

CMOS imagers for low-level light and high-speed biomedical applications

CMOS IMAGERS FOR LOW-LEVEL LIGHT AND HIGH-SPEED BIOMEDICAL APPLICATIONS

By

Munir M. El-Desouki, M.Eng., M.A.Sc., B.Sc.,

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
& THE SCHOOL OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

McMaster University
Hamilton, Ontario, Canada

© Copyright by Munir M. El-Desouki, November 2010

All Rights Reserved

DOCTOR OF PHILOSOPHY (2010)
(Electrical and Computer Engineering)

McMaster University
Hamilton, Ontario

TITLE: **CMOS imagers for low-level light and high-speed
biomedical applications**

AUTHOR: **Munir M. El-Desouki,**
B.A.Sc. (Electrical Engineering)
KFUPM, Dhahran, Saudi Arabia
M.A.Sc. (Electrical Engineering)
McMaster University, Hamilton, Ontario Canada
M.Eng. (Engineering Entrepreneurship and Innovation)
McMaster University, Hamilton, Ontario Canada

SUPERVISOR: **Prof. M. Jamal Deen**

NUMBER OF PAGES: **xxiii, 134.**

Dedicated to

My late mother, Dr. Mayada Alhomsy

My father, Prof. Mahmoud El-Desouky

&

My dear wife, Noora Dabbagh

Abstract

Fluorescence optical imaging is becoming a very important technique for *in vivo* imaging and characterization of biological tissues. In order to add more contrast to the fluorescence image, fluorescence life-time imaging (FLIM) can be used, making it possible to differentiate between molecules with overlapping spectra, such as cancerous and noncancerous cells. However, designing FLIM imaging systems in a compact, complete camera-on-chip solution is a very challenging task and has led to significant research efforts in designing high-speed and high sensitivity imagers.

This work focuses on designing low-light level imagers in CMOS technology for biomedical applications that can be suitable for extremely high-speed imaging applications, such as FLIM. A fully integrated, 256-pixel CMOS camera-on-chip, was fabricated in a standard CMOS 0.18 μm technology. The imager was tested by controlling it with an Altera FPGA board. When clocking the ADC at a frequency of 1 MHz, images were obtained at about 60 frames/s. The design was next improved to achieve ultrahigh-speed imaging using a CMOS imager that can capture 8 frames with a frame capture rate that is higher than 1.25 billion frames per second.

The sensitivity was further improved using avalanche photodiode single-photon counters that were implemented in a standard digital 130 nm CMOS technology. The circuit achieves deadtimes as low as 200 ps, which is at least an order of magnitude less than previous work. The circuit also has a higher fill-factor of 25%, compared to 1-5% in previous work. A novel deadtime reduction technique design for active quench and reset circuits is also discussed.

The dynamic range of the imager was also improved using a novel design that relies on single-photon counting in time-domain. The design can achieve high sensitivity and high dynamic range, while maintaining a speed that is around 1000 times faster than conventional time-domain imagers. In order to further improve the frame rate, an imager that allows for simultaneous pixel counting and threshold detection in time-domain was also designed. The pixel also includes a novel analog counting technique that allows for an increased fill-factor.

Acknowledgements

I would like to express my sincere gratitude to my supervisor and mentor of the past six years, Prof. M. Jamal Deen, for giving me the opportunity to work on these projects and for his continuous support and guidance throughout my career on a research and personal level. It was a great honor to follow into the footsteps of such a great mentor. I have learned, and I continue to learn so much from him and I hope that my future achievements meet and exceed the expectations of being one of his students. I feel that I have excelled as a researcher and have won many scholarships due to his guidance and support.

Also, I would like to thank my committee members Prof. Qiyin Fang and Prof. Steve Hranilovic for their advice and support during my research and for taking the time to review my thesis.

I would next like to thank my team members and colleagues. To mention a number of them, M. Waleed Shinwari, Darek Palubiak, and Mohammad Naser. I would also like to thank Dr. Ognian Marinov for keeping his door always open for advice and support and for help with measurement setup and publications. I am also thankful to the administrative staff at the ECE department, especially Cheryl Gies for all her support throughout the past 6.5 years.

Next I would like to thank the Canadian Microelectronics Corporation (CMC) for providing me with the means of fabricating my designs. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) and the Raymond Moore OGSST of Canada and King Abdulaziz City for Science & Technology (KACST) of Saudi Arabia. I would specifically like to thank everyone at KACST for providing me with this great opportunity by funding my graduate studies. In particular, I would like to thank Dr. Mohammed ibn Ibrahim Al-Suwaiyel, Dr. Turki bin Saud bin Mohammad Al Saud, Dr. Daham Alani, Dr. Fayez Alhargan, and Soud Albatal. I would also like to thank Soha Mansour, my Academic Advisor at the Saudi Arabian Cultural Bureau in Canada.

I would also like to thank my dearest friends, many of which I was blessed with for more than 20 years, especially Ismail Alani, Fahd Bahadailah, Khalid Al-Najar, Hisham

Al-Rowaihi, and Teeba Alkhudairi. In Canada, I would like to thank my close friends Hamed Mazhab Jafari, Mousa Kfourri, Elias Haddad, Viva Nsair, Zuzana Stastna and the Samar, Samer, Bassem, Korin group.

I would like to express my deepest acknowledgements to my dear family, Prof. Mahmoud El-Desouki, Dr. Majid El-Desouki, Carolina Sanz, Mohannad El-Desouki, Dr. Omar Dabbagh, Huda Dabbagh, Jameel Dabbagh and Deema Dabbagh, for always being there when I needed them. I would also like to thank my uncle Dr. Bassam Alhomsy for his continuous encouragement and interest in my research.

Finally and most importantly, I would like to thank my wife, Noora, for all her support and for putting up with the never ending discussions about my work and my research, which I am sure, were very joyful to her. You always put me first, and for that I thank you the most.

Table of Contents

- ABSTRACT.....IV**
- ACKNOWLEDGEMENTS V**
- TABLE OF CONTENTS VII**
- LIST OF FIGURES..... XI**
- LIST OF TABLESXX**
- LIST OF SYMBOLS AND ACRONYMS..... XXI**
- CHAPTER 1 1**
- INTRODUCTION AND APPLICATIONS..... 1**
- 1.1. APPLICATIONS 1**
 - 1.1.1. FLUORESCENCE SPECTROSCOPY 2
 - 1.1.2. FLUORESCENCE LIFETIME IMAGING..... 3
 - 1.1.3. WIRELESS ENDOSCOPY CAPSULE..... 5
- 1.2. EXISTING TECHNOLOGIES 6**
 - 1.2.1. FLUORESCENCE MICROSCOPE..... 6
 - 1.2.2. SOLID-STATE ALTERNATIVES..... 7
- 1.3. SCALING EFFECTS 12**
 - 1.3.1. PIXEL AND ARRAY LEVEL..... 13
 - 1.3.2. DETECTOR LEVEL..... 15
- 1.4. MOTIVATION 16**
- 1.5. CONTRIBUTIONS..... 17**
- 1.6. THESIS ORGANIZATION..... 21**
- CHAPTER 2 22**
- CMOS PIXEL STRUCTURES 22**

2.1.	INTRODUCTION.....	23
2.1.1.	OPTICAL ABSORPTION IN SEMICONDUCTORS	24
2.1.2.	SILICON PHOTODETECTORS.....	28
2.2.	PIXEL ARCHITECTURES	35
2.2.1.	PASSIVE PIXEL SENSORS (PPS).....	36
2.2.2.	THREE TRANSISTOR ACTIVE PIXEL SENSOR (3T-APS).....	36
2.2.3.	FOUR TRANSISTOR ACTIVE PIXEL SENSOR (4T-APS).....	40
2.2.4.	CORRELATED-DOUBLE SAMPLING (CDS).....	40
2.2.5.	PIXEL SHARING	41
2.3.	SPECIAL PURPOSE PIXELS	41
2.3.1.	HIGH DYNAMIC RANGE (DR)	41
2.3.2.	LOW-LIGHT LEVEL	44
2.3.3.	HIGH-SPEED	45
2.4.	APS NOISE.....	46
2.4.1.	NOISE DURING RESET	46
2.4.2.	NOISE DURING INTEGRATION	47
2.4.3.	READ OUT STAGE NOISE	48
2.4.4.	RANDOM TELEGRAPH SIGNAL (RTS) NOISE	49
CHAPTER 3.....	51
CMOS APS COMPLETE CAMERA-ON-A-CHIP	51
3.1.	CAMERA-ON-A-CHIP DESIGN	52
3.2.	PHOTODIODE DESIGN.....	53
3.3.	PIXEL DESIGN.....	56
3.4.	ANALOG-TO-DIGITAL CONVERTER (ADC) DESIGN	58
3.5.	ARRAY DESIGN.....	63
3.6.	IMAGER RESULTS	64
CHAPTER 4.....	69

ULTRAHIGH-SPEED CMOS IMAGER.....	69
4.1. REVIEW OF HIGH-SPEED IMAGERS IN THE LITERATURE.....	70
4.1.1. DIGITAL READOUT ARCHITECTURES	71
4.1.2. ANALOG READOUT ARCHITECTURES.....	75
4.2. ULTRAHIGH-SPEED PIXEL DESIGN AND MEASUREMENTS.....	77
4.3. ARRAY AND IMAGER DESIGN	79
4.4. MEASUREMENTS RESULTS	81
CHAPTER 5.....	89
APD-BASED SINGLE-PHOTON IMAGER.....	89
5.1. APD DESIGN AND MEASUREMENTS.....	90
5.2. GEIGER MODE PASSIVE APD.....	93
5.3. GEIGER MODE APD WITH ACTIVE QUENCH AND RESET.....	100
5.4. DEADTIME REDUCTION TECHNIQUE.....	105
CHAPTER 6.....	107
TIME-DOMAIN SINGLE-PHOTON IMAGER.....	107
6.1. INTRODUCTION.....	107
6.2. TIME-DOMAIN SINGLE-PHOTON IMAGING.....	108
6.3. PIXEL DESIGN.....	110
6.4. ANALOG COUNTER.....	112
6.5. SPTD IMAGER	115
6.5.1. ANALOG COUNTER DESIGN	115
6.5.2. SRAM DESIGN.....	118
6.5.3. ARRAY DESIGN	119
CHAPTER 7.....	122
CONCLUSIONS AND FUTURE WORK.....	122

7.1. SUMMARY AND CONCLUSIONS	122
7.2. FUTURE WORK.....	124
REFERENCES.....	127

List of Figures

Figure 1.1: Fluorescence spectral response showing the excitation pulse and the emission pulse [1].....	2
Figure 1.2: Time-resolved and fluorescence lifetime measurements [1].....	4
Figure 1.3: Changes in tissue metabolism with pre-cancer development in vivo using the same hamster cheek pouch model of oral carcinogenesis [3].....	4
Figure 1.4: Sayaka camera pill from RF Systems Lab [7].....	6
Figure 1.5: Cross-section of a photomultiplier tube (PMT) [9].....	6
Figure 1.6: Solid-state imager containing a two dimensional array of pixels. Each pixel contains a photodetector and may or may not contain electronic circuits within the pixel.....	7
Figure 1.7: The fill-factor in a pixel is the ratio of the photosensitive area to the area of the pixel.....	8
Figure 1.8: Color detection in solid-state imagers using microlens color filters [10].	9
Figure 1.9: (a) The cross-section of a CCD pixel [15], and (b) – (e) charge-bucket brigade illustration of CCD charge transfer mechanism.....	10
Figure 1.10: Different readout architecture in CCD and CMOS systems [1].....	12
Figure 1.11: Image resolution as a function of number of columns for a fixed image size.	13
Figure 1.12: Downscaling of CMOS technology compared to some of the state-of-the-art CMOS imagers, reproduced and modified from [25].	14
Figure 1.13: The maximum SNR of an APS pixel as the technology scales down, reproduced from [32].	15
Figure 2.1: Block diagram of a CMOS image sensor.....	22

Figure 2.2: Photon interaction with valence electrons with a photon energy (a) equal to the bandgap, and (b) higher than the bandgap.	23
Figure 2.3: Electromagnetic spectrum showing the visible light range.....	24
Figure 2.4: Photon absorption in semiconductors [35].....	24
Figure 2.5: Optical absorption coefficient for various semiconductor material, reproduced from [34].	25
Figure 2.6: Band to band absorption in (a) direct and (b) indirect materials.....	26
Figure 2.7: Sensitivity of silicon photodiodes, reproduced from [35], [36].	27
Figure 2.8: An ehp generated at $x = l$. (a) the electron and hole drift times. (b) The generated external current corresponding to electrons and holes, and (c) the total photocurrent, reproduced from [35], [37].	28
Figure 2.9: A photoconductor with a length L and an area A	29
Figure 2.10: Schematic representation of how gain is produced in an NiN photoconductor.....	30
Figure 2.11: Formation of a PN-junction. (a) Carrier diffusion. (b) Thermal equilibrium.	31
Figure 2.12: PN-junction energy band-diagrams under (a) zero bias, (b) reverse and (c) forward bias.....	32
Figure 2.13: PN-junction ideal (a) forward bias total current density, and (b) reverse bias saturation current density.	32
Figure 2.14: (a) Reversed bias PN-junction and (b) IV curve showing different regions of operation.....	33
Figure 2.15: Junction breakdown by (a) Zener and (b) avalanche.	34
Figure 2.16: (a) PPS, (b) 3T-APS, and (c) 4T-APS.....	35
Figure 2.17: Capacitor voltage waveform of a 3T-APS.	37
Figure 2.18: (a) Schematic diagram of a 3T-APS with simplified readout circuit and (b) the equivalent circuit during integration and readout.	38

Figure 2.19: Measured and calculated photodiode voltage of a 3T-APS vs. time. The curve is magnified to show that the analytical model used by Faramarzpour in [39] best matches the measured results. Results reproduced from [39].	39
Figure 2.20: CDS circuit example. (a) Schematic diagram, (b) reset values stored, (b) image with FPN, and (c) difference after CDS.....	40
Figure 2.21: 4T-APS implemented with shared pixels, adopted from [21].....	41
Figure 2.22: (a) Schematic of a log-mode sensor, and (b) the output voltage waveform.	43
Figure 2.23: HDR image generated by merging three images obtained at long, medium and short exposure times using a standard SLR camera.....	44
Figure 2.24: Example of the effect of using a rolling shutter on an image with a moving object. (a) The undistorted object and, (b) the captured image.....	45
Figure 2.25: Example of a digital pixel sensor (DPS).	46
Figure 2.26: The equivalent circuit of a 3T-APS during the reset phase.....	47
Figure 2.27: The equivalent circuit of a 3T-APS during the integration phase.....	48
Figure 2.28: APS during readout, (a) schematic, (b) small-signal equivalent circuit, and (c) noise equivalent circuit.....	48
Figure 2.29: Random telegraph behavior of drain current in MOS transistor [44].	50
Figure 3.1: Block diagram of the CMOS imager setup.	51
Figure 3.2: CMOS camera-on-a-chip block diagram.	52
Figure 3.3: Examples of different diodes that can be implemented in a triple-well CMOS process.....	53
Figure 3.4: Photodiode test structure chip layout (a), and photomicrograph (b). (c) The dark room optical setup used to characterize the devices.	54
Figure 3.5: Measured relative responsivity of the n+/p-sub diode as a function of wavelength.	55
Figure 3.6: Color detection using multiple diodes at different depths. (a) The responsivity of two different diodes, and (b) ratio of the responsivity.....	56

Figure 3.7: 3T-APS layout with a 60% FF (a), and the measured output voltage for different optical powers at a wavelength of 680 nm as a function of time (b).	57
Figure 3.8: APS measured signal-to-noise ratio as a function of optical power at a wavelength of 680 nm for different integration times.	58
Figure 3.9: ADC block diagram.....	59
Figure 3.10: Simulation results of S/\bar{H} circuit.....	61
Figure 3.11: (a) Schematic of implemented operational amplifier, and (b) the simulated gain.....	61
Figure 3.12: Dual-slope waveform for three different input voltages.	62
Figure 3.13: Imager layout screen capture in 180 nm CMOS technology using Cadence Virtuoso software.....	64
Figure 3.14: Fabricated CMOS camera-on-a-chip photomicrograph.	65
Figure 3.15: Photograph of the measurement setup.....	65
Figure 3.16: Target images (resized for comparison) compared to the non-processed captured images shown in their original resolution of 256 pixels (a) and then enlarged 3 times for clarity (b). (c) Shows an example of using CDS in software by subtracting the data acquired at the beginning and the end of the integration times, however, in order to remove the non-uniform illumination, the reset image was taken of a blank white sheet before placing the target. .	66
Figure 3.17: (a) Image data of a black-white-black 3-bar target, and (b) the averaged rows corresponding to (a).	67
Figure 3.18: Modulation transfer function measurements. (a) Horizontal and (b) vertical MTF target, captured images and contrast waveform. (c) The MTF as a function of spatial frequency. (d) The sensor size relative to the target size showing the height of the sensor being less than its width, which results in higher vertical resolution.	68

Figure 4.1: A block-diagram categorizing some of the most relevant published high-speed imagers..... 70

Figure 4.2: Array access in a simple 4×4 pixel-by-pixel (PBP) sequential readout architecture [56]...... 72

Figure 4.3: Array access in (a) a per-column ADC (PC-ADC) readout and (b) a PC-ADC ×2 [56]...... 72

Figure 4.4: Simulation results of equations (4.1)-(4.3) showing the FR of PBP, PC-ADC and a PP-ADC readout architectures with 8-bit resolution ADCs ($b=8$), four 8-bit parallel outputs ($n=32$) and a $\tau_{ADC} = 2 \mu s$. (a) The FR as a function of varying the imager resolution with a fixed clock rate of 50 MHz ($1/\tau_{RO}$). (b) The FR as a function of the clock rate with a fixed imager resolution of $H \times V = 64 \times 64$. Both graphs are shown on a log-log scale [56]. 75

Figure 4.5: The storage and readout of an in situ CCD imager that can store up to N frames [56]. 76

Figure 4.6: (a) The schematic diagram of the ultrahigh-speed in-situ APS containing 8 memory elements and 38 transistors and (b) the layout screen capture of a single pixel [56]. 78

Figure 4.7: (a) Simulated photodiode response for 8 different light samples, (b) corresponding stored voltages, and (c) pixel readout voltage..... 79

Figure 4.8: (a) The layout screen capture of the complete camera-on-a-chip design [79]. (b) The schematic diagram of the cross-coupled VCO..... 79

Figure 4.9: (a) Schematic diagram of the write and reset pulse generator circuit. (b) Simulation results of the write and reset pulse generator circuit showing the start pulse coming in at 3 ns with a clock frequency of 1.25 GHz. The top inset figure shows the 8 reset pulses (active low) and the bottom inset figure shows the generated 8 write pulse signals that have a width of 400 ps. 80

Figure 4.10: Increased number of consecutive frames to 1024 using a 1D line-scan imager and fiber optic coupling. 81

Figure 4.11: Photomicrograph of the ultrahigh-speed camera-on-a-chip fabricated in a 130 nm CMOS technology.	82
Figure 4.12: (a) The simplified schematic of ultrahigh-speed APS, and (b) the measured APS output voltage for 3 different (weak, medium, and strong) incident light powers.	82
Figure 4.13: The measured APS output voltage for 3 different incident light powers showing both a single sample measurement and the envelope of 256 samples.	83
Figure 4.14: Photodiode to storage capacitor readout equivalent circuit (a) and noise equivalent circuit (b). Storage capacitor to column readout equivalent circuit (c).	84
Figure 4.15: (a) The measured APS output voltage for 3 different incident light powers used of a single sample, which was repeated 83 times to measure the SNR. (b) The measured compared to the calculated SNR for 3 different light powers with an inset figure showing a close up to the strong light measurement to show that the SNR drops beyond saturation.	85
Figure 4.16: The measured storage capacitor leakage of the ultrahigh-speed APS.	86
Figure 4.17: Simulation results of equations (4.7) showing the FR of PBP, PC-ADC and a PP-ADC readout architectures with 8-bit resolution ADCs, four 8-bit parallel outputs and a $\tau_{ADC} = 128$ ns, for a fixed clock rate of 50 MHz.	87
Figure 5.1: Possible diode layout cross-sectional view. (a) Typical pn-junction. (b) Typical APD layout and (c) Smaller size APD layout. The dashed lines show the depletion region.	91
Figure 5.2: (a) APD layout cross-sectional view, showing how the avalanche area was confined only under the n+ region when using the n-well guard-ring [86] and (b) layout top view showing APD device dimensions.	92

Figure 5.3: Measured I-V profiles of the APD with the guard-ring and the exact same structure without the guard-ring. The breakdown voltage of the APD is 11.3 V.....	93
Figure 5.4: SPAD passive quenching circuit.....	94
Figure 5.5: Oscilloscope screen capture showing the SPAD operation for two different excess biases with and without light. The figures are shown at the same time and voltage scales of 10 μ s and 200 mV per division, respectively.	95
Figure 5.6: (a) Passive quench equivalent circuit in breakdown mode, and (b) passive reset equivalent circuit in charge mode of the circuit shown in Figure 5.4. ...	96
Figure 5.7: Measured cathode voltage as a function of time, showing (a) a comparison to the calculated waveform at an excess bias of 2.2 V and (b) for several excess bias values.	97
Figure 5.8: Normalized photon-detection efficiency (PDE) as a function of wavelength for an excess bias of 10 mV.....	98
Figure 5.9: Measured dark and light count rates as a function of excess bias, (a) without saturation correction, and (b) with saturation correction. The power of the applied optical signal is 38 pW at a wavelength of 570 nm. (c) The dynamic range or signal-to-noise ratio (SNR) as a function of excess bias.	99
Figure 5.10: (a) Schematic diagram of a Geiger Mode APD with active quench and reset and (b) the APD voltage waveform during photon detection.	101
Figure 5.11: (a) Schematic diagram of a the proposed Geiger Mode APD with active quench and reset and (b) the layout screen capture in 130 nm CMOS technology with a 25% fill-factor.	102
Figure 5.12: Active quench and reset SPAD circuit showing the SPICE simulation equivalent circuit model.....	103
Figure 5.13: Simulation results of the active quench and reset SPAD circuit show in Figure 5.12.	104

Figure 5.14: Active quench SPAD imager (a) layout screen capture and (b) chip photomicrograph.	105
Figure 5.15: (a) The layout cross-sectional view of the multiple APDs, and (b) the modification to the counter for simultaneous counting.	106
Figure 6.1: Calculated APS output voltage for four different optical signals showing the generated photocurrents and the corresponding times required to drop below the (a) constant threshold voltage, and (b) ramp threshold voltage. The threshold voltages are shown as the dashed lines.	108
Figure 6.2: The electron count equivalent of the generated photocurrents from Figure 6.1 and the corresponding times required to drop below the (a) constant threshold count, and (b) a variable threshold count. The threshold counts are shown as the dashed lines.	109
Figure 6.3: Block diagram of the TDSPC pixel.....	112
Figure 6.4: (a) Schematic diagram of the designed analog counter, and (b) waveform operation of the analog counter.....	112
Figure 6.5: (a) Analog counter using multiple SPADs, and (b) the waveform operation of the multiple SPAD analog counter.....	113
Figure 6.6: Block diagram of a SPAD pixel that can achieve simultaneous pixel counting as well as simultaneous pixel analog-to-digital conversion.....	114
Figure 6.7: Schematic diagram of the cascaded analog counter.....	116
Figure 6.8: The Cadence simulation results of (a) the first counter, and (b) the second counter.....	116
Figure 6.9: The schematic diagram of the high-speed comparator.....	117
Figure 6.10: Cadence simulation results of the high-speed comparator. (a) Frequency response showing the gain and bandwidth, and (b) time domain rise and fall time simulations.	118
Figure 6.11: Schematic of a standard 6T-SRAM memory cell.	118

Figure 6.12: Schematic of the dual-port 6T-SRAM [110]..... 119

Figure 6.13: Layout of a TDSPC pixel with in-pixel analog counting and SRAM..... 120

Figure 6.14: The layout of the 24×16 pixel TDSPI with timer, readout and array access circuits..... 121

List of Tables

Table 1.1: Fluorescence and chemiluminescence application requirements, reproduced from [2].	3
Table 1.2: Comparison between PMT, CCD and CMOS imagers, reproduced from [25].	12
Table 2.1: Comparison of gain and response time for various photodetectors [32].	35
Table 2.2: Calculated $v_d(t)$ for different values of m [38].	39
Table 4.1: Summary of the various high-speed CMOS imagers available in the literature.	88
Table 5.1: Summary of the various DSM SPADs available in the literature.	100
Table 5.2: Summary of the various DSM SPADs with active quench and reset available in the literature.	104

List of Symbols and Acronyms

Symbols

V_{DD}	Supply voltage
C_{PH}	Photodiode capacitance
q	Electric charge
E	Energy
h	Planck's constant
ν	Frequency of light
c	Speed of light
E_g	Energy bandgap
η_Q	Quantum efficiency
g'	Generation rate per unit volume
α	absorption coefficient
λ	Wavelength
σ	Conductivity
n	Electron concentration
p	Hole concentration
μ_n	Electron mobility
μ_p	Hole mobility
D_p	Hole diffusion coefficient
D_n	Electron diffusion coefficient
p_{n0}	Thermal equilibrium minority carrier concentrations of holes in the n-region
n_{p0}	Thermal equilibrium minority carrier concentrations of electrons in the p-region
k	Boltzmann's constant
v_{sat}	Saturation velocity
C_d	Diode capacitance

i_{ph}	Photocurrent
i_{dark}	Dark current
V_T	Threshold voltage
φ	Built-in potential
g_m	Transconductance

Acronyms

3T-APS	Three transistor active pixel sensor
4T-APS	Four transistor active pixel sensor
AC	Alternating current
ADC	Analog-to-digital converter
AFS	Analog frame storage
APD	Avalanche photodiode
APS	Active pixel sensor
CCD	Charge coupled devices
CDS	Correlated double sampling
CMOS	Complementary-metal-oxide-semiconductor
CT	Computed tomography
DC	Direct current
DCR	Dark count rate
DNA	Deoxyribonucleic acid
DPS	Digital pixel sensor
DR	Dynamic range
DSM	Deep submicron
EFL	Effective focal length
ehp	Electron-hole pair
FF	Fill-factor
FLIM	Fluorescence lifetime imaging
FPGA	Field-programmable gate array
FPN	Fixed pattern noise
fps	Frames-per-second
FR	Frame rate

GI	Gastrointestinal tract
HDR	High dynamic range
I/O	Input/output
LED	Light emitting diode
MOSFET	Metal-oxide-semiconductor field-effect transistor
MRI	Magnetic resonance imaging
MSM	Metal-semiconductor-metal
OPAMP	Operational amplifier
OTPs	Outputs
PAO	Parallel analog outputs
PBP	Pixel-by-pixel
PC-ADC	Per column analog-to-digital converter
PDE	Photon detection efficiency
PEB	Premature edge breakdown
PMT	Photomultiplier tubes
PP-ADC	Per pixel analog-to-digital converter
PPS	Passive pixel sensor
RF	Radio frequency
ROI	Region-of-interest
RTS	Random telegraph signal
S/\bar{H}	Sample and hold
SNR	Signal-to-noise ratio
SPAD	Single photon avalanche photodiode
SPD	Single photon detector
SRAM	Static random access memory
SRH	Shockley-Read-Hall
TCSPC	Time-correlated single photon counting
TDSPC	Time-domain single photon counting
TDSPI	Time-domain single photon imager
VCO	Voltage controlled oscillator

Chapter 1

INTRODUCTION AND APPLICATIONS

The continuous growth of the billion dollar image sensor market has attracted significant research interest over the past decade. Beyond the conventional use in digital cameras, image sensors play an important role in industrial machine-vision, medical and scientific imaging, automotive navigation, driver parking assistance, and most recently, mobile phones that have become a key industry driver. Miniaturization of image sensors using micro- and nano-fabrication technologies offers many advantages such as low-power, low-cost, portability and incorporation of “intelligence” in the pixels of the image sensor. However, for biomedical applications such as disease screening or detection, these image sensing systems must be capable of detecting very low levels of emitted light from the biological samples. In some cases, high speed imagers are also required in biomedical applications. These challenges have attracted significant research interest in image sensors for biomedical applications. The following section starts by giving some examples of such applications, after which, the most popular image sensor technologies are compared and the effects of technology downscaling are also discussed.

1.1. Applications

Optical molecular imaging systems enable a real time non-invasive visualization of cellular functions in vivo. Such imaging systems have been used in medicine, agriculture, biodefense and environmental testing through techniques such as DNA sequencing, protein detection, gene expression, cell migration and evaluation of animal models of human cancer. These imaging techniques are even more attractive when developed in hand-held, portable devices that can be used for forensics and biohazard studies on-site. Compared with established diagnosis techniques such as x-ray, computed tomography

(CT), magnetic-resonance imaging (MRI) and the gamma-camera in the case of nuclear medicine, non-invasive fluorescence imaging systems have been considered to have many advantages, such as patient safety, high spatial resolution, small size and low equipment cost.

1.1.1. Fluorescence spectroscopy

One of the most common optical imaging techniques used for scientific and medical characterization is fluorescence imaging. The word fluorescence originates from the mineral fluorite, which is composed of calcium fluoride and often exhibits this luminescence or light emission phenomenon. Fluorescence is the property of certain atoms and molecules to absorb light at a particular wavelength (ultraviolet (UV) or visible range) and emitting light at a longer wavelength (Figure 1.1 [1]), over a short interval of time known as the fluorescence lifetime. The shift in wavelength between the absorbed and the emitted waveforms is known as the Stokes shift. Immediately following excitation, the fluorescence intensity decays exponentially, usually over a few nanoseconds for most biological fluorophores (a fluorophore is the component of a molecule that fluoresces).

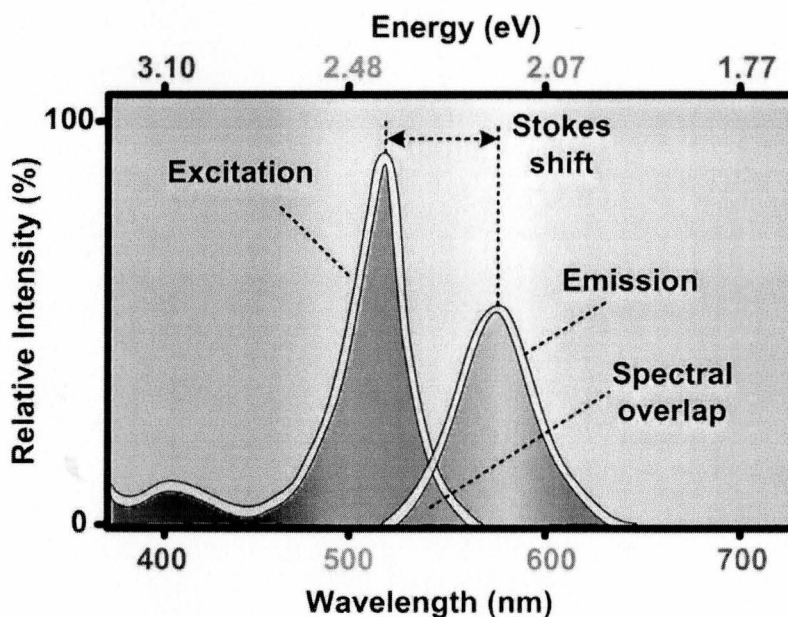


Figure 1.1: Fluorescence spectral response showing the excitation pulse and the emission pulse [1].

When the emission of light takes place at room temperature as a result of a chemical reaction, it is known as chemiluminescence. Bioluminescence is when the reaction takes place inside a living organism. Table 1.1 shows a comparison between the application requirements for fluorescence and chemiluminescence [2].

Table 1.1: Fluorescence and chemiluminescence application requirements, reproduced from [2].

	Fluorescence	Chemiluminescence
Wavelength	Visible, near- IR	Visible (425, 560 nm)
Emission time	Fast (psec ~ msec)	Slow (sec ~ hours)
Emission rate per reporter	High (Up to 10^5 ph/sec)	Low (0.01 - 1 ph/sec)
Sensitivity required	High (single molecule)	Low (1000 molecules)
Background	High (need filters)	Low

There are many applications of fluorescence in medicine. For example, DNA microarrays are used for studying levels of gene expression in living cells. In a microarray experiment, the DNA fragments are tagged with fluorescent dyes before being introduced to the microarray. The fragments that find their match on the surface of the microarray wells get attached to the corresponding probes in a process called hybridization. The DNA microarray is then exposed to light, and the level of fluorescent emission from each well determines the level of expression of the corresponding gene in the sample. Some of the spots can have extremely low levels of fluorescence emission, which need to be detected by ultra-sensitive imaging devices [1]. In addition to the high-sensitivity requirements, in some applications, the tested molecules can have overlapping spectral responses, such as cancerous and non-cancerous cells. One valuable method that can be used in this case is time-resolved measurements such as fluorescence lifetime imaging, which is explained next.

1.1.2. Fluorescence lifetime imaging

Time resolved techniques are used to determine the relaxation times of fluorescence signals, which is the time it takes for the electronically excited fluorophores to relax back to their ground state. Since the signal has an exponential decay over time, integrating approaches that have integration times much longer than the average fluorescent lifetime cannot be used. Rather, averaging a number of repeated measurements in narrow

sampling windows or gates (Figure 1.2 [1]) has been shown to be more effective. This can allow for the collection of a histogram of the detection times, building up the waveform. The background can also be removed by averaging the samples of a number of measurements without excitation. Such high-frame-rate applications require a fast and sensitive CMOS imager. CMOS imagers that can achieve timing resolutions between 150-800 ps from 64×64 pixel imagers with two point per transient waveform sampling and 150 frames/s, have been reported in the literature [1], [57]. However, capturing a lifetime curve without repetition is very challenging. Figure 1.3 shows how FLIM can be used for pre-cancer diagnosis by detecting changes in tissue metabolism [3].

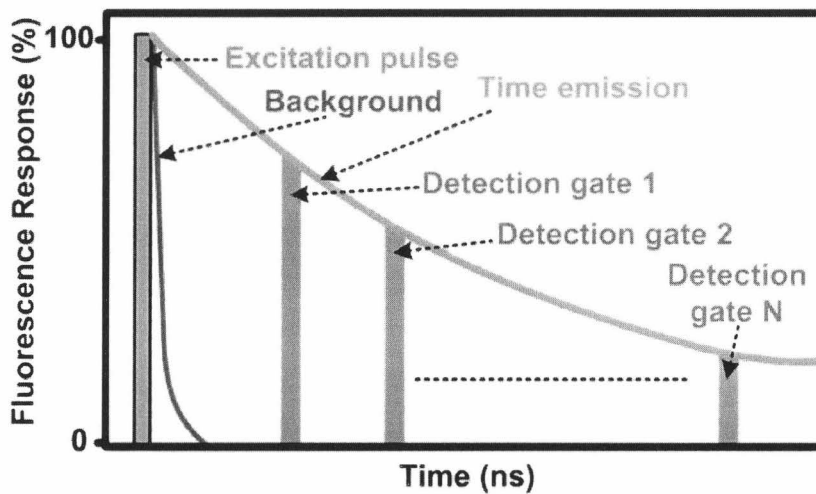


Figure 1.2: Time-resolved and fluorescence lifetime measurements [1].

