made accessible by a 6 μm via through the passivation oxide. The pixel is also surrounded by a 1 μm aluminum border that forms a 2 μm grid throughout the array when the pixels are tiled. The layout for an individual pixel is shown in Fig. 1.11.

IPR2025-00845
Apple-Sony EX1026 Page 31

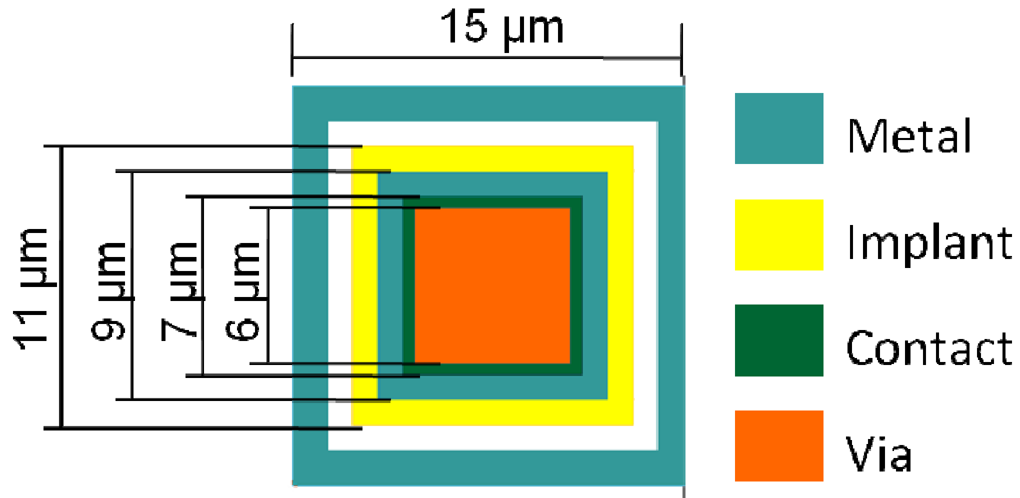

Fig. 1.11 Pixel Design and Layout

The pixel was repeated into a 128 x 256 element array surrounded by a bias ring that consisted of 6 modified pixel elements. Guard ring connections were located along the top and bottom of the array, and connections to the inter-pixel grid were placed at the midline on the left and right sides of the array. Alignment marks and verniers for the hybridization process were included within the 6 pixel guard ring border at the corners of the array. An image of the layout of the photodiode array is shown in Fig. 1.12.

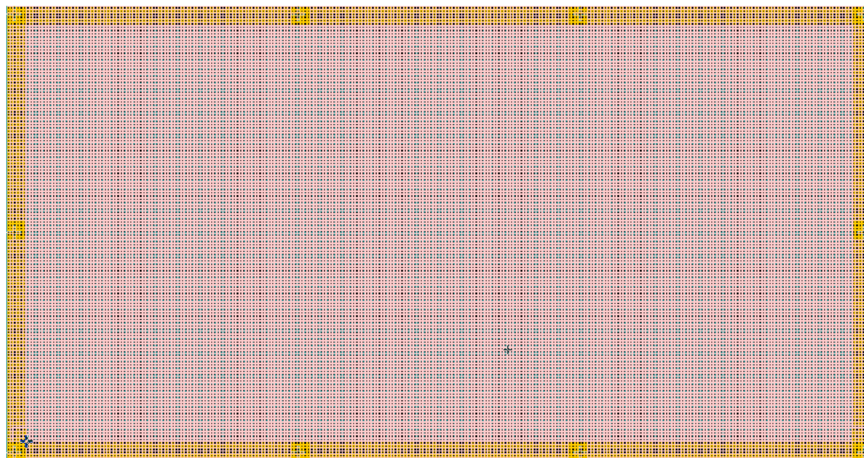
Fig. 1.12 Layout of 256x128 p-i-n photodiode array with guard ring

A substrate thickness of 250 μm was selected to enhance efficiency in the longer wavelength (e.g. 1 μm) regime. The substrates were double-side polished ultra-high purity float-zone silicon substrates doped with phosphorous to a resistivity of 5000 Ω·cm. An operating bias of 50 V was designed to fully deplete the substrate at the chosen thickness. A single-layer anti-reflective coating was used to optimize the quantum efficiency for a selected wavelength range. The hybridization technique was a flip-chip process using indium bump-bonds as the interconnect metal.

IPR2025-00845
Apple-Sony EX1026 Page 33

Chapter 2

p-i-n Photodiode Operation

The *p-n* junction, or diode, is the most basic, fundamental element of all solid state semiconductor devices. In use since 1906 in crystal radios, the theoretical framework behind the operation of the devices was unknown until 1939 when Russel Ohl discovered the role of impurities [15]. In its most simplistic form, the two-terminal device, also known as a rectifier, only allows current to pass in a single direction. The *p-i-n* diode consists of a *p-n* junction separated by a region so lightly doped that for most practical purposes is it assumed to be intrinsic. The intrinsic region most commonly denoted by the letter *i* though sometime the greek letters π or ν are used to denote the lightly doped region as either p or n-type. The static and dynamic characteristics of the device in the absence and presence of illumination will be discussed in the following chapter.

2.1. Junction Electrostatics

The junction electrostatics are discussed assuming a one sided abrupt junction for both of the p^+-n^- and n^+-n^- junctions under thermal equilibrium conditions. The fundamental equation describing the relationships between potential, electric field and

IPR2025-00845
Apple-Sony EX1026 Page 34

charge is known as the Poisson equation. Derived from Gauss's Law, the partial-differential equation in its one-dimensional form is given in (2.1).

$$\frac{d^2\Psi_i}{dx^2} = -\frac{d\mathcal{E}}{dx} = -\frac{\rho}{\varepsilon_s} \tag{2.1}$$

where Ψ_i is the semiconductor potential, □ is the electric field, ρ is the volume charge density and ε_s is the permittivity of the material. The depletion approximation is used to simplify analysis by assuming a rectangular profile for the depleted charge. The total charge on each side of the junction is assumed to be opposite and equal and represented by (2.2)

$$N_A W_{Dp} = N_D W_{Dn} \tag{2.2}$$

where N_A and N_D are the doping densities and W_{Dp} and W_{Dn} are the depletion widths for the p and n-type sides of the junction respectively. The total built-in potential for both junctions can be determined by (2.3) where the intrinsic region is for all intents and purposes ignored.

$$\Psi_{bi} = \frac{kT}{q} ln\left(\frac{p_{p0}}{p_{n0}}\right) = \frac{kT}{q} ln\left(\frac{n_{p0}}{n_{n0}}\right) \tag{2.3}$$

where Ψ_{bi} is the built-in potential, k is the Boltzmann constant, T is the temperature in Kelvin, q is the electronic charge, and the carrier densities are the thermal equilibrium values represented in the traditional fashion. Integration of the Poisson equation, assuming complete ionization of impurities within the depletion, regions yields the

IPR2025-00845
Apple-Sony EX1026 Page 35

electric field distribution, □(x). Further integration of the electric field yields the potential distribution, Ψ(x), of the junction. The width of the depletion region for a one-sided abrupt junction can be calculated by (2.4)

$$W_D = \sqrt{\frac{2\varepsilon_s}{qN}\left(\Psi_{bi} - V - \frac{2kT}{q}\right)} \tag{2.4}$$

The width of the depletion regions is modified by any applied bias, as well as a factor of $2kT/q$ to account for the two minority-carrier distribution tails. The tails are caused by the diffusion of majority carriers into the depletion region causing a rounded corner in the assumed square charge profile.

2.1.1 Current-Voltage Characteristics

The current through an ideal diode is described by the Shockley equation, also known as the ideal diode law. The development of the ideal diode law, shown below in (2.5), is rather exhaustive and provided by many excellent textbooks so it will not be covered here. Interested readers are directed toward the development provided by Sze in "Physics of Semiconductor Devices" [4].

$$J = J_p + J_n = J_0\left[exp\left(\frac{qV}{\eta kT}\right) - 1\right] \tag{2.5}$$

where J is the total current density, J_p and J_n are hole and electron current densities respectively. J_o is the saturation current density, V is the applied voltage, and η is the

IPR2025-00845
Apple-Sony EX1026 Page 36

ideality factor. The ideality factor provides insight into the dominant mechanism of current transport and varies between 1 for diffusion and 2 for recombination.

While the Shockley equation provided a breakthrough in the understanding of device operation, most practical devices do not exhibit ideal operation. Non-ideal effects contribute to deviations from the ideal operation, as shown in Fig. 2.1 [4]. Regions of forward operation are labeled as (a) generation-recombination region, (b) diffusion region, (c) high-injection region and (d) series-resistance effect. The reverse bias regions is labeled as (e) with junction breakdown occurring at a high reverse potential.

