A Study of CMOS Technologies for Image Sensor Applications

by

Ching-Chun Wang

Bachelor of Science in Electrical Engineering
National Tsing Hua University, Taiwan, June 1990

Master of Science in Electrical Engineering
National Tsing Hua University, Taiwan, June 1992

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2001

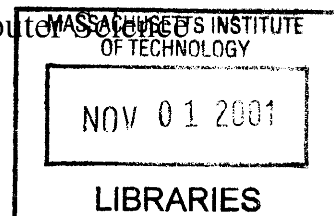
© 2001 Massachusetts Institute of Technology. All Rights Reserved.

Signature of Author ....
Department of Electrical Engineering and Computer Science
August 24, 2001

Certified by ....
Charles G. Sodini
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chair, Department Committee on Graduate Students
Department of Electrical Engineering and Computer Science

BARKER



A Study of CMOS Technologies for Image Sensor Applications

by

Ching-Chun Wang

Submitted to the Department of Electrical Engineering and Computer Science on August 24, 2001, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

CMOS (Complementary Metal-Oxide-Silicon) imager technology, as compared with mature CCD (Charge-Coupled Device) imager technology, has the advantages of higher circuit integration, lower power consumption, and potentially lower price. The advantages make this technology competent for the next-generation solid-state imaging applications. However, CMOS processes are originally developed for high-performance digital circuits. Fabricating high-quality embedded image sensors with CMOS technologies is not a straightforward task. This motivates the study of CMOS technologies for imaging applications presented in this thesis.

The major content of this study can be partitioned into four parts: (a) A two-stage characterization methodology is developed for sensor optimization, including the characterization of large-area photodiodes and comparative analyses on small-dimension sensor arrays with various pixel structures, junction types of the sensors, and other process-related conditions. (b) The mechanism of hot-carrier induced excess minority carriers occurred at the in-pixel transistors is identified and investigated. The influence of the excess carriers on imager performance is analyzed. Suggestions on the pixel design are provided. (c) Signal cross-talk between adjacent pixels is quantified and studied using a sensor array with a specially designed metal shield pattern, which exposes the center pixel and covers the others. The influence of cross-talk on color imager performance is analyzed. Process and layout improvements on cross-talk are also proposed. (d) The trend of pixel size reduction is investigated from the perspective of the achievable optical lens resolution. Using the modulation transfer function (*MTF*) as an index, optical simulations are performed to examine the relation between the lens resolution and the lens complexity.

Thesis Supervisor: Charles G. Sodini
Title: Professor of Electrical Engineering and Computer Science

Acknowledgements

I would like to thank my thesis advisor, Professor Charles Sodini, for his insightful technical guidance, steady encouragement, and unwavering support throughout my Ph.D. journey at M.I.T. I would also like to thank my thesis committee members, Professor Rafael Reif and Dr. Chung Wang at TSMC, for providing invaluable suggestions and advice in the thesis.

Special thanks to Dr. Shang-Yi Chiang, V.P. of Research and Development at TSMC, for offering me a great internship opportunity to work with their CMOS image sensor team. Dr. Chung Wang and Dr. Shou-Gwo Wuu hosted me there for seven months. I'm grateful for their hospitality and technical input. I would also like to thank Dr. Marc Loinaz for instructing me the CMOS imager characterization techniques during the three months of internship at Lucent Technologies.

I want to thank my fellow doctoral students and image sensor group members, Iliana Fujimori and Pablo Acosta Serafini, for the helpful discussion, encouragement and assistance. Special thanks to the former group member, Dr. Joseph Lusky, for providing help and suggestions during the first several rough years. I would also like to thank the other colleagues, including Donald Hitko, Daniel McMahill, Kush Gulati, Ayman Shabra, Andy Wang, Mark Peng, Mark Spaeth, Susan Dacy, and Yao-Joe Yang. Their assistance and friendship enriched my life.

I'm grateful for the endless support and encouragement I have received from my parents and family in Taiwan and also my girl friend, Hui-Ing Ma.

Funding for this work was provided through the M.I.T. Center for Integrated Circuits and Systems. TSMC provided partial support for chip fabrication.

Table of Contents

1 Introduction 17
- 1.1 Thesis Motivation 17
- 1.2 Thesis Contribution 19
- 1.3 Thesis Overview 21

2 Background and Review 23
- 2.1 General Background 23
- 2.2 CMOS Technology Scaling Considerations 25
- 2.3 Review of CMOS Imager Characterization 29
- 2.4 Review of Hot-carrier Induced Minority Carriers 30
- 2.5 Review of Cross-talk on Color Imager Performance 30

3 Characterization of CMOS Processes: Stage I – Large Area Photodiodes 33
- 3.1 Design of Large Area Photodiodes 33
- 3.2 Basic Photodiode Parameters 34
  - 3.2.1 Spectral Response 34
  - 3.2.2 Leakage Current 38
  - 3.2.3 Capacitance 40
- 3.3 Improving Transmission Coefficient of Dielectric Layers 40
  - 3.3.1 Theory 41
  - 3.3.2 Modeling 42
  - 3.3.3 Experimental Results 43
- 3.4 Comparison of Junction Types 44
  - 3.4.1 Quantum Efficiency 44
  - 3.4.2 Leakage Current 46
  - 3.4.3 Capacitance 50
- 3.5 Geometry Considerations 51
- 3.6 Summary 55

4 Characterization of CMOS Processes: Stage II– Small-Dimension Pixel Arrays 57
- 4.1 Design of 64 x 64 Pixel Arrays 57
- 4.2 Experimental Setup 67
- 4.3 Simulation 70
- 4.4 Imager Pixel Parameters 71
  - 4.4.1 Conversion Gain 72
  - 4.4.2 Spectral Response 74
  - 4.4.3 Leakage Current 75
  - 4.4.4 Sensitivity 77
  - 4.4.5 Noise 81
- 4.5 Comparison of Pixel Designs 96
  - 4.5.1 NW/Psub vs. Comb-shape N+/PW/Psub Photodiode APS 97
  - 4.5.2 3-T Photodiode APS vs. 4-T Photodiode APS 98
  - 4.5.3 Photodiode vs. Photogate APS 102

4.6 Comparison of Process Parameters 104
4.6.1 Substrate Types 104
4.6.2 Junction Types 106
4.6.3 Implant Conditions 108
4.7 Test Images 113
4.8 Summary 113

5 The Effect of Hot Carriers on the Operation of CMOS Active Pixel Sensors 117
5.1 Mechanism 117
5.2 Experimental Results 119
5.3 Discussions 129
5.3.1 Hot-carrier Effect in Normal Imager Operation 129
5.3.2 Suggestions on Pixel Design 132
5.4 Summary 133

6 The Influence of Pixel Cross-talk on CMOS Color Imagers 135
6.1 Fundamentals of Color Imagers 135
6.1.1 Basic Principle of Color Imagers 136
6.1.2 Trichromatic Matching Experiment 137
6.1.3 Chromaticity Diagram 142
6.1.4 Color Analysis Functions 144
6.1.5 Color Imagers and Cross-talk 145
6.2 Mechanism of Pixel Cross-talk 145
6.2.1 Minority Carrier Diffusion 145
6.3 Characterization of Cross-talk 146
6.3.1 Test Pattern Design 146
6.3.2 Experimental and Simulation Results 147
6.4 Amplification of Color Error due to Cross-talk 150
6.4.1 Color Error 150
6.4.2 Color Error Amplification 152
6.5 Layout and Process Improvements on Cross-talk 155
6.6 Summary 156

7 Pixel Size Scaling and Optical Lens Design 157
7.1 Scaling Trend of Pixel Size 157
7.2 Fundamentals of Lens Design 157
7.2.1 Focal Length, F-number, and Field of View 157
7.2.2 Lens Optics 159
7.3 Computer Simulation on Lens Systems 161
7.3.1 Modulation Transfer Function 161
7.3.2 Case Study 165
7.4 Discussion 171
7.5 Summary 173

8 Conclusions and Future Work 175
8.1 Conclusion 175
8.2 Future Work 176

Bibliography 179

Appendix A Photometry and Radiometry 185

Appendix B Standard Illuminants 189

Appendix C Lens Data and Specifications 191

C.1 The Two-element Lens Design 191
C.2 The Three-element Lens Design 192
C.3 The Six-element Lens Design 193
C.4 The Two-element Lens Design with One Aspherical Surface 195

List of Figures

Figure 3.1: Normalized optical power (a), and the relative carrier generation rate (b) in silicon 35
Figure 3.2: Silicon absorption coefficient (a), and the corresponding 1/e absorption depth (b) versus wavelength 36
Figure 3.3: Photo-carrier collection mechanism 36
Figure 3.4: Test setup for spectral response characterization 37
Figure 3.5: Transmission coefficient ratio between SiN/Ox/Si (type-A) and Ox/Si (type-B) dielectric layers 43
Figure 3.6: Cross-sections of N^+/Psub and NW/Psub photodiodes 44
Figure 3.7: Q.E. of photodiodes with the 0.5-μm and 2.0-μm processes 45
Figure 3.8: N^+/PW/Psub photodiode structure and the doping profile 46
Figure 3.9: Q.E. of photodiodes with the 0.6-μm twin-well process 47
Figure 3.10: Leakage current density of photodiodes with the 0.5-μm and 2.0-μm processes 48
Figure 3.11: Leakage current density of photodiodes with the 0.6-μm twin-well process 49
Figure 3.12: Arrhenius plot of leakage current density for photodiodes with the 0.5-μm process 50
Figure 3.13: Capacitance of the photodiodes with the 0.5-μm and 2.0-μm processes ...51
Figure 3.14: Top view of the square-shape (left) and grid-shape (right) photodiodes ...52
Figure 3.15: Q.E. ratio between the grid-shape and square-shape N^+/Psub photodiodes 53
Figure 3.16: Arrhenius plot of leakage current for the grid-shape and the square-shape N^+/Psub photodiodes with the 0.5-μm process 54

Figure 4.1: Layout floor plan of the chip (left) and the 64 x 64 test pixel array (right) .58
Figure 4.2: Circuit diagram of the 3-T pixel cell and the associated periphery circuit ..59
Figure 4.3: Timing diagram for the 3-T photodiode APS array 60
Figure 4.4: Circuit diagram of the second-stage source followers 61
Figure 4.5: Hierarchical multiplexer in the column decoder 62
Figure 4.6: Typical multiplexer implementation 62
Figure 4.7: Circuit diagram of the output buffer 63
Figure 4.8: Schematic of the level shifting circuit for reset signal 64
Figure 4.9: Layout of the imager chip 65
Figure 4.10: Layout of a 64 x 64 photodiode APS array 66
Figure 4.11: Experimental setup for sensor array characterization 67
Figure 4.12: Photo of the test board 68
Figure 4.13: Devices for temperature-varying experiment 70
Figure 4.14: Layout and schematic of the 3-T NW/Psub photodiode APS 72
Figure 4.15: Measured Q.E. for the NW/Psub photodiode APS and the large NW/Psub photodiode 75
Figure 4.16: Dark signal versus integration time for the NW/Psub photodiode APS at various temperatures 76

Figure 4.17: Arrhenius plot of leakage current for the NW/Psub photodiode APS 77
Figure 4.18: Transmittance of the visible light filter 79
Figure 4.19: Sensitivity and random noise versus illumination for the NW/Psub photodiode APS 80
Figure 4.20: S/N ratio versus illumination for the NW/Psub photodiode APS 80
Figure 4.21: Data set for processing random noise and fixed pattern noise 81
Figure 4.22: Thermal noise model 84
Figure 4.23: Random noise components versus integration time for the NW/Psub photodiode APS 87
Figure 4.24: P-P FPN and C-C FPN versus illumination 89
Figure 4.25: FPN versus integration time 89
Figure 4.26: Pixel output voltage versus row number with different T_{int} 90
Figure 4.27: Histogram of the dark signal distribution 91
Figure 4.28: Image before and after C-C FPN correction 92
Figure 4.29: C-C FPN measurement before and after C-C FPN reduction circuit 93
Figure 4.30: Random noise and P-P FPN versus illumination 94
Figure 4.31: Dark current shot noise and P-P FPN versus integration time 95
Figure 4.32: Layout and schematic of the N^+/PW/Psub photodiode APS 97
Figure 4.33: Layout and schematic of the 4-T photodiode APS 99
Figure 4.34: S/N ratio versus illumination for the 3-T photodiode APS and the 4-T cascode photodiode APS 100
Figure 4.35: Layout and schematic of the photogate APS 102
Figure 4.36: Timing diagram for the photogate APS 103
Figure 4.37: Q.E. comparison between the NW/Psub photodiode APS and the photogate APS 104
Figure 4.38: Simulation and experiment of Q.E. for the photodiode APSs on epi and bulk wafers 105
Figure 4.39: Simulation and measurement of Q.E. for the photodiode APSs with the three junction types 107
Figure 4.40: Arrhenius plot of leakage current for the photodiode APSs with the three junction types 108
Figure 4.41: Simulated doping profile for the N-type implant dose split 109
Figure 4.42: Simulation and measurement of Q.E. for the N-type implant dose split .110
Figure 4.43: Arrhenius plot of leakage current for the N-type implant dose split 111
Figure 4.44: Simulated doping profile for the N-type implant energy split 111
Figure 4.45: Simulation and measurement of Q.E. for the N-type implant energy split 112
Figure 4.46: Arrhenius plot of leakage current for the N-type implant energy split 112
Figure 4.47: Test images of the 0.35-μm (a) and 0.25-μm (b) NW/Psub photodiode APS arrays 113

Figure 5.1: Positive feedback mechanism for hot-carrier induced excess minority carriers 118
Figure 5.2: Pixel structure used for hot-carrier effect characterization: (a) pixel layout and (b) circuit diagram 120

Figure 5.3: Temporal evolution of photodiode voltage change in a dark environment after reset 121
Figure 5.4: Dark signal versus V_{rst} with different Vdd values 122
Figure 5.5: Dark signal versus V_{ds} of Q2 with different Vdd values 123
Figure 5.6: Dark signal versus simulated maximum E-field in the pinch-off region of Q2 123
Figure 5.7: Dark signal versus V_{rst} with different I_{bias} values 124
Figure 5.8: Dark signal versus Vdd with different V_{rst} values 125
Figure 5.9: Dark signal versus V_{ds} of Q2 with different Vdd values for the cascode mode operation 125
Figure 5.10: Arrhenius plot of dark signal under three different bias conditions 126
Figure 5.11: Dark signal versus q/k_BT under three different bias conditions in a linear plot 127
Figure 5.12: Excess dark signal versus row number with stress at the 33rd row 128
Figure 5.13: Excess dark signal versus row number with a higher stress at the 33rd row 128
Figure 5.14: Center portion of Figure 5.13 in log scale 129
Figure 5.15: Excess V_{out} versus V_{out} without the H-C effect for the APS arrays operating in normal mode and cascode mode 131
Figure 5.16: V_{out-HC} versus V_{out} without the H-C effect for an APS array with a β of two times the characterized value 131
Figure 5.17: Simulation results on maximum electric field in the pinch-off region with different gate lengths 132
Figure 5.18: (a) Typical 3-T APS design; (b) modified pixel design for mitigating the H-C effect 133

Figure 6.1: Procedure of color image sensing, processing, and display 136
Figure 6.2: RGB Bayer pattern 137
Figure 6.3: Schematic of the color matching experiment 138
Figure 6.4: Tristimulus values vs. wavelength corresponding to primary stimuli R/G/B 139
Figure 6.5: The CIE color-matching functions for the 1931 Standard Colorimetric Observer 141
Figure 6.6: The CIE 1931 Chromaticity Diagram 143
Figure 6.7: The CIE 1976 Uniform Chromaticity Scale Diagram 144
Figure 6.8: Minority carrier diffusion mechanism 145
Figure 6.9: Cross-section of the cross-talk test pattern 147
Figure 6.10: Measured cross-talk versus wavelength for the adjacent pixels along the horizontal, vertical, and diagonal directions 148
Figure 6.11: Cross-talk versus wavelength from the 2-D device simulation 149
Figure 6.12: Quantum efficiencies of the pixels 150
Figure 6.13: P-P FPN versus illumination and the normalized P-P FPN with the average signal versus illumination 151
Figure 6.14: Spectral responses of the R/G/B pixels with and without cross-talk 154
Figure 6.15: Color gamuts with and without cross-talk on the 1976 CIE Uniform Chromaticity Scales chart 154

Figure 7.1: Ray bending according to Snell's law 159
Figure 7.2: Light rays and wave front in a lens 160
Figure 7.3: Object and image of a sinusoidal pattern in spatial domain 161
Figure 7.4: Intersection points of rays with entrance pupil 162
Figure 7.5: Spot diagrams associated with two incident angles 163
Figure 7.6: Point and line spread functions 164
Figure 7.7: 2-D lens drawing of the two-element lens design 167
Figure 7.8: MTFs of the two-element lens design 167
Figure 7.9: 2-D lens drawing of the three-element lens design 168
Figure 7.10: MTFs of the three-element lens design 168
Figure 7.11: 2-D lens drawing of the six-element lens design 169
Figure 7.12: MTFs of the six-element lens design 169
Figure 7.13: Distortion of images: (a) no distortion; (b) positive distortion; (c) negative distortion 171

Figure A.1: Normalized luminous efficiency function for photopic vision, V(λ), and scotopic vision, V'(λ). 185
Figure B.1: Normalized spectral energy distribution of CIE standard illuminants 189

List of Tables

Table 3.1: Transmission coefficient of dielectric layer configurations 42

Table 4.1: Characteristics of random noise sources 96
Table 4.2: Characteristics of fixed pattern noise sources 96
Table 4.3: Comparison of NW/Psub and comb-shape N^+/PW/Psub photodiode APSs . 98
Table 4.4: Comparison of 3-T and 4-T cascode NW/Psub photodiode APSs 100
Table 4.5: Trade-off between 3-T and 4-T cascode photodiode APSs 101
Table 4.6: Comparison of the photodiode APSs with three different junction types ... 107

Table 5.1: Characteristics of the 4-T photodiode APS array in the experiments 120

Table 6.1: Characteristics of the APS array in the experiments 146

Table 7.1: Comparison of lens performance between different designs....................... 170
Table 7.2: Comparison of lens performance between 2-element lens designs with and without a aspherical surface 173

Table A.1: Approximate natural illuminance levels 187

Chapter 1

Introduction

1.1 Thesis Motivation

Solid-state imaging technology has extensive applications ranging from consumer electronics (e.g. security cameras, digital still cameras, and camcorders) and machine vision systems (e.g. product quality control, robot vision and automatic cruise control systems for vehicles) to scientific research (e.g. planet and biological cell filming). This market is further nourished by the booming computer networking and wide-band, mobile communication technologies that allow the transmission of high-quality images and real-time video. Such a rapidly growing market necessitates the development of imagers with low price, low power consumption, light weight, and highly integrated functions. Therefore, imager developers actively search for substitutions of the widely used CCD (Charge-Couple Devices) imagers due to their poor circuit integration capability and high power consumption. Among all the imager technologies, CMOS (Complementary Metal-Oxide-Silicon devices) imager technology emerges and becomes the candidate that fulfills the power and integration requirement for the next-generation imaging systems. The availability of tiny transistors in CMOS processes further allows pixel-level innovations for focal-plane signal processing. In addition, CMOS processes are the leading and the most widely available technologies among all the silicon processes. This makes CMOS imagers potentially cheaper alternatives to CCD imagers.

Current CMOS imagers, however, have some drawbacks. The most significant problem is attributed to the fact that CMOS processes are originally developed for efficient logic/mixed-mode circuit operations rather than for high-quality image sensors. Process steps such as doping levels and thermal treatments are all optimized for higher transistor performance. As a result, the available image-sensor structures from a standard CMOS process, such as photodiode or photogate sensors, can only be implemented with existing junctions and structures available to the specific process. Usually the performance of such

sensors is not optimized. Furthermore, as the development of CMOS processes enters the deep sub-micron era, problems such as poor sensitivity, high leakage current, and high fixed pattern noise occur. In order to ensure adequate image sensor performance, process modifications become necessary. The current trend in CMOS imager technology is to develop a specialized process module dedicated for building image sensors. This process module can be integrated with other standard CMOS manufacturing procedures. This modular approach is similar to that for embedded DRAM and RF components. In this way, the performance of the image sensors and the on-chip circuitry can be optimized separately.

To develop an optimized CMOS process for imaging applications, a thorough characterization is essential to fully understand the capability of a specific process. For example, pixels with different sensor structures and process parameters should be evaluated to explore the advantages/disadvantages, and the trade-off on image sensor parameters. Based on the characterization results, the optimized pixels or improvements on critical process steps can be proposed and the aforementioned specialized process module can thus be developed. Such an incentive provokes developing an effective characterization methodology, in which test structures are built to investigate CMOS processes from an imager perspective. This is one major objective of this thesis.

Besides the characterization of the pixels and image sensors, the influence of circuit operations to the sensor performance is another focus of this study. During characterization, it is noticed that a high drain-to-source voltage at the in-pixel transistors induces a higher dark signal. The mechanism of this phenomena is identified to be hot-carrier induced excess minority carriers due to the high electric field near the drain of the transistors. This effect results in nonlinearity of imager response as the transistor size of CMOS processes continues to shrink. It is desired to describe the verification of the hot-carrier mechanism, quantify its influence to the image sensors, and provide design suggestions in the thesis.

Signal cross-talk between adjacent pixels is another topic in this study. As the pixel size continues to reduce, the cross-talk between adjacent pixels is expected to increase and color mismatch between pixels will be degraded. By studying the cross-talk mechanism and utilizing the principles developed in color theory, this thesis quantitatively estimates

the degradation factor of color mismatch and discusses the approaches to alleviate pixel cross-talk.

Pixel size reduction can be achieved with CMOS transistor scaling. However, a reduced pixel size poses a more stringent specification on the optical lens resolution. Using an optical simulation software, this thesis explores the relation between the number of lens elements and its resolution limit. The trade-off between pixel size reduction and the necessity of increased lens complexity is discussed.

To summarize, this thesis investigates the characteristics of pixel designs, sensor types, and process parameters for sensor optimization, the influence of in-pixel circuit operation on image sensor performance, the effect of signal cross-talk on color imagers, and the requirement of enhanced lens resolution with pixel size reduction. It is expected that the findings of this thesis can improve the performance of CMOS imagers and thus their advantages of low power, high integration, and low cost can be realized in the imager market.

1.2 Thesis Contribution

The contributions of the thesis mainly addresses an image-sensor characterization methodology, the physical mechanisms related with image sensor performance including hot-carrier effects and signal cross-talk, and the increased requirement on the optical system to accommodate the down-scaling trend of pixel size. The specific contributions are listed below.

1. A two-stage characterization approach is demonstrated to optimize photosensor and pixel performance for CMOS imagers. The first stage is characterizing the imaging-related junction properties and optimizing the transmission coefficients of the dielectric layers above silicon using large-area photodiodes. The size of the photodiode is 500 x 500 μm^2. Different junction types, layout strategies, and configurations of the dielectric layers are compared. The results provide first-order sensor parameters that are practical reference for making test sensor arrays used in the second-stage characterization. The characterization compares the sensor performance using small-dimension active-pixel-sensor (APS) arrays with various

pixel designs, sensor types, and process parameters. The dimension of the sensor arrays is 64 x 64. Several goals are targeted with this characterization methodology. The first goal is to explore an optimized sensor structure for a specific CMOS process. The second is to locate the process parameters that significantly influence the sensor performance for further improvement. The third goal is to investigate the feasibility of utilizing the test structures for wafer-level testing or lot-to-lot performance monitoring.

2. Excess minority carriers induced by hot carriers at the source follower transistors of active pixel sensors (APS) are experimentally observed using a 4-T NW/Psub sensor arrays fabricated with a standard 0.35-μm process. The number of carriers absorbed by photodiodes depends on bias conditions and consequently becomes optical-signal dependent. A cascode 4-T APS is more sensitive to this effect due to its small sensing capacitance. Temperature-varying experiments are performed to confirm this mechanism. The spatial range of the excess carriers is quantified to be within ~30 μm around the stressed transistors under normal operating conditions. This effect is expected to become more important as the minimum feature of CMOS processes continues to shrink. Suggestions on pixel design are provided.

3. Cross-talk phenomena of CMOS image sensors is studied using a 64 x 64 NW/Psub APS array with pixel size 4 x 4 μm^2. The major cross-talk mechanism is identified to be photo-carrier diffusion in the lightly doped substrate. The cross-talk not only reduces the imager's spatial resolution; it further amplifies the pixel-to-pixel color mismatch due to intrinsic pixel noise. The characterization results and the principles of color theory are utilized to quantify the degradation factor of color mismatch. The lowest illumination level under which the color mismatch is severely degraded is also predicted. The approaches to improve cross-talk are discussed.

4. Pixel size reduction is evaluated from a perspective of optical lens systems. An optical simulation software is utilized to study the relation between the minimum resolvable pixel size and the requirement on the number of lens elements. The Modulation Transfer Function (*MTF*) is used as a measure of resolution.

1.3 Thesis Overview

The remaining portion of this thesis is divided into seven chapters. Chapter 2 is dedicated to a background study and literature review, which describes the state of the art development of CMOS processes for imaging applications. Chapter 3 describes the characterization methodology of the large-area photodiodes. The sensor-related junction properties include quantum efficiency, leakage current, and junction capacitance. The configurations of the dielectric layers on top of silicon is also investigated to improve the optical transmission coefficient. Chapter 4 is dedicated to the characterization and comparison of small-dimension sensor arrays. The characterized parameters include conversion gain, quantum efficiency, leakage current, sensitivity, and noise sources. The characterization approach is utilized to compare pixels with different pixel designs, sensor types, and process parameters. Chapter 5 describes the effect of hot-carrier induced excess minority carriers resulting from the stressed in-pixel transistors. The hot-carrier mechanism is identified with a temperature-varying experiment and supplemented with measurements under various bias conditions. The spatial distribution of the minority carriers is characterized. Suggestions on pixel design is provided. Chapter 6 focuses on the degradation of color mismatch due to pixel cross-talk. The diffusion of photo-carriers in the lightly doped substrate is identified to be the major cross-talk mechanism. The degradation factor of color mismatch is calculated with the principles developed in color theory. Approaches to improve cross-talk are discussed. Chapter 7 evaluates the requirement of enhanced lens performance to accommodate pixel size reduction. Case studies on lens designs with different number of lens elements is performed to explore the relation between lens complexity and the achievable resolution. Chapter 8 concludes the thesis work with a summary of the studies and provides directions on the future work for further improvement of CMOS imagers.

Chapter 2

Background and Review

2.1 General Background

The idea of MOS imagers was first proposed in late 60's [1, 2], about the same time as CCD imagers [3]. Because MOS imagers required the incorporation of transistors into pixels, which was not feasible at that time due to large transistor size, CCD imagers have become the dominant solid-state imaging technology ever since. However, with the device shrinking trend, current CMOS imagers can be fabricated with a reasonable fill factor (20%–50%) while maintaining a reasonable pixel size (4 μm–8 μm). Therefore, the research effort and market share of CMOS imagers are gradually catching up with those of the CCD imagers. State-of-the-art CMOS imager products have successfully demonstrated large format arrays with on-chip analog-to-digital converters [4] and camera-on-a-chip systems which incorporate imagers, digitizers, and signal processors on the same chip [5, 6].

In general, CCD and CMOS imagers have no difference in their photo-sensing principle. Both devices utilize silicon as the sensing material and the collected photo-charges by the image sensors represent the intensity of optical signal. What differentiates these two types of imagers is where and how the photo-charges are converted into an electrical signal. The discrepancies between the manufacturing processes are attributed to the difference on their charge-to-voltage/current conversion schemes. Take CCD imagers for example, the optical signal stays in charge format in the pixel. The charges are transported sequentially through a series of coupled-gates until the final floating diffusion node performs charge-to-voltage conversion [7]. In order to assure the amount of photo-charges is intact during transportation, special device structures, such as coupled poly gates and buried CCD channels, are developed. The CCD process has evolved into a specialized process among silicon technologies to accommodate these device structures. With more than 20 years of research and development, the CCD imagers are able to achieve high-quality

image sensing and have become the dominate product in the solid-state imaging market. Unfortunately this specialized process is not suitable for building efficient transistors and circuits. Therefore, a CCD imager usually requires another supporting chip to provide control signals and perform signal processing. Whereas for CMOS imagers, the charge-to-voltage/current conversion is performed in a much earlier stage. For CMOS active pixel sensors [8, 9], the conversion is performed in the floating-diffusion sensing nodes within the pixel; for CMOS passive pixel sensors [10], the photo-charges is dumped onto to the column line and sensed by the column amplifiers; for CMOS logarithmic pixels [11], the photo-current is converted into voltage format within the pixel. Because of the early charge-to-voltage/current conversion in CMOS imagers, the electrical signal can be processed with efficient circuits which are available from standard CMOS processes.

From a process perspective, it is possible to integrate a CMOS module in a CCD process to build circuits on a CCD chip or to develop a special photo-sensor module in a CMOS process to improve CMOS imager quality. In both ways, high-quality photo-sensors and efficient circuits can be achieved on the same chip. However, building a CMOS process module in a CCD process is not as cost effective as developing an embedded image sensor module in a CMOS process. This is because process conditions for CCD and CMOS are incompatible with each other. It takes many more masking steps to build this hybrid CCD-CMOS process. Usually the processing fee is proportional to the number of masking steps and the yield decreases with it. Therefore, a CCD-CMOS process is rarely developed. On the other hand, the embedded image sensor module can be more easily integrated into a CMOS process since all it takes is an optimized junction to provide high-performance sensors.

Low power consumption is an important criteria for modern mobile electronics. In this area, CMOS imagers provide better performance than CCD imagers. That is because CCD imagers require high frequency clock signals on their coupled gates for charge transfer. Whereas CMOS imagers consumes power only when a particular row or column is addressed. A typical CCD imaging system consumes power on the order of 1W; whereas a CMOS imager with the same function typically consumes power on the order of 100mW.

CMOS imagers have another advantage that is difficult to achieve by the CCDs — the pixel-level innovation. Due to the small transistor size available from modern CMOS pro-

cesses, several transistors or even a small circuit can be integrated within a pixel without severely degrading the fill factor. Novel pixel-level innovation enables new imager functions, such as dynamic-range expansion [12, 13, 14], focal-plane image compression [15], or motion detection [16]. The incorporation of circuits and sensors has expanded the dimension of the system-on-chip (SOC) applications [17].

2.2 CMOS Technology Scaling Considerations

The renaissance of MOS imagers is mainly attributed to the down-scaling of the transistor size so that a pixel can achieve a reasonable fill factor. However, with the scaling trend continuing for CMOS processes, the device characteristics and the available sensor structures in new generations of CMOS processes may be detrimental to imaging applications. Some of the fundamental device problems in standard logic processes may eventually force CMOS imagers to take a different process route. During the period of this study (~1997 – 2001), the mainstream CMOS technology focuses on the 0.35-μm or 0.25-μm process and is gradually shifting towards 0.18-μm process. Process modifications for improving image sensors have already happened on some of the CMOS technologies [18, 19, 20, 21]. A substantial amount of information on this topic is provided in reference [22].

This section discusses the impact of the technology scaling and looks beyond for the future generation of CMOS processes. The analysis is performed on the individual imager parameters and listed below.

1. Spectral Response and Sensitivity:

For deep sub-micron CMOS processes, silicide is applied on the surface of diffusion region and poly gates to reduce sheet resistance and improve RC delay in logic circuits. However, silicide is an opaque material for visible light. More than 90% of the incoming photons are unable to transmit through this layer. Therefore, if the source/drain junction or poly gate is utilized for the photodiode or photogate sensor, a special process procedure is required to remove the silicide. In fact, the silicide removal process has been developed and utilized to fabricate on-chip resistors in some CMOS processes. If this process module is available for image sensor

fabrication, the sensitivity of the sensors will not be degraded by silicide.

As the scaling trend continues, if the diffusion/well junction is used to build a photodiode sensor, the width of the depletion region will decrease since the doping concentrations on both sides of the junction increases. Spectral response and sensitivity of this sensor will thus reduce due to the reduction of the depletion region width. One way to solve this problem is to utilize a well/substrate junction, in which the substrate doping concentration is usually low (e.g. 10^{15} cm^{-3}) and less sensitive to technology scaling. The insensitive substrate doping is due to the fact that the optimization of bulk doping for transistors is made through well implant for a twin-well process. Because of that, the depletion region width of the well/ substrate junctions will roughly keep constant and the imager sensitivity will not be degraded. However, the metallurgical junction depth will become shallower following the scaling trend. Special implants may be required to optimize the junction profile for the image sensors.

The utilization of silicon-on-insulator (SOI) wafers in the future CMOS processes poses challenges on the sensitivity of CMOS imagers. That is because the silicon layer on top of the oxide layer (~40 – 200 nm) is much thinner than the absorption depth (see Figure 3.2) of silicon at visible light. A severely reduced number of photons will be absorbed and converted into electron-hole pairs. As a result, special measures must be taken to improve the sensitivity; otherwise, the CMOS imager process has to continue utilizing epi/bulk wafers and the process development will thus depart from generic logic processes.

2. Leakage Current:

Several leakage mechanisms are expected to become serious problems for future generations of CMOS imager processes. The leakage sources include junction leakage current, subthreshold leakage current, gate current, and hot-carrier induced current. These mechanisms could produce excess current components leaking in or out of the charge sensing node and greatly impact the imager performance.

The junction leakage current is by far the most significant leakage current

component for CMOS imagers. This component is largely due to the excess minority carriers generated from the carrier generation/recombination (G/R) centers within or near the depletion region. This leakage source is dominant for both N^+-diffusion and N-well photodiodes. The density of the G/R centers is a strong function of the field oxide structure, the doping levels, and process conditions such as thermal annealing and passivation layer formation. For N-well photodiodes, as long as the carrier generation issue is addressed properly in process development, the junction leakage current is not expected to degrade severely by CMOS technology scaling. Whereas for N^+-diffusion photodiodes, due to the shallow junction depth and the resultant strong electric field at the diode edges and corners, the junction leakage current is expected to increase dramatically with the scaling trend.

The subthreshold leakage current refers to the current component flowing through the CMOS transistor channel even though the gate voltage is low and the channel is nominally “off”. This current component increases significantly with device scaling due to the non-scalability of the subthreshold slope. As the threshold voltage decreases with the scaling, this leakage current increases exponentially [23]. For the transistors used for resetting the charge-sensing node or as an electronic shutter, this mechanism produces excess current leaking in or out of the sensing node and thus corrupts the signal.

The gate current is due to carriers tunneling through thin gate oxide layers. For a CMOS active pixel sensors (see Figure 4.14), this current component also behaves like an extra leakage current component since the charge sensing node is directly connected to the gate of a source follower transistor. It has been experimentally observed that as the gate oxide thickness approaches 2.8 nm, the tunneling current could reach 10 $fA/\mu m^2$ at 1V gate bias [22]. Compared with typical leakage current of 1 fA/pixel in modern CMOS imagers, this leakage level is unacceptably high.

A high electric field near the drain of the in-pixel transistors could produce hot carriers and generate excess minority carriers in the substrate. These carriers could be absorbed by the photo-sensors and be treated as part of the optical signal. This

phenomena is observed in this study and presented in Chapter 5.

3. Voltage Scaling:

Voltage scaling is utilized in newer generations of CMOS logic processes to alleviate the effects of high electric fields in small devices. For image sensors and analog circuits, the voltage scaling has a detrimental effect to the signal swing. For the current generation of CMOS imager processes (e.g. a 0.25-μm process) with a Vdd of 3.3 V, the maximum output voltage of CMOS active pixel sensors (APS) is roughly on the order of 1V. Once the voltage is shrunk to 1.8V or lower, the signal swing will be shrunk to only hundreds of mV. It is likely that multiple Vdd and oxide thickness options will be available in the new-generation processes to maintain the signal swing for analog applications. However, the design rules for these transistors may not be scaled as much as the transistors in logic circuits so the pixel size or fill factor will not be benefitted as much from the scaling. Under such circumstances, the advantage from device scaling would merely be a smaller size of the periphery logic circuits, which utilize a lower Vdd and stringent design rules. The image sensor arrays would possibly remain a similar size and performance as the previous generation. The determining factor for adopting the newer generation of process would become "whether the reduction of the silicon area occupied by the periphery logic circuit is cost effective to switch to a more advanced process?"

4. Pixel size:

It is reported in the literature [22] that a further decrease in the pixel size below 5 μm x 5 μm is unnecessary due to the diffraction limit of the lens. However, the current trend in CCD/CMOS imager industry is that more and more research effort is devoted into the development of imagers with small pixel-size and large format. To quantitatively clarify the relation between lens resolution limit and pixel size, an optical simulation tool is utilized in this study and the results is reported in Chapter 7.

Current observation shows that special process modules are created specifically for CMOS imagers. Several recently published papers on CMOS imager processes reflect this

fact. In reference [19, 24], a 0.25-μm non-silicide source/drain process is developed for removing silicide. An arsenic and phosphorous double ion implant at the source/drain is utilized to reduce leakage current due to the stacking fault at the STI edge. A pure hydrogen annealing is also used to passivate the defect states at the Si/SiO2 surface to improve junction leakage current. In reference [18], an embedded CMOS imager module is created for a 0.35-μm process. A customized deep P-well implant is performed to produce an internal electric field in the substrate for guiding the photon-generated minority carriers towards the photodiode junction. This approach is proven to improve both the spectral response and signal cross-talk. Another shallow P^+ implant is utilized to perform surface pinning for reducing leakage current generated from surface states. Similar pinned photodiode structures are also developed in reference [20, 21]. The minimum feature sizes in the two processes are 0.35-μm and 0.6-μm respectively. All the above technologies are modified from standard CMOS processes.

Foreseeing the process development in the long run, it seems that CMOS imagers are unlikely to simply follow the scaling trend of the generic logic processes. The degraded device characteristics, the high leakage current of gate oxide, the utilization of SOI wafers, and the scaling of Vdd are all detrimental factors to imager performance. Eventually, the CMOS imager process may need to depart from the generic logic process. In the short term, new process modules, such as H_2 annealing or special photodiode implants, can be inserted into the generic process to improve imager performance. However, as the generic CMOS technologies continue to down-scale the transistor size, the gate leakage current will increase dramatically. The amount of charges stored at the sensing node will be dominated by this leakage current and the charge capacity allocated for photo-carriers at the sensing node is greatly reduced. The shot noise mechanism of the gate leakage current further increases the noise level, which degrades the image sensor performance. On the other hand, once a critical process modification, such as SOI wafers, is adopted for the new generations of processes, either a novel image sensor structure needs to be invented or the CMOS imagers will not be able to follow the scaling trend as demonstrated in the previous generations of processes.

2.3 Review of CMOS Imager Characterization

An extensive study on the noise performance of the CMOS photogate was reported in reference [25]. The imager chip was fabricated with Lucent Technologies 0.8-μm non-silicide process. The pixel schematic was similar to the one shown in Figure 4.35 except that there are no holes in the poly gates since this was a non-silicide process. The pixel size was 16 μm x 16 μm. A two-level correlated double sampling circuit was performed to cancel the fixed pattern noise due to read-out circuit mismatches. The performance of this photogate sensor was comparable to low-end CCD sensors but was inferior to that reported for the high-end CCD sensors. The conclusion of this study described that the performance of this photogate sensor was limited by low quantum efficiency (due to ploy gate absorption), high leakage current (dark current shot noise), high leakage current non-uniformity, and pixel cross-talk. Further improvement of the sensor performance would rely on the advancement on the process conditions instead of circuits or architectures.

Another study on building test structures to perform characterization and comparative analysis of CMOS image sensors was reported in reference [26]. Various kinds of single pixel test structures and 64 x 64 pixel arrays were fabricated with two separate 0.35-μm processes. Some of the characterization results and mathematical analysis on quantum efficiency and fixed pattern noise were reported in [27, 28].

2.4 Review of Hot-carrier Induced Minority Carriers

The hot-carrier mechanism and the induced excess minority carriers have been studied extensively in the '80's and early '90's. Several mechanisms were proposed to explain the generation of the minority carriers in the substrate [29, 30, 31]. However, the influence of the hot carriers induced minority carriers has not been reported in the literature. The impact of the excess minority carriers on imager performance is investigated in this study. The spatial distribution of the excess minority carriers is also characterized to be within several tens of micron meters. A detailed description on this topic is presented in Chapter 5.

2.5 Review of Cross-talk on Color Imager Performance

Pixel cross-talk was characterized and studied in references [18, 19, 25, 32]. Most of the

above characterization reported the cross-talk values at three different color bands – red, green, and blue. A further analysis between cross-talk and the performance of color imagers was given in [25]. The cross-talk was observed by the higher spectral response of the green and blue pixels at the wavelengths corresponding to the peak red-pixel response. It was also noticed that the green pixels in the same row as the red pixels suffered a higher cross-talk than the ones on the same row as blue pixels. The cross-talk distorted the color analysis functions after color correction. The distortion is more severe for blue response due to the poor blue-light quantum efficiency of a photogate sensor. Thus, a poorer color fidelity was resulted. The signal-to-noise ratio of the color signal was also degraded due to cross-talk, which is the same conclusion achieved in this thesis.

Chapter 3

Characterization of CMOS Processes: Stage I – Large Area Photodiodes

This chapter is dedicated to the topic of characterizing large-area photodiodes for image sensor applications. Test structures are implemented in standard CMOS processes for comparing the performance of photodiodes with different junction types and layout strategies. The measured junction properties include spectral response, leakage current, and junction capacitance. To improve the transmission coefficient of photons, different stacking arrangement of dielectric layers on top of silicon is studied. The characterization results provide preliminary information for selecting sensor structures utilized in the second-stage characterization, in which small-dimension pixel arrays are implemented.

3.1 Design of Large Area Photodiodes

A simple photodiode structure allows to directly measure its junction properties. The design and characterization is much easier than that for a pixel arrays. The size of the photodiodes should be determined by the magnitude of the measured parameter and the resolution of the equipment. Leakage current, among all the photodiode parameters, is the critical parameter that dictates the size of the diodes. For example, if the dimension of a photodiode is designed to be as small as 10μm x 10μm, which is roughly the pixel size for imagers, the total leakage current is typically on the order of 1fA(10^{-15}A) or lower. The small magnitude of current is beyond the resolution for most of the instruments. Besides the problem in leakage current measurement, a small diode size also poses challenges on correctly estimating the optical power falling on the diode during spectral response measurement. The simple and effective way to avoid those measurement-related problems and improve accuracy is to build large-area photodiodes.

With the available current meter providing DC current resolution down to sub-pA range, the photodiodes are determined to be as large as 500μm x 500μm to assure sufficient accuracy for leakage current measurement. Each on-chip photodiode is surrounded

with a 100μm wide N-well and a 100μm wide P^+-diffusion guardrings. The N-well guardring is connected to Vdd and the the P^+-diffusion guardring is connected to ground. The N-well guardring serves to capture the stray photo-carriers which are generated outside the active photodiode region and diffuse towards the diode. The P^+-diffusion provides a low-impedance connection from the substrate around the diode to ground.

The whole chip is fully covered with the top layer of metal except for the photodiode region and the bonding pads. This metal shield is necessary since it precisely defines the optically active region of the photodiodes.

3.2 Basic Photodiode Parameters

The basic parameters of the photodiodes include spectral response, leakage current, and capacitance. The properties and measurement setup for these parameters are described in this section.

3.2.1 Spectral Response

Spectral response of a photo-sensor determines the absorption efficiency at different optical wavelengths. It is important for both color imagers and black-and-white (B/W) imagers. For color imagers the spectral response determines the output signals corresponding to different color pixels; whereas for B/W imagers the sensitivity is an integration of the response curve over the visible spectrum (~380nm to 780nm in wavelength). The spectral response is often described as quantum efficiency (Q.E.), which is defined as the number of absorbed photo-carriers divided by the number of the injected photons of a specific wavelength. The full scale of Q.E. is 100%.

Physics

Q.E. is mainly determined by three factors — the transmission coefficient of the dielectric layers above silicon, the absorption coefficient of Si, and the carrier collection efficiency of the sensor. These three factors are all functions of wavelengths and correspond to following physical mechanisms. First, the impinging photons need to penetrate through the dielectric layers. If the photons are reflected or absorbed in those layer, they will not contribute to the optical signal. Second, the penetrated photons need to be absorbed by Si and generate electron-hole pairs (*EHP*). Third, the generated *EHP*s must be separated by the

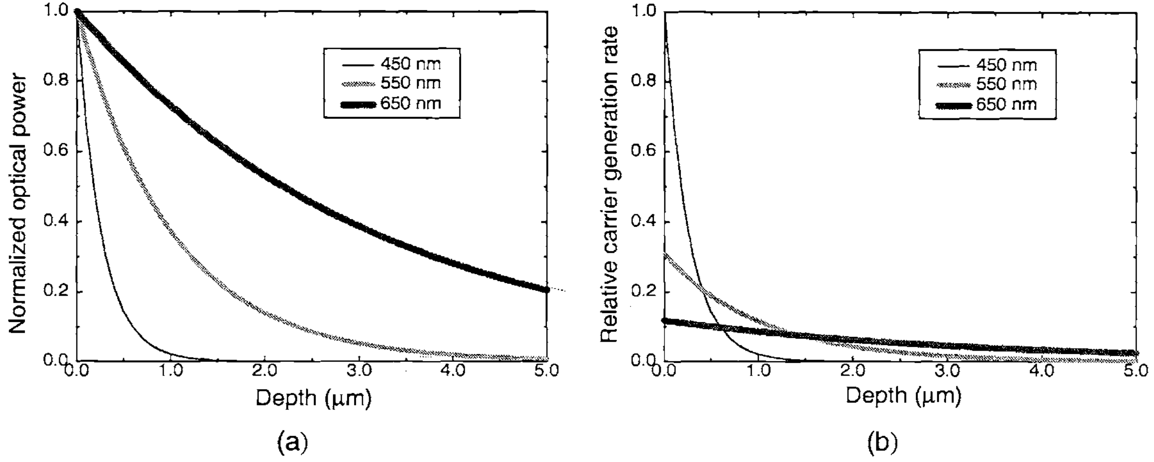


Figure 3.1: Normalized optical power (a), and the relative carrier generation rate (b) in silicon

junction before recombination occurs. These factors is described below.

The transmission coefficient of the dielectric layers depends on the refractive index of the layers used in the CMOS processes. Improvement on the transmission coefficient will be described in Section 3.3.

As the photons penetrate through the dielectric layers and impinge on Si, the optical power decays exponentially with the Si depth due to photon absorption[1]. The normalized optical power versus Si depth is shown in Figure 3.1(a). The three curves correspond to the normalized optical power at three different wavelengths. The slope of the curves represents the Si absorption rate. Since each absorbed photon generates one *EHP*, the absorption rates also represent the *EHP* generation rates, which are presented in Figure 3.1(b). Theoretically, the *EHP* generation rate can be expressed as

$$G = \frac{P'\lambda}{hc}\alpha e^{-\alpha z}, \tag{3.1}$$

where P' is the optical power impinging on Si[2]; h is Planck's constant; c is light speed; λ is optical wavelength; α is the absorption coefficient; and z is Si depth. The absorption coefficient α is a wavelength-dependent parameter. The empirical data [33] of α is presented in Figure 3.2.

1. The exponential behavior is common for physical mechanisms that can be formulated as $-dP/dx = K$, where K is a positive constant. Here, K is associated with the optical absorption coefficient of Si.
2. This is the transmitted optical power instead of the total optical power.

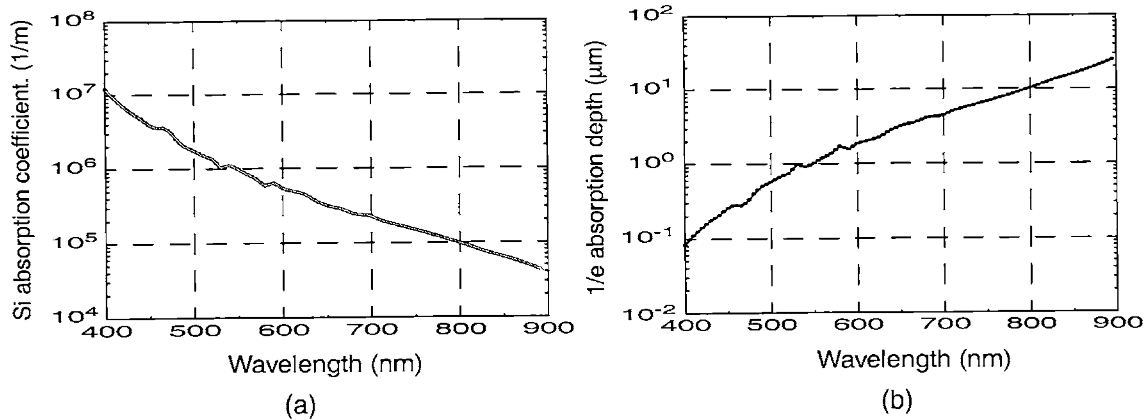


Figure 3.2: Silicon absorption coefficient (a), and the corresponding 1/e absorption depth (b) versus wavelength

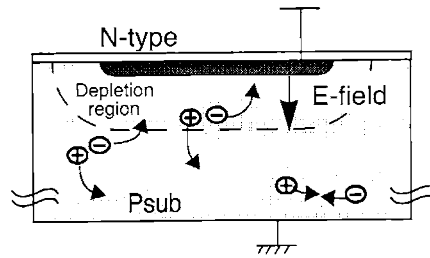


Figure 3.3: Photo-carrier collection mechanism

According to Figure 3.1 and Figure 3.2, blue light, which corresponds to a short wavelength, is absorbed at a shallower region while red light can penetrate much deeper into silicon. This *EHP* generation profile and its wavelength dependence is solely dictated by the sensor material — Si.

The photo-carrier collection mechanism is shown in Figure 3.3. The carriers are driven by the internal electric field at the location they are generated. Therefore, if an *EHP* is generated inside the depletion region, the built-in electric field will separate them and 100% of the carriers will contribute to photo-signal. If the carriers are generated in a region where the local electric field is weak, they can either diffuse to a near-by depletion region and get separated or eventually recombine without contributing to the photo-signal. Considering all the above conditions and the exponential *EHP* generation profile, the carrier collection is a complicated mechanism which involves the internal electric field, the

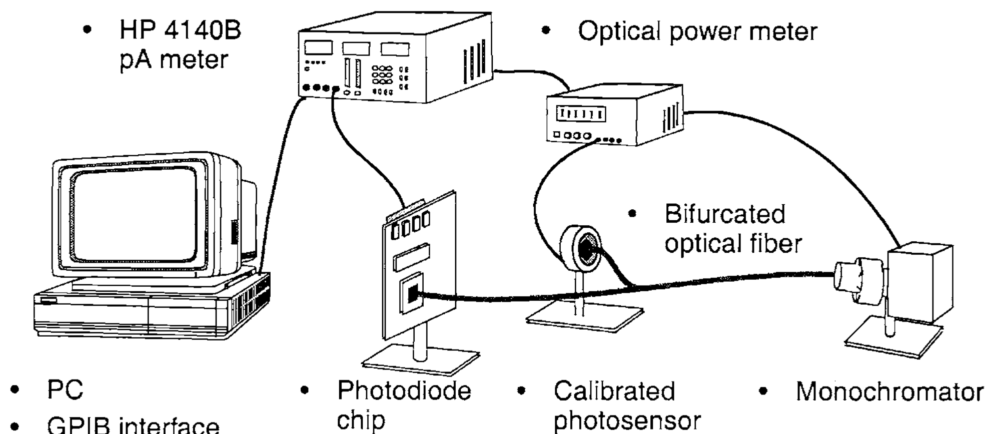


Figure 3.4: Test setup for spectral response characterization

location and the width of the depletion region, and the minority carrier diffusion length. These parameters are mainly determined by the doping profile of the photodiode and are strongly related with the process parameters. One of the objectives in photo-sensor design is to construct a doping profile with an optimized carrier collection efficiency as well as a minimized carrier recombination rate. This issue will be discussed as the Q.E. of the photodiodes with different junction types is compared.

Experimental Setup

The experimental setup is shown in Figure 3.4. The monochromator (model: TRIAX-180 by Instrument, SA) with a 120W tungsten halogen light source provides monochromatic light with variable wavelengths. The optical output is connected to a bifurcated optical fiber with one end shining on a calibrated photo-sensor (model: 818-SL from Newport Corp.) and the other end on the photodiode under test. The monochromator, the optical power meter (model: 1835-C multi-function optical meter from Newport Corp.), and the pico-Amps current meter (model: HP4140B) are connected to a PC through a GBIP interfaces for data acquisition and storage. All the measurements for the photodiodes are conducted in a metal shield box to isolate optical and electrical noise. Each data point represents an average of 20 consecutive measurements to remove random noise.

According to the definition, Q.E. can be calculated by the following equation:

$$\text{Q.E.} = \frac{(h\nu) \cdot I}{P \cdot q}, \tag{3.2}$$

where P denotes total optical power shining on the photodiode; q is electron charge; h is Planck's constant, ν is optical frequency and I is the measured photo-current.

3.2.2 Leakage Current

Leakage current in photodiodes refers to the absorption of carriers that are not generated by photons. Often times it is also called dark current. Since the properties of the leakage-current carriers are intrinsically the same as the photo-generated carriers, there is no physical way to distinguish these two carriers and subtract one component from the mixture. Leakage current reduces the charge capacity dedicated for photo-carriers. More importantly, this current component varies with sensor location and time, producing spatial and temporal variation of the output signal. The spatially varying part contributes to fixed pattern noise of the imager and the temporal part is one of the sources for random noise (this noise component is also called dark current shot noise). The impact of leakage current is more severe under low illuminance levels since its percentage in the total amount of carriers becomes higher. Therefore, it is crucial to minimize leakage current to improve the noise performance under low illumination conditions.