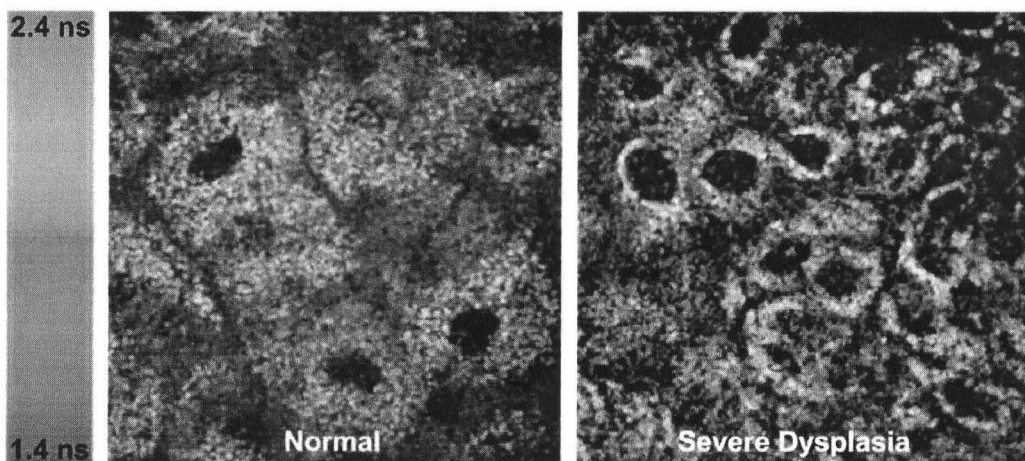


Figure 1.3: Changes in tissue metabolism with pre-cancer development in vivo using the same hamster cheek pouch model of oral carcinogenesis [3].

A typical FLIM system [4] requires a picosecond excitation source laser, with a bandwidth of less than 40 ps, a wavelength below 400 nm, a repetition rate of around 40 MHz and an average power of 0.5 mW. The optical detection gates should have a width in the range of 200 ps to 1 ns, and the imager requires low resolution (32×32) since the samples are usually small areas [4]. The FLIM image can be obtained after capturing the histogram from a series of images (21 used in [4]) that are colored based on different delays with respect to the excitation pulse. Since the FLIM image does not maintain the intensity information, the examples shown in Figure 1.3 was obtained by merging the lifetime map together with the intensity image.

Fluorescence imaging systems are considered attractive imaging tools that are not only limited to the lab to be used with large microscopes. Recent endoscopes come equipped with fluorescence imaging capabilities, such as the Endoscopic SPY Imaging System from Novadaq. Although catheter-based endoscopy is widely used for gastrointestinal (GI) tract cancer screening, it is invasive, expensive and uncomfortable. Wireless endoscopy pill cameras, which are discussed next, have emerged as an alternative.

1.1.3. Wireless endoscopy capsule

These pills are even more important when imaging the more difficult to reach areas of the GI tract such as the small intestine. Pill camera devices usually measure around 20 mm in length and 10 mm in diameter and cost a couple of hundred dollars. The pill is swallowed after which, it will begin to capture images that are wirelessly transmitted to an external computer [5]-[6]. Although most previous capsule designs would capture images either from a dome in front, rear or both, recent pills, such as Sayaka from RF Systems Lab (Figure 1.4 [7]), can capture images in 360 degrees from the side of the pill as it moves in a rotary fashion. The pill consumes roughly 50 mW, which is provided externally through an induction coupling vest, and can capture images at a rate of 30 frames per second. As seen in the figure, the pill provides white LEDs in addition to UV LEDs for fluorescence imaging. These types of pills must be able to operate with low power consumption, which limits the illumination power of the LEDs, requiring very sensitive imagers that can operate with low light levels. These imagers should also be integratable with the processing components in the pill.

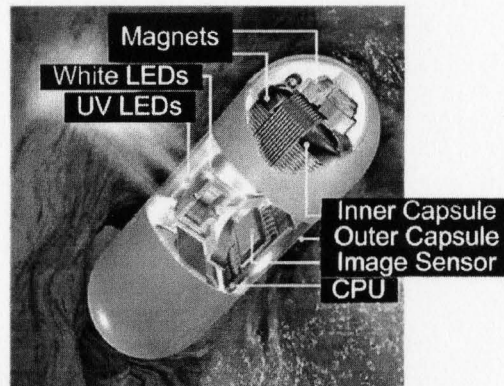


Figure 1.4: Sayaka camera pill from RF Systems Lab [7].

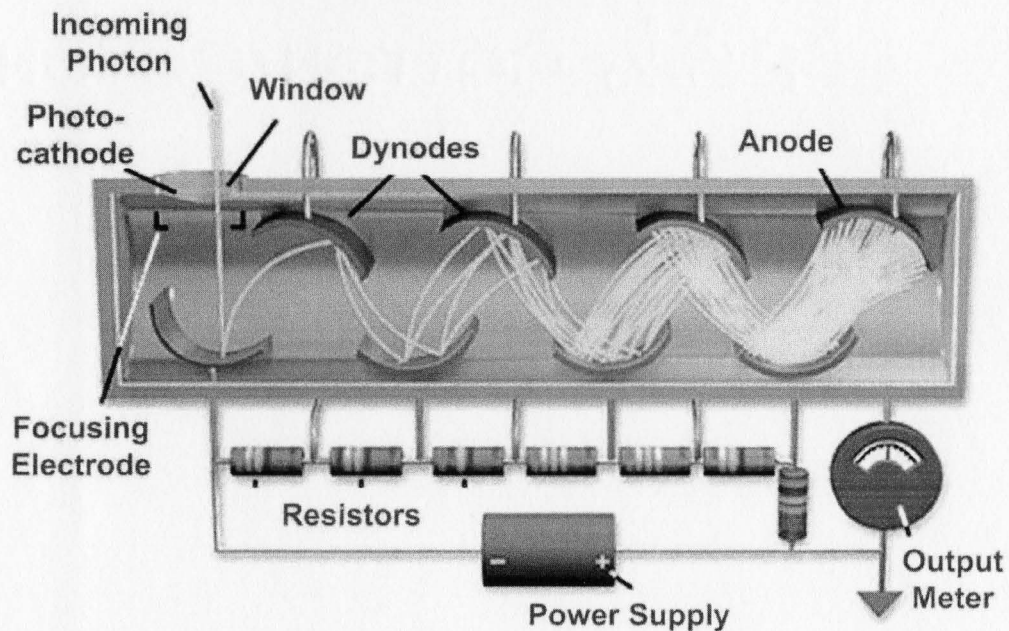


Figure 1.5: Cross-section of a photomultiplier tube (PMT) [9].

1.2. Existing Technologies

1.2.1. Fluorescence microscope

The fluorescence microscope is one of the most common imaging tools used to study fluorescence properties. The light sensing element of the fluorescence microscope, which is the most sensitive light detector that is currently used, is the photomultiplier tube (PMT). A PMT can generate up to one billion electrons for every incident photon. However, PMTs are expensive and require high operating voltages within the range of

1000V to 2000V, thus making them unsuitable for hand-held systems. Also, PMT systems have a limited photon detection efficiency of below 4% and their large size makes multiplexed imaging infeasible and hence, they are not suitable for dense arrays [1], [8]. Figure 1.5 shows a cross-section of a PMT [9]. As the incoming photon passes through the input window and hits the photocathode, an electron is emitted in the vacuum-tube. The emitted electron is then accelerated and focused by the focusing electrode onto the first dynode. Each dynode emits another electron by secondary emission causing a multiplication effect, and based on the number of dynodes used in a PMT, a specific gain can be achieved by collecting the multiplied electrons at the anode, which is located at the end of the dynode chain.

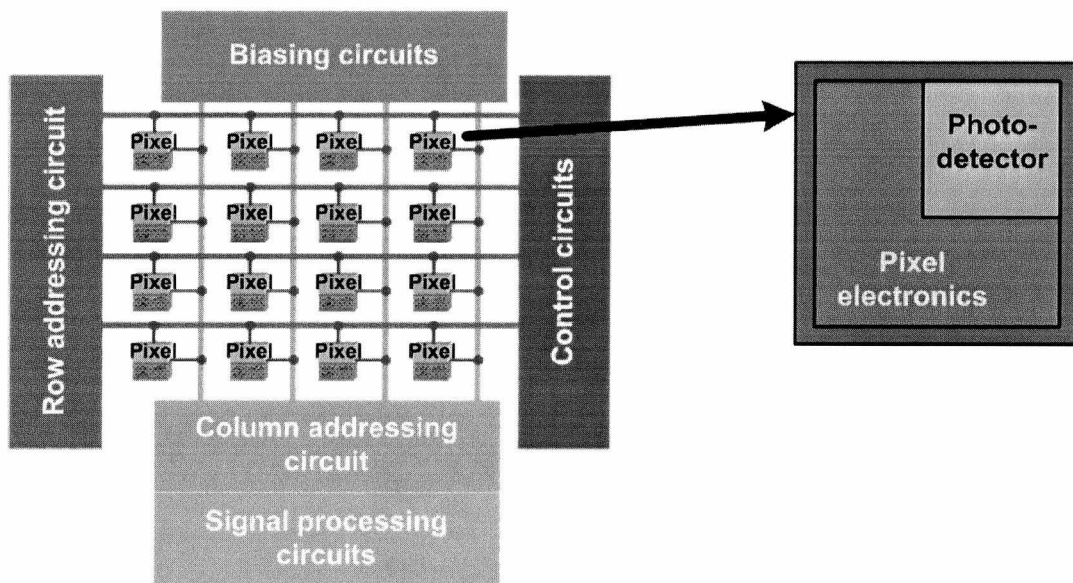


Figure 1.6: Solid-state imager containing a two dimensional array of pixels. Each pixel contains a photodetector and may or may not contain electronic circuits within the pixel.

1.2.2. Solid-state alternatives

A solid-state or silicon imager consists of a one- or two- dimensional array of pixels, with each pixel containing a photodetector to convert incident light into photocurrent (Figure 1.6). The array also includes decoders and multiplexers to access it, and readout circuits to convert the photocurrent into electric charge or voltage and read it out of the array. The photodetector converts incident photon flux to photocurrent, which is then converted to an output voltage. The photocurrent is not readout directly since the current levels produced are very low, in the femto- to nano-amperes range; rather, it is integrated in a

capacitance and read out as charge or voltage at the end of the integration time. The size of the integration capacitor, which is usually the parasitic capacitance of the photodetector, determines the well capacity, which is the maximum amount of charge that can be stored, and also sets the charge-to-voltage conversion gain, which is measured in microvolts per electron [1].

In many cases, a portion of the solid-state pixel may not be photosensitive. The fraction of the area occupied by the photodetector (the photosensitive area) in a pixel, compared to the total area of the pixel, is known as the fill-factor (FF), see Figure 1.7. Array sizes vary from a few tens of pixels for low-resolution sensor applications to megapixels for commercial cameras, while individual pixel sizes can be as small as $2\ \mu\text{m} \times 2\ \mu\text{m}$ [1]. The color detection in an imager is usually done using filters that are typically deposited on top of the pixel array, (Figure 1.8 [10]). Microlenses are also fabricated over the array to increase the amount of light incident on the photosensitive area of each pixel.

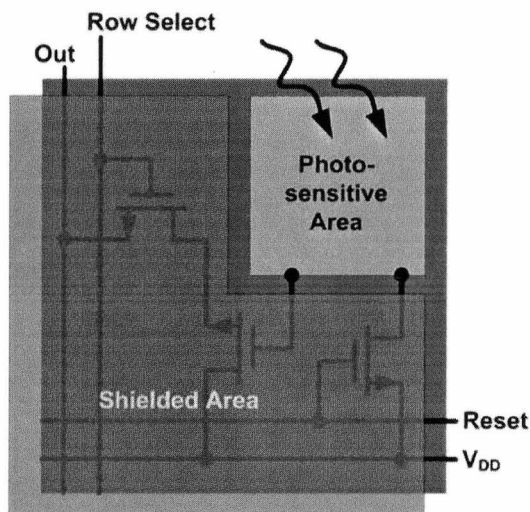


Figure 1.7: The fill-factor in a pixel is the ratio of the photosensitive area to the area of the pixel.

One of two dominant technologies can be used to construct solid-state image sensors, which are charge-coupled devices (CCDs) and CMOS imagers. These technologies are explained in the following subsections, where a comparison between the different readout architectures is also provided.

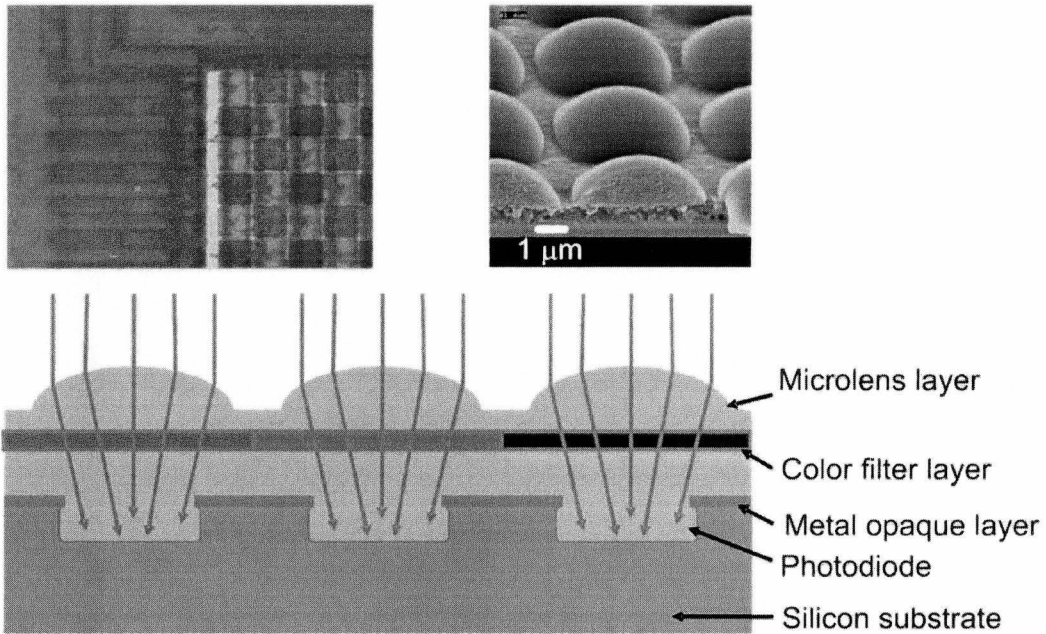


Figure 1.8: Color detection in solid-state imagers using microlens color filters [10].

1.2.2.1. Charge-Coupled Devices (CCDs)

When CCDs were first reported in 1970 [11], they became the most popular image sensor mainly due to their freedom from fixed-pattern noise (FPN), which was considered the major problem with CMOS image sensors [12]. CCDs have simpler and smaller pixel sizes, which also contributed to their market dominance [13]. In addition, CCDs are superior in terms of signal-to-noise ratio (SNR) and dynamic range (DR) especially for high quality still photography [14]. Figure 1.9 (a) shows the cross-section of a CCD pixel [15]. When the energy from an incident photon is absorbed in the detector, charges are generated in a potential well. These charges need to be measured and read out, but since each cell does not contain a charge measuring unit, the charges have to be serially transferred from one detector to the next down a column and then from one column to the next down a row. The CCD readout mechanism is illustrated in Figure 1.9 (b) to (e).

Since CCDs do not contain any charge conversion components in the pixel, they have a very high fill-factor and do not suffer from charge variation from pixel to pixel, which results in FPN. These advantages have enabled CCD imager to produce very high quality images. However, CCDs have some disadvantages when compared to CMOS imagers. The serial charge transfer readout in CCDs results in limited speed. CCDs also consume

more power due to the need for high-rate, high-voltage clocks to guarantee good charge transfer efficiency. Usually, to ensure that the charge transfer efficiency is high, special processes might be used, which prevents CCDs from being integrated with other processing circuit blocks on a single chip. Also, to read out a single pixel, the entire array needs to be read out, which prevents the ability to immediately access a specific region-of-interest (ROI) in an image. In addition, each pixel can only be read out once, since the readout process is destructive. Finally, when using CCDs for high sensitivity, low-level light biomedical applications they must remain cooled, which may limit device portability or increase system cost.

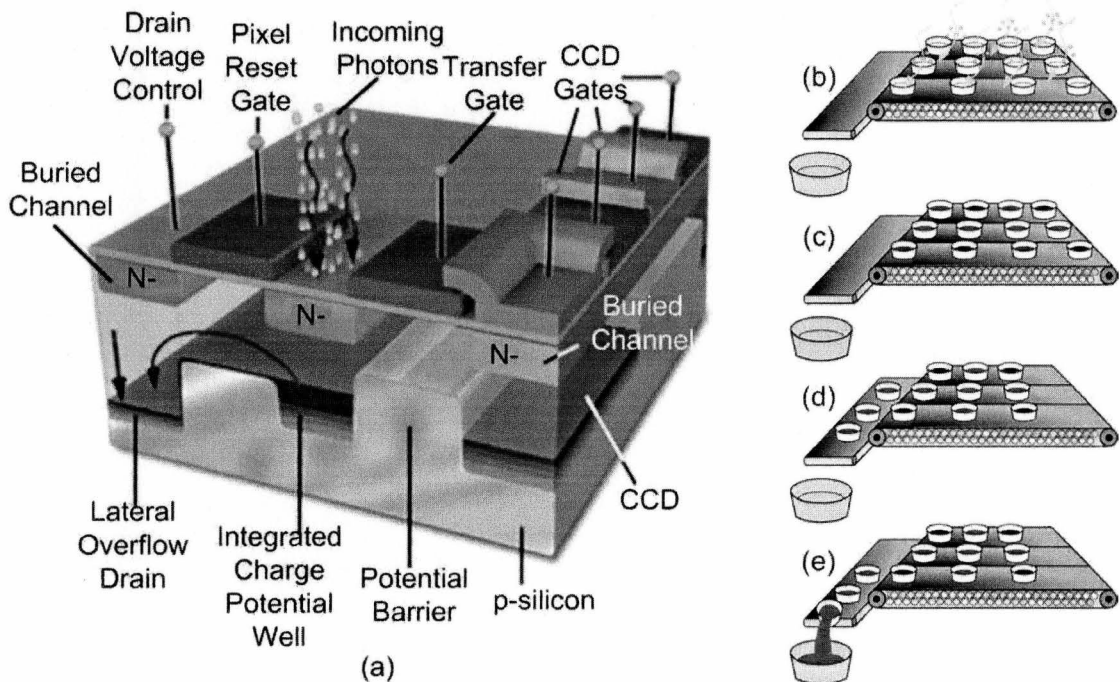


Figure 1.9: (a) The cross-section of a CCD pixel [15], and (b) – (e) charge-bucket brigade illustration of CCD charge transfer mechanism.

1.2.2.2. CMOS Imagers

With the enormously expanding camera phone market, where shipments have doubled from 2003 to 2004 [16], CMOS image sensor shipments have surpassed charge-coupled devices (CCDs). While CCDs still maintain a substantial market share due to their preferred use in digital still cameras and camcorders, 230 million CMOS sensors were shipped in 2004 with an estimated annual growth rate of over 28% [12]. Of the cameras

shipped in 2005, camera phones cover around 70%, of which, most use CMOS sensors [17]. CMOS sensors are expected to see even more growth mainly due to their low-cost and low-power consumption, making it the technology of choice for low-end digital still camera markets such as mobile phones, toys, embedded cameras in PCs and notebooks specifically, which hit the market in 2005, and automotive rear-view cameras in minivans for example [17], [18]. Exceeding 400 million units in 2005, the camera phone market will continue to drive CMOS image sensors [19]. The total imaging market grew over 30% annually by the end of 2008, with CMOS image sensors owning the majority of the growth [18].

CMOS imagers were improved in the early 1990s by two independent drivers, single-chip high functionality imaging systems, where low-cost was the driving factor, and NASA's deep-space exploration spacecraft that needed highly miniaturized, low-power imagers with a high-performance driving factor [13]. With the advances in deep-submicron CMOS technologies and integrated microlenses, CMOS imagers, specifically the active-pixel sensor (APS), have become a practical alternative to the long dominating CCD imaging technology. Perhaps the main advantage of CMOS image sensors is that they are fabricated in standard CMOS technologies, which allows full integration of the image sensor along with the analog and digital processing and control circuits on the same chip. This camera-on-chip system leads to reduction in power consumption, cost and sensor size and allows for integration of new sensor functionalities.

In terms of pixel size, it is true that CCDs offer smaller pixels than CMOS APS sensors, however this advantage is practically limited by both optical physics (the light wavelength) and optics cost [14]. Pixel sizes smaller than 4~5 μm per pixel side are not considered preferable [14]. Small pixels are needed, especially in mobile imaging, to increase the spatial resolution of the imager without increasing the area of the sensor itself. Since a smaller pixel has lower light sensitivity and dynamic range, some novel designs reduce the number of transistors per pixel by sharing transistors among a group of pixels [21], [22], thus increasing the FF. Higher FF is desired since it allows for shorter exposure times for a fixed pixel size or for a smaller pixel size for a given photo-sensitive area. Figure 1.10 shows a comparison between the different readout

architectures in CCD and CMOS systems [1]. CMOS imagers allow for array random access and selective ROI readout and can operate at higher speeds than CCD imagers.

CMOS APS usually have a FF of around 30% and the FF is typically limited by the interconnection metals and silicides that shadow the photosensitive area and recombination of the photo-generated carriers with majority carriers. When comparing CMOS photodetectors to PMTs, aside of price and size, CMOS photodetectors can achieve frame rates that are 16,000 times faster than commercial PMTs, 3,200 frames/s compared to 0.2 frames/s, with densities that can be 1000 times higher, 4590 fluorophores/ μm^2 compared to 4.49 fluorophores/ μm^2 [23], [24]. Table 1.2 shows a summary of the different imaging schemes that were discussed.

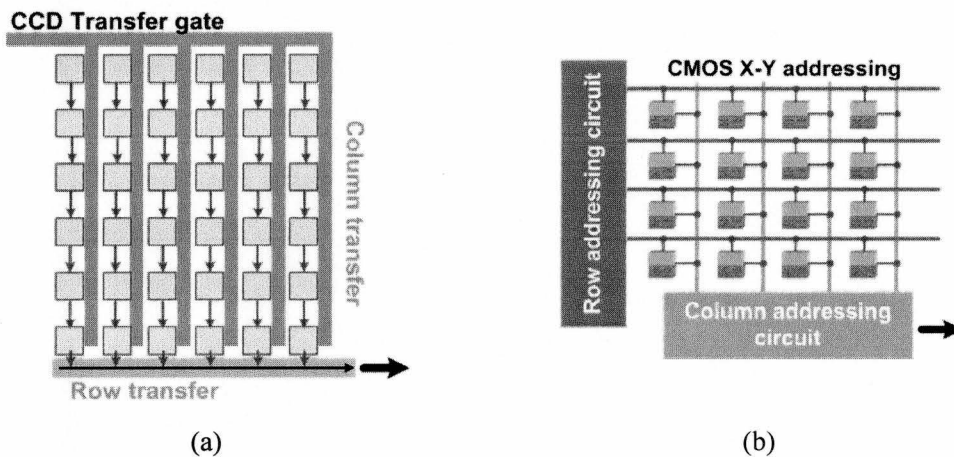


Figure 1.10: Different readout architecture in CCD and CMOS systems [1].

Table 1.2: Comparison between PMT, CCD and CMOS imagers, reproduced from [26].

	PMT	CCD	CMOS
Sensitivity	High	Moderate	Moderate-low
Power	Very high	High	Low
Speed	Moderate	Slow	Very fast
Cost	Very expensive	Moderate	Cheap
Customization	No	No	Yes
Integration	System level	Board level	Chip level
Format	Single	Array	Array

1.3. Scaling Effects

The resolution of a solid-state image sensor array can be increased either by increasing the number of pixels, hence increasing the array size, or by keeping the size of the array

fixed and reducing the size of the pixel. The latter case is only possible by technology downscaling. However, downscaling may affect the detector differently than how it affects the pixel or array, as explained in the following two subsections.

1.3.1. Pixel and array level

For a fixed array size, increasing the number of pixels, by reducing the pixel size, results in an immediate increase in resolution, as shown in Figure 1.11. Downscaling can also allow for an increased number of transistors to be integrated within a pixel, while keeping a fixed pixel size, which can increase the functionality within the pixel, hence allowing for increased frame rates, dynamic range, or for real time hardware compression. Reducing the pixel size also helps in reducing the power consumption, cost, lens volume, camera volume and camera weight.

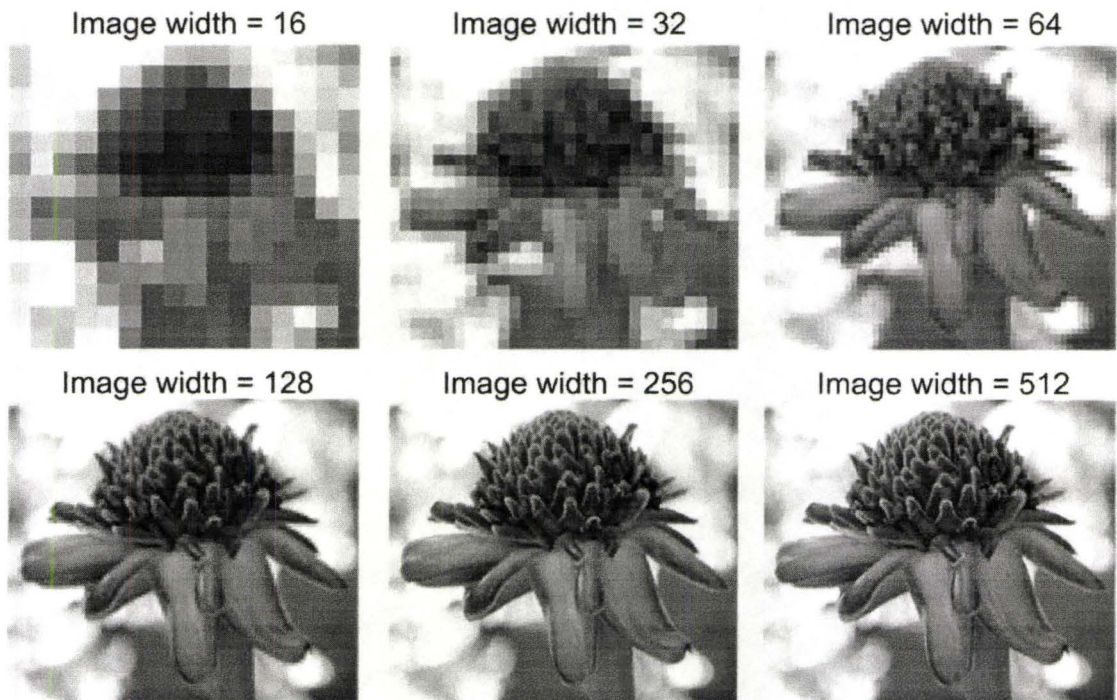


Figure 1.11: Image resolution as a function of number of columns for a fixed image size.

Downscaling however, requires the design of optics to match the smaller pixel size, which is much more challenging as the diffraction limit is reached. Diffraction occurs when light passes through a small aperture, which sets a practical limit on the size of microlenses to somewhere between 1-5 μm [27], [28], [29], [30], [31]. Once diffraction is reached, the image will begin to blur, making any further increase in resolution

ineffective. The optical efficiency can be reduced due to light diffraction by about 40% for a $3.2\ \mu\text{m}$ pixel and 75% for a $1.45\ \mu\text{m}$ pixel [27].

Figure 1.12 shows the downscaling of CMOS technology based on the ITRS roadmap compared to some of the state-of-the-art CMOS imagers showing the pixel size and the technology node used [25]. CMOS imagers usually do not use the most advanced technology available since it is not imager friendly, as will be explained in the following section showing the impact of downscaling on photodetectors. The pixel size is usually 20 times the minimum feature size of the technology used. The figure also shows an optical limit for visible light. This limit is due to the diffraction caused by light passing through the lens of the imager that causes the light to interfere with itself. This sets a limit on the spatial resolution of the imager, limiting how small the pixels can be ($\sim 5\ \mu\text{m}$). Now that the technology nodes have scaled well below the optical diffraction limit, the main use of pixel downscaling will be designing smart pixels, such as the ones that will be shown in this thesis. For example, when using a CMOS $0.18\ \mu\text{m}$ technology with a $5\ \mu\text{m} \times 5\ \mu\text{m}$ pixel and a 30% FF, 8 analog transistors or 32 digital transistors can be integrated within the pixel [30].

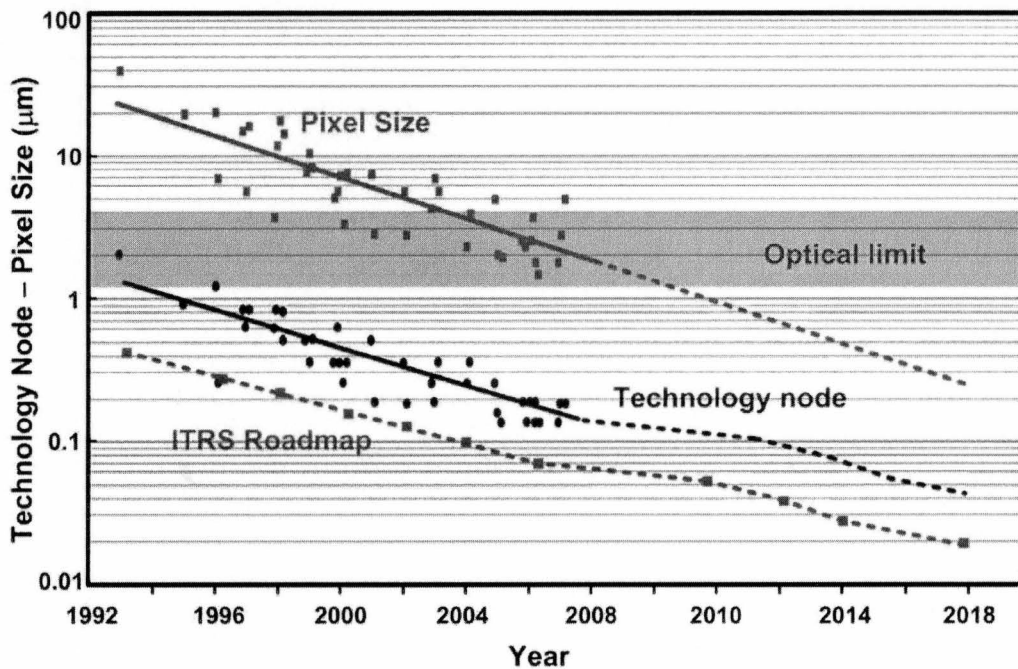


Figure 1.12: Downscaling of CMOS technology compared to some of the state-of-the-art CMOS imagers, reproduced and modified from [25].

1.3.2. Detector level

The SNR and DR of a photodetector are reduced with pixel downscaling and size reduction. When the area of a photodetector is reduced, the DR is reduced since the potential well can store less amounts of charge. The SNR will be explained in Chapter 2 to be given approximately by

$$\text{SNR}_{\max} \approx 10 \log \left(\frac{V_{\text{DD}} C_{\text{PH}}}{q} \right), \quad (1.1)$$

where C_{PH} is the photodiode's capacitance and V_{DD} is the source voltage. C_{PH} depends on the diode area in addition to its doping concentration and profile, which are affected with downscaling and pixel size reduction. V_{DD} is also reduced with downscaling. The maximum SNR can be evaluated as pixel size decreases with technology downscaling, as shown in Figure 1.13 [32].

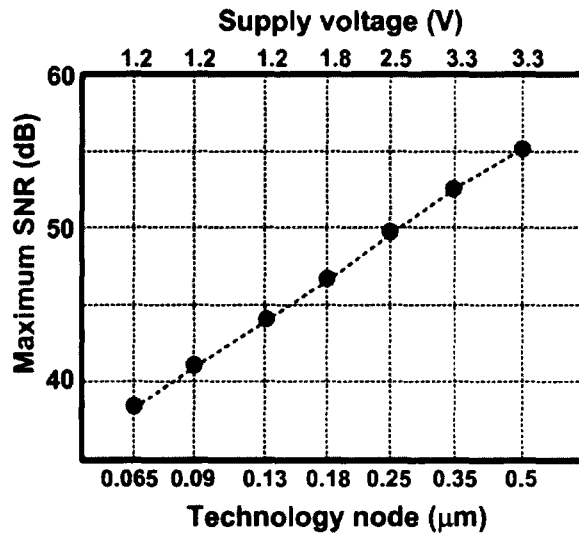


Figure 1.13: The maximum SNR of an APS pixel as the technology scales down, reproduced from [32].

The internal quantum efficiency of a photodiode, which depends on process parameters and diode geometry, varies as the diode is scaled. Also, the tunnel that light passes through from the surface of the photodiode to the light absorption region becomes narrower, while the depth of field does not scale as much, causing a reduction in quantum efficiency. Finally, the peak quantum efficiency photo-response shifts towards shorter wavelengths as the junction gets shallower. This however, may be beneficial for

biomedical applications such as fluorescence imaging where the light response tends to be in the blue region.

The increased doping levels as technology downscales also reduce the chance that minority carriers that are generated outside of the depletion region of the photodiode will diffuse to the depletion region and be detected. The increased doping levels decrease the carrier lifetime and diffusion coefficient of electrons. As a result, carriers generated deep in the substrate have a less chance of being detected, which reduces the quantum efficiency and slightly shifts the spectral response towards the blue region. This is mostly because of the shallower junctions.

Reducing the area of the photodetector also increases the effect of mismatch between pixels, which will cause an increase in FPN. Shot noise and dark current also increase with downscaling as doping levels increase resulting in narrower depletion regions. The narrow depletion region also results in reducing the breakdown voltage of avalanche photodiodes (APD) that provide gain based on high electric field avalanche breakdown carrier multiplication. At high doping levels used in CMOS 180 nm or 130 nm technologies, tunneling breakdown is more likely to occur before avalanche, causing a large increase in dark count noise [32]. This predicts that CMOS 180 nm might be the smallest scale that can allow for APDs to be easily implemented using main stream technology.