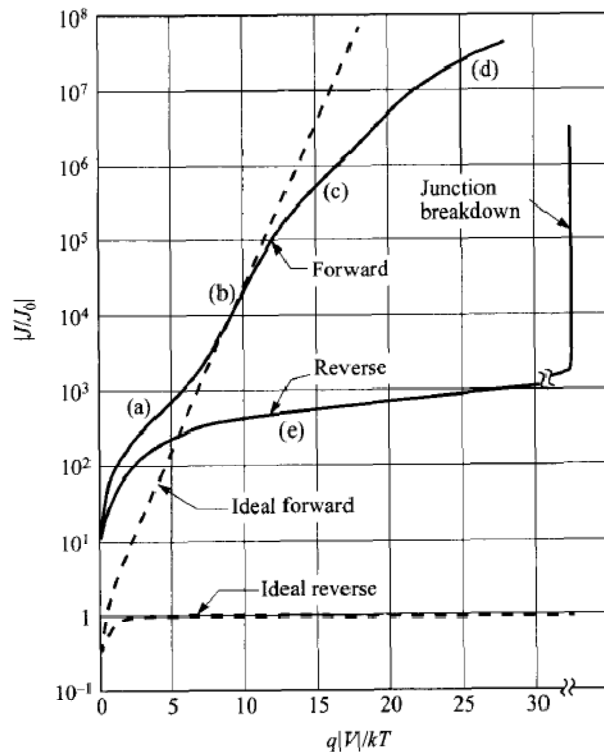

Fig. 2.1 Current-voltage characteristics of a practical Si diode showing ideal and experimental behavior in forward and reverse modes of operation [4]

Forward Bias

The forward bias operation of a practical *p-i-n* diode is largely determined by recombination rate in the large intrinsic region. Thus, for good forward operation, long carrier lifetimes are desirable. A high quality junction is also needed to prevent defects from creating recombination centers in the vicinity of the junction. Due to the extremely low doping levels on the intrinsic side of the junction, high-injection effects start to occur at relatively low voltage levels. The ideality factor provides an insight into the dominant current mechanism.

Reverse Bias

In the absence of light, an ideal device would have an extremely small reverse bias current due only to the thermal generation of carriers within the depletion region. Leakage currents in practical devices are always larger than theoretically predicted due to mid-level traps in the vicinity of the metallurgical junction. These defects are typically due to the methods of dopant introduction used and the inability to completely heal all crystal imperfections in the annealing process. Metal-ion contamination can also be a major contributor to leakage currents, although the precautions taken in semiconductor fabrication are specifically designed to address contamination concerns. The net

IPR2025-00845
Apple-Sony EX1026 Page 38

transition rate (U) in a semiconductor is the difference between the recombination rate and generation rate and is determined by (2.6) [4].

$$U = \frac{\sigma_n \sigma_p v_{th} N_t (np - n_i^2)}{\sigma_n \left(n + n_i e^{\left(\frac{E_t - E_i}{kT} \right)} \right) + \sigma_p \left(p + n_i e^{\left(\frac{E_t - E_i}{kT} \right)} \right)} \tag{2.6}$$

The numerator is the relative change in carrier concentrations compared to thermal equilibrium levels, σ_n and σ_p are the electron and hole capture cross sections respectively, v_{th} is the thermal velocity, and N_t is the density of bulk traps with corresponding energy levels E_t. A positive value corresponds to an excess of carriers resulting in recombination and a negative value indicates a deficit of carrier and leads to generation. Image sensors are typically operated in reverse bias with relatively large depletion regions so the defect levels in the substrate are an extremely important parameter.

As the energy of the trap level deviates from the mid-gap value its efficiency as a generation/recombination center falls of dramatically due to the exponential dependence, consequently only mid-gap traps are typically considered. Minority carrier lifetimes (τ) are defined as the inverse product of the capture cross section, thermal velocity, and bulk trap density for each of the carrier types. Substituting these values for lifetimes and using the above assumption (2.6) can be simplified to (2.7) [4].

IPR2025-00845
Apple-Sony EX1026 Page 39

$$U = \frac{(np - n_i^2)}{\tau_p(n+n_i) + \tau_n(p+n_i)} \tag{2.7}$$

2.1.2 Minority Carrier Lifetime

Leakage currents are highly dependent upon minority carrier lifetimes, which are determined by initial doping concentration and the methods used to introduce additional dopants to create the desired profiles. The minority carrier lifetime (holes in n-type silicon) as a function of donor density is shown in Fig. 2.2 taken from [16]. Defects can be created during the doping process that form allowed energy levels within the band gap known as generation/recombination centers. Ion implantation has been shown to cause higher leakage levels than thermal doping due to lattice damage that is not fully healed during the activation anneal, however thermal doping can require additional process steps [16].

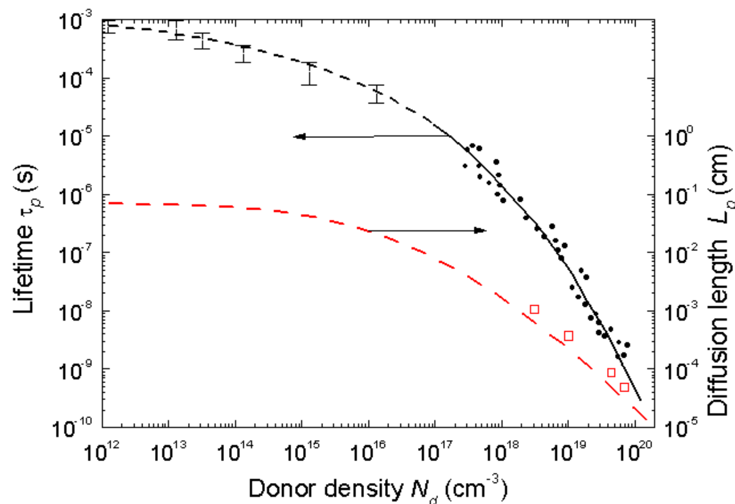

Fig. 2.2 Lifetime (τ_p) and Diffusion length (L_p) of Holes in n-type Si as a function of Donor density [17]

IPR2025-00845
Apple-Sony EX1026 Page 40

There are several methods to electrically measure the lifetime of carriers in a p-i-n diode, although they all rely on similar principals [18, 19]. The diode is first forward biased to inject minority carriers across the junction into the base. In this condition the recombination of minority carriers in the quasi-neutral region is the primary contribution to total current flow through the device. The method by which the forward bias condition is removed leads to several techniques for measuring carrier lifetimes.

2.1.3 Open-Circuit Voltage Decay (OCVD) Method

In the open-circuit voltage decay (OCVD) method the voltage across the junction is monitored as a switch is opened removing the bias from the device [18]. Typical current and voltage transients in the device right before and after the switch is opened are shown in Fig. 2.3 (a). An initial drop in voltage is observed due to ohmic potential losses that vanish as the external current is removed. The remaining voltage across the diode is the junction voltage caused by the presence of excess carriers. As there is no current flow, the decay of this voltage is directly related to the recombination of these carriers and can be used to determine a carrier lifetime in the neutral bulk region. The lifetime can be shown to be related to the time-varying voltage by (2.8).

$$\tau = -\frac{kT/q}{dV(t)/dt} \tag{2.8}$$

IPR2025-00845
Apple-Sony EX1026 Page 41

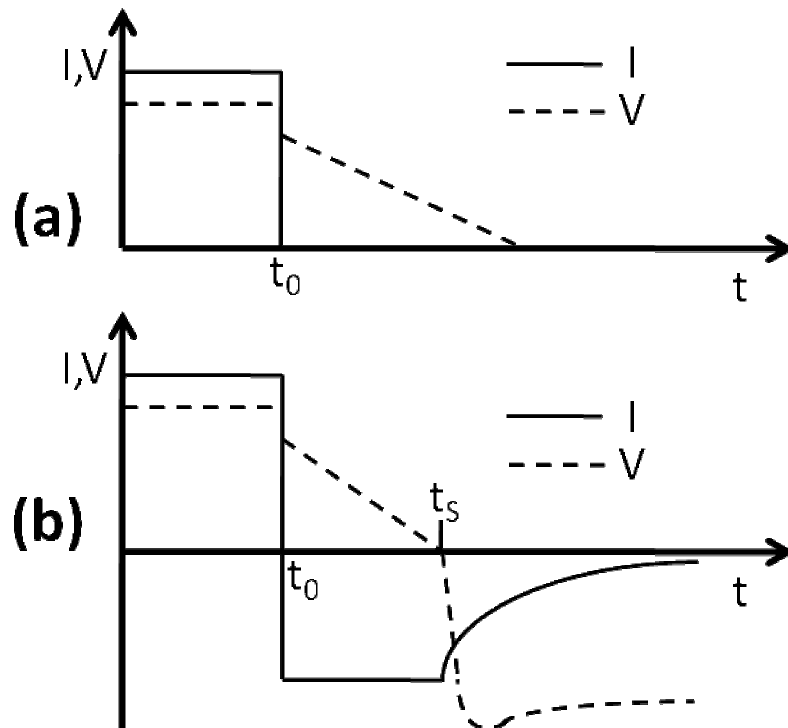

Fig. 2.3 Current and voltage transients observed in the methods used for carrier lifetime measurements: (a) OCVD; (b) Junction Recovery Method [18]

2.1.4 Reverse Recovery Method

In the reverse recovery method, the polarity of the voltage across the diode is switched from forward to reverse bias [20]. Excess carriers in the junction are removed by both recombination and a drift current caused by the electric field. Initially the device is still forward biased and the junction voltage is observed across the device as the ohmic losses are eliminated.