Physics

The mechanism of leakage current for a photodiode is essentially the same as that of the reverse-bias current for a p-n junction diode. From semiconductor physics, the reverse-bias current density of a n$^+$-p abrupt junction diode consists of two terms [34] and can be written as

$$J_R = q\sqrt{\frac{D_n}{\tau_n}}\frac{n_i^2}{N_a} + \frac{qn_iW}{\tau_e}, \tag{3.3}$$

where D_n and τ_n represent the diffusion coefficient and the minority carrier lifetime of electrons in the p-type region; n_i is the intrinsic carrier concentration; N_a is the doping concentration in the p-type region; W is the width of depletion region; and τ_e denotes the effective lifetime in the depletion region. The first item in the equation comes from minority carrier diffusion from the charge neutral region; the second item is the generation cur-

rent inside the depletion region. Both items have strong temperature dependence through n_i and the second item has a bias voltage dependence through W. For Si junction diodes at room temperature (~ 25 °C), the reverse-bias current is usually dominated by the component of generation current. On the contrary, the diffusion current dominates at high temperature because the n_i^2 term increases faster with temperature[1].

For imagers of commercial applications, the operating temperature is around room temperature; therefore, minimizing the generation current becomes a high-priority task. According to eq.(3.3), only W and τ_e are variables for the generation current. Since the depletion region width W is also related to the photo-carrier absorption and is proportional to Q.E. to the first order, it is not desired to reduce W. Whereas for τ_e, it is not related with any other image sensor parameter and deserve the effort to maximize it.

The effective lifetime τ_e, according to Shockley-Hall-Read generation/recombination (G/R) model [35], is directly related to the density of the G/R centers within silicon bandgap, especially for the ones with energy states near the center of the bandgap. Those energy states are created largely due to lattice defects. The defect density is much higher near the silicon surface since the transition of material from Si to SiO_2 creates dangling bonds and generates a huge amount of states within the bandgap. Therefore, process development effort is required to minimize this component of leakage current originated at the Si surface.

Experimental Setup

The problem associated with the small magnitude of leakage current density is solved by utilizing large-area photodiodes. However, direct measurement of leakage current with packaged chips produces error. That is because the resistivity of the package material, either ceramic or plastic, does not necessarily provide sufficient electrical isolation. Thus, the measurement can only be performed through a probe station with the HP4140B pico-Amps current meter. The meter also provides necessary bias voltage for the photodiodes. A thermal chuck is utilized to measure the temperature dependence of leakage current.

1. $n_i^2 \propto \exp\left(\frac{-E_g}{k_B T}\right)$

3.2.3 Capacitance

Junction capacitance at the charge sensing node determines the charge capacity and the charge-to-voltage conversion gain of the image sensors. A smaller capacitance at the charge sensing node is usually favored since it provides a higher signal-to-noise ratio of the output signal. However, the drawback is a lower charge capacity and thus a lower dynamic range. The trade-off on the sensing node capacitance will be re-visited in Section 4.5.2. In this chapter, the objective is to understand the capacitance density associated with different junction types.

Physics

The photodiode capacitance is essentially a p-n junction capacitance. Therefore, the width of the depletion region and the capacitance value are mainly determined by the doping profile. Using abrupt junction approximation, the width of depletion region can be calculated as

$$x_d = \left[\frac{2\varepsilon_s}{q}\left(\frac{1}{N_a}+\frac{1}{N_d}\right)(\phi_i - V_a)\right]^{1/2}, \tag{3.4}$$

where ε_s is permittivity of silicon; N_a and N_d are doping concentrations of p-type and n-type region respectively; ϕ_i is the built-in potential, and V_a is the applied bias voltage. The capacitance per unit area can be calculated as $C = \frac{\varepsilon_s}{x_d}$. Generally, the higher the doping levels, the higher the capacitance per unit area.

Experimental Setup

Similar to the leakage current measurement, the capacitance is directly measured with a probe station for better accuracy. An HP4275 LCR meter is utilized for capacitance measurement. Capacitance of the dummy bonding pad is also measured and subtracted from the measured data.

3.3 Improving Transmission Coefficient of Dielectric Layers

There are multiple dielectric layers on top of Si wafers. The functions of these layers are to passivate the oxide-silicon interface, to electrically isolate metal layers, to isolate humidity and impurities from active devices, and to provide scratch protection. The materials

commonly used in the layers include oxide, nitride and oxy-nitride. These materials typically exhibit nearly 100% transmittance at visible light but their difference on refractive index causes reflection at the layer interface. The reflection reduces the number of photons impinging on Si and degrades the Q.E. of the image sensor.

In this section, the reflection phenomena of the dielectric layer is studied and the objective is to explore an optimized stacking configuration to improve photon transmission from air into Si.

3.3.1 Theory

The behavior of electromagnetic waves is governed by Maxwell's equations. For a single dielectric interface, the reflection coefficient (assuming a 90° incident angle) can be calculated [36] as

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2, \tag{3.5}$$

where n_1 and n_2 represent the refractive index of media on both sides of the interface. The transmission coefficient is simply $1 - R$.

According to eq.(3.5), the larger the difference of refractive index between n_1 and n_2 is, the higher the optical energy is reflected. In imager operations, the photons need to transmit through the air (n ~ 1), pass through the multiple dielectric layers and then reach silicon (n ~ 3.44). If the dielectric layers are stacked properly so that the refractive index increases monotonically along the light path, mathematically the overall reflection can be minimized. Thus, the best way to stack the dielectric materials from top to bottom would be oxide (n ~ 1.46), oxy-nitride (n ~ 1.7), and then nitride (n ~ 2.0) and finally silicon (n ~ 3.44).

In reality, this approach is impractical because there are other process concerns. For example, oxide is the best material for passivating Si interface and it needs to be in direct contact with Si. Furthermore, nitride is the best material for humidity isolation and needs to be applied in the stacking structure. Therefore, the alternative is to apply buffer layers between the nitride layer and oxide/Si interface with a proper arrangement of the refractive index to reduce the reflection caused by large refractive index jumps.

3.3.2 Modeling

For simplicity, interference phenomena and multiple reflection between the dielectric layers are not considered in the calculation. The assumptions can be justified by the following facts. First, the interference phenomena produces constructive and destructive response on the transmission curves, making it difficult to compare the average performance between different dielectric configurations. Second, the transmission components due to multiple reflection is usually much smaller than the directly injected component. Therefore, the total transmission coefficient of the multiple-layer structure can be calculated as

$$T_{total} = T_{12} \cdot T_{23} \ldots T_{(k-1)k} = \left[1 - \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2\right] \cdot \left[1 - \left(\frac{n_2 - n_3}{n_2 + n_3}\right)^2\right] \ldots \left[1 - \left(\frac{n_{k-1} - n_k}{n_{k-1} + n_k}\right)^2\right]. \quad (3.6)$$

Utilizing eq.(3.6), transmission coefficients are calculated with respect to the stacking configurations shown in Table 3.1. The ideal configuration described previously is also shown at the last row for comparison.

The Ox/Si configuration is considered as a basic structure and its transmission coefficient is used as the normalization factor in the third column of the table. It is observed that only 10% difference of transmission coefficient exists between the best case (Ox/Si) and the worst case (SiN/Ox/Si). That is because a significant amount of optical reflection happens at the final Ox/Si interface due to the large refractive index jump. For the practical configurations, the transmission coefficient of the standard case (SiN/Ox/Si) can be

Dielectric stacking configuration	Transmission coefficient	Normalized transmission coefficient
Air/Ox/Si	0.808	1.00
Air/SiN/Ox/Si	0.726	0.899
Air/Ox/SiN/Ox/Si	0.769	0.952
Air/Ox/SiN/SiON/Ox/Si	0.778	0.963
Air/Ox/SiON/SiN/SiON/Ox/Si	0.788	0.975
Air/Ox/SiON/SiN/Si (* impractical*)	0.890	1.10

Table 3.1: Transmission coefficient of dielectric layer configurations

improved from 0.726 to 0.788 (the Ox/SiON/SiN/SiON/Ox/Si case), a 8.5% improvement with the drawback of process complexity. Another observation from the table is that the Q.E. of a photodiode fabricated with a typical dielectric-layer structure (SiN/Ox/Si) will not exceed 73% due to the reflection phenomena.

3.3.3 Experimental Results

The transmission coefficients of the dielectric layers are difficult to measure directly by electrical methods. That is because the measured photo-current is convolved with the charge collection efficiency of the photodiode junctions. Nevertheless, the ratio of the transmission coefficient between two dielectric layer configurations can be calculated by taking the ratio of the corresponding photo-current. Since the photodiodes are of the same structure (NW/Psub photodiode), the only factor that produces photo-current variation is the transmission coefficient difference of the dielectric layers.

The photo-current measurements are performed on photodiodes with Ox/Si and SiN/ Ox/Si configurations. The utilized technology is a 0.25-μm CMOS process. The ratio between the photo-current is shown in Figure 3.5. The DC levels of the measured and sim-

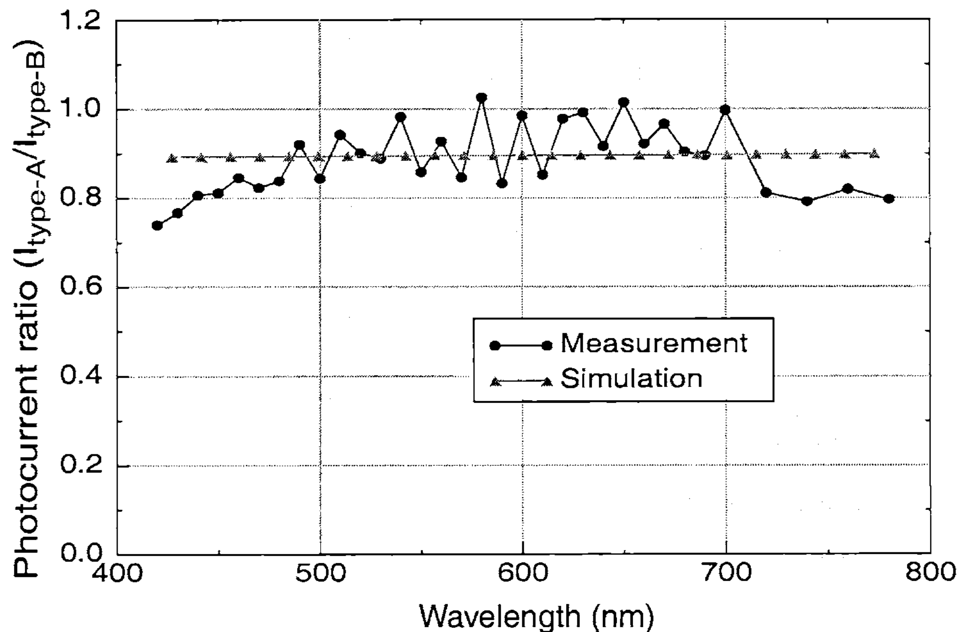


Figure 3.5: Transmission coefficient ratio between SiN/Ox/Si (type-A) and Ox/Si (type-B) dielectric layers

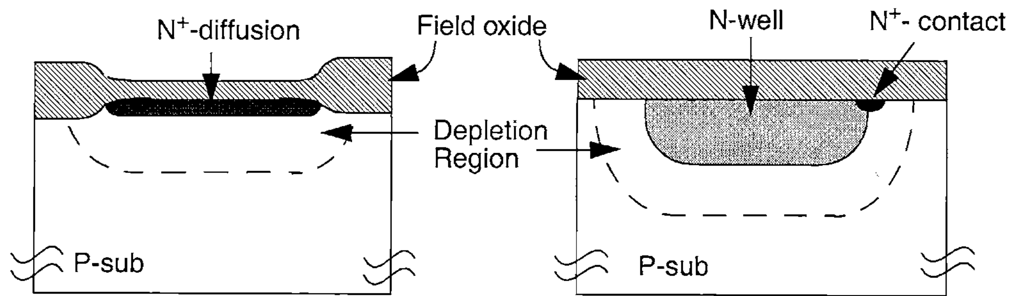


Figure 3.6: Cross-sections of N$^+$/Psub and NW/Psub photodiodes

ulated data match well except for the wavy behavior on the measured curve due to minor optical interference effects. This result validates the first-order calculation of the transmission coefficient as described in eq.(3.6).

3.4 Comparison of Junction Types

Standard CMOS processes offer two junction types that can be used as photodiodes — NW/Psub and N$^+$/Psub (or N$^+$/PW/Psub) photodiodes. Cross-sections of the photodiodes are shown in Figure 3.6. The NW/Psub photodiode requires a N$^+$-diffusion/N-well butting junction to electrically connect the N-well region, which violates the design rules for typical CMOS circuits and is thus not as well-accepted. However, this diode structure is valid from a process perspective. This section compares the Q.E., leakage current, and junction capacitance of both structures fabricated with standard CMOS processes.

3.4.1 Quantum Efficiency

Figure 3.7 shows Q.E. measurement on NW/Psub and N$^+$/Psub photodiodes with two different CMOS processes — a 2.0-μm and a 0.5-μm processes. The former is a non-silicided process and the latter applies a silicide block mask on optically active regions. The applied reverse bias voltage is 2.0 V. Comparisons are made mainly between NW/Psub and N$^+$/Psub photodiodes from the same process. Inter-process comparison is not as reasonable since there are too many process-dependent variables, such as the process conditions, involved in the fabrication. According to the figure, both diode structures

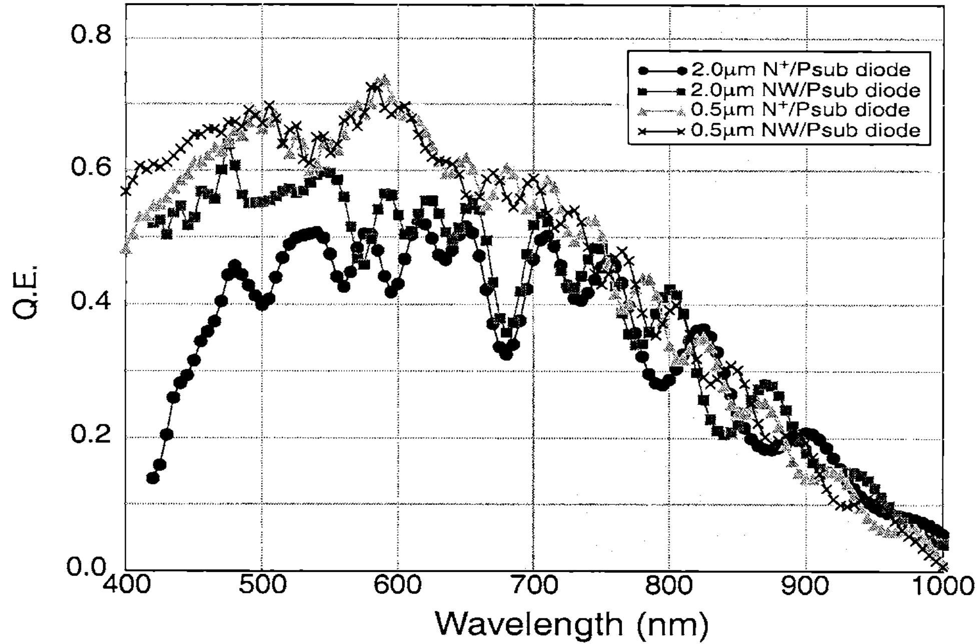


Figure 3.7: Q.E. of photodiodes with the 0.5-μm and 2.0-μm processes

demonstrate comparable Q.E. for wavelengths longer than 600nm. This can be explained by the fact that long-wavelength photons have lower Si absorption coefficient (refer to Figure 3.2) and most of the photo-carriers are generated deeper in the substrate. Since the carrier lifetime and the internal E-field are determined by the same substrate doping profile at such depth, both structures exhibit similar charge collection efficiency and Q.E.

At short wavelengths, however, the NW/Psub diodes demonstrate better Q.E. This outcome is attributed to a much longer minority carrier diffusion length in the N-well than the N^+-diffusion region. Since the short-wavelength photons tend to generate *EHP*s in a shallow region, most of the photo-carriers are generated within the N-well or N^+-diffusion regions. The short minority-carrier diffusion length in the N^+-diffusion recombines the carriers more efficiently and thus results in a lower Q.E. According to empirical data [42], the minority-carrier diffusion length of holes (L_h) is roughly 100μm if $N_d = 10^{17} cm^{-3}$, a typical doping concentration for N-well. Whereas, for N^+-diffusion region with N_d on the order of $10^{20} cm^{-3}$, L_h is on the order of sub-micron. The huge difference in the diffusion

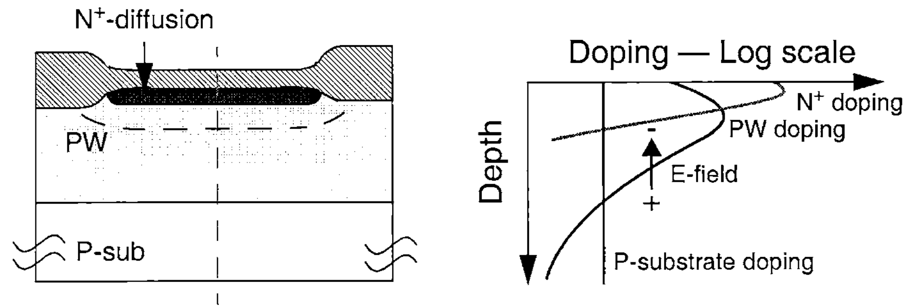


Figure 3.8: N^+/PW/Psub photodiode structure and the doping profile

length recombines more photon-generated holes in the N^+-diffusion region than in the N-well region.

The wavy Q.E. curves at longer wavelengths is due to the effect of optical constructive and destructive interference in the dielectric layers. The phenomena is more evident for long-wavelength photons due to better phase matching.

For some CMOS technologies, a P-well implant is performed to optimize the performance of nMOS transistors. This allows one to create the N^+/PW/Psub junction. The cross-section and a typical doping profile of the diode structure is shown in Figure 3.8. The existence of the P-well doping profile degrades the collection efficiency of the junction in two ways. First, the higher doping level of P-well over P-substrate reduces the width of the depletion region. As a result, the region with the highest charge-collection efficiency shrinks. Second, the doping profile of P-well produces an internal E-field in the charge neutral region in which the field repels photo-generated electrons from diffusing back into the depletion region. Both mechanisms degrade the charge collection efficiency and the Q.E. Further, the repelled carriers by the internal E-field could also contribute to signal cross-talk between adjacent pixels, which is the topic of Chapter 6. A comparison between the Q.E. of N^+/PW/Psub and NW/Psub photodiodes is shown in Figure 3.9. The diodes are fabricated with a standard 0.6-μm CMOS process. The Q.E. degradation for N^+/PW/Psub photodiode is significant.

3.4.2 Leakage Current

Leakage current, as described previously, is a characteristic closely related with process

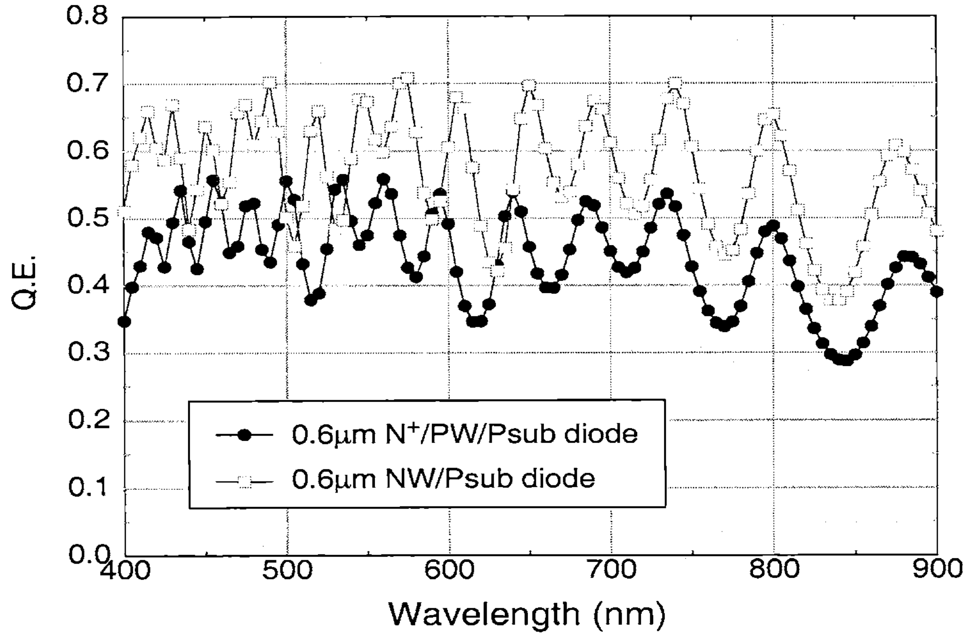


Figure 3.9: Q.E. of photodiodes with the 0.6-μm twin-well process

parameters. It is desired to achieve a photodiode with the lowest leakage current. This section presents the measurement of leakage current versus bias voltage and temperature.

The measurement of leakage current density versus reverse bias voltage for the NW/Psub and N^+/Psub photodiodes with the 0.5- and 2.0-μm processes is shown in Figure 3.10. For the 2.0-μm process, it is found that the leakage current density of the N^+/Psub photodiode is 10 times higher than that of the NW/Psub diode on the same chip. Whereas for the 0.5-μm process, the discrepancy is not as much but the N^+/Psub photodiode still exhibits 40% higher leakage current than the NW/Psub diode at 2.5V reverse bias.

The lower leakage current of NW/Psub than N^+/Psub photodiodes can be explained from a process perspective. First, the N-well implant is performed at an earlier stage than the N^+-diffusion implant in the process. The N-well implant usually requires a high-temperature drive-in step. Therefore, the thermal annealing is better performed than the N^+-diffusion. Second, the LOCOS process creates a highly stressed region at the periphery of the diode. In this region this is a large amount of carrier G/R centers which contribute to the leakage current. This effect becomes more significant for a diode with a higher periph-

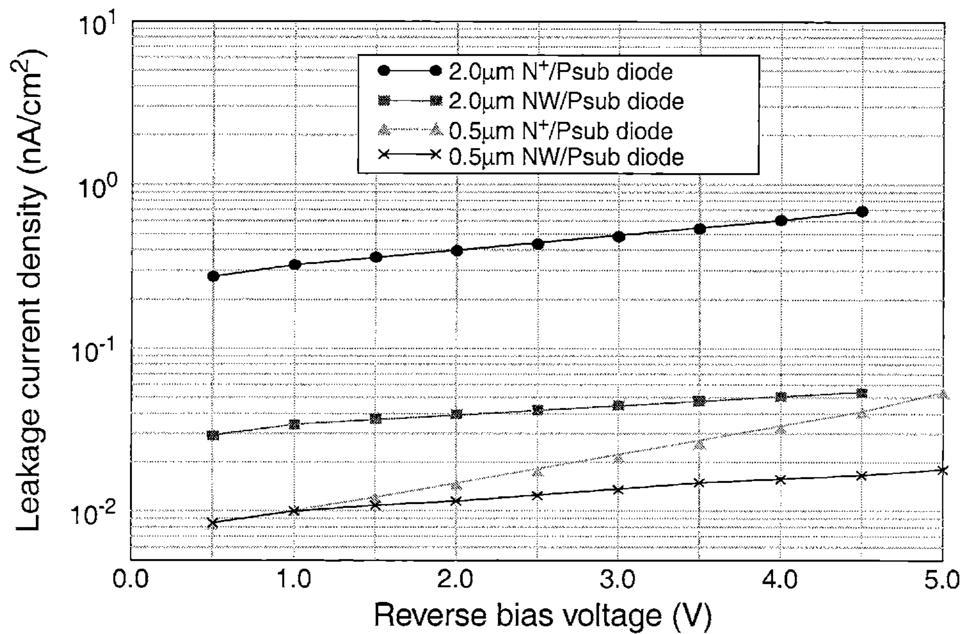


Figure 3.10: Leakage current density of photodiodes with the 0.5-µm and 2.0-µm processes

ery/area ratio as will be presented in Section 3.5. Third, the doping level of N^+-diffusion is much higher than the N-well doping level. This potentially creates more impurity states within silicon bandgap in the depletion region. The higher bias voltage dependence of the leakage current for the N^+/Psub photodiode in Figure 3.10 suggests that the generation current component in the depletion region is higher than that of the N-well photodiode.

The comparison of leakage current density between a NW/Psub photodiode and a N^+/PW/Psub photodiode in the 0.6-µm twin-well process shows a slightly different story. The measurement result is presented in Figure 3.11. The N^+/PW/Psub photodiode shows lower leakage current at low reverse bias. This can be explained with eq.(3.3). The diffusion current component of leakage current for the N^+/PW/Psub photodiode is lower than that of the NW/Psub photodiode due to the lower electron concentration in the P-well region than in the P-substrate region. On the other hand, the width of the depletion region is narrower for the N^+/PW/Psub photodiode so the total amount of G/R centers may become lower than the amount for the NW/Psub diode. Under such conditions, the N^+/PW/Psub photodiode presents lower leakage current than NW/Psub diode. As the applied bias exceeds

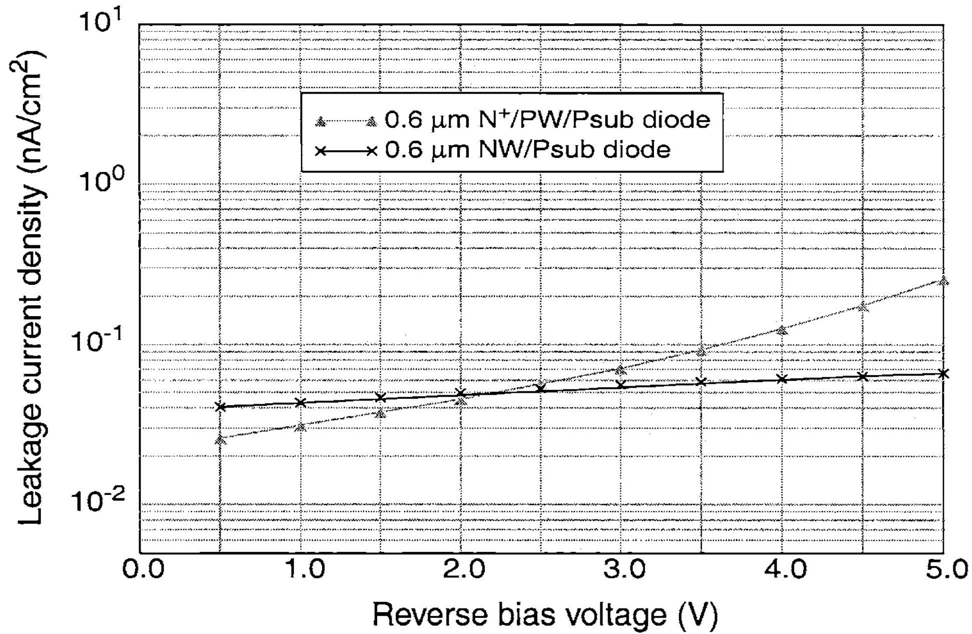


Figure 3.11: Leakage current density of photodiodes with the 0.6-μm twin-well process

2.5V, the generation current component for the N^+/PW/Psub photodiode increases faster and therefore the leakage current density exceeds that of the N-well photodiode.

Leakage current measurement is also performed under different temperatures. The temperature dependence of leakage current is important for two reasons. First, the chip temperature usually depends on the power consumption of the on-chip circuitry. The highest tolerable leakage current for an imager chip poses an upper limit on the power dissipation. Second, the temperature-varying experiments are helpful for distinguishing the dominant leakage mechanism since the two items on the right-hand side of eq.(3.3) have different temperature dependence through n_i^2 and n_i.

Figure 3.12 shows the Arrhenius plot of the leakage current density for the NW/Psub and N^+/Psub photodiodes at three different bias voltages — 1V, 3V, and 5V. The temperature range is from 25°C up to 100°C. It can be deduced from this figure that the diffusion current component of leakage current dominates at high temperature since all the curves merge together with a slope roughly equal to E_g. A slope of E_g on the Arrhenius plot illustrates that the leakage current has the same temperature dependence as n_i^2, same as the diffusion current component. Once the diffusion current component becomes dominant,

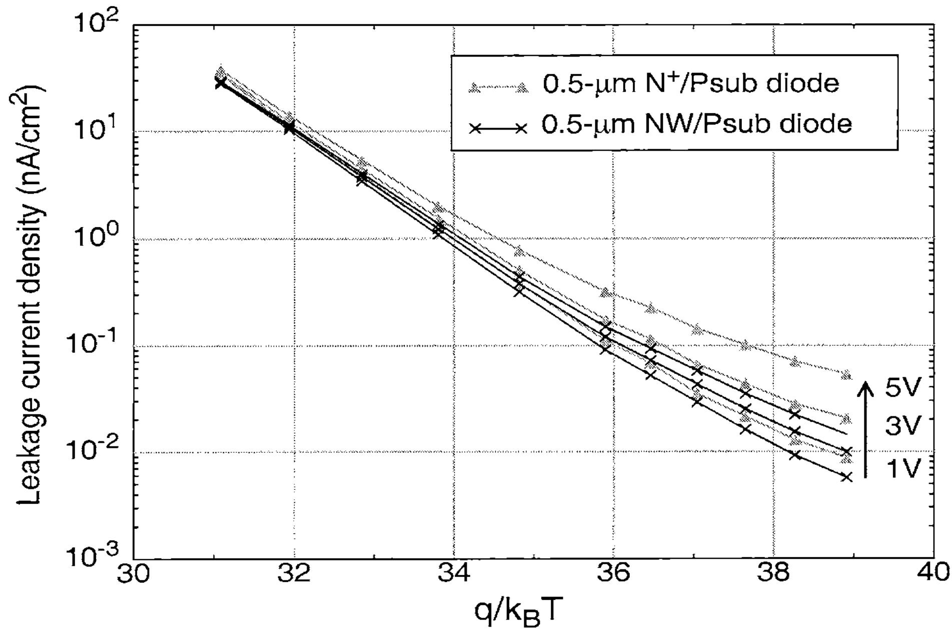


Figure 3.12: Arrhenius plot of leakage current density for photodiodes with the 0.5–μm process

theoretically both photodiodes should demonstrate similar leakage current performance since this diffusion current component only depends on the P-substrate doping, which is the same for both photodiodes. This is confirmed by Figure 3.12. Whereas at low temperature, the slope of the curves is close to $E_g/2$, showing the same temperature dependence as n_i. Besides, the curves demonstrate stronger bias-voltage dependence. The phenomena shows that the generation current component dominates in this temperature region. Therefore, the curves demonstrate strong voltage dependence through the depletion region width W. In this temperature region, the N^+/Psub photodiode, which potentially carries more G/R centers in the depletion region, exhibits stronger bias dependence than the NW/ Psub photodiodes. From the results of the temperature measurements, it can be concluded that the diffusion current component dominates at high temperature and the generation current component dominates at low temperature for the large-area photodiodes.

3.4.3 Capacitance

Figure 3.13 shows the measured capacitance per unit area for the photodiodes with the

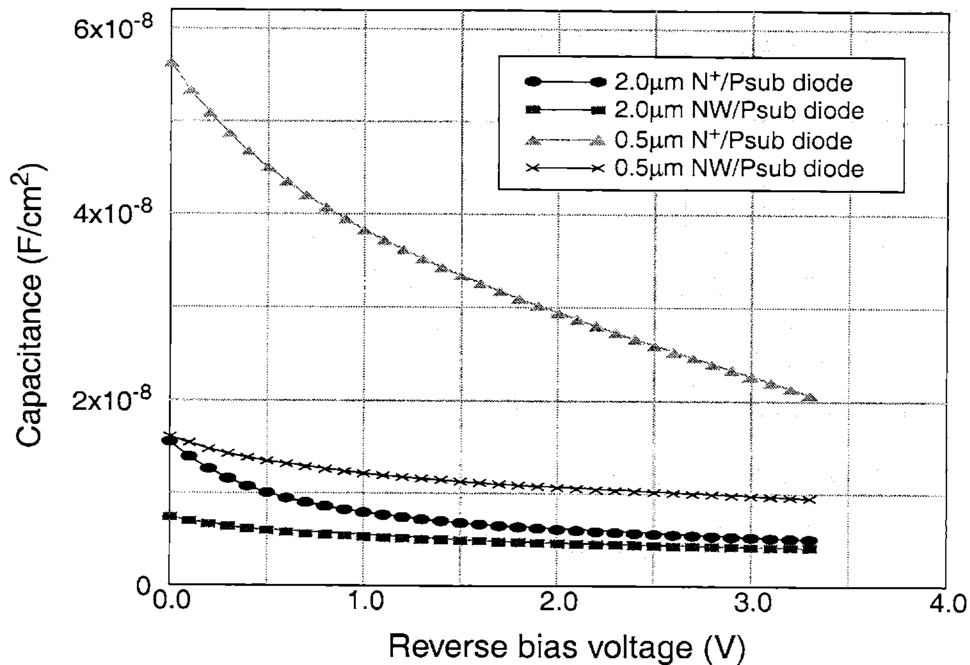


Figure 3.13: Capacitance of the photodiodes with the 0.5-μm and 2.0-μm processes

0.5-μm and 2.0-μm processes. As expected, the N-well photodiodes on both processes have lower capacitance than the N-diffusion photodiodes due to the lower doping concentration. Therefore, the N-well photodiodes should have a higher charge-to-voltage conversion gain in the pixels under the assumption that the capacitance of the charge-sensing node is dominated by junction capacitance.

3.5 Geometry Considerations

This section discusses the geometric effects of image sensors on Q.E. and leakage current. Two N^+/Psub photodiodes with different layout strategies are compared. One photodiode is designed with a typical square-shape layout and the other has a grid-shape layout. The layouts are shown in Figure 3.14. The latter structure has a higher periphery/area ratio than the former. The original idea to design such a grid-shape diode is for Q.E. improvement, which is described in the following paragraph.

According to the previous discussion on Q.E., a depletion region is a high carrier-collection-efficiency zone. If the depletion region can somehow be expanded in a photodiode and placed close to Si surface, Q.E. should be improved. Both targets can be achieved by

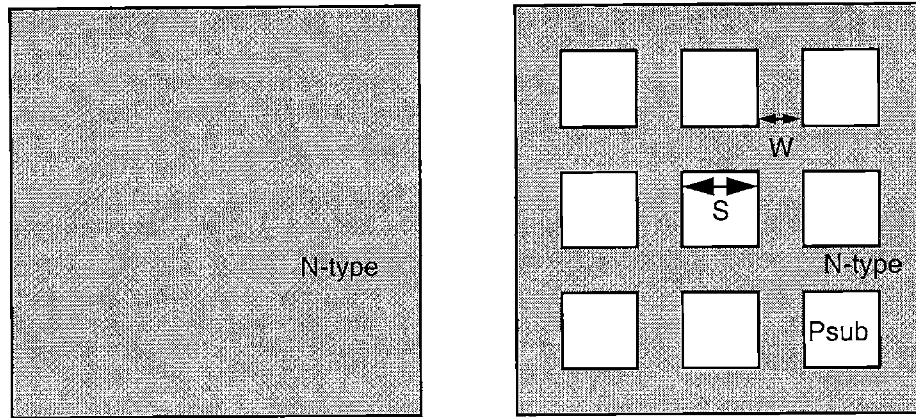


Figure 3.14: Top view of the square-shape (left) and grid-shape (right) photodiodes

laying out photodiodes in a grid-shape structure as opposed to a regular square or rectangular shape. This design is more appropriate for N^+-diffusion photodiodes than the N-well diodes because the long lateral diffusion length of N-well dopants caused by the high-temperature drive-in step makes it difficult to implement such N-well photodiodes within a 10µm x 10µm pixel.

As the grid-shape photodiode shown in Figure 3.14, the width of the N-type region is W and the separation is S. If W and S are designed properly, depletion region in the lateral direction will be wide enough to cover the exposed P-substrate region, providing a higher charge collection efficiency near the surface. Thus, Q.E. can be improved.

This photodiode is designed and fabricated with the 0.5-µm CMOS process described before. The width of the N^+-diffusion (W) is 0.9µm and the separation between the grids (S) is 2.7µm. On average, the P-substrate region occupies 56% of the total diode area. Assuming the doping concentration of the P-type substrate is $10^{15}cm^{-3}$ and the N^+-diffusion doping is $10^{20}cm^{-3}$, the exposed P-substrate region is expected to be entirely covered by depletion region under 2.0V reverse bias voltage.

The Q.E. of the grid-shape photodiode is characterized and the ratio to the square-shape diode is shown in Figure 3.15. The Q.E. is improved at wavelengths between 400–500nm. The improvement is higher towards shorter wavelengths. For wavelengths longer than 500nm, the Q.E. ratio, on average, is nearly unity. Similar to the Q.E. measurement,

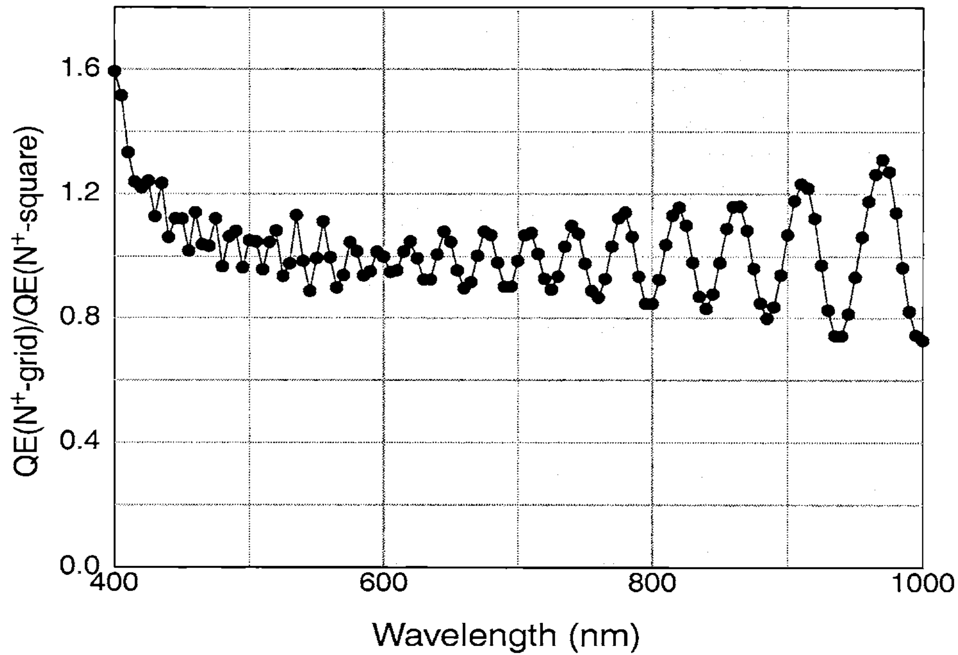


Figure 3.15: Q.E. ratio between the grid-shape and square-shape N^+/Psub photodiodes

the wavy phenomena of the curve is due to the optically constructive and destructive interference in the dielectric layers. Since the two photodiodes have slightly different dielectric thickness, the Q.E. peaks and valleys are at different wavelengths for the two diodes.

The Q.E. improvement at short wavelengths is anticipated. Since the short-wavelength photons tend to generate *EHP*s near Si surface, more carriers would be collected if the depletion region is placed closer to the surface. On the other hand, the Q.E. discrepancy of the two diodes at short wavelengths is a proof that a substantial amount of photo-carriers are recombined in the N^+-diffusion region.

An estimation on the Q.E. ratio is performed to verify the data. According to Figure 3.7, the Q.E. of the square N^+-diffusion diode at 400nm is 48%. For the grid-shape diode, the area percentages of N^+-diffusion and P-substrate are 44% and 56% respectively. Since the P-substrate region near Si surface is depleted, the carrier collection efficiency is assumed to be 100% because the Si absorption length at 400nm is less than 0.1μm. The Q.E. ratio is estimated to be $\frac{0.48 \cdot 0.44 + 1 \cdot 0.56}{0.48} = 1.61$, which matches pretty well with the measured data at 400nm.

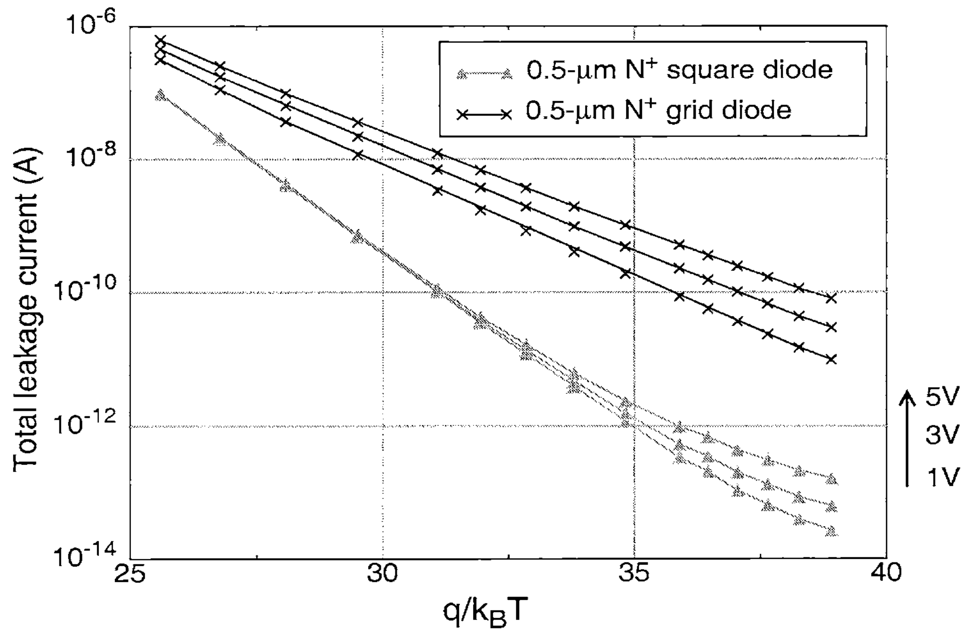


Figure 3.16: Arrhenius plot of leakage current for the grid-shape and the square-shape N^+/Psub photodiodes with the 0.5-μm process

In spite of the Q.E. enhancement at short wavelength, the utilization of a grid-shape photodiode is not always advantageous. The leakage current of the grid-shape photodiode, however, is much higher than that of the square-shape diode due to the high density of G/R centers at the field edge. The Arrhenius plots of leakage current for both structures are presented in Figure 3.16. The temperature sweeps from 25°C up to 180°C. The grid-shape photodiode shows more than two orders of magnitude higher leakage current than the square-shape diode. The strong voltage dependence for the grid-shape diode and the slope of the curve near $E_g/2$ suggest that the generation component of leakage current absolutely dominates in the whole temperature range.

It can be derived from the above measurement that the leakage current generated from the field edge at the N^+-diffusion periphery is of much greater importance than the area component. This is also true for small-dimension N^+-diffusion photodiodes because of the high periphery/area ratio. Therefore, a good photodiode design should try to minimize the periphery length. Also the process effort should focus on reducing the generation component of leakage current from the field edge.

3.6 Summary

This chapter describes the characterization of CMOS processes for image sensor applications using large-area photodiodes. Three major parameters of photodiodes — spectral response (Q.E.), leakage current, and capacitance are discussed, both theoretically and experimentally. Dielectric-layer optimization and photodiode layout strategies are discussed. The characterization results are summarized as the follows.

1. The improvement on transmission coefficient of the dielectric layers can be realized by inserting buffer layers with refractive index between SiN and Si. For the best case, the transmission coefficient can be improved from 0.73 (SiN/Ox/Si) to 0.79 (Ox/SiON/SiN/SiON/Ox/Si). The drawback is the increased process complexity.

2. The NW/Psub photodiodes demonstrate lower dark current density, higher quantum efficiency, and higher charge-to-voltage conversion gain than the N^+/Psub diodes. The results can be explained by a better-annealed p-n junction, a longer minority carrier diffusion length, and a wider depletion region width in the NW/Psub photodiodes.

3. An N^+/PW/Psub photodiode could possibly present a lower leakage current than the NW/Psub photodiode due to the narrower depletion region and the lower minority-carrier concentration in the lightly doped side. However, the narrow depletion region and the internal E-field resulted from P-well doping profile significantly reduces the Q.E.

4. The Arrhenius plot of leakage current is effective in observing the dominate leakage-current mechanism. For large-area photodiodes, the generation current component from the depletion region dominates at lower temperature and the minority-carrier diffusion component dominates at higher temperature.

5. A photodiode with grid-shape layout can be used to improve Q.E. at short wavelength because of the wider depletion region than a square-shape photodiode. However, the higher G/R density at the field edge produces more than two orders of magnitude higher leakage current. This high leakage current prohibits this structure to be used for typical imaging applications. The measurement result also suggests that a photodiode should be laid out with less periphery length as possible. Process effort should also be addressed on reducing the G/R centers at the field edge.

6. From the characterization results, an NW/psub photodiode is recommended to be

the sensor structure used for the second-stage characterization on small-dimension pixel arrays.

Chapter 4

Characterization of CMOS Processes: Stage II– Small-Dimension Pixel Arrays

The characterization of large-area photodiodes compares the junction types that are suitable for imaging applications. In the second stage of characterization, small-dimension pixel arrays are implemented for further analysis and optimization on the pixel level. Building small-dimension pixel array as test devices is an cost-effective way to extract pixel-level parameters and is suitable for imager development.

The pixel arrays are designed with three objectives. First, it is utilized to identify the optimized pixel design. Second, from a pixel perspective, it locates the process parameters that need to be improved. Third, the arrays serve as test patterns for confirming process improvement or for monitoring lot-to-lot variations on the sensor parameters.

The characterization of small-dimension pixel arrays is performed on sensor parameters such as Q.E., leakage current, and conversion gain, which are similar to the ones measured from large-area photodiodes except for the small size of the pixels. There are essential parameters that can only be extracted from pixel arrays but not large diodes. For example, the pixel-level noise sources, which include temporal and spatial components, are unable to be characterized with large photodiodes. The pixel-level characterization method is presented using a three-transistor (3-T) NW/Psub photodiode active pixel sensor (APS) array as an example. The approaches are utilized to compare arrays with different pixel designs, sensor junctions, and process parameters.

4.1 Design of 64 x 64 Pixel Arrays

An accurate characterization on the pixel parameters is the goal of the pixel array design. Therefore, proper buffering of the analog signal become the top priority. Because of that, on-chip circuitry is reduced to minimize switching noise and substrate bouncing. The control signals, analog signal amplification, and digitization are provided with off-chip components.

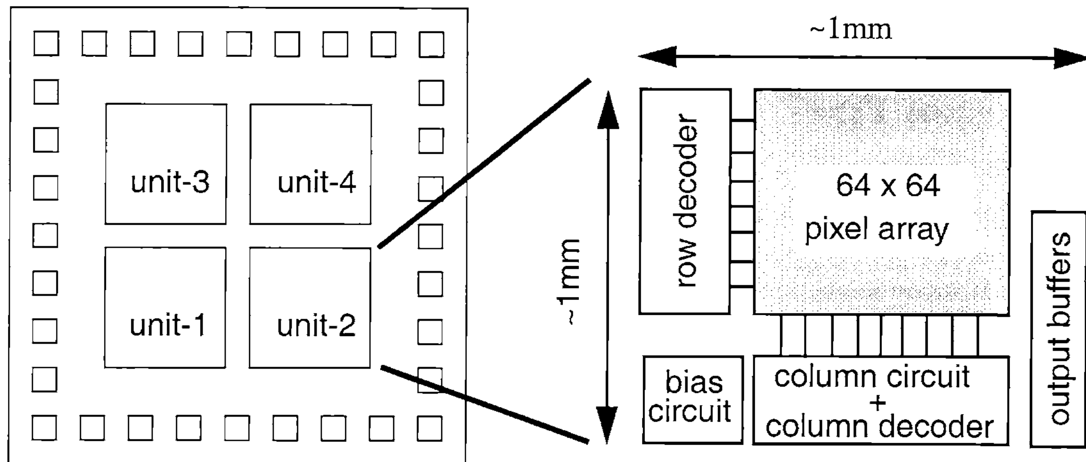


Figure 4.1: Layout floor plan of the chip (left) and the 64 x 64 test pixel array (right)

Figure 4.1 shows the floor plan of the test chip. There are four units on the chip. Each unit contains a pixel array with different pixel designs. Since the process conditions are identical for the four on-chip units, the comparison of pixel parameters between the designs would be fair. The dimension of the pixel array is determined to be 64 x 64, which provides sufficient statistical data with a small Si area. Besides the pixel array, each unit contains a 6-bit row decoder, 6-bit column decoder, column circuitry, analog output buffers, and biasing circuits. The four units share the same input control signals such as row/column addressing bits and others. Each unit also has its own analog output pads so the analog outputs are independent.

The schematic of a 3-T photodiode pixel and its analog signal path are shown in Figure 4.2. The presented signal path starts from the pixel cell and extends all the way down to the output pads. Inside the pixel, there is one p-n junction photodiode and three nMOS transistors. Transistor Q1 is used for resetting the voltage of the photodiode to a pre-determined value. Transistor Q2 and transistor Q4 down in the column circuit form a source follower so that V_{out} will track the photodiode voltage V_d. Transistor Q3 is used as a switch (row-select transistor) to turn on/off the branch current through the source follower transistor Q2. For all the pixels connected on the same column line, only one of the pixels

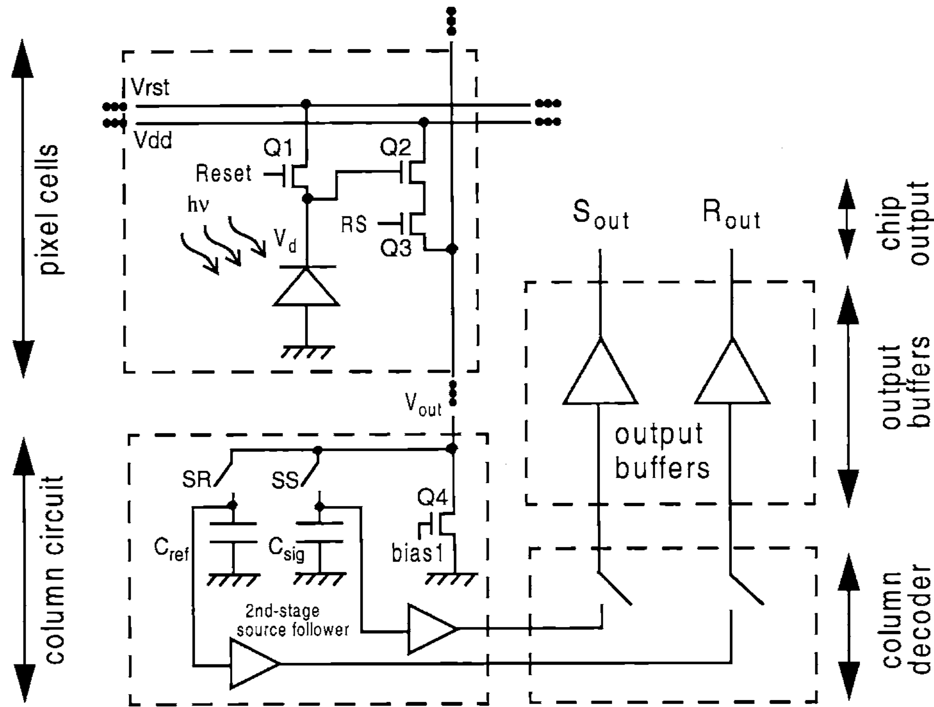


Figure 4.2: Circuit diagram of the 3-T pixel cell and the associated periphery circuit

will be turned on at any time. Vdd and V_{rst} are common to all pixels in the array. *Reset* and *RS* (row select) signals are generated from the row decoder circuit and are common to pixels on the same row. At the bottom of each column there is a column circuitry which consists of the current source transistor Q4, two sampling capacitors — C_{sig} and C_{ref}, two sets of CMOS switches which are controlled by signals *SS* and *SR*, and two second-stage source followers. The capacitors C_{sig} and C_{ref} are implemented with pMOS gate capacitors. The column decoder selects a particular column and connects signals from the second-stage source followers to the output buffers.

The operation of the pixel is described with the timing diagram of the control signals shown in Figure 4.3. First, a particular row is selected when its corresponding *RS* signal turns "high". The reset signal, *Reset,* on that row is pulsed "high" at time t_1 to charge up the voltage of the photodiode to a pre-determined value. The photodiode can be treated as a junction capacitor in this case. The depletion region of the p-n junction increases to sus-

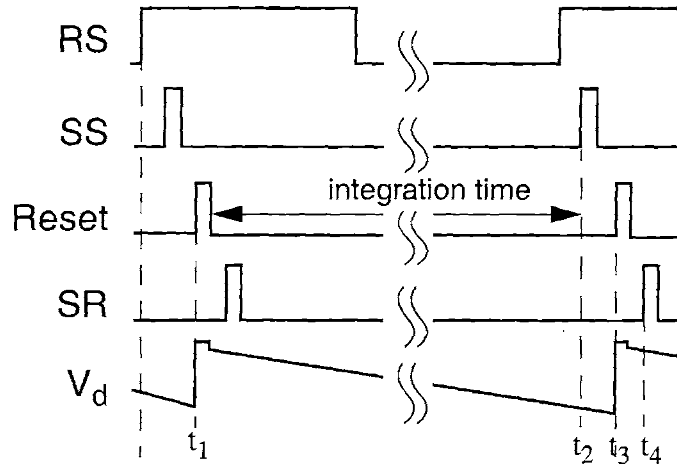


Figure 4.3: Timing diagram for the 3-T photodiode APS array

tain the voltage drop. Due to illumination, the photon-generated carriers steadily accumulate on the junction capacitor and the photodiode voltage V_d continues to drop. The magnitude of the photo-current determines the slope of the V_d curve. The stronger the optical signal, the faster the voltage drops. A certain amount of time is allocated to integrate the photo-carriers. At the end of the integration time t_2, control signal *SS* pulses "high" and samples $V_{out}(t_2)$ onto capacitor C_{sig}. The photodiode is then reset again at t_3. The reset operation is followed by a *SR* pulse to sample $V_{out}(t_3)$ onto C_{ref}. In this way, C_{ref} stores a pre-determined reset voltage and C_{sig} stores the reset voltage plus the optical signal integrated during the integration period. The voltage difference between the two capacitors, (V_{ref} - V_{sig}), represents the magnitude of the optical signal. After the *SR* pulse, the column decoder is then multiplexing through all columns to connect V_{ref} and V_{sig} on each column to the output buffers. The above operations are performed in a row-by-row basis to scan the output signals throughout the whole array. The differential output signal is then converted into a single-ended signal with an off-chip instrumentation amplifier.