1.4. Motivation

Recent advances in deep submicron CMOS technologies and improved pixel designs have enabled CMOS-based imagers to surpass CCD imaging technology for main stream applications. The parallel outputs that CMOS imagers can provide offer complete camera-on-a-chip solutions due to being fabricated in standard CMOS technologies, which result in compelling advantages in speed and system throughput. CMOS technology scaling can allow for an increased number of transistors to be integrated into the pixel to improve both detection and signal processing. Such smart pixels truly show the potential of CMOS technology for imaging applications, allowing CMOS imagers to achieve the image quality and global shuttering performance necessary to meet the demands of ultrahigh-speed applications.

This work focuses on designing low-light level imagers in CMOS technology for biomedical applications that can be suitable for extremely high-speed applications, such as FLIM. The design and implementation of ultrahigh acquisition rate CMOS imagers that can take 8 frames at a rate of 1.25 billion frames/s, resulting in one of the fastest imagers in the world at present, is described. The high acquisition rate that this work achieved is necessary for high-speed applications such as FLIM or high energy physics experiments.

Increasing the sensitivity of CMOS imagers was addressed by using single-photon avalanche photodetectors (APD) with an extremely short deadtime. The deadtime of a single-photon APD is what will limit the rate at which photons can be counted. By successfully fabricating an avalanche photodiode in a deep-submicron CMOS technology, an active quench and reset single-photon detector pixel that has a deadtime 40 times lower than previous published work was designed. A novel time-domain single-photon counter that can capture images with high dynamic range and high sensitivity while maintaining high speed, is also presented.

The final impact of this work would result in designing a complete, portable, cheap, small size and low-power camera-on-a-chip that is suitable for high-speed FLIM. Such an imager can be added on to the tip of an endoscope or encapsulated within a swallowable endoscopy capsule. This will make FLIM characterization more available to the medical and scientific research community and better the quality of healthcare specifically for cancer patients.

1.5. Contributions

The work discussed in this thesis has resulted in a number of publications, patents and awards. The major contributions from this work are:

1. An image sensor complete camera-on-a-chip measurement setup was designed that is suitable for controlling and processing CMOS imagers. The setup was tested by designing and measuring a fully integrated camera-on-a-chip implemented in a standard 0.18 μm CMOS technology.

2. An ultrahigh-speed CMOS imager was designed that can capture 8 frames with a frame capture rate that is higher than 1.25 billion frames per second, which to the author's best knowledge, is currently one of the fastest cameras.
3. Avalanche photodiodes have been implemented in this work, in a mainstream standard digital CMOS 0.13 μm technology. The devices have been characterized and their single-photon counting (SPC) behavior has been modeled into the circuit simulator.
4. Based on the measurements of the single-photon counter, an integrated active quench and reset circuit has been designed. The circuit achieves deadtimes as low as 200 ps, which is significantly lower than previous work. The circuit also has a higher fill-factor of 25%, compared to 1-5% in previous work.
5. A novel deadtime reduction technique was designed for active quench and reset circuits that slightly sacrifices fill-factor but maintains high sensitivity.
6. A novel high dynamic range imager was designed using single-photon counting in time-domain. The design can achieve high sensitivity and high dynamic range, while maintaining a speed that is around 1000 times faster than conventional time-domain imagers.
7. An imager that allows for simultaneous pixel counting and threshold detection in time-domain was also designed in order to further improve the frame rate.
8. A novel analog counter was designed for in-pixel counting in order to simplify the pixel design and increase the fill-factor.
9. The analog counter can also be used in a novel pixel design that allows simultaneous pixel counting and analog-to-digital conversion for single-photon counters.

Awards Received:

1. The Natural Science and Engineering Research Council of Canada (NSERC) two year Industrial Research and Development Fellowship (April 2010)
2. Saudi Cultural Bureau Student Travel Grant, San Francisco, USA (Jan. 2010)
3. The Dean's Award for Excellence in Communicating Graduate Research (Sep. 2009)

4. The Ontario Graduate Science and Technology (OGSST) Raymond Moore two year scholarship (July 2009)
5. The Society of Solid-state and Electrochemical Science and Technology Student Travel Grant, Hawaii, USA (Oct. 2008)
6. Seven Canadian Microelectronics Corporation (CMC) chip allocation grants (2006-2009)
7. The NSERC three year doctorate scholarship (April 2006)
8. The Ontario Graduate Scholarship (OGS) (April 2006)
9. The KACST of Saudi Arabia Scholarship (Jan. 2006 - Dec. 2010)
10. The Saudi Ministry of Higher Education Scholarship (Jan. 2006 - Dec. 2010)

Patents Submitted to USPTO:

1. Reduction of delay between subsequent capture operations of a light-detection device. Application number: 12769988. April 29th, 2010.
2. High speed imaging through in-pixel storage. Application number: 12814443. June 12th, 2010.
3. High-speed analog photon counter and method, Sep. 1st, 2010.
4. Coordinated in-pixel intensity detection method and apparatus, Sep. 1st, 2010.
5. Single photon counting image sensor and method, Sep. 1st, 2010.

Published/Accepted Refereed Journal Papers:

1. M. M. El-Desouki, O. Marinov, M. J. Deen, and Q. Fang, "CMOS active-pixel sensor with in-situ memory for ultrahigh-speed imaging," *IEEE Sensors Journal*, 5 pages, accepted Oct. 15th, 2010, published online Oct. 21st, 2010.
2. M. M. El-Desouki, D. Palubiak, M. J. Deen, Q. Fang, and O. Marinov, "A Novel, high dynamic range, high speed and high sensitivity CMOS imager using time-domain single photon counting and avalanche photodiodes," *IEEE Sensors Journal*, 7 pages, accepted June 22, 2010, published online Sep. 23rd, 2010.
3. (Invited) M. M. El-Desouki, M. J. Deen, Q. Fang, L. Liu, F. Tse and D. Armstrong. "CMOS image sensors for high speed applications," *Sensors*, vol. 9, (1), pp. 430-444, 2009.

4. M. M. El-Desouki, S. M. Abdelsayed, M. Jamal Deen, N. K. Nikolova and Y. M. Haddara, “The impact of on-chip interconnections on CMOS RF circuits,” *IEEE Transactions on Electron Devices*, vol. 56, pp. 1882-1890, May 2009.
5. N. Faramarzpour, M. M. El-Desouki, S. Shirani, Q. Fang, and M. J. Deen “CMOS imaging for biomedical applications,” *IEEE Potentials Magazine*, vol. 27, (3), pp. 31-36, May-June 2008.
6. (Invited) N. Faramarzpour, M. M. El-Desouki, M. J. Deen, S. Shirani and Q. Fang, “CMOS photodetector systems for low-level light applications,” *Journal of Materials Science: Materials in Electronics*, (Special issue), vol. 20, pp. 87-93, Jan. 2009.

Published Refereed Conference Papers and Abstracts:

1. M. M. El-Desouki, D. Palubiak, M. J. Deen, and Q. Fang, “A Novel CMOS image sensor using time-domain single-photon counting,” *The Electrochemical Society 217th Meeting*, Vancouver, Canada, April 25-30, 2010.
2. S. Majumder, M. M. El-Desouki, O. Marinov, and M. J. Deen, “Random telegraph signal noise in CMOS imagers and its impact on image quality,” *The Electrochemical Society 217th Meeting*, Vancouver, Canada, April 25-30, 2010.
3. M. M. El-Desouki, M. J. Deen, and Q. Fang. “CMOS camera-on-a-chip for fluorescence lifetime imaging,” *SPIE Photonics West (BIOS10)*, San Francisco, California, USA, Jan. 2010.
4. M. El-Desouki, M. J. Deen, S. Shirani, S. Sirouspour, F. Tse, D. Armstrong, Q. Fang, “Integrated CMOS sensors for fluorescence spectroscopy and imaging,” *Communications and Photonics Conference Exhibition (ACP)*, Shanghai, China, pp. 1-2, Nov. 2009.
5. M. M. El-Desouki, M. Jamal Deen, and Q. Fang. “A CMOS single-photon avalanche photodetector camera-on-a-chip for biomedical applications,” *14th Canadian Semiconductor Technology Conference (CSTC09)*, Hamilton, Ontario, Canada, Aug. 2009.
6. M. M. El-Desouki, M. Jamal Deen, Q. Fang and F. Tse. “Ultrahigh speed CMOS camera-on-a-chip for biomedical applications,” *European Conferences on Biomedical Optics (ECBO)*, Munich, Germany, 3-pages, June 2009.

7. (Invited) M. J. Deen, M. M. El-Desouki and N. Faramarzpour, "CMOS camera-on-a-chip: image sensors for low-light-level biomedical applications," *IEEE International Conference on Electron Devices and Solid-State Circuits 2008*, Hong Kong, pp. 1-6, December 8-10, 2008.
8. (Invited) M. M. El-Desouki, M. J. Deen, Q. Fang, F. Tse and L. W-C. Liu, "CMOS camera-on-chip image sensor for biomedical applications," *The Joint International meeting of ECS 214th Meeting and The Electrochemical Society of Japan 2008 Fall Meeting*, Honolulu, Hawaii, October 12-17, 2008.

1.6. Thesis Organization

A brief introduction to different CMOS pixel structures is presented in Chapter 2. In this chapter, the basics of light interaction with silicon semiconductors are first explained, followed by some of the most common pixel architectures. Special purpose pixel designs are also discussed. Finally, the different noise sources and noise analysis are described.

In Chapter 3, the design, simulations and measurement results of the implemented 256 pixel CMOS APS complete camera-on-a-chip in CMOS 180 nm technology, is presented. This imager was designed in order to provide a working platform and measurement setup using an FPGA interfaced to a PC and a monitor. The design was then improved using a 130 nm CMOS technology in order to achieve ultrahigh-speed imaging for FLIM that is presented in Chapter 4.

The APD based single-photon imager implemented in 130 nm CMOS technology is presented in Chapter 5. Both the passive and active quench and reset single-photon counting circuits are presented and compared to APDs implemented in deep-submicron technologies. A novel deadtime reduction and parallel pixel processing technique, using in-pixel counting, is also discussed in this chapter.

In Chapter 6, the implementation of a novel CMOS imager is discussed. The imager uses time-domain single-photon counting to achieve high-speed, high dynamic range and high sensitivity at the same time in a single frame. Finally, Chapter 7 will conclude with a summary of this work and a discussion of future work.

Chapter 2

CMOS PIXEL STRUCTURES

The basic operation of an imager is to reproduce a scene under specific illumination conditions. This is mainly done by sensing and converting photons into electrons. A simplified imager block diagram is shown in Figure 2.1. Chapter 3 will explain the design of a complete camera-on-a-chip, whereas this chapter focuses on the pixel level. Section 2.1 starts with an introduction to the basic principles of light interaction with silicon semiconductors. Section 2.2 presents the various pixel architectures while the special purpose structures are discussed in Section 2.3. Finally, the noise sources of CMOS APS are presented in Section 2.4.

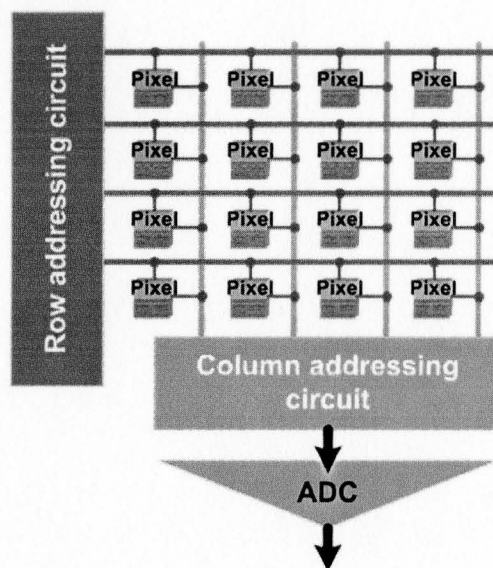


Figure 2.1: Block diagram of a CMOS image sensor.

2.1. Introduction

A photon can interact with a piece of silicon semiconductor in a number of different ways. If a photon interacts with the semiconductor lattice, it will generate heat, which is the basis of infrared (IR) thermal detectors. Photons can also interact with impurity atoms and defects. However, the interaction of interest when detecting light, is photon interaction with valence electrons, which can happen in three ways (Figure 2.2). If the energy of the photon ($E = h\nu$, and $\lambda\nu = c$) is less than the silicon bandgap (E_g), light will pass through and no photons will be absorbed by the valence electrons. If the photon energy is equal to the bandgap of silicon, which is 1.12 eV, an electron-hole-pair (ehp) will be generated, as shown in (Figure 2.2 (a)), where an electron will transit from the valence band to the conduction band, leaving behind a hole. Finally, if the photon energy is higher than the bandgap, then the excess energy will cause heat generation (Figure 2.2 (b)). The incident photons should have at least the bandgap energy, or a wavelength below $1.1 \mu\text{m}$. This corresponds to the visible light spectrum and near-IR, as shown in Figure 2.3. Light absorption is also limited if the energy of the incident photon is too high. This is because most of the photons will be absorbed immediately at the surface, without being able to penetrate to the depth of the photo-detector [33], [34]. This section explains the optical absorption properties of semiconductors as well as the different photodetectors that can be designed in silicon.

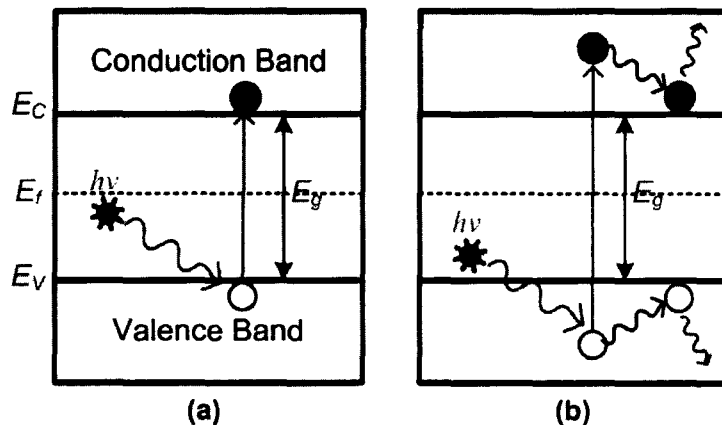


Figure 2.2: Photon interaction with valence electrons with a photon energy (a) equal to the bandgap, and (b) higher than the bandgap.

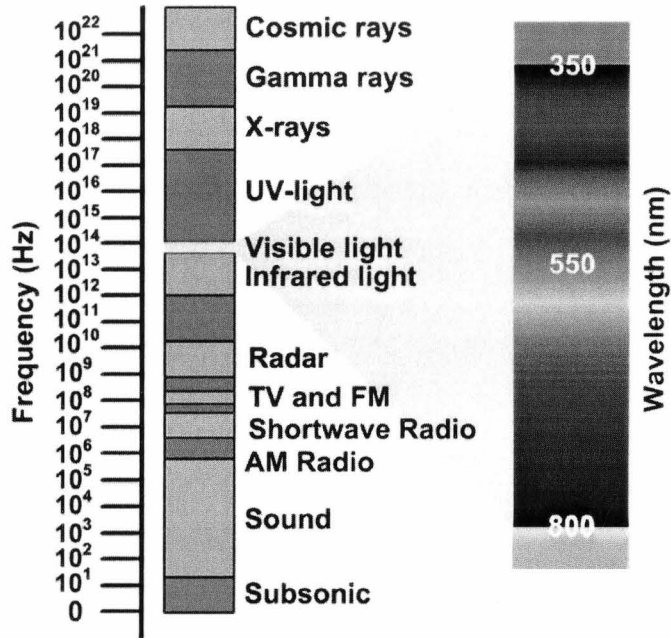


Figure 2.3: Electromagnetic spectrum showing the visible light range.

2.1.1. Optical absorption in semiconductors

Not all the light that is incident on a semiconductor will be absorbed, since some of it will be reflected, as shown in Figure 2.4 [35]. In Chapter 1, it was mentioned that the number of layers in CMOS technology increases with downscaling, causing more light reflection and photon loss from layer to layer.

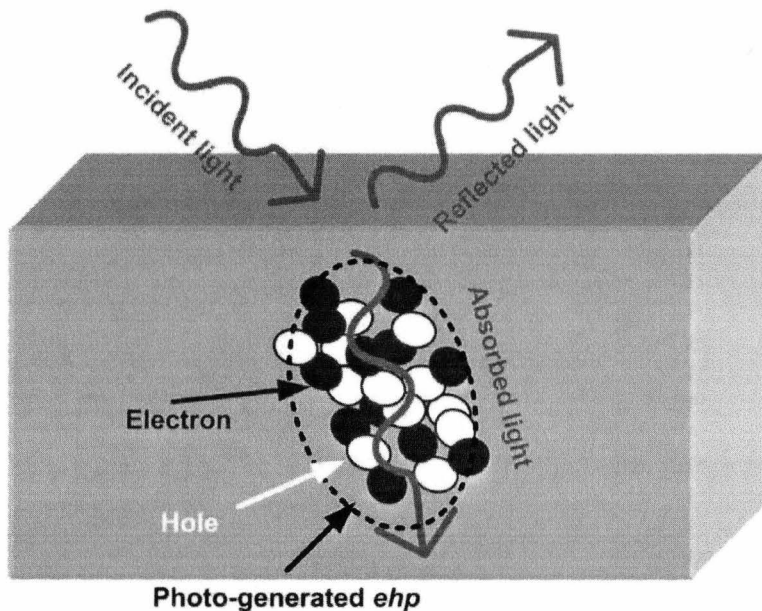


Figure 2.4: Photon absorption in semiconductors [35].

The amount of photo-generated ehp in a semiconductor depends on the absorption coefficient (α) of the material. The absorption coefficient is described by

$$\alpha(\lambda) = \frac{-1}{\Delta Z} \frac{\Delta P}{P}, \quad (2.1)$$

which is the ratio of decrease in light power (P) as it travels a distance ΔZ . From this equation, the following can be obtained

$$P(Z) = P_0 \exp(-\alpha Z). \quad (2.2)$$

From equation (2.2), the absorption length, or penetration depth can be defined as

$$L_{abs} = \frac{1}{\alpha}. \quad (2.3)$$

It was mentioned in the previous section that downscaling tends to shift the spectral response more into the blue region, which is why the absorption length is an important parameter and would usually lie in the range of 0.1-10 μm [35]. The absorption coefficient is a function of the material and the wavelength or energy of the incident photons, as shown in Figure 2.5.

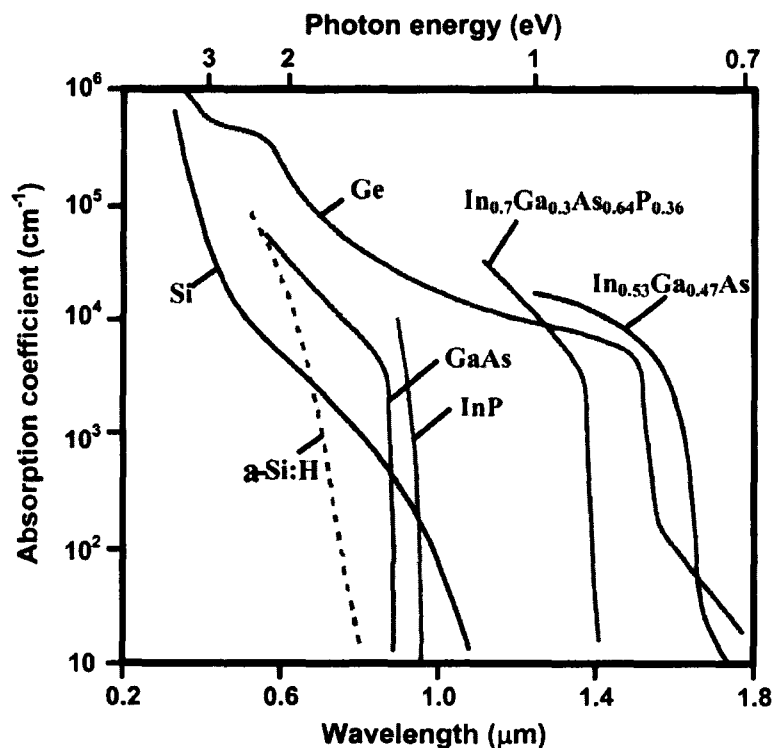


Figure 2.5: Optical absorption coefficient for various semiconductor material, reproduced from [34].

In Figure 2.5, a rapid increase can be seen at the bandedge of each material since the photon energy is high enough to generate an *ehp*. Si and Ge however, have a weaker absorption near the bandedge, because they are indirect bandgap materials that require phonon assisted transitions, as shown in Figure 2.6. The absorption coefficient is also temperature dependent since the bandgap of a material can increase or decrease with temperature, depending on the material [33].

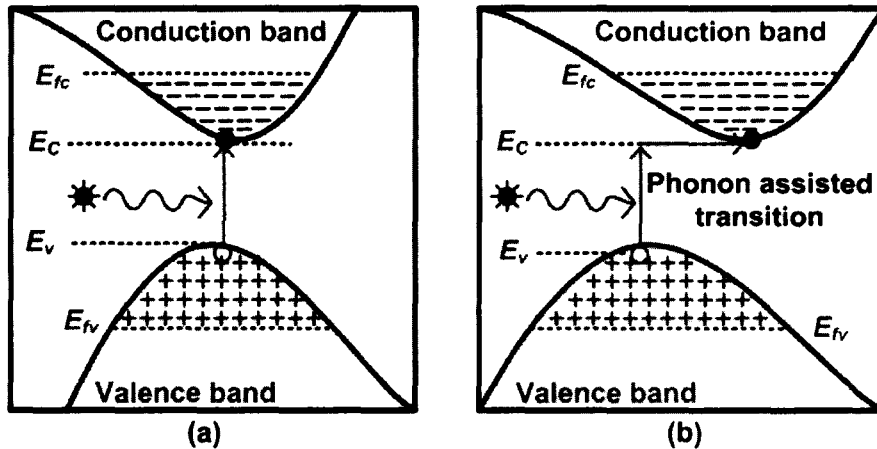


Figure 2.6: Band to band absorption in (a) direct and (b) indirect materials.

The rate at which electron-hole pairs are generated is known as the generation rate. If we assume that each photon generates one *ehp*, then the generation rate per unit volume g' is given by

$$g' = \frac{\alpha P(Z)}{h\nu}, \quad (2.4)$$

where h and ν are Planck's constant and the frequency of the light. The sensitivity of the detector (R_{ph}) is the amount of photocurrent (I_L) that is produced when one unit of light power (P_0) is incident, given by

$$R_{ph} = I_L / P_0. \quad (2.5)$$

Based on the definition of sensitivity, the quantum efficiency of the detector can be defined as external quantum efficiency or internal quantum efficiency. The internal quantum efficiency considers only the photons that are absorbed, whereas the external quantum efficiency (η_Q) takes the ratio of the number of generated photo-carriers to the number of incident photons, given by

$$\eta_Q \equiv \frac{I_L/e}{P_0/h\nu} = R_{ph} \frac{h\nu}{e}, \quad (2.6)$$

where e is the charge of an electron. The maximum sensitivity is found when the external quantum efficiency is equal to one, where $R_{ph,max} = \lambda[\mu m]/1.23$. Figure 2.7 shows the sensitivity of silicon compared to ideal maximum sensitivity that increases linearly to the bandgap cutoff above $1.1 \mu m$ [35], [36].

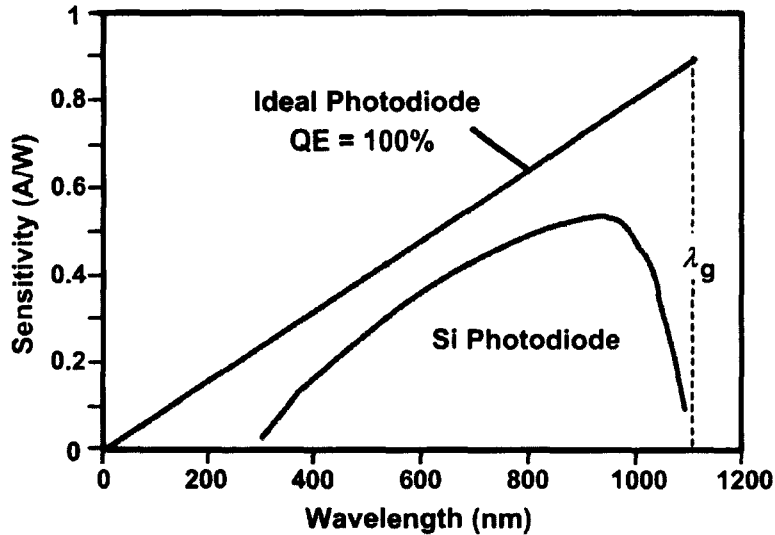


Figure 2.7: Sensitivity of silicon photodiodes, reproduced from [35], [36].

Finally, it is worth noting that the external photocurrent measured from the source that biases the photodiode (Figure 2.8 (a)) depends only on the flow of electrons and not holes, which can be shown using Ramo's theorem [37], as follows. The transit time it takes an electron and a hole to drift from the generation point to the electrodes where they recombine with carriers coming from the battery, can be found as

$$t_e = \frac{L-\ell}{v_e}, \quad t_h = \frac{\ell}{v_h}, \quad (2.7)$$

where the velocity of the carriers is related to the mobility ($\mu_{e,h}$) and the electric field as

$$v_e = \mu_e E, \quad v_h = \mu_h E. \quad (2.8)$$

Assuming a uniform electric field, to move a charge e a distance dx in a time dt by a force eE , the battery must do work that is equal to the force times the distance, which is provided by the battery as average power [37]

$$V i_e(t) dt = e E dx \rightarrow i_e(t) = \frac{e E dx}{V dt}. \quad (2.9)$$

This equation leads to

$$i_e(t) = \frac{e v_e}{L}, \quad i_h(t) = \frac{e v_h}{L}, \quad (2.10)$$

Using these equations in Figure 2.8 (b) and summing the two currents as shown in Figure 2.8 (c), we can integrate the total external current to obtain the total charge as:

$$Q_{tot} = \int_0^{t_h} \frac{e v_h}{L} dt + \int_0^{t_e} \frac{e v_e}{L} dt = t_e \left[\frac{e v_h}{L} + \frac{e v_e}{L} \right] + (t_h - t_e) \left[\frac{e v_h}{L} \right],$$

$$Q_{tot} = t_e \frac{e v_e}{L} + t_h \frac{e v_h}{L} = \frac{e(L-l)}{L} + \frac{e l}{L} = e. \quad (2.11)$$

The total collected charge, which can also be shown from the area under the curve in Figure 2.8 (c), is only e , rather than $2e$.

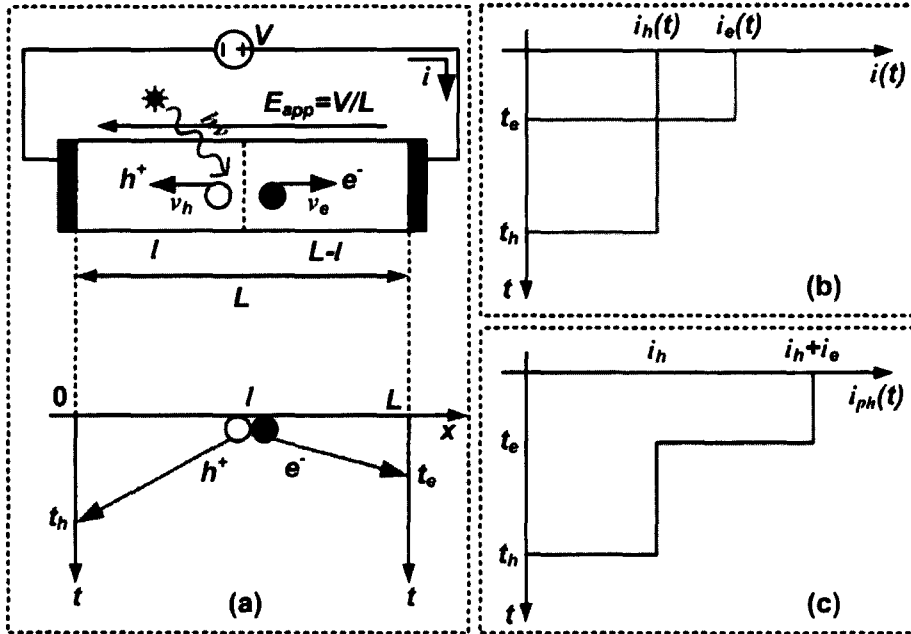


Figure 2.8: An ehp generated at $x = l$. (a) the electron and hole drift times. (b) The generated external current corresponding to electrons and holes, and (c) the total photocurrent, reproduced from [35], [37].

2.1.2. Silicon photodetectors

There are a number of different photodetection devices, which can be categorized as photoemission detectors that emit an electron once a photon is received, thermal detectors that give rise to temperature with incident photons, and photoelectric detectors that generate electron-hole pairs. The later is the focus of this section.

2.1.2.1. The photoconductive detector (PCD)

The photoconductive detector is a bar of semiconductor material (Figure 2.9) that has a photoconductivity, which is a conductivity that can change with optical excitation. This will allow the device to operate as a light variable resistor that can be placed in a voltage divider network to obtain a light varying output voltage.

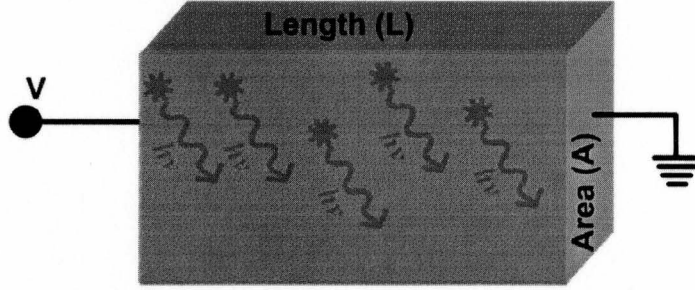


Figure 2.9: A photoconductor with a length L and an area A .

The dark conductivity (σ_{dark}) of a photoconductor is given by

$$\sigma_{dark} = e \left[\mu_n n_{dark} + \mu_p p_{dark} \right], \quad (2.12)$$

where n_{dark} and p_{dark} are the electron and hole concentrations, and excess carriers (δn , δp) are generated when light shines on the PCD, changing the conductivity to σ_{light} given by

$$\sigma_{light} = e \left[\mu_n (n_{dark} + \delta n) + \mu_p (p_{dark} + \delta p) \right]. \quad (2.13)$$

If we assume that the excess carrier concentration is equal for electrons and holes [34], and with $\delta p = G_L \tau_p$, where G_L is the generation rate of excess carriers and τ_p is the excess carrier lifetime, equation (2.13) can be modified to

$$\sigma_{light} = e \left(\mu_n n_{dark} + \mu_p p_{dark} \right) + e (\delta p) (\mu_n + \mu_p). \quad (2.14)$$

Finally, the photoconductivity (change in conductivity due to light) can be given as

$$\Delta \sigma = e \delta p (\mu_n + \mu_p). \quad (2.15)$$

By using the photoconductivity equation with Figure 2.9, the total current passing through the device based on the electric field (E) that is produced due to the applied voltage, is

$$I_{total} = A (\sigma_{dark} + \Delta \sigma) E, \quad (2.16)$$

and the generated photocurrent, assuming uniform generation, is

$$I_{ph} = e G_L \tau_p (\mu_n + \mu_p) A E. \quad (2.17)$$

The electron transit time (t_{tr}) is given by $L/\mu_n E$, which gives us

$$I_{ph} = eG_L \left(\frac{\tau_p}{t_{tr}} \right) \left(1 + \frac{\mu_p}{\mu_n} \right) AL. \quad (2.18)$$

If each ehp contributes one charge, we get

$$I_{ph'} = eG_L AL, \quad (2.19)$$

which implies that there will be a gain given by

$$gain = \left(\frac{\tau_p}{t_{tr}} \right) \left(1 + \frac{\mu_n}{\mu_p} \right). \quad (2.20)$$

This is the main advantage of using a PCD since it can provide gain of up to a thousand in Si. Figure 2.10 explains how gain can be produced in a PCD. Each time an ehp is produced (Figure 2.10 (a)), both carriers drift to the contacts, however, the electron reaches the contact before the hole recombines with an electron in the i -region or at the contact (Figure 2.10 (b)). This requires additional electrons to be injected from the contact in order to maintain charge neutrality (Figure 2.10 (c)), which gives rise to gain since each electron passing through the contacts contributes to the photocurrent [36]. The gain can be controlled by increasing τ_p , however, this trades off with device speed. Also, PCDs have very high dark current since they are always conducting, unlike PN-photodiodes that operate in the reverse biased region, which are explained next.

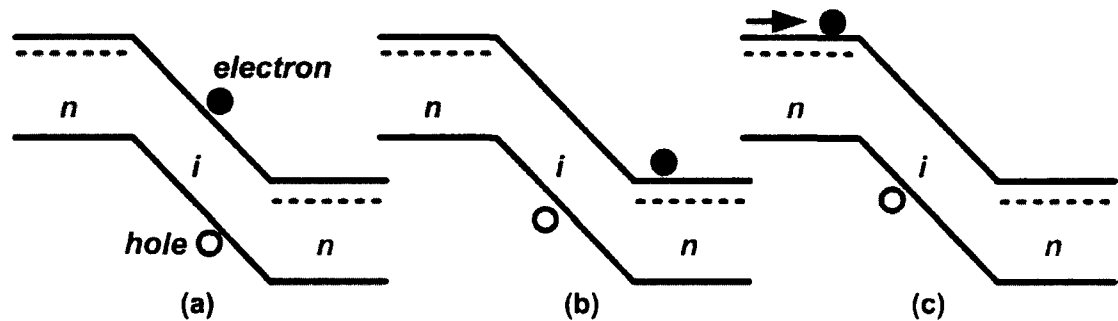


Figure 2.10: Schematic representation of how gain is produced in an NiN photoconductor.

2.1.2.2. The PN-photodiode

A PN-junction is formed by placing p-type and n-type semiconductors next to each other, where majority carriers will diffuse and electrons will leave behind positive donor ions while holes will leave behind negative acceptor ions (Figure 2.11 (a)). The diffusion will give rise to an electric field that will oppose the diffusion force and carriers will recombine creating a depletion region (Figure 2.11 (b)). Finally, at thermal equilibrium, both fluxes balance out and the net current flow is equal to zero.