A reverse current begins to flow, as shown in Fig. 2.3 (b), and maintains a fairly constant magnitude as the junction voltage decays. This continues until the excess carriers at the edge of the space charge regions are approximately zero. This time,

IPR2025-00845
Apple-Sony EX1026 Page 42

indicated as t_r, is known as the storage time for the device. At this point the reverse current decreases to its leakage level as the depletion region widens and the remaining excess carriers deep within the quasi-neutral region recombine. A charge storage analysis during the constant current phase results in (2.9). Recombination lifetimes (τ_R) can be determined from the slope of a best fit line on a plot of t_r versus $ln(1+ I_F/I_R)$.

$$t_r = \tau_R ln\left(1 + \frac{I_F}{I_R}\right) \tag{2.9}$$

The reverse recovery technique was one of the first methods available for electrically measuring lifetimes of carriers in devices. It was used widely in industry, however it can become inaccurate for small charge storage times or when the amount of charge stored in the device, when is has recovered, is significant. The OCVD method is much simpler, requiring only a single measurement to obtain carrier lifetimes, and the assumptions made during the derivation are less likely to be invalid in practical situations. For these reasons the OCVD method is a common test for solar cells created in the photovoltaic industry [18].

2.2. Absorption and Photogeneration

The first event that must occur in the photogeneration of carriers is the absorption of a photon by the silicon. Absorption is governed by several different mechanisms and

IPR2025-00845
Apple-Sony EX1026 Page 43

the optical properties of the materials. These properties are determined by the complex refractive index of the material shown in (2.10)

(2.10)

where n is the real portion of the refractive index and k is the imaginary portion (also known as the extinction coefficient). Both parts of the refractive index are a function of the wavelength of the incident radiation. The complex refractive index as a function of wavelength for silicon is shown in Fig. 2.4.

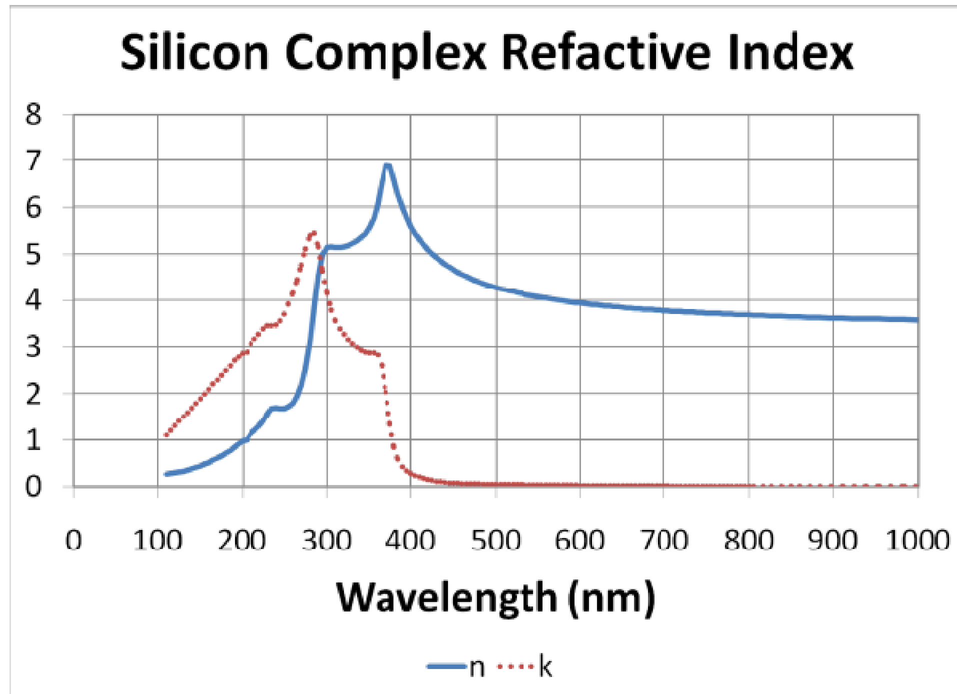

Fig. 2.4 Complex refractive index vs. wavelength for silicon

The Fresnel equations determine the amount of light that is either reflected or transmitted at the interface of two materials. It is dependent on the refractive index of the two materials forming the interface and the angle the incident radiation makes with the

IPR2025-00845
Apple-Sony EX1026 Page 44

normal to the interface. The reflectance can depend on the polarization state of the light. For an angle of incidence nearly perpendicular through a transparent media, the equations simplify to (2.11) and (2.12).

(2.11)

(2.12)

The reflectance and transmittance at the silicon-interface have been calculated and are shown in Fig. 2.5 as a function of wavelength. It can be seen that there are significant reflection in the sub-400 nm region that limits the performance in the ultra violet (UV) region, but they decay to ~ 30% above 400 nm. The reflections at this interface will have a large affect on the device performance.

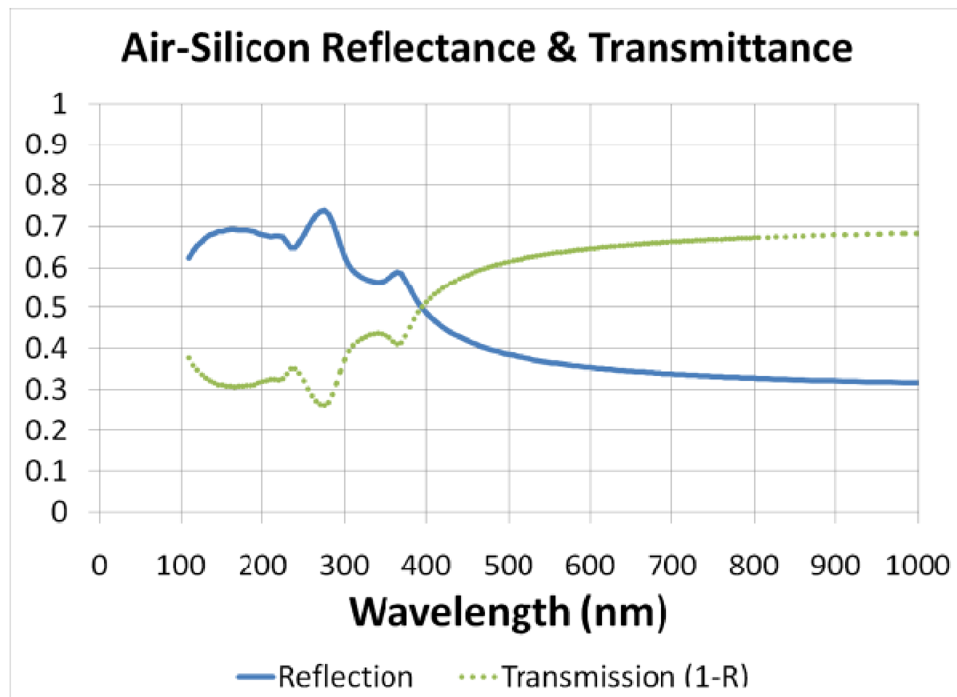

Fig. 2.5 Reflectance and Transmittance at the air-silicon interface

IPR2025-00845
Apple-Sony EX1026 Page 45

Light that is transmitted at the interface must then be absorbed by the silicon. The absorption of light is determined by the absorption coefficient, α, as calculated by (2.13)

— (2.13)

where λ is the wavelength of the light. The intensity of the light is attenuated exponentially with distance in the silicon as shown by (2.14)

(2.14)

where I_0 is the incident intensity and x is the distance into the substrate. A characteristic penetration depth can be defined as the inverse of the absorption coefficient and is the depth at which 63.2% of the incoming radiation has been absorbed. A graph of the absorption coefficient and penetration depth as a function of wavelength for silicon is shown in Fig. 2.6.

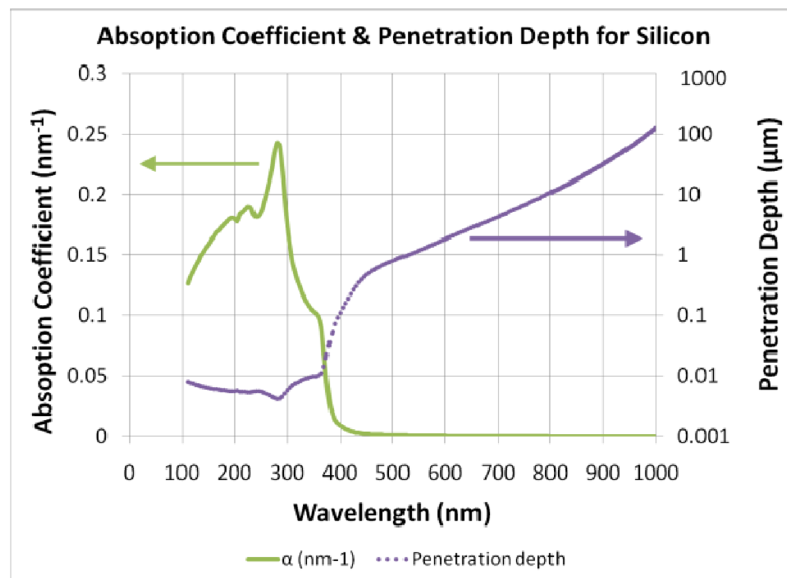

Fig. 2.6 Absorption coefficient and penetration depth vs. wavelength for silicon

Carrier Generation

In order to generate carriers in a semiconducting material the absorbed photons must have sufficient energy to excite a carrier from the valence band to the conduction band; that is to say the energy must be greater than the band gap of the material. The generation of carriers is given by (2.15) and follows the same exponential decay as the absorption but is modified by the absorbed photon flux per unit area given in (2.16)

$$G_e(x) = \Phi_0 \alpha e^{-\alpha x} \tag{2.15}$$

$$\Phi_0 = \frac{P_{opt}(1-R)}{Ahv} \tag{2.16}$$

where P_{opt} is the incident optical power, R is the reflectivity, A is the area of the device and hv is the energy of the photon. P_{opt}/hv can be recognized as the incident photon flux that is modified by $(1\text{-}R)$ to include only those photons that are not reflected.