The two second-stage source followers on each column are utilized to properly buffer the differential analog signals, V_{ref} and V_{sig}. The circuit diagram is shown in Figure 4.4. PMOS source followers are implemented in this case in order to avoid back-gate effects in nMOS source-followers. Therefore, the voltage gain is nearly unity. The pMOS source

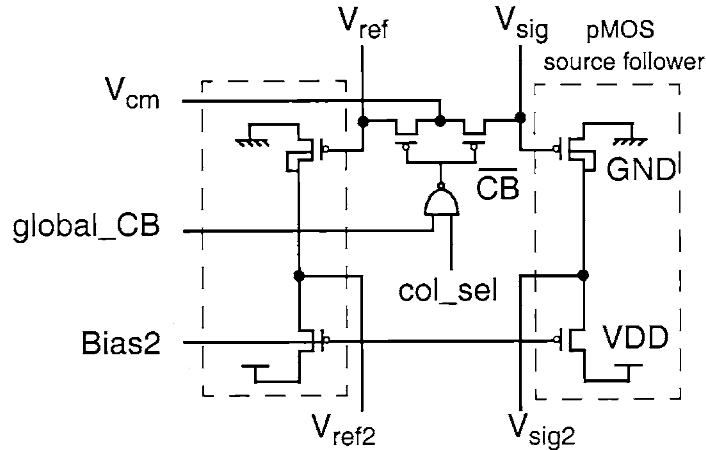


Figure 4.4: Circuit diagram of the second-stage source followers

followers are also used as voltage level shifters to compensate for the voltage drop in the first stage of nMOS source follower. In this way, a better signal swing can be achieved.

Some extra devices in the column circuitry are used for cancelling the offset voltage difference between the two pMOS source followers. As shown in Figure 4.4, there are two pMOS transistors connected across the two input nodes — V_{ref} and V_{sig}. The NAND gate controls the on/off state of the pMOS switches. Normally the switches are "off" and the output voltage difference (V_{ref2} - V_{sig2}) is equal to (V_{ref} - V_{sig})+ΔV_{off}, where ΔV_{off} represents the offset voltage difference between the two pMOS source followers. The ΔV_{off} varies from column to column and is considered as the major source of column-to-column fixed pattern noise (C-C FPN) for this type of readout circuit. It is desired to sample this ΔV_{off} and subtract it from the differential output. In this circuit, the sampling of ΔV_{off} is performed by switching *global_CB* to a "high" state when a specific column is selected for readout (*col_sel* on this column is "high"). Thus, the two input nodes, V_{ref} and V_{sig}, will be shorted to a common voltage V_{cm}. Consequently, V_{ref} equals V_{sig} and (V_{ref2}-V_{sig2}) becomes ΔV_{off}. To utilize this ΔV_{off} cancellation feature, two off-chip sample-and-hold (S/H) buffers are required. One is used to sample the original output (V_{ref} - V_{sig})+ΔV_{off} when the two pMOS switches are "off". The other S/H is for sampling ΔV_{off} when the two branches are shorted. The subtraction of the two S/H outputs corrects the column-to-column fixed pattern noise.

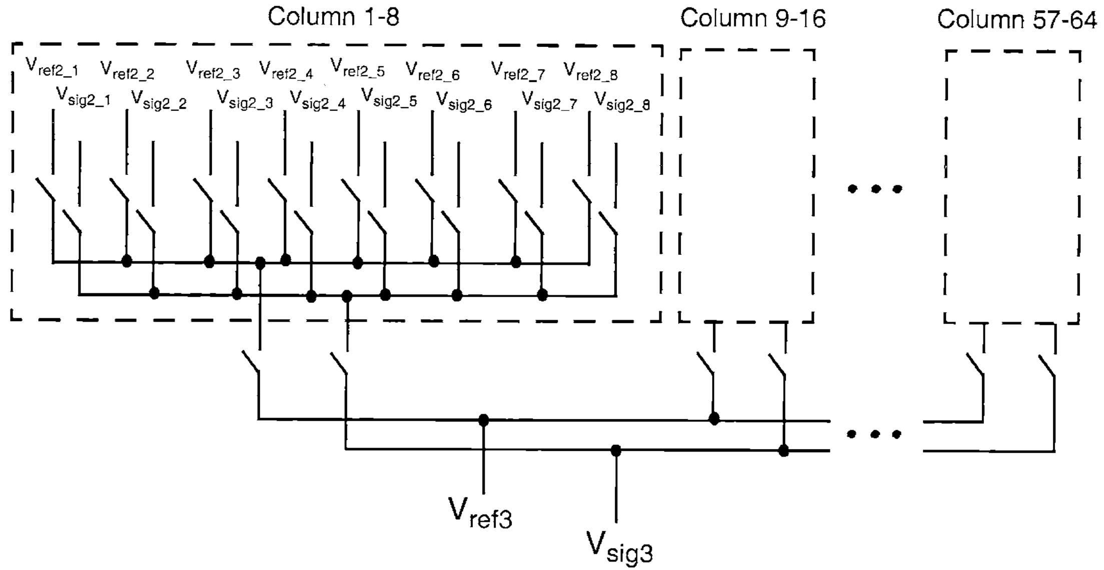


Figure 4.5: Hierarchical multiplexer in the column decoder

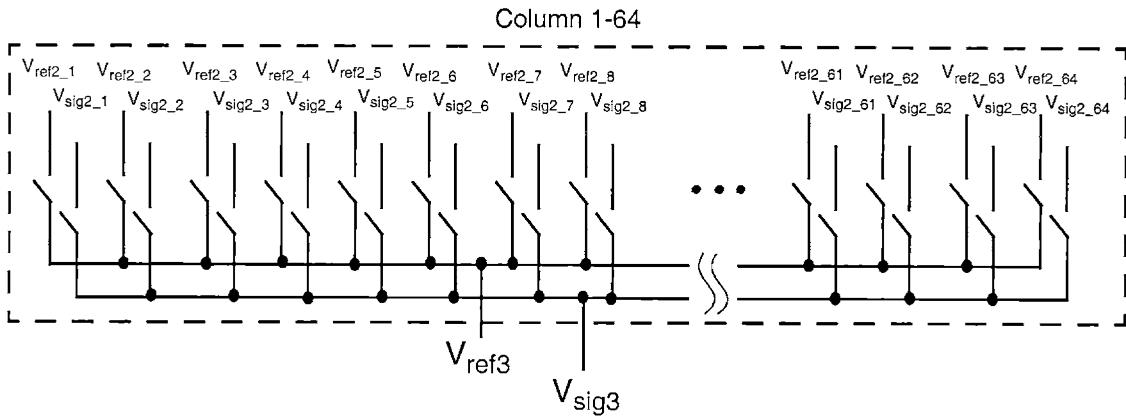


Figure 4.6: Typical multiplexer implementation

A hierarchical multiplexer in the column decoder is utilized to select one pair of the second-stage source-follower outputs, e.g. V_{ref2_n} and V_{sig2_n} on the n-th column, to the output buffers. The schematic of the multiplexer is shown in Figure 4.5. The advantage of the hierarchical structure is a minimized parasitic capacitance on the signal path for the second-stage source followers to drive. The parasitic capacitance is mainly contributed from the N^+-diffusion junctions at the MOS switches. For a typical multiplexer shown in Figure 4.6, the total parasitic capacitance on the signal path consists of the N^+-diffusion

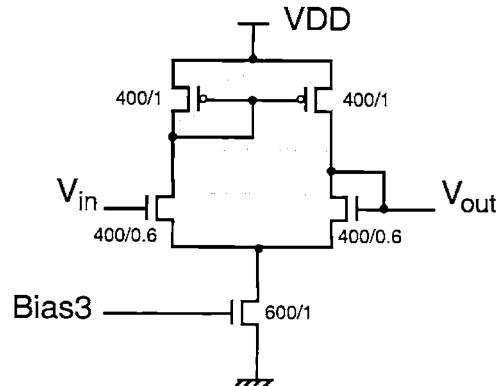


Figure 4.7: Circuit diagram of the output buffer

junctions from the 64 column switches. Whereas for the presented 2-level hierarchical structure, there are only $2 \times 8 = 16$ junction nodes on the signal path. The parasitic capacitance reduction is more significant for imagers with a large number of columns.

Two output buffers are connected to the outputs of the hierarchical multiplexer for driving the off-chip parts. The circuit diagram for the buffer is shown in Figure 4.7. The two pMOS transistors on top splits the bottom tail current into half. Since the two nMOS transistors below the pMOS transistors are of the same size, V_{gs} of the two nMOS transistors should be the same. Because the source nodes of the two nMOS transistors are shorted, V_{out} would simply follow V_{in} with a unity gain. This circuit avoids voltage-gain reduction due to back-gate effects in nMOS source followers. However, this circuit requires two times the branch current in a typical nMOS source follower so the power consumption is higher. Since maintaining the analog-signal accuracy is the major concern in the experiments, power consumption is not optimized for this design.

The on-chip digital circuits, row decoders and column decoders of the array, are implemented with typical NAND-NOR logic circuits to realize functions similar to the following:

$$R_1 = \overline{a_5} \cdot \overline{a_4} \cdot \overline{a_3} \cdot \overline{a_2} \cdot \overline{a_1} \cdot \overline{a_0},$$

$$R_2 = \overline{a_5} \cdot \overline{a_4} \cdot \overline{a_3} \cdot \overline{a_2} \cdot \overline{a_1} \cdot a_0,$$

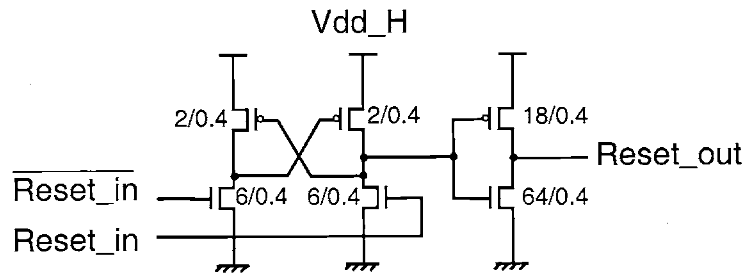


Figure 4.8: Schematic of the level shifting circuit for reset signal

$$R_3 = \overline{a_5} \cdot \overline{a_4} \cdot \overline{a_3} \cdot \overline{a_2} \cdot a_1 \cdot \overline{a_0},$$

$$\vdots$$

$$R_{64} = a_5 \cdot a_4 \cdot a_3 \cdot a_2 \cdot a_1 \cdot a_0,$$

where R_n represents row-select or column-select signal on the n-th row or column. The reset signal (*Reset*) on each row is implemented by an AND operation of the row-select (*RS*) signal and a global reset signal. In order to obtain a flexible reset voltage level, a level-shifting circuit is implemented before *Reset* is connected into the pixels. The level-shifting circuit is shown in Figure 4.8 The "high"-level voltage at the output is converted to Vdd_H.

In order to minimize the influence of the injected photons to the circuit operation, the top metal layer is utilized as a metal shield to cover up the whole chip except for the photo-sensing area. The layout of the test chip is shown in Figure 4.9. A 400 x 400μm^2 NW/Psub photodiode is also placed on the chip to characterize the junction properties. NMOS and PMOS test devices are also built. Layout for one of the pixel arrays is shown in Figure 4.10. The configuration of the other three arrays are similar.

Two sets of small-dimension imager chips are fabricated with different CMOS processes. The first set is fabricated with a HP 0.35-μm standard CMOS process through MOSIS VLSI Fabrication Service [43]. Pixel structures on this chip include several designs of 3-T photodiode, 4-T photodiode, and photogate APS arrays. The purpose of this chip is to compare pixel parameters between different pixel structures. The second set

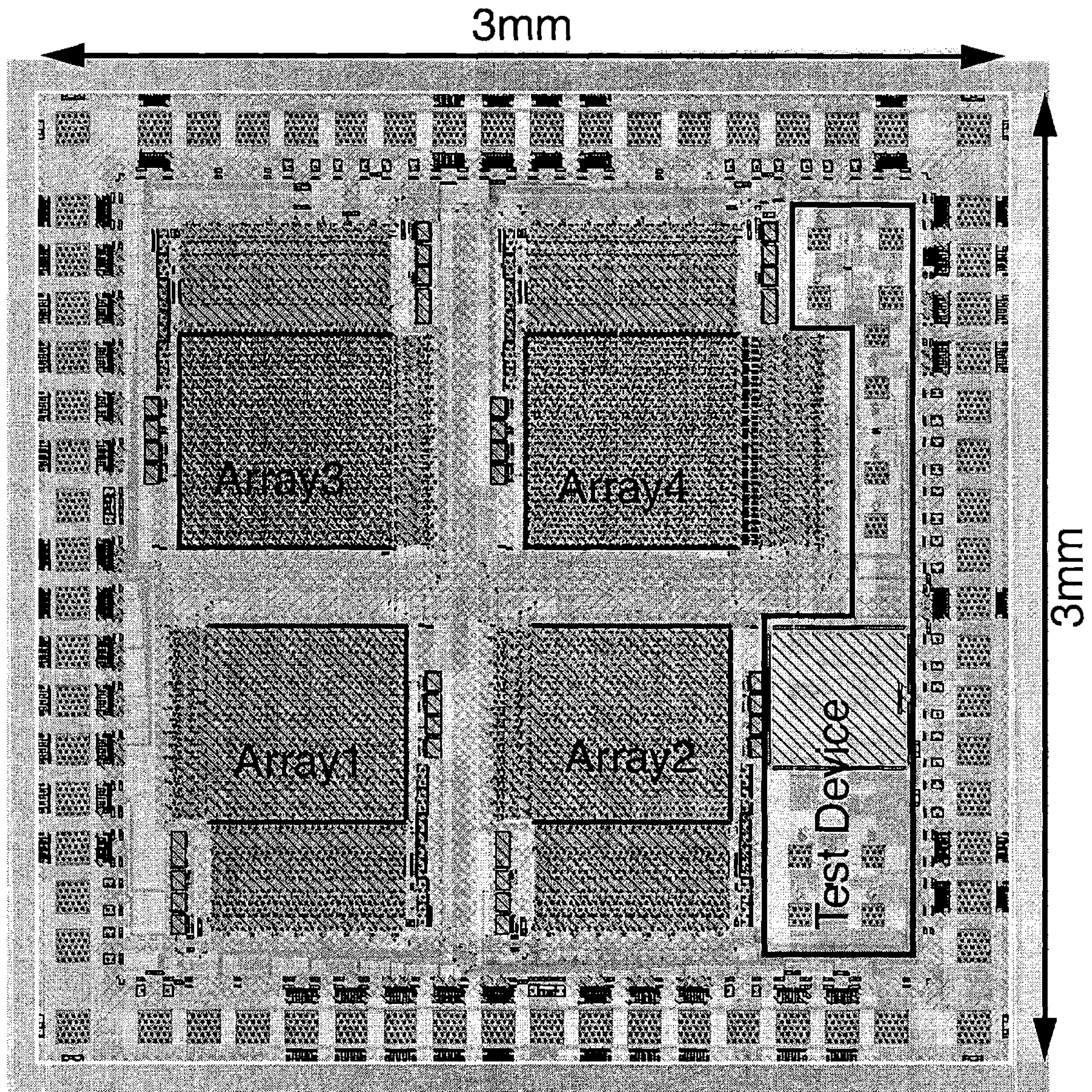


Figure 4.9: Layout of the imager chip

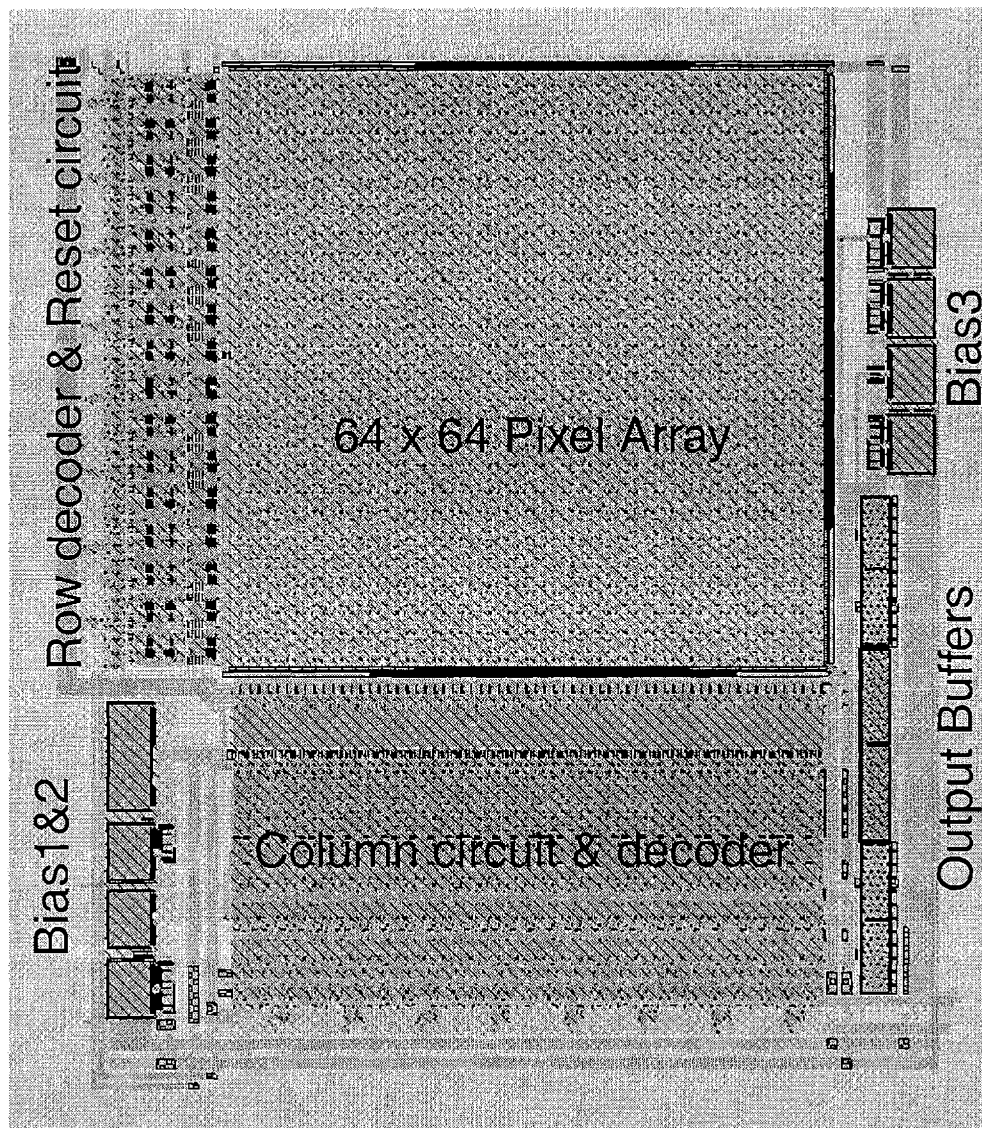


Figure 4.10: Layout of a 64 x 64 photodiode APS array

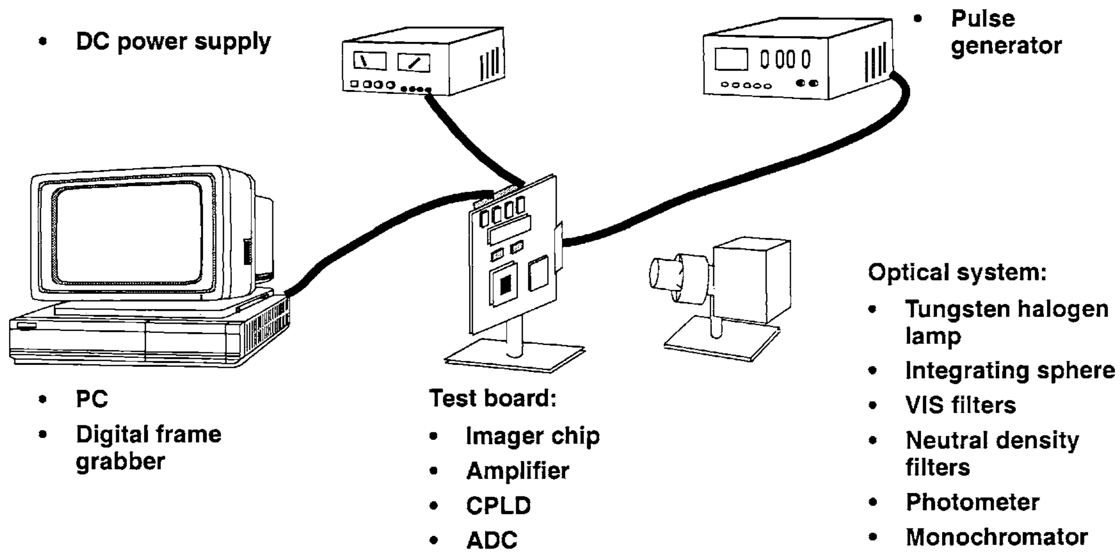


Figure 4.11: Experimental setup for sensor array characterization

of imager chips is fabricated with a TSMC [44] 0.25-μm non-silicided source/drain shallow-trench-isolation (STI) CMOS process. Chips fabricated with this process include the variation of sensor junctions, and process parameters such as substrate types, photodiode implant dose and energy. A special metal shield is further designed for a crosstalk study, which is the topic for Chapter 6. All the chips are packaged with Kyocera 84-pin ceramic PGA packages. The pin outputs between the two sets of chips are identical so that they can be characterized with the same test board.

4.2 Experimental Setup

The experimental setup for characterizing the pixel arrays is shown in Figure 4.11. The major parts consist of a test board, an optical illuminating system, and a PC with a digital frame grabber (model: Genesis-LC from Matrox). Periphery equipment such as pulse generators and power supplies are also required.

A photo of the test board is shown in Figure 4.12. The major components on the board includes the imager chip, two instrumentation amplifiers, a 12-bit analog-to-digital converter (ADC), several digital output (RS-422) drivers, and some voltage regulators. The voltage regulators provides ±9V, ±5V, 3.3V for analog chips and +3.3V, +5V for digital

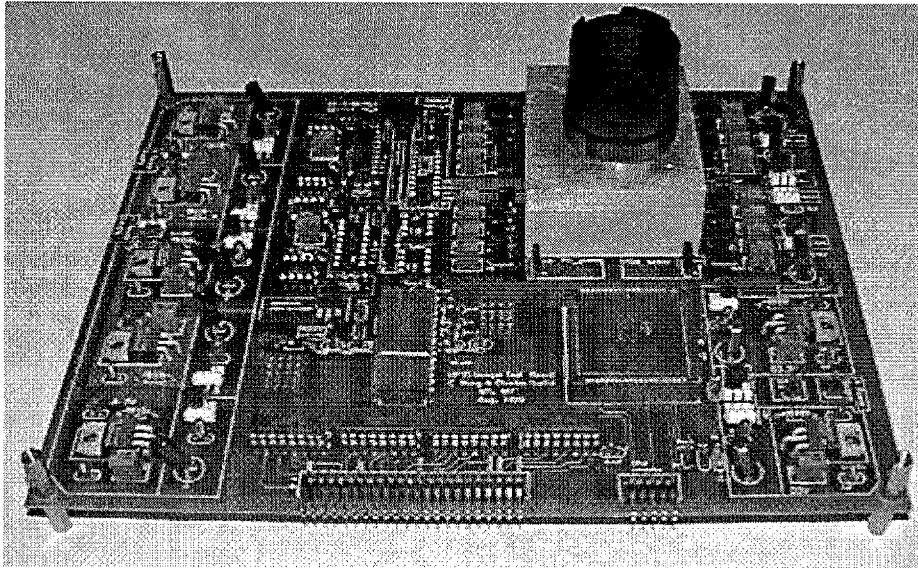


Figure 4.12: Photo of the test board

chips. The imager chip is located near the center of the board. Drill holes are prepared for assembling the lens mount. There are four pairs of analog differential outputs coming out from the imager chip. Each pair corresponds to one of the four pixel arrays. One of the output pair is selected by a 4-differential-channel analog multiplexer chip (model: ADG-609 from Analog Devices, Inc.). The differential outputs are amplified by a low noise, low distortion instrumentation amplifier (model: INA-103 from Burr-Brown). The gain of the amplifier can be adjusted with an external resistor R_{ext} and calculated from $G = 1 + 6K\Omega/R_{ext}$. The output of this amplification stage is single-ended and the offset level can be adjusted with a potentiometer. The signal path is followed by two S/H chips. The S/H chips are utilized for canceling the offset voltage difference between the two on-chip second-stage source followers as described in the previous section. The subtraction of the two S/H outputs is performed by another instrumentation amplifiers. The board is designed so that the two S/H chips can be by-passed in order to measure the column-to-column fixed pattern noise. The final analog signal is then digitized by a 12-bit 5MSPS analog-to-digital converter (model: AD-871 from Analog Device, Inc.) The digital outputs are buffered by four RS-422 drivers before feeding into the frame grabber. All the control

signals are generated by a CPLD (complex programmable logic device, model: CY37064P84 from Cypress Semi. Corp.). The programming code is written with VHDL and loaded into the CPLD through a RS-232 interface. The in-system programming ability of the CPLD provides high flexibility for generating the timing pattern of the control signals. Special timing patterns, such as the ones for varying the integration time or addressing a specific row or column, can be programmed into different files and easily download into the CPLD within seconds.

The illumination system includes optical instruments utilized for spectral response and sensitivity measurements. The monochromator, the bifurcated fiber, and the optical power meter are used for Q.E. measurement. For sensitivity measurements, an integrating sphere (model: US-080-SF from Labsphere, Inc.) with an external tungsten halogen light source is used. The integrating sphere is connected to a filter holder for holding a visible-light filter (filter# 59062 from Oriel) and a set of neutral density filters with density from 1 to 3 (filter#: 59350, 59360, and 59370 from Oriel). The neutral density filters are for wide-range intensity adjustment. Small-range intensity adjustment is performed with a manual shutter at the input port of the sphere. Varying the voltage setting of the light source during measurement is not recommended because this will affect the output optical spectrum. A silicon photometric sensor (model: SED033 from International Light, Inc.) is mounted on the integrating spheres and connected to a research radiometer (model: IL-1700 from International Light, Inc.) for light intensity measurement. All optical measurements are conducted in a metal shield box to reduce optical and electrical noises.

A PC is used for real-time image display, data storage, and data analysis. It is equipped with a GPIB board and Labview software, which automates the test instruments such as the monochromator, the optical power meter, and a temperature measurement module.

For temperature varying experiments, a Peltier Junction device is used. The assembly of the device with the imager chip is shown in Figure 4.13. This Peltier Junction device acts as a thermal pump which utilizes thermal-electric effects to convey heat from one side of the junction to the other. The pumping direction depends on the polarity of the current flowing through the junction. This device is mounted on a heat sink and attached to the package of the imager chip during measurements. A K-type thermal couple is attached on

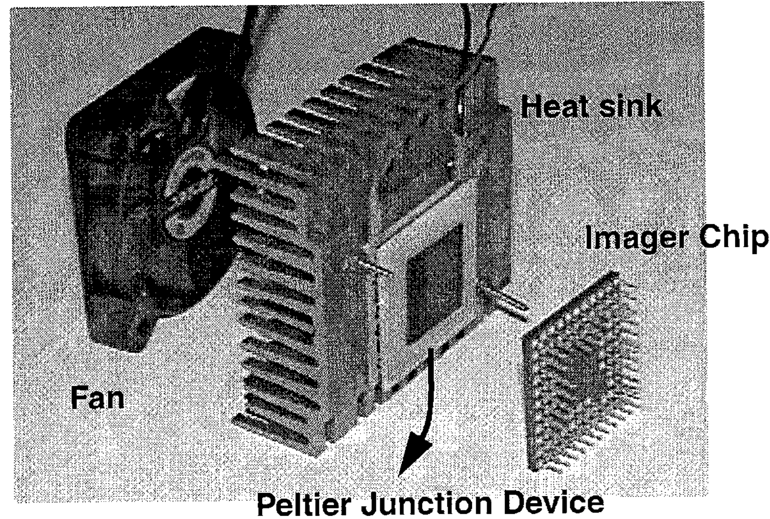


Figure 4.13: Devices for temperature-varying experiment

the package close to the chip for temperature measurement. A Keithley M1321 K-thermocouple calibration module with RS-232 interface is used for acquiring temperature data into the PC. This setup allows temperature variation of the imager chip from 5°C to 75°C without affecting the temperature and the characteristics of the other on-board parts.

4.3 Simulation

A device simulator is utilized to predict the Q.E. of junctions with different process parameters such as implant dose or energy. The main purpose of the simulation is to optimize the doping profile of photodiodes for a better quantum efficiency. The doping profile can be created with a specialized implantation to create the desired junction. The extra process steps involve one more mask, one extra photo-exposure step, and one or several implantation operations. Since a typical CMOS process consumes 12–20 masks, one extra mask step should have minor impact on the total fabrication cost.

A doping profile is the input parameter for the Q.E. simulation. If the process steps are known, the profile can be simulated from a process simulator such as *TSUPREM4* from Avant! Corp.[38]. For some standard implantation conditions, the doping profile is calculated from the measured spreading resistance data. The doping profile is then fed into a 2-D device simulator for the Q.E. simulation. The device simulator is *Atlas* from Silvaco

International [39]. The simulator utilizes the well-known solid-state physics equations for the simulation. The following equations are some of the fundamental ones.

(1). Poisson's equation:

$$\nabla \cdot (\varepsilon \nabla \psi) = -\rho \,, \tag{4.1}$$

where ψ is the electrostatic potential, ε is the local permittivity, and ρ is the local space charge density.

(2). Carrier continuity equations:

$$\frac{\partial n}{\partial t} = \frac{1}{q}(\nabla \cdot \vec{J_n}) + G_n - R_n \,, \tag{4.2}$$

$$\frac{\partial p}{\partial t} = \frac{-1}{q}(\nabla \cdot \vec{J_p}) + G_p - R_p \,, \tag{4.3}$$

where n and p represent electron and hole concentrations, $\vec{J_n}$ and $\vec{J_p}$ are electron and hole current density; G_n and G_p are the generation rates for electrons and holes; R_n and R_p are the recombination rates for electrons and holes; q is electron charge.

(3). Transport equations from Drift-Diffusion Transport Model:

$$\vec{J_n} = qn\mu_n \vec{E_n} + qD_n \nabla n \,, \tag{4.4}$$

$$\vec{J_p} = qp\mu_p \vec{E_p} - qD_p \nabla p \,, \tag{4.5}$$

where μ_n and μ_p are electron and hole mobilities; D_n and D_p are electron and hole diffusion coefficients.

The carrier generation rate can be calculated from the injected optical power and the silicon absorption coefficient shown in Figure 3.2. Reflection on dielectric interfaces is considered and calculated using ray-tracing method [40, 41]. The applied carrier-lifetime model is according to formulas derived from empirically fitted data [42]. Most of the other coefficients associated with silicon are default parameters built in the simulator.

The Q.E. simulation is applied to the imager chips fabricated with the 0.25 μm process. The results are presented in Section 4.5.

4.4 Imager Pixel Parameters

This section provides a general description on pixel parameters and the approaches used to characterize the parameters. The parameters include conversion gain, spectral response,

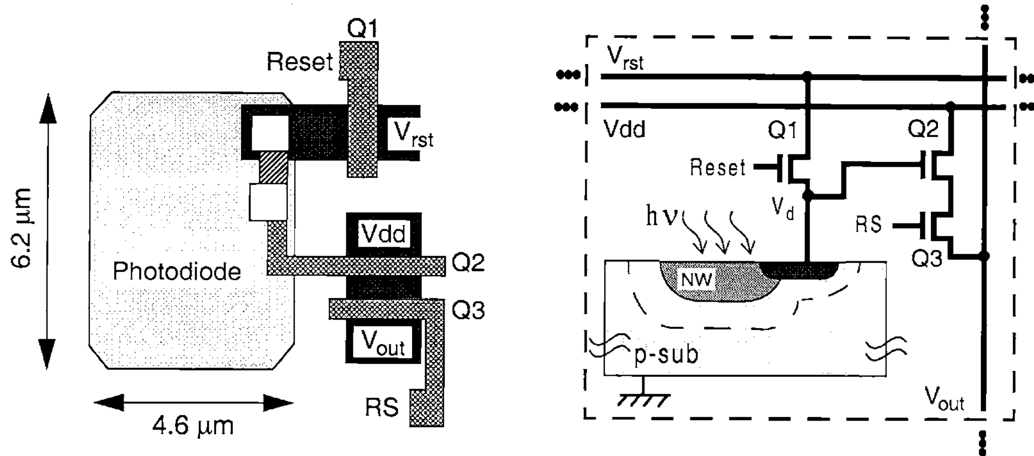


Figure 4.14: Layout and schematic of the 3-T NW/Psub photodiode APS

leakage current, sensitivity, and noise. The characterization approaches are demonstrated using a 3-T NW/Psub CMOS APS array with a pixel size of 9.2 x 9.2 μm^2. Layout and schematic of the pixel is shown in Figure 4.14. This chip is fabricated with a standard HP 0.35-μm silicide process. The results demonstrate an example of modern standard CMOS processes utilized for imaging applications.

4.4.1 Conversion Gain

Conversion gain typically refers to the output voltage variation of an image sensor when a photo-carrier is captured by the charge sensing node. As this voltage variation is referred back to the charge sensing node, the conversion gain becomes essentially a measurement of the capacitance of the charge sensing node. The relation between conversion gain (CG) and capacitance (C) is $CG = q/C$, where q is unit charge. For a 3-T photodiode sensor shown in Figure 4.14, the photodiode acts as the charge sensing node. The capacitance of the sensing node mainly consists of the junction capacitance of the photodiode, the source-to-gate coupling capacitance of Q1, gate capacitance of Q2 and other parasitic capacitance associated the node. The capacitance of the sensing node can be estimated with the following equation:

$$C = C_{gs-Q1} + C_{g-Q2} + C_{sw} \cdot L + C_j \cdot A + C_p, \tag{4.6}$$

where C_{sw} represents side-wall capacitance of the photodiode, L is the periphery length, C_j is the junction capacitance, A is the photodiode area, and C_p represents other parasitic

components such as the metal-to-sensing node coupling capacitance.

A direct measurement of the sensing-node capacitance is difficult since the capacitance is typically on the order of fF (10^{-15} F). Two approaches are typically adopted for measuring the sensing-node capacitance. The first one is to inject a known amount of charge into the node and measure the voltage difference before and after the charge injection [26]. The capacitance can be calculated from $C = \frac{\Delta Q}{\Delta V}$. This method requires extra circuitry for controlling charge flowing into the sensing node. The other approach utilizes an assumption that the photo-carrier absorption is a Poisson process. Namely, the absorption of each carrier is treated as an independent probabilistic event. As the events are integrated for a certain amount of time, statistically the mean value of the total number of events is N and the standard deviation of the number of events will be $\sqrt{N}$ [37]. This uncertainly of carrier absorption represents *photonic shot noise*, which is originated from the particle property of photons. The shot-noise process is a well-accepted assumption in the imaging industry. In this work, the shot-noise approach is adopted for conversion gain estimation.

The measurement is performed by applying an illumination with sufficient intensity on the imager so that photonic shot noise dominates the noise mechanism. One hundred repeated measurements is performed. The mean value of the voltage output is denoted as V_{out} and the standard variation as V_{noise}. From the photonic-shot-noise assumption, the two voltage values can be written as

$$V_{out} = N \cdot CG \cdot G_{SF}, \tag{4.7}$$

$$V_{noise} = \sqrt{N} \cdot CG \cdot G_{SF}, \tag{4.8}$$

where CG is conversion gain and G_{SF} is the voltage gain for the on-chip source followers. G_{SF} can simply be obtained by turning on all the switches on the analog signal chain and measuring the slope between V_{out} and V_{rst}. Apply some algebra on eq.(4.8) and eq.(4.7), the conversion gain can be written as

$$CG = \frac{V_{noise}^2}{V_{out}} \cdot \frac{1}{G_{SF}}. \tag{4.9}$$

This approach is applied to the 3-T NW/Psub APS array. The illumination level is adjusted so that V_{out} is at half of the saturation level. From the signal and noise plot presented in Figure 4.19, the imager is assured to be operating in the photonic-shot-noise

dominate region. Applying eq.(4.9), the conversion gain is calculated to be 10 μV/e and the corresponding capacitance on the sensing node is 16fF.

A correct estimation of conversion gain is critical in pixel array characterization since there are many other sensor parameters derived from it. In the estimation, the voltage dependence of the sensing node capacitance is not considered for simplicity. In reality, the sensing node capacitance dynamically increase with the integrated charges due to the p-n junction.

4.4.2 Spectral Response

Measurement

The measurement of spectral response on pixel arrays is similar to the case on large-area photodiodes but with a slight difference of the meaning. Since the photo-sensor does not cover 100% of the pixel area, there is a built-in area factor associated with the Q.E. of the pixel arrays. This factor is called *fill factor* and is defined as the optically active area divided by the pixel area. Because of that, the Q.E. of pixel arrays is smaller than that of large-area diodes roughly by the ratio of fill factor.

The Q.E. of a pixel array can be calculated from

$$QE = \frac{h\nu}{P \cdot T_{int}} \cdot \frac{V_{out}}{CG \cdot G_{SF}}, \tag{4.10}$$

where P denotes the optical power illuminating on a pixel at a specific wavelength and T_{int} represents the integration time.

The Q.E. of the NW/Psub APS array measured with eq.(4.10) is shown in Figure 4.15. The Q.E. of the large-area NW/Psub diode is also presented for comparison. The Q.E. of the large diode is scaled down with a multiplication factor of 0.4 so that a better comparison can be made. This figure shows that the Q.E. of the pixel is quite similar to that of the large diode with a scaling factor slightly larger than the fill factor (~0.35 in this case). The matching of both curves is expected since both devices have the same junction profile and therefore should exhibit the same carrier collection efficiency. The slightly larger scaling factor (0.4) than the fill factor (0.35) is considered to be due to the depletion region at the

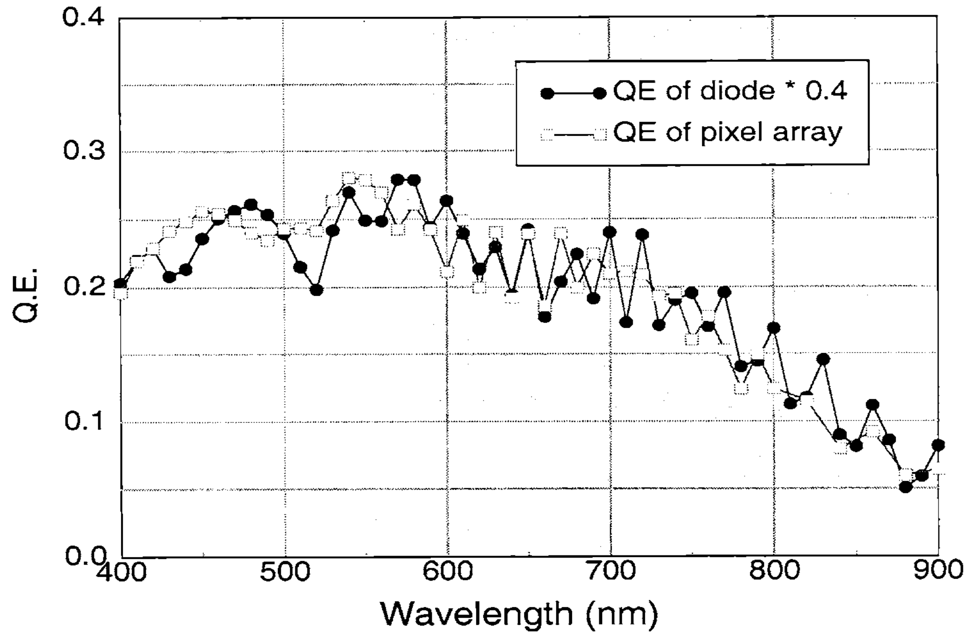


Figure 4.15: Measured Q.E. for the NW/Psub photodiode APS and the large NW/Psub photodiode

periphery of the pixel photodiode, producing a larger optically active area than the size of the drawn feature.

4.4.3 Leakage Current

Leakage current is a sensor parameter highly dependent on process conditions. Due to the complex process dependence, it is difficult to derive a universal model that fits all the CMOS processes. On the other hand, the small magnitude of leakage current density imposes challenge on most of the device simulators since a much higher resolution on the numerical representation is required. Due to the above problems, the most reliable way to quantify leakage current is through measurements on real devices.

Direct measurement of leakage current on the pixels is a formidable task. Usually, an integration approach is utilized. The approach is performed by accumulating leakage current at the charge sensing node. Since the capacitance of the charge sensing node converts the collected carriers into voltage, the leakage current can be calculated by

$$I_{Leak} = \frac{q \cdot V_{out}}{CG \cdot G_{SF} \cdot T_{int}}, \tag{4.11}$$

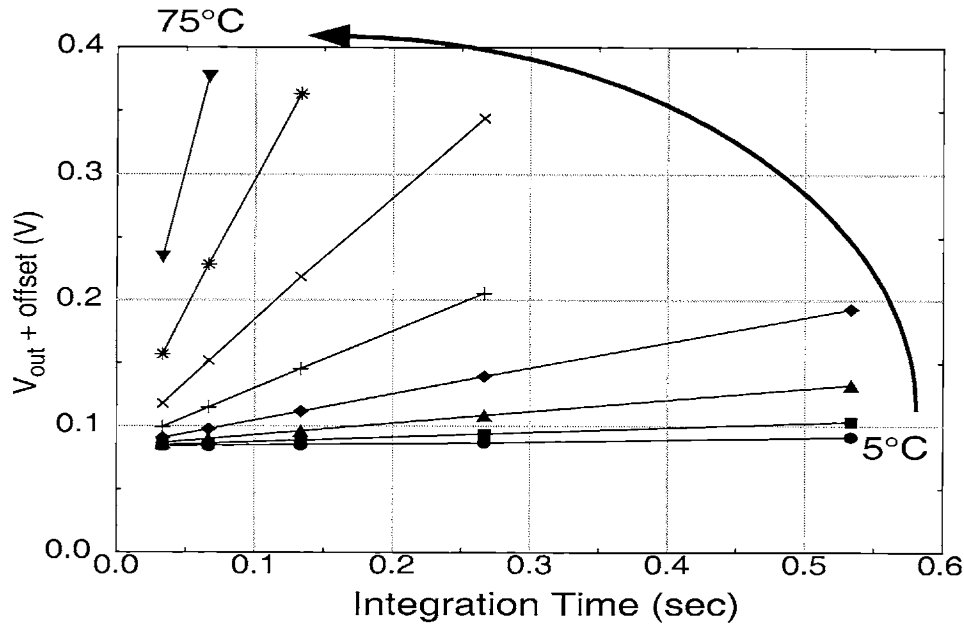


Figure 4.16: Dark signal versus integration time for the NW/Psub photodiode APS at various temperatures

where V_{out} is the output voltage with measurement performed in a dark environment; T_{int} is the integration time.

In the dark current measurement, the integration time T_{int} is set to 2^n multiples of 1/30 second, where n varies from 0 to 4. The 1/30 second is used as a time unit since it corresponds to a typical motion-picture frame rate of 30 frames/sec. The results on the NW/Psub pixel array are shown in Figure 4.16. Each data point is an average value of the frame with one hundred times of measurements. The output voltage consists of a DC offset from the on-chip source followers and the amplification stages. Each curve represents a specific temperature setting, which is varied from 5°C to 75°C with steps of 10°C. The slope of the curves is plugged in (4.11) to calculate the leakage current. The Arrhenius plot is shown in Figure 4.17. The slope of the Arrhenius plot represents the activation energy of the leakage charge generation. The energy is calculated to be 0.68eV, a value close to $E_g/2$. According to the findings in Chapter 3, this result shows that the leakage current source is dominate by the G/R centers residing near the center of Si bandgap. In order to determine whether the leakage current is dominated by area or periphery components, the temperature-varying measurement is also performed on the large-area NW/Psub photo-

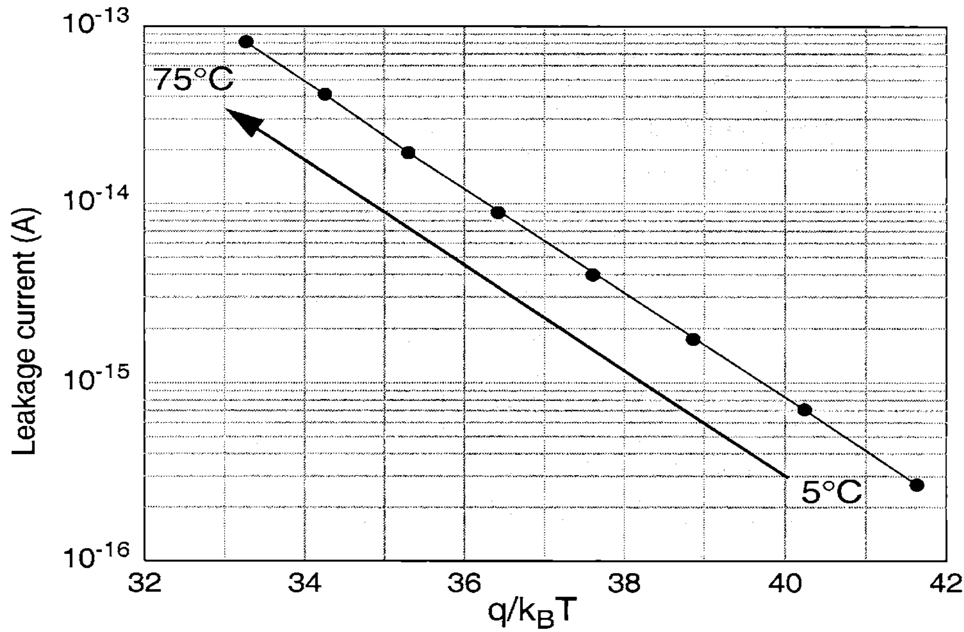


Figure 4.17: Arrhenius plot of leakage current for the NW/Psub photodiode APS

diode and the activation energy is calculated to be 0.9eV, which is between E_g and $E_g/2$. This result suggests that the leakage current in the pixel array is dominated by the periphery component, which should be highlighted for process improvement.

Utilizing the conversion gain estimated previously, the total leakage current of a pixel at room temperature is calculated to be 1.74fA, corresponding to a dark signal of 3.5mV for an integration time equal to 1/30 sec. This signal level is roughly equivalent to 1 LSB (Least Significant Bit) for an 8-bit ADC with 1V full scale.

4.4.4 Sensitivity

Sensitivity determines the output signal of an image sensor illuminated by a certain light level within a specific integration time. For CMOS APSs, the unit is usually in *volts per lux per second* (V/lx/sec) or *amps per lux per second* (A/lx/sec). Conceptually, sensitivity is similar to spectral response in the sense that both parameters represent how the image sensors respond to light. In fact, sensitivity can be derived from spectral response if the spectral energy distribution of the illumination is known.

Sensitivity is a sensor parameter with certain ambiguities. The ambiguity needs to be clarified in order to make a fair comparison between the quoted data from different imagers.

The first ambiguity comes from the photometric property, *lux*, built in the unit. Unlike radiometric units which are defined clearly in terms of energy or power, the photometric unit is a bio-physical term. The response of the human visual system to light is taken into consideration for the photometric unit because our eyes only respond to a specific range of the optical spectrum — the visible light (~380–780nm). The spectral sensitivity function[1] of human visual system and the conversion between the photometric and radiometric quantities are described in Appendix A. This spectral response function (shown in Figure A.1) is quite different from the spectral response of silicon image sensors (Figure 3.7). If the incoming photons are all near-infrared (λ ~1 μm), the illumination is almost zero lux but the image sensor will have finite output. Under such circumstance, the sensitivity of the image sensor becomes unreasonably high compared with the case of green-light illumination. To solve this problem, the spectral distribution of the light sources needs to be specified. Standard light sources are defined by the Commission Internationale de l'Éclairage (CIE), an international organization in charge of defining illumination standards. The spectral distribution of some standard light source is shown in Appendix B. In this thesis, the utilized light source is a 120W tungsten halogen light source. The color temperature is around 3000K. A visible light filter is utilized as an IR-cut and the external transmittance is shown in Figure 4.18.

The second ambiguity comes from the probe point where the sensitivity is defined. If the sensitivity is defined at chip output, on-chip analog amplifiers can be used to magnify the output signal. As a result, the sole description of sensitivity becomes meaningless since a fair comparison can not be made. To solve the flaw, the accompanied noise performance should also be described since the amplifier will magnify the noise as well. The signal-to-noise (S/N) ratio would become a reasonable parameter for comparison. However, this complicates the scenario since the noise performance of the amplifiers needs to be considered. This problem can be solved if the sensitivity and noise is taken at the

1. This function is called spectral luminous efficiency function, usually denoted as $V(\lambda)$.

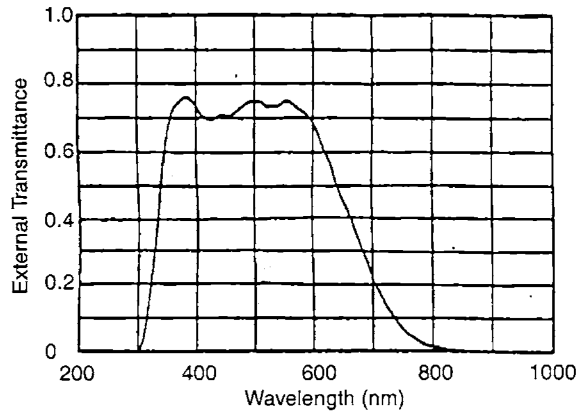


Figure 4.18: Transmittance of the visible light filter

charge sensing node. The measured parameters would be generically from the image sensors and independent of any other on-chip electronic components. This approach is taken throughout the rest of the thesis.

The sensitivity is calculated from 100 consecutive frames of data at various illumination levels. The integration time is set to 1/30 second. The mean values and standard deviations are computed for each pixel, which represent the signal and random noise of each pixel. The signal and the noise values are then averaged throughout the array to represent the signal and random noise of the sensor. The signal and random noise of the NW/Psub APS array is shown on a log-log plot in Figure 4.19. The slope of the signal is unity and the slope of the noise is 1/2 at higher illumination levels. This square-root behavior of the random noise is due to the photonic shot noise. Dynamic range of a sensor is defined as the full range of signal divided by the noise floor. The full signal range for APS chips is typically determined by the circuit operating range. For the NW/Psub APS, the full signal swing is roughly 1.2V. The measured sensitivity is 1.3V/lx/sec, and the noise floor is about 0.4mV. Therefore, the intrinsic dynamic range of the sensor is 70dB. The S/N ratio versus illumination of the pixel is shown on a log-log plot in Figure 4.20. It is observed that the S/N curve increases linearly with illumination at low illumination and shifts to a square-root relation at higher illumination. This plot determines the minimum illumination needed for the imager performance to meet the S/N specification.