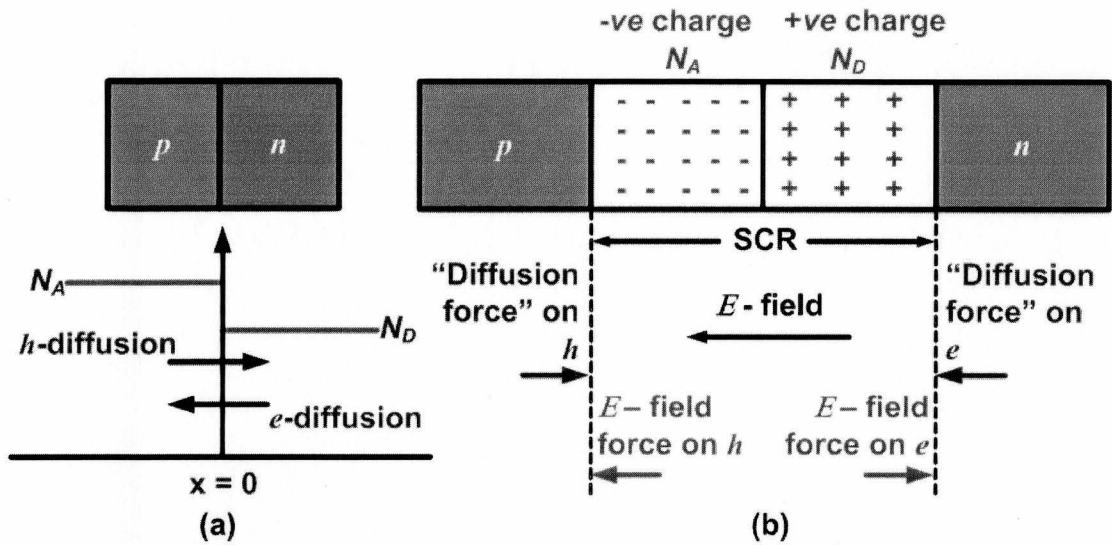


Figure 2.11: Formation of a PN-junction. (a) Carrier diffusion. (b) Thermal equilibrium.

Figure 2.12 shows the operation of a PN-junction under different biasing conditions. When no external bias is applied (Figure 2.12 (a)), electrons see a built-in potential barrier that maintains equilibrium and no current is produced. As soon as any bias is applied, the device is no longer in thermal equilibrium. If a reverse bias is applied (Figure 2.12 (b)), the charges in the space charge region (SCR) will increase, which results in an increase in the width of the SCR since the concentration is constant. In this case, the potential barrier height increases, resulting in negligible diffusion. However, drift currents still exist due to ehp generation in the SCR, which result in leakage. Finally, under forward bias conditions (Figure 2.12 (c)), the potential barrier, as well as the electric field and the SCR width are reduced, allowing diffusing currents.

Figure 2.13 (a) shows the ideal forward bias junction current density components, which give a total current by summing the minority hole and electron diffusion currents at the junction boundaries ($x_n, -x_p$). Minority drift in the bulk is ignored since the electric field in the bulk is assumed to be zero. The total current density is given by [34]

$$J_{total} = J_p(x_n) + J_n(-x_p) = \frac{eD_p p_{n0}}{L_p} \left[\exp\left(\frac{eV_A}{kT}\right) - 1 \right] + \frac{eD_n n_{p0}}{L_n} \left[\exp\left(\frac{eV_A}{kT}\right) - 1 \right], \quad (2.21)$$

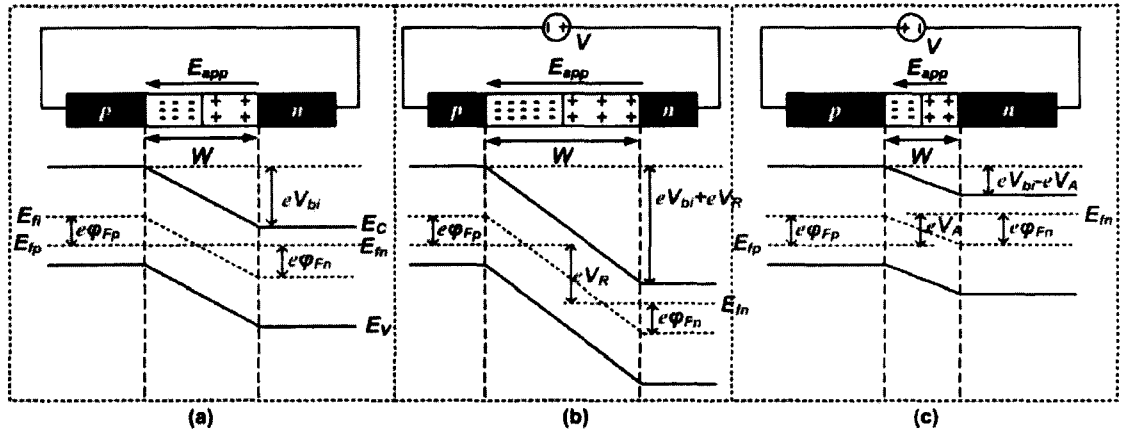


Figure 2.12: PN-junction energy band-diagrams under (a) zero bias, (b) reverse and (c) forward bias.

which can be rewritten as

$$J_{total} = J_s \left[\exp\left(\frac{eV_A}{kT}\right) - 1 \right], \tag{2.22}$$

where p_{n0} and n_{p0} are the thermal equilibrium minority carrier concentrations of holes in the n-region and of electrons in the p-region, respectively, D_p and D_n are the hole and electron diffusion coefficients, L_n and L_p are the electron and hole diffusion lengths, k is Boltzmann’s constant, and T is absolute temperature. The reverse bias saturation current (J_s - Figure 2.13 (b)) adds dark (always conducting) current to a photodiode, and it is given by

$$J_s = \left[\frac{eD_p p_{n0}}{L_p} + \frac{eD_n n_{p0}}{L_n} \right]. \tag{2.23}$$

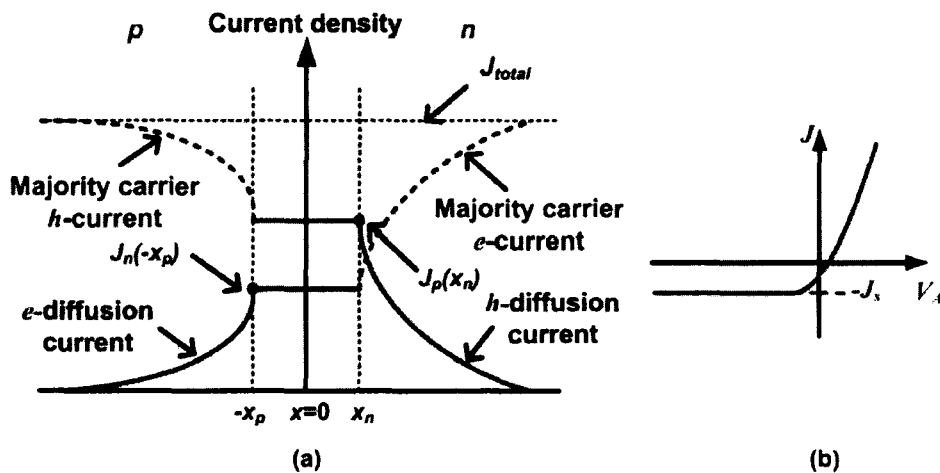


Figure 2.13: PN-junction ideal (a) forward bias total current density, and (b) reverse bias saturation current density.

Although the ideal saturation current shown in Figure 2.13 (b) seems to have no bias dependence, in reality, many other components add to the dark current that have bias dependences. Examples include generation-recombination currents, band-to-band and trap assisted tunneling currents, impact ionization currents, surface leakage currents, Frankel-Poole currents that emit trapped electrons into the conduction band, and surface recombination currents, which are specifically important for short wavelengths [33].

When a light is shun on a reverse biased PN-junction, ehp are generated in the depletion region as well as within the diffusion length of the bulk n- and p- regions (Figure 2.14 (a)), which can diffuse to the SCR and then drift by the electric field. The effect of the added photocurrent on the IV-curve of a diode is shown in Figure 2.14 (b), which shows the different regions of operation of a photodiode. The generated photocurrent is given by

$$I_{ph} = eG_L (L_p + L_n + W) A, \quad (2.24)$$

which shows that the photocurrent or device sensitivity clearly increases with the area of the photodiode. Finally, the photocurrent is also the difference between the total current and the saturation current, given by

$$I_{ph} = I_{total} - I_s \left[\exp\left(\frac{eV_A}{kT}\right) - 1 \right] \rightarrow I_{ph} \approx I_{total} - I_s. \quad (2.25)$$

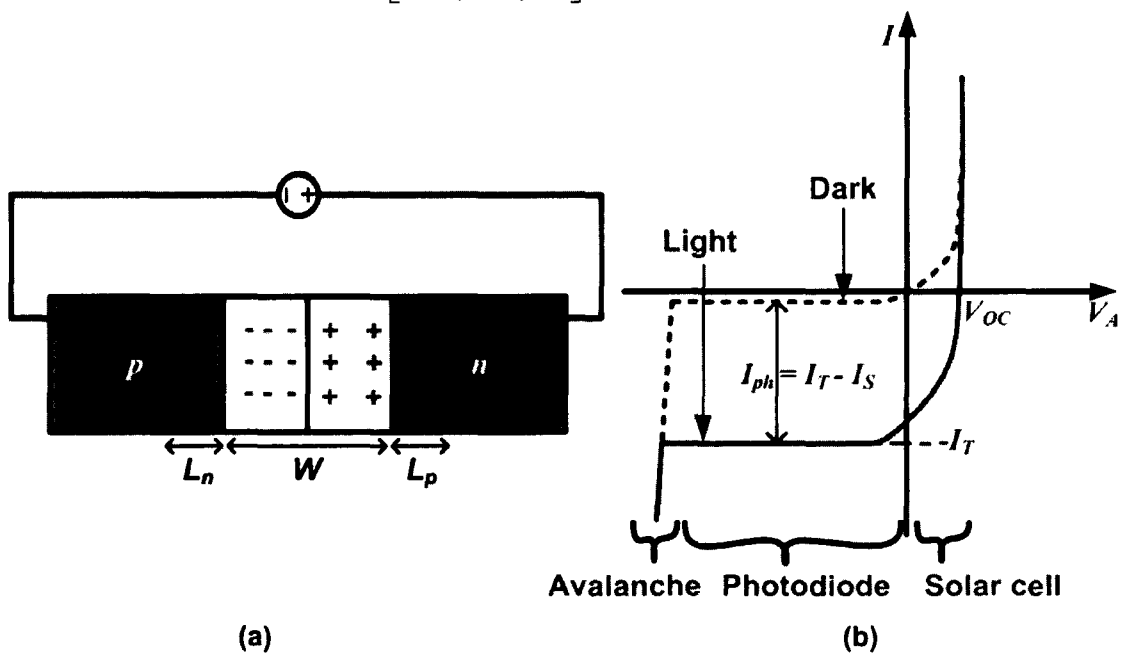


Figure 2.14: (a) Reversed bias PN-junction and (b) IV curve showing different regions of operation.

2.1.2.3. The PiN-photodiode

The PiN diode is similar to the PN-diode with the addition of a wide intrinsic semiconductor region between the p- and the n-regions. Although this helps reduce the dark current and surface leakage, the main advantage of adding the intrinsic region is to limit photo-generation to the *i*-region making drift the dominant current. This will greatly increase the speed of the device for applications such as fiber optic telecommunications, where the maximum frequency of operation of the device can be approximated as

$$f_{\max} = \frac{1}{\text{carrier transit time across intrinsic region}} \approx \frac{1}{(W_i / v_{\text{sat}})} \quad (2.26)$$

2.1.2.4. The avalanche photodiode (APD)

Generally, junction breakdown is considered a negative effect since it leads to unexpected high current values that can destroy the circuit. The two dominant types of junction breakdown are shown in Figure 2.15, which are Zener breakdown and avalanche breakdown. Zener breakdown occurs in highly doped PN-junctions through the tunneling mechanism. Avalanche breakdown occurs when carriers move across the SCR under a very large electric field, giving them sufficient energy to create *ehp* by collision. If an incident photon generates an *ehp* that triggers an avalanche, gain can be obtained by avalanche multiplication, and hence, increased sensitivity. This is also a very fast process, which gives APDs a high gain-bandwidth product. However, if an *ehp* is generated by thermal generation without an incident photon, the same process can be triggered, which means that the dark current noise is also multiplied. For this reason, APD design, specifically in CMOS technology requires a lot of care, as will be discussed in Chapter 5.

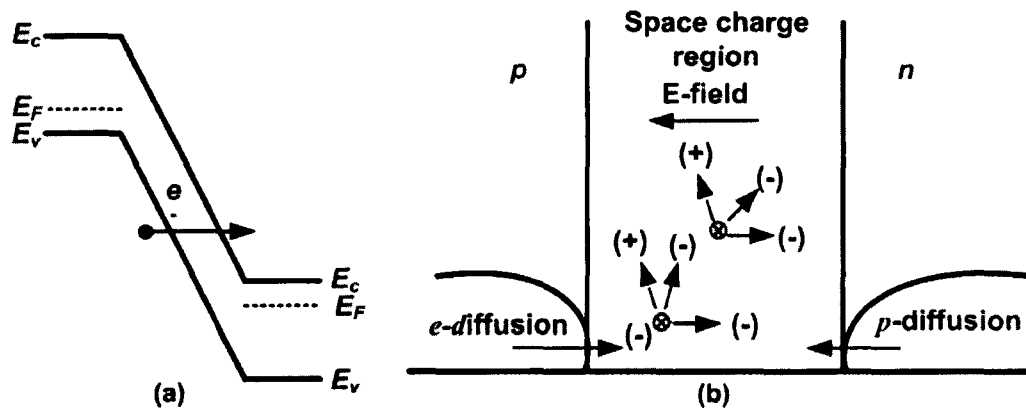


Figure 2.15: Junction breakdown by (a) Zener and (b) avalanche.

Table 2.1: Comparison of gain and response time for various photodetectors [33].

Photodetector	Gain	Response Time (s)
Photoconductor	$1-10^6$	$10^{-8}-10^{-3}$
PN-junction	1	10^{-11}
PiN-junction	1	$10^{-10} - 10^{-8}$
MSM	1	10^{-11}
APD	10^2-10^4	10^{-10}
Phototransistor	10^2	10^{-6}

Table 2.1 shows a summary that compares the performance characteristics of different photodiodes [33]. The metal-semiconductor-metal (MSM) photodetector and the phototransistors were not discussed. The phototransistor is basically a bipolar-junction transistor that generates ehp at the base and amplifies them to the collector and emitter. Whereas, the MSM detector relies on detection between slots of metal and semiconductor to reduce the effect of slow minority carrier lifetime delays and operate at a very high speed. This is at the expense of reduced efficiency due to the metal layers and increased dark current from the Schottky barriers.

2.2. Pixel Architectures

This section describes some common pixel architectures implemented in CMOS technology, such as the ones shown in Figure 2.16.

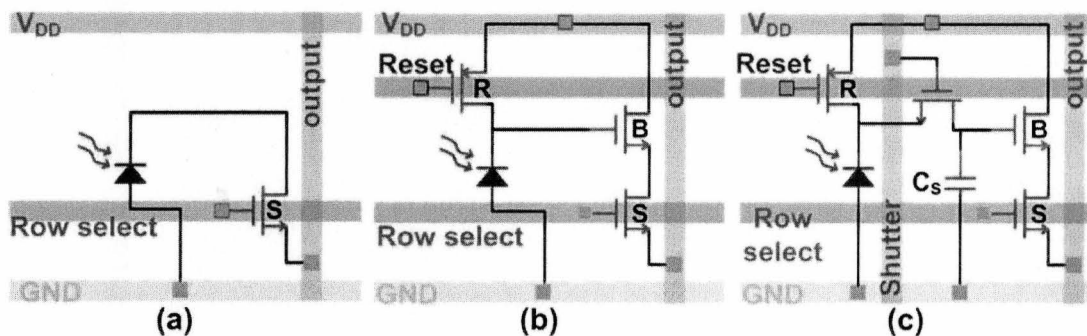


Figure 2.16: (a) PPS, (b) 3T-APS, and (c) 4T-APS.

2.2.1. *Passive pixel sensors (PPS)*

The passive pixel sensor (PPS), shown Figure 2.16 (a), is the earliest and most simple CMOS pixel structure. Each pixel consists of a photodiode and a row-select transistor. During integration, the internal capacitance of the photodiode integrates the generated photocurrent. At the end of integration, rows are selected one at a time, and connected to the column read buses. Then, the photodiodes are reset, by providing a reset voltage through the output line, and ready for the next integration cycle. PPS has only one transistor per pixel, and thus has the highest fill-factor (FF). However, column readout of the small charge that is integrated in the photodiodes causes significant loading that reduces the performance of the PPS [35]. This large column loading capacitance also results in a large thermal noise (k_bTC).

2.2.2. *Three transistor active pixel sensor (3T-APS)*

Although the three transistor active pixel sensors (3T-APS), shown in Figure 2.16 (b), originally traces back to 1968 [38], they only began to gain popularity in the mid 1990s [12], where now, they are the most popular CMOS pixel structures. Unlike the PPS, the reset in this pixel is done internally. The sense node is also isolated from the readout column by a buffering source follower transistor, which improves the SNR of this pixel compared to the PPS. This comes at the cost of a reduced FF, in addition to increased circuitry in each pixel, which increases the non-uniformity of the pixels outputs under the same illumination levels (known as fixed pattern noise, FPN). A large part of FPN is due to the variation of the sense node voltage after reset, which is known as reset noise. The dynamic range of a 3T-APS is set by the full-well capacity, which is the number of charges that can be accumulated in the photodetector. This capacity can be increased by increasing the supply voltage that the diode resets to, and by increasing the storage capacitance (diode capacitance plus parasitic capacitance of the transistors in the circuit). The supply voltage is usually limited to a specific range by the technology, whereas, increasing the capacitance (diode area) will result in a larger pixel area and lower conversion gain, which measures the change in accumulated voltage in relation to the accumulated charge [35]. This presents a tradeoff, which can be avoided by using a 4T-APS.

Figure 2.17 shows an example of the voltage integrated onto the photodiode of a 3T-APS ($v_d(t)$ of Figure 2.18). The light applied in sample 1 corresponds to a strong optical power signal, which results in a significant discharge of the stored voltage. Whereas, when applying a weak signal, such as sample 2, the discharge is lower resulting in the sample maintaining a higher voltage.

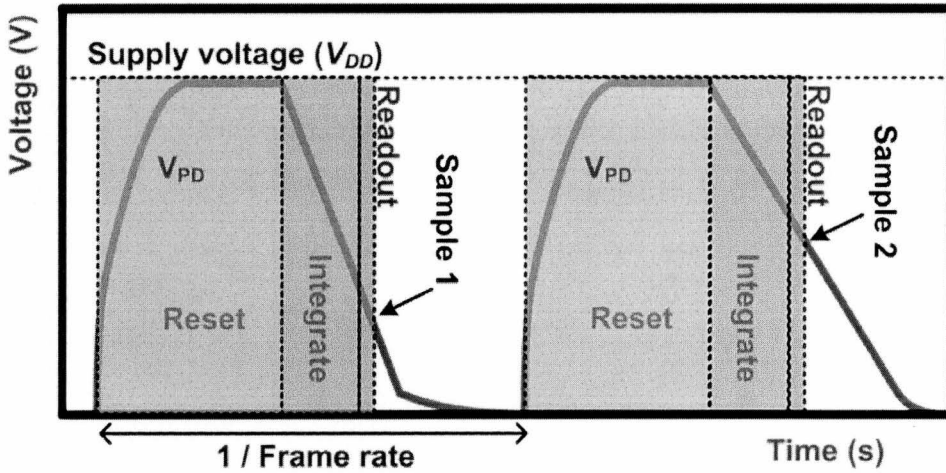


Figure 2.17: Capacitor voltage waveform of a 3T-APS.

Figure 2.18 shows a schematic diagram of a 3T-APS that is used for large signal analysis of the readout voltage. During the reset phase, the capacitance of the photodiode charges up to a full V_{DD} since a PMOS transistor (M_1) is used. Once the integration phase starts, transistor M_1 is open and the photodiode begins to discharge by the dark current and the photocurrent. The photodiode voltage is given by

$$v_d(t) = V_{DD} - \left(\frac{i_{ph} + i_{dark}}{C_d} \right) t. \quad (2.27)$$

If we assume that M_4 provides a constant bias current I_b , and if we neglect the on-resistance of M_3 and the charging time of C_s , the output voltage can be obtained from

$$I_b = K_n (v_d - v_o - V_T)^2, \quad (2.28)$$

where K_n is equal to $[(1/2)\mu_n C_{ox}(W_2/L_2)]$, and V_T is the threshold voltage of transistor M_2 .

The output voltage can be obtained as

$$v_o = v_d - V_T - \sqrt{\frac{I_b}{K_n}} \approx v_d - V_T. \quad (2.29)$$

Equation (2.29) shows that the photodiode voltage drops by a transistor threshold voltage as it passes through a source follower. For this reason, it is desirable to increase the

supply voltage of the pixel to increase the dynamic range, which can be done when using thick oxide transistors.

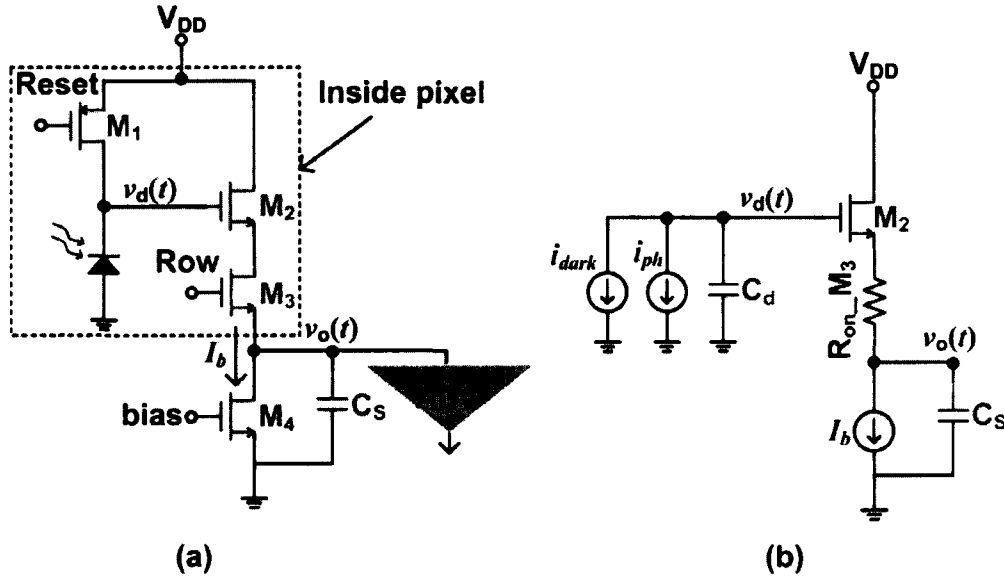


Figure 2.18: (a) Schematic diagram of a 3T-APS with simplified readout circuit and (b) the equivalent circuit during integration and readout.

Finally, the equivalent circuit used in Figure 2.18 (b) assumes that the capacitance of the diode is constant. This assumption is not correct since the capacitance is voltage dependent [39], [34]. The voltage dependence of the photodiode capacitance, assuming an initial voltage and capacitance of V_{DD} and C_0 , can be given by

$$C_d(v) = C_0 m \sqrt{\frac{V_{DD} + \varphi}{v + \varphi}}, \quad (2.30)$$

where φ is the junction built-in potential, and m is a technology parameter that specifies the type of junction, where $m = 2$ corresponds to an abrupt junction, and $m = 3$ to a linear junction. The following photodiode voltage equation

$$\frac{dv_d}{dt} = -\frac{i_{ph} + i_{dark}}{C_d(v_d(t))}, \quad (2.31)$$

was evaluated for different values of m , and the closed form solutions are shown in Table 2.2 [39]. In [39], the value of $m = 4$ was used as a good approximation for an exponential junction. The choice of m has significant effect on the analytically derived photodiode voltage, especially as integration time increases, as shown in Figure 2.19, which shows different analytic solutions of equation (2.31), assuming different values of m .

Table 2.2: Calculated $v_d(t)$ for different values of m [39].

m	Junction type	Photodiode voltage (v_d)
∞	Ideal	$V_{DD} - \left(\frac{i_{ph} + i_{dark}}{C_0} \right) t$ (2.32)
2	Abrupt	$V_{DD} - \left(\frac{i_{ph} + i_{dark}}{C_0} \right) t + \left(\frac{(i_{ph} + i_{dark})^2}{4C_0^2} \right) (V_{DD} + \phi)^{-1} t^2$ (2.33)
3	Linear	$\left[(V_{DD} + \phi)^{2/3} - \left(\frac{2(i_{ph} + i_{dark})}{3C_0} \right) (V_{DD} + \phi)^{-1/3} t \right]^{-3/2} - \phi$ (2.34)
4	Ref. [39]	$\left[(V_{DD} + \phi)^{3/4} - \left(\frac{3(i_{ph} + i_{dark})}{4C_0} \right) (V_{DD} + \phi)^{-1/4} t \right]^{-4/3} - \phi$ (2.35)

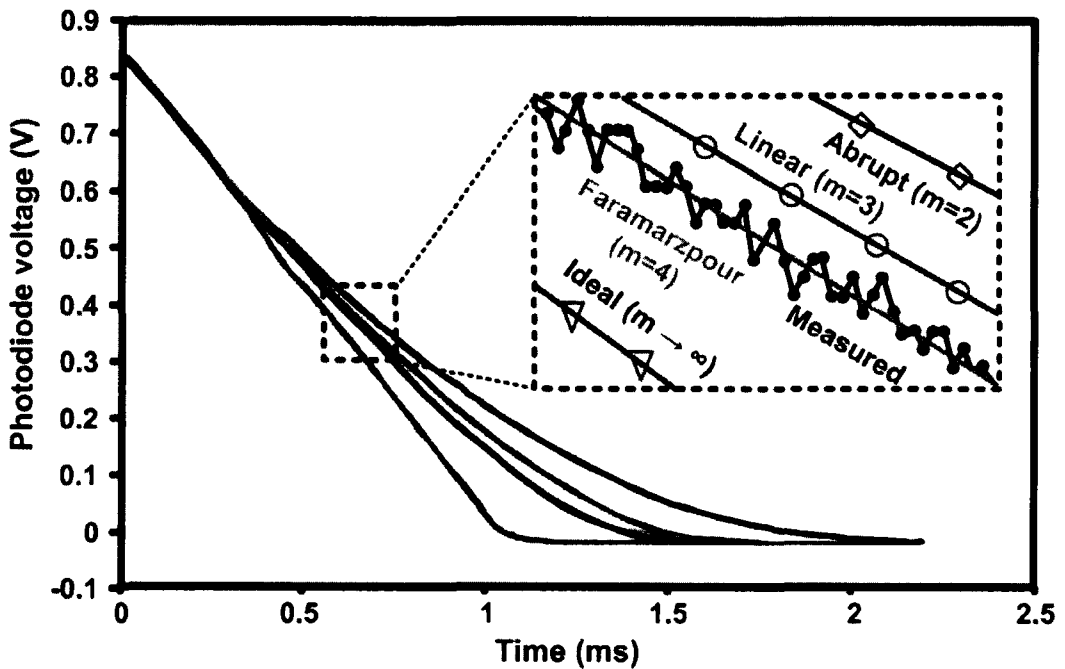


Figure 2.19: Measured and calculated photodiode voltage of a 3T-APS vs. time. The curve is magnified to show that the analytical model used by Faramarzpour in [39] best matches the measured results. Results reproduced from [39].

2.2.3. Four transistor active pixel sensor (4T-APS)

The 4T-APS, shown in Figure 2.16 (c), allows for separation of the photodetection and photo-conversion phases by adding a transfer shutter transistor. This will allow for designing the low capacitance detector required for high conversion gain, as well the large capacitor required for a large potential well. Also, the 4T-APS is necessary in order to remove FPN and to reduce the thermal noise by using correlated-double sampling (CDS). The thermal (reset) noise can be removed by ensuring that the reset operation occurs when the transfer transistor is off, avoiding passing the noise to the storage capacitor. Finally, the 4T-APS can also help reduce motion blurs, as will be explained in the high-speed section. The 4T-APS however, suffers from a reduced FF due to the extra transistor and if the charge accumulated from one frame is not completely transferred by the next frame, image lag can occur [35].

2.2.4. Correlated-double sampling (CDS)

As explained previously, this technique helps in removing reset noise as well as FPN. Figure 2.20 (a) shows an example of a CDS implementation. The circuit samples each pixel twice, once before integration and a second time after integration. The first sample stores the reset values, capturing the variation from pixel to pixel (Figure 2.20 (b)), while the second sample stores the actual image data including the FPN (Figure 2.20 (c)). Finally, the difference of the two samples is obtained using a differential amplifier and the FPN can be removed as shown in Figure 2.20 (d).

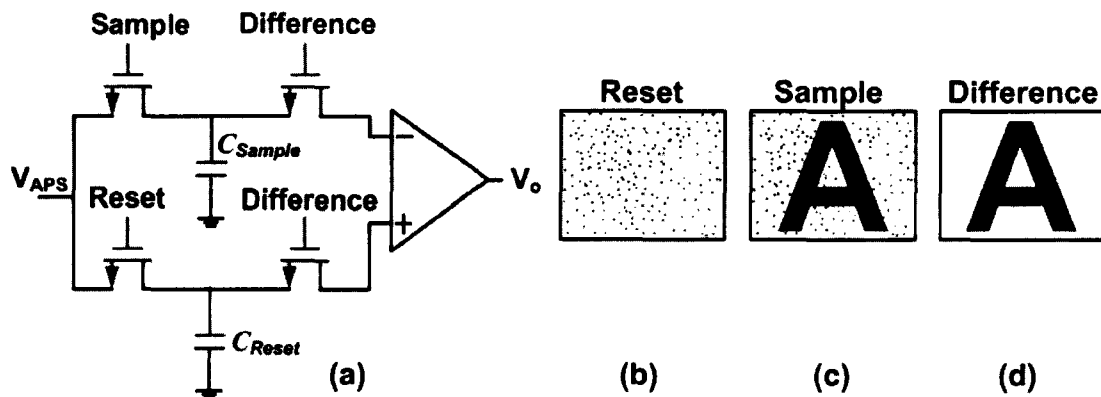


Figure 2.20: CDS circuit example. (a) Schematic diagram, (b) reset values stored, (b) image with FPN, and (c) difference after CDS.

2.2.5. Pixel sharing

One of the drawbacks of the 4T-APS is the reduced FF due to the additional transistors. Pixel sharing is a method to share a number of transistors between neighboring pixels. Figure 2.21 shows an example of how pixel sharing can be adopted to four neighboring pixels [21]. By controlling transistors S1-S4, the desired photodiode can be connected to the reset transistor and column buffer. This means that each of these photodiodes has to be read out sequentially, which decreases the frame-rate, however, the current pixel count is now 1.75 transistors per pixel rather than 4 transistors per pixel in the standard 4T-APS.

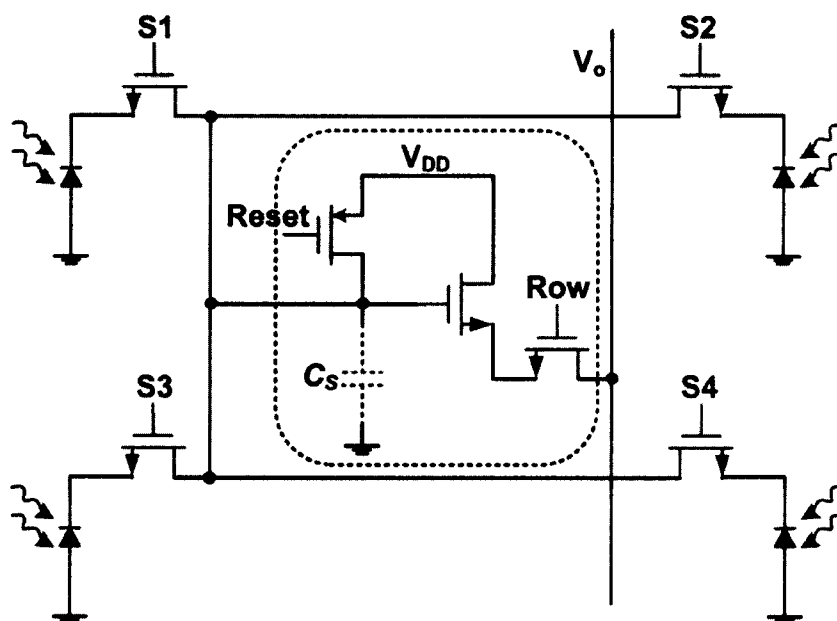


Figure 2.21: 4T-APS implemented with shared pixels, adopted from [21].

2.3. Special Purpose Pixels

With the many emerging applications of CMOS imagers, special purpose pixels were designed to meet the special demands of these applications, such as high dynamic range, high sensitivity, or high speed.

2.3.1. High dynamic range (DR)

The dynamic range of an imager is defined by the ratio of the strongest light that will cause the imager to saturate, to the weakest light that the imager can detect. As

previously mentioned, the DR can be increased by increasing the size of the potential well or the range of the supply voltage, which are both not easy to achieve. Emerging high dynamic range (HDR) imagers have adopted techniques from the human eye, since the eye can achieve a DR of 200 dB [35], whereas a CMOS sensor would typically have a DR of around 70 dB. The eye achieves this high DR by using three different mechanisms. The first one is the use of different detectors (cones and rods), one sensitive to weak light and the other sensitive to strong light. This technique is known as dual-sensitivity and it can be achieved in CMOS imaging using multiple photodetectors that are sensitive to different light powers, such as a photodiode and a MOSFET photo-gate. The second technique the human eye uses is saturation detection, where the response curve can shift according to the level of ambient light, which is why it takes our eyes some time to adjust when moving from a bright room to a dark room. Saturation detection has also been implemented in CMOS by comparing the integrated charges to the saturation limit and if the limit is reached, resetting the photodiode and counting the number of resets. Finally, the human eye also relies on nonlinear response, where saturation occurs slower at higher levels of illumination. This is implemented in CMOS imaging using the log-mode sensor (Figure 2.22). The idea of the logarithmic sensor is to make use of the MOSFET subthreshold region of operation, where the current and gate-source voltage are exponentially related. Figure 2.22 (a) shows an example of a log-mode sensor, while the output voltage waveform is shown in Figure 2.22 (b), which can be expressed by

$$V_O = V_{DD} - V_{th} - \frac{mk_B T}{e} \ln \left(\frac{I_{ph}}{I_0} \right), \quad (2.36)$$

where V_{th} is the transistor threshold voltage, I_0 is the saturation current that occurs when $V_o = V_{DD} - V_{th}$, and m is the body effect coefficient given by [35]

$$m = 1 + \frac{C_{depletion_layer}}{C_{oxide}}, \quad (2.37)$$

The main advantage of the log-mode sensor is the very high DR that can be achieved due to the nonlinear sensitivity. However, log-mode sensors suffer from reduced sensitivity, especially for low light. Also, the MOSFET has a slower response when operating in the subthreshold region, which reduces the speed of the sensor. The voltage swing of the output is also reduced, as seen from equation (2.36), where the signal drops

by at least a threshold voltage. Finally, since this pixel depends more on device characteristics, it will have an increased FPN level than the standard 3T-APS.