2.3. Illuminated Operation

The main function of a photodiode is the efficient conversions of photons into electrons. Three events must occur to achieve the conversion of a signal from optical to electrical. A photon must be absorbed within the semiconducting material, charged carriers must be elevated to an excited state, and the charged carriers must be collected at the terminals of the device. The charge is then converted into a potential and read out by circuitry attached to the image sensor. During each of these steps there are opportunities for loss of signal and the incorporation of noise.

IPR2025-00845
Apple-Sony EX1026 Page 47

2.3.1 Charge Collection

Once the carriers are generated they must be transported to the collection nodes before they recombine, giving rise to a current in the device. This current can be divided into two components, drift and diffusion. Carriers generated inside the depletion region will experience an immediate acceleration due to the electric field present, while those generated outside the depletion region must diffuse into the junction before they can be swept away. The total current density in a p-i-n diode is the sum of the individual components and is represented by (2.17)

$$J_{tot} = q\Phi_0\left[1 - \frac{e^{-\alpha W_D}}{1+\alpha L_p}\right] + \frac{q p_{no} D_p}{L_p} \tag{2.17}$$

where W_D is the depletion width, L_p is the minority carrier diffusion length, p_{no} is the equilibrium minority carrier concentration and D_p is the minority carrier diffusion constant [4]. The first term in (2.17) represents the carriers generated inside and within one diffusion length of the depletion region. A dependence on bias is not seen as it is assumed that all generated carriers are collected. The second term in Equation (8) is the dark current and is relatively insignificant in the presence of light. The minority carrier diffusion length is determined by (2.18)

$$L_p = \sqrt{D_p \tau_p} \tag{2.18}$$

where τ_p is the minority carrier lifetime.

IPR2025-00845
Apple-Sony EX1026 Page 48

2.3.2 Quantum Efficiency

The quantum efficiency (QE) of a detector is a measure of the effectiveness at converting photons into electrons. The QE can be separated into internal QE and external QE, where internal QE is the ratio of absorbed photons to collected electrons. The external quantum efficiency is determined by taking the ratio of the collected carriers to the incident photon flux as shown in (2.19).

$$\eta = \frac{AJ_{tot}/q}{P_{opt}/h\nu} = (1-R)\left[1-\frac{e^{-\alpha W_D}}{1+\alpha L_p}\right] \quad (2.19)$$

The external quantum efficiency is reduced due to both reflections at the silicon air interface and the generation of carriers outside the depletion region. To maximize the quantum efficiency it is desirable to have a large depletion region ($\alpha W_D \gg 1$) and a large diffusion length ($\alpha L_p \ll 1$).

2.3.3 Point Spread Function

Carriers generated in the device diffuse laterally as they are swept away by the electric field and collected at the terminals. The amount of lateral diffusion is determined by the diffusion rate and the amount of time the carrier has to diffuse. If the lateral diffusivity is large enough, or the transit time is too long, carriers generated in one pixel can be collected by an adjacent pixel. This is an extremely important effect for thick detectors, such as the one used in this project, as the transit time is considerably longer

than for thin film detectors. The amount of lateral diffusion can be characterized by the point spread function (PSF), which also provides information about the minimum resolution of the device.

IPR2025-00845
Apple-Sony EX1026 Page 50

Chapter 3

SILVACO Atlas TCAD Simulations

The physics based device simulation package Atlas from the technology computer-aided design (TCAD) software suite by SILVACO is used to simulate the electrical behavior of a defined structure. Numerical simulations are an extremely important tool to the modern day engineer. They allow new products to be developed faster and cheaper than traditional experimentation and prototyping methods. In the past century over a trillion dollars has been collectively invested into semiconductor research creating a vast pool of empirical data from which theoretical models have been created to describe virtually every aspect of device operation. In the early 1980's a group of researchers at Stanford developed a two-dimensional, two-carrier semiconductor simulation program known as PISCES-II [21]. This program became the basis for the S-PISCES module of Atlas, where the "S" implies it is specifically for silicon. The Luminous module is used to simulate the optoelectronic behavior of the device.

3.1. Numerical Poisson Solver Theory

The finite element method is a numerical technique for the discretization and solution of partial differential equations. The numerical simulation of the electrical behavior of semiconducting devices is accomplished by the application of the finite

IPR2025-00845
Apple-Sony EX1026 Page 51

element method to a set of partial differential equations derived from Maxwell's Laws. The electrostatic potential is determined from Poisson's equation as a function of the space charge distribution. Carrier continuity and transport equations determine the concentration and motion of carriers within the devices as a result of generation/recombination and transport processes. In this specific application the simulation program is often referred to as a numerical Poisson solver.

3.1.1 Electrostatic Equations

The relationship between electrostatic potential and spatial charge distribution is described by the Poisson equation as

$$\nabla^2\varphi = -\frac{\rho}{\varepsilon_S} = \frac{q(n - p + N_A - N_D)}{\varepsilon_S} \tag{3.1}$$

where φ is the electrical potential, ρ is the space charge density, ε_S is the permittivity of the semiconductor, q is the electronic charge, n is the electron concentration, p is the hole concentration, N_A is the acceptor-like dopant density and N_D is the donor-like dopant density. The electric field is determined from the gradient of the potential.

$$E = -\nabla\varphi \tag{3.2}$$

3.1.2 Continuity Equations

To determine the space charge density the continuity equations relate the change in carrier concentrations over time due to generation/recombination events and low-level current injection.

IPR2025-00845
Apple-Sony EX1026 Page 52

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot J_n + G_n - R_n \quad (3.3)$$

$$\frac{\partial p}{\partial t} = \frac{1}{q}\nabla \cdot J_p + G_p - R_p \quad (3.4)$$

where J_n and J_p are the current densities, G_n and G_p are the generation rates, and R_n and Rp are the recombination rates for electrons and holes respectively, for each set of terms.

3.1.3 Drift-Diffusion Transport Equations

Carriers move about within the semiconductor due to two phenomena known as drift and diffusion. Drift is the movement of carriers in response to an electric field, or the desire of carriers to minimize their electric potential. Diffusion is the tendency of carriers to redistribute to achieve uniform concentrations or resist concentration gradients. The current density equations resulting from these processes are shown below

$$J_n = qn\mu_n E + qD_n\nabla n \quad (3.5)$$

$$J_p = qp\mu_p E - qD_p\nabla p \quad (3.6)$$

where μ_n and μ_p are the electron and hole mobility, and D_n and D_p are the electron and hole diffusion constants.

3.2. Atlas Simulation Organization

The code for an Atlas simulation follows an organizational structure outlined in Fig. 3.1. Specific excerpts of the code will be presented in this chapter for discussion as it relates to device operation and the validity of the results. A typically simulation file can be found in Appendix I.

IPR2025-00845
Apple-Sony EX1026 Page 53

Group	*Statements*
1. Structure Specification	MESH REGION ELECTRODE DOPING
2. Material Models Specification	MATERIAL MODELS CONTACT INTERFACE
3. Numerical Method Selection	METHOD
4. Solution Specification	LOG SOLVE LOAD SAVE
5. Results Analysis	EXTRACT TONYPLOT

Fig. 3.1 Atlas Input Code Organizational Structure

First the structure of the device to be simulated is described by a grid of points known as a mesh. Regions within the mesh are specified as particular materials and electrodes and doping profiles are defined. Next the material properties are set for each region and the desired physical models are invoked. Contact characteristics for the terminals are specified, and any interface parameters such as surface generation rates, interface charge, or anti-reflective layers are defined. With the structure ready to be simulated, the numerical methods to be used for simulation are chosen. Finally the solution specification section allows the definition of the external stimulus applied during the simulation, and the location where results are saved. When a consistent solution has

IPR2025-00845
Apple-Sony EX1026 Page 54

been reached parameters can be extracted or the structure can be plotted with overlaid distributions.

3.2.1 Device Structure Specification

The first set of commands in an Atlas simulation defines the size of the structure and spacing of the grid points. The '`loc`' and '`spacing`' commands specify lines in the x and y directions that intersect to form nodes, where the x-direction is horizontal and the y-direction is vertical. Locations can be positive or negative but must be in ascending order. The spacing can be uniform or non-uniform and Atlas will automatically grade the spacing from one location to the next. Nodes are also connected diagonally creating triangular domains. Abrupt changes in grid spacing and obtuse triangles are avoided as they can make convergence difficult and cause inaccurate results.

It is difficult to obtain the optimum mesh as there is an inherent trade-off between the resolution of grid points and the resulting accuracy of the simulation, and the computational efficiency of the simulation. Fine resolution is needed for areas with large doping gradients or electric fields but not in areas of uniform doping or constant fields. To avoid excessively large structures there is a 20000 node limit for a two-dimensional simulation.