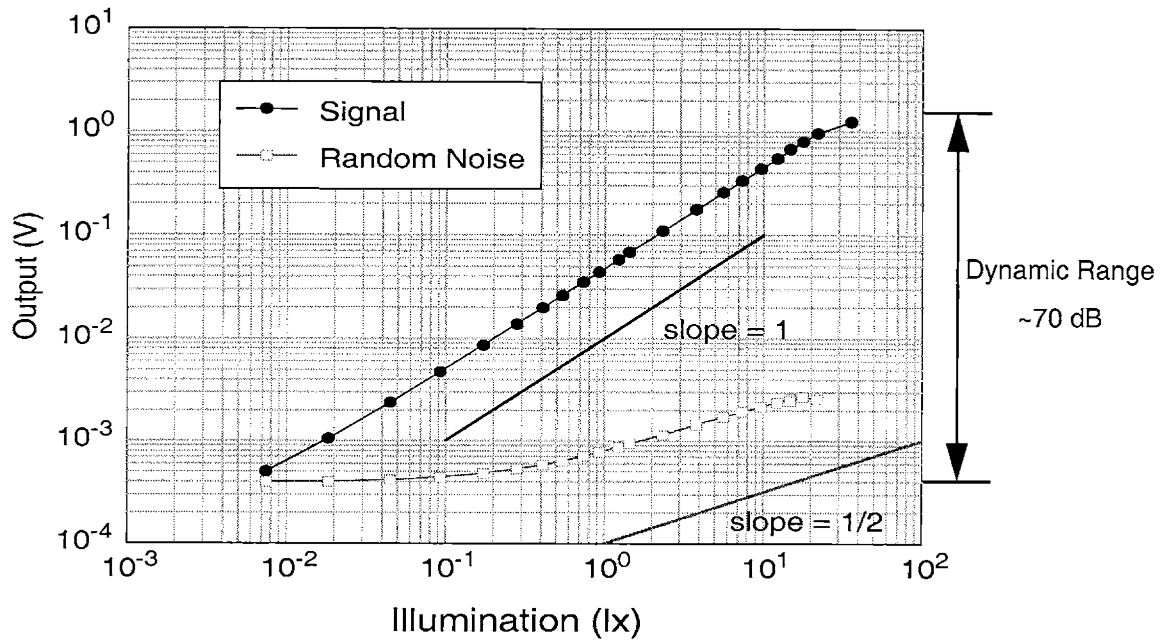


Figure 4.19: Sensitivity and random noise versus illumination for the NW/Psub photodiode APS

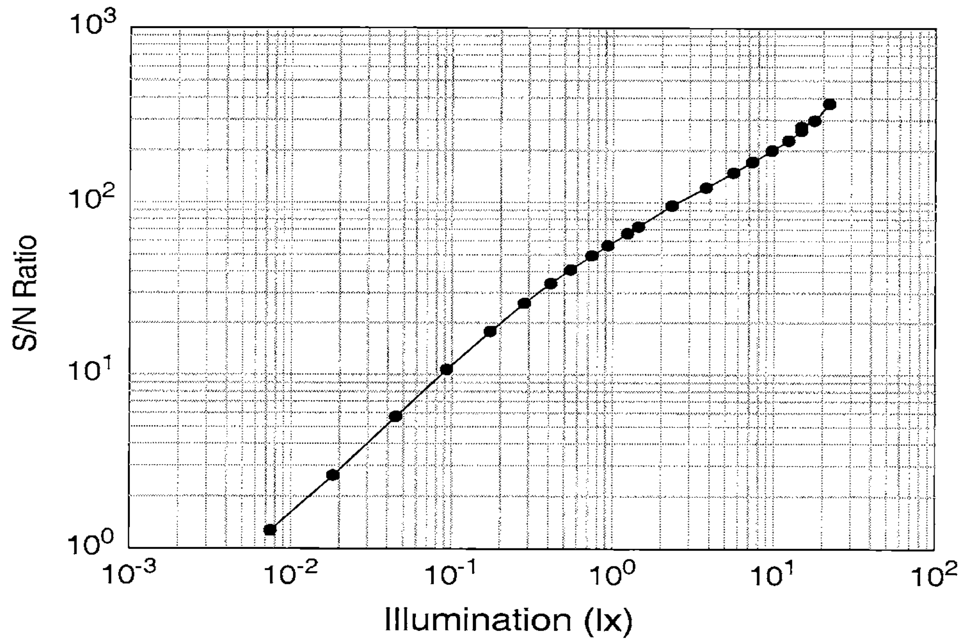


Figure 4.20: S/N ratio versus illumination for the NW/Psub photodiode APS

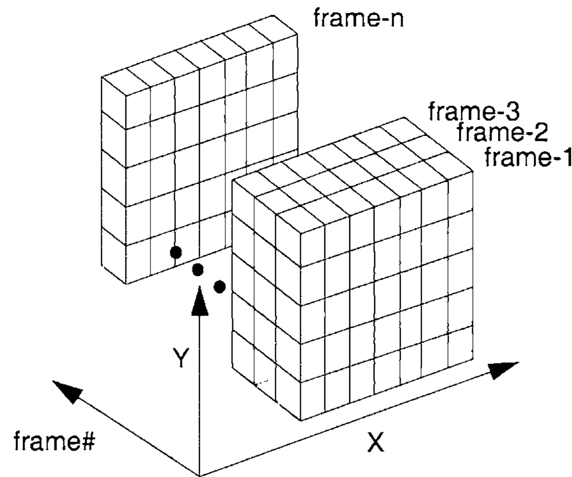


Figure 4.21: Data set for processing random noise and fixed pattern noise

4.4.5 Noise

Noise is a critical performance parameter for all sensors. For analog sensors such as CMOS imagers, any noise generated at the charge sensing node will be shown on the display unless signal processing technique, e.g. frame averaging, is performed. Basically, the noise components for CMOS imagers can be classified into two categories — random noise and fixed pattern noise. The random noise describes the temporal variation of the output signal; whereas the fixed pattern noise described the spatial variation of the 2-D imager plane as an uniform illumination is applied. The former is purely probabilistic and the latter is deterministic. Due to their difference in property, one component can be distinguished from the other with the following mathematical approach.

First, many frames of signal is taken under identical measurement conditions and stored in a data buffer. This data buffer can be viewed as a three-dimensional data set as shown in Figure 4.21. Two of the axes correspond to x-/y- location of the pixels and the third dimension corresponds to frame number. Mathematically, this data set can be denoted as $S(x, y, f)$, where x and y represent column and row numbers; f is the frame number.

Random noise represents the fluctuation of the output signal from frame to frame. It can be calculated as the standard deviation for each pixel along the f-axis. Assuming 100 consecutive frame data is taken, the calculation is shown as below.

$$R(x, y) = \sigma^2 \cdot \{S(x, y, f)\}\big|_{f = 1\ldots100}, \tag{4.12}$$

where σ^2 denotes the mathematical operation of taking the variance of the data set in the parenthesis; the lower right corner shows that the operation is taken effect along *f*-axis. An average of the noise is taken throughout the 2-D array to represent the random noise of the pixel as shown below:

$$RN = \sqrt{M \cdot \{R(x, y)\}\big|_{x, y = 1\ldots64}}, \tag{4.13}$$

where M represents the mathematical operation of taking the average for the data set within the parenthesis. In this case, the average is done throughout the 2-D pixels.

To calculate the fixed pattern noise, the random noise component is removed by averaging pixel signals throughout the 100 frames as shown in the following equation:

$$F(x, y) = M \cdot \{S(x, y, f)\}\big|_{f = 1\ldots100}. \tag{4.14}$$

Assuming the random noise exhibits a Gaussian distribution, this averaging operation reduces the magnitude of random noise by $\sqrt{100}$ so the fixed pattern noise becomes the dominate noise source. The fixed pattern noise can further be classified as column-to-column fixed pattern noise (C-C FPN) and pixel-to-pixel fixed pattern noise (P-P FPN). They can be calculated from the following equations:

$$C\text{-}CFPN = \sqrt{\sigma^2 \cdot \{M \cdot \{F(x, y)\}\big|_{y = 1\ldots64}\}\Big|_{x = 1\ldots64}}, \text{ and} \tag{4.15}$$

$$P\text{-}PFPN = \sqrt{M \cdot \{\sigma^2 \cdot \{F(x, y)\}\big|_{y = 1\ldots64}\}\Big|_{x = 1\ldots64}}. \tag{4.16}$$

The classification of the two FPN components is related with the circuit architecture of the CMOS imager. This will be addressed in a latter section.

Random Noise

Random noise of CMOS image sensors is typically composed of photonic shot noise, reset noise (also known as KT/C noise), and dark current shot noise. Each noise component is originated from a specific mechanism. Therefore, these components can be considered independent to each other and the variance of total random noise voltage can be written as

$$V_{RN}^2 = V_{ph}^2 + V_{RESET}^2 + V_{D\text{-}S}^2, \tag{4.17}$$

where V_{ph}, V_{RESET}, and $V_{D\text{-}S}$ denote photonic shot noise, reset noise, and dark current shot noise respectively. The dominate component depends on the operating condition of the image sensor.

The photonic shot noise originates from the particle property of photons. Each photon emission and absorption is considered as an independent event and the process is a shot noise mechanism as described in Section 4.4.1. As an imager is operated under photonic-shot-noise-dominated region, the absorption of N carriers on average for a photo-sensor is accompanied with a standard deviation of $\sqrt{N}$ carriers. As a result, the photonic shot noise component can be formulated as

$$V_{ph} = CG \cdot \sqrt{N_{ph}} = CG \cdot \sqrt{\frac{I_{ph} \cdot T_{int}}{q}}, \tag{4.18}$$

where I_{ph} represents the average photo-current.

An imager operating under photonic-shot-noise dominate regime achieves the ultimate noise limit since the particle property of photon is nature. To achieve this, the illumination level should be high enough for V_{ph} to become the dominant noise source. The random noise versus illumination for the NW/Psub APS array is shown in Fig.4.18. As expected, the photonic shot noise has a square root relation with illumination at higher illumination. Conversely, under low light levels, the importance of photonic shot noise decreases and the other two components — reset noise and dark current shot noise become more important.

Reset noise (kT/C noise) comes from the reset operation at the sensing node of the pixel. When the sensing node is preset to a designated voltage (V_{rst}), there is a voltage uncertainty associated with the node after reset. This noise component is random and in fact, has the same origin as thermal noise on electronic devices. The equivalent circuit model for the reset operation is shown in Figure 4.22. It consists of a capacitor, a resistive thermal noise source, a perfect switch and a constant reset voltage source V_{rst}. Since we are only considering noise voltage here, the constant voltage source V_{rst} can be replaced with a short circuit to ground. The capacitor represents the sensing node capacitance; the

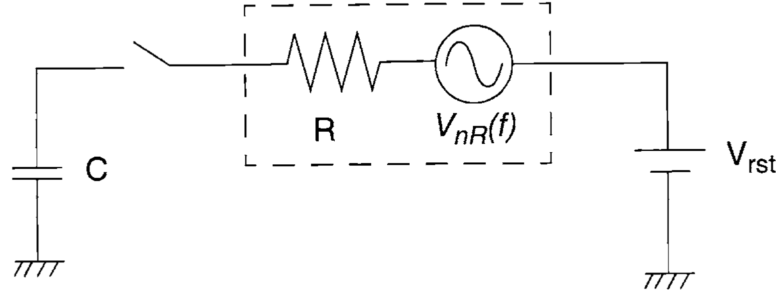


Figure 4.22: Thermal noise model

switch represents the reset transistor; and the resistor R contains the resistance from the voltage source to the sensing capacitor when the switch turns on. According to thermal noise theory [45], a resistor with value R under absolute temperature T has noise spectral density

$$V_{nR}^2(f) = 4k_B TR, \tag{4.19}$$

where k_B is Boltzmann's constant. The noise spectral density is proportional to R and T and the magnitude is uniform over an extremely wide frequency range (so called white noise). This frequency range can extend to as high as 10^{13}Hz [46]. Due to this ultra wide-band nature, the bandwidth of the thermal noise is usually assumed to be infinite so that the computation can be simplified. According to Figure 4.22, the voltage transfer function from the noise voltage to the sensing node can be written as

$$T(f) = \frac{1}{1 + j2\pi RCf}, \tag{4.20}$$

where j represents $\sqrt{-1}$.

At the moment as the reset pulse transits from "high" to "low", the instantaneous noise voltage on the sensing node is sampled. This instantaneous sampling can be modelled as a convolution operation of an impulse function $\delta(t)$ with the noise voltage at the sensing node in time domain. The computation can be easily conducted in frequency domain. As Fourier Transform is applied, the convolution operation becomes a multiplication and the noise voltage on the sensing node becomes

$$V_{nC}^2(f) = 1 \cdot T^2(f) \cdot V_{nR}^2(f). \tag{4.21}$$

The variance of noise voltage is an integration over the frequency domain. After simplification, the result is

$$V_{nC}^2 = \int_0^\infty V_{nC}^2(f) \cdot df = \int_0^\infty \frac{4k_B TR}{1 + 4\pi^2 R^2 C^2 f^2} \cdot df = \frac{k_B T}{C}. \tag{4.22}$$

The noise voltage on the sensing node only depends on temperature and its capacitance, not the resistance R. The independence of the resistance R can be explained as the follows. As R increases, the magnitude of the noise spectral density also increases but the effective bandwidth of the transfer function $T(f)$ decreases. As both effects are considered in the integration, they cancel with each other and end up with a constant noise voltage which is independent of the resistance R.

Theoretically, the reset noise can be completely removed by using a technique called correlated-double sampling (CDS) [47]. The idea of CDS is described as follows. Since the reset noise is due to the reset operation, the noise component would remain constant after the reset operation until the following reset is performed. If the sampling pulses SR and SS can be arranged in the same reset cycle, the reset noise component will be 100% correlated for the V_{ref} and V_{sig}. The subtraction of both voltages will completely remove the reset noise. This technique is widely adopted in CCD readout circuit to remove the reset noise. Unfortunately this technique is not easily implemented for a 3-T photodiode APS. That is because the V_{ref} and V_{sig} correspond to different reset cycles (refer to Figure 4.3). The noise voltage becomes uncorrelated after each reset operation. Unless a frame buffer is utilized to store the V_{ref} for each pixel, the reset noise is unable to be removed. In fact, the subtraction of V_{ref} and V_{sig} in an uncorrelated case produces reset noise $\sqrt{\frac{2k_B T}{C}}$, which is $\sqrt{2}$ times the derived kT/C noise.

The above derivation of reset noise applies to the case that V_{rst} is lower than Vdd by at least the threshold voltage V_{th}. Under such condition, the sensing node can be completely reset to V_{rst} when the reset transistor turns on. This reset operation is called "hard reset". Whereas in another popular reset scheme, V_{rst} is shorted with Vdd, causing the reset transistor Q1 (refer to Figure 4.14) to operate in sub-threshold regime during reset. The sensing node voltage will not be pulled up completely to Vdd since the current flowing through the reset transistor for charging up the sensing node decreases exponentially with

the increases of the sensing node voltage. This type of reset operation is called “soft reset”. The reset noise from soft reset is theoretically proven to be smaller than the reset noise from the hard reset [48]. That is because when the reset transistor operates in sub-threshold regime, the mentioned noise mechanism is no longer valid since the conducting channel is replaced by a potential barrier. Thus, the noise model is more similar to a shot noise mechanism since the carriers need to pass the potential carrier to travel between source and drain. According to the derivation result in [48], the variance of the reset noise under soft reset can be written as

$$\sigma_{nC}^2 = \frac{mk_BT}{2C}, \tag{4.23}$$

where m is the non-ideality factor in the subthreshold I-V relation: $I = \exp\left(\frac{q \cdot (V_{gs} - V_{th})}{mk_BT}\right)$. Typically m is around unity. Therefore, the noise associated with the soft reset would be roughly half of the noise value in the hard reset case.

Considering the three reset schemes — CDS, hard reset, and soft reset, a general reset noise value can be written as

$$V_{RESET} = \alpha \cdot \sqrt{\frac{2k_BT}{C}}, \tag{4.24}$$

where α depends on how the reset operation is performed. If an ideal CDS is performed, α is zero. Otherwise, α is equal to unity for the hard reset scheme and less than one for the soft reset scheme.

The dark current shot noise is due to the temporal randomness of dark carrier generation from the G/R centers. This carrier generation process is usually modelled as a Poisson process, which demonstrates shot noise property similar to the photonic shot noise. Consequently, as an average of N dark carriers are generated and collected in a specific integration time, there would be a randomness with a standard deviation of $\sqrt{N}$ carriers associated with it. The dark current shot noise can be written as

$$V_{\text{D-S}} = CG \cdot \sqrt{N_{dark}} = CG \cdot \sqrt{\frac{I_{dark} \cdot T_{int}}{q}}. \tag{4.25}$$

Thus, it is desired to minimize the total dark current in order to reduce the dark current shot noise.

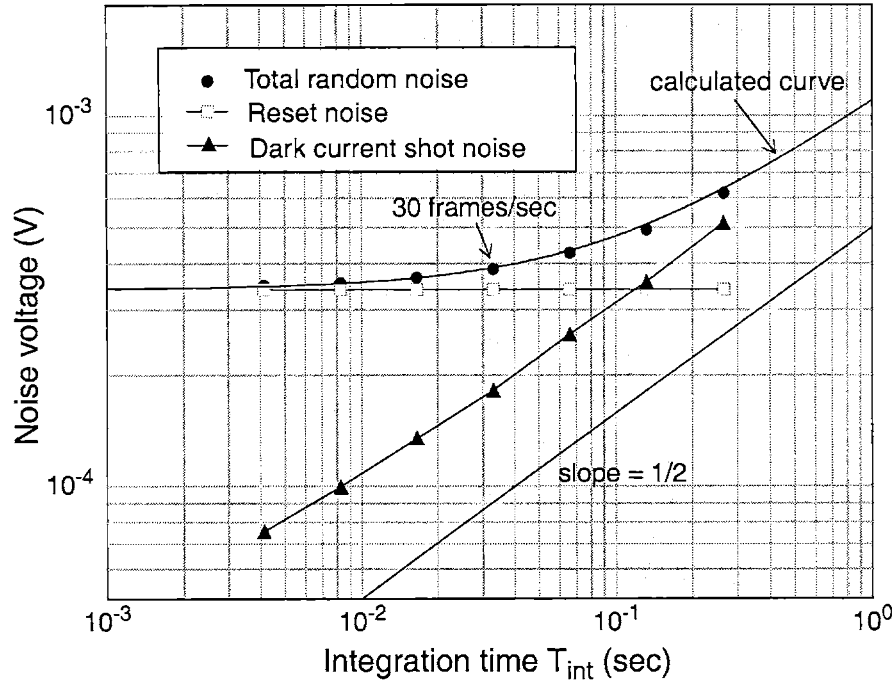


Figure 4.23: Random noise components versus integration time for the NW/Psub photodiode APS

At low illumination, the reset noise and dark current shot noise become dominant. Since their values are independent of illumination, the noise value at low illumination is roughly constant (Figure 4.19). These two noise components can be separated by changing the integration time and measure the random noise in a dark environment. That is because the reset noise is independent of the integration time but the dark current shot noise increases with it according to eq.(4.25). The measurement results of the total random noise in a dark environment is shown on a log-log plot in Figure 4.23. The reset noise and dark current shot noise are extracted using eq.(4.17) with v_{ph} equal to zero. Both noise components are also shown on the same plot. The reset noise is constant and the dark current shot noise demonstrates a square-root relation with the integration time, proving the shot noise assumption is valid. The calculated curve in the figure represents the random noise calculated from the extracted reset noise and the total dark current using eq.(4.17) and eq.(4.25). The calculated curve matches the measured data quite well. The α value used in eq.(4.24) is less than unity and is calculated to be 0.48, which is a result due to the "soft reset". It is also observed from this figure that the reset noise is the dominate compo-

nent if the integration time is shorter than 0.1sec. For a longer integration time, due to the shot noise mechanism, the dark current shot noise increases with $\sqrt{T_{int}}$ and gradually becomes a major component of the random noise. At low illumination conditions, the dark current shot noise is the ultimate bottleneck of random noise since this noise cannot be removed with circuit techniques such as the CDS circuit to the reset noise.

Pixel-to-pixel Fixed Pattern Noise (P-P FPN)

P-P FPN is mainly due to the mismatch of sensor properties on pixel level. The mismatch is originated from process non-uniformity and is highly process- and layout-dependent. Mismatches on leakage current, optically-active area, sensing node capacitance, and the gain of the in-pixel source followers are all sources of the P-P FPN. In this section, P-P FPN is characterized by varying illumination and integration time to determine the major source of P-P FPN under different operating conditions. The numerical values of C-C FPN is also demonstrated on the plot for comparison.

The FPN characterization is performed on the NW/Psub APS array. Uniform illumination with varying intensity is applied on the imager plane and the P-P FPN and C-C FPN are measured and presented in Figure 4.24. As shown in the graph, the P-P FPN stays constant as illumination is lower than 1 lux; whereas at higher illumination it increases linearly with the intensity. The integration time is 1/30 second. The linear behavior at high illumination can be explained by the mismatch of the optically-active area, the sensing-node capacitance, and the gain of the in-pixel source followers. These mismatches produce mismatches proportional to the number of the injected photons. A ratio is calculated between the P-P FPN and the output signal at higher illumination. The P-P FPN is roughly 0.6% of the total output signal.

On the other hand, the nearly constant P-P FPN behavior at low illumination suggests that there is another P-P FPN mechanism which has no illumination dependence. This mechanism is identified to be the leakage current mismatch. It is verified by measuring the P-P FPN in a dark environment versus integration time. The measurement result is shown in Figure 4.25. The P-P FPN presents a linear dependence on integration time, which can be explained by a non-uniform distribution of the G/R centers over the pixels. A detailed observation of the leakage current mismatch is shown in Figure 4.26. The three curves on

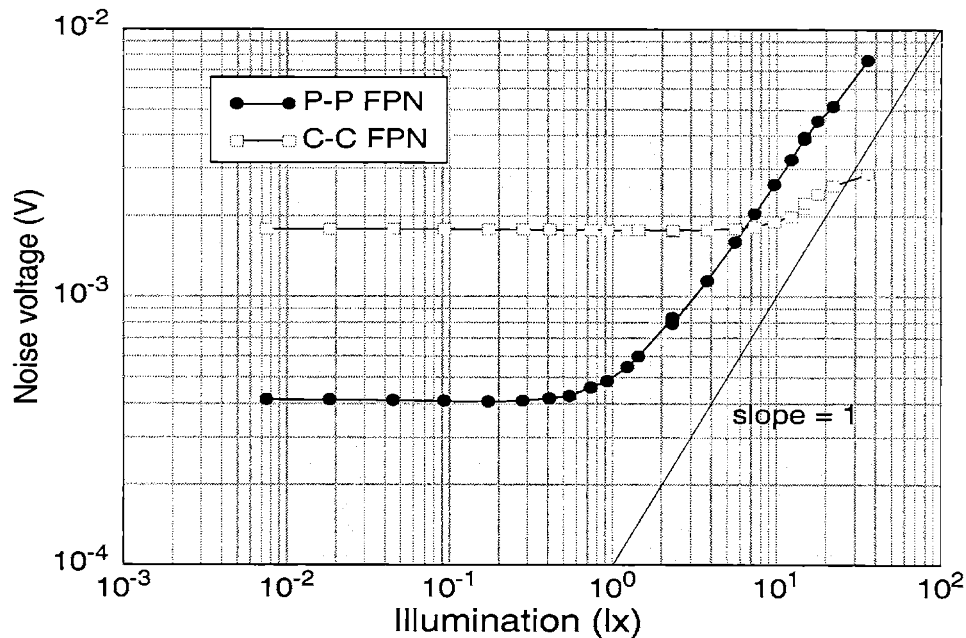


Figure 4.24: P-P FPN and C-C FPN versus illumination

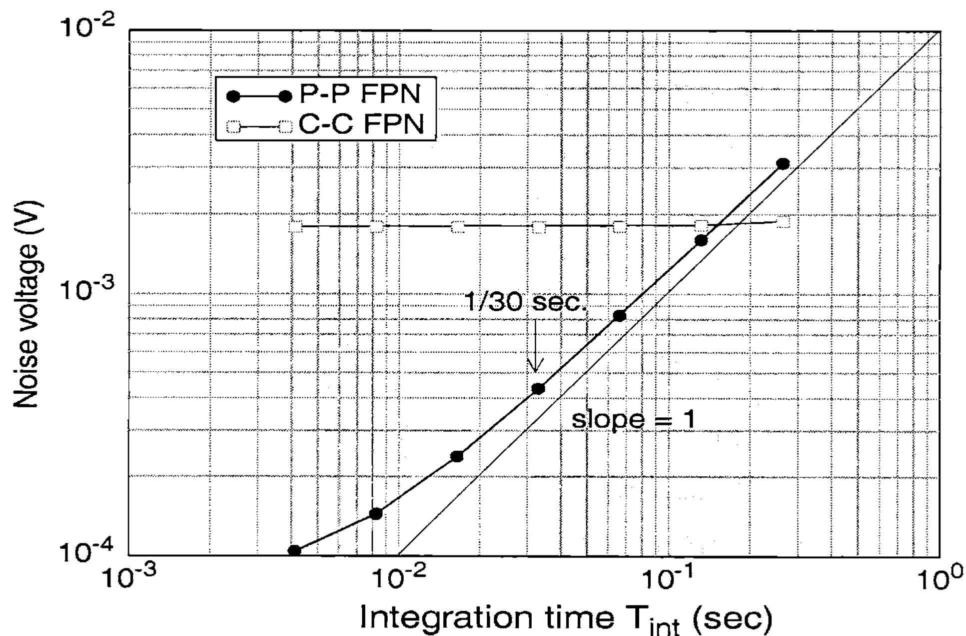


Figure 4.25: FPN versus integration time

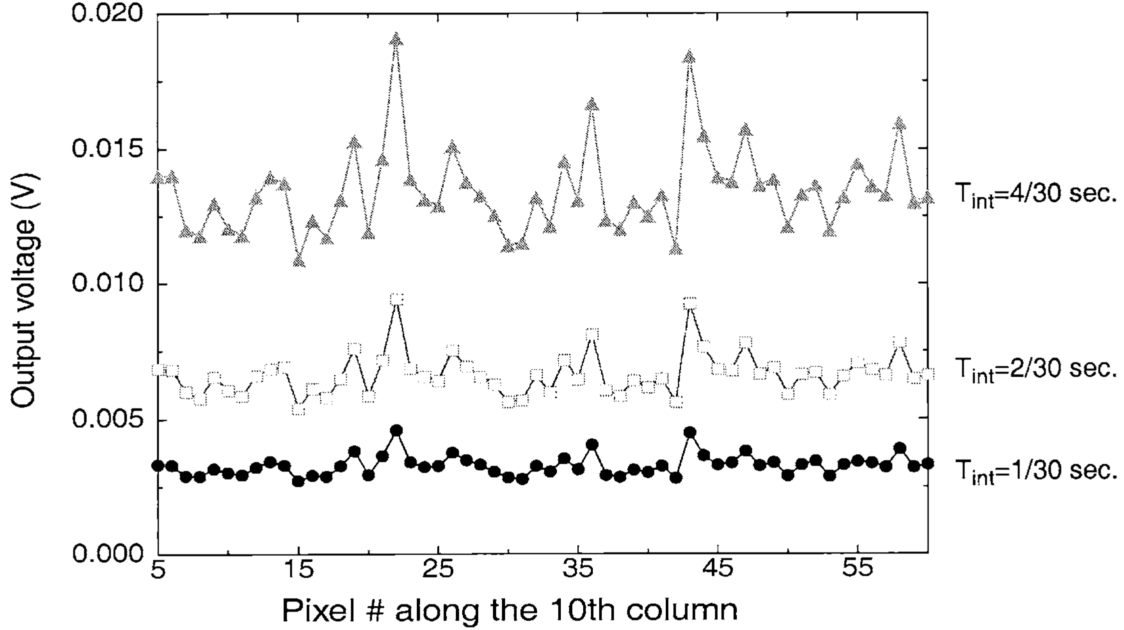


Figure 4.26: Pixel output voltage versus row number with different T_{int}

the plot correspond to dark voltage measurement along a specific column with integration time equal to 1/30, 2/30 and 4/30 second. The figure shows that not only the mean value of the dark signal increases proportionally to T_{int}; the voltage discrepancy between pixels does too. This result is a direct proof that P-P PFN is dominated by leakage current mismatch at low illumination. For the NW/Psub APS array, the P-P FPN under low illumination is 0.4mV as T_{int} is equal to 1/30 second. Combining the P-P FPN characterized at both high and low illumination conditions, a general P-P FPN can be calculated from $V_{\text{P-P FPN}} = \sqrt{(0.006 \cdot V_{out})^2 + (0.4\text{mV})^2}$.

If the P-P FPN caused by dark current mismatch is severe, some pixels always present higher output than the others and show white spots on the display. These pixels are usually called "white pixels". The existence of the white pixels is also related with Si process conditions and should be eliminated. A characterization technique is needed to quantify the number of white pixels and provide an index for process improvement. The P-P FPN described above provides a rough idea about the distribution of dark signal but it does not

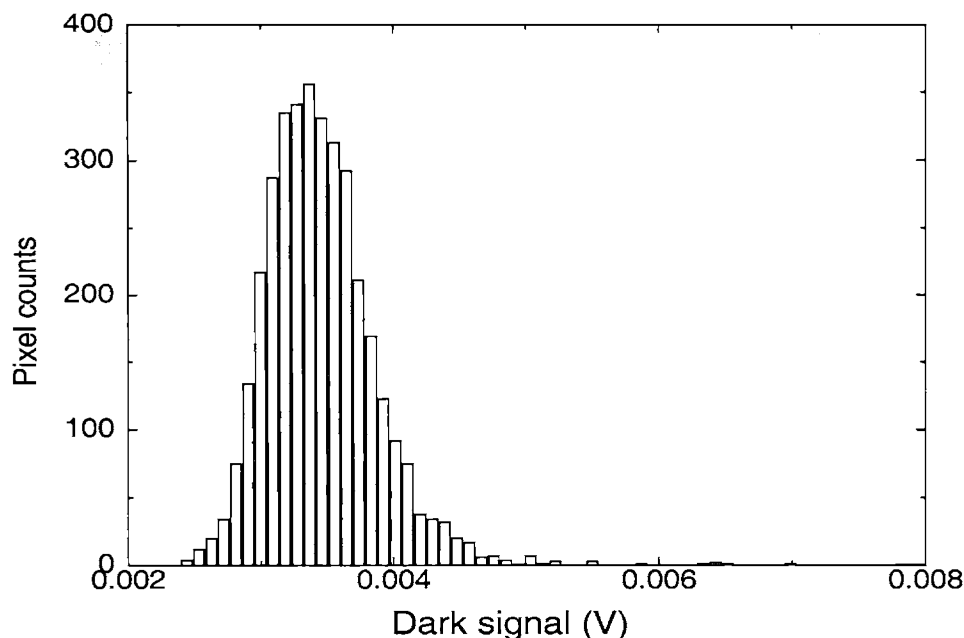


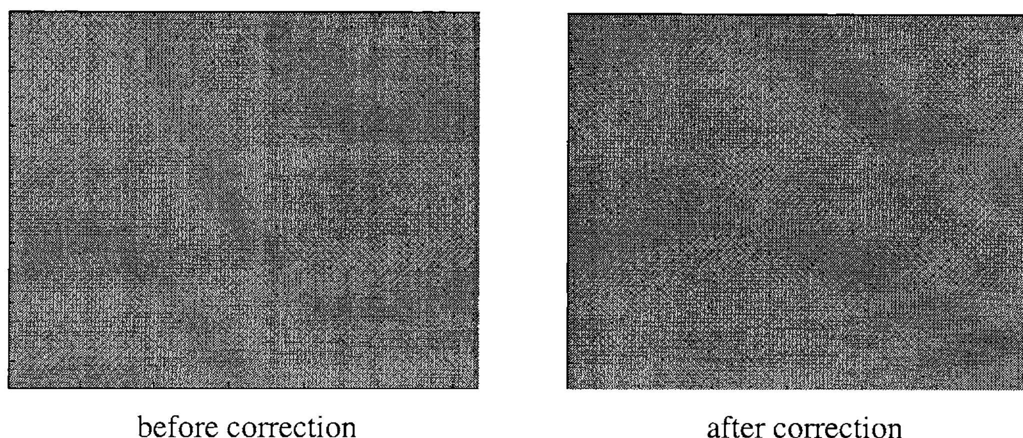
Figure 4.27: Histogram of the dark signal distribution

provide specific information about white pixels. To characterize the density of white pixels, a histogram of dark signal is a handy way for presentation.

Figure 4.27 shows the histogram of the dark signal for the NW/Psub APS array with integration time 1/30 sec. The mean value of dark signal is calculated to be 3.5mV and the standard deviation 0.4mV. A dark signal threshold can be defined with the mean value and standard deviation. Pixels with dark signal higher than this threshold can be classified as white pixels. For example, two times the mean dark signal is defined as the white-pixel threshold. Under such definition, one white pixel out of 64 x 64 is contained in this pixel array. In a more sophisticated characterization, the white pixels can be categorized according to the severity of dark signal.

Column-to-column Fixed Pattern Noise (C-C FPN)

C-C FPN is mainly due to the characteristic mismatch of the column processing circuitry, which usually consists of voltage/current buffers, CDS circuits, analog-to-digital converters, or a combination of the above. Mismatches on any above circuits can produce either voltage/current offset or gain discrepancies between column circuits. Without mismatch



before correction after correction

Figure 4.28: Image before and after C-C FPN correction. Illumination: 0.2 lux; gain:12

correction some of the columns are always brighter or darker than the others and thus form a stripe pattern. Since human eyes are quite sensitive to stripe patterns, the tolerance for C-C FPN is small and a correction is generally required.

For the 64 x 64 NW/Psub APS array, the C-C FPN is due to the offset voltage mismatch of the two second-stage source followers in the column circuit as shown in Figure 4.2. The construction of the second-stage source followers is essential since the driving ability of a sole APS is insufficient to steer the analog signal through the multiplexer and settle the output voltage in time. For column-parallel analog amplifiers or analog-to-digital converters, the mismatch on device characteristics produces either offset error or gain error from column-to-column.

There are several approaches to correct the C-C FPN. The first one is to calibrate the mismatch while the imager is powered on and store the mismatch value in a line memory. The mismatch value is then corrected. The second approach is to design a FPN correction circuit built-in the processing circuitry so that the FPN can be sampled and then subtracted dynamically. The second method is better than the first one since it does not suffer from the variation of operating conditions such as temperature drift. In the 64 x 64 APS arrays, the second approach is adopted. The operating principle of the C-C FPN correction circuit is presented in Section 4.1. An example of the 64 x 64 images before and after C-C FPN correction is presented in Figure 4.28. The stripe pattern can be observed at low illumina-

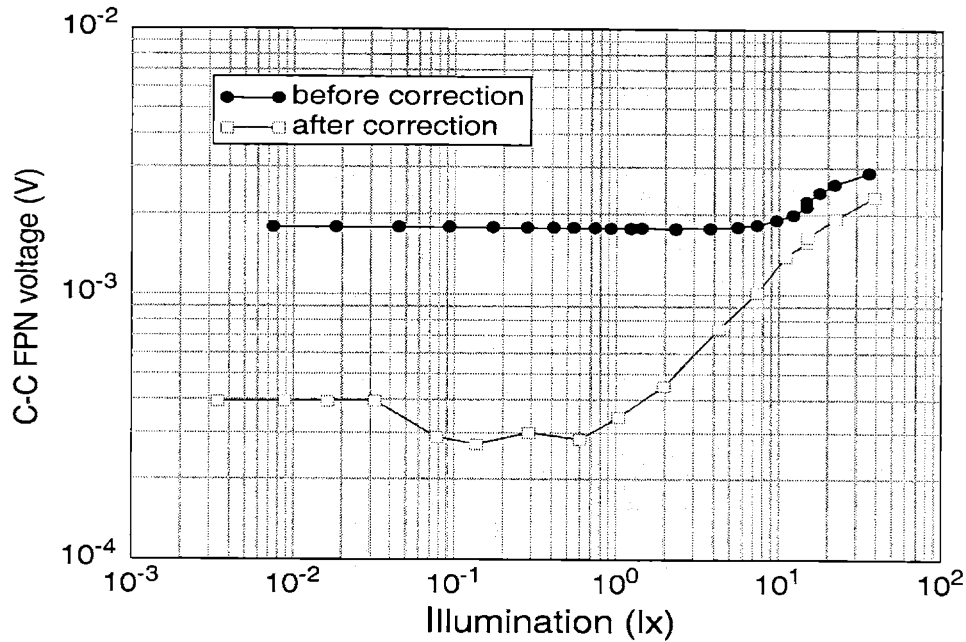


Figure 4.29: C-C FPN measurement before and after C-C FPN reduction circuit

tion (0.2 lux) with an amplifier gain of 12. After the correction, the image is more pleasant. The C-C FPN versus illumination before and after correction is also measured and shown in Figure 4.29. The integration time is 1/30 second. With the correction, the C-C FPN drops from 1.8mV down to 0.3mV at low illumination. For an imager with 1V signal swing, this correction improves the resolution from 9-bit to 12-bit. This plot demonstrates that the correction circuit successfully suppresses the C-C FPN. It is also observed from the figure that the C-C FPN remains constant for a certain range of illumination and finally increases at higher illumination. The increase of C-C FPN at high illumination is because the averaged column outputs are calculated from the 64 pixels in a column. The numerical significance of the averaged P-P FPN eventually surpasses the C-C FPN at high illumination. If a larger array with more columns is utilized to calculate the C-C FPN, the C-C FPN should remain constant for a wider illumination range.

Comparison of Random Noise and Fixed Pattern Noise

Though random noise and fixed pattern noise are caused by temporal and spatial variations respectively, it is still interesting to compare these two noise parameters to determine

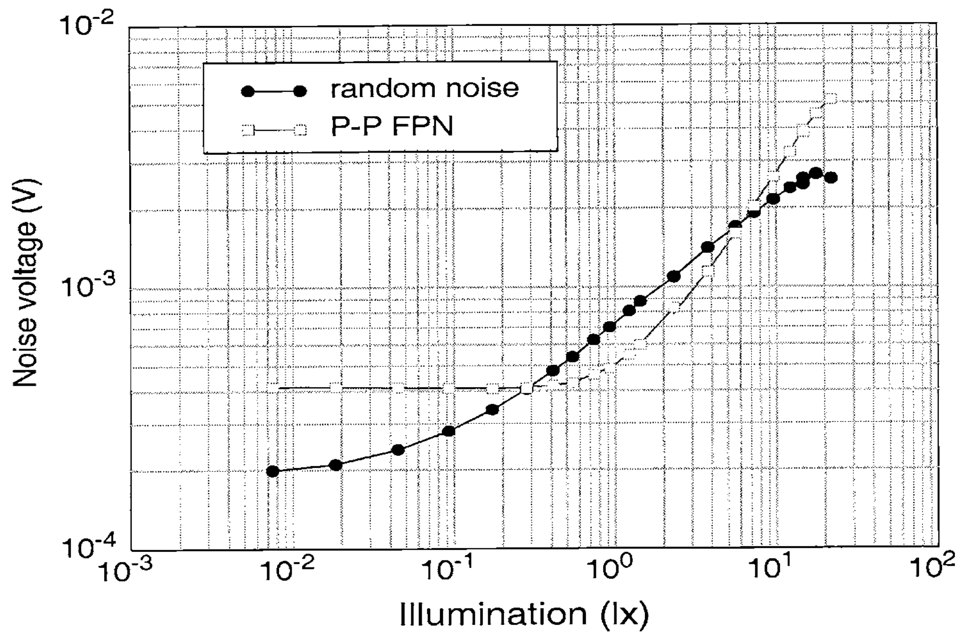


Figure 4.30: Random noise and P-P FPN versus illumination

the most significant noise source. In the comparison, it is assumed that a CDS circuit and a C-C FPN correction circuit are utilized so that there is no reset noise component in the random noise and no C-C FPN component in the FPN. In this way, the noise performance will be independent of any circuit related issues and the noise sources are all from the intrinsic process characteristics. Thus, the comparison presents an ultimate noise limit between random noise and P-P FPN.

The comparison between the random noise and the P-P FPN versus illumination is shown in Figure 4.30. The NW/Psub APS array remains the test device and the integration time utilized in the measurement is 1/30 second. It is observed that at low illumination (illumination < 0.3 lux) P-P FPN is greater than random noise. In this region the leakage current mismatch is the dominate noise source. Between 0.3 lux to 5 lux the photonic shot noise increases and the random noise surpasses the P-P FPN. Beyond 5 lux, the P-P FPN induced by optically-active area mismatch, the sensing-node capacitance mismatch, and the gain mismatch of the in-pixel source followers take over the random noise and become dominate. Since the quality of an imager is typically determined by the low-illumination

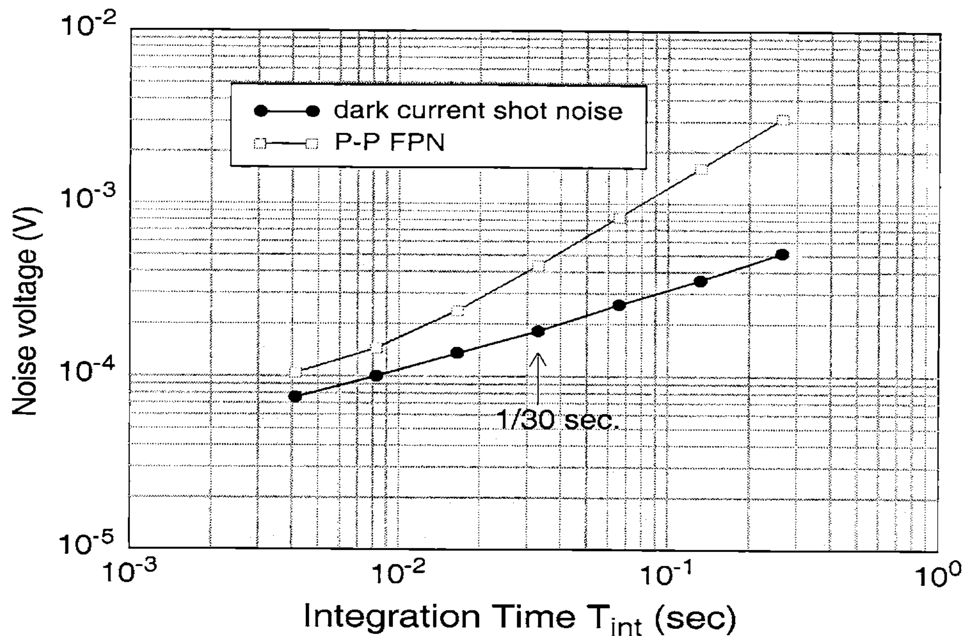


Figure 4.31: Dark current shot noise and P-P FPN versus integration time

performance, the leakage current mismatch is identified to be the ultimate performance hindrance for this APS array.

Since the P-P FPN induced by leakage current mismatch is proportional to integration time, the significance of P-P FPN may be mitigated by reducing the integration time. It is thus desired to compare the random noise and P-P FPN versus integration time in a dark environment. In such environment, random noise is solely determined by dark current shot noise. The comparison between dark current shot noise and P-P FPN versus integration time is shown in Figure 4.31. The P-P FPN shows higher magnitude than the dark current shot noise in the measured T_{int} ranges. Therefore, it is concluded that the dark current mismatch is the ultimate noise mechanism that determines the imager performance. The improvement of the dark current mismatch should have the top priority in the CMOS imager process development.

To conclude the discussion of noise in this section, the characteristics of all the noise sources are listed in Table 4.1. and Table 4.2.

Random noise source	Mechanism	Value	Dominate regime
Photonic shot noise	Poisson process characteristics for carrier absorption	$CG \cdot \sqrt{N_{ph}}$	High illumination
Reset (KT/C) noise	Brownian motion of electrons	$\alpha \cdot \sqrt{\frac{2kT}{C}}$	Low illumination
Dark current shot noise	Poisson process characteristics for dark carrier generation	$CG \cdot \sqrt{N_{dark}}$	Low illumination

Table 4.1: Characteristics of random noise sources

FPN noise source	Mechanism	Value	Dominate regime
P-P FPN	a. Leakage current mismatch b. Capacitance mismatch, optically-active area mismatch, or gain mismatch of the in-pixel source followers	a. Mismatch of $CG \cdot \frac{I_{\text{dark}} \cdot T_{int}}{q}$ b. Mismatch of $CG \cdot \frac{I_{\text{ph}} \cdot T_{int}}{q}$	a. Low illumination b. High illumination
C-C FPN	Mismatch of column-circuit characteristics	Depending on the utilized column circuit	Always existing

Table 4.2: Characteristics of fixed pattern noise sources

4.5 Comparison of Pixel Designs

The characterization approaches described previously are utilized to compare the pixel performance between different pixel designs. The 3-T NW/Psub APS array is used as a basic device to compare with a N^+/PW/Psub comb-shape photodiode APS array, a photogate APS array, and a 4-T NW/Psub photodiode APS array. The four arrays are on the same chip so that they went through exactly the same process conditions. The only difference is their pixel structures so the comparison is objective. The comparison highlights the advantages and disadvantages of the pixel structures fabricated with the standard 0.35-µm silicide CMOS process.

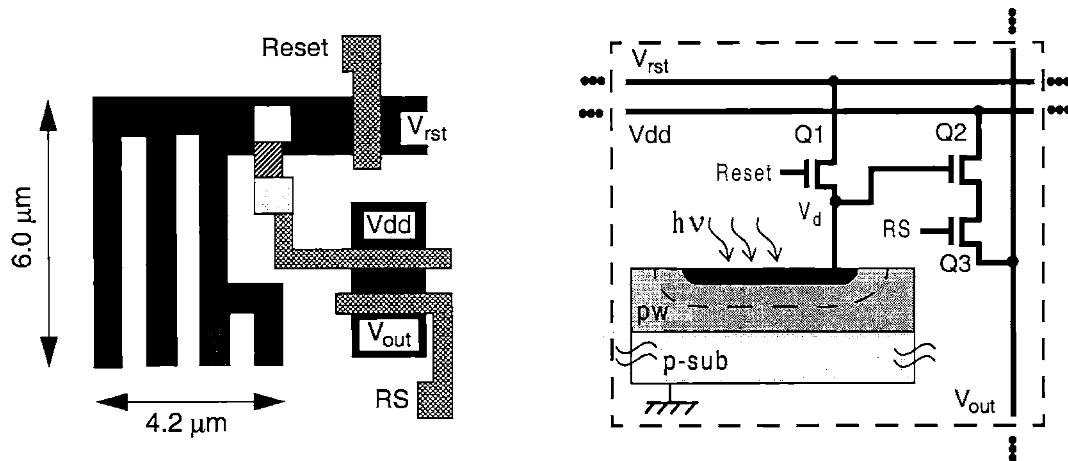


Figure 4.32: Layout and schematic of the N+/PW/Psub photodiode APS

4.5.1 NW/Psub vs. Comb-shape N+/PW/Psub Photodiode APS

For a standard CMOS process with silicide applied on source/drain/gate surface, the Q.E. of the N+/PW/Psub photodiode is very low since the silicide is an opaque material for visible light [22]. One way to get around the silicide problem without modifying the process is to either use a NW/Psub photodiode, which has no silicide on the surface, or lay out the N+/PW/Psub diode in a comb-shape or grid-shape layout to allow photons injecting through the field region. According to the result in Chapter 3, a grid-shape diode leaks more than 100 times than a square-shape diode. In this comparison, a comb-shape layout of the photodiode is utilized instead, trying to achieve a compromise between a high leakage current and a better Q.E.

The layout and schematic of the NW/Psub APS is shown in Figure 4.14 and the same plot for the N+/PW/Psub APS is shown in Figure 4.32. Pixel size for both structures is 9.2 x 9.2 μm^2. The layout of the three in-pixel transistors are identical. The only difference is the layout of the photodiode. From a perspective of junction capacitance, though the capacitance density of the N+/PW/Psub diode is higher, the effective diode area for the comb-shape diode is less than the rectangular diode used in the NW/Psub diode. Therefore, the discrepancy between the two sensing-node capacitances (or conversion gain) is expected to be marginal. On the other hand, the Q.E. of both structures should be on the same order of magnitude according to the characterization result in Chapter 3. The final

determining factor is still the leakage current. Table 4.3 lists the sensor parameter compar-

Photodiode	Capacitance/ conversion gain	Sensitivity	Dark signal @ 300K	Leakage current @ 300K
NW/Psub	16fF (10μV/e)	1.3 V/lx/sec	110 mV/sec	1.74 fA
comb-shape N^+/PW/Psub	22fF (7.3μV/e)	0.34 V/lx/sec	380 mV/sec	8.2 fA

Table 4.3: Comparison of NW/Psub and comb-shape N^+/PW/Psub photodiode APSs

ison between both structures. The five times higher leakage current in the comb-shape N^+/ PW/Psub pixel is the most significant disadvantage compared with the NW/Psub pixel. Besides high leakage current, the four times less sensitivity is another disadvantage. From the ratio of sensitivity and capacitance, it can be understood that the silicide on N^+-diffusion surface still block much of the photon flux and results in a lower Q.E. The only advantage of the N^+/PW/Psub pixel is the 50% higher charge capacity than the NW/Psub pixel. However, this advantage is unable to compensate for the low sensitivity and high leakage current.

4.5.2 3-T Photodiode APS vs. 4-T Photodiode APS

In a 3-T photodiode APS, the charge sensing node and the photo-carrier collection region are implemented with the same photodiode junction. This dual roles of the photodiode makes it difficult to improve the conversion gain by reducing the junction capacitor since the photo-carrier collection efficiency will also be reduced. This problem can be solved by placing one more transistor in the pixel [49]. Layout and schematic of the 4-T photodiode APS is shown in Figure 4.33. The pixel size is 10 x 10 μm^2, which is slightly larger than the 3-T APS due to the design rule restriction. The transistor Q4 separates the carrier-collecting area and the charge-sensing node. As the gate of Q4 is biased at a voltage lower than V_{rst} - V_{th}, this transistor is operated in cascode mode. Namely, the photo-carriers collected by the photodiode will be transferred to the charge sensing node at V_d. The photodiode voltage remains constant around V_{ref} - V_{th}, at which the channel just vanishes. The constant photodiode voltage is due to a negative feedback mechanism built in this struc-

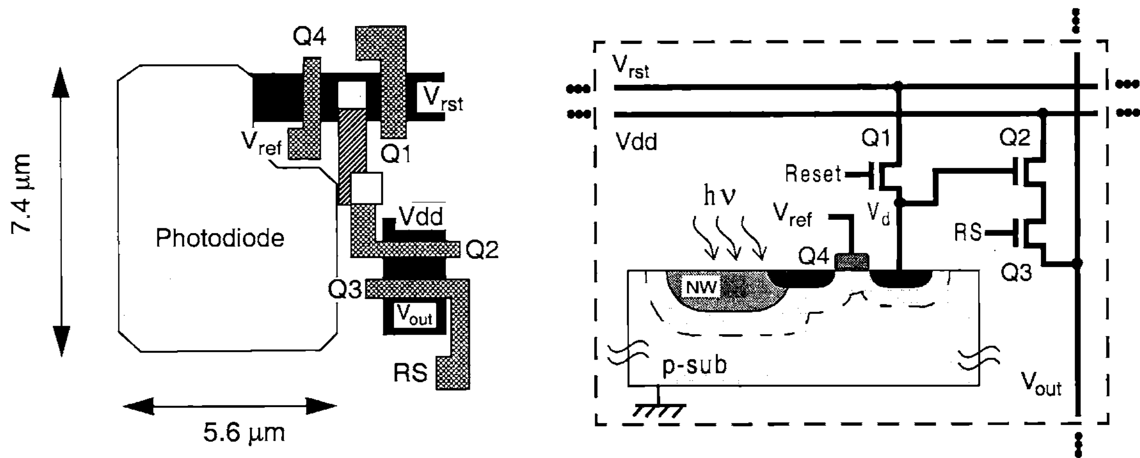


Figure 4.33: Layout and schematic of the 4-T photodiode APS

ture. Once the voltage of the photodiode is pulled down by photo-current or leakage current, the gate-to-source voltage of Q4 is increased and turns on the channel. Current will flow from the sensing node V_d to the photodiode to transport the integrated charge. This current continues to flow until the photodiode voltage returns to V_{ref} - V_{th} and then the channel turns off. In this way the carriers collected in the photodiode are transported to the sensing node, where the charge-to-voltage conversion is performed.

The separation of the photodiode and the charge sensing node allows to optimize the conversion gain without affecting the photo-carrier collection efficiency. This is advantageous in improving the S/N ratio of the sensor. The reason is as follows. The Q.E. for both pixels is expected to be similar since they utilize the same photodiode junction. Assuming the reset noise (KT/C noise) is the dominate random noise source (which is true at lower illumination), the random noise is roughly inversely proportional to the square root of the sensing node capacitance as described in eq.(4.24). Whereas the output signal voltage is proportional to the conversion gain and thus inversely proportional to the sensing node capacitance. Consequently, the S/N ratio becomes inversely proportional to the square root of the sensing node capacitance. This result suggests that the 4-T APS with a smaller sensing node capacitance will have a better S/N than the 3-T APS. This is confirmed with the measurement results of the S/N ratios for both pixels as shown in Figure 4.34. As illumination increases, the random noise becomes dominated by photonic shot noise. Under

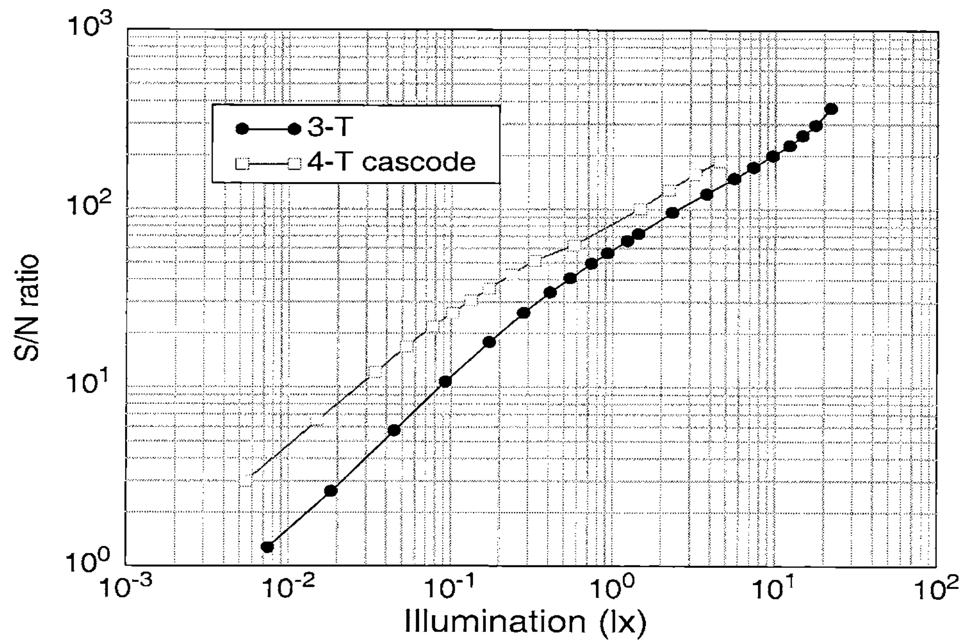


Figure 4.34: S/N ratio versus illumination for the 3-T photodiode APS and the 4-T cascode photodiode APS

APS type	Sensing-node capacitance (conversion gain)	Sensitivity	Dark signal @ 300K	Leakage current @ 300K
3-T	16fF (10μV/e)	1.3 V/lx/sec	110 mV/sec	1.74 fA
4-T cascode	7.2fF (22μV/e)	4.5 V/lx/sec	307 mV/sec	2.21 fA

Table 4.4: Comparison of 3-T and 4-T cascode NW/Psub photodiode APSs

such condition, both signal and noise have the same capacitance dependence and the two S/N curves tend to merge.

The characteristics for both APSs are listed in Table 4.4. As expected, the 4-T cascode APS has a lower sensing node capacitance, higher sensitivity, and higher dark signal due to the smaller sensing capacitance. The higher leakage current is in fact due to the larger photodiode area than the 3-T case. As the area ratio is considered, the leakage current den-

sity is roughly the same. The higher sensitivity ratio than the capacitance ratio is also due to the different size of the photodiodes.