High DR can also be achieved in CMOS technology using pulse-frequency modulation (PFM) in digital-pixel sensors using multiple resets to control a counter. This achieves light-to-frequency encoding since the output frequency is linearly proportional to the light intensity. In [41], a DPS was used with multiple-reset PFM to achieve a DR of 115 dB. Although this technique requires a larger pixel size and causes increased FPN, the pixel can operate under a low supply voltage, it offers a programmable response, and can potentially reduce reset noise, depending on the number of resets [41].

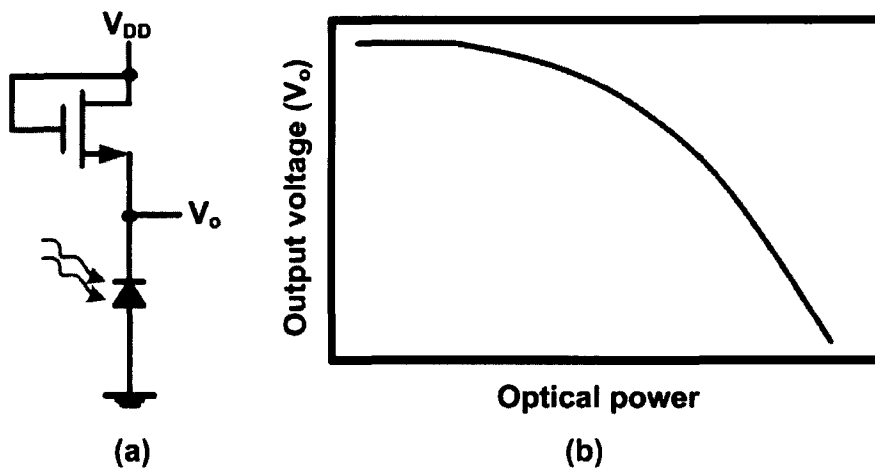


Figure 2.22: (a) Schematic of a log-mode sensor, and (b) the output voltage waveform.

Another method that can increase the DR, which can be implemented in hardware or software, is multiple sampling. In this method, samples can be captured repetitively with different integration times and then combined together to achieve HDR. Figure 2.23 shows an example of an image captured with an SLR camera at three different exposure times and then merged together to achieve a HDR image. In the HDR image, dark areas between the trees, which require long integration times, show clearly, as well as bright areas such as the clouds, which require short integration times.

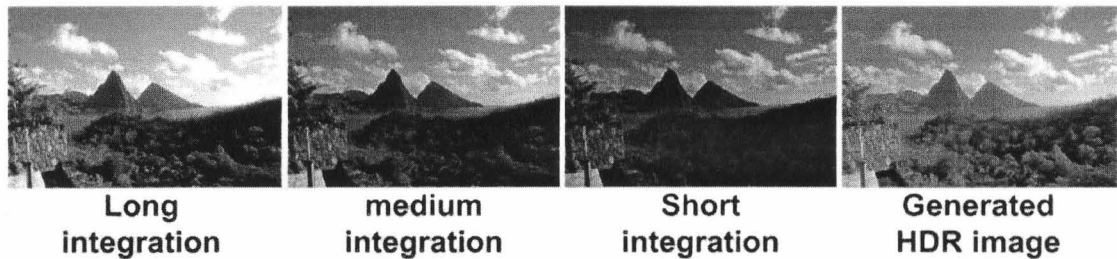


Figure 2.23: HDR image generated by merging three images obtained at long, medium and short exposure times using a standard SLR camera.

2.3.2. Low-light level

When an application requires low-light level imaging at relatively low speeds, it can be seen from the previously shown APS output curves, such as Figure 2.17, that increasing the integration time can result in an increased SNR as long as the signal does not saturate. Sensitivity or SNR can also be enhanced by reducing the noise, or dark current. This can be done by cooling, however, one of the advantages of using CMOS technology is the portability and reduced size and cost. The dark current also has a small dependence on the reverse bias voltage, which is why near zero bias voltage circuits have been developed to operate the photodiode under low biasing conditions [35]. To reduce the thermal noise caused by hard switching of the reset transistor, active reset can be used, where feedback is used to stabilize the photodiode voltage [35]. In order to reduce common-mode noise, a differential APS structure can be used, which would add to the number of transistors in the pixel, and hence, reduce the FF.

As mentioned in the previous section, using an avalanche photodiode (APD) can help increase the sensitivity due to the availability of gain. In order to avoid increasing the noise, since it is also multiplied by the gain, when the electric field is increased beyond the breakdown voltage, the APD will operate in Geiger mode. In this mode, once an avalanche process starts, it will generate a macroscopic current that can generate a digital pulse. Since this mode of operation does not maintain a linear gain relationship between the generated photocurrent and the number of photons incident, rather, it gives a digital pulse per avalanche event, the Geiger mode APD is considered a single-photon detector (SPD). If the photon events are counted over a period of time, a photon flux can be obtained, which can be converted into light intensity. The sensitivity of Geiger mode APDs is limited by the false events that occur due to thermal generation of *ehp* rather

than incident photons, which is known as dark count. The detailed implementation of an SPD is explained in Chapter 5.

2.3.3. High-speed

The frame-rate and speed of an imager is very important for applications such as FLIM, machine vision, and military imaging. One of the advantages of the 4T-APS is that it can act as a global shutter, whereas, in the 3T-APS, the integration continues while the pixels are being readout, which can cause motion blurs if the objects in the image are moving faster than the frame-rate. Figure 2.24 shows an example of the effect of using a rolling shutter on an image with a moving object. In this case, the object is a triangle that is assumed to be moving to the right at a speed of one pixel per row readout time. So as the imager reads out one row after the other, the object would have moved one pixel per row, showing how the image can get distorted. This effect is more noticeable when capturing night images with long exposure times.

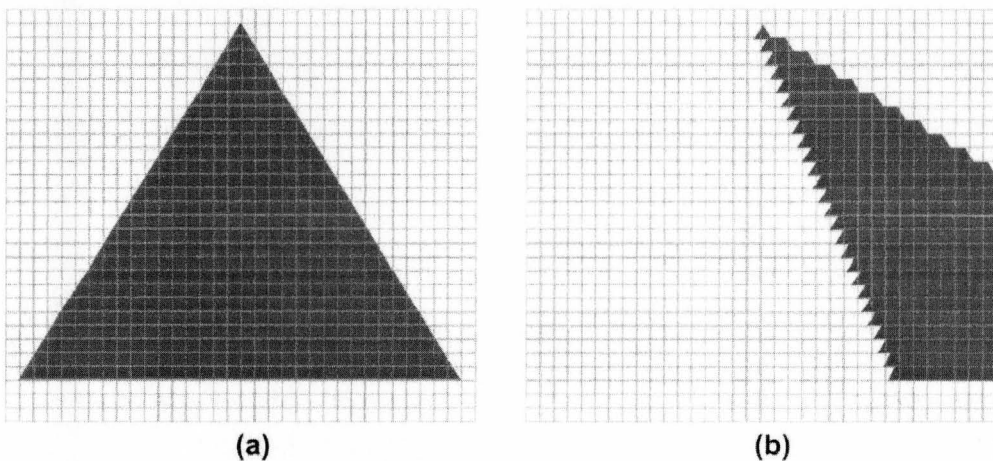


Figure 2.24: Example of the effect of using a rolling shutter on an image with a moving object. (a) The undistorted object and, (b) the captured image.

More will be explained on high-speed imaging in Chapter 4. However, with the downscaling of CMOS technology, a new type of pixel has emerged that can increase the frame-rate of CMOS imagers, which is the digital pixel sensor (DPS). Figure 2.25 shows an example of a DPS which contains a reset transistor and a photodiode, in addition to an 8-bit memory and an operational amplifier that acts as a comparator. The pixel allows simultaneous analog-to-digital conversion of the entire array by comparing the

photodiode voltage to a ramp generated reference voltage. Once the photodiode voltage exceeds the ramp voltage, the time in digital bits will be latched from a common array counter to the in-pixel memory, which will correspond to the converted digital value. Using a DPS, a frame rate of 10,000 frames-per-second was achieved [41]. However, the drawbacks were the large increase in power consumption and reduction in FF.

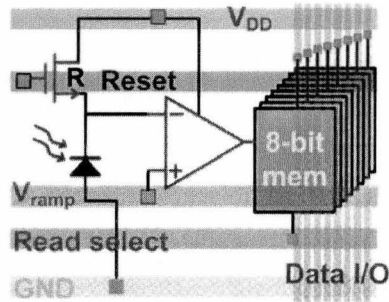


Figure 2.25: Example of a digital pixel sensor (DPS).

2.4. APS Noise

Active pixel sensors suffer from a number of noise sources that limit the performance of the pixel. These noise sources are mainly divided into pixel-to-pixel variation and temporal noise. An example of pixel-to-pixel variation is FPN, which was previously explained. The following subsections explain the sources of APS temporal noise.

2.4.1. Noise during reset

The dominant noise during the reset phase is the thermal noise from the on-resistance of the reset transistor. Figure 2.26 shows the equivalent circuit of the APS during the reset phase where the row select transistor is off, which also turns off the source follower. The thermal noise voltage (v_n), over a bandwidth Δf , is given by

$$\overline{v_n^2} = 4k_b T R_{on} \Delta f, \quad (2.38)$$

and the output voltage relates to the thermal noise voltage as

$$\frac{v_{out}}{v_n}(s) = \frac{1}{R_{on} C_{PD} s + 1}, \quad (2.39)$$

where the output noise can be obtained as

$$\overline{v_{out}^2} = \int_0^{\infty} \frac{4k_b T R_{on}}{(2\pi R_{on} C_{PD} f)^2 + 1} \cdot df = \frac{k_b T}{C_{PD}}, \quad (2.40)$$

this gives the noise power of the charge to be

$$q_{out}^2 = (C_{PD}V_{out})^2 = k_bTC_{PD}, \quad (2.41)$$

which depends on the diode capacitance and temperature, but not on the on-resistance of the transistor. This is because although the increase in on-resistance results in an increase in thermal noise voltage, it also results in a reduction in bandwidth, which cancels out. Using an NMOS reset transistor rather than a PMOS, to achieve soft reset, can reduce the thermal noise by ~half [41], however, soft reset can lead to image lag from frame to frame [43].

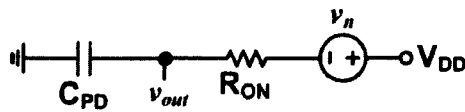


Figure 2.26: The equivalent circuit of a 3T-APS during the reset phase.

2.4.2. Noise during integration

During the integration phase, the dominant noise is shot noise due to the dark current and the photocurrent, and the equivalent circuit is shown in Figure 2.27. Assuming that the photodiode capacitance is constant over the integration period, with a PSD of shot noise given by

$$S_{I_n}(f) = q(i_{PD} + i_{dark}) A^2/\text{Hz}, \quad (2.42)$$

the following equation can be obtained

$$\frac{d(v_s + v_n)}{dt} = -\frac{i_{PD} + i_{dark} + I_n(t)}{C_{PD}}, \quad (2.43)$$

and the noise voltage can be obtained as

$$v_n(t) = -\frac{\int_0^{t_{int}} I_n(\tau) \cdot d\tau}{C_{PD}} \rightarrow v_n^2 = \frac{\int_0^{t_{int}} I_n(\tau)^2 \cdot d\tau}{C_{PD}^2}. \quad (2.44)$$

Finally, the mean square value of the noise voltage at the end of the integration time is given by

$$\overline{v_n^2} = \frac{q(i_{PD} + i_{dark})}{C_{PD}^2} t_{int}. \quad (2.45)$$

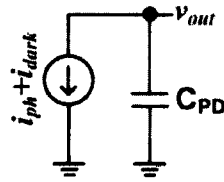


Figure 2.27: The equivalent circuit of a 3T-APS during the integration phase.

2.4.3. Read out stage noise

During the readout phase (Figure 2.28 (a)), the noise contributions from within the pixel are due to the source follower (M_2) and the row select transistor (M_3). Additional readout noise is added on an array level from the column biasing transistor (M_4) and the column capacitance (C_o). The small-signal equivalent circuit is shown in Figure 2.28 (b), where g_m is the transistor transconductance, g_d is the channel conductance, $r_{2,4}$ are the output resistances of the transistors, and v_s is the photodiode voltage. The gain can be given by

$$\frac{v_o}{v_s} = \frac{1}{\frac{r_2 + g_{d3} + r_4}{g_m r_2 r_4} + \frac{r_4 + g_{d3}}{r_4}} \tag{2.46}$$

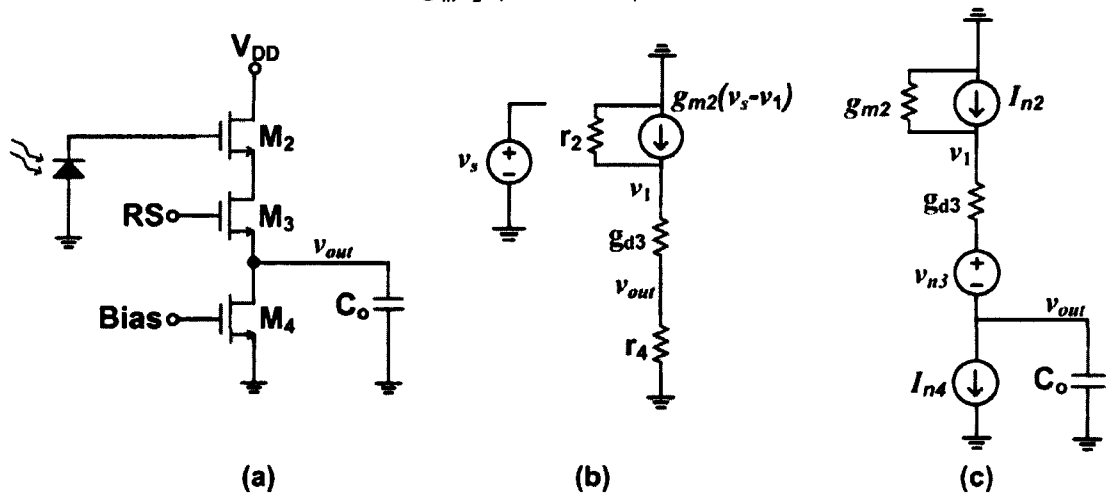


Figure 2.28: APS during readout, (a) schematic, (b) small-signal equivalent circuit, and (c) noise equivalent circuit.

The equivalent noise circuits is shown in Figure 2.28 (c), where the MOSFETs are assumed to have a thermal noise PSD of [43]

$$S_{i_n}(f) = 4k_b T \gamma g_m \text{ A}^2/\text{Hz}, \tag{2.47}$$

where γ is assumed to be 2/3. The output total noise power can be obtained as

$$\overline{V_{n_out}^2} = \underbrace{\frac{2kT}{3C_o} \frac{1}{1 + \frac{g_{m2}}{g_{d3}}}}_{M_2} + \underbrace{\frac{kT}{C_o} \frac{1}{g_{d3} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)}}_{M_3} + \underbrace{\frac{2kT}{3C_o} g_{m4} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)}_{M_4}. \quad (2.48)$$

2.4.4. Random telegraph signal (RTS) noise

With the continuous downscaling of pixel pitch, random telegraph signal (RTS) noise has become a major concern in CMOS APS imagers. The RTS noise may have a significant impact on the quality of the image as it becomes the leading source of noise in pixels that have transistors with an active area in the sub-micron range [44]. In a 3T-APS, the source follower transistor is the major contributor to RTS noise. Since, RTS noise is also observed in photodiodes [45], they may also be one of the sources of RTS noise in an imager.

RTS is caused by the defects in the semiconductor bulk [46], which act as traps for electrons. The trapping and de-trapping of an electron modulates the channel current of the transistor and thus causes voltage fluctuation across the gate and source terminals. The mean trapping and de-trapping time, which are commonly referred to as capture and emission times, respectively; are temperature dependent and remain constant at a given temperature [47]. The capture time and emission time as well as the magnitude of the voltage fluctuation determine the characteristics of the RTS noise and thus affect the performance of the imager. Figure 2.29 shows the typical behavior of RTS noise [44]. The drain current of the MOS transistor switches to a lower value when an electron is captured by a trap, after which, it goes back to the higher value, when the electron is released. Using Shockley-Read-Hall (SRH) theory [48], the mean capture (τ_c) and emission (τ_e) times can be given as [51]:

$$\tau_c = \frac{1}{v_{th} \sigma_n n}, \text{ and} \quad (2.49)$$

$$\tau_e = \frac{\exp\left[\frac{(E_f - E_T)}{kT}\right]}{g v_{th} \sigma_n n_t}, \quad (2.50)$$

where v_{th} is the speed of the electron, σ_n is the electron capture cross section at the oxide semiconductor interface, n and n_t are the concentration of electrons at the interface and in

the conduction band, respectively, E_T is the trap energy, and g is the degeneracy factor. The capture time can be decreased by increasing the gate voltage [47], [51].

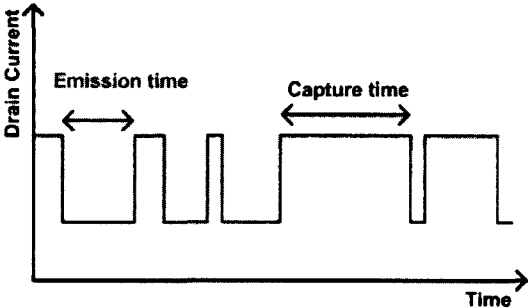


Figure 2.29: Random telegraph behavior of drain current in MOS transistor [44].

Chapter 3

CMOS APS COMPLETE CAMERA-ON-A-CHIP

The research conducted in this work required having a setup that is suitable for testing and processing data from CMOS imagers. A simplified imager measurement setup block diagram is shown in Figure 3.1. The CMOS imager is controlled by an Altera field programmable gate array (FPGA) board through the expansion slot. Image settings, such as integration time, averaging or correlated-double sampling, could be adjusted by the switches on the board. The FPGA also interfaces the imager to a VGA monitor to display the captured images in video mode, as well as interfaces to the PC to upload the captured images serially through the RS-232 link, for further processing. In order to test the system, a fully integrated, 16×16 pixel 3T-APS CMOS camera-on-a-chip, was designed and fabricated in a standard CMOS $0.18 \mu\text{m}$ technology. The camera is explained in this chapter. An overview of the camera-on-a-chip implementation is given in Section 3.1. The photodiode design is discussed in Section 2.2, while the pixel design is explained in Section 2.3. The analog-to-digital converter (ADC) design is discussed in Section 2.4, which is followed by the array design in Section 2.5. Finally, the imager measurement results are presented in Section 2.6.

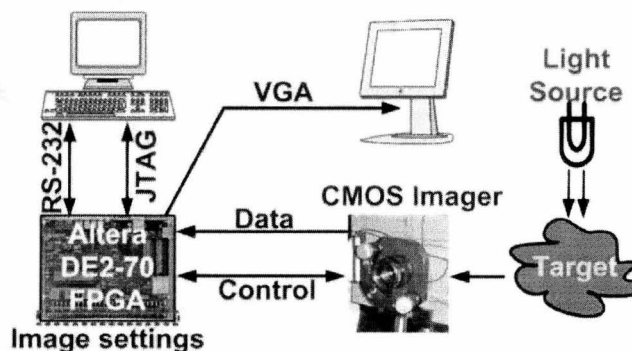


Figure 3.1: Block diagram of the CMOS imager setup.

3.1. Camera-on-a-Chip Design

The block diagram of the camera-on-a-chip is shown in Figure 3.2. The imager consists of an array of 256 3T-APS pixels, controlled by row and column addressing circuits. The design provides an analog output that comes directly from the buffered APS, in addition to a digital output using an on-chip ADC. The output of the ADC is provided as both serial (to reduce the number of I/O pads if needed) and parallel for higher readout speed. The first phase of this implementation was to test a number of different photodiode designs, after which, the APS can be designed. The following section explains the photodiode design and measurements.

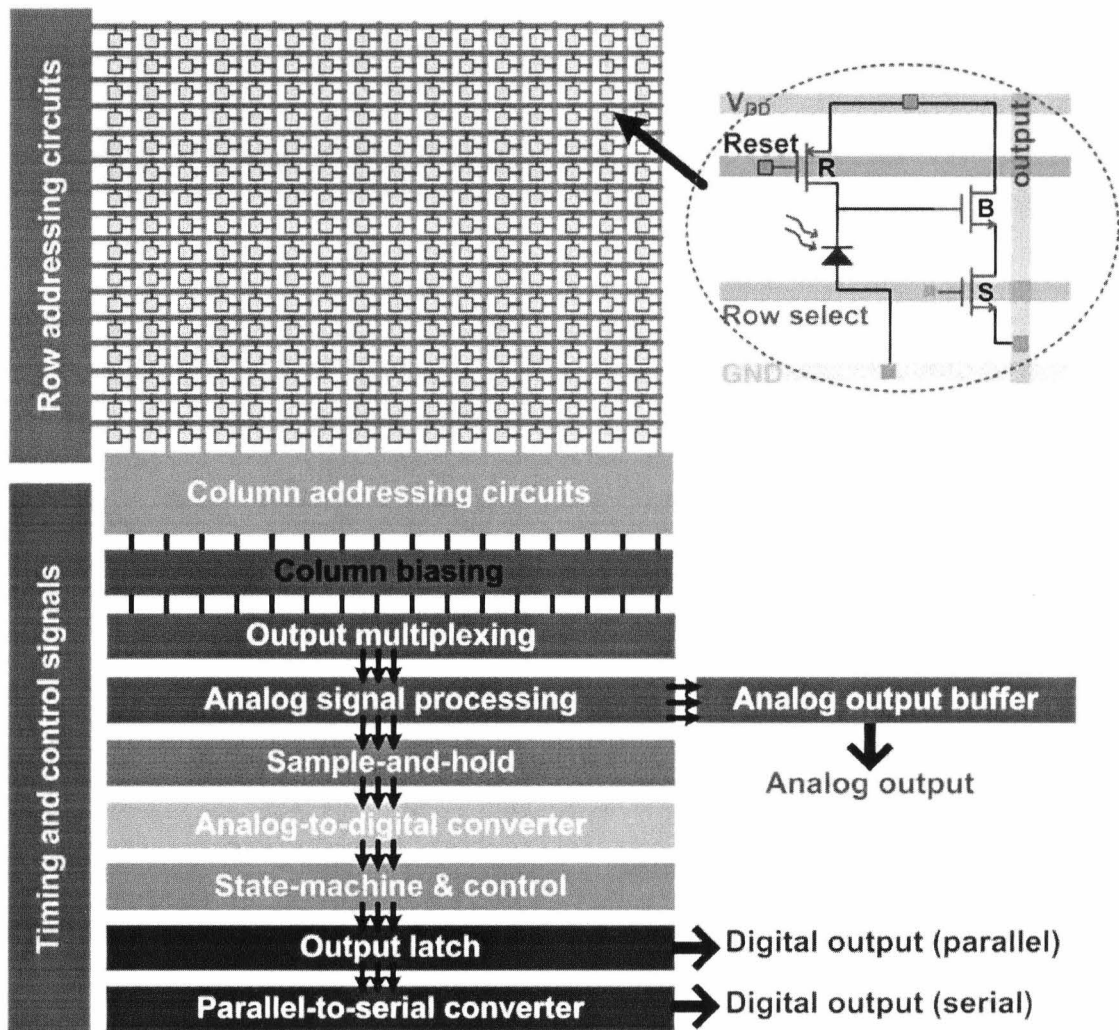


Figure 3.2: CMOS camera-on-a-chip block diagram.

3.2. Photodiode Design

In a triple-well CMOS process, a number of different photodiode structures can be implemented. Figure 3.3 shows the cross-section of one of our CMOS layouts. In the figure, the six diodes that can be implemented in a triple-well process are shown. The purpose of the triple-well in the process is to allow isolation of different parts of the substrate, allowing for each MOSFET to have different body potentials, which is necessary, for example, in a cascode amplifier. Diodes can be formed between the n+ layer and the p-substrate (D1), the p+ layer and the n-well (D2), the n+ layer and the p-well (D3), the p-well layer and the deep n-well layer (D4), the deep n-well layer and the p-substrate (D5), and finally, between the n-well layer and the p-substrate (D6). The major differences between these diodes are in the doping levels, junction depths, and junction capacitances. Also, depending on the layers used, the minimum layout area needed for the diode could differ. For the prototype presented in this chapter, high speed was not a priority; rather, the focus was on design simplicity and high FF. The simple n+/p-sub (D1) diode was used in this design. A microchip was fabricated in a CMOS 180 nm technology from TSMC, shown in Figure 3.4 (a) and (b), in order to test the performance of various diodes. The measurement setup is shown in Figure 3.4 (c), which was placed in a dark room to minimize background illumination. The setup was placed on a floating vibration isolation measurement table, which was necessary for noise measurements.

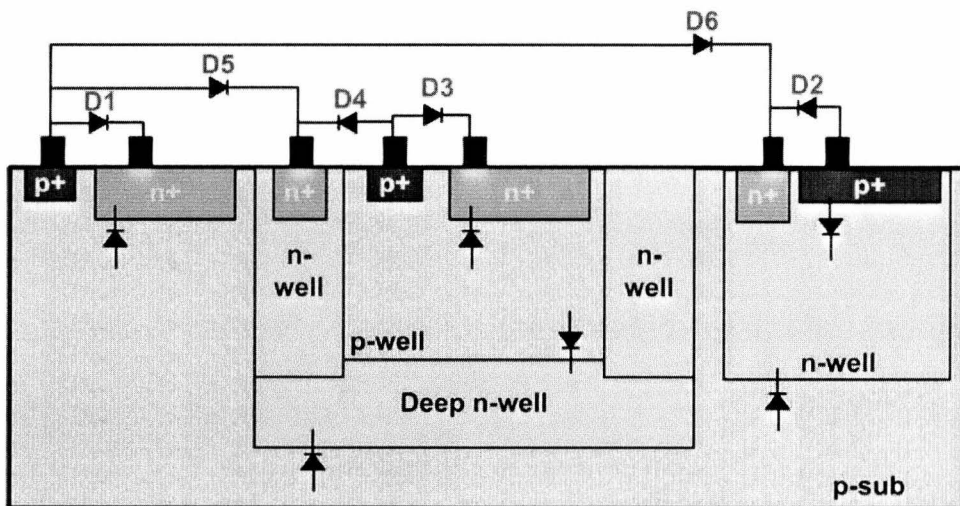


Figure 3.3: Examples of different diodes that can be implemented in a triple-well CMOS process.

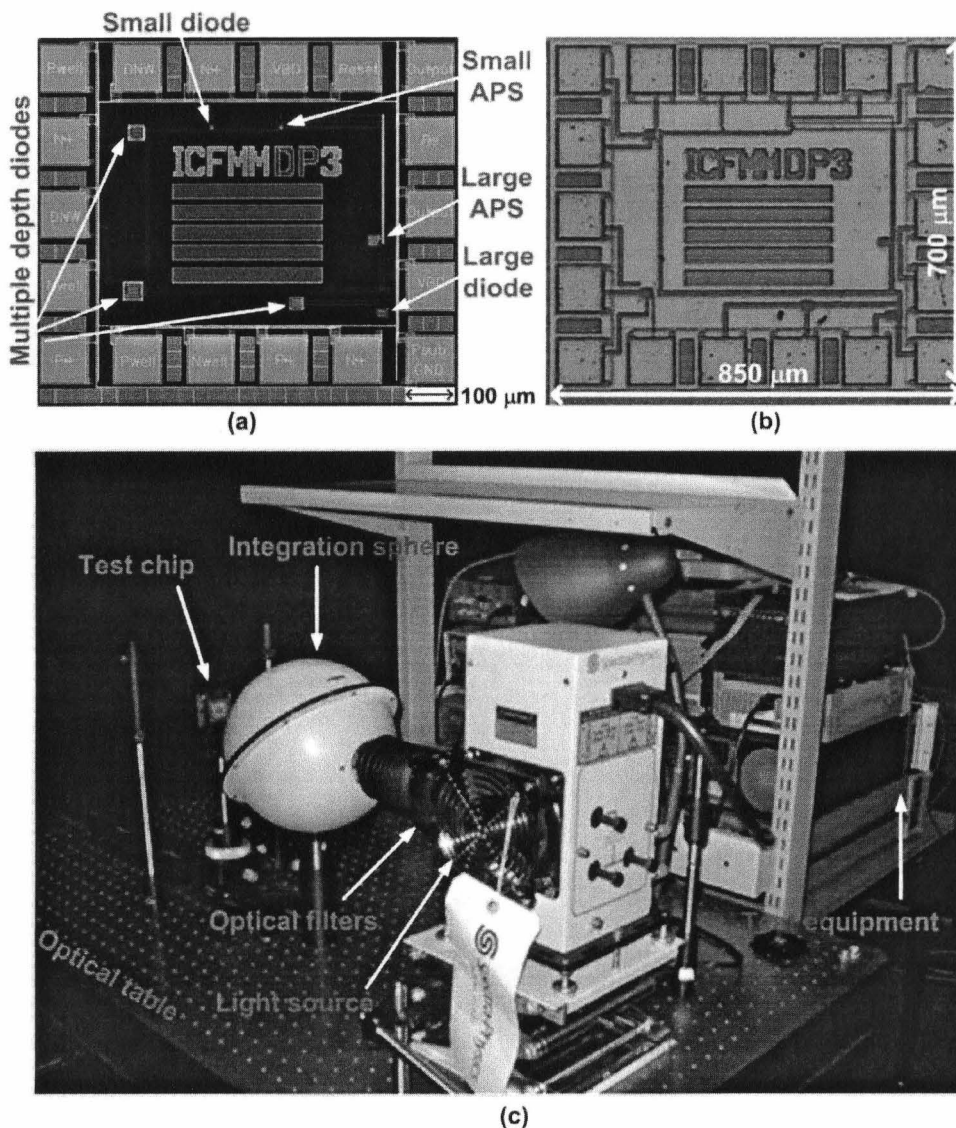


Figure 3.4: Photodiode test structure chip layout (a), and photomicrograph (b). (c) The dark room optical setup used to characterize the devices.

The setup shown Figure 3.4 (c) was used for characterization of all photodiodes that are described in this thesis. A 100 W Xenon lamp was used for its uniform irradiance and stable optical power over a wide spectrum range from ultraviolet to near infrared. The lamp was connected to an integration sphere through an optical filter box, where different filters can be placed to set the wavelength of the light incident on each photodiode. The integration sphere provides nearly uniform light flux at its output, to ensure that the optical power on various parts of the microchip is equal. The test chip was placed on a rail to vary the incident optical power on the chip by changing its distance from the

integrating sphere. Finally, the optical power was measured with a calibrated reference photodiode and optical meter from Newport (2835-C, 818-UV).

The spectral response of the n+/p-sub diode is shown in Figure 3.5. The measurement was obtained using different wavelength filters in increments of 10 nm. At each wavelength, the generated photocurrent was measured using a semiconductor parameter analyzer, and the optical power was measured using the calibrated photodiode and scaled to the area of n+/p-sub diode. A peak responsivity can be seen at 680 nm.

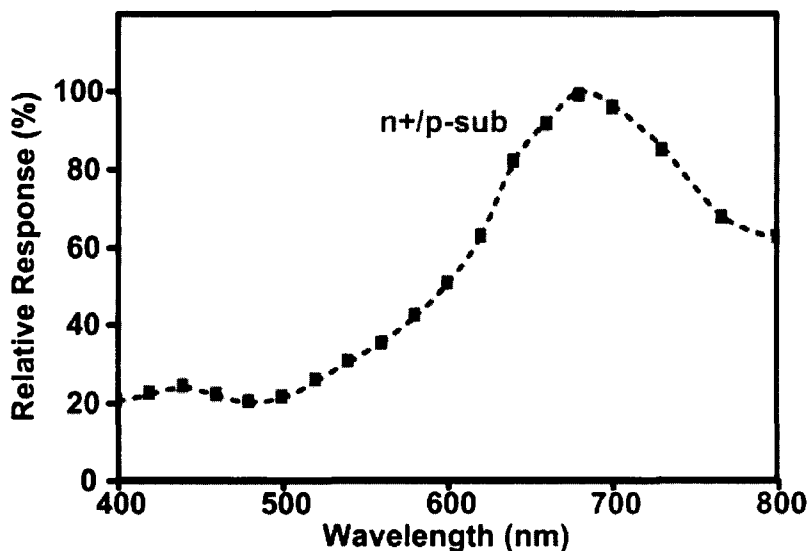


Figure 3.5: Measured relative responsivity of the n+/p-sub diode as a function of wavelength.

Different diodes can have a peak spectral response at a different wavelength, depending on the depth of the depletion region. This can be used for color detection to avoid the use of expensive color micro-filters. Figure 3.6 (a) shows the measured relative responsivities of a shallow p+/n-well diode and a deep n-well/p-sub diode. If the current from both diodes is measured simultaneously, a ratio can be obtained that indicates the color of the incident light, as shown in Figure 3.6 (b). Color detection in a typical imager requires a demosaicing algorithm to combine information from the RGB color filters of neighboring pixels, which reduces the effective resolution of an array. By using multiple depth diodes, each pixel can detect its color independent of the neighboring pixels, unlike the typical color Bayer filter case, explained previously in Chapter 1. This can allow for providing an actual 1:1 ratio between the number of pixels in the image and the number

of pixel locations in the array. This technique is used by a number of companies, such as Sigma Corporation that uses Foveon X3 image sensors.