IPR2025-00845
Apple-Sony EX1026 Page 55

Fig. 3.2 Atlas example mesh definition code

The mesh definition for the simulations performed in this study are shown in Fig. 3.2. The structure was defined to have a variable width and thickness with constant node spacing in the x-direction and graded spacing in the y-direction. A fine resolution is specified at the light absorbing surface that gets coarser into the bulk of the device. The majority of the device thickness contains a rather coarse grid spacing of 5.0 microns before becoming fine again at the back surface. The thickness of the simulated device was typically set to 250 microns as determined by the substrate thickness, and the width was set to simulate the largest device possible within the available number of nodes.

A fair amount of time during the initial investigations centered around optimizing the vertical mesh spacing. Ensuring an adequate spacing of nodes at the light absorbing surface to accurately describe the absorption of short wavelength light was a particularly difficult problem and will be described further in Section 5.4 on Quantum Efficiency.

IPR2025-00845
Apple-Sony EX1026 Page 56

Fig. 3.3 Atlas simulation code defining electrode name as positions and doping profiles

Once the mesh has been defined, areas within the mesh are specified as a region and material. For this simulation the entire structure was defined as a continuous region of silicon. Next electrodes are added to the structure and the doping contours are defined as shown in Fig. 3.3. The N+ contact was defined to cover the entire surface of the structure and five P+ electrodes are defined on the backside of the device to contact individual pixels. The electrodes are given arbitrary names and numbers as identifiers.

The substrate is defined to be uniformly doped with an n-type impurity to a concentration of $1x10^{-12}$ cm^{-3}. The species of the dopant is not important, and the concentration is set to match the substrate manufacturer specification. Both the N+ and P+ regions are specified with Gaussian doping profiles, although the N+ region is defined

IPR2025-00845
Apple-Sony EX1026 Page 57

to have specific junction depth, and the P+ regions are defined with characteristic lengths that are related the standard deviation of the Gaussian. After the doping profiles have been defined a '`regrid`' statement is included to increase the mesh resolution in the regions where doping concentration varies rapidly. The smoothing of the mesh in these areas helps to prevent the solutions creating large discontinuities between nodes.

3.2.2 Material Models

With the basic structure of the device defined, the materials properties and physical models to be used during the simulation are shown in Fig. 3.4. The only material properties of the silicon that were adjusted from default values are '`TAUP0`' and '`TAUN0`,' the minority carrier lifetimes. They were defined as 1 millisecond as specified by the substrate manufacturer and shown in literature for float-zone silicon [17], but were also varied in some simulations to see the effect they have on device performance.

Fig. 3.4 Atlas code defining material and interface properties and specifying numerical models

In the '`models`' statement the temperature is defined as a variable because the device is designed to operate at low temperatures, so simulations were performed from

IPR2025-00845
Apple-Sony EX1026 Page 58

room temperature to as low as 170K. The remaining terms in the '`models`' statement activate specific physical models as follows: '`srh`' Shockley-Reed-Hall (SRH) Recombination, '`auger`' Auger Recombination, '`cvt`' Lombardi CVT Model, '`fermi`' Fermi-Dirac statistics, '`bgn`' Band-Gap Narrowing.

Fermi-Dirac Statistics

Fermi-Dirac statistics are typically approximated by Boltzmann statistics in situations where the Fermi level is greater than a few kT away from the band edge. The approximation begins to fail at low temperatures and high doping concentrations. As both will be present in this work, Fermi statistics have been enabled for all simulations. The probability that an available electron energy state (ε) in a semiconductor is occupied by an electron is

$$f(\varepsilon) = \frac{1}{1 + exp^{\left(\frac{\varepsilon - E_F}{kT_L}\right)}} \tag{3.7}$$

where E_F is the Fermi level, k is Boltzmann's constant, and T_L is the lattice temperature. Evaluation of the Fermi-Dirac integral by Atlas is handled using a Rational Chebyshev approximation scheme [22].

Bandgap Narrowing

In heavily doped semiconductors the conduction and valence bands broaden slightly causing a narrowing of the bandgap. This primary effect of this phenomenon is

IPR2025-00845
Apple-Sony EX1026 Page 59

to locally increase the effective intrinsic carrier concentrations and is modeled by Atlas using a spatially varying intrinsic concentration n_{ie} calculated as

$$n_{ie}^2 = n_i^2 exp^{\left(\frac{\Delta E_g}{kT}\right)} \tag{3.8}$$

where ΔE_g is the bandgap reduction calculated as

$$\Delta E_g = BGN.E\left[ln\frac{N}{BGN.N} + \sqrt{\left(ln\frac{N}{BGN.N}\right)^2 + BGN.C}\right] \tag{3.9}$$

where N is the dopant concentration and $BGN.E$, $BGN.N$ and $BGN.C$ are empirical values determined by Slotboom and de Graaf [23].

Once the amount of bandgap narrowing has been determined, the effect is introduced into the other models by adjusting the value of the bandgap used in their calculations accordingly. The primary motivation for the inclusion of this model was to account for any effects the heavy doping of the N+ surface may have on the absorption of light and photogeneration of carriers.

Shockley-Reed-Hall Recombination

The '`srh`' term invokes the Shockley-Reed-Hall Recombination statistics model shown in (3.10) to account for trap-assisted phonon transitions. The model uses the minority carrier lifetimes defined in the '`material`' statement to calculate the generation or recombination of carriers as

IPR2025-00845
Apple-Sony EX1026 Page 60

$$R_{SRH} = \frac{pn - n_{ie}^2}{TAUP0\left[n + n_{ie}e^{\left(\frac{ETRAP}{kT_L}\right)}\right] + TAUN0\left[p + n_{ie}e^{\left(\frac{-ETRAP}{kT_L}\right)}\right]} \quad (3.10)$$

where n_{ie} is the intrinsic carrier concentration, 'ETRAP' is the difference between the trap energy level and the intrinsic energy level, and T_L is the lattice temperature in degrees Kelvin. The default value of 'ETRAP' is 0 corresponding to the center of the band gap. This value was used for all simulations as the actual identity of contamination and defect states was not under investigation. The value for the lifetime combines the effects of the trap capture cross section, the thermal velocity, and trap density as

$$\tau_p = \frac{1}{\sigma_p v_{th} N_t} \quad (3.11)$$

An analogous statement can be written for electrons.

Lombardi CVT Model

Atlas contains many mobility models incorporating a variety of different phenomena appropriate for various situations. For the simulations presented here the Lombardi CVT Mobility models was used. Despite the fact that the model is intended for simulating the inversion layer of a MOSFET, it is the only model that includes the effects of doping concentrations, lattice temperature, and velocity saturation all in one model. It also includes effects due to transverse electric fields, although those won't be applicable to these simulations. The Lombardi model combines the effects of acoustic phonon

IPR2025-00845
Apple-Sony EX1026 Page 61

scattering (μ_{AC}), surface roughness (μ_{sr}), and optical inter-valley phonon scattering (μ_b) using Matthiessen's rule as

$$— \quad — \quad — \qquad (3.12)$$

where μ_T is the total effective mobility. The additional effects of coulombic scattering are incorporated using the same reciprocal method akin to adding resistors in parallel.

3.2.3 Numerical Methods

To perform a simulation, Atlas uses an internal discretization procedure to convert one to six coupled, continuous non-linear partial differential equations (PDE) to a discrete non-linear algebraic system that is solved at each point in the mesh. Following an initial guess, an iterative procedure successively refines estimates of the solution until the changes are smaller than convergence criteria.

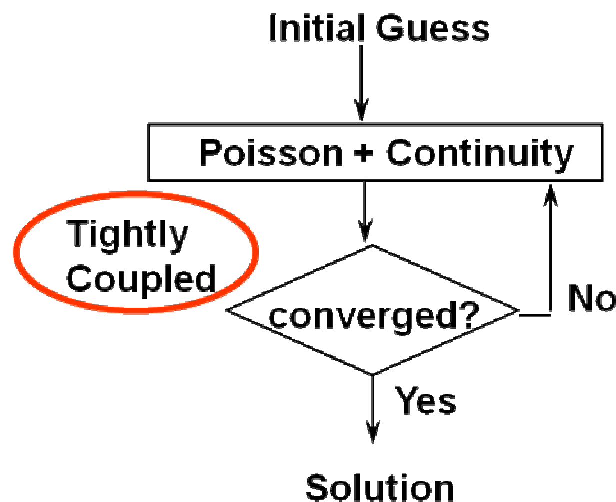

Fig. 3.5 Flow diagram for Newton method [24]

IPR2025-00845
Apple-Sony EX1026 Page 62

The discretization process in Atlas is performed using the Box Integration Method that creates a polygon around each node within which the equations are integrated. The fluxes at the edges of a polygon are equated with sources and sinks within. The non-linear algebraic system is solved by Atlas using Newton's method, by evaluating a linearized version of the entire non-linear algebraic system. A block diagram of Newton's method is shown in Fig. 3.5. This makes the problem rather large and may take a while to solve, however the final solution can typically be reached in only a few iterations. This method requires computers with large amounts of processing power and memory.