In spite of the higher S/N ratio, however, utilizing a small sensing-node capacitance is not always advantageous. The drawbacks include a smaller dynamic range and a worse P-P FPN. The reduction in dynamic range is due to the increase of noise floor from a higher reset noise. As soft reset is applied in the characterization, the dynamic range is lower for the 4-T cascode APS (67 dB) than the 3-T APS (70 dB). On the other hand, the P-P FPN degrades due to the small sensing-node capacitance, which amplifies the leakage current mismatch by the capacitance ratio. As a result, the P-P FPN at low illumination increases from 0.4mV (3-T) to 0.9mV (4-T cascode). Whereas at high illumination, the P-P FPN for the 4-T APS stays the same as the 3-T APS at 0.6% of the output voltage. According to the above characterization results, the S/N ratio enhancement technique by reducing the sensing-node capacitance should be used wisely with the consideration of the dynamic range and P-P FPN degradation. The trade-off is listed in Table 4.5 with measured data.

APS type	S/N @ 0.1 lux	Dynamic range	P-P FPN
3-T	10	70 dB	0.4mV+0.6% V_{out}
4-T cascode	25	67 dB	0.9mV+0.6% V_{out}

Table 4.5: Trade-off between 3-T and 4-T cascode photodiode APSs

Another concern of the cascode configuration is image lag, which occurs because of incomplete charge transfer from the photodiode to the charge-sensing node. This is due to the slow response of the cascode transistor Q4 since it is typically operated in subthreshold regime. The slow response causes some residual photo-charges left behind in the photodiode region at the end of integration time. Those charges could contribute to the output of the following frames and thus produce image lag. The image lag is expected to become worse at low light levels since the transistor Q4 remains a longer time operating in subthreshold regime. Characterization of image lag requires an optical shutter with precise timing control and is beyond the scope of this study. Discussion on image lag of CMOS APS is available in reference [48, 50].

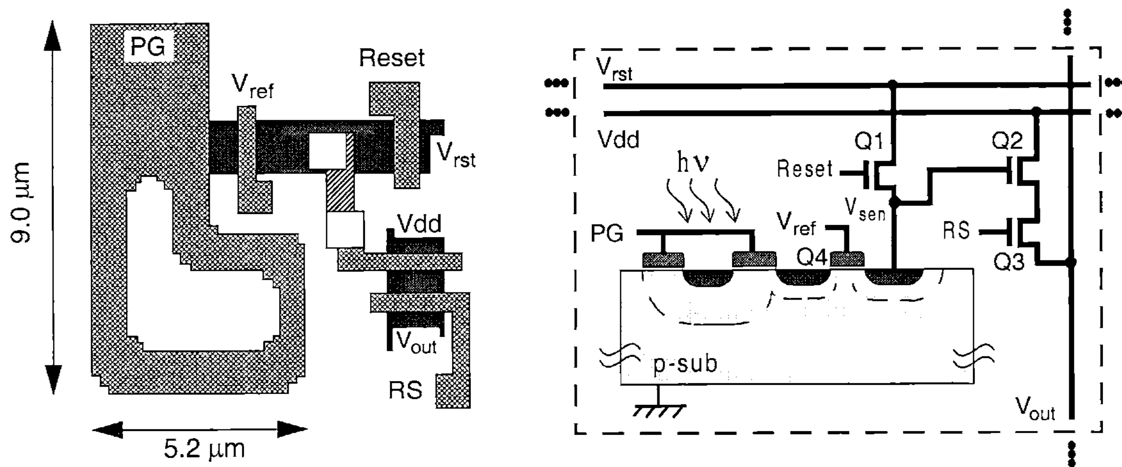


Figure 4.35: Layout and schematic of the photogate APS

4.5.3 Photodiode vs. Photogate APS

Photogate is another widely adopted APS structure in CMOS image sensors. Layout and schematic of a photogate APS is shown in Figure 4.35. The gate with an opened dielectric window is used to improve Q.E. because the gate is silicided. Compared with the typical photodiode APS, the advantage of photogate is the capability of implementing a real CDS to completely remove reset noise.

The reason why the photogate APS is able to implement CDS but the photodiode APS cannot is because the photogate can be used to temporarily hold the photo-charges without affecting the voltage of the charge sensing node. However, a typical 3-T photodiode APS is unable to achieve that because it shares the same photo-carrier collection site and the charge sensing node. The CDS operation of the photogate APS can be explained with the timing diagram shown in Figure 4.36. At the beginning of the integration time t_0, *PG* is pulsed to a high voltage to generate a depletion region on silicon surface for charge collection. Around the end of the integration time, the reset signal pulses at t_1 and sets the sensing node voltage V_{sen} to a high value. The voltage at V_{out} is then sampled onto the capacitor C_{ref} with a *SR* pulse at t_2. Following that, *PG* pulses low at t_3 to dump the photo-charge to the sensing node through the cascode transistor Q4. The voltage of V_{sen} drops due to the transported charges. *PG* then returns back “high” to start the following integra-

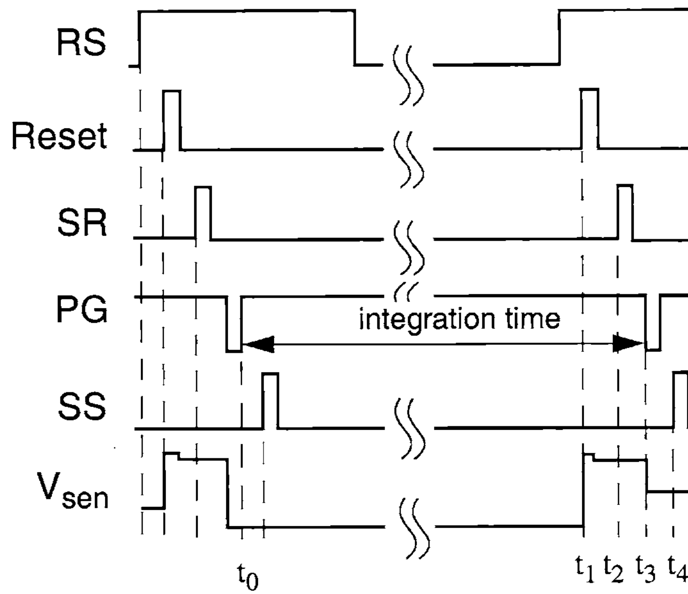


Figure 4.36: Timing diagram for the photogate APS

tion period. And the signal voltage on V_{out} is then sampled with *SS* onto the capacitor C_{sig} at t_4. This completes one sampling cycle. It can be observed that *SR* and *SS* are both activated within the same reset cycle so the voltage of C_{ref} and C_{sig} contain correlated reset noise. After subtraction, the reset noise is removed.

In spite of the ability to realize CDS, the photogate APS suffers from a low quantum efficiency and high leakage current. The reason for low Q.E. is because the photons need to transmit through the poly-gate in order to arrive at the photo-carrier collection area. Unfortunately poly-Si has a strong absorption at visible light similar to single crystal Si. The Q.E. is further degraded if there is silicide on top of the gate. In order to improve the Q.E., a dielectric window is opened on the gate to improve photon transmission. A comparison between the Q.E. of the NW/Psub photodiode and the photogate APS is shown in Figure 4.37. A degradation with a factor of 4 is observed for the photogate. If there is no dielectric window on the poly gate, the degradation is expected to become worse.

High leakage current is another problem of this photogate structure. The reason is because the gate edge is a highly stressed region. This stress is due to the LOCOS process. Therefore, a large amount of G/R centers reside there. According to the measurement result, the leakage current of the photogate APS is 16 fA/pixel/sec at room temperature,

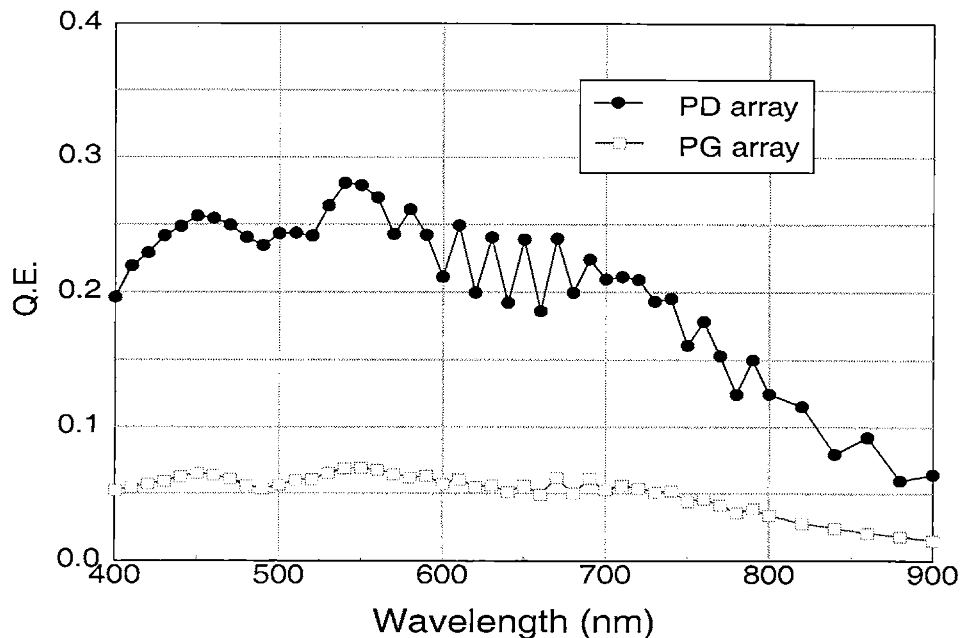


Figure 4.37: Q.E. comparison between the NW/Psub photodiode APS and the photogate APS

which is 10 times higher than that of the NW/Psub APS. This high leakage current is a more significant problem than the lower Q.E. since the dark current shot noise and the P-P FPN due to leakage current mismatch will all get worse by a large factor.

4.6 Comparison of Process Parameters

This section studies the effects of CMOS process parameters to the performance of photodiode APSs. The process variables include substrate types, photodiode junction profiles, and implant conditions. The utilized process is a 0.25-μm non-silicide source/drain shallow trench isolation (STI) CMOS process [19]. The pixel size is designed to be 4μm x 4μm to accommodate the pixel shrinking trend. The circuit structure is roughly the same as the one utilized in the 0.35-μm process described above. Modifications are made on the column and row circuits to fit in the 4 μm pixel pitch.

4.6.1 Substrate Types

Epi wafers and bulk wafers are two often used Si substrates. The major difference between these two types of wafers comes from the doping profile. For the utilized p-type epi

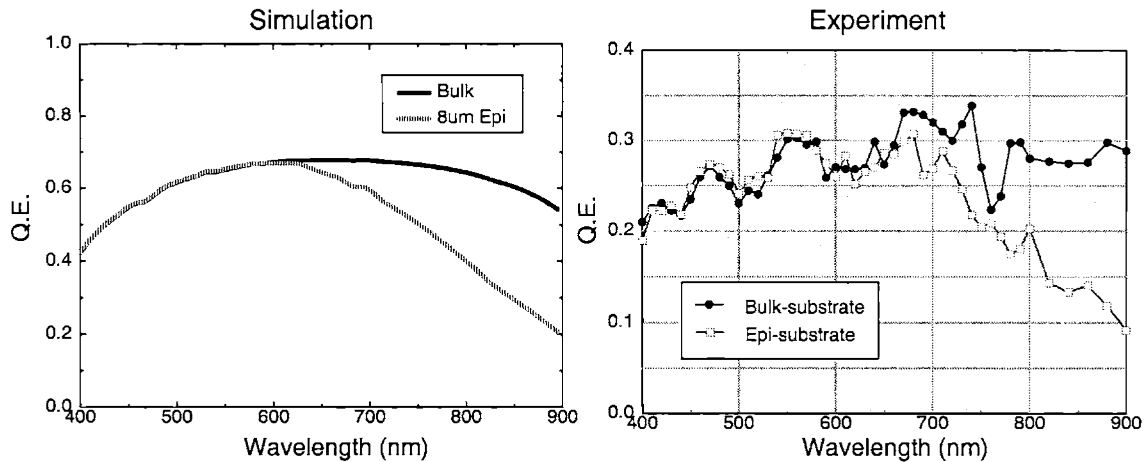


Figure 4.38: Simulation and experiment of Q.E. for the photodiode APSs on epi and bulk wafers

wafers, the thickness of the epi layer is 8-μm and the resistance in the epi layer is 8–12 ohm-cm, which corresponds to a doping level of roughly 10^{15} cm^{-3}. The doping level beyond the epi layer is roughly 10^{19} cm^{-3}. This high doping concentration in the substrate is utilized for preventing latch-up in digital circuits. For the p-type bulk wafers, the doping concentration is uniformly 10^{15} cm^{-3}.

The effects of the substrate type is mainly reflected on Q.E. of the photodiode since the minority carrier lifetime and internal electric field is a strong function of the doping profile. In fact, the doping profile difference only exists beyond 8μm in depth. It is expected that the absorption of short-wavelength photons should not be affected due to their shallow absorption depth in Si. A 2-D simulation is performed to compare the Q.E. of the NW/Psub photodiode APSs using the epi and bulk wafers. The spreading resistance profile (SRP) from the foundry is utilized as an input to the simulation program. The Q.E. measurement is also performed on both photodiode APSs. Both simulation and measurement results are shown in Figure 4.38. The scales for both figures are different because fill factor is not considered in the 2-D Q.E. simulation. Both the simulation and measurement results show that the Q.E. for both substrate types are almost identical for wavelength from 400nm to 650nm. Beyond 650nm, the Q.E. of the bulk wafer becomes larger than that of the epi wafer. This discrepancy is due to the short carrier lifetime in the highly

doped substrate underneath the epi layer. This can be confirmed by turning on and off the doping-dependent lifetime model in the simulation.

Nominally a higher Q.E. is favored since it means a higher output signal and better S/N ratio. However, the longer minority carrier lifetime in the substrate could cause a higher probability for the photo-carriers to diffuse to adjacent pixels and cause a higher cross-talk performance. The higher cross-talk not only reduces the imager resolution, it also impacts the performance of a color imager. The topic on cross-talk and its effects on color imagers will be described in Chapter 6. Therefore, from the consideration of cross-talk, epi wafers are better choices than the bulk wafers. On the other hand, the wavelength region where the bulk wafer exhibits higher Q.E. is almost beyond the visible light range. Thus, the Q.E. advantage for bulk wafers is not as significant for visible-light imagers.

4.6.2 Junction Types

Sensor performance of different junction types are compared using a 3-T photodiode APS structure. The junction types utilized in the comparison include N^+/PW/Psub, N^+/Psub, and NW/Psub photodiodes. These are the photodiode junctions available from the 0.25-μm CMOS processes without any process modifications. Compared with Chapter 3, this section conducts a more detailed investigation on the pixel-level. The compared parameters include sensing capacitance, sensitivity, leakage current, and Q.E.

For the sensing capacitance, the NW/Psub and N^+/Psub photodiodes utilize the same p-type substrate as the lightly doped region and should have similar pixel capacitance. Whereas for the N^+/PW/Psub photodiode, the capacitance is expected to be higher due to the higher doping concentration in the P-well region. The capacitance ratio may be different from the value extracted from the characterization of large-area photodiodes. That is because parasitic capacitance such as gate-to-source coupling capacitor also contribute to the sensing capacitance. The significance of the parasitic capacitance is expected to become higher for a smaller pixel size due to the shrinkage of the photodiode.

Comparison of the sensor parameters is show in Table 4.6 As expected, the N^+/Psub and NW/Psub photodiode APSs have similar properties on capacitance and sensitivity; the N^+/PW/Psub photodiode APS has a higher capacitance and a lower sensitivity. The leakage current performance are different for these junctions due to the difference of field edge

Photodiode	Capacitance/ conversion gain	Sensitivity	Dark signal @ 300K	Leakage current @ 300K
N^+/PW/Psub	8fF (20μV/e)	0.64 V/lx/sec	130 mV/sec	1.0 fA
N^+/Psub	5fF (32μV/e)	1.7 V/lx/sec	100 mV/sec	0.5 fA
NW/Psub	5fF (32μV/e)	1.7 V/lx/sec	90 mV/sec	0.45 fA

Table 4.6: Comparison of the photodiode APSs with three different junction types

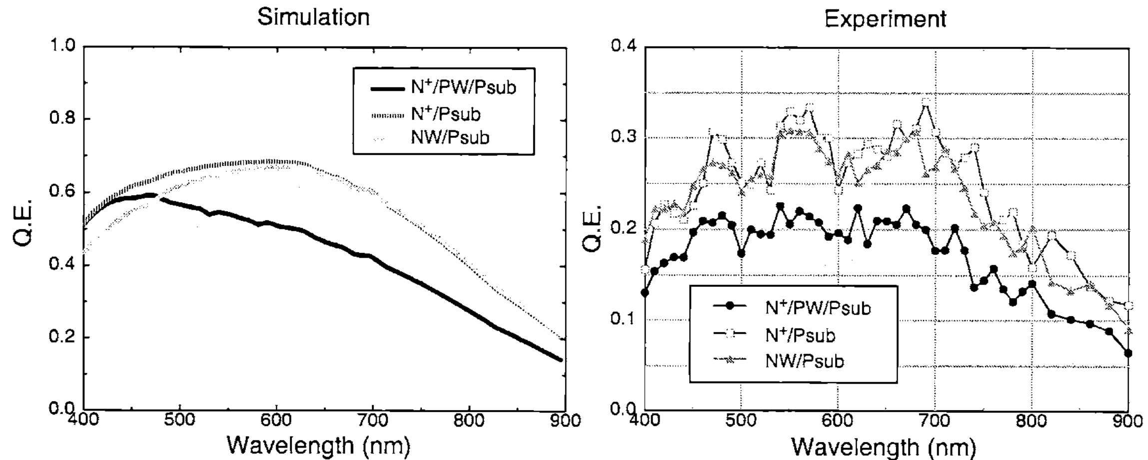


Figure 4.39: Simulation and measurement of Q.E. for the photodiode APSs with the three junction types

and junction properties. The leakage current from the field edge of the N^+-diffusion periphery is not significantly worse than the N-well structures. This could be due to the utilized STI structure [19], which produces less G/R centers than the LOCOS process utilized in the described 0.35-μm process. In general, the NW/Psub photodiode APS still hold a better sensitivity and leakage current performance than the other two.

The Q.E. of the three photodiode structures are simulated and measured. The results are presented in Figure 4.39. The N^+/Psub and NW/Psub photodiode APSs show similar Q.E. performance due to the similarity in doping profile at the lightly doped side. The N^+/PW/Psub photodiode APS, due to the narrower depletion width and the internal electric

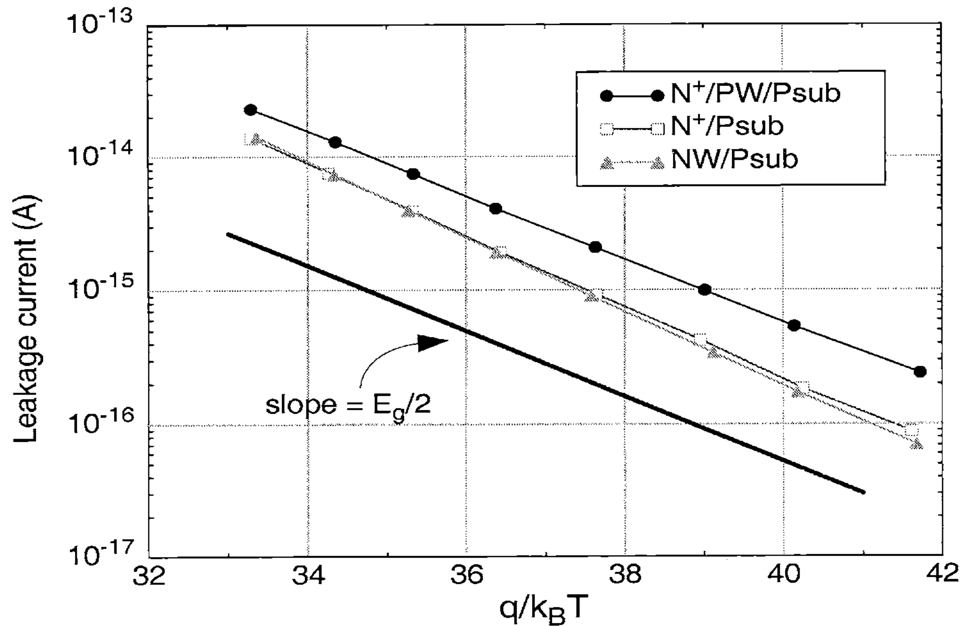


Figure 4.40: Arrhenius plot of leakage current for the photodiode APSs with the three junction types

field against carrier collection as described in Chapter 3, has a lower Q.E. By comparing both simulated and measured Q.E. at short wavelengths, it is noticed that the simulated Q.E. in this wavelength region is higher for the N^+/PW/Psub and N^+/Psub photodiode APSs. This suggests that the carrier recombination speed in the N^+ region should be higher than the parameter used in the simulation.

The Arrhenius plot of the leakage currents is shown in Figure 4.40. The activation energy of leakage current for N^+/PW/Psub, N^+/Psub, and NW/Psub photodiode APSs are 0.54 eV, 0.61 eV, and 0.64 eV respectively.

4.6.3 Implant Conditions

According to the above characterization results, the NW/Psub photodiode APS shows optimal performance than the other types of photodiodes. In order to further optimize the sensor performance, a specifically designed mask is used at the photodiode area to generate a specific doping profile for the sensor. This implant is performed right before the nominal N-well implant and has the same thermal treatment as the N-well implant. Implant dose and energy are the two control variables in this experiment. For the process

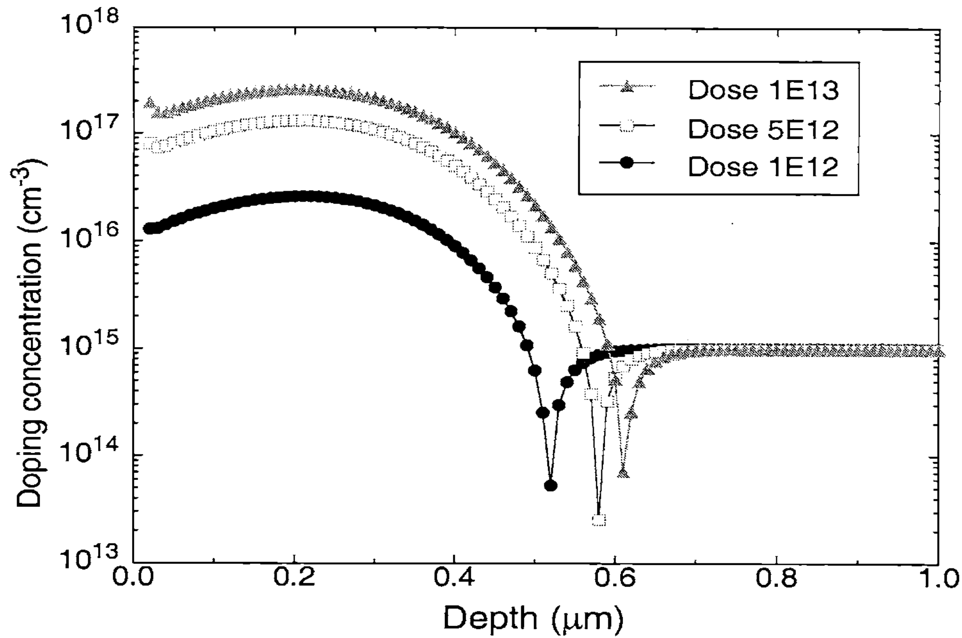


Figure 4.41: Simulated doping profile for the experiment of N-type implant dose split

split on dose, the dose conditions are 10^{12}, $5x10^{12}$, $10^{13}cm^{-2}$ and the implant energy is 460 KeV, the same as the N-well implant. The utilized dopant is phosphorus. Whereas for the implant energy split, the dose is $10^{13}cm^{-2}$ and the applied implant energies are 300, 380, and 460 KeV. The test APS arrays for the dose splits are fabricated on epi wafers and the ones for energy splits are on bulk wafers. The observed sensor parameter is mainly Q.E., which is compared with simulations and measurements. The measured leakage currents are also compared.

Implant Dose Split

In order to perform the Q.E. simulation, the doping profile is required. Since the SRP data for these special implant conditions are not available, an alternative approach is thus taken by conducting process simulation to estimate the profile. The utilized process simulator is *Tsuprem4* from Avant!. The simulated doping profiles after thermal treatment are shown in Figure 4.41. As shown on the plot, varying dose basically changes the magnitude of the N-type profile. The pattern remains the same. The junction depths also differ slightly (~ 0.2μm). The simulated profiles are then used to perform the Q.E. simulations. The simula-

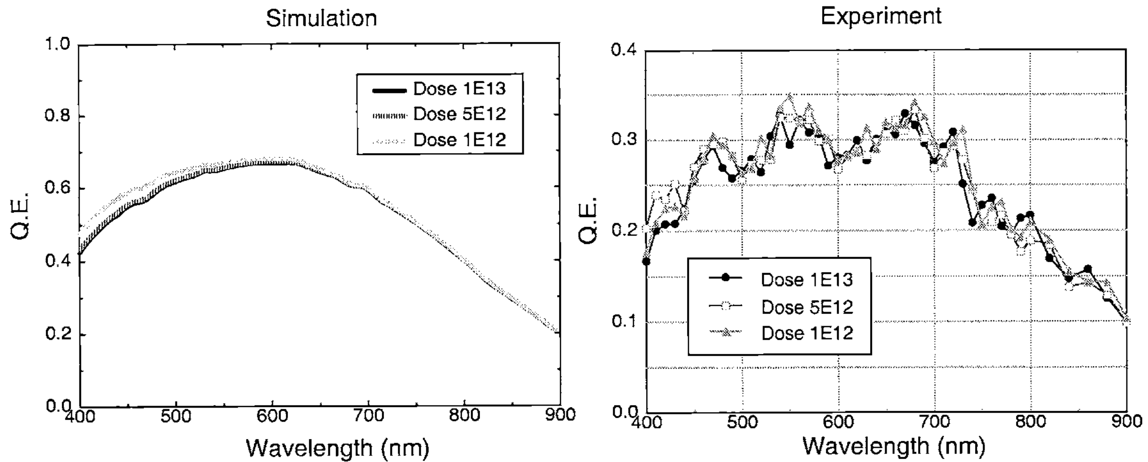


Figure 4.42: Simulation and measurement of Q.E. for the N-type implant dose split

tion and experimental results are shown in Figure 4.42. Similarity between the three dose conditions for both simulation and measurement suggests that Q.E. is not a strong function of implant dose for this type of photodiode sensor. This can be explained by the fact that the depletion region is determined by the lightly doped p-substrate region, which is identical for all the three test devices.

The dependence of leakage current on dose conditions is compared. If the implant dose has influence on the density of G/R centers, it will be reflected on the magnitude of leakage current and the slope of the Arrhenius plot. The Arrhenius plot for the leakage current measurement is shown in Figure 4.43. Similar to the Q.E. results, the dose has little impact on the leakage current. Thus, the density of G/R centers are roughly the same.

Implant Energy Split

An experiment on implant energy split is also performed to explore its influence on Q.E. and leakage current. The simulated junction profiles and Q.E. are shown in Figure 4.44 and Figure 4.45 Once again, both simulation and experiment prove that the Q.E. have little dependence on the doping profile in the more heavily doped side.

As for the leakage current, the temperature varying experiment is conducted and the Arrhenius plot is shown in Figure 4.46. The curves do show some implant energy dependence at higher temperature. However, the difference is not significant enough to draw

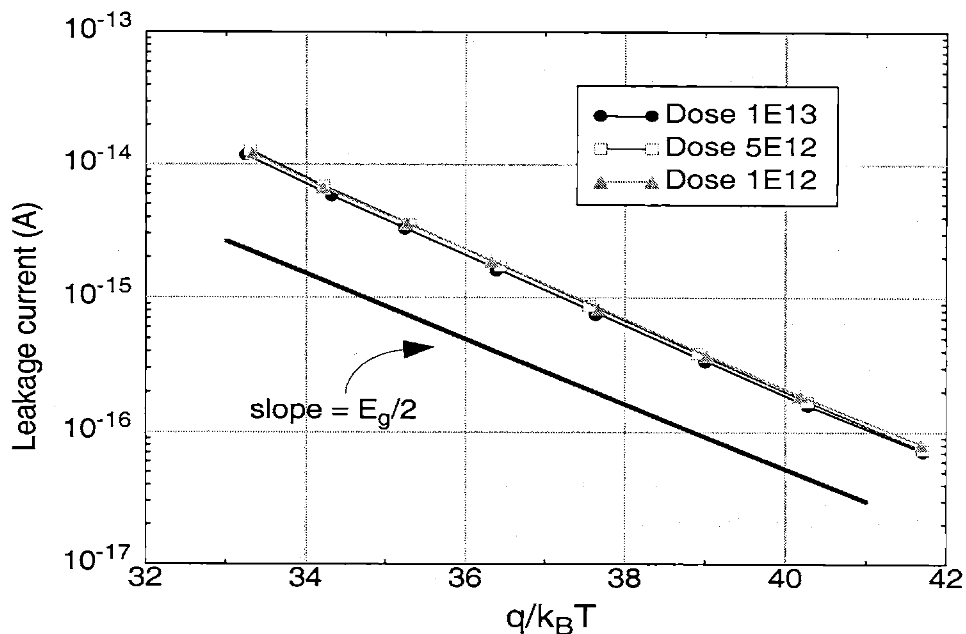


Figure 4.43: Arrhenius plot of leakage current for the N-type implant dose split

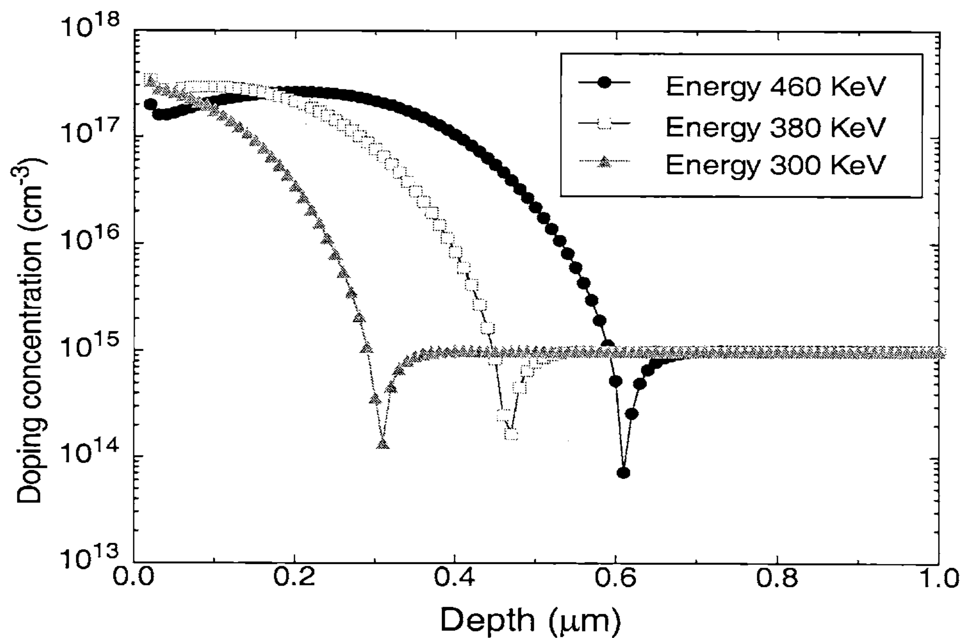


Figure 4.44: Simulated doping profile for the N-type implant energy split

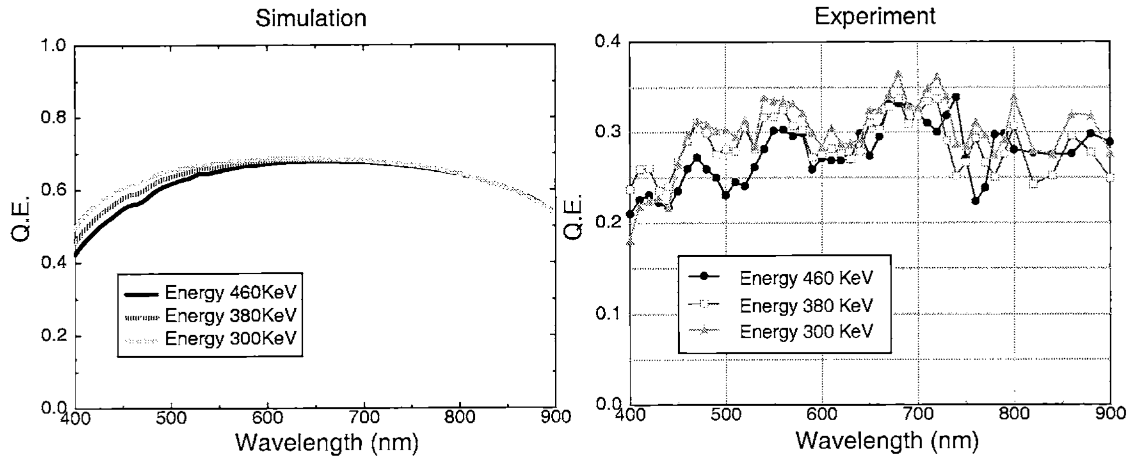


Figure 4.45: Simulation and measurement of Q.E. for the N-type implant energy split

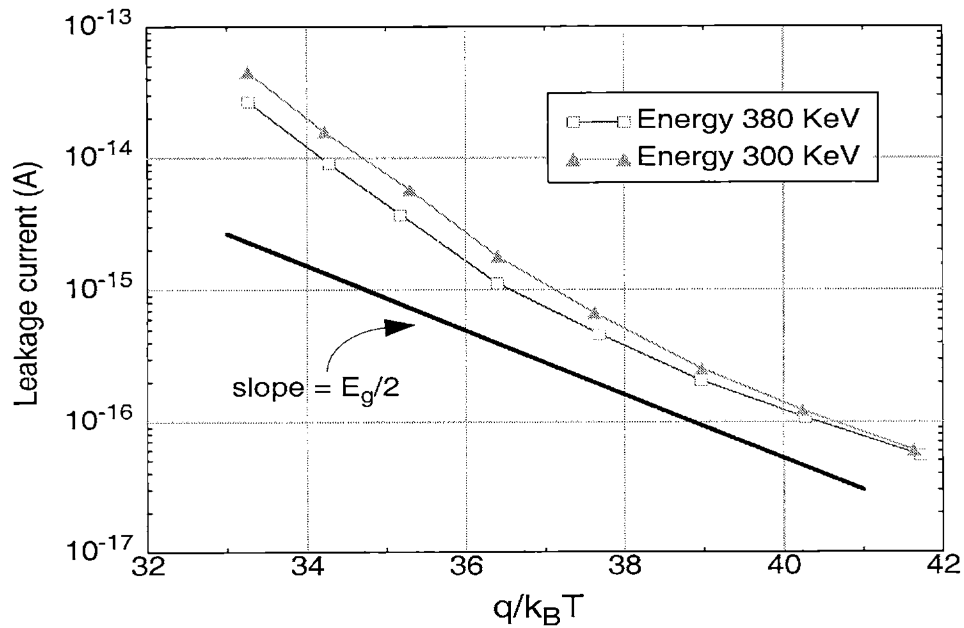
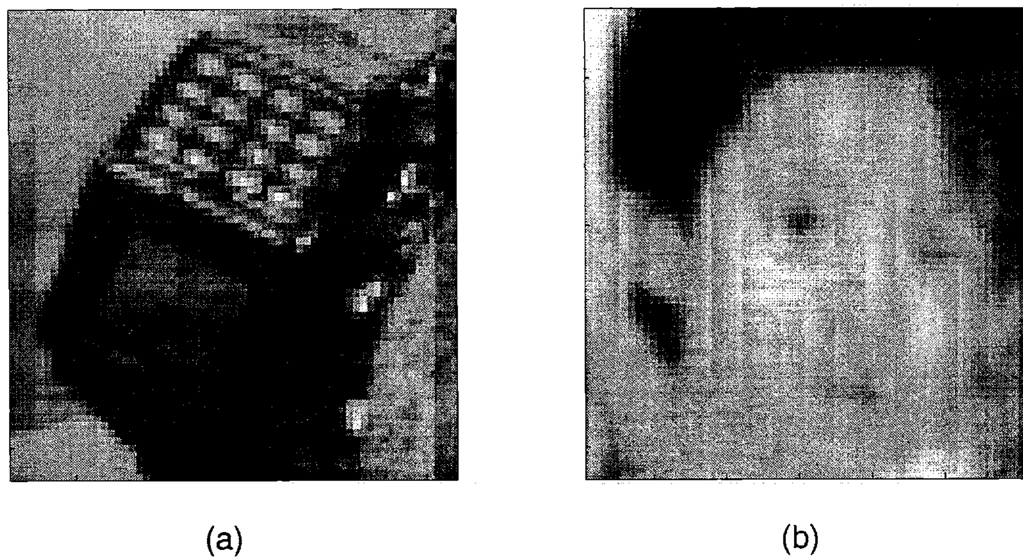


Figure 4.46: Arrhenius plot of leakage current for the N-type implant energy split



(a) (b)

Figure 4.47: Test images of the 0.35-μm (a) and 0.25-μm (b) NW/Psub photodiode APS arrays

any conclusion. It is suspected that the curving-up behavior at high temperature is due to some hot-carrier mechanism, which is the topic of Chapter 5.

4.7 Test Images

The test images of the 0.35-μm and 0.25-μm CMOS NW/Psub photodiode APS array are shown in Figure 4.47. The pixel sizes are 9.2 x 9.2 μm^2 and 4 x 4 μm^2 for the 0.35-μm and 0.25-μm CMOS APS arrays respectively. The dimension of the array is 64 x 64. A 1/2" C-mount lens with a focal length of 12mm and an f-number of 1.4 is utilized for taking the images.

4.8 Summary

This chapter presents the characterization methodology of pixel parameters and the comparison between different pixel designs, junctions types, and process conditions. The pixel parameters include conversion gain, spectral response, leakage current, sensitivity and noise. A 3-T NW/Psub photodiode APS array is characterized and the characterization approach is demonstrated as an example of modern CMOS process utilized for imaging applications. The results show that the dark current non-uniformity limits the ultimate noise performance for the active pixel sensor and should be addressed in the process. The

characterization approaches are utilized to compare different APS structures. The comparison results are summarized as the follows:

1. For a N^+/PW/Psub photodiode APS in a silicided CMOS process, a grid-shape or comb-shape diode structure can be utilized to improve Q.E. degradation due to silicide. However, the sensitivity and the leakage current performance are much worse than that of the NW/Psub photodiode APS in the same process.

2. A 4-T cascode photodiode APS, compared with a typical 3-T APS counterpart, is superior in higher conversion gain and a better S/N ratio. However, the drawback is a lower dynamic range and worse P-P FPN due to the small sensing capacitance. A potential image lag is another concern. This technique should be used widely according to the applications.

3. A photogate APS is capable of realizing CDS to remove reset noise. However, the Q.E. degradation due to optical absorption at the poly gate is a concern. Besides, the high leakage current induced at the gate edge significantly increases the dark current shot noise and P-P FPN and consequently degrades the overall performance.

4. A photodiode APS fabricated with bulk wafers demonstrates a higher Q.E. at long wavelength than the ones with epi wafers. The lower Q.E. for the epi wafers is due to the short diffusion length of carriers in the highly doped substrate beneath the epi layers. The high Q.E. for bulk wafers may not be an advantage if the high Q.E. region is beyond the visible range. Besides, the long minority-carrier diffusion length in the bulk substrate may cause higher pixel cross-talk and consequently degrade imager performance.

5. The N^+/PW/Psub, N^+/Psub, and NW/Psub photodiode APSs are compared using a 0.25-μm nonsilicided source/drain STI process. The NW/Psub and N^+/Psub photodiode APSs are observed to have similar Q.E. and leakage current performance. The N^+/PW/Psub photodiode APS has a lower Q.E. due to the narrower depletion width and the built-in electric field against photo-carrier collection. The N^+/PW/Psub photodiode APS also demonstrates a relatively higher leakage current than the other two structures due to the junction property.

6. A specifically designed mask is utilized to vary the implant dose and energy for the N-well implant at the NW/Psub photodiode APS. The variation on dose have little influ-

ence on the sense parameters. The implant energy also has little impact on Q.E. since the depletion region is mainly determined by the p-type substrate.

Chapter 5

The Effect of Hot Carriers on the Operation of CMOS Active Pixel Sensors

CMOS active pixel sensors (APS), different from CCD imagers, place transistors inside pixel cells for signal resetting, steering and buffering. Some complex pixel designs even place a circuit within a pixel for pixel-level signal processing. Ideally the transistors should conduct voltage and current according to the design without affecting the photo-sensor performance. However, this situation is no longer valid if a high electric field is generated by biasing the in-pixel transistors in the saturation region and produces hot carriers. The hot carriers generate excess minority carriers in the substrate that can be absorbed by the nearby photo-sensors and consequently distort the sensor response. If the hot carrier mechanism happens to the source follower transistor of an APS, the absorbed excess carriers will intensify the electric field. More excess carriers are generated and a positive feedback loop is thus formed. In a severe case, the charge bucket of the photo-sensor will be filled up instantly and saturates the pixel.

In this chapter, the mechanism of the induced excess minority carriers by hot carriers is described. This mechanism is identified by varying bias and temperature in the experiments. A 4-T photodiode APS array fabricated with a standard 0.35-μm process is used as the test structure. The spatial distribution of the excess minority carriers is also quantified. The influence to normal imager operation is estimated. From the mechanism, pixel design modifications for mitigating this effect are provided. As imager fabrication continues to shift towards deep sub-micron processes and the pixel size continues to shrink, this hot-carrier effect is expected to emerge.

5.1 Mechanism

Figure 5.1 shows the circuit diagram and the positive feedback loop of the excess carriers in a typical 3-T photodiode APS. The source follower transistor Q_{n2} is operated in saturation region and a high electric field region resides in the pinch-off region near the drain of

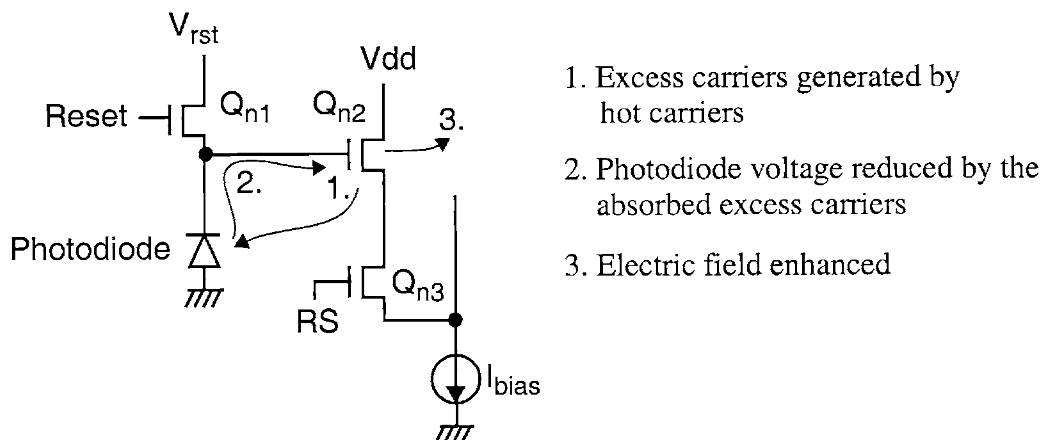


Figure 5.1: Positive feedback mechanism for hot-carrier induced excess minority carriers

Q_{n2}. As current passes through this region, electrons are accelerated by this strong electric field and become "hot" electrons before their energy is released to the lattice. According to the literature [29, 30, 31], as the energy of the hot carriers reaches a threshold, excess minority carriers (electrons in this case) are generated in the substrate either through photon emission or impact-ionization (step1). Those minority carriers can be easily absorbed by the nearby photodiodes and lower the photodiode voltage, which is also the voltage at the gate of Q_{n2} (step2). Due to the source follower structure, the source voltage of Q_{n2} follows the photodiode voltage and decreases. Since the drain of Q_{n2} is connected to Vdd, the lowered source voltage increases V_{ds} and enhances the E-field in the pinch-off region (step3). The hot-carrier mechanism is strengthened and more minority carriers are generated. A positive feedback loop is thus sustained.

This effect can be influential because the sensing capacitance of the pixel is pretty small. Typically the capacitance is on the order of femto-Farad (10^{-15} F). Even the hot-carrier induced excess current as low as in femto-amps range is sufficient to hamper the imager resolution for an 8-bit output at a frame rate of 30 frames/sec. For pixel designs that utilize a small sensing capacitance for sensitivity improvement, the influence of the hot carriers is more significant. Moreover, as the geometry of the transistors continues to shrink, the electric field in the pinch-off region under the same bias condition is expected

to increase. This hot-carrier effect is expected to play an important rule in CMOS imager scaling.

In the operation of an APS imager, the voltage at the sensing node is also controlled by the intensity of optical signal. Equivalently, the optical signal determines V_{ds} of Q_{n2} and modulates the amount of excess minority carriers generated in the substrate. Therefore, the amount of excess carriers absorbed by the photo-sensors become dependent on the optical signal and produces nonlinear imager response. For image processing highly relying on sensor linearity, e.g. color processing, the hot-carrier effect becomes a source of distortion and should be minimized.

This hot-carrier effect does not solely impact the active pixel sensors. Any pixel design with transistors biased in deep saturation and located near the photo-sensors is subject to this effect. For example, if a high-speed digital circuit is incorporated in a pixel, the logic transistors will be operating in deep saturation during on/off transitions. Depending on the switching activity of the circuit, the amount of excess carriers absorbed by the photo-sensor varies and the output signal will depend on the switching activity of the logic circuits.

5.2 Experimental Results

In order to verify and quantify the hot carrier effect for a modern CMOS process, a 64 x 64 4-T NW/Psub APS array is utilized. This array is fabricated with a standard 0.35-μm CMOS processes as described in the previous chapter. Pixel layout and circuit diagram are shown in Figure 5.2. The sensor can either operate in a cascode mode (gain-enhancement mode) [49] if the gate voltage of Q4 is at a lower value (e.g. 1.5V) or it can operate in normal mode (similar to a 3-T pixel) if the gate voltage of Q4 is at Vdd. Typical sensor characteristics for these two operating modes is shown in Table 5.1. The reason for running the APS array in the two modes is that they have the same charge collection efficiency but different sensing node capacitance. The sensitivity of the hot-carrier effect to the sensing node capacitance can be examined.

The control variables in this pixel structure are I_{bias}, Vdd, V_{rst}, V_{ref}, and temperature. I_{bias} controls the current flowing through the source follower transistor Q2 and is adjust-

Process	0.35-μm standard CMOS logic process
Format	64 x 64
Pixel size	10 x 10 μm^2
Sensor type	NW/Psub photodiode
Operation mode	normal mode (V_{ref} = 3.3V), cascode mode (V_{ref} = 1.5V)
Fill factor	~ 40%
Sensitivity	1.4 V/lx/sec (normal), 4.5 V/lx/sec (cascode)
Leakage current	2.2 fA/pixel/sec @ 300K
Capacitance	23 fF (normal), 7.2 fF (cascode)

Table 5.1: Characteristics of the 4-T photodiode APS array in the experiments

able with an off-chip resistor. Vdd, V_{rst}, and I_{bias} determines the drain-to-source voltage (V_{ds}) of Q2 and therefore affect the electric field in the pinch-off region. Temperature is also an important factor since the mean free paths of the hot carriers depend on it.

In the experiments, a randomly-selected pixel on the array is addressed to observe the temporal evolution of the photodiode voltage right after reset. The same measurement is also performed on other pixels to assure the behavior is common to all the pixels. The gain of the on-chip source followers has been characterized so that the measured output voltage can be referred back to the photo-sensing node. All the measurements are conducted in a dark environment.

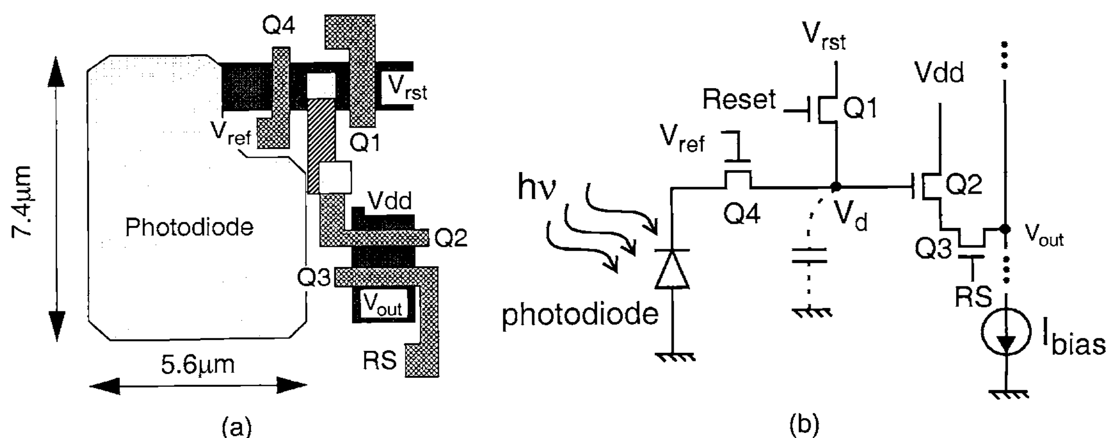


Figure 5.2: Pixel structure used for hot-carrier effect characterization: (a) pixel layout and (b) circuit diagram

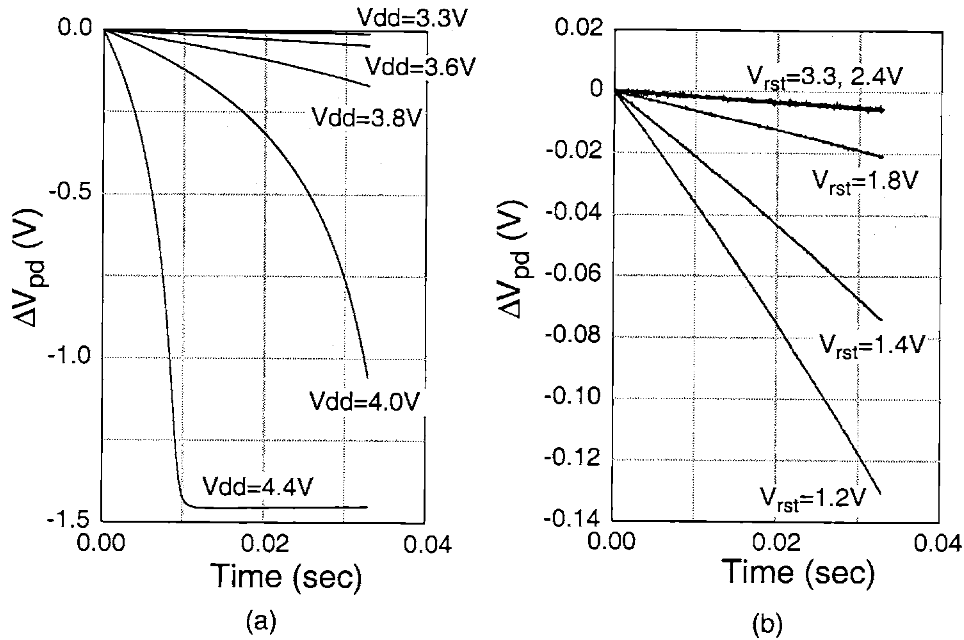


Figure 5.3: Temporal evolution of photodiode voltage change in a dark environment after reset: (a) with different Vdd (V_{rst}=3.3V, V_{ref}=3.3V, I_{bias}=15μA) (b) with different V_{rst} (Vdd=3.3V, V_{ref}=3.3V, I_{bias}=1.5μA)

The first set of measurements are performed to observe the dark signal dependence on Vdd and V_{rst}. According to the described hot-carrier mechanism, a higher Vdd or a lower V_{rst} will raise V_{ds} of the source follower. The density of excess minority carriers is expected to increase and induces a higher dark signal. Figure 5.3 shows the time evolution of the photodiode voltage as various Vdd or V_{rst} is applied. Notice the I_{bias} difference for the two plots. A strong positive feedback behavior is observed as raising Vdd. An increase of dark signal is also observed for lowering V_{rst}. The behaviors agree with the mentioned hot-carrier mechanism.

To further quantify the bias dependence, the dark signal is calculated by taking the slope of the curves near t = 0. The results with the case of varying V_{rst} are shown in Figure 5.4. The measurements are conducted with three Vdd settings. As V_{rst} is varies from 3.3V to 2.4V, the curves remain roughly the same. As V_{rst} continues to decrease, the dark signal increases exponentially. Before this phenomena is fully attributed to the hot-carrier effect, other mechanisms that could possibly cause the dependence of dark signal on V_{rst} are examined. The increase of junction leakage current and the decrease of junction capaci-

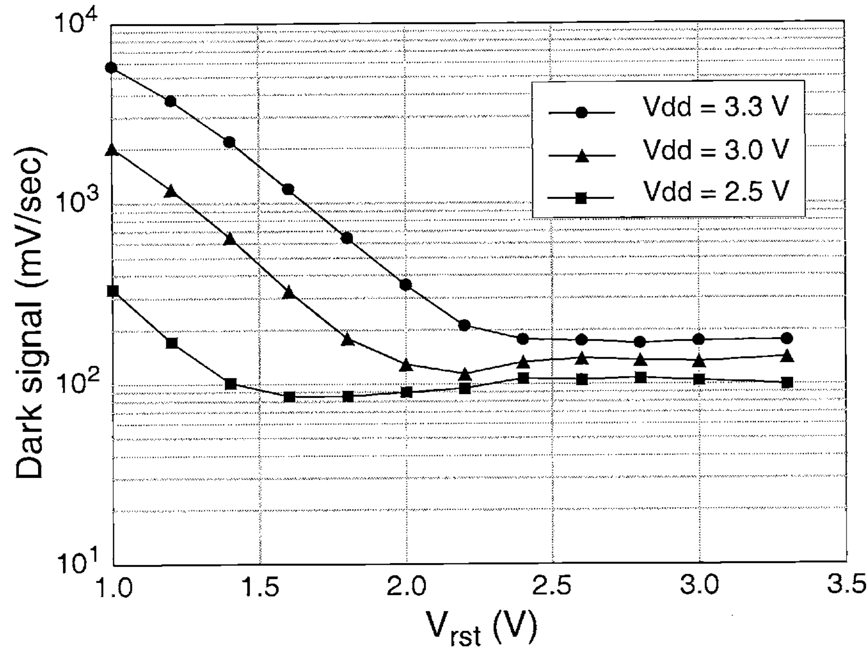


Figure 5.4: Dark signal versus V_{rst} with different Vdd values (V_{ref}=3.3V, I_{bias}=1.5μA)

tance are two other V_{rst} dependent mechanisms. However, a lower reverse-bias voltage should decrease the junction leakage current and increase the junction capacitance, both are in opposite with the observed phenomena. This leads to the conclusion of an increased minority carrier density due to the hot-carrier effect.