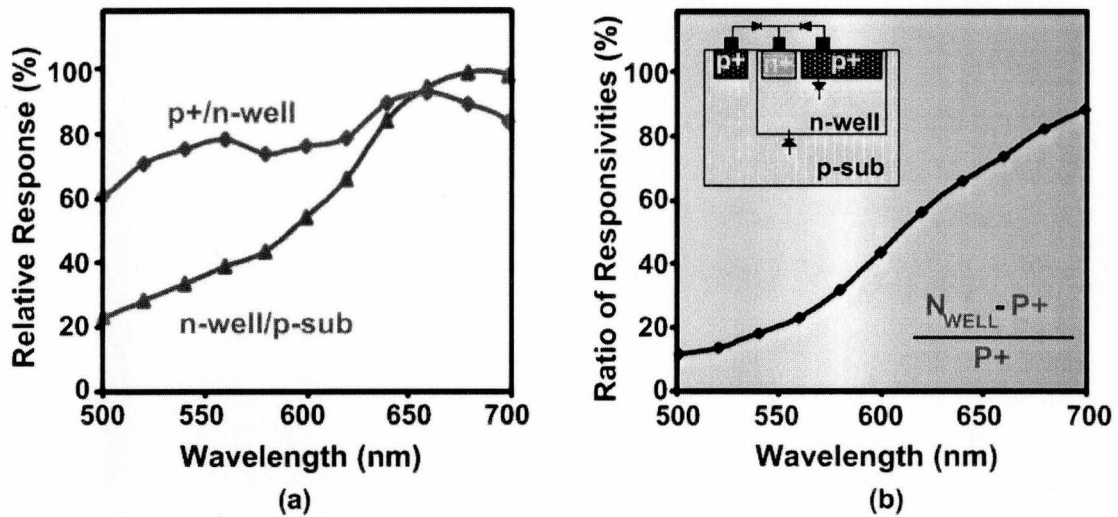


Figure 3.6: Color detection using multiple diodes at different depths. (a) The responsivity of two different diodes, and (b) ratio of the responsivity.

3.3. Pixel Design

The pixel used in this imager is the standard 3T-APS, which was explained in Chapter 2, with the exception of using a PMOS reset transistor. The PMOS transistor provides faster reset, higher dynamic range since it can reset to a full V_{DD} , and lower noise. However, it requires extra layout area due to the additional n-well. A large area photodiode is used because optical signals in biomedical applications are usually weak. In this case, the extra layout area of the n-well used for the PMOS device will not cause a significant FF reduction. In Figure 3.4 (a), two APS test structures were fabricated with different photodiode areas, where the area of the large one was $30 \mu\text{m} \times 20 \mu\text{m}$, with a 60% FF, and the area of the second was $5 \mu\text{m} \times 5 \mu\text{m}$, with a 45% FF. The layout of the larger pixel, which was used in the imager array, is shown in Figure 3.7 (a). The pixel is covered with metal everywhere except for the light sensitive area. All transistors were implemented with a $1 \mu\text{m}$ and a 180nm channel width and length, respectively. Figure 3.7 (b) shows the measured output voltage of the APS for different optical powers. The active low PMOS transistor resets the photodiode voltage to a full supply of 1.8V . The output of the pixel has a maximum voltage of 1V , which is due to the drop in the source

follower and row select transistor. Also, the pixel is a test structure that was biased off-chip by applying 0.4V to the output through a resistor, which prevents the output voltage from dropping below 0.4 V. The figure shows how the slope changes with the optical power and how a larger integration time is needed in order to achieve a good SNR for low-level light.

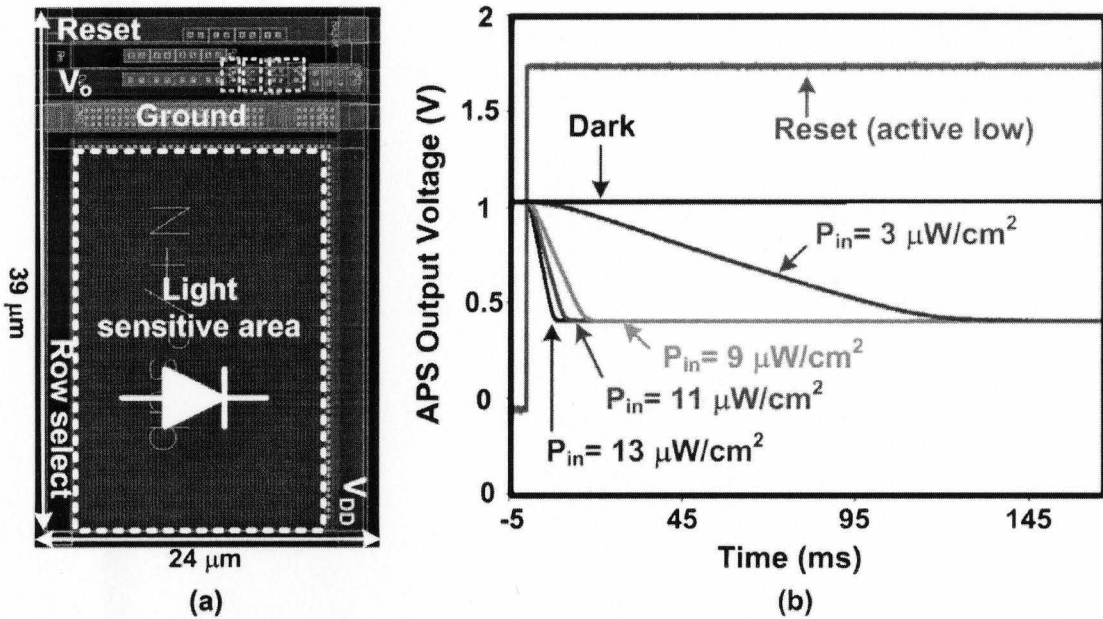


Figure 3.7: 3T-APS layout with a 60% FF (a), and the measured output voltage for different optical powers at a wavelength of 680 nm as a function of time (b).

The SNR can be obtained from the APS output by calculating the mean (signal) and the standard deviation (noise) of around 40 measurements. Figure 3.8 shows the measured SNR as a function of incident optical power (corresponding to Figure 3.7 (b)) at three different integration times. The standard error bars, shown in the figure, were calculated as the standard deviation over the square root of the number of samples. The error appears to be high due to using only 40 samples. Later, this was improved using 83 samples and the results are presented in Chapter 4. The SNR seems to peak at higher power levels after a certain integration time since the signal saturates. This figure indicates that an increased integration time can result in an improved SNR. This is true for most cases. However, in some biomedical applications, such as fluorescence lifetime imaging (FLIM), the power of the incident light decreases with time as the fluorescence decays during its lifetime. In this case, integrating beyond the lifetime will result in

increased noise only, degrading the signal. For this reason, in biomedical applications, it is common to have repeated experiments that use averaging and accumulation.

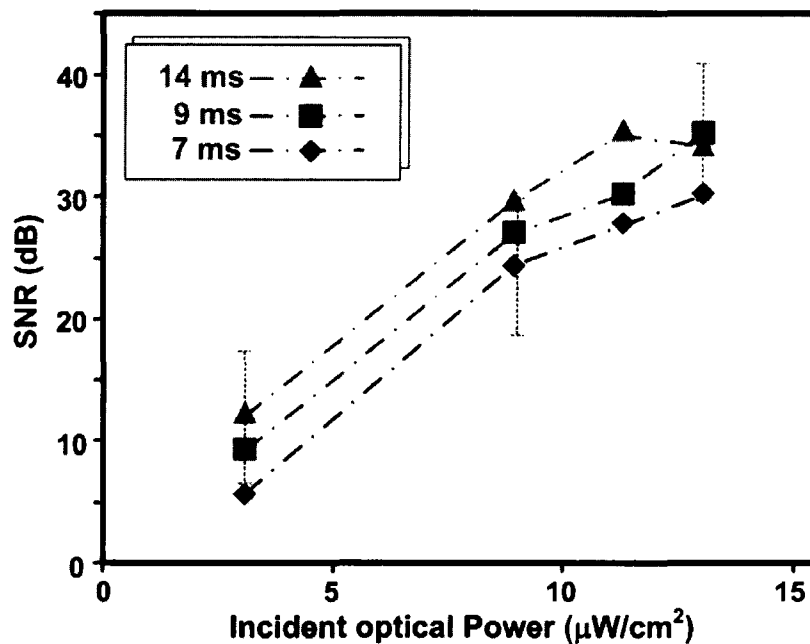


Figure 3.8: APS measured signal-to-noise ratio as a function of optical power at a wavelength of 680 nm for different integration times.

3.4. Analog-to-Digital Converter (ADC) Design

The process of converting an analog signal to a digital signal requires three steps, which are sampling, quantization and encoding. The analog signal is first sampled, then a decision is made to identify the corresponding digital level, and finally, the sample is given a digital code. The digital circuits require some time to perform the quantization and encoding steps. For this reason, the sampled value must be held for a specific time that is long enough for the following digital circuitry to complete any required processing; at least until the digital code equivalent of the sample is generated. This is done using a sample-and-hold (s/\bar{H}) circuit, also referred to as a track-and-hold circuit since the sampling is not instantaneous. A block diagram of the dual-slope integrating ADC that was used in this imager is shown in Figure 3.9. The s/\bar{H} circuit was implemented using an open-loop parallel mode design, with an output opamp unity gain buffer in order to prevent loading the hold capacitor C_H , which can result in varying the amplitude of the held value of V_{in} . An input opamp unity gain buffer was also used to

prevent loading the source and to speed up the charging or discharging of the hold capacitor, rather than depending on the source. The switch used was a PMOS-NMOS transmission gate.

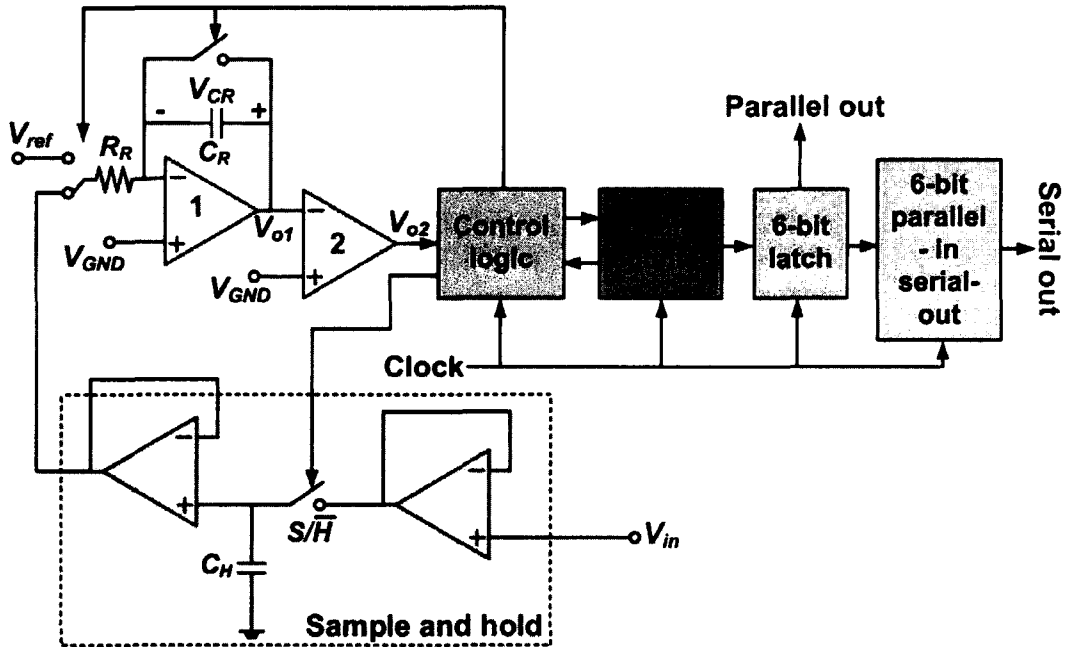


Figure 3.9: ADC block diagram.

Figure 3.10 shows the Cadence simulation results of the S/\bar{H} circuit. One of the major drawbacks of the open-loop parallel mode S/\bar{H} circuit is charge and clock feed-through. During the transition from sample to hold, the switch is turning off and the input is being isolated from the storage capacitor. At this point, charges that feed-through the overlapping gate capacitance of the MOSFET switches can change the value held in the storage capacitor. This is known as clock feed-through and usually has a small effect on the held value. Since clock feed-through only depends on the amplitude of the clock, it is considered signal independent, which means that it has a constant value that can be predicted and cancelled using a differential topology, for example. The variation caused in the held value as a result of clock feed-through can be calculated using

$$\Delta V_{CH} = \frac{C_{overlap}}{C_{overlap} + C_H} \times V_{DD}, \quad (3.1)$$

where V_{CH} is the voltage held in the capacitor, V_{DD} is the supply voltage, C_H is the value of the hold capacitor and $C_{overlap}$ is the value of the overlap gate to source/drain capacitance. Also during the sample-to-hold transition, the charges that created the

channel of the MOSFET switch are discharged into the substrate; however, some of them are also discharged into the source and drain. The charge going into the input source of V_{in} is not a problem since the input source is assumed to have a low source resistance. However, the charges that go into the storage capacitor may significantly vary the amplitude of the held value causing a pedestal error. This is known as charge injection and it is one of the most challenging design issues faced when dealing with S/\bar{H} amplifiers. The effect that charge injection has on the amplitude of the held value can be calculated as follows [51], [52], [53], [54]

$$\Delta V_{C_H} = \frac{Q_{injected}}{C_H} = k_Q \times \frac{C'_{ox}WL(V_{GS} - V_T)}{C_H}, \quad (3.2)$$

where $Q_{injected}$ is the charge injected into the storage capacitor, k_Q is the fraction of the injected charge, C'_{ox} is the oxide capacitance, W and L are the width and length of the MOSFET switch (S/\bar{H} in Figure 3.9), respectively, V_T is the threshold voltage and V_{GS} is the gate-source voltage, which in this case is equal to $V_{DD} - V_{in}$. If the switch is assumed to be an NMOS device, the charge is assumed to be negative. Also, if the clock is assumed to turn off very fast, the factor k_Q can be assumed to be equal to one half. Unlike clock feed-through, the voltage V_{GS} depends on the input, making the pedestal error introduced by charge injection signal dependent. The threshold voltage also has a dependence on the input (source-body voltage), which is non-linear, resulting in signal distortion that is even more difficult to eliminate. Since both clock feed-through and charge injection are inversely proportional to the size of the hold capacitor, a large 50 pF hold capacitor was used.

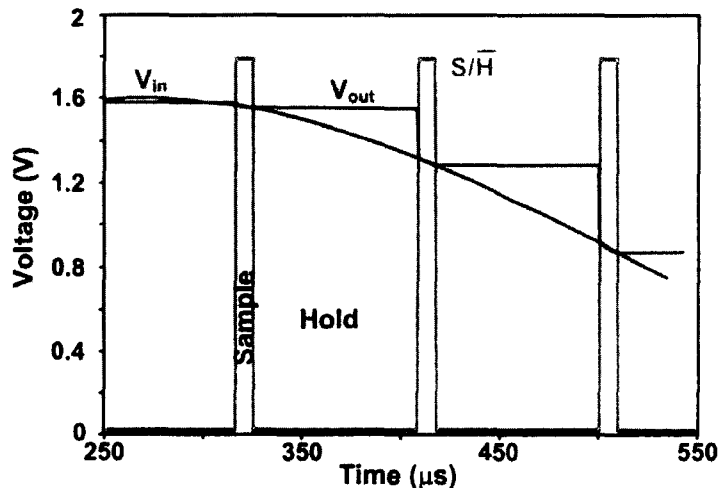


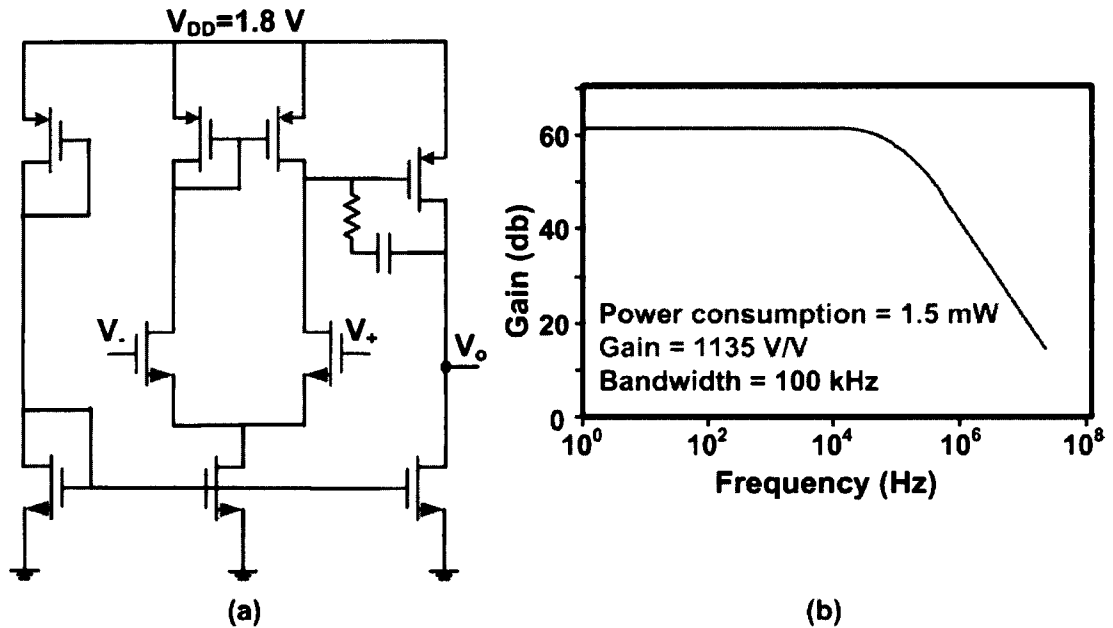
Figure 3.10: Simulation results of S/\bar{H} circuit.


Figure 3.11: (a) Schematic of implemented operational amplifier, and (b) the simulated gain.

The schematic of the opamp used is shown in Figure 3.11 (a), which uses an NMOS differential stage with a PMOS load. The simulated gain (Figure 3.11 (b)) was over 60 dB with a bandwidth of 100 kHz, consuming a power of 1.5 mW.

The dual-slope integrator ADC, previously shown in Figure 3.9, offers high accuracy at slow speed. The voltage (V_{CR}) across capacitor (C_R) can be expressed as

$$V_{CR}(t) = \frac{V_{GND} - V_{ref.in}}{R_R C_R} t, \quad (3.3)$$

which is a constant slope function of time. The accuracy of the dual-slope is due to both slopes being generated using the same integrator and counter, cancelling out any component non-idealities due to temperature or process variation. Assuming initially that V_{GND} is zero, the ADC operates as follows. First the input voltage coming from the S/\bar{H} circuit is connected to resistor (R_R), charging the capacitor C_R with a slope equal to $-V_{in}/R_R C_R$. In order to have a positive slope, causing a charge increase in the capacitor, V_{in} should be a negative voltage, which is difficult to obtain on chip. For this reason, the design uses a reference voltage of 1 V rather than zero. This sets an upper limit of 1 V on the input signal to remain negative compared to the reference. This was a valid assumption since, as previously mentioned, the APS output voltage does not exceed 1 V

due to the voltage drop in the source follower. While the capacitor voltage is increasing, the counter will count starting from zero until it overflows, after which, the capacitor charging is stopped and the switch connects the reference voltage to resistor (R_R) instead of the input voltage. Once the reference voltage is connected, assuming it has a potential that is larger than V_{GND} , the capacitor voltage begins to discharge with a slope $(V_{GND} - V_{ref})/R_R C_R$, which is negative if V_{ref} is be larger than V_{GND} , causing the capacitor to discharge. During the discharge, the counter will count until the capacitor voltage drops below V_{GND} , after which, the second opamp will generate a negative pulse, indicating that the conversion is complete and stopping the counter. The final count reached by the counter will be the digital code that is equivalent to the analog input and will be latched into the output latch and loaded into the parallel-to-serial shift register for readout. The control logic is a state-machine that is in charge of synchronizing all the necessary control signals. Figure 3.12 shows an example of the waveform of a dual-slope ADC for three different input voltages.

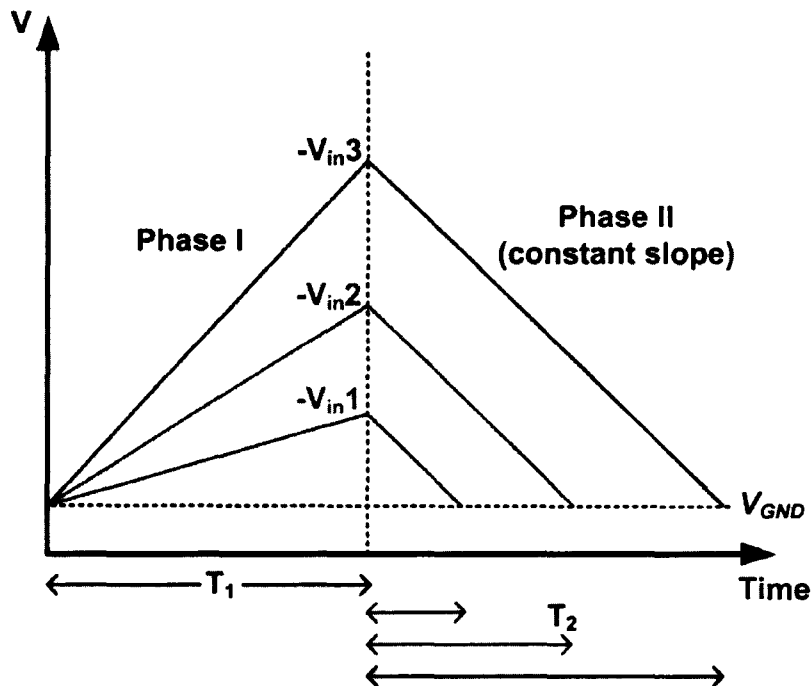


Figure 3.12: Dual-slope waveform for three different input voltages.

The voltage across capacitor C_R at the end of phase II is equal to V_{GND} , and can be expressed as

$$V_{CR}(T_2) = \frac{V_{GND} - V_{in}}{R_R C_R} T_1 - \frac{V_{GND} - V_{ref}}{R_R C_R} T_2 = V_{GND}, \quad (3.4)$$

The time T_2 is the time corresponding to the code word, and can be found from equation (3.4) as

$$T_2 = \frac{2^N / f_{clk} (V_{in} - V_{GND}) + V_{GND} R_R C_R}{(V_{ref} - V_{GND})}, \quad (3.5)$$

where $T_1 = 2^N / f_{clk} = 2^N t_{clk}$, N is the number of bits in the counter, and $0 < V_{in} < V_{GND} < V_{ref}$. The maximum T_2 corresponding to the highest input voltage (1 V) should be below 2^{N+1} , otherwise the counter will overflow before the conversion is complete. For a 6-bit counter, with a 1 MHz clock, T_2 should be less than 128 μ s. The design used a 200 k Ω resistor (R_R) and a 150 pF capacitor (C_R), so $R_R C_R = 30 \mu$ s.

3.5. Array Design

The block diagram of the imager was previously shown in Figure 3.2. The layout screen capture of the imager that measures an area of 1.5 mm \times 3 mm, is shown in Figure 3.13. The control is done entirely on-chip, by providing a single external master clock from the FPGA. Column and row scanners are used instead of decoders, which have the advantage of requiring significantly less inputs; however, they do not allow immediate addressing of a specific pixel. Rather, the entire array must be scanned sequentially. A scanner is a shift register that shifts one bit from a D-flip-flop to the next, scanning the array. The external clock controls the column scanner, and after reading out the last column in a row, the row scanner is clocked by the first bit of the column scanner. The column scanner selects which of the columns to connect to the common output bus, through a multiplexer that uses NMOS switches. Each column is biased by a long channel NMOS current source transistor that provides biasing to the source follower inside the APS. The FPGA is in charge of initiating the ADC conversion process, where the ADC issues a “conversion complete” signal to the FPGA once it is done so that the FPGA can read the data, either by serial or parallel readout.

In the pixel array, everything except for the photosensitive area was covered with metal layers. The large capacitors used for the s/\bar{H} and the ADC circuits can be seen on the chip. Also, a number of other capacitors have been added to filter all DC supply and

bias lines. Based on the number of pixels in the array, and the ADC speed, this imager is designed to operate at around 60 frames per second.

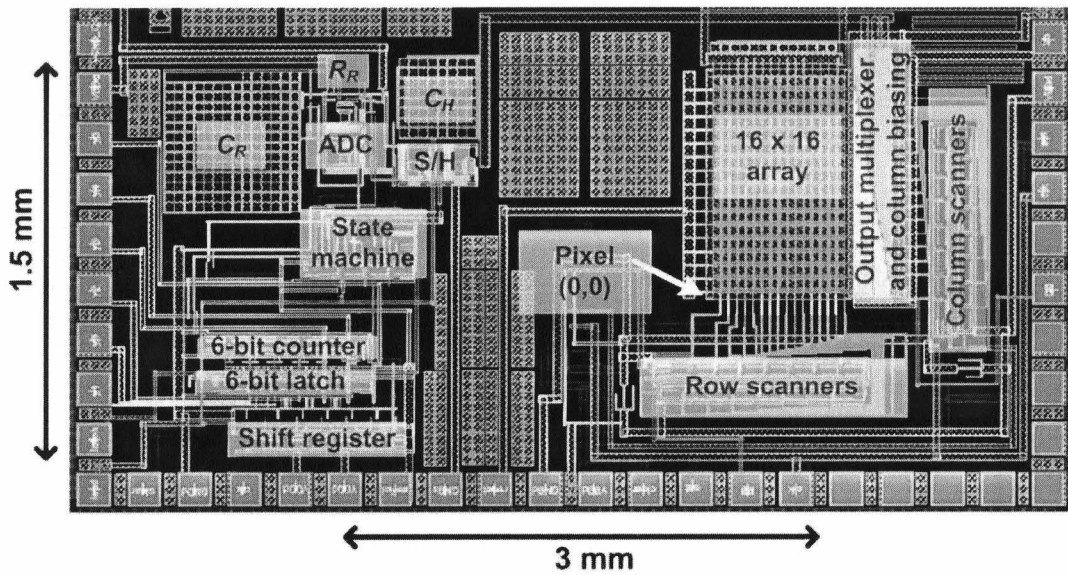


Figure 3.13: Imager layout screen capture in 180 nm CMOS technology using Cadence Virtuoso software.

3.6. Imager Results

The photomicrograph of the fabricated CMOS camera-on-a-chip is shown in Figure 3.14. The chip was enclosed in a 68-PGA package and connected onto a PCB. The experimental setup previously shown in Figure 3.1 was used for testing. The Altera DE2-70 FPGA was programmed using Verilog hardware description language on Quartus II. The FPGA controlled the user defined reset and integration times, synchronizing the row and column scanners and moving from pixel to pixel, starting and clocking the ADC, and retrieving the converted digital values. Also, a VGA controller was implemented in the FPGA to display the image at ~ 60 frames per second and an enlarged version of the image (16 times) was also displayed on the monitor. Finally, the captured images were transmitted, when needed, for further processing and storage to a PC through the serial RS-232 port. An RS-232 controller was also implemented to transfer the images.

A photograph of the measurement setup on the optical table is shown in Figure 3.15. A single lens was used to focus the light onto the image sensor array. Based on the size of the optical table, the target object was placed 44 cm away from the lens holder, and a lens with an effective focal length (EFL) of 15 mm was used. The imager was placed roughly

13 mm away from the lens, which gives an object to image magnification of $1/33$. Based on the array size (Figure 3.14), the object size should be within $21 \times 14 \text{ mm}^2$.

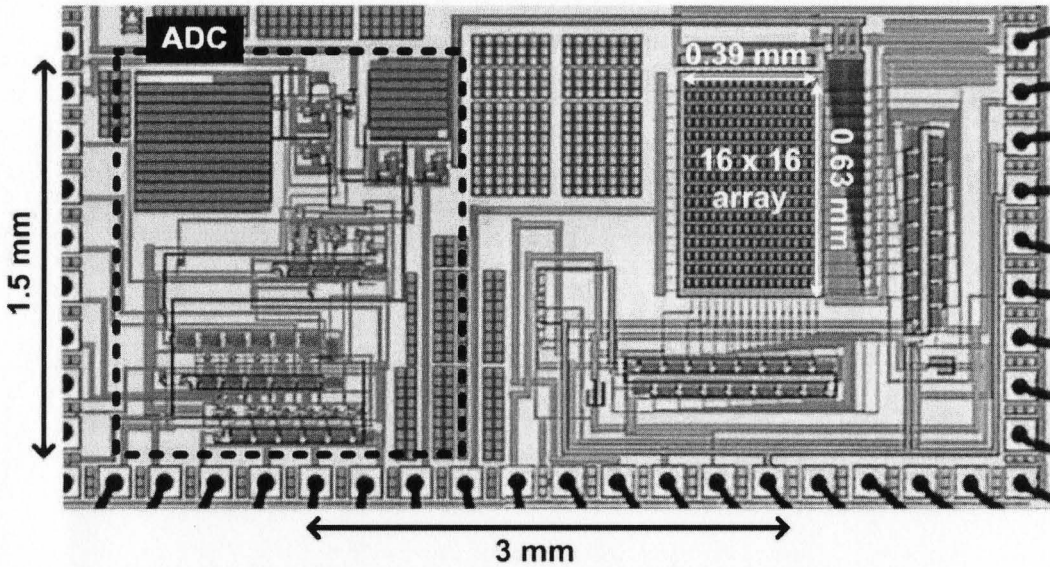


Figure 3.14: Fabricated CMOS camera-on-a-chip photomicrograph.

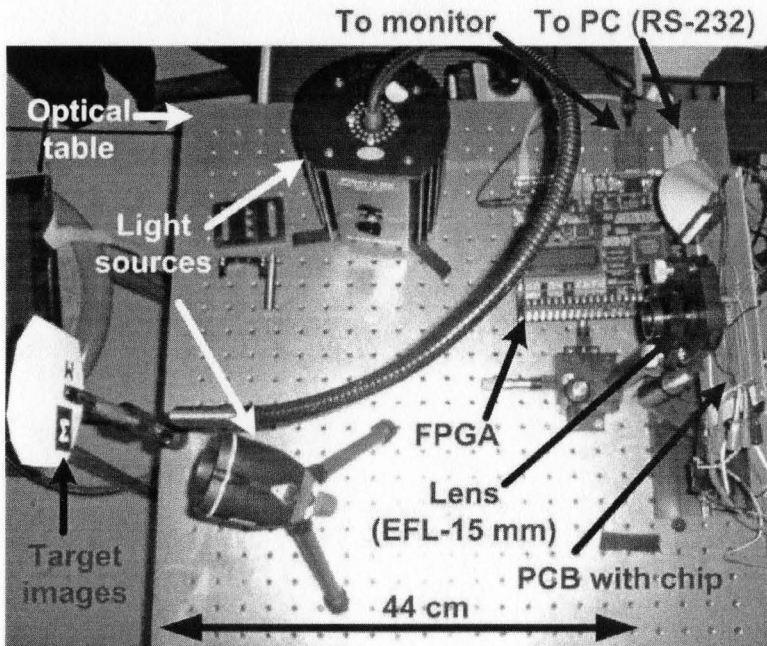


Figure 3.15: Photograph of the measurement setup.

Figure 3.16 (a) shows some examples of the captured images that were uploaded to the PC using the RS-232 link in the FPGA. The images shown are not processed or averaged, nor was correlated-double sampling used. In the enlarged images (Figure 3.16

(b)), it is easy to see the noise that appears due to FPN and non-uniform illumination. An example of how CDS can reduce the noise (done in software) is shown in Figure 3.16 (c).

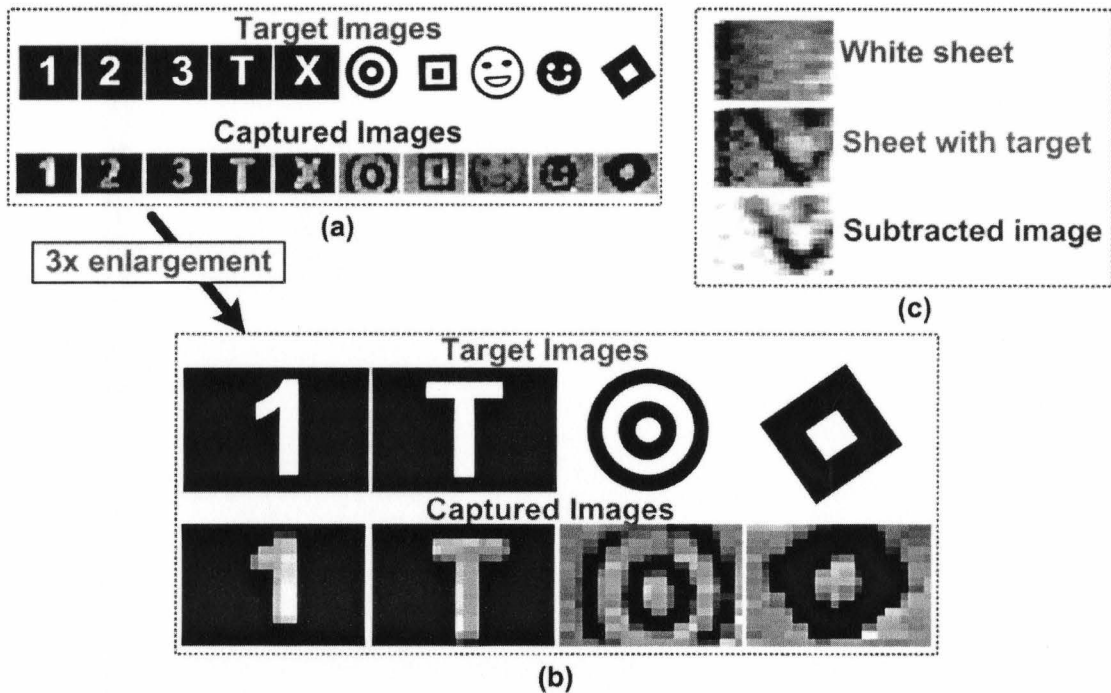


Figure 3.16: Target images (resized for comparison) compared to the non-processed captured images shown in their original resolution of 256 pixels (a) and then enlarged 3 times for clarity (b). (c) Shows an example of using CDS in software by subtracting the data acquired at the beginning and the end of the integration times, however, in order to remove the non-uniform illumination, the reset image was taken of a blank white sheet before placing the target.

In order to measure the resolution of the imager, contrast was measured as a function of spatial frequency using a modulation transfer function (MTF) measurement [55]. To obtain the MTF curve, black and white bars (line-pairs) with varying the number of line-pairs per millimeter were imaged. The maximum number of lines is reached when each pixel has one bar passing over it. The contrast ratio is then calculated and plotted as a function of spatial frequency to obtain the MTF curve. The contrast between black and white bars was calculated as

$$MTF = \frac{I_{white} - I_{black}}{I_{white} + I_{black}}, \quad (3.6)$$

where I_{white} and I_{black} are the light intensity of the white and the black bars. Figure 3.17 (a) shows the image of a vertical 3-bar object in a 3-dimensional graph showing all pixels. The pixels from all 16 rows were averaged and Figure 3.17 (b) was obtained, after which,

the pixels corresponding to the black bar were averaged together and the pixels corresponding to the white bar were averaged together, and the contrast ratio was obtained using equation (3.6).