Fig. 3.6 Atlas specification of solution method and parameters

Figure 3.6 shows the Atlas command that invokes the numerical models used to evaluate the solution. The '`trap`' command specifies that if the solution does not converge the magnitude of the bias step will be reduced. A factor of ½ is the default, although this term can also be adjusted. The '`clim.dd`' term specifies a cut-off value for carrier concentrations below which the convergence of equations is not required. This factor is included as the bulk of the material is very lightly doped to begin with, and is

IPR2025-00845
Apple-Sony EX1026 Page 63

then depleted at reduced temperatures creating extremely low carrier concentrations in regions that are important to device operation.

3.2.4 Solutions Specification

The last section of code in an Atlas program specifies the external stimulus to be applied and the location where solution files are to be saved. A sample of the code used to define the solution conditions and specify output formats is shown in Fig. 3.7. The illumination sources for optical excitation are defined by the 'beam' statement. A location for the origination of the source is specified as well as the angle the incident beam makes with the surface of the structure.

Fig. 3.7 Atlas solution specification code sample

By default the illumination is monochromatic with uniform spatial intensity, although spectral sources and Gaussian intensity profiles can be specified. Atlas also

IPR2025-00845
Apple-Sony EX1026 Page 64

contains spectral intensity profiles for both AM0 and AM1.5 illumination. The 'rays' command splits the uniform illumination into many different rays that can be applied individually to evaluate the spatial dependence of illumination. The '`front.refl`' and '`back.refl`' commands specify that reflections off the first surface and back surface should be included in the calculations. These reflections are not included in the default ray tracing performed by Atlas.

Simulations are finally performed when the program reaches the first '`solve`' statement. An initial solution is performed using the '`init`' command with no applied bias on any of the terminals. The initial guess for potential and carrier concentrations is made from the doping profiles. A file where extracted parameters are saved is specified using the '`log`' command. On subsequent '`solve`' statements the '`prev`' command is used to specify that the previous solution should be used as the initial guess for the following solution. This improves chances of convergence for small bias steps and prevents the program from performing a new solve init returning the device to zero applied bias. Once the device has reached the operating conditions the log file storing the I-V data is closed using the '`log off`' command and a new log file is opened.

In the final solve statement the optical beam is enabled using the '`b1=1`' command which specifies beam1 is to have an intensity of 1 W/cm^2. In the solve statement above

the command '`scan.spot=1`' specifies that beam1 should be solve for each of the defined rays separately, much like scanning the beam across the surface of the wafer. The wavelength of the beam was also varied for uniform illumination much like the bias was incremented to determine spectral response.

3.3. Simulation Results

Simulations provide results in three different forms. Extract statements produce a single numerical result such as the junction depth or the photo-generation rate. Log files store the runtime output of a solve command including the potential and currents at all terminals and the result of any probe statements at every incremented step. When the solutions have been determined at operating conditions, the simulated structure can be visualized using Tony Plot to overlay contours of distributions including potential, electric field, current densities, and carrier concentrations.

3.3.1 Current Voltage Characteristics

The electrostatic solution for the device provides the current-voltage characteristics and an estimation of reverse bias leakage, or dark current. The thermal generation rate is largely determined by the minority carrier lifetime of the material. The minority carrier lifetime of $1x10^{-3}$ sec specified by the substrate manufacturer was implemented in Atlas. The current-voltage characteristics of the device are highly

IPR2025-00845
Apple-Sony EX1026 Page 66

dependent on the starting substrate doping level and the methods of processing. Since the substrate conditions are fixed and the effects of processing are hard to predict, the un-illuminated results provide simulation baseline.

Forward Bias

The forward bias current density of the *p-i-n* diode was simulated from an applied bias of 0 to 1.5 Volts. A graph of the simulation results in Fig. 3.8 show the device did not exhibit the traditional diode behavior. Even with high minority carrier lifetimes, at low bias conditions recombination is the dominant current transport mechanism due to the extremely wide intrinsic region. As the device turns on further it enters a forward bias region with an ideality factor of nearly one. At even higher biases the injected carriers begin to overwhelm the background ion concentration ($N_D \sim 10^{12}$ cm^{-3}) of the intrinsic region and the ideality factor degrades rapidly due to high level injection and series resistance.

IPR2025-00845
Apple-Sony EX1026 Page 67

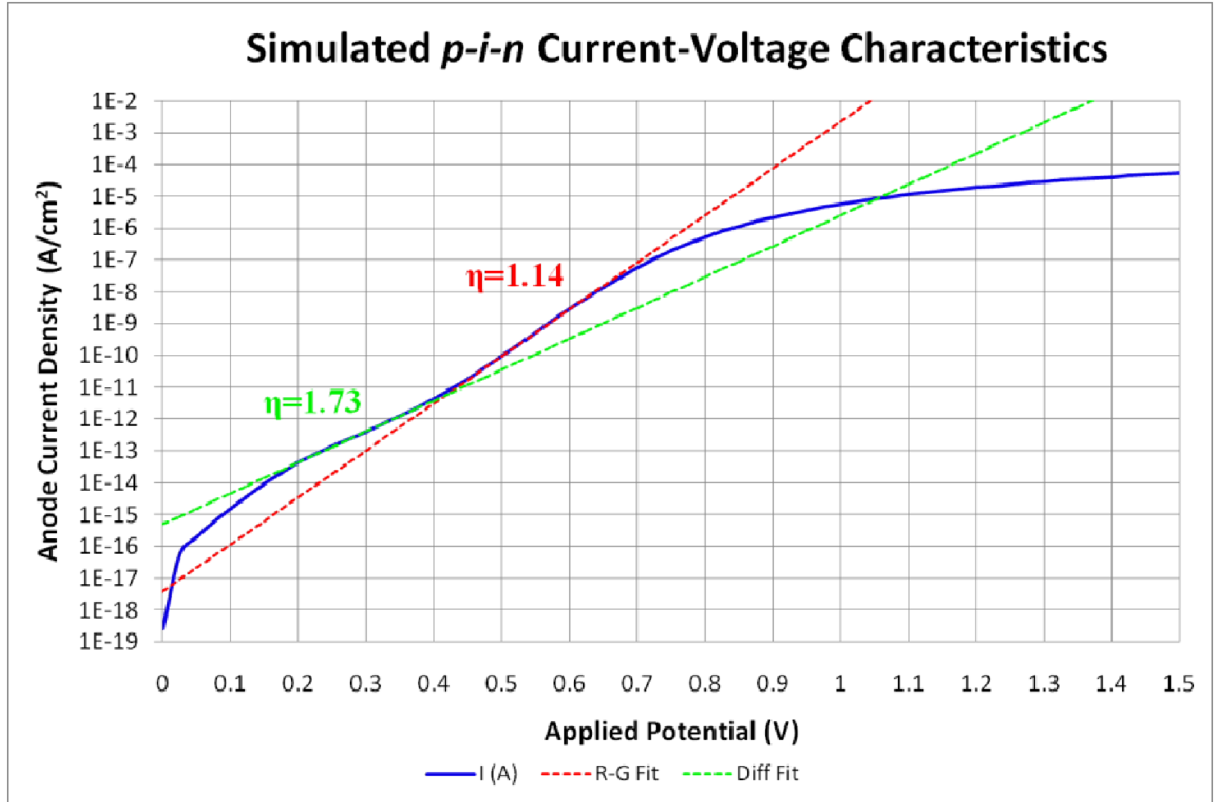

Fig. 3.8 Simulated *p-i-n* diode forward bias current density

Reverse Bias

The reverse bias leakage current presented another challenge for the simulation package. The results shown in Fig. 3.9 were generated for a structure of silicon with dimensions of 15 μm wide by 250 μm thick to represent a single pixel. Due to the very low doping levels, the resulting depletion width is hundreds of microns, and extends equally in all direction. The generation current is underestimated when attempting to simulate individual devices as the actual depleted region would create a hemisphere extending beyond the simulated structure.