A strong correlation between the dark signal and V_{ds} of Q2 is presented in Figure 5.5. The bias conditions are the same as the case in Figure 5.4 and the source voltage is determined by SPICE simulation. To further demonstrate the relation between dark signal and electric-field in the pinch-off region, the maximum E-field in the channel is simulated and plotted in Figure 5.6. The dark signal is observed to be mainly determined by the maximum E-field and increases rapidly as the field becomes higher than 2.5 x 10^5 V/cm, which corresponds to a V_{ds} around 2V.

Dark signal dependence on the magnitude of I_{bias} is measured and shown in Figure 5.7. The dark signal increases with the amount of I_{bias} since it determines the amount of electrons flowing through the high E-field region, where the hot electrons contribute to the excess carrier generation. A higher I_{bias} value also increases V_{gs}, which enlarges V_{ds} of

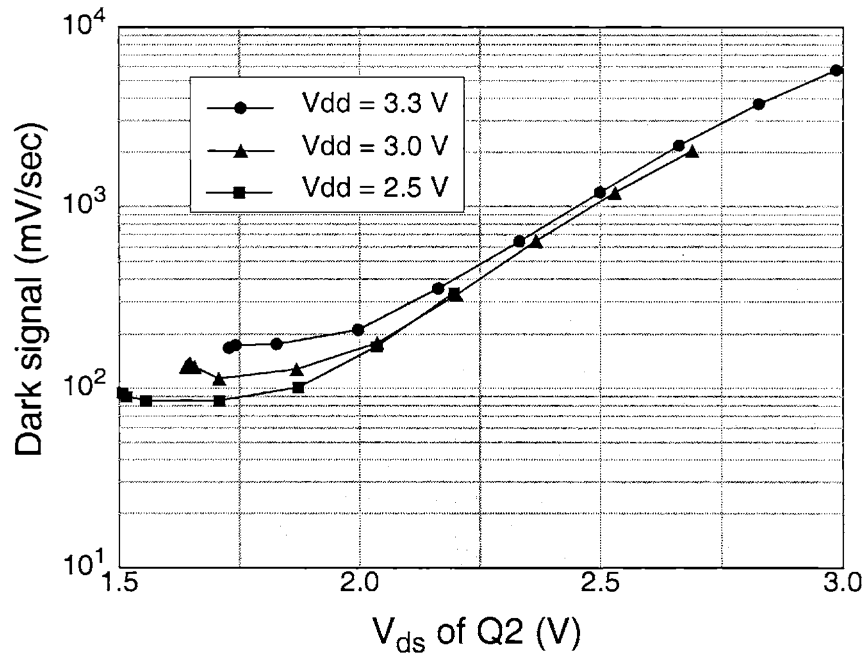


Figure 5.5: Dark signal versus V_{ds} of Q2 with different Vdd values (V_{ref}=3.3V, I_{bias}=1.5μA)

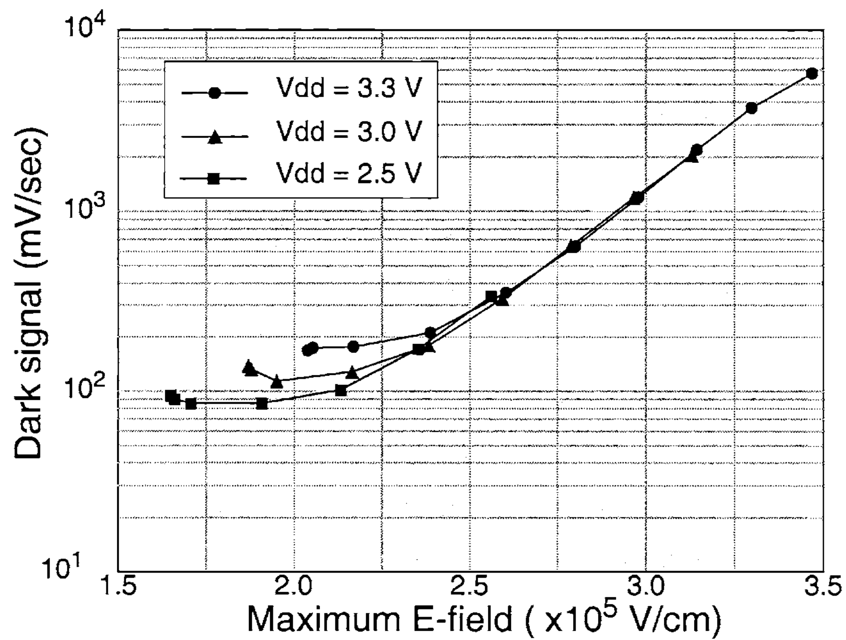


Figure 5.6: Dark signal versus simulated maximum E-field in the pinch-off region of Q2

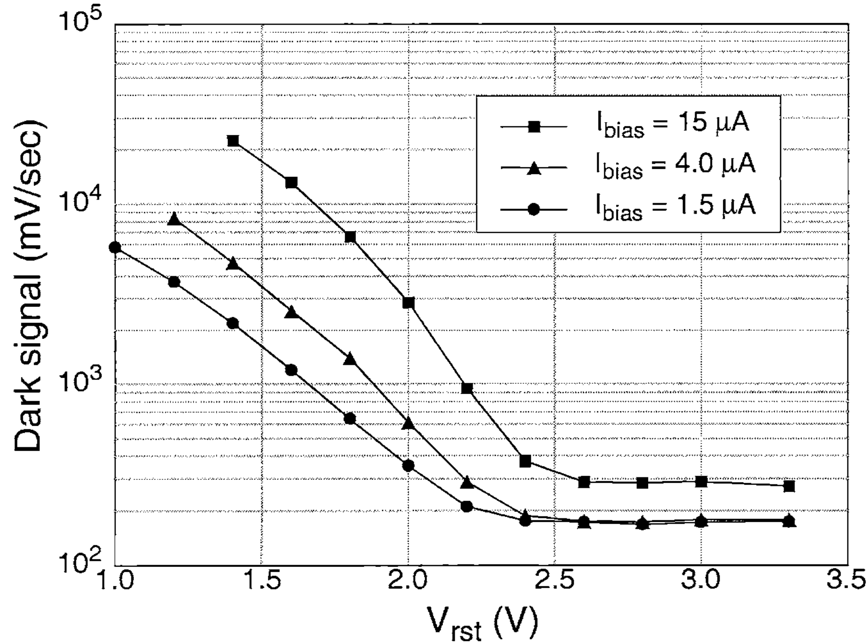


Figure 5.7: Dark signal versus V_{rst} with different I_{bias} values (Vdd=3.3V, V_{ref}=3.3V)

Q2 and favors the hot-carrier mechanism. In another bias-varying experiment, Vdd is varied and the dark signal results are shown in Figure 5.8. As expected, the dark signal increases with a higher Vdd.

A cascode operation mode on the 4-T APS can be utilized to improve imager sensitivity. However, the small sensing capacitance associated with the structure makes it more vulnerable to the collection of excess minority carriers. The dark signal versus V_{ds} of Q2 is shown in Figure 5.9. Compared with the normal-mode operation, the dark signal of the cascode mode is much higher for the same bias conditions. The same mechanism is also detrimental to the operation of dynamic logic circuits with nearby transistors operating in saturation region. It is noticed from Figure 5.9 that the onset V_{ds} where the hot-carrier mechanism starts to affect is lower than that for the normal-mode case. This could be due to the capacitance coupling and charge-injection effect when the reset transistor switches from "high" to "low". The voltage of the charge sensing node will be driven down by this effect and the V_{ds} of Q2 increases slightly. This phenomena is more significant when the switch is connected to a node with a smaller capacitance. Since the V_{ds} on Figure 5.9 shows the drain-to-source voltage before reset, the actual V_{ds} after reset should be higher

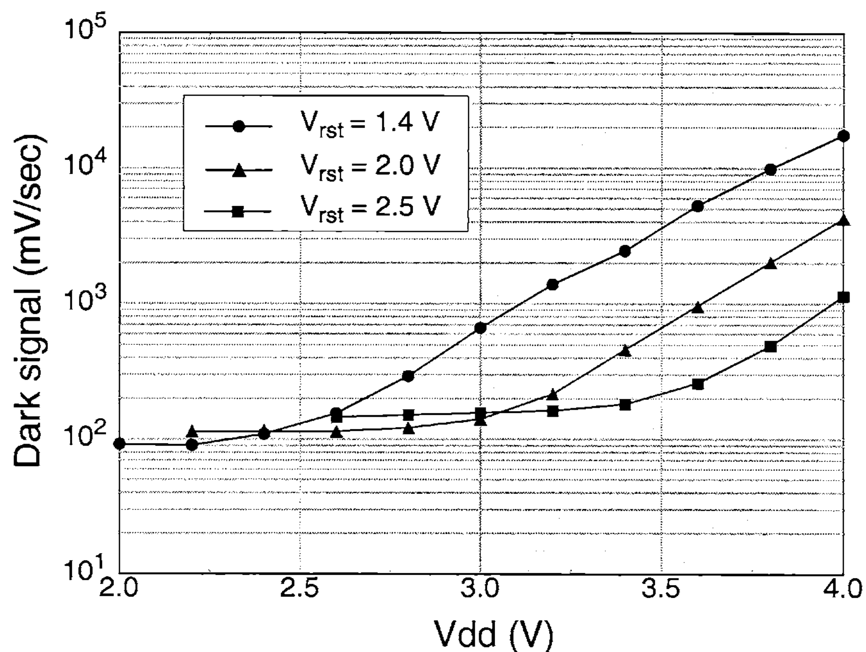


Figure 5.8: Dark signal versus Vdd with different V_{rst} values (V_{ref}=3.3V, I_{bias}=1.5μA)

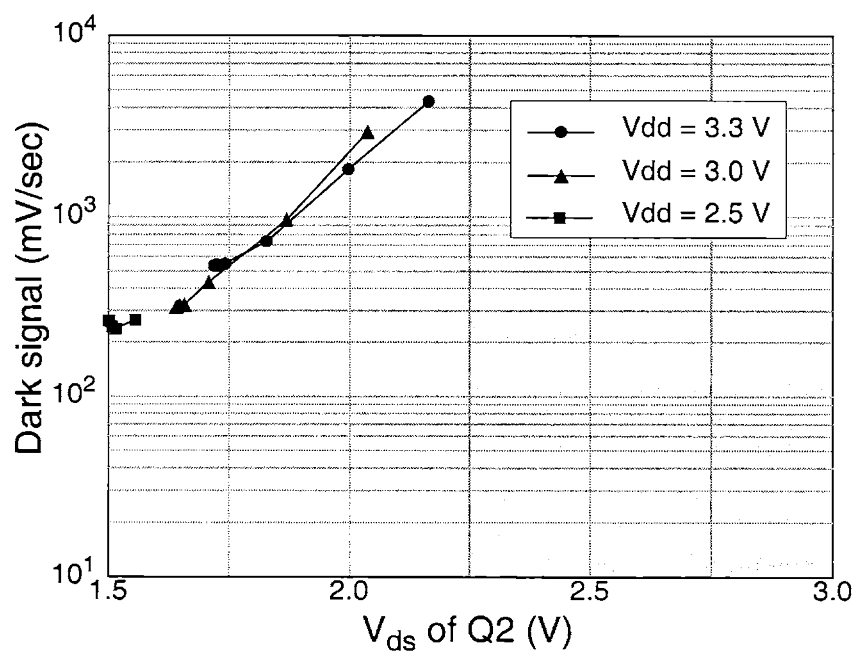


Figure 5.9: Dark signal versus V_{ds} of Q2 with different Vdd values for the cascode mode operation (V_{ref}=1.5V, I_{bias}=1.5μA)

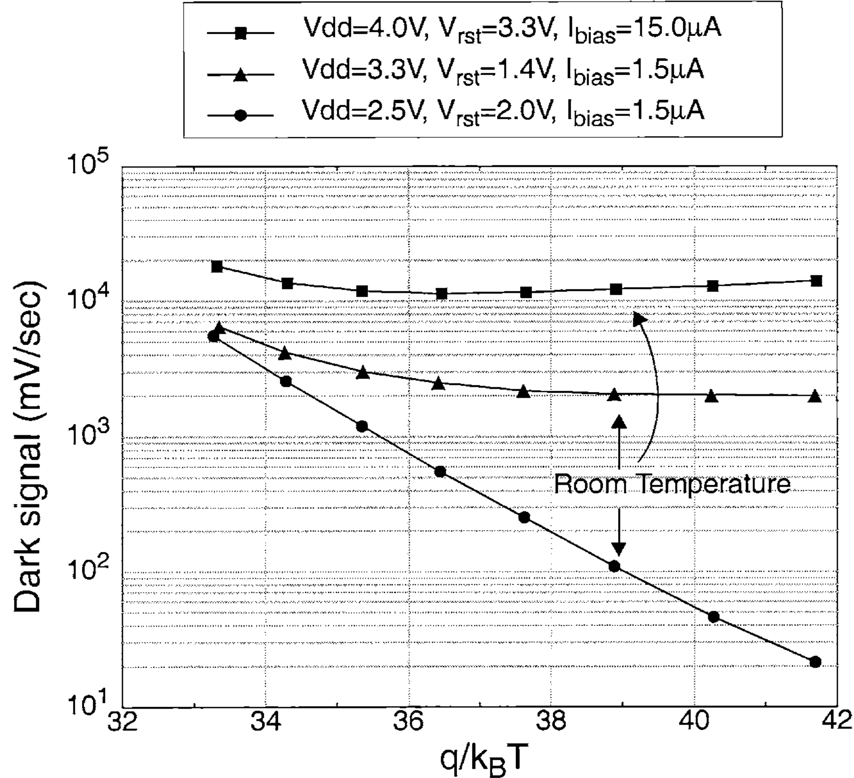


Figure 5.10: Arrhenius plot of dark signal under three different bias conditions. Square: strong hot-carrier (H-C) effect; Triangle: medium H-C effect; Circle: negligible H-C effect

than the posted values. According to SPICE simulation, the V_{ds} difference before and after reset for the 4-T cascode mode can be as large as 200–300 mV. Taking that into consideration, the onset V_{ds} for the cascode mode is around the same value as the normal-mode case.

To verify the hot-carrier (H-C) mechanism, a temperature-varying experiment is conducted under three bias conditions that present different levels of the H-C effect — strong, medium, and negligible. Chip temperature is varied from 5°C to 75°C with 10°C increments. The Arrhenius plot of the dark signals is shown in Figure 5.10. At higher temperature (lower q/k_BT), the dark signal is dominated by junction leakage current for the medium and the negligible H-C cases and the two curves tend to merge. At lower temperature, the curve with negligible H-C effect remains a typical exponential behavior with temperature; whereas the dark signal of the medium H-C effect saturates at a higher value. For the strong H-C case, the dark signal also increases at higher temperature due to junc-

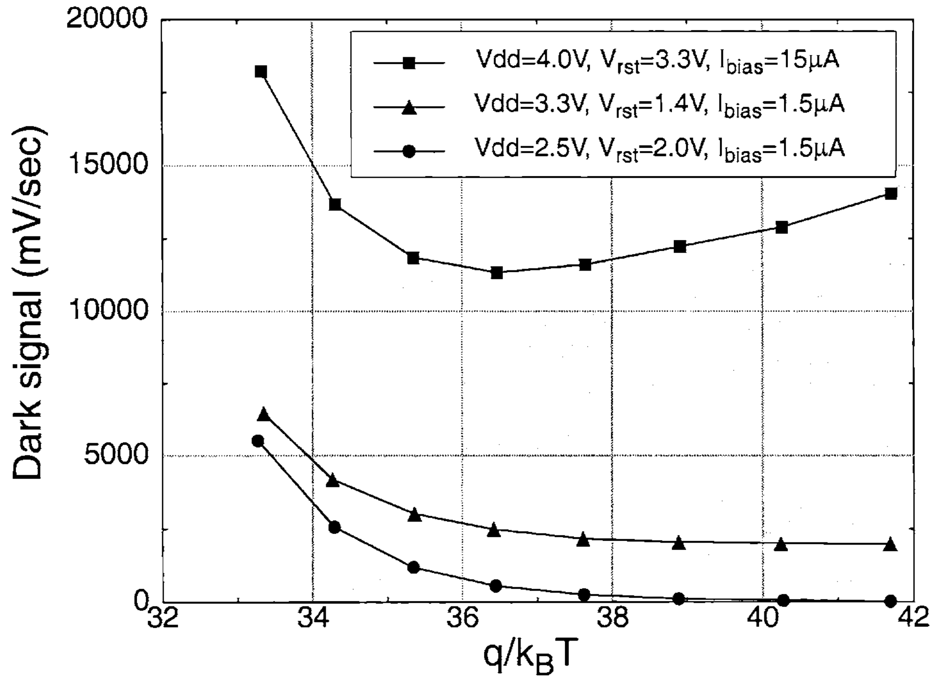


Figure 5.11: Dark signal versus q/k_BT under three different bias conditions in a linear plot. Square: strong hot-carrier (H-C) effect; Triangle: medium H-C effect; Circle: negligible H-C effect

tion leakage. At lower temperature, however, the dark signal increases as well. This phenomena can be more clearly observed with a linear plot shown in Figure 5.11. This is a strong evidence of the hot-carrier mechanism since the mean-free-path of hot carriers is longer at lower temperature and the mechanism is more effective.

A measurement is performed to quantify the spatial distribution of the excess minority carriers. It is conducted by applying a long row-select signal (30ms) on a specific row (the 33rd row in this experiment) while keeping the row-select time of the other rows at 50μs. More excess minority carriers are generated on that particular row due to the long row-select time. The charge integration time for each row is 33ms. The frame signal of the normal timing case is subtracted from the frame signal of this special timing case to show the spatial distribution of the excess dark signal. The excess dark signal versus row number is presented in Figure 5.12. According to this figure, the affected region by the excess carriers is within ~30μm from the stressed row. Another high bias stress experiment is performed to examine the bias dependence on the spatial distribution of the excess minority

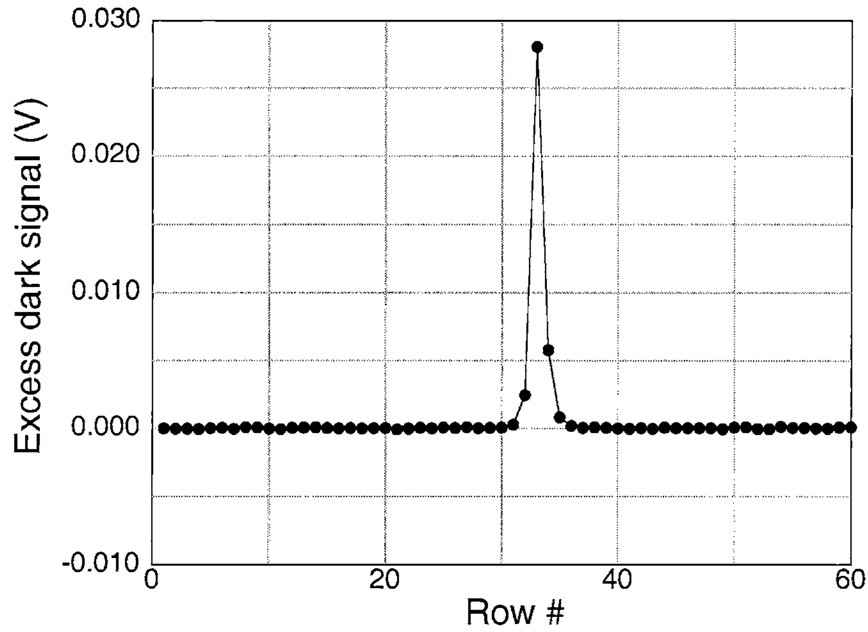


Figure 5.12: Excess dark signal versus row number with stress at the 33rd row ($Vdd=3.3V$, $V_{rst}=1.6V$, $V_{ref}=3.3V$, $I_{bias}=1.5\mu A$)

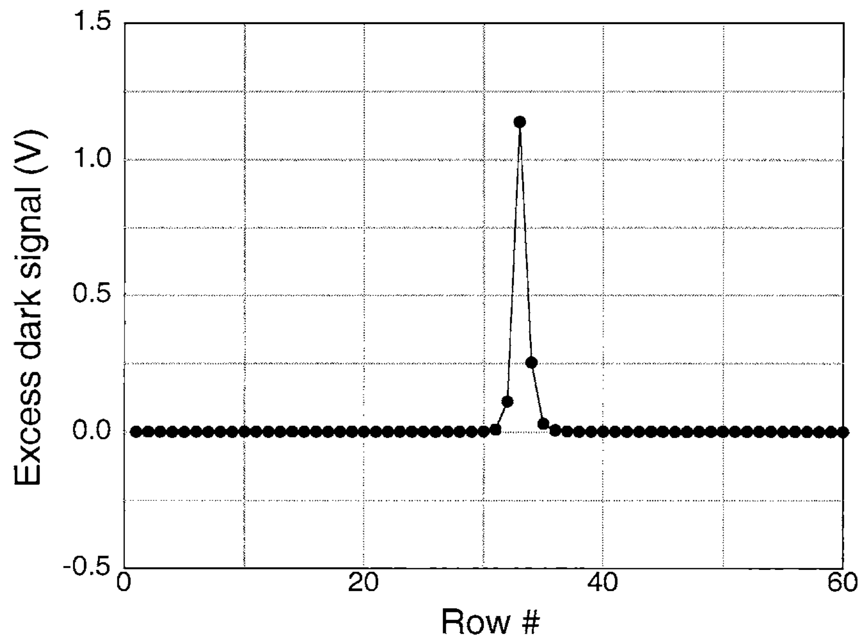


Figure 5.13: Excess dark signal versus row number with a higher stress at the 33rd row ($Vdd=4.4V$, $V_{rst}=2.5V$, $V_{ref}=3.3V$, $I_{bias}=15.0\mu A$)

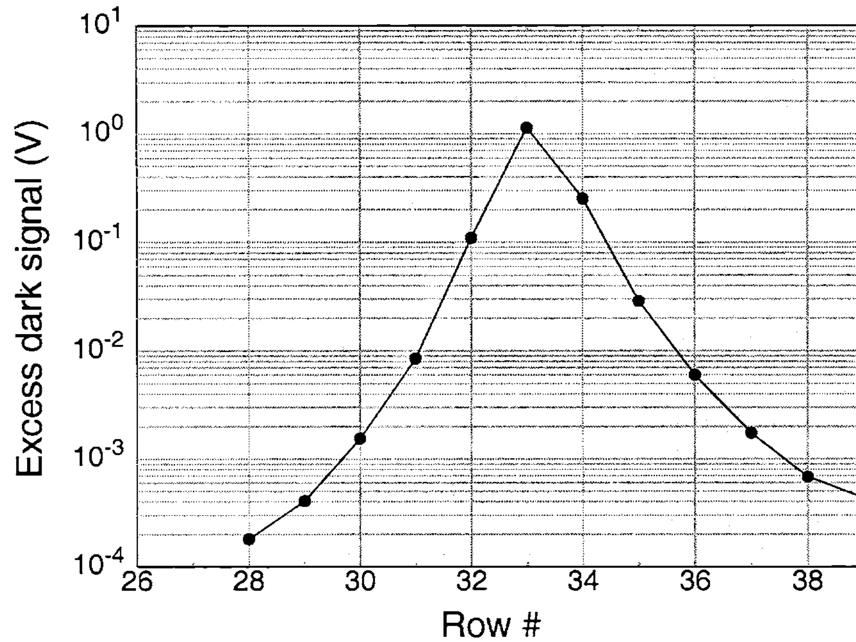


Figure 5.14: Center portion of Figure 5.13 in log scale

carriers. The result is shown in Figure 5.13. Comparing the figures of both stress cases, the spatial distribution patterns are essentially the same except for a scaling factor. This result ensures that the mechanism producing excess minority carriers is the same for typical imager operating conditions. The excess dark signal around the stressed row for the high stress case is plotted in log scale and presented in Figure 5.14. The response curve presents a exponential decay relation with pixel location: $V_{out}(y) = V_{peak} \cdot e^{-y/Y}$. The 1/e characteristic decay length is roughly 5µm. This demonstrates that the influence of the excess minority carriers is a short-range effect.

5.3 Discussions

5.3.1 Hot-carrier Effect in Normal Imager Operation

In the above single-pixel measurements, the current source in the column circuit is always on and the source follower transistor is continuing under stress. This is quite different from the real case. In imager readout sequence, the row-select signal only turns on during the readout phase of that particular row. Most of the time the row-select signal is off. Therefore, the influence of the hot-carrier induced excess minority carriers will not be as

significant as in the measurements. Based on the measurement results, the excess output voltage due to the hot-carrier effect can be estimated. The estimation is described below.

If the hot-carrier effect is not considered, the output voltage of the pixels can be written as

$$V_{out} = (I_{ph} + I_{dark}) \cdot T_{int} \cdot CG \cdot G_{SF}, \tag{5.1}$$

where I_{ph} denotes photo-current, and I_{dark} represents dark current. Once the hot-carrier effect is taken into account, another term will be inserted and the equation looks like

$$V_{out-HC} = \left[(I_{ph} + I_{dark}) \cdot T_{int} + \int_{T_{RS}} I_{HC}(V_{\mathrm{ds}})\, dt \right] \cdot CG \cdot G_{SF}, \tag{5.2}$$

where $I_{HC}(V_{\mathrm{ds}})$ represents the current component from the hot-carrier effect and T_{RS} denotes the period of the row-select time. $I_{HC}(V_{\mathrm{ds}})$ can be extracted from Figure 5.5 and written as

$$I_{HC}(V_{\mathrm{ds}}) = I_0 \cdot e^{\beta(V_{\mathrm{ds}} - V_0)}, \tag{5.3}$$

where β represents the slope of curve on Figure 5.5; I_0 and V_0 represent the leakage current and V_{ds} at the onset of the hot-carrier effect.

In the calculation, the integration time is set to be 1/30 second and the row-select time is 50µs. From Figure 5.5, β is fitted to be 3.4. The normal-mode (V_{ref}=3.3V) and cascode-mode (V_{ref}=1.5V) operations are considered in the calculation for comparison. Vdd and V_{rst} are both 3.3V and I_{bias} is 1.5µA in the estimation. The excess V_{out}, which is the voltage difference between the output voltages with and without the hot-carrier effect, is calculated by subtracting eq.(5.1) from eq.(5.2) and the result on a linear-log plot is shown in Figure 5.15. Both excess V_{out} curves increase exponentially with the output signal. The excess V_{out} of the cascode mode is higher than that of the normal mode due to the small sensing capacitance. Under the prescribed bias and row-select time, it is found that the magnitude of the excess V_{out} is only a small percentage of V_{out}. Therefore, this hot-carrier effect seems to be minor under normal operating conditions in current CMOS APS designs. However, as the transistor size continues to shrink and the separation between

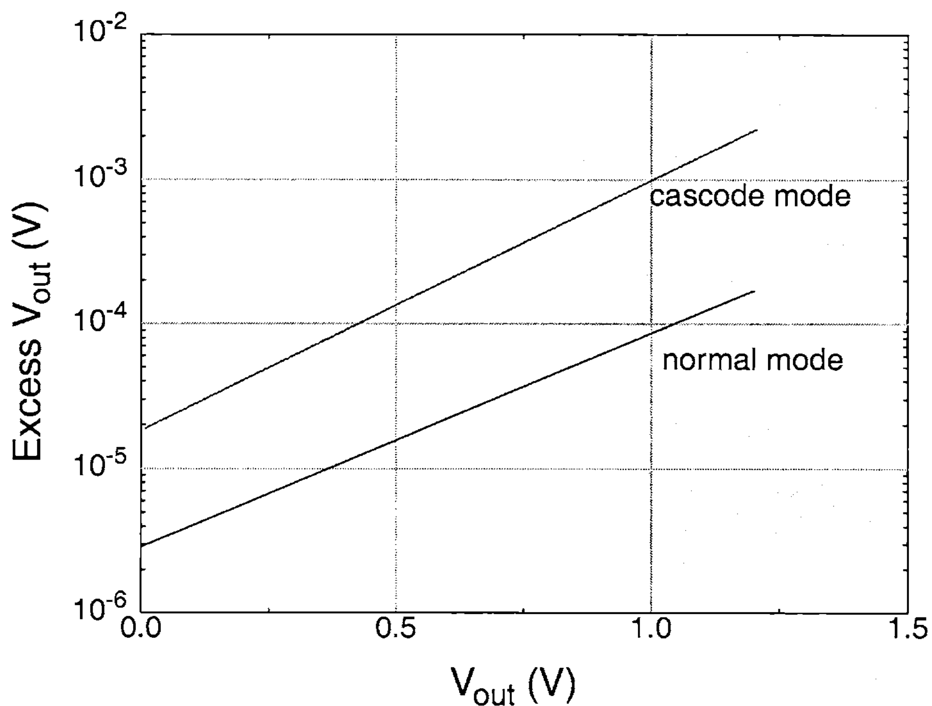


Figure 5.15: Excess V_{out} versus V_{out} without the H-C effect for the APS arrays operating in normal mode and cascode mode (Integration time: 1/30 sec.)

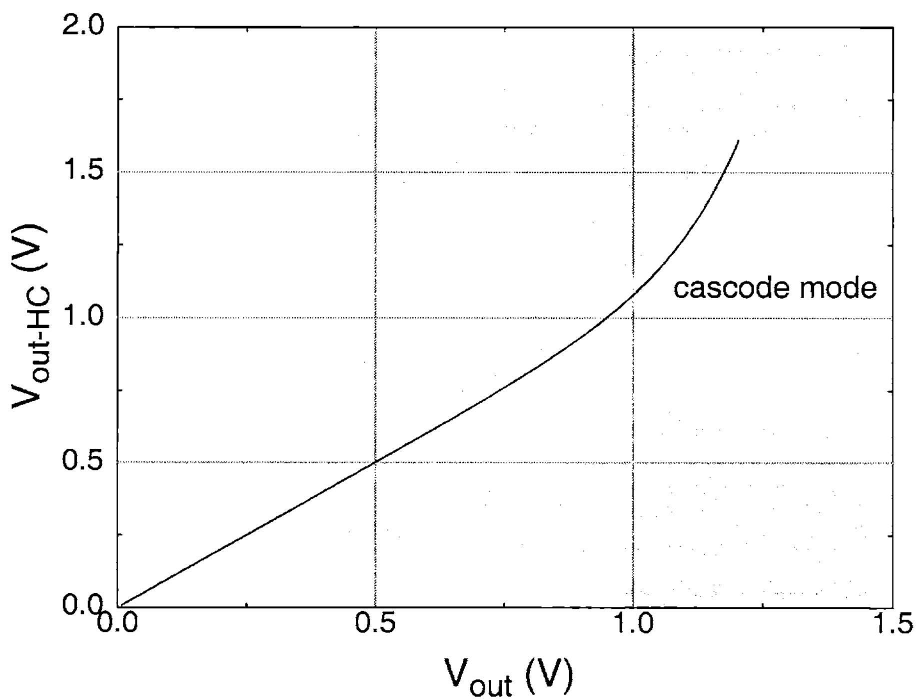


Figure 5.16: V_{out-HC} versus V_{out} without the H-C effect for an APS array with a β of two times the characterized value

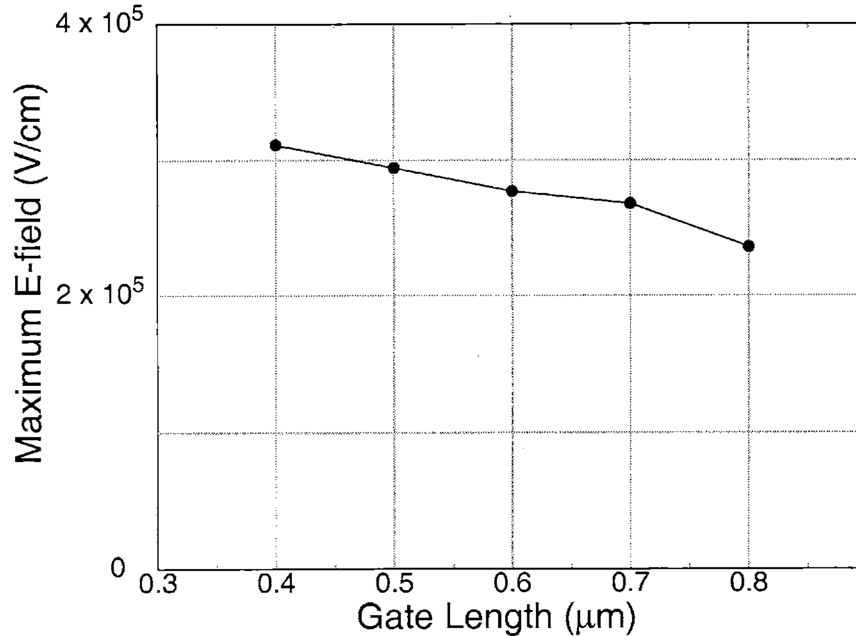


Figure 5.17: Simulation results on maximum electric field in the pinch-off region with different gate lengths (The bias condition corresponds to the case with medium H-C effect in Figure 5.10)

photo-sensors and pixel transistors reduces, this hot-carrier effect is expected to emerge soon. Figure 5.16 shows V_{out} versus $V_{out\text{-}HC}$ for the cascode-mode APS operation with a β of two times the characterized value. All the other parameters and conditions hold the same values as the previous simulation. The output voltage presents a strong nonlinearity. For a large-dimension imager with a higher I_{bias} for faster signal readout, the extent of nonlinearity is expected to increase too.

5.3.2 Suggestions on Pixel Design

The hot-carrier effect is expected to emerge as the device size continues to shrink. Several measures can be taken to mitigate the effect. According to device simulation results shown in Figure 5.17, a longer gate length on the source follower transistor Q2 can be utilized to reduce the maximum electric field. However, the maximum E-field does not have significant dependence on gate length. A more effective approach is to apply a lower voltage at the drain of Q2. For a typical 3-T APS design shown in Figure 5.18(a), this approach requires separate junction nodes for Vdd and Vdd' as shown in Figure 5.18(b). Further, one more wire is required to provide an extra voltage for the pixel. This approach poten-

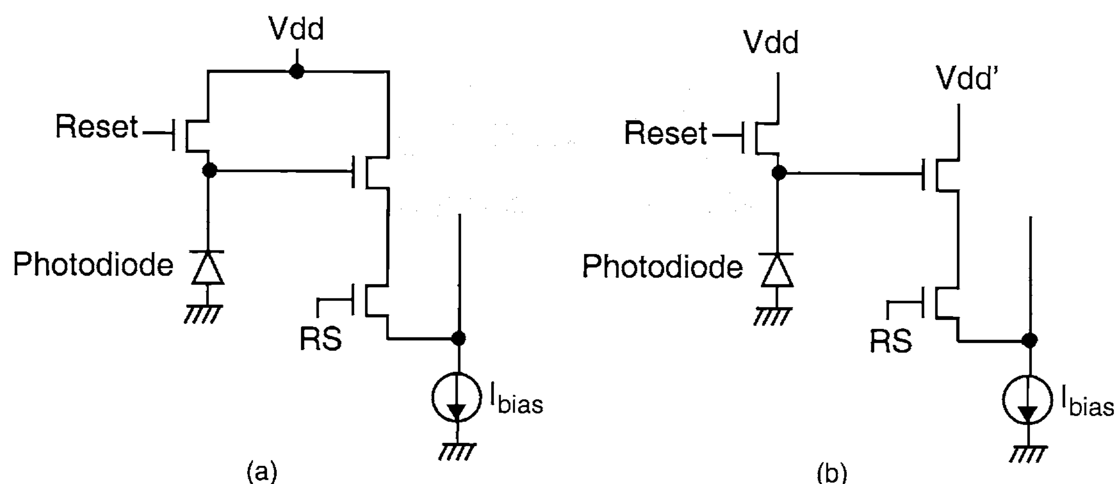


Figure 5.18: (a) Typical 3-T APS design; (b) modified pixel design for mitigating the H-C effect

tially reduces fill factor for pixel designs.

5.4 Summary

The effect of excess minority carriers induced by hot carriers at the source follower transistors of an APS is experimentally observed and quantified using a 4-T NW/Psub APS array fabricated with a standard 0.35-μm CMOS process. The amount of excess minority carriers absorbed by the photo-sensors depends on bias conditions of the transistors as well as the intensity of illumination. Nonlinear sensor response can result. A 4-T APS operating in a cascode mode is more sensitive to this effect due to its small sensing capacitance. A temperature varying experiment is performed to confirm the hot-carrier mechanism. The spatial distribution of the excess minority carriers is quantified to be ~30 μm around the stressed transistors. This distribution range of the excess carriers is the same for typical bias conditions. Under normal row-select timing and bias, the influence of the excess minority carriers to the output signal is minor for the APS. However, as the transistor size continues to shrink and the separation between the photo-sensors and the stressed transistors reduces, this hot-carrier effect is expected to emerge. Suggestions on pixel designs to alleviate this effect is provided.

This hot-carrier phenomena also suggests to avoid any in-pixel transistor operating in

deep saturation region. If the pixel circuit is too complex to achieve that, special attention should be paid to properly isolate the photo-sensors from the excess minority carriers.

Chapter 6

The Influence of Pixel Cross-talk on CMOS Color Imagers

The latest development on imagers show decreasing pixel size in order to simultaneously achieve higher spatial resolution and maintain a reasonable chip size. As pixel size continues to shrink, significant cross-talk between pixels emerges. This cross-talk phenomena reduces the effective spatial resolution and amplifies the pixel-to-pixel color mismatch for an imager with a mosaic color filter pattern. The cross-talk phenomena and its influence on color imagers are the topic to be discussed in this chapter.

In this chapter, the fundamental concepts on color imager is presented first. It is followed by the description of the major cross-talk mechanism. The cross-talk mechanism is then studied using an N-well APS array with a hollow metal shield, which exposes a single pixel in the center of the array and covers the others. Monochrome light is then applied on the exposed pixel to examine the 2-D distribution of the cross-talk signal. Device simulations confirm that minority carrier diffusion is the major cross-talk mechanism. The measured cross-talk data is then used to study the amplification effect of the pixel-to-pixel color mismatch induced by cross-talk. From the data and the characterization results on imager noise, the degradation factor of the color mismatch can be estimated. Finally, process and layout improvements are proposed for reducing the cross-talk effect.

6.1 Fundamentals of Color Imagers

A background knowledge on color science is provided in this section to assist in apprehending the color signal processing in imagers. The content touches upon the basic principles of color imagers, the trichromaticity theorem, the color matching function, and the chromaticity diagrams. The description in this section is not intended to cover all aspects of color science. More extensive and detail information on color science can be found in reference [51, 52].

6.1.1 Basic Principle of Color Imagers

Color perception is one of the important senses of human beings. The reason why people recognize color is because there are three types of photo-receptors on the retinas of the eyes. Those photo-receptors carry peak spectral responses corresponding to long (red), medium (green) , and short (blue) wavelengths. The excitation ratio between the three photo-receptors determines the chromaticity of the stimulus and the magnitude represents the light intensity. Therefore, the signal dimension for color perception is only three. For example, if two optical stimuli display identical color, it does not necessarily mean the spectral energy distribution for both stimuli are identical. instead, what is identical for the stimuli is their color coordinates, which can be transformed from their spectral distribution onto a three-dimensional color space. For a specific color coordinates, there are many possible spectral distributions that map to it. Using official terminologies, those optical stimuli are "metameric" to each other.

In order to identify the color coordinates of light with silicon photo-sensors, at least three types of color filters with linearly independent spectral responses are required. The concept of color sensing and displaying is shown in Figure 6.1 First, the optical stimulus is analyzed by the three types of color sensors to identify the stimulus's color. With some signal processing, the color signal is transformed into a color space that is recognizable by the display and then be presented on the screen.

Typically there are two ways to arrange the color filters and imager chips. The first approach utilizes off-chip color filters and several imager chips. The incoming light is split into three different paths. The light on each path is filtered by a different color filter (R,G,B filters for example) and then projected onto individual imager chip. The resolution

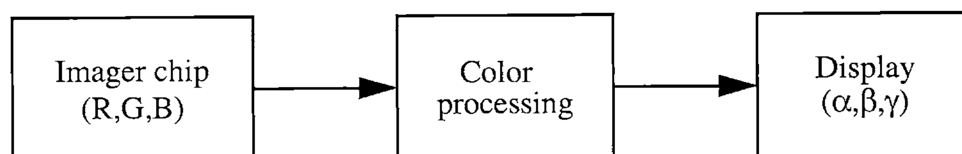


Figure 6.1: Procedure of color image sensing, processing, and display

R	G	R	G
G	B	G	B
R	G	R	G
G	B	G	B

Figure 6.2: RGB Bayer pattern

of this color imaging system is good. However, it requires many optical components and at least three imager chips to extract color signal. This approach is expensive and therefore is typically adopted for high-end, high-quality imaging applications. The second approach, which is widely utilized in consumer imaging systems, is to arrange on-chip color filters in a mosaic pattern. Figure 6.2. shows one of the popular RGB color filter pattern called Bayer pattern. For this type of imager each pixel has one color filter on top. Because of that, some color interpolation algorithm is required to generate the other two lacking color signal for each pixel. Usually the interpolation algorithm utilizes the signal from the neighbor pixels with different color filters to generate the lacking color signal. Because of the interpolation operation, the effective resolution for this type of imager is not be as good as the previous one. Another observation from this color filter pattern is that the density of the green filters is two times the density for red or blue filters. This arrangement takes the advantage of the fact that human eyes has higher sensitivity for green light. Therefore the higher accuracy of green signal improves the effective sensitivity and the higher density of green pixels enhances the effective resolution.

6.1.2 Trichromatic Matching Experiment

From the above discussion, it is understood that color information can be encoded with three parameters. In order to quantify the three parameters for color science research, a scientific way for defining a color is desired. The basic idea is as follows. Since the dimension of color is three, it is plausible to pick up three well defined and widely available optical stimuli as the primary basis. Any other color stimulus can be matched by mixing proper proportions of the three primary stimuli. Therefore, the coefficients of the three

primary stimuli will uniquely define the matched color. Experiments show that color matching obey linearity laws so the concepts of linear systems and linear algebra is applicable. This principle is called Grassmann's laws in color science [51, 52]. The pioneering experimental work was conducted by two groups of scientists in the early 20th century. The first group was led by J. Guild [53] at National Physical Laboratory at Teddington and the other was led by W. D. Wright [54] at Imperial College, Kensington. Both teams independently performed similar experiments and their results were consistent. With proper modifications, the results were announced by the Commission Internationale de l'Éclairage (CIE, the English translation is "International Commission of Illumination") to become a internationally accepted standard.

The designated primary stimuli in the experiments are monochromatic light with wavelengths at 700 nm, 546.1 nm, and 435.8 nm, which are denoted as R, G, and B resectively. The radiant power ratio for the $R/G/B$ primary stimuli are 72.1: 1.4: 1.0, which is equivalent to luminous ratio 1: 4.5907: 0.0601 at the corresponding wavelengths. These numbers are selected so that the three chromaticity coordinates for equal-energy stimulus[1] are the same.

Figure 6.3 presents the idea of the experiment. The output of $R/G/B$ stimuli are mixed with an optical diffuser. The intensity of the $R/G/B$ sources is adjustable. Light source U represents the color stimulus to be matched and its output is guided into another diffuser. The aperture defines a specific observing angle because the human color sense is angle

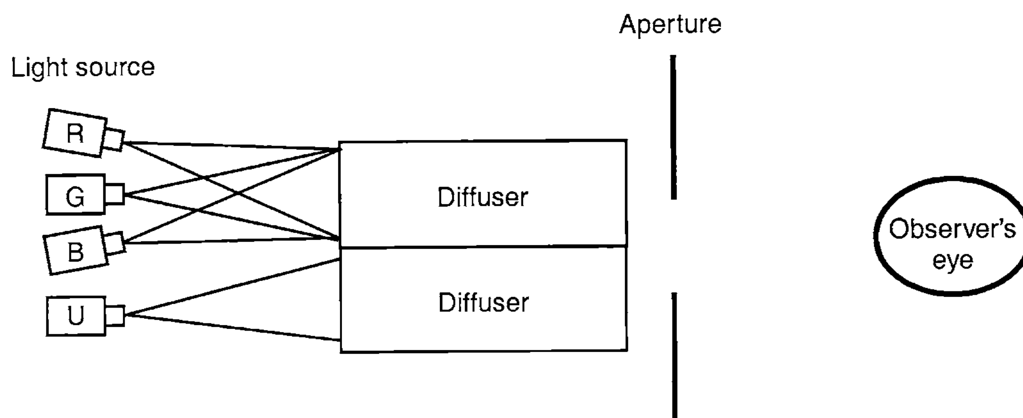


Figure 6.3: Schematic of the color matching experiment

1. The equal-energy stimulus has a uniform spectral energy distribution.

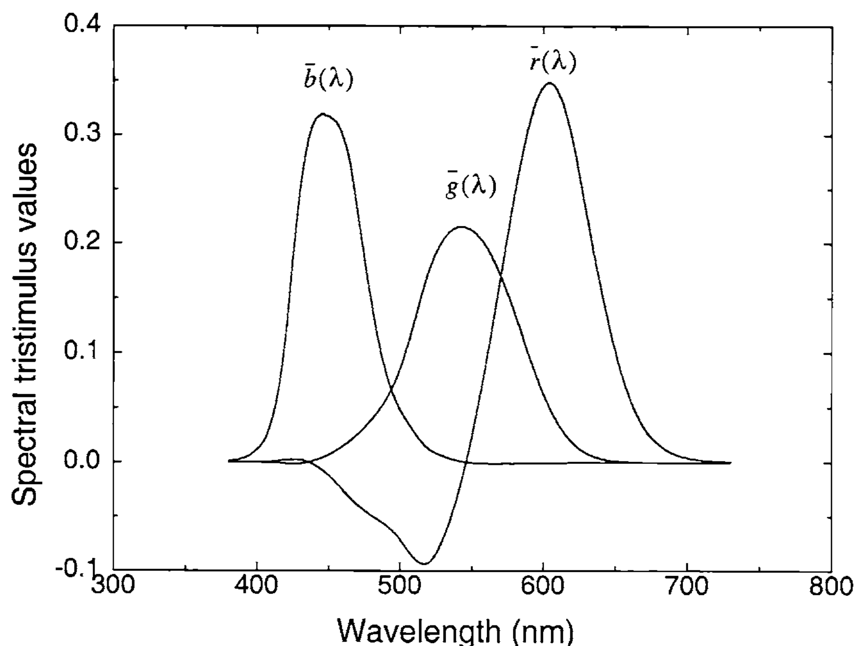


Figure 6.4: Tristimulus values vs. wavelength corresponding to primary stimuli $R/G/B$

dependent. Typical observing angles are 2° or 10°. The observer watches the output from the two side-by-side diffusers through the aperture and judges if the colors on the two half-planes are matched. The intensity of the $R/G/B$ stimuli are adjusted until color match is achieved. The intensity coefficients of the $R/G/B$ color stimuli are then recorded. These coefficients are called tristimulus values and this test is called trichromatic matching experiment. Mathematically the color matching relation can be equated as $U = R_U \cdot R + G_U \cdot G + B_U \cdot B$, where R_U, G_U, B_U represent the tristimulus values and R, G, B are the primary color stimuli. The "=" sign in the equation means the *color* of the optical stimuli on both sides is the same.

As monochromatic light source is applied for the light source U and the wavelength sweeps across the visible spectrum, the corresponding tristimulus values, $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ can be measured and denoted as functions of wavelength. Figure 6.4. shows the results from monochromatic color matching experiments. The spectral tristimulus values, $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$, are called the *color-matching functions* with primary stimuli R, G, and B.

It is desired to acquire the above color-matching functions since the data can be used to derive the tristimulus values of any other optical stimulus as long as the spectral energy distribution is known. Assuming the spectral energy distribution of a color stimulus is $P(\lambda)$, the tristimulus values for the color stimulus can be calculated from the following integration equations:

$$R_p = \int_{\lambda_1}^{\lambda_2} P(\lambda) \cdot \bar{r}(\lambda)\, d\lambda \ , \tag{6.1}$$

$$G_p = \int_{\lambda_1}^{\lambda_2} P(\lambda) \cdot \bar{g}(\lambda)\, d\lambda \ , \tag{6.2}$$

$$B_p = \int_{\lambda_1}^{\lambda_2} P(\lambda) \cdot \bar{b}(\lambda)\, d\lambda \ , \tag{6.3}$$

where the integral range is the visible spectrum $[\lambda_1\ \lambda_2]$. Applying the above method, any optical stimulus P obtains a set of tristimulus values (R_p, G_p, B_p) associated with it. If two color stimuli P_1 and P_2 have different spectral distribution but the same tristimulus values, their color will be identical and the two stimuli are called *metameric stimuli*.

An observation of Figure 6.4 shows that some tristimulus values are negative in certain wavelength regime. The negative tristimulus value means that the component of the associated primary stimulus needs to be mixed with light source U in the other diffuser in order to match the colors on both sides of the aperture. This phenomena is related with how the color information is processed, encoded, and transmitted from the photo-receptors on the retina to the brain. In fact, none of any combination of three physical color stimuli is able to generate all the perceivable colors with positive coefficients.

Though the above *R*/*G*/*B* primary stimuli provide an accurate system for color specification, this color system is not used as widely. The reason is because the negative values of the color-matching functions make it difficult to implement the color-measuring equipment such as a colorimeter in the early 20th century. The required computation power was too expensive to deal with integration with negative values. The cost would be lower if the color-matching functions are all positive. In 1931, the CIE announced a new set of tristimulus values X, Y, and Z, that can be obtained from the original tristimulus values R, G, and B with the following linear transformation:

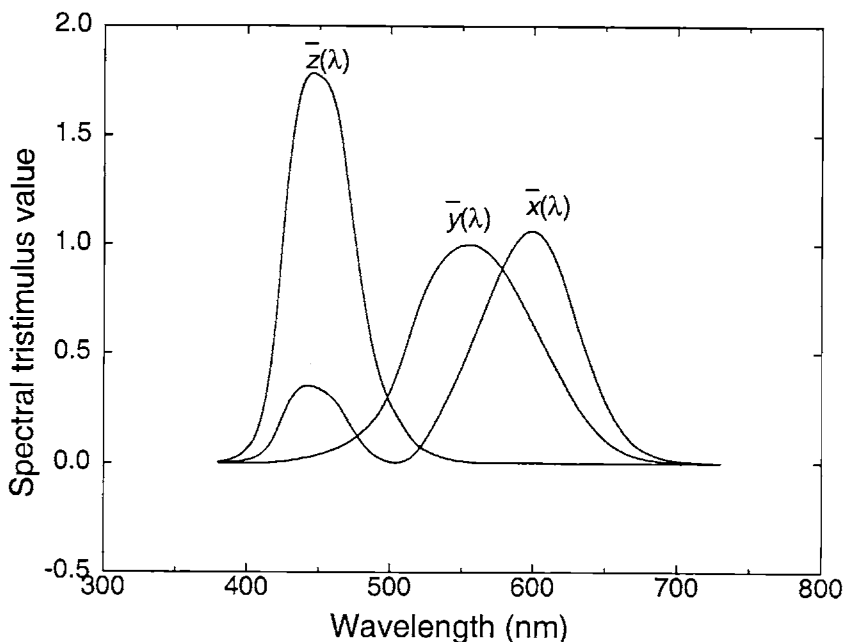


Figure 6.5: The CIE color-matching functions for the 1931 Standard Colorimetric Observer

$$X = 0.49R + 0.31G + 0.20B, \tag{6.4}$$

$$Y = 0.17697R + 0.81240G + 0.01063B, \tag{6.5}$$

$$Z = 0.00R + 0.01G + 0.99B. \tag{6.6}$$

Since color mixing is a linear mechanism, this linear transformation is valid to transfer the tristimulus space constructed with the $R/G/B$ primaries into another tristimulus space by $X/Y/Z$ primaries. This process is similar to coordinate transformation in a three-dimensional space. The color-matching functions after the transformation are denoted as $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$. This set of functions is shown in Figure 6.5 and is called the *CIE Color-matching Functions for the 1931 Standard Colorimetric Observer.* An important characteristic of these functions is that all the values are positive. In order to achieve that, the primary stimuli used in the $X/Y/Z$ space are non-physical optical stimuli. This is valid since the primary stimuli are just used as units of measurements. Except for the positive coefficients, there are other considerations associated with the selection of the $X/Y/Z$ primaries. First, the primary stimuli X and Z contain zero luminance. That means the tristimulus value Y is proportional to the luminance of the color stimulus. If the power spectral distri-

bution of a color stimulus is $P(\lambda)$, the luminance can be calculated from $Y = K \cdot \int_{\lambda_1}^{\lambda_2} P(\lambda) \cdot \bar{y}(\lambda)\, d\lambda$, where K is a conversion factor. This integral provides both the luminous information and one of the color coordinates, which saves computation power as both properties are generally needed. Second, the magnitudes of $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are scaled so that the tristimulus values for an equal-energy stimulus are the same.