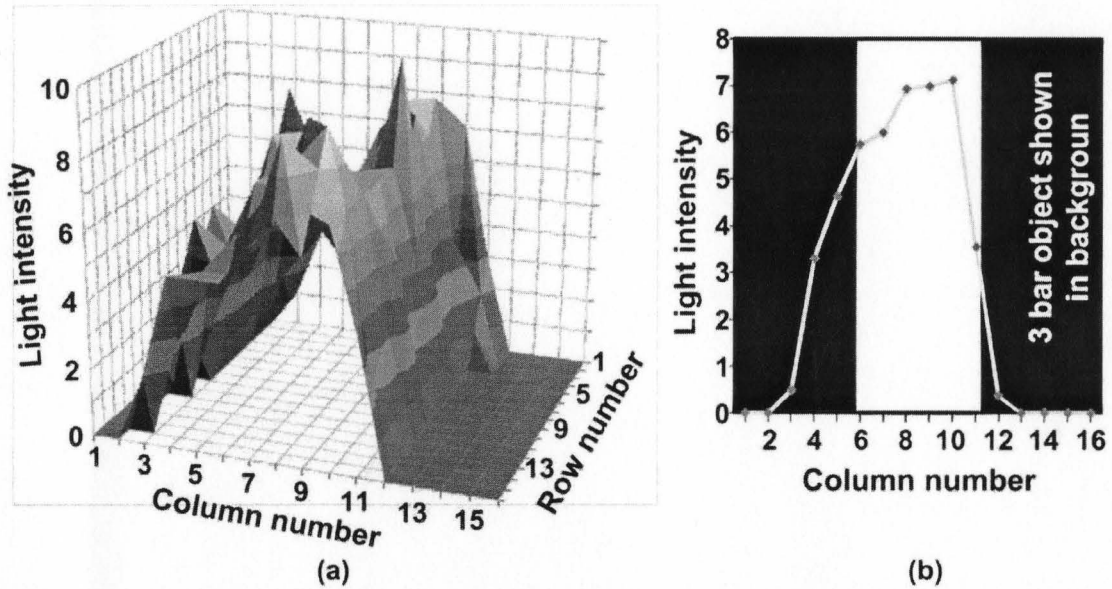


Figure 3.17: (a) Image data of a black-white-black 3-bar target, and (b) the averaged rows corresponding to (a).

The complete MTF measurement (for both the lens and the sensor) for the imager is shown in Figure 3.18 (a) and (b). Here, the vertical and horizontal targets and the captured images (resized), in addition to the MTF waveforms, are shown. Figure 3.18 (c) shows the MTF curve where the 50% points are at around 12 and 15 line-pairs/mm for the horizontal and the vertical MTFs, respectively. The 50% point is important since an object with a spectral resolution higher than this value will result in significant data loss. The reason for the vertical and horizontal MTFs not being equal, showing a higher vertical resolution, although both dimensions have the same number of pixels (16×16) is because the pixel is not a square, as seen in Figure 3.18 (d), where the vertical pitch is smaller than the horizontal pitch. This result illustrates the importance of FF and pixel size in an imager.

The camera-on-a-chip implementation presented in this chapter shows a good platform setup that can be used for testing subsequent imagers that follow in Chapters 4, 5 and 6. The frame-rate achieved by this imager is 60 frames/s, which will be improved

in Chapter 4. The sensitivity will be addressed in Chapter 5. And finally, the dynamic range will be increased using a novel technique that is presented in Chapter 6.

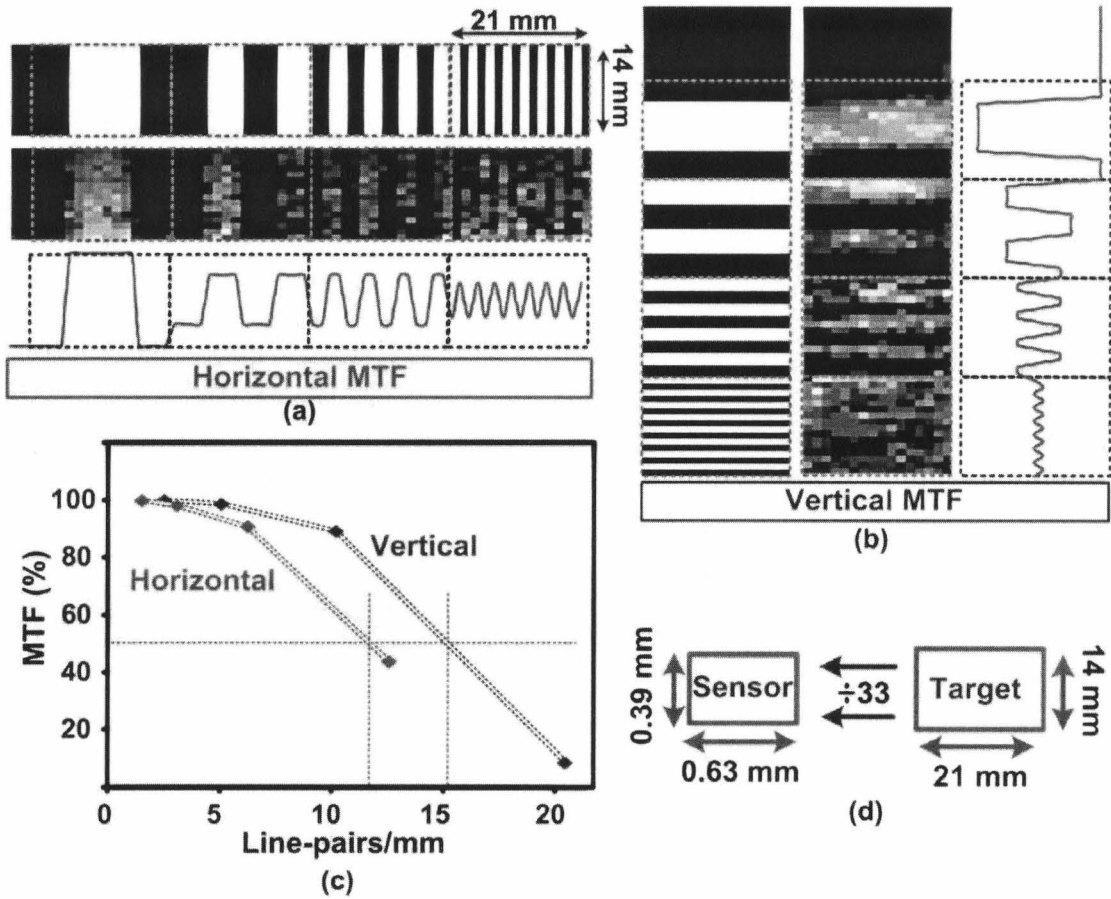


Figure 3.18: Modulation transfer function measurements. (a) Horizontal and (b) vertical MTF target, captured images and contrast waveform. (c) The MTF as a function of spatial frequency. (d) The sensor size relative to the target size showing the height of the sensor being less than its width, which results in higher vertical resolution.

Chapter 4

ULTRAHIGH-SPEED CMOS IMAGER

Emerging imaging applications, such as machine vision, time-of-flight (TOF) imaging, topographic imaging, three-dimensional high-definition television (3D-HDTV) and optical molecular imaging systems, specifically fluorescence life-time imaging (FLIM), have resulted in significant research efforts in designing high-speed imagers [56]. The advances in deep submicron CMOS technologies have especially made such high-speed imaging possible. As mentioned in Chapter 1, one of the main advantages of CMOS image sensors is that they are fabricated in standard CMOS technologies, which allows for full integration of the image sensor along with the processing and control circuits on the same chip and at a low cost. This camera-on-chip system leads to a reduction in power consumption, cost and sensor size, and allows for integration of new sensor functionalities, leading to smart pixel design. Such smart pixels truly show the potential of CMOS technology for imaging applications, allowing CMOS imagers to achieve the image quality and global shuttering performance necessary to meet the demands of ultrahigh-speed applications [56]. Such applications include biometric analysis, robotic visions systems, material analysis, *in-vivo* bio-imaging, and geological surveying.

In order to meet the ultrahigh-speed demands of high frame-rate applications such as FLIM, fast and sensitive CMOS imagers are required. CMOS imagers that can achieve timing resolutions between 150-800 ps from 64×64 pixel imagers with two point per transient waveform sampling and 150 fps, have been reported in the literature [57]. In order to sample a fluorescence lifetime curve without using repeated experiments, a CMOS imager that can capture a number of consecutive frames at sub-nanosecond resolution would be required. The photodiode must be very sensitive as well, which may require the use of avalanche-photodiodes.

Another high-speed imaging application is proton radiography [58], which is a new tool for advanced hydro-testing. Proton radiography had recently become an attractive imaging tool when the blurry images that would result from proton scattering were improved by using a magnetic lens to focus the protons. Proton radiography is especially attractive when imaging thick objects and acquiring images at high frame-rates (5 million fps), which is required for nuclear weapons hydrotests. Image sensors used for proton radiography must be able to capture images at rates of thousands to millions of frames-per-second, even if only for a few frames [58].

In the next section, a review of the existing CMOS high-speed imager designs is given, in addition to a discussion of the various implementations that target ultrahigh-speed imaging. After that, the design of an ultrahigh acquisition rate CMOS active-pixel sensor imager that can take 8 frames at a rate of more than a billion frames per second (fps) is discussed. The ultrahigh-speed design is implemented in a standard 130 nm CMOS technology from IBM.

4.1. Review of High-Speed Imagers in the Literature

Figure 4.1 shows a block-diagram categorizing some of the high-speed imagers in the literature, showing the frame rates (FR) that can be achieved with the two classes of readout architectures. The digital readout architectures that are discussed in this section include the standard pixel-by-pixel (PBP) sequential readout, the per-column analog-to-digital converter (PC-ADC) readout and the per-pixel ADC (PP-ADC) – also called the DPS. The ultrahigh-speed imaging analog techniques are also discussed in this section.

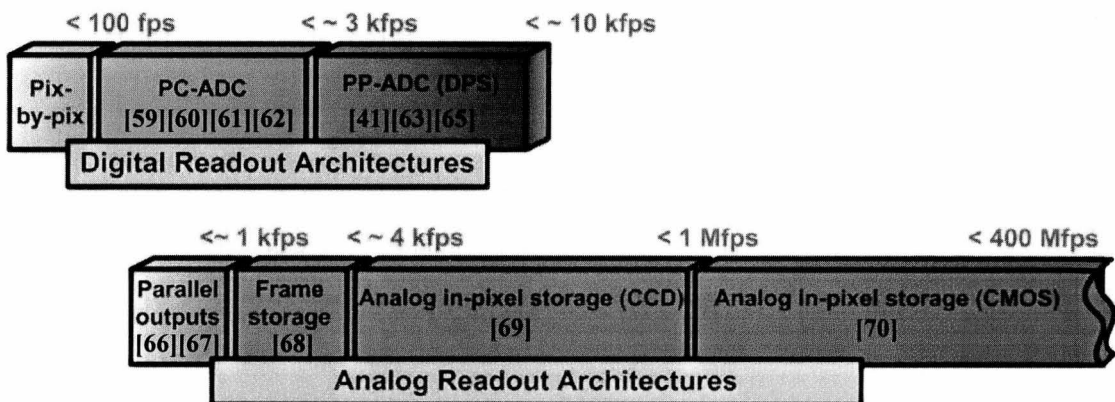


Figure 4.1: A block-diagram categorizing some of the most relevant published high-speed imagers.

4.1.1. Digital readout architectures

Over the past few years, a number of readout architectures ([41], [59]-[65]) have been used for CMOS imagers. Referring back to Figure 4.1, the simplest and slowest form of readout is sequential pixel-by-pixel (PBP) array access. Figure 4.2 shows the sequence of pixel access in such an array [56]. The frame rate (FR) in this case can be calculated as [56]

$$FR_{PBP} = \left[H \times V \left(\tau_{ADC} + \frac{b}{n} \times \tau_{RO} \right) \right]^{-1}, \quad (4.1)$$

where H and V are the number of rows and columns in the array respectively, τ_{ADC} is the time it takes the ADC to complete one conversion, τ_{RO} is the time it takes the chip I/O to send out the converted digital result, b is the number of digital bits, and n is the number of parallel outputs. The dominating factors in equation (4.1) are the $H \times V$ product and τ_{ADC} , which shows that this architecture cannot be used for large arrays. For example, with a τ_{ADC} of 2 μs , the FR drops below 30 fps for an imager of 128×128 and drops below 0.5 fps for a 1M-pixel imager. It is worth mentioning, however, that the pixel-by-pixel readout architecture has the lowest fixed-pattern-noise (FPN) of the three readout architectures, as well as the lowest power consumption. The following subsections discuss the efforts presented in the literature to achieve high-speed imaging on the array-level as well as the pixel-level.

Sequential Pixel by Pixel (PBP) Readout

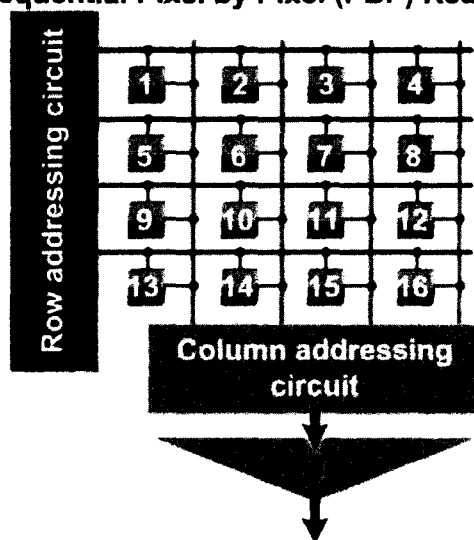


Figure 4.2: Array access in a simple 4×4 pixel-by-pixel (PBP) sequential readout architecture [56].

4.1.1.1. Array-level techniques

One of the most common techniques to increase the FR is to process in parallel as many pixels in the array as possible. The most common technique to do so is to have an ADC per every column of the array [59], [60], [61], [62], as shown in Figure 4.3 (a). The FR of a PC-ADC design is almost V times faster than a sequential readout array, which comes at the expense of an increase in power consumption and silicon area. The FR is calculated as [56]

$$FR_{PC-ADC} = \left[H \times V \left(\frac{\tau_{ADC}}{V} + \frac{b}{n} \times \tau_{RO} \right) \right]^{-1}. \quad (4.2)$$

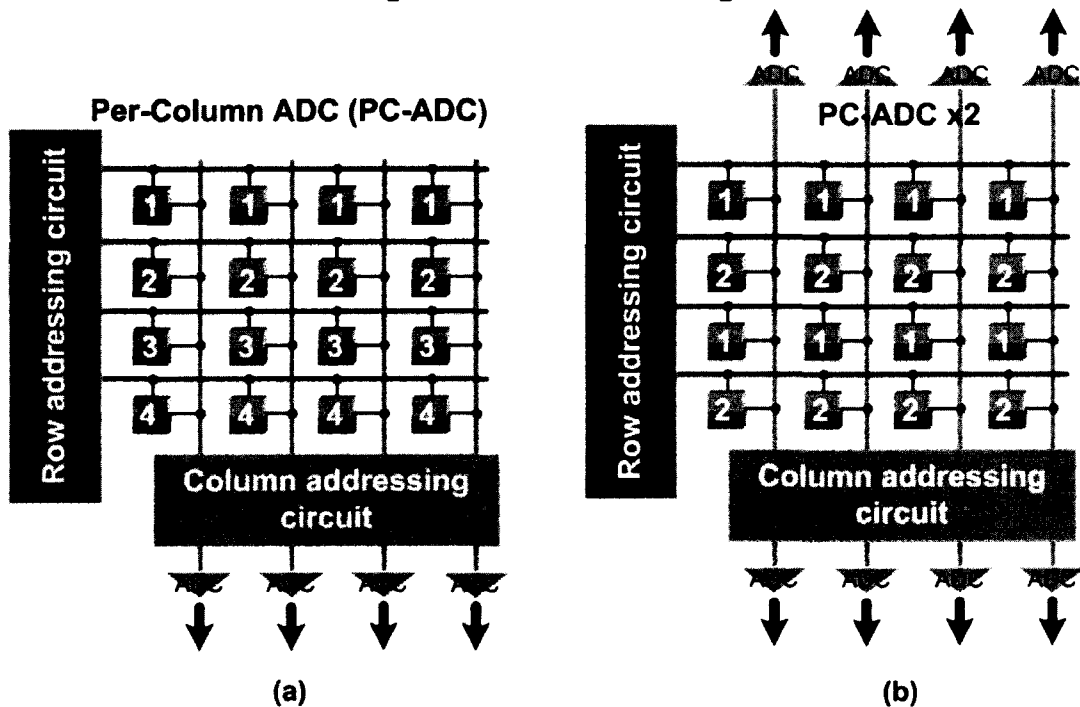


Figure 4.3: Array access in (a) a per-column ADC (PC-ADC) readout and (b) a PC-ADC x2 [56].

Krymski *et al.* [59] described a 1M-pixel imager in 1999 using a 0.5 μm CMOS technology that has a FR of 500 fps. The array of 1024(H)×1024(V) pixels had a PC-ADC architecture that was divided into two groups of ADCs on the top and the bottom of the array since the pixel pitch was very small (10 μm). The authors used a dual-port RAM to double the readout speed since writing to the RAM from the 8-bit ADCs ($b=8$) and readout can be done simultaneously. The authors also used 8 output ports of 8-bits

each in parallel ($n=64$) to send out the data clocked at a master clock rate of 66 MHz ($1/\tau_{RO}$). The ADC conversion time, in addition to the sample time, was 2 μs (τ_{ADC}). By substituting these numbers into equation (4.2), a FR of 248 fps can be found, which is doubled due to the dual-port RAM [59]. This high FR comes at the expense of a power consumption of 350 mW from a 3.3 V supply. If the imager in [59] was designed using a pixel-by-pixel architecture, from equation (4.1), then the FR can be calculated to be 0.5 fps, which would be doubled due to the dual-port RAM to only 1 fps. In 2003, the same group improved on their work described in [59] by using a smaller feature size of 0.35 μm CMOS technology and increased the imager and ADC resolution to 4.1M-pixels and 10-bits with a FR of 240 fps [60]. With such high resolution, although the FR is not very high, this imager delivers 9.75 Gb/s of data. This shows how the bottleneck for high resolution imagers can be the chip I/O transfer. Nishikawa *et al.* [61] reported an on-chip parallel image compression circuit to address the I/O bottleneck. With the proposed compression technique and a master clock rate of 53 MHz, the authors proposed a 3,000 fps 1M-pixel imager [61].

Another way to increase parallelism and improve the FR is to split the array into two groups and have a group of top ADCs and a group of bottom ADCs, each in charge of reading out half of the array, as shown in Figure 4.3 (b). This technique is more feasible in biomedical arrays where the pixel pitch is large due to needing large photodiodes to increase the sensitivity [8], [57]. A more attractive approach is discussed in the following subsection.

4.1.1.2. Pixel-level techniques

Digital pixel sensors (DPS) have become very attractive, since transistors used for digital applications take more advantage of CMOS scaling properties than transistors used for analog circuits. A DPS integrates an ADC into each pixel [63], resulting in a massively parallel readout and conversion that can allow very high speed operation, where the converted digital code is read out of each pixel. Figure 2.25 in Chapter 2 showed a schematic representation of a DPS. In this example, only part of the ADC is included within the pixel to maximize the FF, where an integrating ADC scheme requiring only one ramp generator and one counter that are common to all pixels can be used. The in-pixel opamp compares the photodiode voltage to the ramp voltage (V_{ramp}) and once V_{ramp}

exceeds the photodiode voltage, the 8-bit memory cells will latch the count value that is coming in from the common counter. Using a DPS will only require one ADC conversion cycle for all pixels in parallel, which results in a great increase in FR, assuming that the readout circuits are fast enough to handle the extremely large amounts of data. The high-speed readout makes CMOS image sensors suitable for very high-resolution imagers (multi-megapixels), especially for video applications [56].

The extra circuitry within the pixel in a DPS comes at the expense of a reduced FF. However, the low FF of DPS sensors is no longer an issue for CMOS technologies of 0.18 μm and below [1], [71]. In 2001, Kleinfelder *et al.* [41] described a 352 \times 288 pixel DPS imager in a 0.18 μm CMOS technology, with 37 transistors per pixel. The imager is capable of operating at 10,000 fps (1 Gpixel/s) with a power consumption of 50 mW and a pixel FF of 15%. Ghannoum *et al.* [63] improved the FF in 2007 to 26% by using a 90 nm CMOS technology with 57 transistors per pixel.

The FR of a DPS-based imager can be calculated as [56]

$$FR_{PP-ADC} = \left[\tau_{ADC} + H \times V \times \frac{b}{n} \times \tau_{RO} \right]^{-1} \quad (4.3)$$

Unlike the pixel-by-pixel readout imager, the dominating factor that affects the FR of a DPS-based imager is the I/O transfer speed. Figure 4.4 shows a comparison between the frame rates of the pixel-by-pixel, PC-ADC and PP-ADC readout architectures based on equations (4.1)-(4.3) [56]. From the figures, it can be seen that the FR of the pixel-by-pixel readout architecture is strongly affected by the array size and there is an insignificant effect of the master clock rate. The PP-ADC and PC-ADC on the other hand are mainly affected by the readout speed and the advantage of using a DPS in PP-ADC readout as opposed to PC-ADC readout cannot be realized unless the chip I/O speed can handle the large data rates being generated.

Figure 4.4 (a) shows that the PP-ADC has a constant FR versus resolution, until some point where the FR drops rapidly after the readout speed becomes too slow. Due to the bottleneck in chip I/O readout and ADC conversion times, even with PP-ADC, the published frame rates that use digital techniques are reaching their saturation limits [56]. A number of researchers ([66], [67], [68], [69], [70], in addition to the work presented in

this chapter) explored analog readout methods for CMOS imagers, which is explained in the following sub-section.

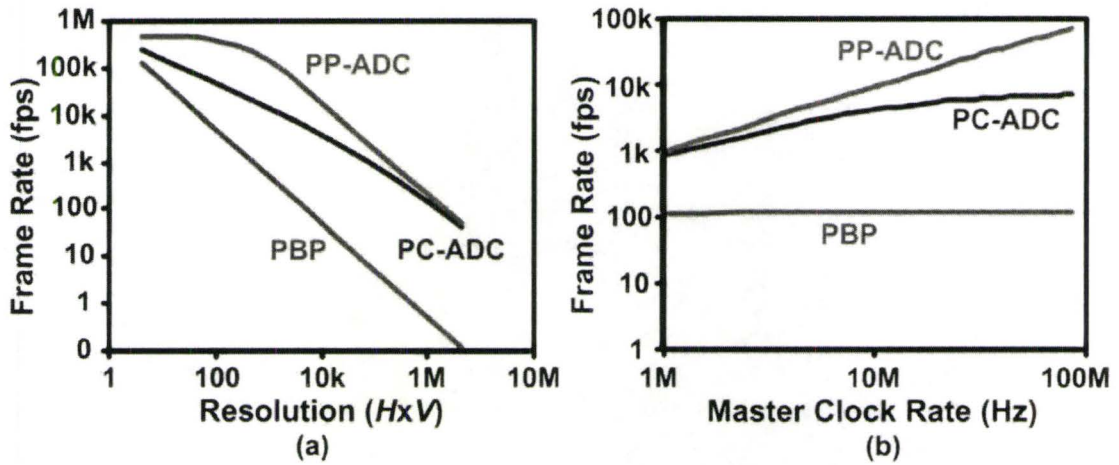


Figure 4.4: Simulation results of equations (4.1)-(4.3) showing the FR of PBP, PC-ADC and a PP-ADC readout architectures with 8-bit resolution ADCs ($b=8$), four 8-bit parallel outputs ($n=32$) and a $\tau_{\text{ADC}} = 2 \mu\text{s}$. (a) The FR as a function of varying the imager resolution with a fixed clock rate of 50 MHz ($1/\tau_{\text{RO}}$). (b) The FR as a function of the clock rate with a fixed imager resolution of $H \times V = 64 \times 64$. Both graphs are shown on a log-log scale [56].

4.1.2. Analog readout architectures

Stevanovic *et al.* [62] and Hosticka *et al.* [72] used imagers with 4 parallel analog output channels and 256×256 pixels achieving over 1,000 fps, while Lauxternann *et al.* [67] used 16 parallel analog output channels and 256×256 pixels to get a FR of 5,000 fps. Another approach to reduce the high-speed requirements of the ADC is to use an analog frame memory array [64], [69]. By using an analog memory, the captured frame can be stored to separate the image capture and data conversion steps from each other. Sugiyama *et al.* [69], [68] used this method for 3-D sensing and implemented a 320×240 pixel CMOS imager that can capture images at 3,300 fps.

Analog memory techniques are even more interesting when the analog memory cell is included within the pixel. Including an analog memory unit within the pixel has been used in many imaging systems, with features such as motion detection [73], high dynamic range with pixel-level integration time control [74], ambient light suppression [75] and cancellation of FPN or offset correction [76]. In-pixel memory is also used for high-speed applications to achieve imaging with a global shutter ([61], [65], [66], [77])

rather than a rolling shutter, as previously explained in Chapter 2. Chapinal *et al.* [73] used the in-pixel storage capacitor to store the captured image for tens of seconds, avoiding the need for an external RAM. Dubois *et al.* [65] used two capacitors per pixel where one captures the current frame while the other holds the previous frame for processing, which increases the FR.

In order to achieve the fastest FR possible for a certain high-speed experiment, a number of extremely fast consecutive images can be captured and stored in analog form. By doing so, the inter-frame delay caused by the ADC conversion time and array readout can be avoided. In this case, the FR only depends on the speed of the devices and transistors used within the pixel, assuming high enough illumination exists on the object being imaged.

The concept of *in situ* storage has been implemented in both CCD imagers [69] and CMOS imagers [70]. Figure 4.5 shows the concept of an *in situ* CCD imager that stores up to N frames [56]. By placing the storage elements in a very small area within or beside a pixel and increasing as much as possible the number of storage elements, which is equal to the number of consecutive frames, the theoretical maximum FR can be achieved [69]. The CCD *in situ* 312×260 pixel imager presented by Etoh *et al.* [69] can capture 100 consecutive images at a FR of 1M fps with a pixel FF of 13%.

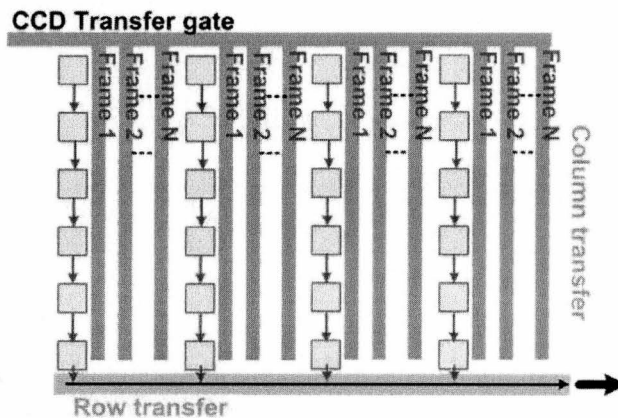


Figure 4.5: The storage and readout of an *in situ* CCD imager that can store up to N frames [56].

A 1-D linear array CMOS implementation of the *in situ* imager has been presented by Kleinfelder *et al.* [70] in a CMOS 0.35 μm technology. The design has a 150 photodiodes with a 150-frame analog storage array and is capable of capturing images at a FR of 400M fps. The authors in [70] suggest using a 3-D packaging technique to achieve a 2-D

imager design where chips are arranged standing on end with a separate photodiode array bonded on top.

4.2. Ultrahigh-Speed Pixel Design and Measurements

In this work, the design of an ultrahigh-speed APS that can capture 8 frames at an acquisition rate of 1.25 billion fps is proposed. The schematic diagram of the pixel, which contains 38 transistors, is shown in Figure 4.6 (a). The basic idea is to utilize 8 analog memory units *in situ* to temporarily hold 8 frames at a very high speed, avoiding the delay time in analog-to-digital conversion and readout. The write switches (WT) that select which storage element to use also serve as global shutters. The storage elements C_{S1} to C_{S8} were implemented using MOS capacitors to reduce layout area; they have a capacitance of 60 fF and were designed using thick-oxide (5.2 nm) devices to reduce leakage.

The pixel was designed using a CMOS 0.13 μm technology kit from IBM. Even though all devices used thick-oxide to increase the dynamic range, this kit can allow for a smaller pixel compared to the CMOS 0.18 μm technology. Figure 4.6 (b) shows a screen capture of the APS layout, which occupies an area of $37 \mu\text{m} \times 30 \mu\text{m}$. The photodiode used has an area of $10 \mu\text{m} \times 10 \mu\text{m}$, which gives a FF of 9%. The low 9% FF is comparable to smart pixel designs such as the digital pixel sensors in [41], with a FF of 15%. The photodiode used is an n+/p-well with a guard ring, which was necessary to increase the speed of the photodiode by eliminating the slowly diffusing substrate carriers [78].

A simulation test of the ultrahigh-speed APS, showing the photodiode voltage variation for 8 different light samples is presented in Figure 4.7 (a). The different light samples were generated using 8 switchable parallel ideal current sources. Note that the 1st and 5th samples are the same and the reset frequency is 1.25 GHz. Figure 4.7 (b) shows the values that are stored in the 8 storage capacitors corresponding to the 8 light samples. The 8 frames were sampled within roughly 7 ns, which makes this imaging technique suitable for FLIM. If repeated experiments were needed for low-light level measurements, the consecutive frames can be accumulated in the capacitor without

clearing the previous frames. Figure 4.7 (c) shows the values read out of the pixel at a readout frequency of 50 MHz where the pattern of test light samples can be seen.

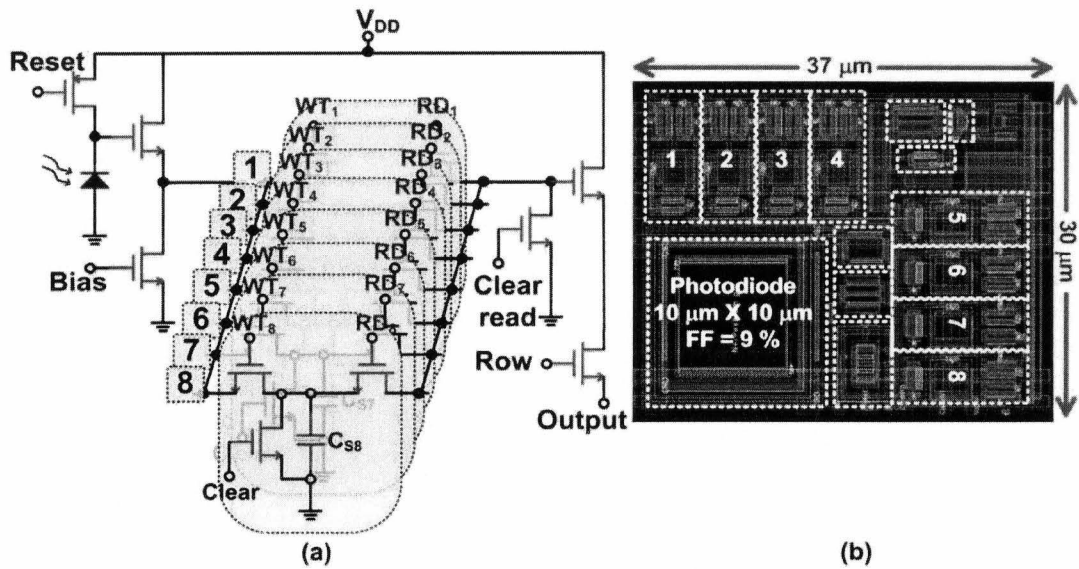


Figure 4.6: (a) The schematic diagram of the ultrahigh-speed in-situ APS containing 8 memory elements and 38 transistors and (b) the layout screen capture of a single pixel [56].

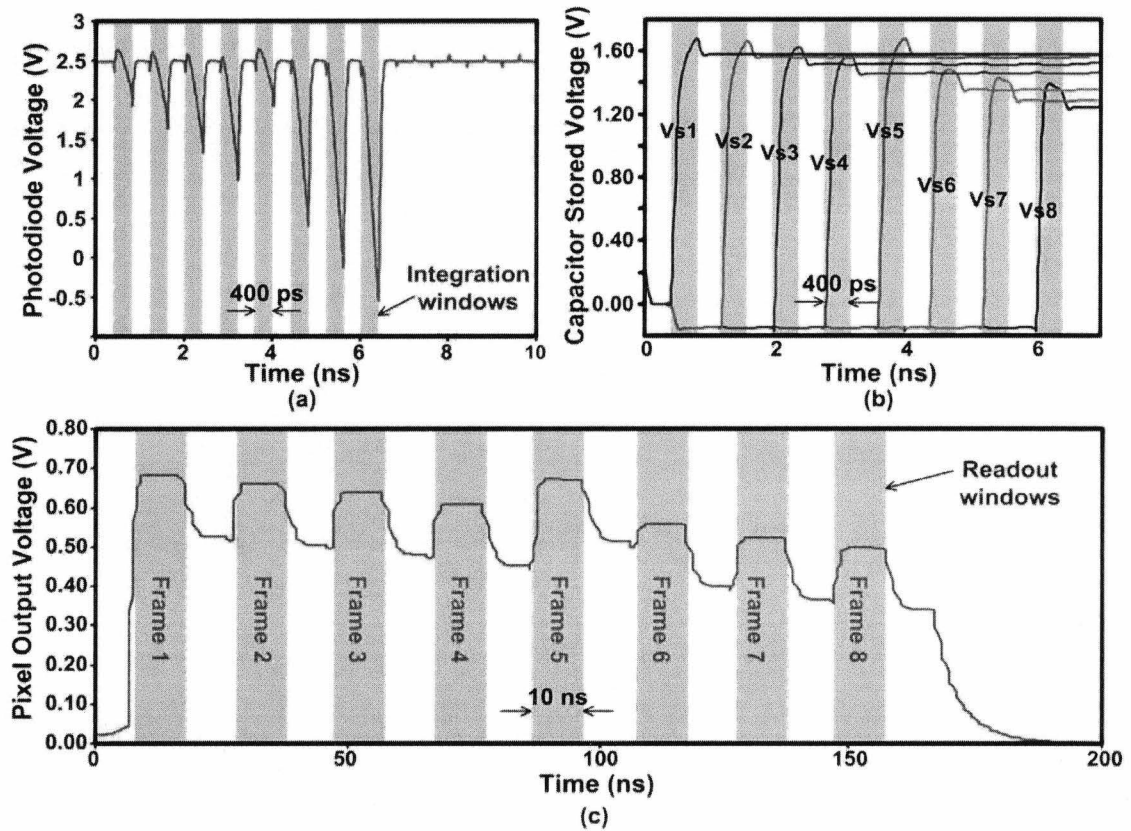


Figure 4.7: (a) Simulated photodiode response for 8 different light samples, (b) corresponding stored voltages, and (c) pixel readout voltage.