IPR2025-00845
Apple-Sony EX1026 Page 68

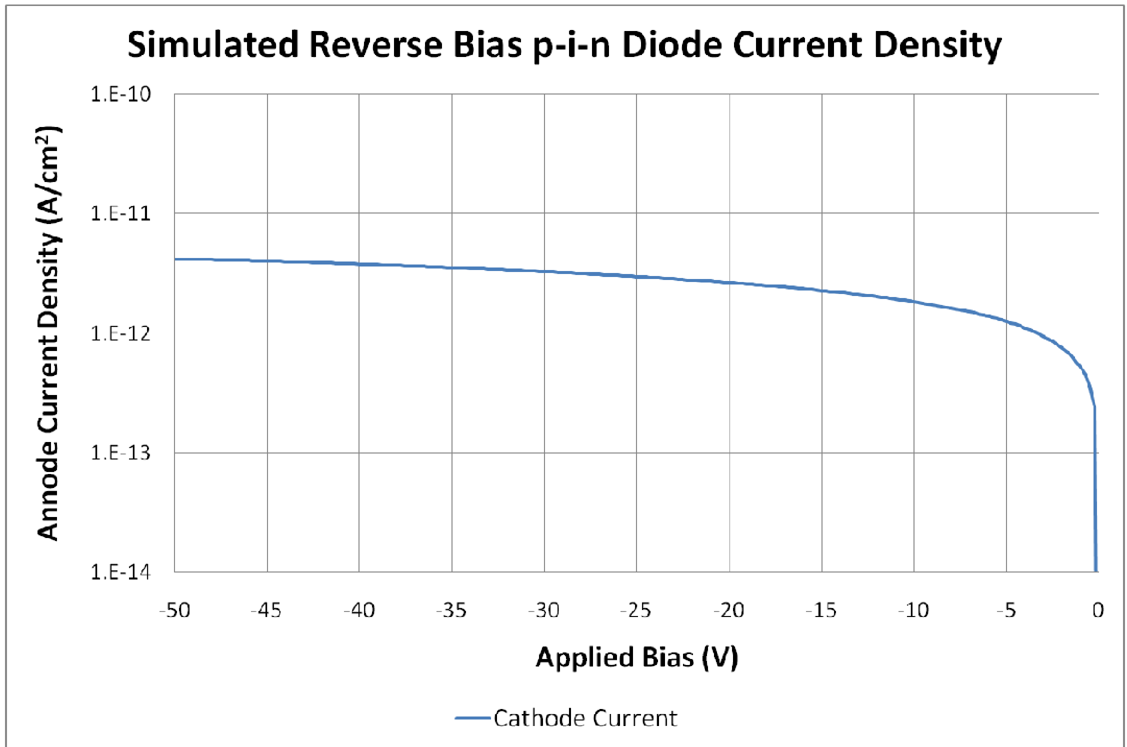

Fig. 3.9 Simulated *p-i-n* diode reverse bias leakage current vs. applied bias

The leakage current was also examined as a function of the minority carrier lifetime to confirm the expected relationship within the simulation and the choice of doping levels in the substrates. The results in Fig. 3.10 show that for a lifetime of 1 msec a current density of ~ $1x10^{-8}$ A/cm^2 can be expected. Although this seems rather large compared to the leakage goal, when scaled to low temperature the result is within an order of magnitude. As this lifetime represents the upper limit of lifetimes that can be realistically achieved in silicon, it shows that the goals set were realistic and can be attained with high quality manufacturing.

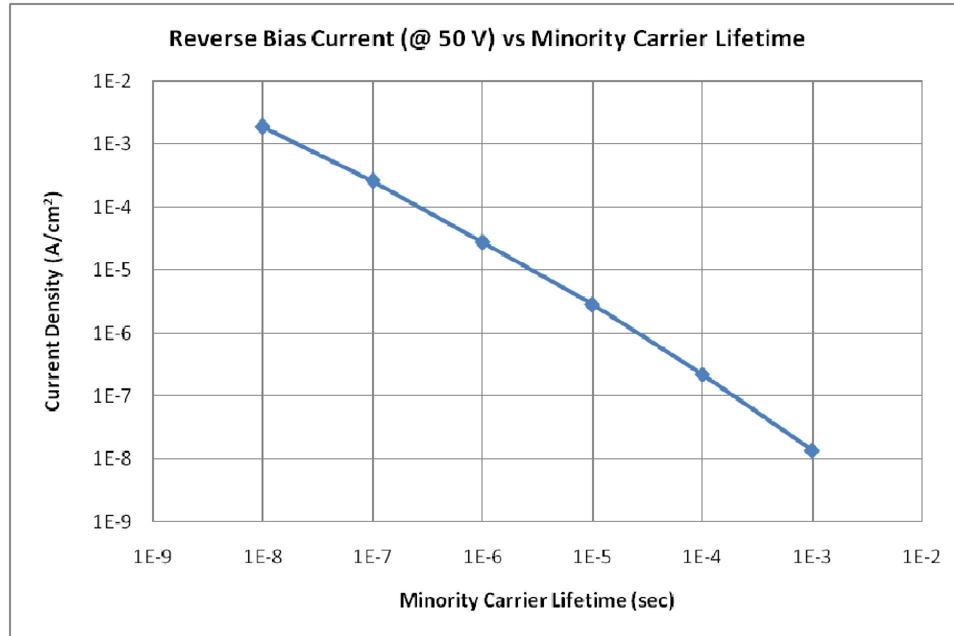

Fig. 3.10 Simulated reverse bias current density at 50 V vs. minority carrier lifetime

3.3.2 Quantum Efficiency

The quantum efficiency (QE) is a measure of the fraction of electrons extracted from a device per incident photon as a function of wavelength. It is an important metric used to quantify the performance of devices across different technologies. In this context QE can be separated into internal quantum efficiency (IQE), the ratio of collected carriers to absorbed photons, and external quantum efficiency (EQE), the ratio of collected carriers to incident photons. The IQE is only affected by losses within the device such as carrier recombination, or generation of carriers beyond a diffusion length away from the depletion region. EQE includes internal losses but also accounts for losses due to reflection and incomplete absorption. All further mentions of QE in this document will

IPR2025-00845
Apple-Sony EX1026 Page 70

refer to external QE unless specified otherwise. Both are affected by many aspects of the device design such as the distribution of the electric field and the location where carriers are generated.

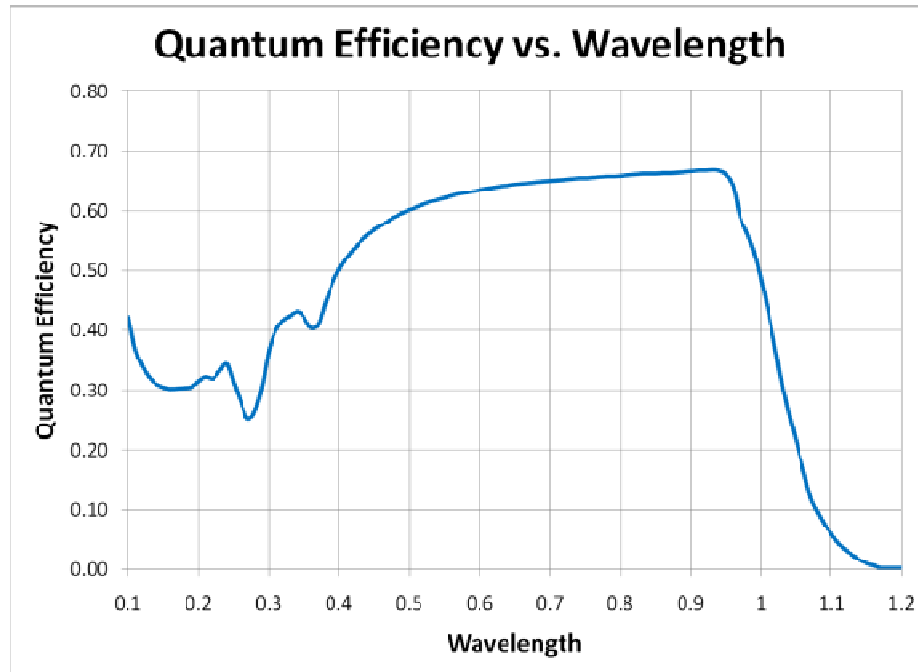

Fig. 3.11 Typical quantum efficiency curve from SILVACO Atlas

The biased device is illuminated with a uniform intensity monochromatic source as the wavelength is varied over the desired range (200 – 1200 nm). A typical quantum efficiency curve for a 250 μm thick silicon p-i-n diode operated at 2 kV/cm is shown in Fig. 3.11. In addition to providing a framework to characterize theoretical device performance, the simulations also provided an opportunity to optimize several process related parameters and explore the operating design space.

IPR2025-00845
Apple-Sony EX1026 Page 71

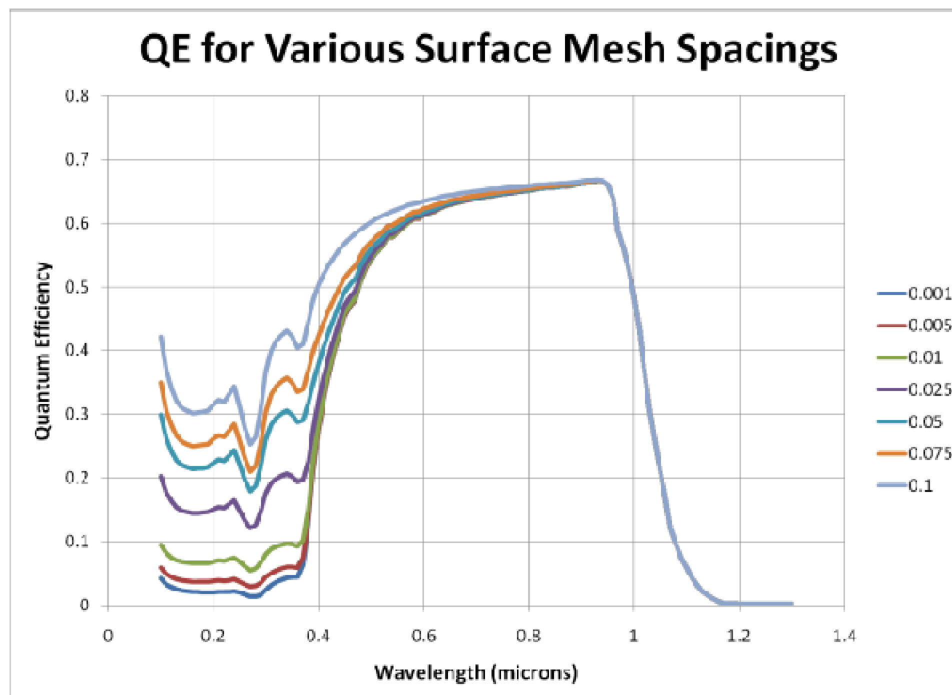

Fig. 3.12 QE curves for surface grid resolution from 0.001 to 0.1 µm

During the course of performing the simulations, a significant loss of signal for wavelengths less than 400 nm was observed. Recalling Fig. 2.6, this is where the absorption coefficient for silicon begins to increases rapidly reducing the penetration depth to about 10 nm. Further investigation found the resolution of the mesh at the surface of the device had a strong affect on the quantum efficiency in the short wavelength range. The simulated QE curves for surface grid resolutions ranging from 1 nm to 0.1 µm is shown in Fig. 3.12. The fine mesh spacing appears to overestimate recombination, reducing the amount of carriers that reach the depletion region. Unfortunately this effect was not noticed until after the completion of simulations.