6.1.3 Chromaticity Diagram

The tristimulus values, X, Y, and Z, define the color coordinates of an optical stimulus in a three-dimensional color space. Sometimes this three-dimensional space is difficult to visualize. To provide a two-dimensional representation of the color space, chromaticity diagrams are developed. For the $X/Y/Z$ color space we can do the following calculation:

$$x = \frac{X}{X+Y+Z}, \tag{6.7}$$

$$y = \frac{Y}{X+Y+Z}, \tag{6.8}$$

$$z = \frac{Z}{X+Y+Z}. \tag{6.9}$$

Since $x+y+z=1$, there are only two independent dimensions for the three parameters. Normally (x, y) are the used parameters. After the above normalization, the luminance information is suppressed and the color coordinates on the (x, y) plot contains only chromaticity information, which is the origin of its name. In 1931, CIE announces the chromaticity diagram of the *CIE 1931 Standard Colorimetry Observer*. The diagram is shown in Figure 6.6. The coordinates of the monochromatic light on this diagram are calculated from the spectral tristimulus values in Figure 6.5. As the wavelength sweeps through the visible spectrum, the coordinates forms a horse-shoe shape curve. The straight line connecting the end points of the shortest and longest wavelengths is called the *line of purple*. The two curves form a closed area, which is the color gamut. Each coordinate on the gamut represents a specific color. The color gamut contains all the color coordinates of the physical color stimuli. That is because any physical color stimuli can be decomposed into a linear combination of monochromatic stimuli with positive coefficients. As the property of linearity is applied, the color coordinates of any color stimulus must also be a

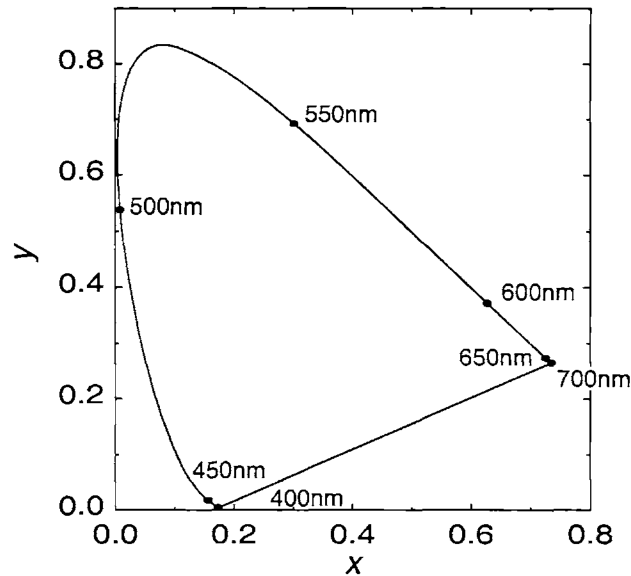


Figure 6.6: The CIE 1931 Chromaticity Diagram

linear combination of the coordinates of the monochromatic light. Therefore, it must be enclosed by the color gamut.

The CIE 1931 Chromaticity Diagram is useful in locating the color coordinates for color mixing. For example, if one component of color A is mixed with color B with the same amount of luminance, the resultant color coordinates will be located at the mid-point between point A and point B.

One problem with the CIE 1931 Chromaticity Diagram is that the color distribution on the diagram is not uniform. Often times there is a necessity to do color processing on a uniform chromaticity diagram. Because of that, the *CIE 1976 Uniform Chromaticity Scale (UCS) Diagram* was announced. The color coordinates on the diagram can be calculated from

$$u' = \frac{4X}{X + 15Y + 3Z}, \text{ and} \tag{6.10}$$

$$v' = \frac{9Y}{X + 15Y + 3Z}. \tag{6.11}$$

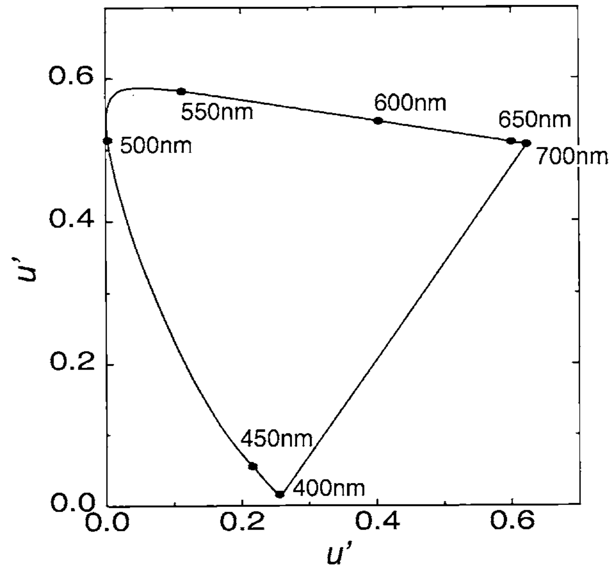


Figure 6.7: The CIE 1976 Uniform Chromaticity Scale Diagram

This diagram is shown in Figure 6.7. Since the color distribution on this diagram is relatively uniform, the distance between two coordinates can be used as a measurement of color difference. This property will be used latter in this chapter.

6.1.4 Color Analysis Functions

The spectral responses of the color pixels are used as the color analysis functions to identify the color coordinates of optical stimuli. The responses are not solely determined by the color filters but also the spectral response of the silicon sensors. Usually the response curves are far from ideal curves such as the ones shown in Figure 6.5. Some color correction operations are usually required to transform the color signal to a standard color space. The topic on the optimization of color correction algorithm is beyond the scope of this thesis. Here, the color analysis functions are assumed to be $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ to simplify the calculation. Due to the simplification, the color coordinates of the stimuli can be easily transformed between standard color spaces and chromaticity diagrams according to the transformation matrix developed in color science. In real cases, the transformation matrix needs to be calibrated and corrected.

6.1.5 Color Imagers and Cross-talk

Cross-talk induces signal exchange between neighboring pixels. If this happens to an imager with a mosaic color filter pattern as shown in Figure 6.2, the color signal will be distorted and the color performance of the imager will be affected. The rest of this chapter is devoted to quantitatively characterizing the degradation of color performance of the imagers due to cross-talk.

6.2 Mechanism of Pixel Cross-talk

There are two possible cross-talk mechanisms. One is the electronic cross-talk; the other is the optical cross-talk. Our discussion focuses on the former since it's the major component according to the experiment. The mechanism of electronic cross-talk is identified to be mainly from minority carrier diffusion, which is described below.

6.2.1 Minority Carrier Diffusion

Minority carrier diffusion is considered the major cross-talk mechanism. The mechanism is schematically shown in Figure 6.8. This figure shows a cross-section diagram of a typical CMOS imager with n-type photodiode sensors. As photons impinge on Si, photo-carriers are generated and the generation profile depends on the wavelength of the photons. Long-wavelength photons tend to generate carriers deeper in the substrate and the carriers from the short-wavelength photons are generated near the surface. Those carriers generated in the depletion region are driven by the electric field and get collected by the photodiode. The carriers generated deep in the P^+-substrate are recombined efficiently due to the high hole concentration. The carriers generated outside the depletion region in the epi

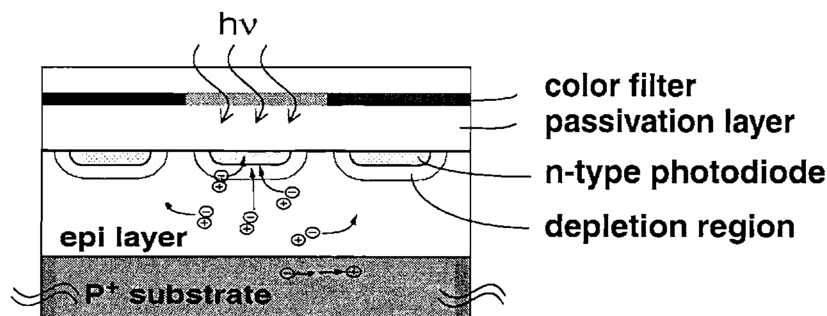


Figure 6.8: Minority carrier diffusion mechanism

layer could diffuse either back to the "correct" photodiode or to the adjacent photodiodes. Those carriers diffusing to the adjacent diodes are the components that contribute to cross-talk.

This cross-talk mechanism induces carrier exchange between adjacent pixels. For a black-and-white imager, the effect reduces the effective resolution. For a color imager, not only the resolution is degraded, the color signal is further modulated since the amount of carriers diffusing from red to green pixels is not the same as the one in the opposite direction. And vice versa for the other color pixel pairs. Thus, the color signals are distorted.

6.3 Characterization of Cross-talk

6.3.1 Test Pattern Design

In order to quantify the cross-talk phenomena, a test pattern is specifically designed to measure the cross-talk. A 64 x 64 3-T NW/Psub active pixel sensor (APS) array is fabricated using a non-silicide source/drain 0.25-μm STI CMOS process. The utilized wafer contains an 8μm lightly-doped p-type epi layer (~10^{15}cm^{-3}) on top of a heavily-doped p-type substrate. The NW/Psub photodiode structure is utilized in this experiment for its superior performance over the N^{+}/PW/Psub photodiode in this process [19]. Typical characteristics of the pixel array are shown in Table 6.1. The pixel size is 4 x 4 μm^2. The on-

Process	0.25μm non-silicide source/drain STI CMOS process
Format	64 x 64
Pixel size	4 x 4 μm^2
Sensor type	NW/Psub photodiode
Fill factor	~ 25%
Sensitivity	1.7 V/lx/sec
Leakage current	0.45 fA/pixel/sec @ 300K
Conversion gain	33 μV/electron

Table 6.1: Characteristics of the APS array in the experiments

chip circuitry for row/column addressing and analog signal buffering is the same as described in Chapter 4.

The specialty about this chip is that all pixels, with the exception of the center pixel,

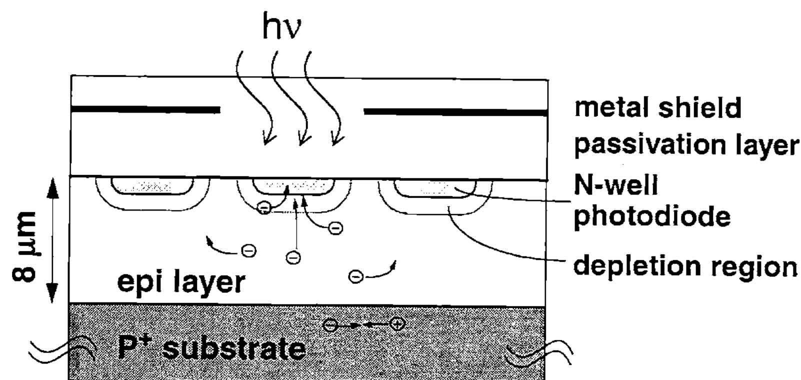


Figure 6.9: Cross-section of the cross-talk test pattern

are optically shielded with the metal-3 layer. The cross-section diagram of the chip is shown in Figure 6.9. By measuring the signal of the exposed pixel and the adjacent pixels, cross-talk is characterized.

The utilized light source is a monochromator with tungsten halogen light source. The color temperature is roughly 3000K. The wavelength of the output light beam is varied to characterize the spectral response of cross-talk.

6.3.2 Experimental and Simulation Results

The cross-talk measurement is performed with optical wavelengths varying from 400nm to 900nm. Varying optical intensity shows little impact on the cross-talk data, which suggests that the cross-talk mechanism is linear. Figure 6.10 shows the cross-talk value versus wavelength for the nearest neighbors of the exposed pixel. The three curves in the figure represent the cross-talk for the direct vertical, horizontal, and diagonal neighbors. The cross-talk value is defined as the ratio between the signal of the shielded pixels and that of the exposed pixel. Clearly the value is higher at longer wavelengths. Since the long-wavelength photons penetrate deeper into Si, the wavelength dependence of cross-talk suggests that minority carrier diffusion is the major mechanism. Another evidence is that the diagonal pixel shows lower cross-talk value than that of the vertical and horizontal pixels, revealing that the separation distance between the shielded pixel and the exposed pixel plays an important role. This can also be explained by the minority-carrier diffusion mechanism. On the other hand, the cross-talk discrepancy between the vertical and horizontal

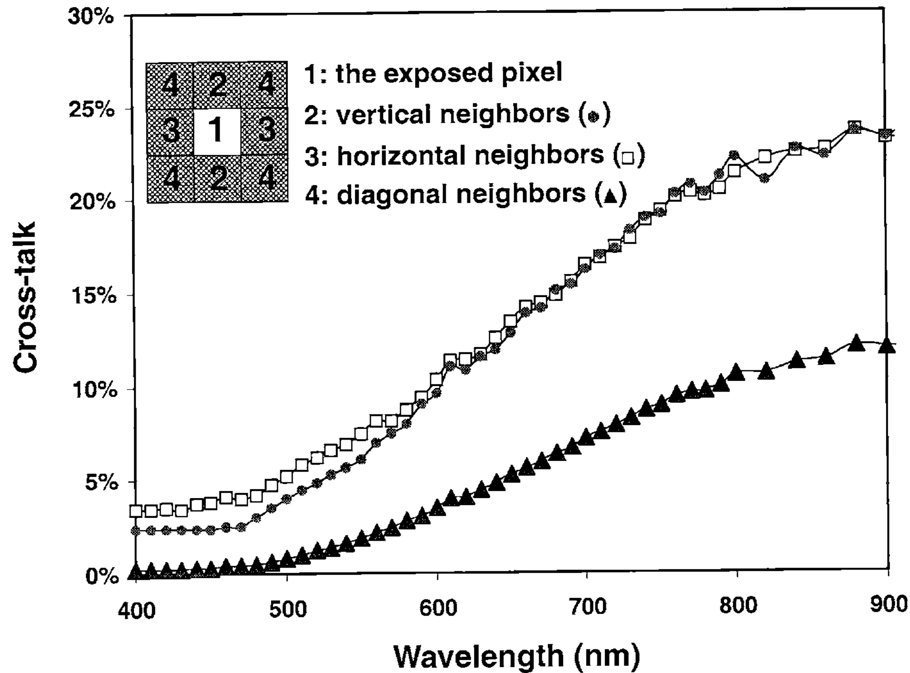


Figure 6.10: Measured cross-talk versus wavelength for the adjacent pixels along the horizontal (empty square), vertical (solid circle), and diagonal directions (triangle)

neighbors at short wavelengths is due to pixel layout asymmetry along the two axes. Beyond certain wavelength, the generated carriers are deep enough so the carriers see the same environment in all directions. Thus, the two cross-talk curves merge.

A 2-D device simulation is performed to confirm the minority-carrier diffusion mechanism. The simulation utilizes the doping profile estimated from the spreading resistance data in the process. The simulated cross-talk is shown in Figure 6.11. The bumps on the curve are due to the utilized empirical Si absorption coefficient, which is not a monotonic function with wavelength. The trend of the simulated cross-talk matches well with the measured data — the cross-talk is higher at longer wavelengths. As expected, the simulated cross-talk is generally higher than the measured cross-talk since the impact of optical blockage and carrier absorption by the pixel transistors is not considered in the simulation. In the simulation, the non-zero cross-talk at short wavelengths is due to the larger exposed window than the size of the photodiode. As the window size is shrunk to the size of the photodiode, the cross-talk value between 400 to 450nm decreases to nearly zero. At short

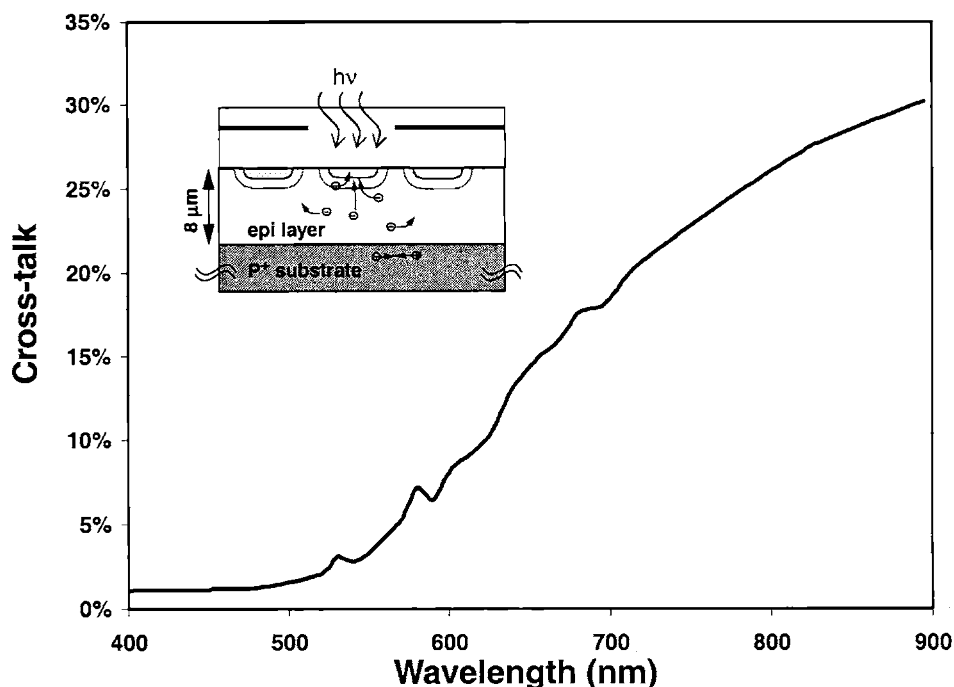


Figure 6.11: Cross-talk versus wavelength from the 2-D device simulation

wavelengths, the measured cross-talk value is higher than the simulated value. This can be attributed to optical cross-talk mechanisms such as the obliquely injected light from optical diffraction effect or metal layer reflection. However, the cross-talk from these optical effects, according to the measurement, is much lower (~ 3 – 4%) compared with the cross-talk from the minority-carrier diffusion.

The quantum efficiency (Q.E.) of the exposed pixel is measured and the data is shown in Figure 6.12. The three curves correspond to different measures of Q.E. The curve with triangle symbols denotes the Q.E. of the exposed pixel. The one with solid circles represents the aggregate Q.E. of the exposed pixel plus the cross-talk components from the eight nearest pixels. The curve with empty squares shows the Q.E. of the pixels on an array with no hollow metal shield pattern. There are two observations in the measurement. First, the aggregate Q.E. of the exposed pixel plus the eight nearest neighbors (solid circle) matches with the Q.E. taken from the array with no metal shield pattern (empty square). This result shows that the total amount of photo-carriers is conserved and thus validates the cross-talk measurement. The peaks of the two curves differ by roughly 20nm due to

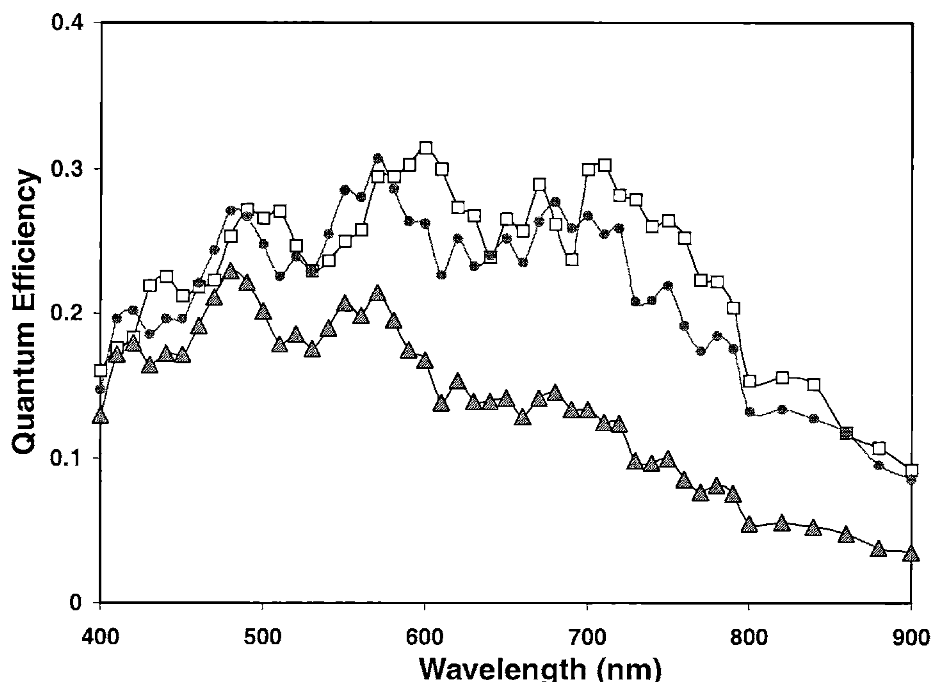


Figure 6.12: Quantum efficiencies (Q.E.) of the pixels. Triangle: QE of the exposed pixel; solid circle: QE of the exposed pixel plus cross-talk components; empty square: QE of the pixels for the array with no metal shield pattern

the interference effect by the various dielectric thickness on the two chips. Second, the large discrepancy between the Q.E. of the exposed pixel (triangle) and the aggregate Q.E. from the nine pixels (solid circle) shows that significant carrier diffusion happens for long-wavelength photons. This phenomena also reveals that a large percentage of Q.E. measured with the typical way actually comes from cross-talk.

6.4 Amplification of Color Error due to Cross-talk

Imagers produce pixel-to-pixel color mismatch even if a uniform color stimulus is projected on the imaging plane. This mismatch is due to all kinds of noise sources that induce non-uniform signal output on the frame. With the cross-talk phenomena, the color mismatch is amplified. The amplification factor will be estimated under several reasonable assumptions.

6.4.1 Color Error

The measured cross-talk data is utilized to study the amplification effect of the pixel-to-

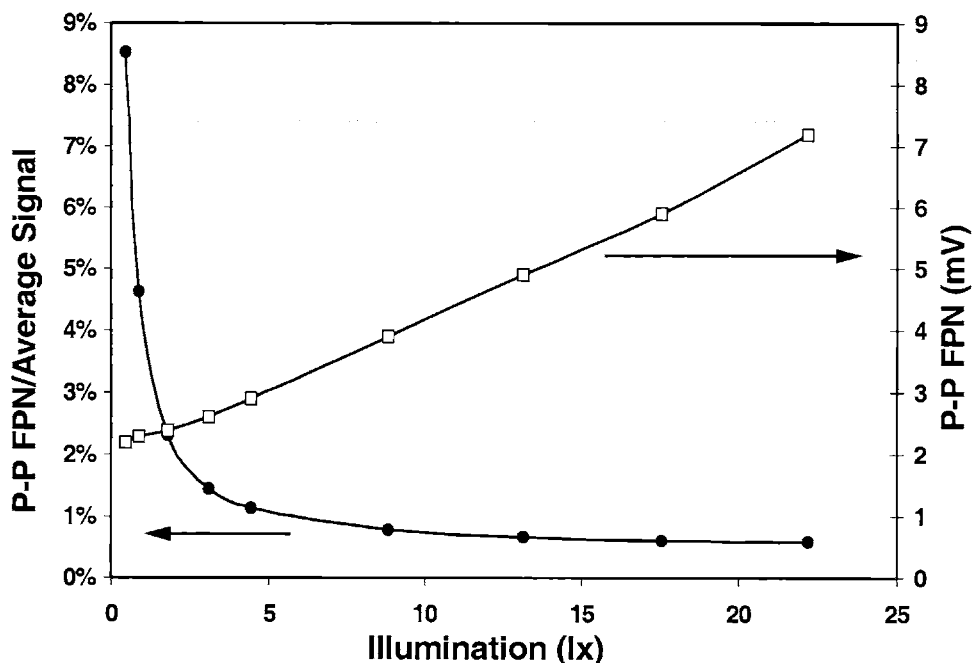


Figure 6.13: P-P FPN (empty square) versus illumination and the normalized P-P FPN with the average signal (solid circles) versus illumination

pixel color mismatch. Conceptually, color mismatch is the result of several noise mechanisms, which include temporal and spatial noise. However, the mismatch produced by fixed pattern noise (FPN) usually dominates and is more significant than that produced by random noise. That is because color mismatch by FPN cannot be averaged out with fast refreshed image frames by human eyes. FPN can be classified as column-to-column fixed pattern noise (C-C FPN) and pixel-to-pixel fixed pattern noise (P-P FPN). C-C FPN is typically caused by circuit mismatch and is easier to fix. (as described in Chapter 4) P-P FPN is typically determined by process characteristics and is difficult to correct without a frame buffer. In order to quantify the P-P FPN, an un-shielded pixel array is utilized so that the amount of cross-talk is uniform across the array and has no impact on the P-P FPN measurement. The measured P-P FPN and its ratio with the average signal at various illumination is shown in Figure 6.13. The integration time is 1/30sec. According to the characterization result depicted in Section 4.4.5, the non-zero P-P FPN at low illumination is mainly due to dark current mismatch of the pixels; the linear response of P-P FPN versus illumination is due to the mismatch of the photo-sensing area, the capacitance mis-

match, or the gain mismatch of the in-pixel source followers. The ratio between the P-P FPN and the average signal defines a normalized noise level. This level increases significantly as the illumination on the imager plane falls below 5 lux. Therefore, the color mismatch is more critical under low illumination.

6.4.2 Color Error Amplification

The color mismatch is further amplified by the cross-talk mechanism. To estimate the amplification factor, two assumptions are made. First, it is assumed that the R/G/B Bayer filter pattern is utilized. Second, the spectral responses of the R/G/B pixels are assumed to be proportional to the color-matching functions of the CIE 1931 standard colorimetric observers as shown in Figure 6.5. The first assumption is based on a well-known example of color filter configuration; the second is made to simplify color coordinate mapping from the pixel outputs onto a standard color space. In real cases, an optimized transformation matrix is required to perform the color space mapping.

Based on the assumptions and the measured cross-talk data, the spectral responses for the R/G/B pixels with the cross-talk effect can be calculated. In the calculation, cross-talk from the eight nearest neighbors are considered. The calculation procedure is described with the following example.

According to the characterization result, the cross-talk phenomena for the nearest neighbors can be described by a 2-D discrete-space impulse response function, which can be written as a 3 x 3 matrix as follows:

$$C(\lambda) = \frac{1}{K} \cdot \begin{bmatrix} c_{11}(\lambda) & c_{12}(\lambda) & c_{13}(\lambda) \\ c_{21}(\lambda) & 1 & c_{23}(\lambda) \\ c_{31}(\lambda) & c_{32}(\lambda) & c_{33}(\lambda) \end{bmatrix}, \tag{6.12}$$

where the row and column index represents the position of the pixel; c_{12} and c_{32} are the vertical cross-talk value; c_{21} and c_{23} are the horizontal cross-talk values; all the diagonal coefficients correspond to the diagonal cross-talk values; and K is the summation of the nine matrix coefficients. To further simplify the calculation, the vertical cross-talk value is assumed to be the same as the horizontal cross-talk value.

As the Bayer pattern shown at the upper right corner of Figure 6.14, a blue pixel is surrounded by four green pixels as adjacent neighbors and four red pixels as diagonal neighbors. Therefore, the spectral response of the blue pixel under cross-talk mechanism becomes

$$B'(\lambda) = \frac{1}{K} \cdot B(\lambda) + \frac{4 \cdot c_{12}(\lambda)}{K} \cdot G(\lambda) + \frac{4 \cdot c_{11}(\lambda)}{K} \cdot R(\lambda), \tag{6.13}$$

where $B(\lambda)$, $G(\lambda)$, and $R(\lambda)$ represent the spectral responses of the B/G/R color pixels with zero cross-talk. Similar equations can be achieved for the green and red pixels:

$$G'(\lambda) = \frac{1 + 4 \cdot c_{11}(\lambda)}{K} \cdot G(\lambda) + \frac{2 \cdot c_{12}(\lambda)}{K} \cdot G(\lambda) + \frac{2 \cdot c_{12}(\lambda)}{K} \cdot R(\lambda), \text{ and} \tag{6.14}$$

$$R'(\lambda) = \frac{1}{K} \cdot R(\lambda) + \frac{4 \cdot c_{12}(\lambda)}{K} \cdot G(\lambda) + \frac{4 \cdot c_{11}(\lambda)}{K} \cdot B(\lambda). \tag{6.15}$$

The physical meaning of the above equations is the following. Due to cross-talk, the blue pixel exchanges photo-carriers with its neighbors. The amount of carrier exchange is further modulated by the spectral responses of the corresponding color filters. As a result, the amount of the net exchanged carriers could be positive or negative depending upon the wavelength. At the wavelength corresponding to the peak response of the blue pixels, the response is lowered since the out-going carriers are more than the in-coming ones. Those out-going carriers end up increasing the response of the green and red pixels at this wavelength. Similar phenomena happens to the green and red pixels. Eq.(6.13) to eq.(6.15) are utilized to calculated the distorted spectral responses of the color pixels and the results are shown in Figure 6.14. The ideal spectral response curves without cross-talk are also shown for comparison.

The spectral response curves with and without cross-talk are used as the color analysis function and inserted in eq.(6.10) and eq.(6.11) to plot the color gamuts that can be analyzed by the imager. The color gamuts are plotted on the 1976 CIE Uniform Chromaticity Scales chart and shown in Figure 6.15. The reason for choosing the UCS chart is because color distribution is relatively uniform on this chart so the distance of two color coordinates on this chart can be used to present the color error.

The two closed areas in Figure 6.15 show the color gamuts for the ideal color pixels

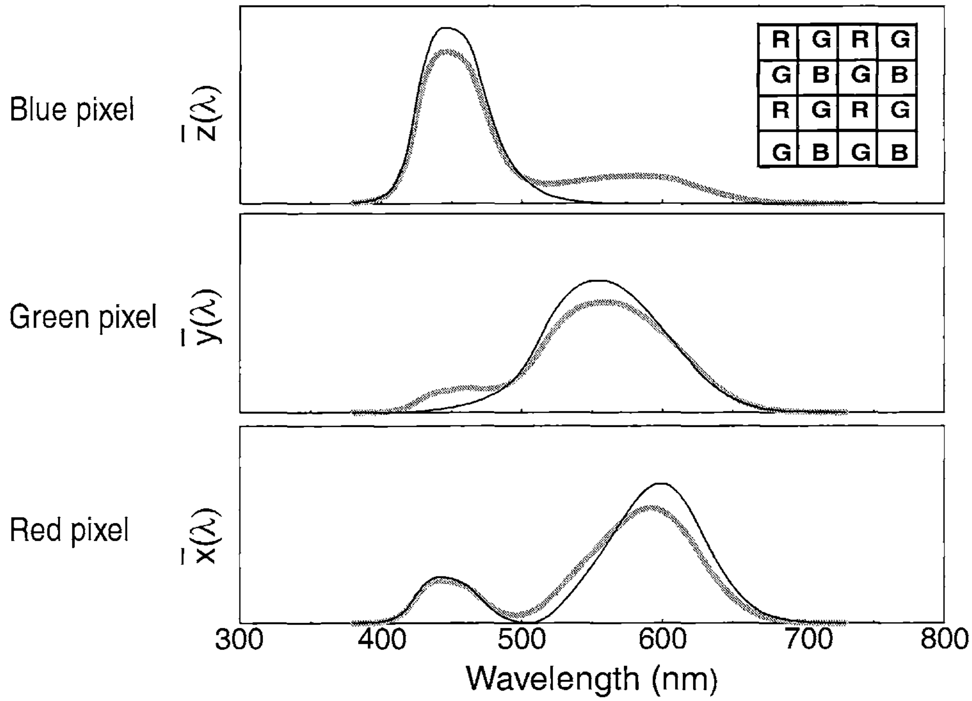


Figure 6.14: Spectral responses of the R/G/B pixels with (thick line) and without cross-talk (thin line)

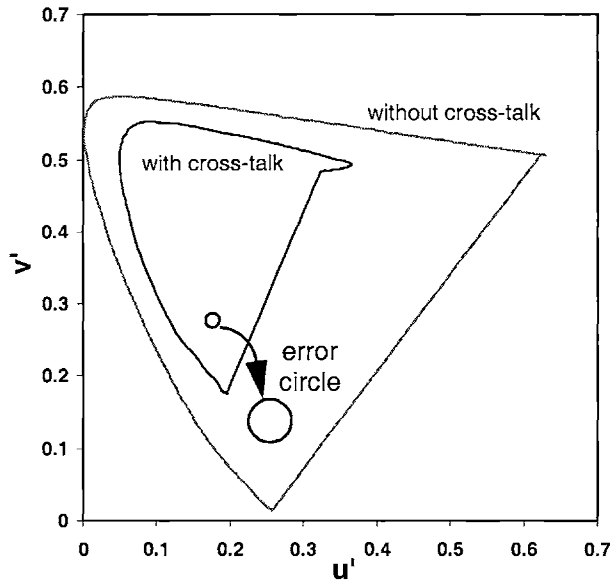


Figure 6.15: Color gamuts with and without cross-talk on the 1976 CIE Uniform Chromaticity Scales chart

and the pixels with the cross-talk effect. Due to cross-talk, the color gamut shrinks. According to the discussion in Section 6.1.1, a standardized color space is necessary in order for the displaying systems to correctly present color. That means an ideal color correction process is required to map the distorted color coordinates back to the original. This operation is essentially an amplification process, as the error circle expands after the correction as shown in Figure 6.15. Since the area ratio of the two color gamuts is approximately 3:1, this color correction process will amplify pixel-to-pixel color error due to intrinsic pixel noise, on average, by a factor of $\sqrt{3}$.

The above derivation demonstrates the estimation of the color-error amplification factor using the CIE 1976 UCS chart for ideal color pixel responses. For practical color pixel responses, if the optimized transformation matrix is known, the same evaluation procedure can be applied to calculate the color-error amplification factor. Further information on color filter considerations can be found in [55, 56]. On the other hand, a uniform 3-D color space, instead of a 2-D CIE 1976 UCS chart, can be used as the processing color space. In contrary to the 2-D case, the luminance error information will be contained in the 3-D color space. Information about some commonly used uniform 3-D color spaces, such as CIELUV or CIELAB color spaces, can be found in [51, 52, 57].

6.5 Layout and Process Improvements on Cross-talk

Several approaches are proposed to minimize the cross-talk. First, a shallow epi wafer forces the carriers generated deeper in the substrate to recombine efficiently and thus reduces the probability for the carriers to diffuse to other pixels. This approach is more effective for long-wavelength photons and consequently reduces the Q.E. for red light. For imagers not requiring high red-light response, this approach should be feasible. Second, a deep p-type implant can be used to generate an internal electric field to guide the photo-carriers back to the sensor and thus reduces cross-talk. This approach is adopted in [18] and proven to be effective. One of the concerns for applying such a deep p-type implant is the potentially increased leakage current. If thermal annealing is properly performed, this approach is also viable. As the electronic cross-talk is minimized, the problem of optical cross-talk will emerge. A metal shield pattern with opening windows of the same size as the photo-sensors can be utilized to assures all photons falling on the sensors. This method

minimizes the number of photon-generated carriers near the pixel boundary and is effective for reducing cross-talk at all wavelengths. However, the Q.E. of the pixel will be reduced since the incoming photon flux is lowered. A micro-lens technology, which is utilized to focus photons at the center of the sensors, has a similar cross-talk reduction function without sacrificing the photon flux. The drawback is higher process complexity and cost. In general, the approaches for minimizing cross-talk should be considered from all aspects of imager performance. It is certainly undesired to severely degrade other sensor parameters just for fixing the cross-talk.

6.6 Summary

The cross-talk mechanism of CMOS image sensors is studied using a 64 x 64 NW/Psub APS array with a hollow metal shield pattern. The pixel size is 4 x 4 μm^2 and the array is fabricated using a 0.25-μm STI CMOS process with 8μm thick epi wafers. Minority carrier diffusion is identified to be the major cross-talk mechanism by simulation and experiment. Due to cross-talk, the pixel-to-pixel color mismatch is degraded. The impact is more severe under low illumination conditions. The degradation factor is estimated with reasonable assumptions from color science. Finally, process and layout improvements for minimizing cross-talk are proposed and discussed.

Chapter 7

Pixel Size Scaling and Optical Lens Design

7.1 Scaling Trend of Pixel Size

There exists a strong incentive of pushing the process limit towards achieving a smaller pixel size. The driving force comes from two major objectives. First, a better image quality is desired by packing more pixels within a certain chip size to improve imager resolution while maintaining a reasonable chip cost and utilizing lenses with the same optical format. Second, a smaller pixel size reduces chip size for a specific imager resolution, producing higher chip count for each wafer. Thus, the chip cost can be reduced. The first objective aims towards a better imager quality and the second aims at the cost.

With the down-scaling trend of sub-micron CMOS processes, pixel size reduction can be more easily achieved. However, the aforementioned objectives can not be achieved effectively without the compatibility of optical lenses. The reduced pixel size poses challenges on the resolving ability of the lens systems. The financial benefit from reducing pixel size can only be justified if the reduced cost from pixel size reduction is more than the increased cost for a better lens system.

In this chapter, the fundamentals of lens designs and the principles of optical simulation tool is briefly described. Since it is not possible to cover all the details on lens design, only general concepts are provided. A computer simulation tool is utilized to estimate the minimum resolvable pixel size for different lens systems. The comparison is made between lens systems with different lens elements under certain specifications and restrictions. The results provide a guideline for the requirement on lens performance in order to keep up with pixel size reduction.

7.2 Fundamentals of Lens Design

7.2.1 Focal Length, F-number, and Field of View

The function of a lens is to collect the light emitting or reflecting from the imaged object,

apply proper bending for the diverging light rays, and form a clear two-dimensional image on the imaging plane, where an imager chip or a chemical film is located. An ideal lens produces precise point-to-point correspondence between the object and its image. Unfortunately, such an ideal lens does not exist. Each lens has its resolution limit due to physical imperfections [58, 59] such as diffraction, monochromatic aberrations, chromatic aberration, and mechanical mis-alignment. Some of the imperfections can be corrected by assembling various lens elements into a lens system to alleviate the effects. The idea is to minimize the overall imperfection with lens elements equipped with positive and negative imperfection values. The imperfection values associated with different lens elements cancel with each other and therefore a high-performance lens system is achieved. For such multiple-element lens systems, the design factors include the number of elements, lens material (with different refractive index), surface radius, lens separation, and so on. Generally, a better performance can be achieved if more elements are used. That is because more degrees of freedom are granted to minimize the overall imperfection. However, realistic factors such as cost and physical dimension of the lens system usually restrict the number of elements to be implemented.

For camera lenses, the fundamental parameters include effective focal length (EFL), f-number, and field of view (FOV). These parameters are briefly described as follows:

1. Effective focal length (EFL) is the most fundamental parameter for any lens system. It determines the on-axis focal point and the focal plane where the imager chip should be placed. If the object is at an infinite distance, the focused image will be located at the focal plane. But if the object is at a finite distance, the relative position of the image plane will change. The focus ring on a camera lens needs to be adjusted to obtain a clear image. The focal length also determines the amplification factor of the image. The focal length is proportional to the amplification factor of the image and inversely proportional to the square root of the light intensity. F-number and FOV are both functions of the EFL.

2. F-number is also denoted as f/#. It is defined as the EFL divided by the diameter of the clear aperture of the lens. $1/(\text{f-number})^2$ indicates a measurement of the light intensity on the imaging plane under a certain illumination condition. The f-number of a lens system is usually adjustable with an aperture. The smallest f-number is reached when the aperture is fully open. At this f-number, the highest light intensity is achieved at the image

plane. Lenses with small f-numbers are difficult to fabricate since a large curvature is required at the lens surface and the resolution of the curvature is more difficult to control. Besides the difficulty of fabrication, distortion and aberration are usually larger for lenses with small f-numbers.

3. Field of view (FOV) determines the viewing angle of the scene projected onto the imaging device. In order to mimic a normal viewing angle for human vision, the FOV of camera systems typically ranges from 40° to 50°. Some of the photographic lenses have larger (wide-angle lens) or smaller (zoom lens) FOV depending on the focal length. Mathematically, the FOV is determined by the size of the sensing device and EFL. It can be calculated from the equation: $2\tan^{-1}\left(\frac{H_i}{EFL}\right)$, where H_i is the radius of the imager array.

7.2.2 Lens Optics

In modern optical engineering, the behavior of light in a lens system is usually described by geometrical optics, in which light is treated as rays spreading from the source. As the rays hit an interface of the lens, their trajectories change according to the incident angles, the curvature of the surface, and the refractive index on both sides of the interface. This phenomena of ray bending is physically described by Snell's law of refraction: $n_1 \cdot \sin\phi_1 = n_2 \cdot \sin\phi_2$. The schematic of the refraction phenomena is shown in Figure 7.1 [58]. The guideline on lens design is to choose the right lens materials and design the right curvatures of the lens surfaces so that light rays originated from a point source merge to a corresponding point on the image plane. The schematic is shown in Figure 7.2.

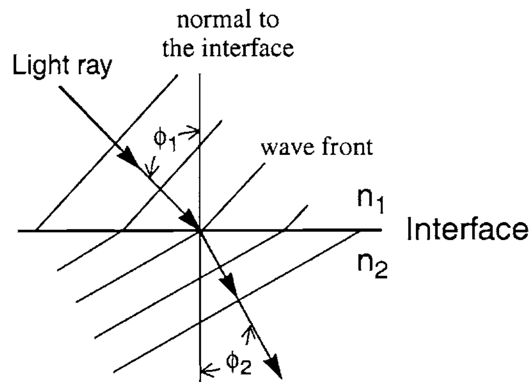


Figure 7.1: Ray bending according to Snell's law

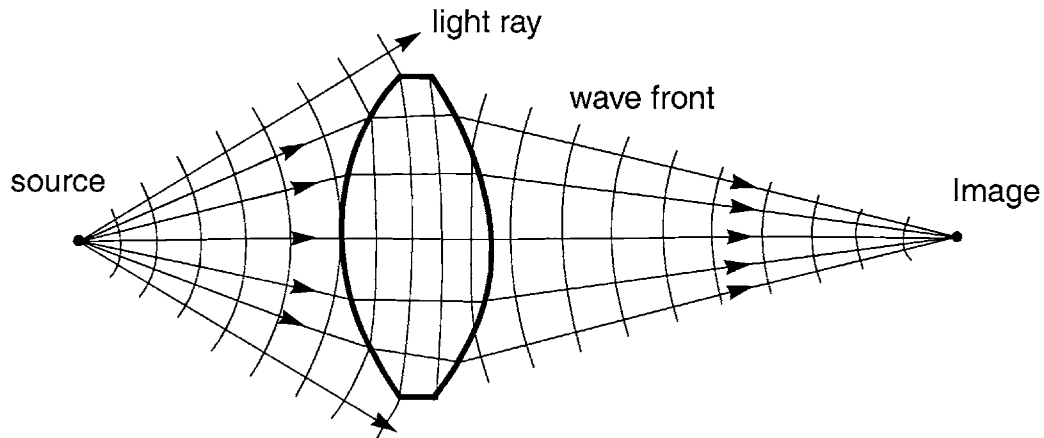


Figure 7.2: Light rays and wave front in a lens

Unfortunately, point-to-point correspondence between the object and the image is not viable in reality. One of the reasons is because the refractive index of lens material is a function of wavelength. The focal point on the image plane for one wavelength is different from the focal point for another wavelength. The spectral components of light spreads out spatially on the image plane and the resolution is degraded. This phenomena is called chromatic aberration [59]. Even for a monochromatic light source, such ideal mapping does not exist either. That is because the degrees of freedom on the design parameters is simply insufficient to satisfy all the boundary conditions requested by the point-to-point mapping. As a result, lens design essentially becomes a compromise between all the imperfections under certain specifications such as EFL, f-number, and FOV. Other considerations such as the available materials, manufacturing ability, and cost should also be taken. Generally, if more lens elements are allowed, more degrees of freedom are granted for the system to minimize one or several of the critical imperfections.

The principles of geometrical optics works properly as long as the resolution of a lens system is not diffraction-limited. The diffraction effect becomes important as the aperture of the lens closes down to a certain level. In a computer simulation tool for optical systems, the diffraction effect is considered so that the outcomes reflect the actual behavior of the lens system.

7.3 Computer Simulation on Lens Systems

A computer simulation tool[1] is utilized to evaluate the resolution of the lens system by computing the system's modulation transfer function (*MTF*), which is a widely adopted parameter for evaluating lens resolution. The calculation of *MTF* is briefly described in this section. The tool is utilized to evaluate fixed focal length lens systems with a different number of lens elements.

7.3.1 Modulation Transfer Function

The concept of the modulation transfer function is analogous to the frequency response for a linear system reacting to a sinusoidal input. The difference between the two cases is that one is in the temporal domain and the other is in the spatial domain. The idea of modulation can be described with Figure 7.3. Assuming the light intensity emitting from the object is of a bright/dark sinusoidal pattern as shown in the figure, the modulation of this pattern is calculated as $M_o = (E_{max_o} - E_{min_o})/(E_{max_o} + E_{min_o})$. A similar definition applies to the light intensity of the image pattern and the modulation is calculated as $M_i = (E_{max_i} - E_{min_i})/(E_{max_i} + E_{min_i})$. From the above two equations, *MTF* is defined as the modulation of the image divided by the modulation of the object, which is equal to M_i/M_o. As the spatial frequency of the bright/dark pattern varies, *MTF* changes accordingly. Generally, *MTF* is a function of the spatial frequency.

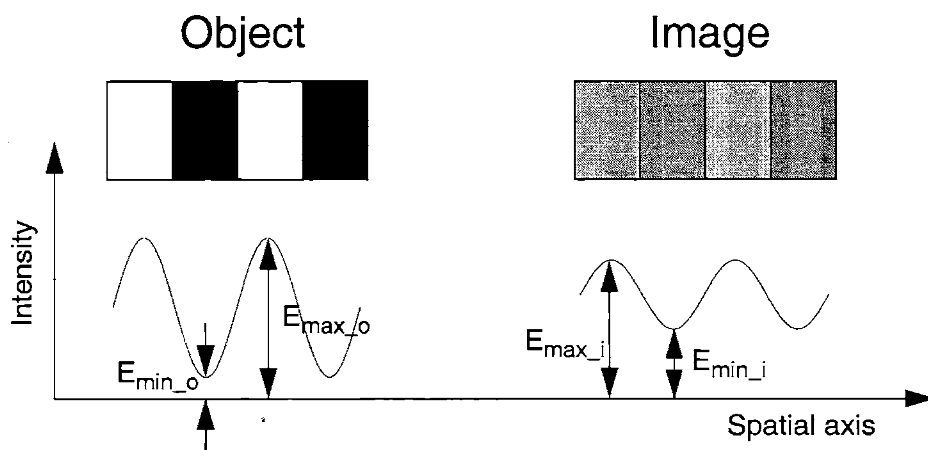


Figure 7.3: Object and image of a sinusoidal pattern in spatial domain

1. The utilized simulation tool is *Code V* from Optical Research Associates [60].

To estimate *MTF*, geometrical ray-tracing technique [58, 59] is an effective tool for a lens system operated far from the diffraction-limited region. The first step is to calculate the point spread function, which is the energy spreading distribution on the image plane associated with a point source on the object plane. Consider a bundle of rays emanating from a single object point and traced through the lens system to the image plane. The arrangement of the rays can be done by dividing up the entrance pupil of the optical system into many small grid areas. Figure 7.4 shows an example of the rectangular placement for the ray intersections with the entrance pupil. Each ray tracing through the center of the grid represents the optical energy passing through this small area. Since the curvatures of the surfaces and the refractive index of the lens materials are pre-determined, the trajectory of each ray can be fully described by the ray-tracing technique. Due to the imperfections of the lens, the intersections of the rays on the image plane will form a pattern. This pattern is a *spot diagram*. Figure 7.5 shows the spot diagrams associated with an object plane at infinite distance. The top diagram is for rays injecting in a 14° incident angle with the lens axis; the bottom diagram is for 0° incident angle. The pattern of the spot diagram, as shown in the figure, depends on the incident angle.

As more rays are traced, the spot diagram will more accurately represent the energy distribution of the point source on the image plane. The density of the spots on the 2-D image plane forms a 3-D function, which is the *point spread function*. The integrations of the point spread function along x-/y-axis directions are the *line spread functions*, which represent the optical energy distribution of the image for a line source with infinitely nar-

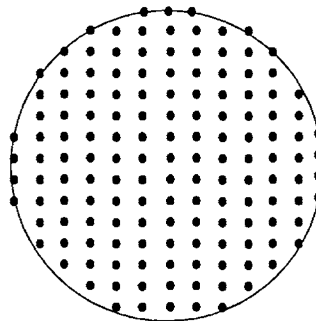


Figure 7.4: Intersection points of rays with entrance pupil

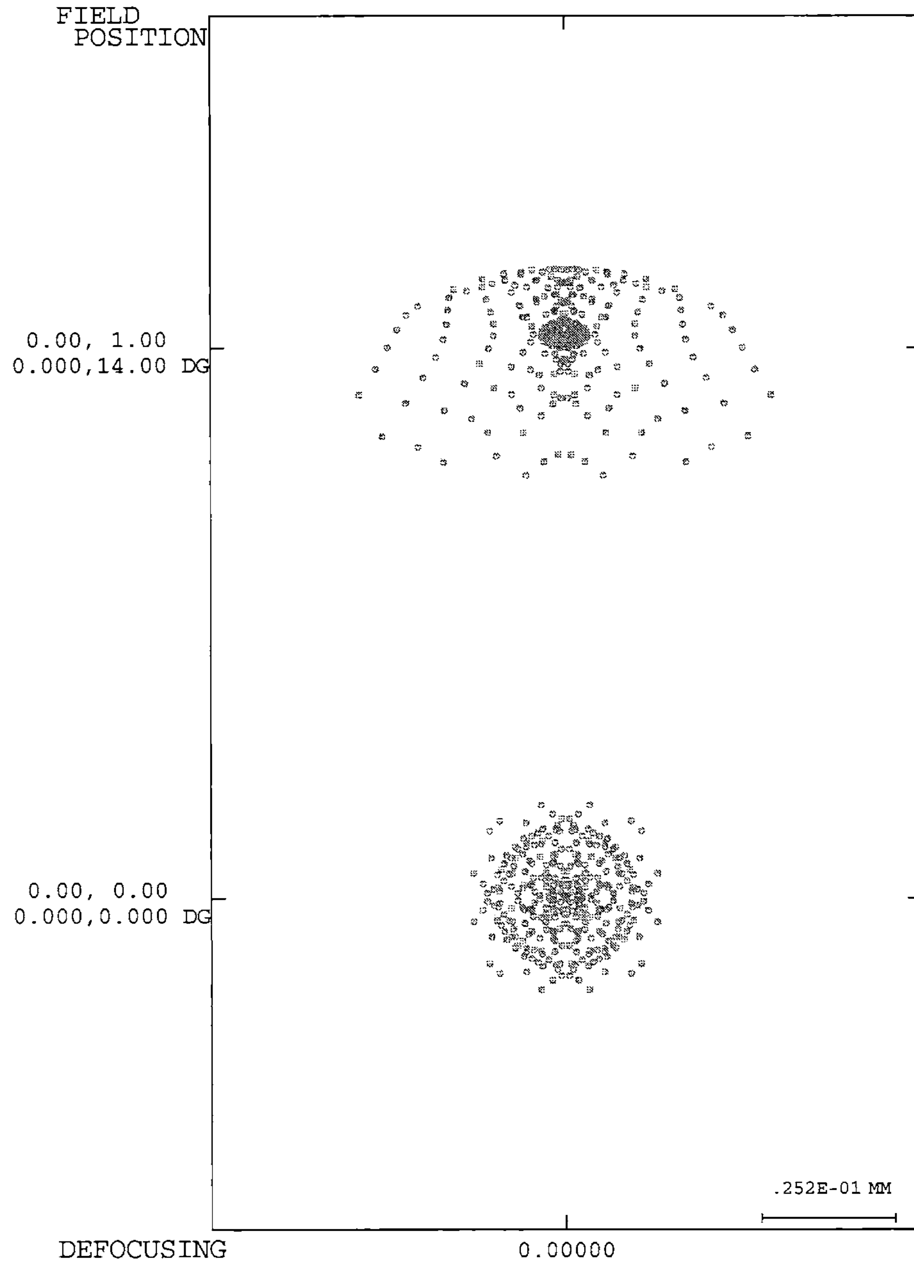


Figure 7.5: Spot diagrams associated with two incident angles

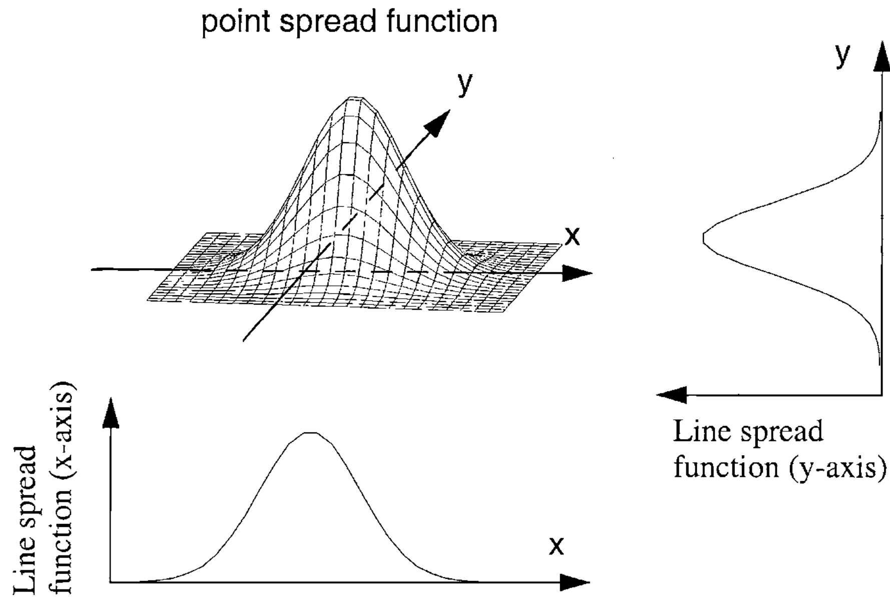


Figure 7.6: Point and line spread functions

row width. Examples of a point spread function and the associated line spread functions are shown in Figure 7.6.

The line spread function can be used to compute the spatial intensity of the image associated with the sinusoidal bright/dark pattern as shown in Figure 7.3. The calculation procedure is essentially a convolution operation between the vertical line spread function $L_v(x)$ and the brightness distribution function $B(x)$ on the object plane. It is calculated by

$$F(x) = \int L_v(\delta) \cdot B(x-\delta) d\delta, \tag{7.1}$$

where $F(x)$ represents the brightness distribution on the image plane.

This calculation is familiar in linear-system analysis. The line spread function is analogous to the impulse response and the brightness distribution function of the object $B(x)$ is analogous to the input signal. Apply Fourier transform to eq.(7.1), this convolution operation can be transformed into a multiplication operation in spatial frequency domain:

$$\boldsymbol{F}(\nu) = \boldsymbol{L}(\nu) \cdot \boldsymbol{B}(\nu), \tag{7.2}$$

where ν denotes the spatial frequency. If the light intensity of the object has a sinusoidal pattern, $\boldsymbol{B}(\nu)$ becomes a delta-function centered at the specific spatial frequency. Thus, the

frequency response of the image, $\boldsymbol{F}(\nu)$, would be solely determined by the Fourier transform of the line spread function. Apply some algebra, the modulation transformation function can be computed as

$$MTF(\nu) = |\boldsymbol{L}(\nu)|. \tag{7.3}$$

As shown in Figure 7.6, there are two line spread functions associated with a point source. These two functions correspond to orthogonal stripe patterns. Since typical lenses are made with circular symmetry, the orthogonal axes can be chosen with one in the radial direction and the other in the transversal direction. Consequently, the *MTF*s are classified into radial *MTF* and transversal *MTF*. These two functions are generally different.

Ideally, a perfect *MTF* is a function with unitary magnitude and linear phase at all spatial frequencies. In such case the brightness variation on the object plane can be faithfully reflected on the image plane. This condition requires the line spread function to be a delta-function in spatial domain, which is not viable due to lens imperfections. As long as the line spread function is not a delta-function, the *MTF* will eventually drop from unity as the spatial frequency increases.

7.3.2 Case Study

An optical simulation tool is utilized to compare the *MTF*s of the test lens systems with two, three, and six lens elements. In order to make a fair comparison, all the lens systems are given an unified specification, which are an EFL of 10 mm, an f-number of 3, and a FOV of 40°. From the EFL and FOV, the size of the imager chip is calculated to be roughly 7 x 7 mm^2, which is equivalent to a lens system with 2/3" optical format. These parameters are selected to represent a typical imaging application. It is worth mentioning that the adopted specifications only represent one sample point in the lens design space. The results deduced from the comparison should not be quantitatively extrapolated for other specifications. A constant focal length is utilized to simplify the simulation. Besides, a fixed-local-length lens system tends to have a higher resolution than that with a variable focal length. The data and the specifications of the lens systems are listed in Appendix C.

In the simulations, several patented lens systems are utilized as the preliminary designs. The dimensions of the lens systems are modified first to conform to the specifications. An internal optimization procedure built in the simulation tool is then used to mod-

ify the curvature of the surfaces and the thickness of the lens elements to achieve a locally-optimized design under the constraints of the specifications. All the surfaces utilized in the simulations are spherical surfaces. Three optical wavelengths, 656.3 nm (red), 587.6 nm (green), and 486.1 nm (blue)[1], are considered for optimizing chromatic aberration. A weighting ratio of 1:2:1 is applied to the three wavelengths.

Figure 7.7 shows the 2-D cross-section of the two-element lens design. The three groups of rays on the figure correspond to incident angles[2] of 0°, 14°, and 20°. The simulated *MTF*s are presented in Figure 7.8. The curves correspond to the radial and transversal *MTF*s with the three incident angles. The *MTF* associated with the diffraction limit of the lens system is also shown for comparison. The comparison reveals how far the performance of the lens system is from the diffraction limit.

In order to determine the resolving ability of the lens systems, a minimum *MTF* value of 0.3 is set as the criteria. The selection of this *MTF* value is a general rule of thumb. An averaged *MTF* is calculated from the *MTF*s with different incident angles and orientations. The maximum resolvable spatial frequency is determined to be 25 cycles/mm. Since it takes at least two pixels to map up a bright/dark cycle, the minimum resolvable pixel size of this lens system is 20 μm.

Similar procedures are applied to a three-element lens system and a six-element lens system. Figure 7.9 and Figure 7.10 show the lens drawing and the *MTF*s for the three-element lens system; Figure 7.11 and Figure 7.12 show the same plots for the six-element lens design. From the simulation results, the minimum resolvable pixel is 4 μm for the three-element design and 2.2 μm for the 6-element design. Table 7.1 lists the performance parameters for the three different designs. The second column of the table shows that a smaller resolvable pixel size can be achieved as more lens elements are used. The third column of the table lists the relative illumination on the image plane. As more lens elements are used, the uniformity of illumination becomes worse. This is because some of the rays with larger incident angles cannot reach the image plane. This phenomena is called

---

1. The wavelengths correspond to hydrogen C-line (656.3 nm), helium d-line (587.6 nm), and hydrogen F-line (486.1 nm). These wavelengths are typically used for specifying the refractive index and dispersion of lens materials.
2. The incident angles are selected so that the enclosed image area are equally partitioned.

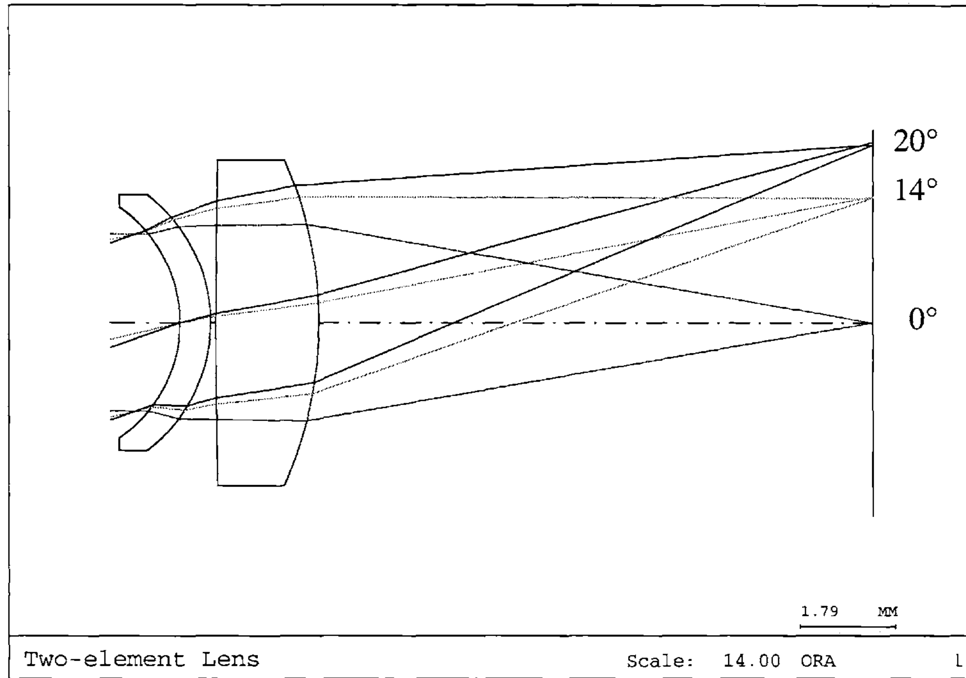


Figure 7.7: 2-D lens drawing of the two-element lens design

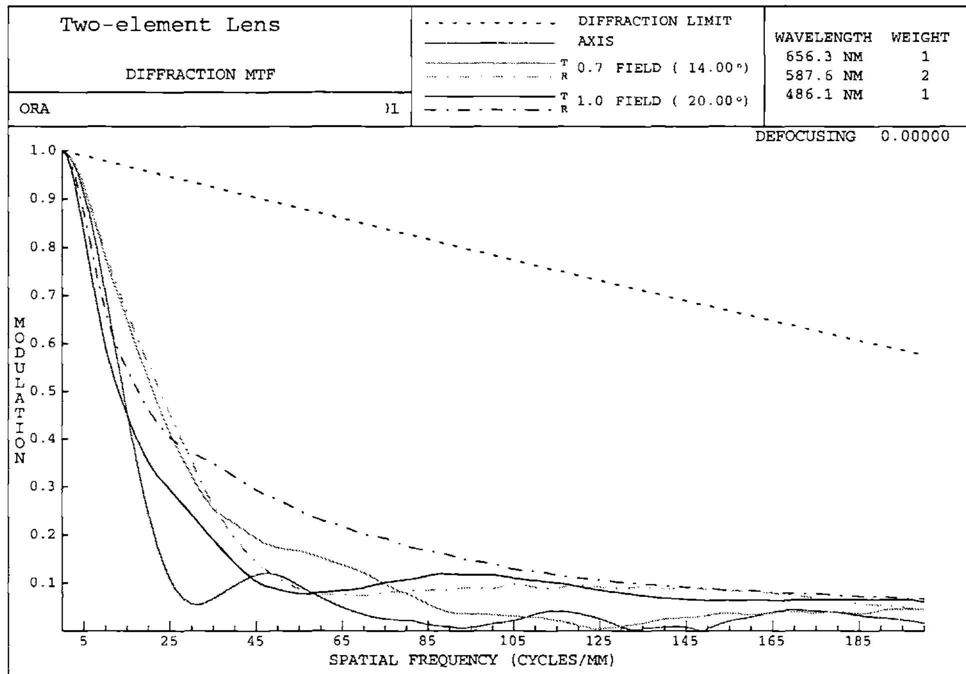


Figure 7.8: *MTF*s of the two-element lens design

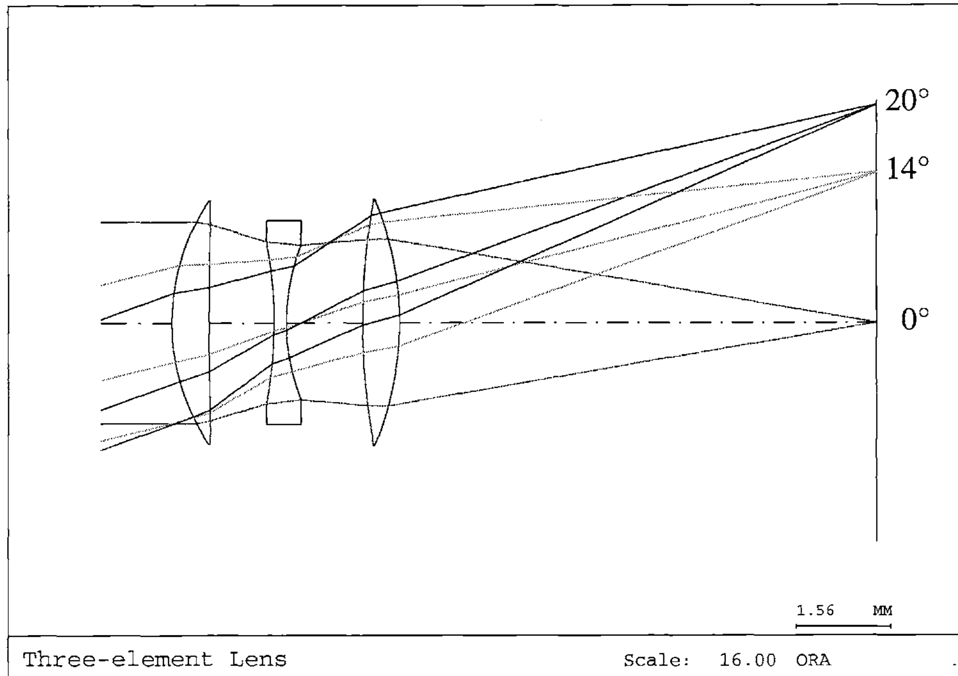


Figure 7.9: 2-D lens drawing of the three-element lens design

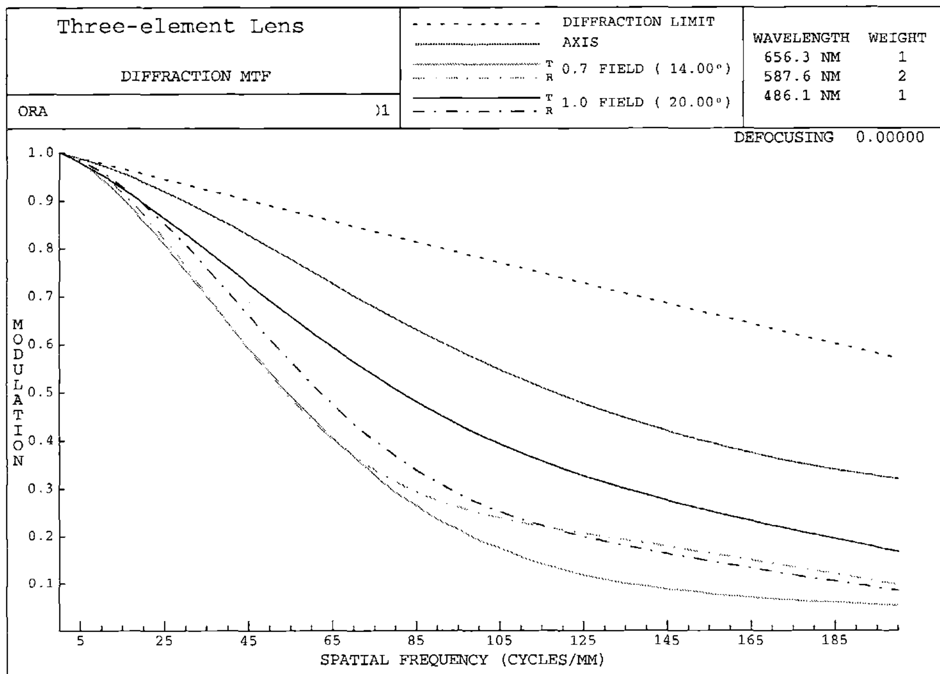


Figure 7.10: *MTF*s of the three-element lens design

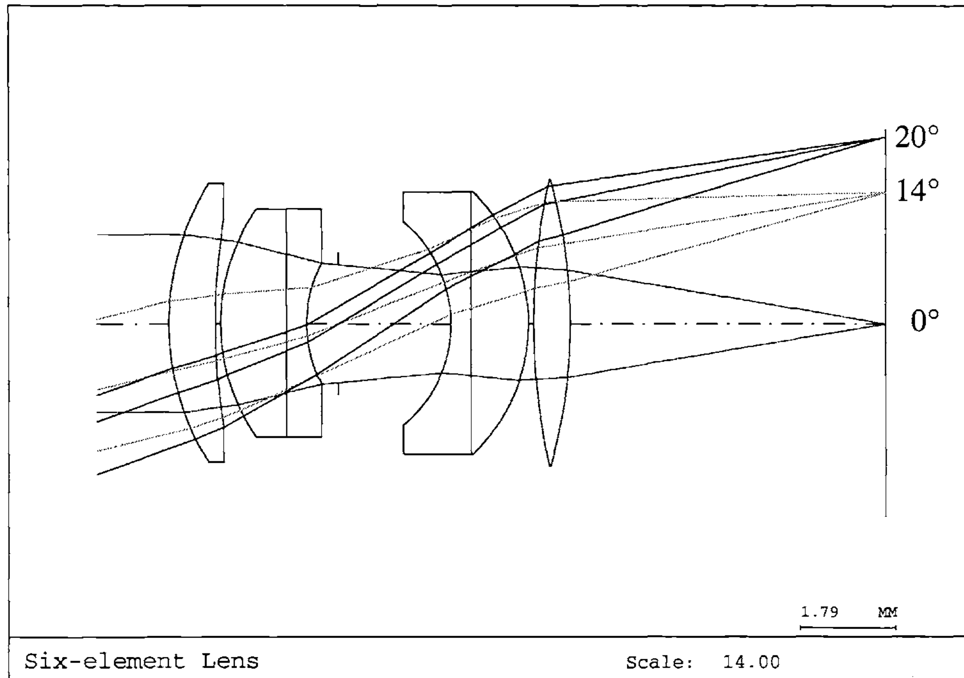


Figure 7.11: 2-D lens drawing of the six-element lens design

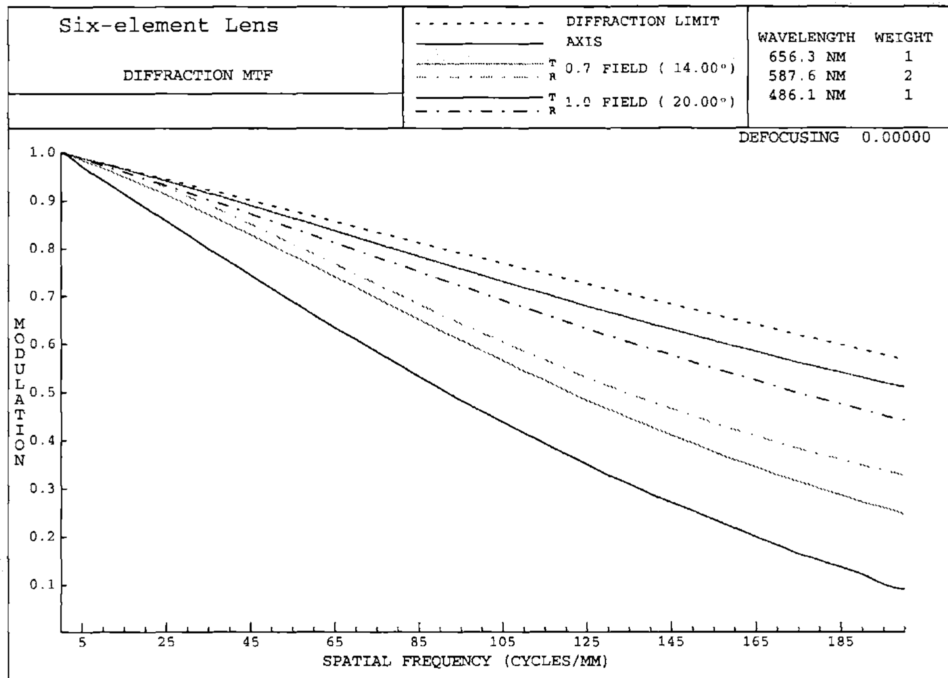


Figure 7.12: *MTF*s of the six-element lens design

# of lens element	minimum resolvable pixel size with *MTF* > 0.3	relative illumination versus incident angles	image distortion versus incident angles
2-element	20 μm	100% (0°), 95.2% (14°), 91% (20°)	0% (0°), -1.71% (14°), -3.55% (20°)
3-element	4 μm	100% (0°), 75.6% (14°), 53.5% (20°)	0% (0°), -0.41% (14°), -1.05% (20°)
6-element	2.2 μm	100% (0°), 72.3% (14°), 36% (20°)	0% (0°), -1.5% (14°), -3.78% (20°)

Table 7.1: Comparison of lens performance between different designs

vignetting [58]. Take the 6-element lens (Figure 7.11) as an example. Considering the optical rays with 20° incident angle, only those rays intersecting the first lens at the lower portion are able to reach the image plane. The rays entering the upper half of the first lens, due to the larger amount of bending, are unable to be collected by the last element of the lens system. This vignetting effect depends on the design but it tends to become more serious as more lens elements are used in this comparison. Therefore, even though the resolution is better for the 6-element lens, the illumination uniformity is worse. One way to solve this problem is to block some of the light rays corresponding to smaller incident angles as well so that the illumination difference between the edge and the center of the image is reduced. This can be done by placing some diaphragms in the lens. This method, however, decreases the overall light intensity on the image plane and equivalently increases the f-number of the lens system. The image distortion on the fourth column of the table is another important parameter for photographic lenses. The effect of distortion can be explained by Figure 7.13. This figure shows images with no distortion, positive distortion and negative distortion. In the positive distortion case, the image of an off-axis point is formed farther away from the axis and the reverse happens to the negative-distor-

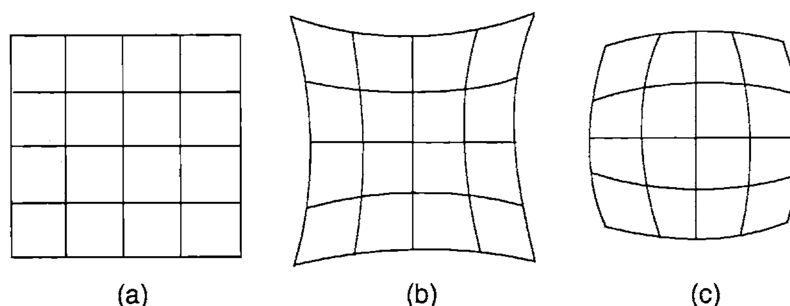


Figure 7.13: Distortion of images: (a) no distortion; (b) positive distortion; (c) negative distortion

tion case. According to Table 7.1, image distortion is not a strong function of the number of lens elements.

The simulation results show that the resolution of a lens system does improve with the increase of lens elements and the corresponding resolvable pixel size decreases. This means that pixel size reduction is feasible by choosing a lens system with the resolution matched with the pixel size. It is likely that the complexity of the lens system will increase to accommodate the smaller pixel size. If the objective of pixel size reduction is mainly for cost reason, the savings from a smaller chip size should be justifiable with the increase of lens cost. Besides, resolution should not be the only consideration on lens performance. As the lens elements are increased to improve resolution, other issues such as illumination uniformity and distortion must be evaluated.

7.4 Discussion

In the previous simulations, constant lens parameters such as EFL, f-number, and FOV are specified. A general discussion is invoked to examine the variation of these parameters to the lens performance. The utilization of aspherical surfaces is also discussed.

As described in Section 7.2.1, if EFL is increased, a higher f-number and a lower FOV result. To accommodate the EFL change and maintain a constant f-number and FOV, the aperture of the lens system and the imager size need to increase. Generally a lens with a short EFL tends to have higher distortion at the edge of the image. That is because a lens with a short EFL has to provide stronger deflection for the outer rays to focus at a shorter

distance. Therefore a larger curvature at the surface is generally required. This large curvature tends to cause resolution degradation and image distortion.

The f-number of a lens system can be varied using an adjustable aperture. As the aperture closes to a smaller diameter, the f-number increases and the resolution of the lens improves. This can be explained from the ray-tracing concept. When the aperture closes down, the rays at the outer portion of the lens are blocked and do not contribute to image formation. Typically these outer rays intersect the image plane farther away from the focus center. The spot diagram on Figure 7.5 demonstrates this effect. These rays are the components that degrades the point spread function and the resolution. Therefore, increasing f-number improves lens resolution. This technique is commonly used to expand the depth of focus in photography. However, increasing the f-number with a factor of F would reduce the light intensity of the image by F^2. To compensate the intensity loss the imager has to improve the sensitivity without amplifying the noise so that a higher S/N ratio of the imager is required. This is the drawback for increasing f-number to improve resolution.

FOV is determined by EFL and the size of the sensing device as described in Section 7.2.1. Assuming EFL is fixed, FOV increases with the size of the sensing device. However, as the size of the sensing device is increased, the imperfections near the boarder of the device usually get worse and the overall resolution is degraded. In photography FOV is controlled by varying EFL since the film size or the imager size is fixed.

The utilization of aspherical surfaces on the lens increases the difficulty in lens fabrication, but the lens resolution can be improved substantially. A 2-element lens system is used for a demonstration. As an aspherical surface is utilized at the first surface and the curvature allows a correction polynomial up to the 8th-order, the resolution can be improved by a factor of two without major degradation on other parameters. The comparison of the lens parameters with and without the aspherical surface is shown in Table 7.2.

lens type	minimum resolvable pixel size with $MTF > 0.3$	relative illumination versus incident angles	image distortion versus incident angles
2-element	20 μm	100% (0°), 95.2% (14°), 91% (20°)	0% (0°), -1.71% (14°), -3.55% (20°)
2-element with one aspherical surface	10 μm	100% (0°), 97.3% (14°), 94.8% (20°)	0% (0°), -2.18% (14°), -4.5% (20°)

Table 7.2: Comparison of lens performance between 2-element lens designs with and without a aspherical surface

7.5 Summary

An optical simulation tool is utilized to study the requirement on lens resolution to accommodate pixel size reduction. Lens systems with different amount of lens elements are compared. Three of the fundamental lens parameters — effective focal length (EFL), f-number, and field of view (FOV) are briefly described and specified in the lens system comparison. In the simulation, the modulation transfer function (*MTF*) is used as an index of resolution and the calculation procedure is presented. According to the simulation results, lens resolution can be improved with an increase on the number of lens elements. However, other lens parameters such as illumination uniformity and optical distortion may be degraded and should be considered for the overall lens performance. From a cost perspective, the saving on chip area from pixel size reduction should be higher than the extra cost of a upgraded lens system. The advantage of pixel size reduction ceases when the lens system becomes too expensive.

Chapter 8

Conclusions and Future Work

8.1 Conclusion

A study on CMOS technologies for image sensor applications is presented in this thesis. The topics in this study can be divided in four major parts, which include a study on the optimization methodology for CMOS imager processes, a study on hot-carrier induced excess minority carriers and its influence to image sensors, a study on pixel cross-talk to color imager performance, and a study on optical lens resolution for pixel size reduction.

A two-stage characterization methodology is implemented for image sensor optimization. In the first stage, a large-area photodiode structure is utilized to study the fundamental junction properties such as spectral response, leakage current, and capacitance. Improvement on the transmission coefficient of the stacking dielectric layers is also studied. The characterization results provide first-order performance estimation for the image sensors. The recommended sensor structures are further implemented in the small-dimension pixel arrays for the second-stage characterization. A detailed comparative analysis is performed on different pixel structures, junction types, and process parameters. The approach is proven to be effective in selecting an optimized image sensor among the ones that are available from a specific CMOS process. Imager parameters that fundamentally limit the image sensor performance are also identified for further process improvement.

The phenomena of hot-carrier induced excess minority carriers is observed using a 64 x 64 NW/Psub active pixel sensor arrays. The influence of the excess carriers depends on the bias condition and temperature. Nonlinearity of imager response results from this mechanism. The spatial distribution of the excess minority carriers is characterized to be a short-range effect under normal operating conditions. Modifications of pixel layout to improve this effect are provided.

Signal cross-talk between adjacent pixels is characterized using an NW/Psub active pixel sensor arrays with a specially design metal shield. The major cross-talk mechanism is identified to be photo-carrier diffusion in the lightly doped substrate. The cross-talk

degrades pixel-to-pixel color mismatch, especially under low illumination conditions. Modifications on process parameters and pixel layout are discussed.

The requirement on optical lens resolution by pixel size reduction is studied using an optical simulation tool. Theoretically the resolution of the lens system can be improved with the number of lens elements. However, other lens parameters such as illumination uniformity or distortion may be degraded as the amount of lens elements increases. The cost advantage from a reduced pixel size will cease when the increase of the lens cost surpasses the saving from the reduced chip area.

8.2 Future Work

This study contains broad aspects of CMOS imager including pixel structures, processing parameters, and the related physical phenomena. However, to achieve a CMOS imager with quality equivalent to the high-end CCDs, a substantial amount of efforts is still needed to improve the imager processes. The characterization framework in this study should provide sufficient testing tools to perform process improvements on leakage current, spectral response, and cross-talk. To the imager industry, pursuing higher imager quality is never-ending.

Suggestions on the future work is listed below:

The testing cycle of the imagers can be improved by using wafer testers to automate the characterization. With a proper optical illuminating system setup, the testers can quickly feed the processing results back to the engineers for further improvements. Batch testing can also be performed to collect statistical data or to monitor the lot-to-lot process variation.

With the mainstream CMOS process shifting to 0.18-μm feature size and below in the near future, the influence of the hot-carrier induced excess minority carriers should continue to be examined. The up side is that it is well-accepted by the process engineers that the hot-carrier effects should be addressed in the new generations of CMOS processes. However, imagers do have a more stringent leakage tolerance than logic devices. Since scaling Vdd to reduce electric field is undesired for imagers because of signal swing reduction, eventually the pixel designs may need to be modified as suggested in the thesis to mitigate this hot-carrier effect.

After the process is modified so that cross-talk is no longer dominated by photo-carrier diffusion, optical cross-talk is expected to emerge and become the major cross-talk mechanism. The optical crosstalk could be due to obliquely injected light, photons reflecting and bouncing between silicon and metal layers, or the fundamental diffraction effects. Further studies on constructing better light shielding structures are required to address this optical cross-talk problem until the diffraction limit is achieved.

Finally, some research effort is deserved to be spent on micro-lens technologies [61]. Micro-lens is a technology that places small len-shape polymer material on top of individual pixels for collecting more photons to improve imager sensitivity. The geometric structures, the materials, and the location of the micro-lenses are the parameters to be optimized. This study can be performed by inserting several micro-lens masks into the current design of the small-dimension sensor arrays.

Bibliography

[1] S. G. Chamberlain, "Photosensitivity and scanning of silicon image detector arrays," *IEEE Journal of Solid-State Circuits*, vol. SC-4, June 1969, pp. 333–342.

[2] P. Noble, "Self-scanned silicon image detector arrays," *IEEE Transactions on Electron Devices,* vol. ED-15, April 1968, pp. 202–209.

[3] W. S. Boyle and G. E. Smith, "Charge-coupled semiconductor devices," *Bell Syst. Tech. J.*, vol. 49, pp. 587–593, 1970.

[4] B. Mansoorian, H.-Y. Yee, S. Huang, and E. Fossum, "A 250mW, 60 frames/s 1280 x 720 pixel 9b CMOS digital image sensor," *ISSCC Digest of Technical Papers*, 1999, pp. 312–313.

[5] M. Loinaz, K. Singh, A. Blanksby, D. Inglis, K. Azadet, and B. Ackland, "A 200mW 3.3V CMOS color camera IC producing 352 x 288 24b video at 30 frames/s," *ISSCC Digest of Technical Papers*, 1998, pp. 168–169.

[6] W. Yang, O.-B. Kwon, J.-I. Lee, G.-T. Hwang, and S.-J. Lee, "An integrated 800 x 600 CMOS imaging system," *ISSCC Digest of Technical Papers*, 1999, pp. 304–305.

[7] A. J. P. Theuwissen, *Solid-State Imaging with Charge-Coupled Devices*, Dordrecht/Boston/London: Kluwer Academic Publishers, 1995.

[8] S. Mendis, S. E. Kemeny, R. C. Gee, B. Pain, C. O. Staller, Q. Kim, and E. R. Fossum, "CMOS active pixel image sensor for high integrated imaging systems," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 2, February 1997.

[9] R. D. McGrath, V. S. Clark, P. K. Duane, L. G. McIlrath, and W. D. Washkurak, "Current-mediated, current-reset 768 x 512 active pixel sensor array," *ISSCC Digest of Technical Papers*, 1997, pp. 182–183.

[10] I. L. Fujimori, C.-C. Wang, and C. G. Sodini, "A 256 X 256 CMOS Differential Passive Pixel Imager with FPN Reduction Techniques," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, December 2000, pp. 2031–2037.

[11] D. Scheffer, B. Dierickx, and G. Meynants, "Random addressable 2048 x 2048 active pixel image sensor," *IEEE Transaction on Electron Devices*, vol. 44, no. 10, October 1997, pp. 1916–1720.

[12] S. Decker, R. McGrath, K. Brehmer, C. Sodini, "A 256x256 CMOS imaging array with wide dynamic range pixels and column-parallel digital output," *ISSCC Digest of Technical Papers*, 1998, pp. 176–177.

[13] D. Yang, A. El Gamal, B. Fowler, H. Tian, "A 640x512 CMOS image sensor with ultra wide dynamic range floating-point pixel-level ADC," *ISSCC Digest of Technical Papers*, 1999, pp. 308–309.

[14] O. Vietze and P. Seitz, "Image sensing with programmable offset pixels for increased dynamic range of more than 150dB," *SPIE* vol. 2654, 1996, pp. 93–98.

[15] K. Aizawa, Y. Egi, T. Hamamoto, M. Hatori, M. Abe, H. Maruyama, and H. Otake, "Computational image sensor for on sensor compression," *IEEE Transactions on Electron Devices*, vol. 44, No. 10, October 1997, pp. 1724–1730.

[16] F. Okamoto, Y. Fujimoto, T. Nagata, M. Furumiya, K. Hatano, Y. Nakashiba, and M. Yotsuyanagi, "A CMOS Imager with New Focal-Plane Motion Detectors," *1999 Symposium on VLSI Circuits Digest of Technical Papers*, 1999, pp.139–140.

[17] E. Fossum, "CMOS Image Sensors: Electronic Camera-On-A-Chip," *IEEE Transactions on Electron Devices*, vol. 44, No. 10, October 1997, pp. 1689–1698.

[18] M. Furumiya, H. Ohkubo, Y. Muramatsu, S. Kurosawa, and Y. Nakashiba, "High sensitivity and no-cross-talk pixel technology for embedded CMOS image sensor", *Digest of International Electron Device Meeting*, 2000, pp.701–704.

[19] S.-G. Wuu, D.-N. Yaung, C.-H. Tseng, H.-C. Chien, C.-S. Wang, Y.-K. Fang, C.-C. Wang, C. G. Sodini, Y.-K. Hsiao, C.-K. Chang, and B.-J. Chang, "High performance 0.25-μm CMOS color imager technology with non-silicide source/drain pixel," *Digest of International Electron Device Meeting*, 2000, pp.705–708.

[20] I. Inoue, H. Nozaki, H. Yamashita, T. Yamaguchi, H. Ishiwata, H. Ihara, R. Miyagawa, H. Miura, N. Nakamura, Y. Egawa, and Y. Matsunaga, "New LV-BPD (Low Voltage Buried Photo-Diode) for CMOS imagers," *Digest of International Electron Device Meeting*, 1999, pp.883–886.

[21] R. M. Guidash, T.-H. Lee, P. P. K. Lee, D. H. Sackett, C. I. Drowley, M. S. Swenson, L. Arbaugh, R. Hollstein, F. Shapiro, amd S. Domer, " A 0.6-μm CMOS pinned photodiode color imager technology," *Digest of International Electron Device Meeting*, 1997, pp.927–929.

[22] H.-S. Wong, "Technology and device scaling considerations for CMOS imagers," *IEEE Transactions on Electron Devices*, vol. 43, No. 12, December 1996, pp. 2131–2142.

[23] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*, Cambridge: Cambridge University Press, 1998.

[24] D.-N. Yaung, S.-G. Wuu, H.-C. Chien, C.-H. Tseng, and C.-S. Wang, "Effects of hydrogen annealing on 0.25-μm CMOS image sensor," presented at *2001 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, June 2001.

[25] A. J. Blanksby, M. J. Loinaz, "Performance analysis of a color CMOS photogate image sensor," *IEEE Transactions on Electron Devices*, vol. 47, No. 1, January 2000, pp. 55–64.

[26] D. Yang, H. Min, B. Fowler, A. El Gamal, M. Beiley, and K. Cham, "Test structures for characterization and comparative analysis of CMOS image sensors," *SPIE* vol.2950, 1996, pp. 8–17.

[27] A. El Gamal, B. Fowler, H. Min, X. Liu, "Modeling and estimation of FPN components in CMOS image sensors," *SPIE* vol.3301, 1998, pp. 168–177.

[28] B. Fowler, A. El Gamal, D. Yang, and H, Tian, "A method for estimating quantum efficiency for CMOS image sensors," *SPIE* vol.3301, 1998, pp. 178–185.

[29] K. Sakui, S. S. Wong, and B. A. Wooley, "The effects of impact ionization on the operation of neighboring devices and circuits," *IEEE Transactions on Electron Devices*, vol. 41, No. 9, September 1994, pp. 1603–1607.

[30] H.-S. Wong, "Experimental verification of the mechanism of hot-carrier-induced photon emission in N-MOSFET's using an overlapping CCD gate structure," *IEEE Electron Device Letters*, vol. 13, no.8, August 1992, pp. 389–391.

[31] S. Tam and C. Hu, "Hot-electron-induced photon and photocarrier generation in silicon MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-31, No. 9, September 1984, pp. 1264–1273.

[32] H. C. van Kuijk, J. T. Bosiers, A. C. Kleimann, L. Le Cam, J. P. Maas, H. L. Peek, C. R. Peschel, "Sensitivity improvement in progressive-scan FT-CCDs for digital still camera applications," *Digest of International Electron Device Meeting*, 2000, pp.689–692.

[33] E. Palik, *Handbook of Optical Constants of Solids*, Orlando: Academic Press, 1985.

[34] S. M. Sze, *Physics of semiconductor devices*, second edition, New York: John Wiley & Sons, 1985.

[35] R. D. Muller and T. I. Kamins, *Device electronics for integrated circuits*, second edition, New York: John Wiley & Sons, 1986.

[36] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics*, vol. 2, chapter 33, Redwood city, California: Addison-Wesley, 1964.

[37] R. E. Walpole, R. H. Myers, and S. L. Myers, *Probability and statistics for engineers and scientists*, sixth edition, Upper Saddle River, N. J. : Prentice Hall, 1998.

[38] Avant! Corporation, 46871 Bayside Parkway, Fremont, California 94538. http://www.avanticorp.com/.

[39] Silvaco International, 4701 Patric Henry Drive, Bldg.1, Santa Clara, CA 94054. http://www.silvaco.com/.

[40] P. Mouroulis and J. Macdonald, *Geometrical optics and optical design*, New York; Oxford : Oxford University Press, 1997.

[41][Silvaco International, *ATLAS Device Simulation Software User's Manual*, vol. 2., November 1998.

[42] M. E. Law, E. Solley, M. Liang, and D. E. Burk, "Self-consistent model of minority-carrier lifetime, diffusion length, and mobility.", *IEEE Electron Device Letters*, vol. 12, no. 8, August 1991, pp. 401–403.

[43] MOSIS VLSI Fabrication Service, 4676 Admiralty Way, Marina del Rey, California (U.S.A.) http://www.mosis.org/.

[44] Taiwan Semiconductor Manufacturing Co., Ltd. 9, Creation Rd. 1, Science-Based Industrial Park, Hsin-Chu, Taiwan, R.O.C. http://www.tsmc.com/.

[45] A. van der Ziel, *Noise*, New York: Prentice-Hall, 1954.

[46] P. R. Gray, and R.G. Meyer, *Analysis and design of analog integrated circuits*, third edition, New York: John Wiley & Sons, 1993.

[47] M. H. White, D. R. Lampe, F. C. Blaha, and I. A. Mack, "Characterization of surface channel CCD image arrays at low light levels," *IEEE Journal of Solid-State Circuits*, vol. SC-9, no. 1, 1974, pp. 1-13.

[48] B. Pain, G. Yang, M. Ortiz, C. Wrigley, B. Hancock, and T. Cunningham, "Analysis and enhancement of low-light-level performance of photodiode-type CMOS active pixel imagers operated with sub-threshold reset," presented at *1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, June 1999.

[49] M. Kyomasu, "New MOS imager using photodiode as current source," *IEEE Journal of Solid-State Circuits*, vol. 16, August 1991, pp. 1116–1122.

[50] H. Tian, B. Fowler, and A. El Gamal, "Analysis of Temporal Noise in CMOS Photodiode Active Pixel Sensor," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 1, Jan. 2001, pp. 92–101.

[51] G. Wyszecki and W. S. Stiles, *Color science – concepts and methods, quantitative data and formulas*, second edition, New York: John Wiley & Sons, 2000.

[52] R. W. G. Hunt, *Measuring Colour*, third edition, Fountain Press, England.

[53] J. Guild, "The colorimetric properties of the spectrum," *Phil. Trans. Roy. Soc. (London)*, A **230**, 149, 1931.

[54] W. D. Wright, "A trichromatic colorimeter with spectral stimuli," *Trans. Opt. Soc.* **29**, 225, 1927-1928.

[55] M. J. Vrhel and H. J. Trussell, "Filter considerations in color correction," *IEEE Transaction on Image Processing*, vol. 3, no. 2, March 1994, pp. 147–157.

[56] K. Engelhardt and P. Seitz, "Optimum color filter for CCD digital cameras," *Applied Optics*, vol. 32, no. 16, June 1993, pp.3015–3023.

[57] M. D. Fairchild, *Color Appearance Models*, Reading, MA: Addison-Wesley, 1998.

[58] W. J. Smith, *Modern Optical Engineering*, third edition, McGraw-Hill, 2000.

[59] R. Kingslake, *Lens Design Fundamentals*, New York: Academic Press, 1978.

[60] Optical Research Associates, 3280 E. Foothill Boulevard, Pasadena, CA 91107. http://www.opticalres.com.

[61] V. Korobov, C. Cork, H. Wolf, and S. Fainleib, "Characterization methodology for micro-lens performance in CMOS image sensors," presented at *2001 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors,* June 2001.

Appendix A

Photometry and Radiometry

Human visual systems only respond to a small section of the electro-magnetic wave spectrum. The corresponding wavelengths spans roughly from 380nm to 780nm. Even in this small spectrum section, the visual sensitivity of human beings is not uniform. The spectral sensitivity can be described with the CIE spectral luminous efficiency function $V(\lambda)$. There are two sets of such functions, which correspond to photopic vision (high light level) and scotopic vision (low light levels) denoted as $V(\lambda)$ and $V'(\lambda)$ respectively. The functions are shown in Figure A.1[1]. The peak luminous efficiency occurs at 555nm for photopic curve $V(\lambda)$ and 507nm for scotopic curve $V'(\lambda)$.

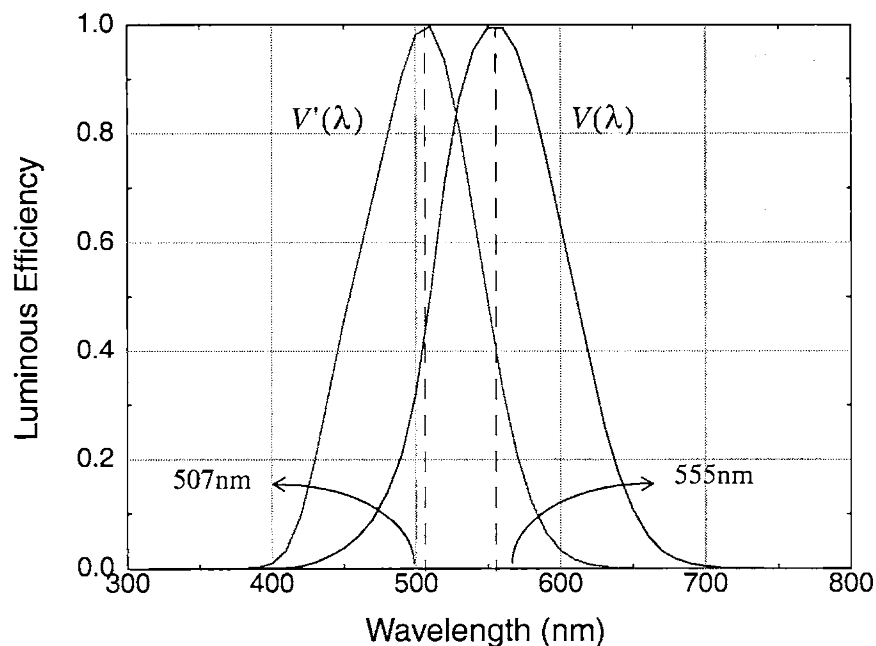


Figure A.1: Normalized luminous efficiency function for photopic vision, $V(\lambda)$, and scotopic vision, $V'(\lambda)$.

1. Data is available in reference [51].

The radiometric quantities can be converted into the photometric quantities via spectral integration, as shown in eq.(H.1).

$$\phi_p = K \cdot \int_{\lambda} \phi(\lambda) \cdot V(\lambda)\, d\lambda, \tag{H.1}$$

where $\phi(\lambda)$ represents the spectral function of the optical stimulus in radiometric unit; ϕ_p is the corresponding photometric quantity; $V(\lambda)$ is the luminous efficiency function, and K is a conversion factor. For example, if $\phi(\lambda)$ is the spectral irradiance of a surface in units of Watt/m^2/nm and ϕ_p is intended to be the illuminance in units of lux (lumen per square meter), the conversion factor K would be 683 lumens/watt for the photopic curve.

It is noticed from eq.(H.1) that the unit conversion from radiometric quantities to photometric quantities is weighted by the luminous efficiency functions $V(\lambda)$. Since the luminous efficiency function falls off rapidly at red light and blue light, an optical stimulus with a large number of red-light or blue-light photons actually contains a relatively small illuminance than the green-light case. However, the spectral response of silicon image sensor extends over a wider spectrum range than the human visual system. It is possible to produce a high signal output with little illuminance on the imager plane if long-wavelength photons dominate the optical stimulus. Thus, the spectral energy distribution of the optical source becomes important.

As a reference, examples of commonly encountered illuminance levels are listed in Table A.1.

Direct sunlight	100,000 lx
Full daylight (excluding direct sunlight)	10,000 lx
Overcast sky	1,000 lx
Typical office illuminance	500 lx
Heavy overcast	100 lx
Twilight	10 lx
Late twilight	1 lx
Full moon	0.1 lx
Quarter moon	0.01 lx
Clear night sky	0.001 lx
Overcast night sky	0.0001 lx

Table A.1: Approximate natural illuminance levels

Appendix B

Standard Illuminants

The normalized spectral energy distribution for some of the CIE standard illuminants is presented in Figure B.1. The CIE standard illuminant A usually refers to the output of tungsten filament lamp, which is an inexpensive and compact artificial light source. The corresponding color temperature is about 2856K. Another important light source is natural daylight. Standard illuminant B with color temperature 4874K is intended to represent direct sunlight at noon. Standard illuminant C with color temperature 6774K is for the average mixture of sunlight plus skylight. Standard illuminant D65 is important for the display industry. Its color temperature is 6500K and represents the white reference used for balancing television monitors and computer displays.

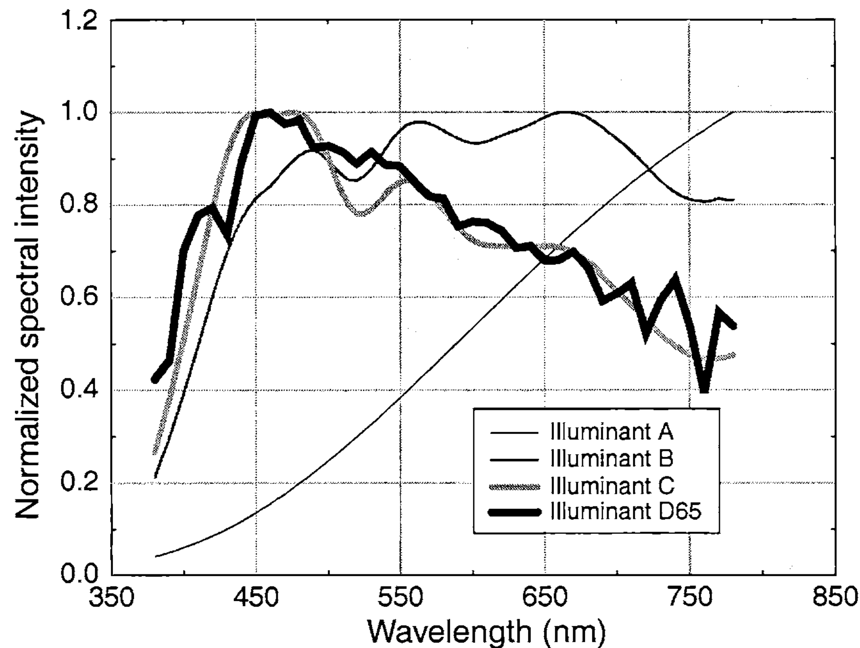


Figure B.1: Normalized spectral energy distribution of CIE standard illuminants

Appendix C

Lens Data and Specifications

C.1 The Two-element Lens Design

This lens design is modified from Japanese patent 60_49296 851101.

LENS DATA:

	RDY	THI	GLA
>OBJ:	INFINITY	INFINITY	
STO:	-2.64418	0.575605	487000.704000
2:	-3.02839	0.100000	
3:	137.14742	1.938336	744000.447000
4:	-7.56906	10.792457	
IMG:	INFINITY	-0.375493	

SPECIFICATIONS:

FNO	3.00000		
DIM	MM		
WL	656.30	587.60	486.10
REF	2		
WTW	1	2	1
INI	ORA		
XAN	0.00000	0.00000	0.00000
YAN	0.00000	14.00000	20.00000
WTF	1.00000	1.00000	1.00000
VUX	0.00000	0.00815	0.01686
VLX	0.00000	0.00815	0.01686
VUY	0.00000	-0.12739	-0.18262
VLY	0.00000	0.12739	0.18288

REFRACTIVE INDICES

GLASS CODE	656.30	587.60	486.10
487000.704000	1.484850	1.486999	1.491770
744000.447000	1.739045	1.743997	1.755696

INFINITE CONJUGATES

EFL	10.0000
BFL	10.7925
FFL	-7.8571
FNO	3.0000
IMG DIS	10.4170
OAL	2.6139
PARAXIAL IMAGE	
HT	3.6397
ANG	20.0000
ENTRANCE PUPIL	
DIA	3.3333
THI	0.0000
EXIT PUPIL	
DIA	4.2424
THI	-1.9348

C.2 The Three-element Lens Design

This lens design is modified from U.S. patent 2,645,157.

LENS DATA:

	RDY	THI	GLA
> OBJ:	INFINITY	INFINITY	
1:	3.62586	0.609756	BSM18_OHARA
2:	53.00226	1.07499	
3:	-7.27045	0.203252	PBM22_OHARA
4:	3.47739	0.245896	
STO:	INFINITY	1.016963	
6:	11.45161	0.609756	BSM18_OHARA
7:	-5.24561	7.964133	
IMG:	INFINITY	-0.072743	

SPECIFICATIONS:

FNO	3.00000		
DIM	MM		
WL	656.30	587.60	486.10
REF	2		
WTW	1	2	1
INI	ORA		
XAN	0.00000	0.00000	0.00000

YAN	0.00000	14.00000	20.00000
WTF	1.00000	1.00000	1.00000
VUY	0.00000	0.05000	0.10000
VLY	0.00000	0.40000	0.60000

REFRACTIVE INDICES

GLASS CODE	656.30	587.60	486.10
BSM18_OHARA	1.635052	1.638537	1.646586
PBM22_OHARA	1.642088	1.647685	1.661257

INFINITE CONJUGATES

EFL	10.0001
BFL	7.9641
FFL	-7.7099
FNO	3.0000
IMG DIS	7.8914
OAL	3.7606
PARAXIAL IMAGE	
HT	3.6397
ANG	20.0000
ENTRANCE PUPIL	
DIA	3.3334
THI	2.5728
EXIT PUPIL	
DIA	3.2418
THI	-1.7611

C.3 The Six-element Lens Design

This lens design is modified from U.S. patent 2,532,751.

LENS DATA:

	RDY	THI	GLA
> OBJ:	INFINITY	INFINITY	
1:	4.95701	0.874666	BSM24_OHARA
2:	12.07116	0.100000	
3:	3.76041	1.242423	SK1_SCHOTT
4:	INFINITY	0.377697	F15_SCHOTT
5:	2.43323	0.595685	
STO:	INFINITY	2.128667	

7:	-2.48391	0.377697	F15_SCHOTT
8:	INFINITY	1.083393	SK16_SCHOTT
9:	-3.42244	0.100000	
10:	12.53735	0.685817	SK16_SCHOTT
11:	-9.96856	5.931335	
IMG:	INFINITY	-0.021162	

SPECIFICATIONS:

FNO	3.00000		
DIM	MM		
WL	656.30	587.60	486.10
REF	2		
WTW	1	2	1
XAN	0.00000	0.00000	0.00000
YAN	0.00000	14.00000	20.00000
WTF	1.00000	1.00000	1.00000
VUY	0.00000	0.20000	0.70000
VLY	0.00000	0.30000	0.40000

REFRACTIVE INDICES

GLASS CODE	656.30	587.60	486.10
BSM24_OHARA	1.614254	1.617644	1.625478
SK1_SCHOTT	1.606991	1.610248	1.617756
F15_SCHOTT	1.600935	1.605648	1.616951
SK16_SCHOTT	1.617271	1.620408	1.627559

INFINITE CONJUGATES

EFL 10.0000
BFL 5.9313
FFL -3.4458
FNO 3.0000
IMG DIS 5.9102
OAL 7.5660
PARAXIAL IMAGE
HT 3.6397
ANG 20.0000
ENTRANCE PUPIL
  DIA 3.3333
  THI 3.5460
EXIT PUPIL

DIA	4.7675
THI	-8.3713

C.4 The Two-element Lens Design with One Aspherical Surface

LENS DATA:

	RDY	THI	GLA
OBJ:	INFINITY	INFINITY	
> STO:	-3.34584	1.738372	616356.604878

ASP:

K : 0.000000 KC : 100

IC : YES CUF: 0.000000 CCF: 100

A :-.173871E-02 B :0.193466E-04 C :-.481140E-04 D :0.000000E+00

AC : 0 BC : 0 CC : 0 DC : 100

2:	-3.67546	0.100000	
3:	11.55521	2.282547	487000.704000
4:	-16.19894	10.934090	
IMG:	INFINITY	-0.087394	

SPECIFICATIONS:

FNO	3.00000		
DIM	MM		
WL	656.30	587.60	486.10
REF	2		
WTW	1	2	1
INI	ORA		
XAN	0.00000	0.00000	0.00000
YAN	0.00000	14.00000	20.00000
WTF	1.00000	1.00000	1.00000
VUX	0.00000	0.00815	0.01686
VLX	0.00000	0.00815	0.01686
VUY	0.00000	-0.12739	-0.18262
VLY	0.00000	0.12739	0.18288

REFRACTIVE INDICES

GLASS CODE	656.30	587.60	486.10
616356.604878	1.613244	1.616354	1.623438
487000.704000	1.484850	1.486999	1.491770

INFINITE CONJUGATES	
EFL	10.0000
BFL	10.9341
FFL	-7.0049
FNO	3.0000
IMG DIS	10.8467
OAL	4.1209
PARAXIAL IMAGE	
HT	3.6397
ANG	20.0000
ENTRANCE PUPIL	
DIA	3.3333
THI	0.0000
EXIT PUPIL	
DIA	4.7585
THI	-3.3415