IPR2025-00845
Apple-Sony EX1026 Page 72

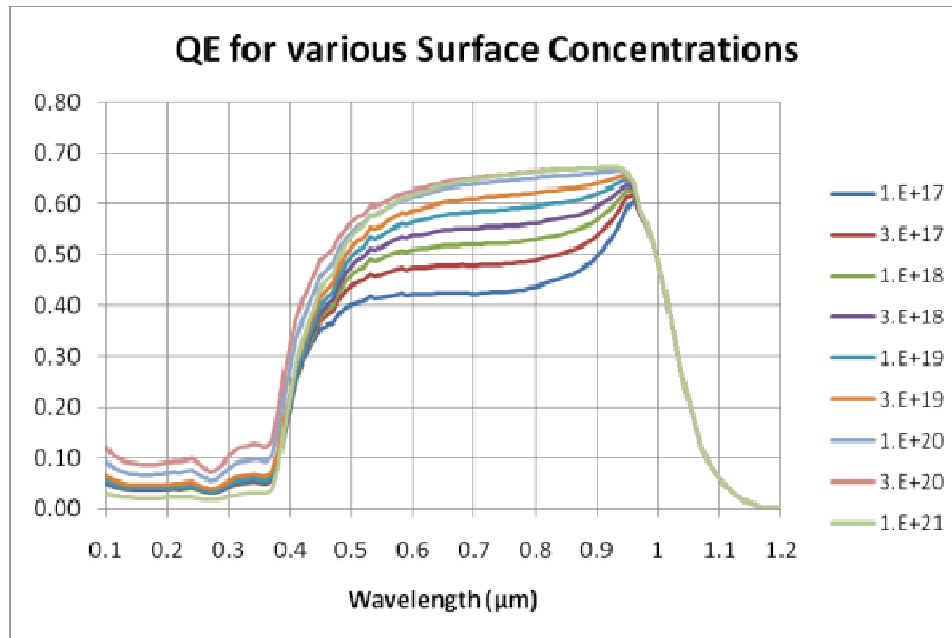

Fig. 3.13 Simulated QE curves for surface dopant concentrations ranging from $1x10^{17}$ cm^{-3} to $1x10^{21}$ cm^{-3}

The lack of a transparent conductive window on the light absorbing side of the detector placed an important constraint on the resistivity of the n^+ surface. The resistivity must be low enough that series resistance does not cause significant potential losses across the 1.9 x 3.8 mm detector. The peak surface concentration and homo-junction depth were varied to determine the optimum profile for device operation. The effect of varying the surface dopant concentration is shown in Fig. 3.13. Below a critical concentration of ~$1x10^{20}$ cm^{-3} the efficiency in the visible to near infra-red (NIR) range (500 – 900 nm) drops of significantly.

IPR2025-00845
Apple-Sony EX1026 Page 73

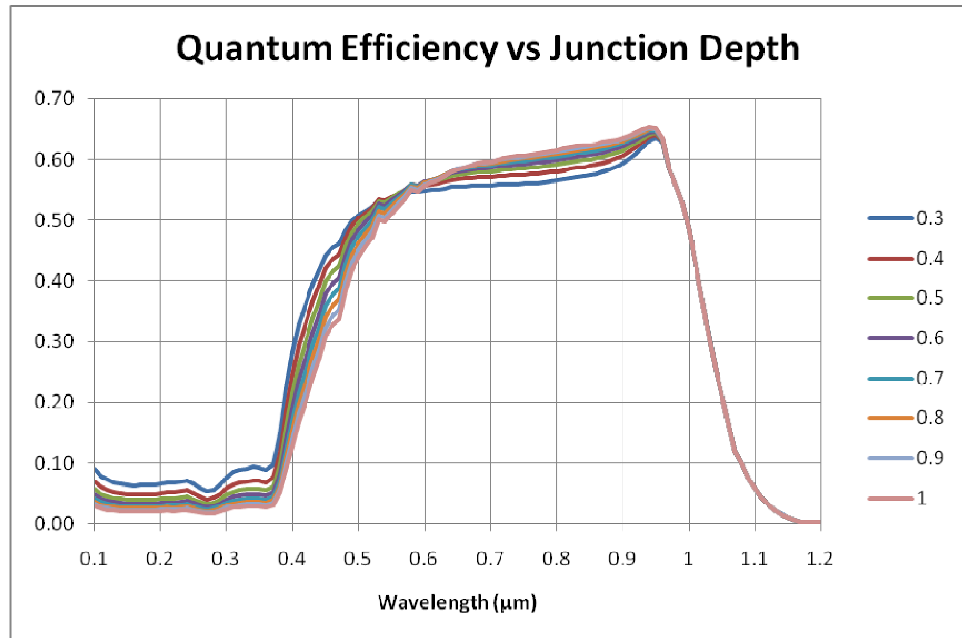

Fig. 3.14 Simulated QE curves for $n^{+} - n^{-}$ homo-junction depths of 0.3 – 1.0 μm

Junction depth is an important parameter because it determines thickness of the un-depleted surface layer. Carriers generated in this surface layer must diffuse into the depleted region before they can be collected. For the devices under investigation in this work the intersection of the dopant profile with the background concentration is referred to as a homo-junction because the dopant on the light absorbing side is n-type, the same as in the substrate, so no metallurgical junction is formed. By specifying the junction depth the steepness of the Gaussian profile can be controlled. The junction depth was varied from 0.3 to 1.0 μm and the resulting QE curves can be seen in Fig. 3.14.

The junction depth did not have a significant effect on device performance throughout the entire spectrum. The small negative effect reducing the junction depth

IPR2025-00845
Apple-Sony EX1026 Page 74

had at longer wavelengths is most likely due to the slight increase in resistance that would be caused by forcing the current flow through a shallower portion of the silicon. A positive effect can be seen at ~450 nm where the QE is increased by 10 %, although this does not translate to smaller wavelengths as was intended. The poor efficiency below 400 nm is most likely due to electron diffusion. The relative difference in electron concentration between the surface layer and the bulk silicon is ~8 orders of magnitude. The result is a significant tail to the profile created from the diffusion of electrons into the substrate region that increases the thickness of the un-depleted surface layer.

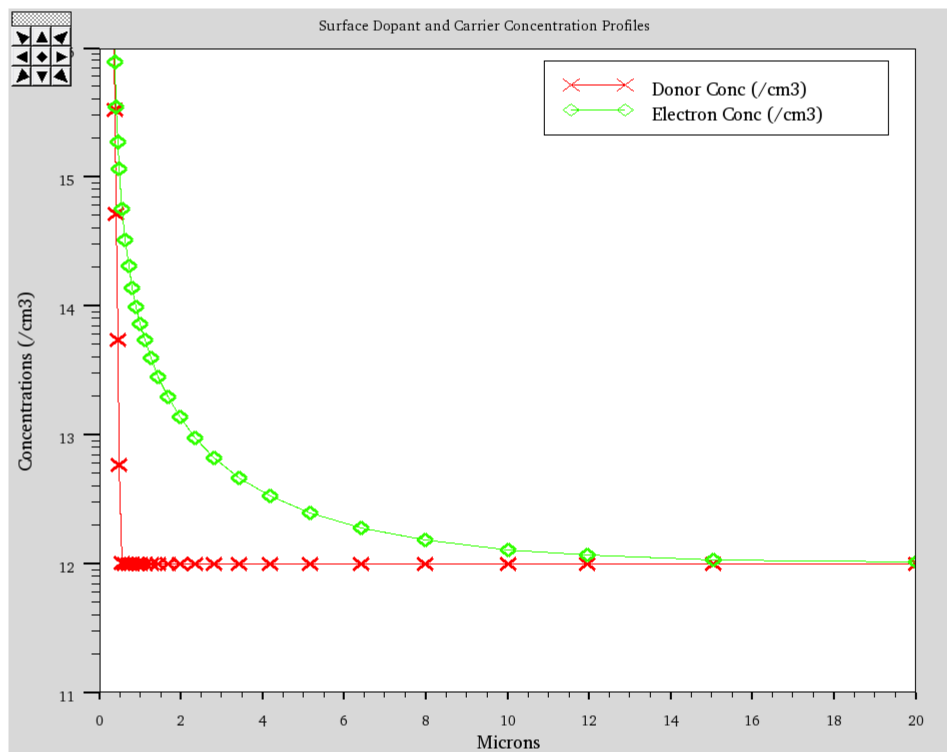

Fig. 3.15 Dopant and electron concentration profiles on *n*-type light absorbing side