4.3. Array and Imager Design

The pixel was used in a 32×32 pixel array, for a complete-camera-on-a-chip design that occupied an area of $2 \text{ mm} \times 2 \text{ mm}$, as shown in Figure 4.9 (a) [79]. In addition to the row and column scanners that were used in Chapter 3, this imager required high-speed circuitry to be implemented on-chip. The imager includes a read sequence generator circuit, write pulse generator circuit, high speed ADC, and on-chip voltage controlled oscillator (VCO) to generate the high speed clocks. The schematic of the VCO is shown in Figure 4.9 (b), which is a cross-coupled negative- g_m 3-GHz oscillator. The VCO clocks the ADC and the write pulse generator circuit. The read circuit is a shift register that is clocked externally to provide the 8 read control signals to the pixels. This was done to reduce the number of I/O pads needed.

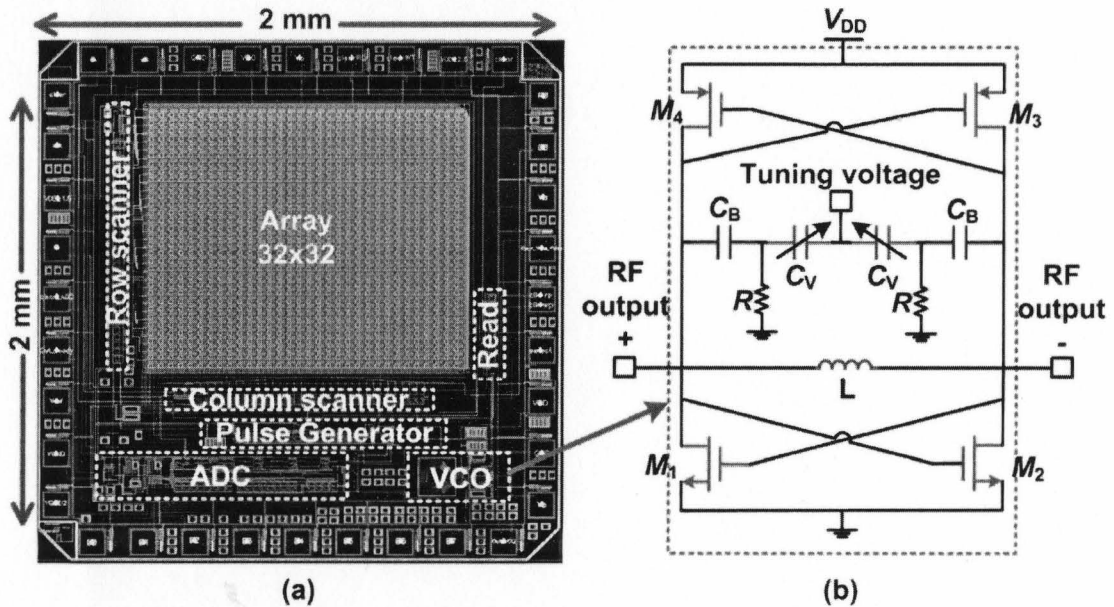


Figure 4.8: (a) The layout screen capture of the complete camera-on-a-chip design [79]. (b) The schematic diagram of the cross-coupled VCO.

The implemented ADC uses the same topology explained in Chapter 3; however, the integrator circuit has a resistor R_R and a capacitor C_C equal to $20 \text{ k}\Omega$ and 2 pF , respectively. These values are much smaller than those used in Chapter 3 since this ADC operates with a 3 GHz clock rather than 1 MHz . The maximum conversion time is 128

ns, which gives a frame rate of 7,629 fps for a single frame. An 8-bit counter was used in this ADC instead of a 6-bit counter, which was used in Chapter 3, to achieve a higher bit depth.

Since the image capture rate is very high, the reset and write pulses cannot be provided externally. Rather, an external trigger is used to initiate the write processes and the write pulses are synchronized and generated on-chip. This is done using an edge-triggered circuit that accepts an external start pulse as an input, and when the start pulse is received, the circuit generates the 8 write pulses that have a width of 400 ps each, as well as generating the reset signal in between the pulses. The logic schematic diagram of the write pulse generator circuit is shown in Figure 4.9 (a). The input is edge-triggered and the circuit will only generate a set of 8 write pulses one time, even if consecutive start signals are inputted. In order to generate the next set of write pulses, the circuit must first be reset, then, the input start signal can be applied. Figure 4.9 (b) shows the simulation results of the pulse generator circuit clocked at a frequency of 1.25 GHz. The actual clock will be roughly 1.5 GHz, which is the VCO frequency divided by 2 using a digital divider. When the start pulse is received at 3 ns, the circuit generates the first write pulse (shown in the lower inset figure) and disables the reset transistor of the pixel. The reset of the pixel is active low since a PMOS device was used.

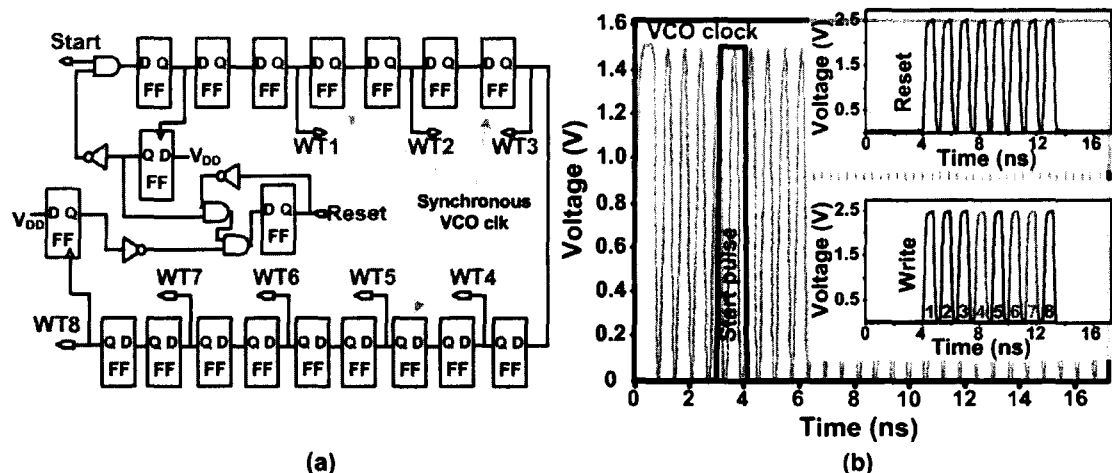


Figure 4.9: (a) Schematic diagram of the write and reset pulse generator circuit. (b) Simulation results of the write and reset pulse generator circuit showing the start pulse coming in at 3 ns with a clock frequency of 1.25 GHz. The top inset figure shows the 8 reset pulses (active low) and the bottom inset figure shows the generated 8 write pulse signals that have a width of 400 ps.

In order to increase the number of frames that can be captured consecutively, a 1D line-scan imager can be used where the imaging array can be arranged into a line of APS pixels followed by an array of memory elements, as shown in Figure 4.10. A 2D image can be coupled to the 1D line-scan imager using fiber coupling to achieve ultrahigh-speed imaging without sacrificing the array fill-factor.

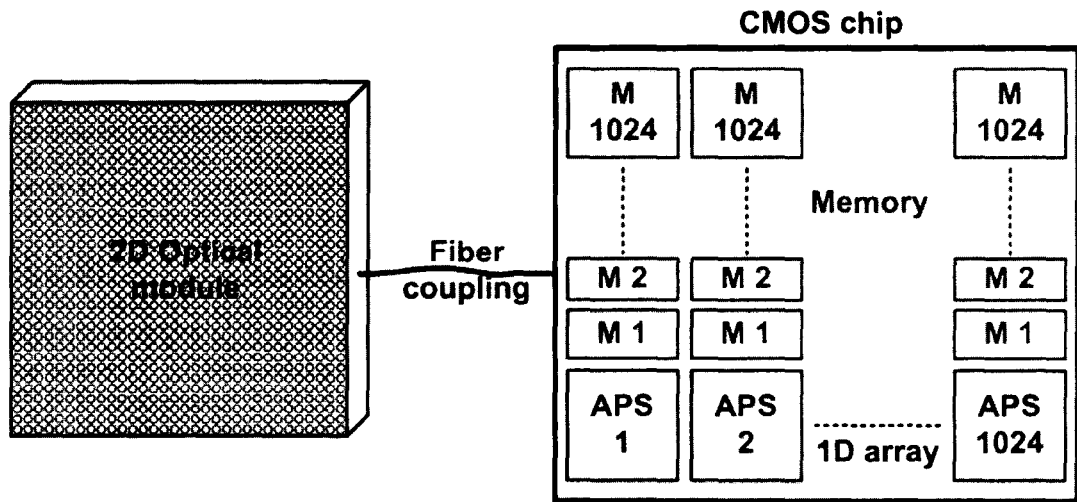


Figure 4.10: Increased number of consecutive frames to 1024 using a 1D line-scan imager and fiber optic coupling.

4.4. Measurements Results

A complete-camera-on-a-chip design occupying an area of $2 \text{ mm} \times 2 \text{ mm}$, shown in Figure 4.11, was fabricated with an array of 32×32 ultrahigh-speed pixels. The microchip was fabricated in IBM's 130 nm CMOS technology (8RF-DM CMOS), which has 8 metal layers, making it easier to connect the complicated smart pixel design. The technology provides thick top metal layers, with the topmost metal having a thickness of $4 \mu\text{m}$. This is needed for all high frequency signals and reactive components, such as the inductor in the VCO. Care was taken to isolate the high-frequency signals from the DC or low-frequency signals. The substrate of the VCO was isolated from the rest of the chip by using guard rings and deep-n-wells. Also, multiple DC filtering capacitors were implemented on-chip to reduce ripples on the bias lines and separate DC, AC, and RF bias lines were used. The pixel-level measurement results are shown next.

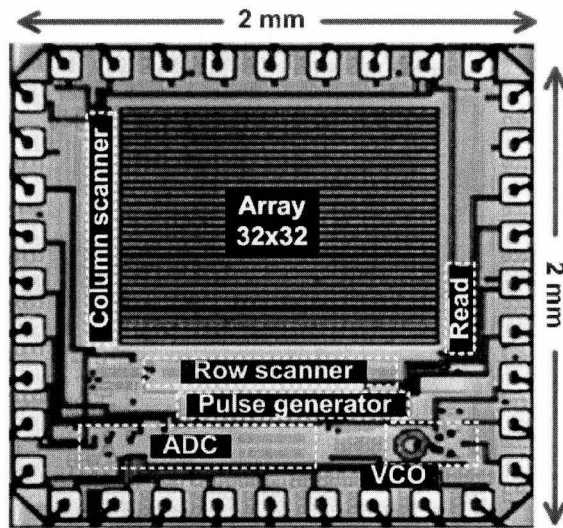


Figure 4.11: Photomicrograph of the ultrahigh-speed camera-on-a-chip fabricated in a 130 nm CMOS technology.

The output of the ultrahigh-speed APS pixel, shown in Figure 4.12 (a), was measured by controlling one of the memory elements. The measured output, shown in Figure 4.12 (b), was obtained by keeping both the write (WT_1) and the read (RD_1) switches closed at the same time, bypassing the storage element and testing the light response of the APS. The figure shows how the response changes with increasing the light intensity. The reset pulse that was applied was generated from the FPGA’s 50 MHz clock. The measurement shows how the pixel responds well to a high speed pulse.

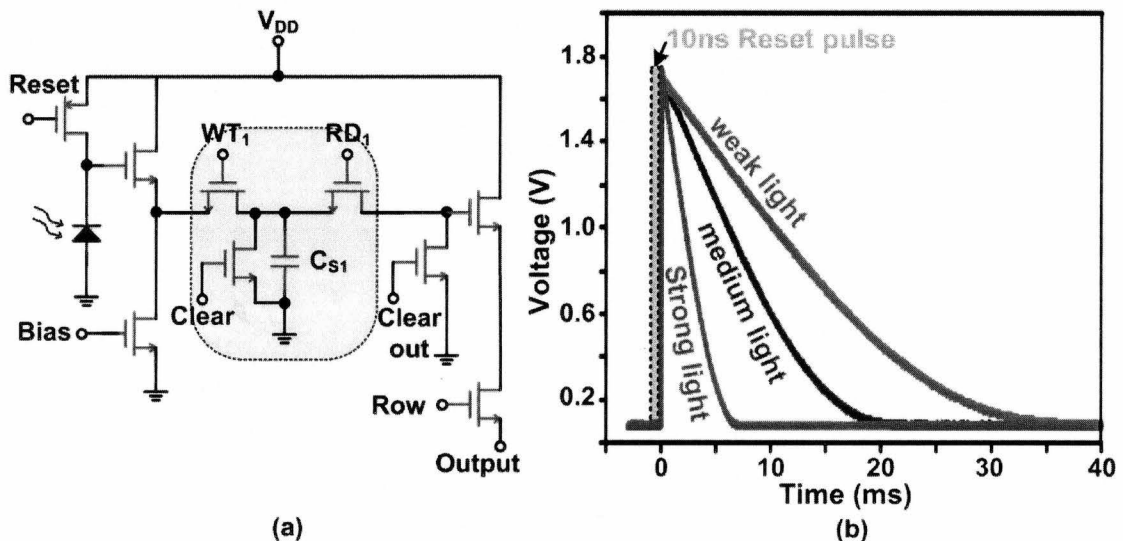


Figure 4.12: (a) The simplified schematic of ultrahigh-speed APS, and (b) the measured APS output voltage for 3 different (weak, medium, and strong) incident light powers.

As explained in Chapter 2, there are a number of noise sources in an APS pixel. As a quick estimate for the amount of noise available at the output of the APS, the envelope of the output voltage was recorded (for 256 measurements) and overlapped with a single sample measurement for comparison, shown in Figure 4.13. This shows an estimate of the maximum noise in a 256 population sample. It can also be seen that the noise increases with integration time, which is mainly due to the integration noise (Chapter 2).

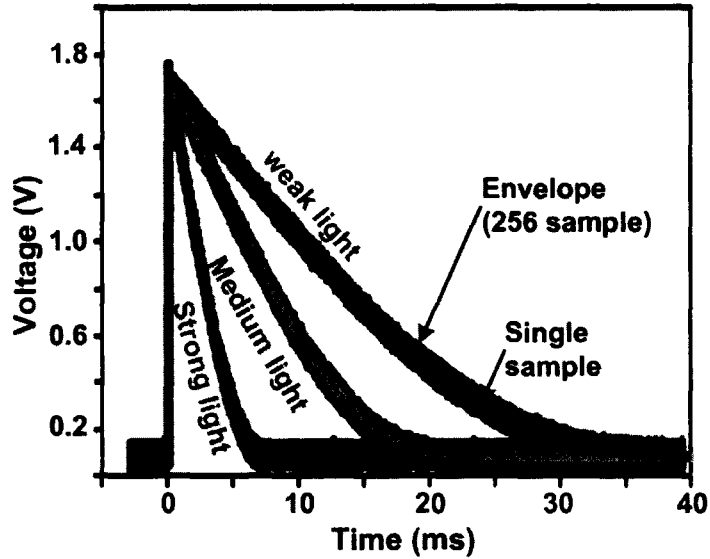


Figure 4.13: The measured APS output voltage for 3 different incident light powers showing both a single sample measurement and the envelope of 256 samples.

The noise sources present in this APS structure are similar to the standard 3T-APS explained in Chapter 2 in terms of reset noise and integration noise. However, the readout noise is different since there is a transfer from the photodiode to the storage capacitor phase, as well as a transfer from the storage capacitor to the column phase. The readout transfer to storage capacitor phase is shown in Figure 4.14 (a) and the small-signal noise equivalent circuit is shown in Figure 4.14 (b). The readout noise for this stage, assuming a constant storage capacitance C_{S1} , can be given by

$$\overline{V_{n_{out}}^2} = \underbrace{\frac{2kT}{3C_{S1}} \frac{1}{1 + \frac{g_{m2}}{g_{d3}}}}_{M_2} + \underbrace{\frac{kT}{C_{S1}} \frac{1}{g_{d3} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)}}_{M_3} + \underbrace{\frac{2kT}{3C_{S1}} g_{m4} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)}_{M_4}. \quad (4.4)$$

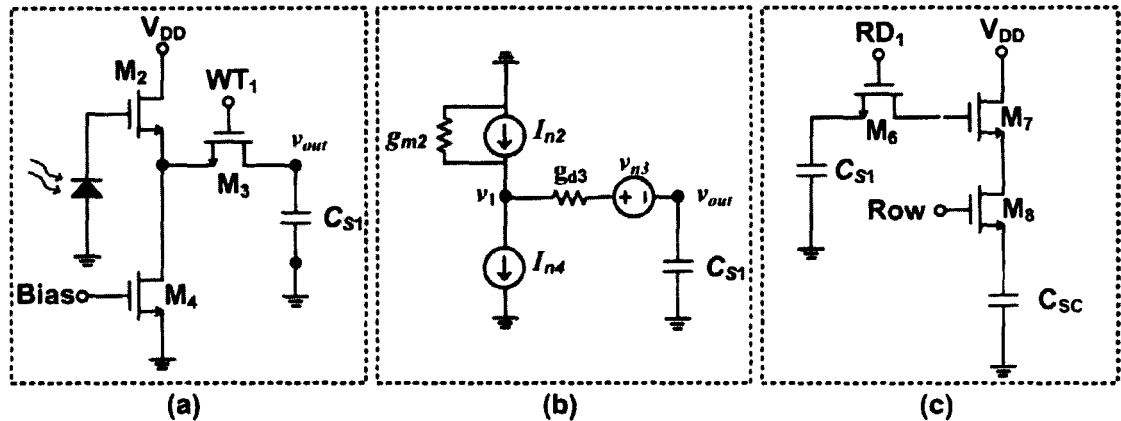


Figure 4.14: Photodiode to storage capacitor readout equivalent circuit (a) and noise equivalent circuit (b). Storage capacitor to column readout equivalent circuit (c).

The readout transfer from the storage capacitor to the column capacitance (C_{SC}) phase is shown in Figure 4.14 (c). The readout noise for this stage, assuming a constant column capacitance, is given by

$$\overline{V_{n_out}^2} = \underbrace{\frac{kT}{C_{SC}} \frac{1}{g_{d6} \left(\frac{1}{g_{d6}} + \frac{1}{g_{m7}} \right)}}_{M_6} + \underbrace{\frac{2kT}{3C_{SC}} \frac{1}{1 + \frac{g_{m7}}{g_{d8}}}}_{M_7} + \underbrace{\frac{kT}{C_{SC}} \frac{1}{g_{d8} \left(\frac{1}{g_{d8}} + \frac{1}{g_{m7}} \right)}}_{M_8} \quad (4.5)$$

Based on the analysis of the noise sources in the APS, the SNR can be calculated by summing up all the noise sources. The SNR was measured using 83 samples and plotted as a function of the integration time for three different light powers. Figure 4.15 (a) shows an example of one of the samples used for SNR calculations. These curves were sampled at different integration times and the values were compared to a population of 83 samples where the noise was obtained as the standard deviation and the signal was obtained as the average. Figure 4.15 (b) shows a comparison between the “semi”-calculated (the noise was calculated but the signal was measured) and measured SNR, showing the standard error bars. The maximum achievable SNR is 45 dB, which is achieved at a specific integration time that depends on the power of the incident light. The inset figure shows that the SNR decreases beyond the saturation point of the pixel, which happens since the signal no longer increases while the noise increases mainly due to shot noise that is a function of the integration time (equation (2.45)).

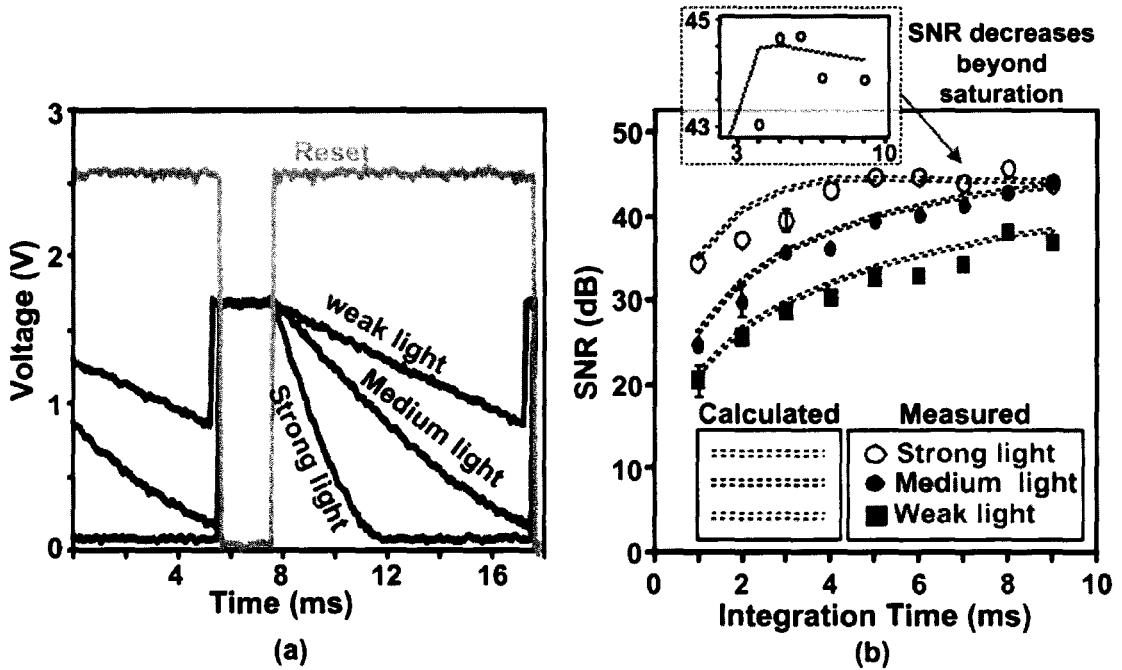


Figure 4.15: (a) The measured APS output voltage for 3 different incident light powers used of a single sample, which was repeated 83 times to measure the SNR. (b) The measured compared to the calculated SNR for 3 different light powers with an inset figure showing a close up to the strong light measurement to show that the SNR drops beyond saturation.

An important question that rises about this ultrahigh-speed design is whether or not the design is scalable. The only limitation on the scalability of this design is on how long it will take to read out the array before the stored values drop significantly. The leakage of the storage capacitor was measured as follows. Referring back to the pixel shown in Figure 4.12 (a), the charge leakage of the storage capacitor can be measured by applying a short reset period, followed by a short write period (WT_1 is on), and then turning the write switch off while keeping the read switch (RD_1) on the whole time. Figure 4.16 shows the measured curve in dark and with applying light. The slope of the dark current was obtained as -23 V/s, whereas the slope of the applied light signal was measured as -154 V/s. During the hold period where the write switch is off, the storage capacitor was leaking at a rate of -78 V/s.

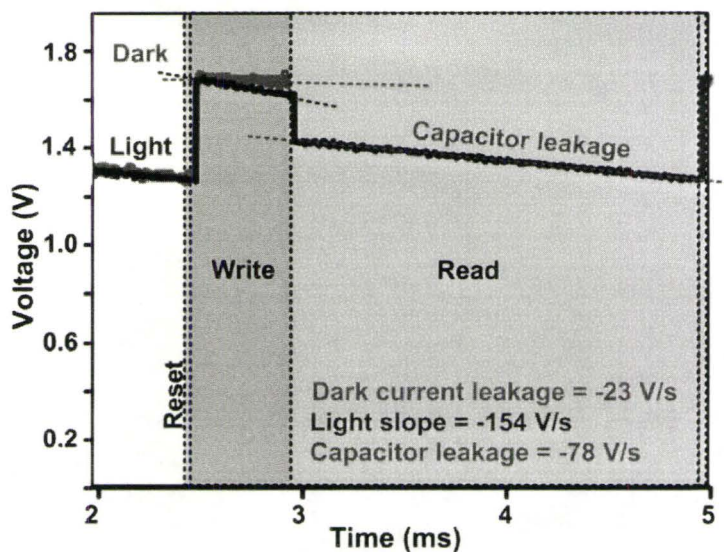


Figure 4.16: The measured storage capacitor leakage of the ultrahigh-speed APS.

The time elapsed starting from pixel $(0,0)$ to pixel (H,V) of frame 8, which corresponds to reading out all 8 stored samples, is given by

$$T_{f_1(0,0) \rightarrow f_8(H,V)} = \frac{8}{FR}, \quad (4.6)$$

which gives a maximum leakage from storage capacitor 1 in pixel $(0,0)$, to storage capacitor 8 in pixel (H,V) of

$$V_{drop} = \frac{78 \times 8}{FR}. \quad (4.7)$$

The voltage drop can be simulated as a function of the array size $(H \times V)$, as shown in Figure 4.17. Assuming that an acceptable drop is less than 10% of the voltage range, which would be 100 mV, the figure shows that using a pixel-by-pixel (PBP) readout architecture would limit the resolution to an array of roughly 32×32 , whereas, a per-column ADC (PC-ADC) readout would be limited to an array size of roughly 168×168 . Finally, the per-pixel ADC (PP-ADC) would allow an array size of roughly 180×180 . This shows that if a higher resolution is required and a voltage drop less than 100 mV is to be maintained, then a readout clock that is faster than 50 MHz is required. The number of parallel outputs and the speed of the ADC should also be increased. However, this shows that the ultrahigh-speed pixel design is limited to small array sizes that are suitable for imaging small areas in biomedical and FLIM applications.

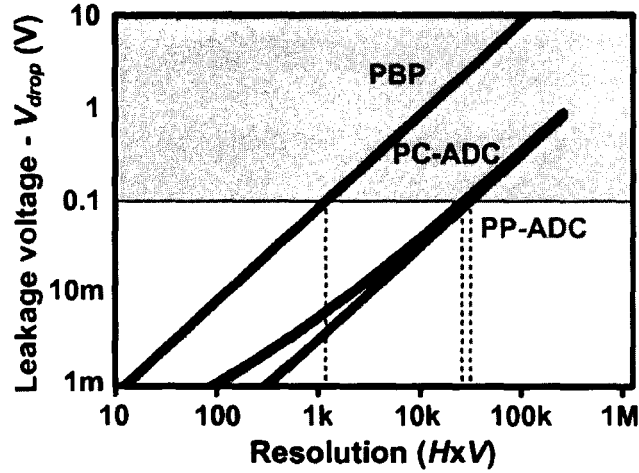


Figure 4.17: Simulation results of equations (4.7) showing the FR of PBP, PC-ADC and a PP-ADC readout architectures with 8-bit resolution ADCs, four 8-bit parallel outputs and a $\tau_{ADC} = 128$ ns, for a fixed clock rate of 50 MHz.

Table 4.1 shows a summary of the different high-speed imager architectures discussed in this chapter. The power consumption (P_{DC}) is not applicable to the *in-situ* techniques, since it is not a property or tradeoff of the pixel and depends on the readout method used. The following figure of merit (*FoM*) has been used to compare the tradeoffs of the different techniques

$$FoM \left[1/\mu m^2 \right] = \frac{FR \left[MHz \right] \times FF \times \text{Array Size}}{\text{Pixel Area} \left[\mu m^2 \right] \times \#outputs \times Clk \left[MHz \right]} \quad (4.8)$$

where the *FoM* increases with frame rate (*FR*), fill-factor (*FF*) and array size, and decreases with pixel area, number of parallel output bits (*#outputs*) of the imager and the readout clock (*Clk*). It was mentioned earlier in this chapter that these parameters tradeoff with high speed. The tradeoff of *FF* and speed is specific to the cases of smart-pixels or *in-situ* imagers. A more general image sensor *FoM* can include sensitivity, expressed as the dark current (I_{dark}) and dynamic range (*DR*) as follows

$$FoM \left[1/\mu m^2 \right] = \frac{FR \left[MHz \right] \times FF \times \text{Array Size} \times DR}{\text{Pixel Area} \left[\mu m^2 \right] \times \#outputs \times Clk \left[MHz \right] \times I_{dark}}, \quad (4.9)$$

this *FoM* was not used however, since *DR* and I_{dark} are not usually reported for high speed imagers, since the main focus is the speed and resolution.

Table 4.1: Summary of the various high-speed CMOS imagers available in the literature.

REF.	TEC. (μm)	SCHEME	# OUT-PUTS	CLK (MHz)	FR (FPS)	ARRAY SIZE	PIXEL AREA (μm^2)	FF (%)	P _{DC} (mW)	FOM ($10^{-3}/\mu\text{m}^2$)
[59]	0.5	PC-ADC & RAM	64	66	0.5k	1024×1024	100	45	350	0.56
[60]	0.35	PC-ADC & RAM	160	66	0.24k	2352×1728	49	43	700	0.81
[61]	0.25	PC-ADC & compression	32	16.8	3k	256×256	225	50 ⁴	--	0.81
[62]	0.25	PC-ADC	160	68	3.5k	512×512	400	50 ⁴	220	0.11
[42]	0.18	PP-ADC	64	167	10k	352×288	88.4	15	50	0.16
[64]	0.09	PP-ADC	64 ⁴	167 ⁴	0.4k	64×48	81	26	--	0.00
[65]	0.35	DPS	64 ⁴	167 ⁴	10k	64×64	1225	25	250	0.00
[66]	1.0	PAO [*]	4	22	1.04k	256×256	900	40	320	0.34
[67]	0.5	PAO [*]	16	24	5k	256×256	162	42	270	2.21
[68]	0.35	AFS ^{**}	8 ⁴	50 ⁴	3.3k	320×240	125.4	53	82	2.68
[69]	CCD	<i>In-situ</i>	8 ⁴	50 ⁴	1M ¹	312×260	4395	13	NA	6.00
[70]	0.35	<i>In-situ</i>	8 ⁴	50 ⁴	10.5M ²	12×12	40,000	10 ⁴	NA	0.01
This work	0.13	<i>In-situ</i>	8	50	1.25B ³ simulated ⁵	32×32	1110	9	NA	259
					100M ³ measured					20.8

*PAO: parallel analog outputs

¹Captures 100 consecutive frames³Captures 8 consecutive frames⁴Extrapolated/assumed

**AFS: analog frame storage

²Captures 64 consecutive frames

NA: not applicable

⁵The simulated result requires a more sensitive photodiode

Chapter 5

APD-BASED SINGLE-PHOTON IMAGER

The ability to detect a single photon is the ultimate level of sensitivity in the acquisition of optical radiation, and allows for maximum speed. Therefore, photon counting is the technique of choice for accurate measurements of very weak optical signals and fast light pulses that are in the nanoseconds and picoseconds range. In particular, time-correlated single photon counting (TCSPC) is one of the best ways of measuring fluorescence decay times and is extremely useful for FLIM. Currently, only photomultiplier tubes (PMT) and solid-state avalanche photodiodes (APD) can achieve the levels of sensitivity required for FLIM. However, for emerging miniaturized and portable imaging applications, PMTs are unsuitable because of their large size, high operating voltages, and very high cost. Since APDs operated in the Geiger mode are the semiconductor equivalent of PMTs, then they are much more suitable for miniaturized imaging systems. This is because the APDs possess the typical advantages of microelectronic devices, such as small size, low operating voltages, and low cost. With recent advances in mainstream RF deep-submicron (DSM) CMOS technologies, it is now possible and advantageous to implement the single-photon avalanche photodiodes (SPAD) with the necessary peripheral circuits on the same chip. In this way, we can realize an integrated, ultra-sensitive, high-speed, and low-cost biomedical imaging system that can be used in miniaturized sensing and imaging devices such as wireless endoscopy capsules or endoscopes.

Two approaches prevail in the design of low-light level CMOS imagers. The first one is a hybrid approach that utilizes highly optimized CMOS processes to fabricate high-performance SPAD devices, which can be then connected to a separate detection and control CMOS chip. The second is the monolithic approach, where the SPAD and all

peripheral circuitry are integrated on a single chip. In the first approach, a wide range of CMOS SPAD devices have been fabricated in a number of different CMOS technologies with characteristics that are consistent with the requirements for high-performance SPADs, such as low-doped diffusions, lower electric fields, high-quality material and limitation of contaminants. As such, these devices exhibit performance comparable to PMTs for photon counting applications, such as high photon detection efficiencies (PDE) and high timing resolutions in the range of several tens of picoseconds [80]-[85]. However, the main disadvantages of such systems are increased cost, lower levels of miniaturization, and larger parasitics, which can limit the performance. To increase the use of SPAD-based detector systems for state-of-the-art biomedical applications, integration of the SPAD and the high-speed peripheral circuitry on the same chip using mainstream DSM CMOS technology is highly desired.

This chapter focuses on the design and measurement of a passively quenched SPAD fabricated in a deep-submicron standard RF 0.13 μm CMOS process, and the design of a low deadtime active quench and reset Geiger mode single-photon counter that uses novel deadtime reduction techniques. In Section 5.1, the design and measurement results of the avalanche photodiode, implemented in a standard deep-submicron technology, are discussed, followed by the passive and active Geiger mode SPAD designs in Sections 5.2 and 5.3, respectively. Finally, in Section 5.4 deadtime reduction techniques is discussed.

5.1. APD Design and Measurements

Fabricating an APD in a standard CMOS technology that is not optimized for imaging can be quite challenging. Figure 5.1 (a) shows an example of the layout cross-section of a standard pn-junction. Although this simple structure can form an APD, it has one major drawback, the electric field around the junction peripheral edges (arrows in the figure) reaches breakdown before the lateral area of the pn-junction. This phenomenon is known as premature edge breakdown (PEB). However, since almost all the light passes through the lateral area of the junction, it is necessary to have a laterally uniform electric field and multiplication. PEB occurs before the lateral part, thus, most of the charge multiplication events will be triggered by thermal generation, which gives rise to large values of dark count and would prevent light detection. For a similar reason, it is preferable to have a

circular APD structure to eliminate sharp junction corners or edges. However, circular shapes might not be possible in standard CMOS technologies, which is why hexagons or squares are usually used.

When using a triple well CMOS process, PEB can be prevented by having a lower doping concentration and electric field at the edges of the pn-junction, which would result in increasing the PEB voltage so that lateral breakdown occurs before edge breakdown. The layout cross-section of an APD with reduced edge doping is shown in Figure 5.1 (b). By using a deep n-well layer, a p-well can be isolated from the substrate. By violating the design rules of the technology and placing an n-well within the p⁺ layer, the edges of the junction will be connected to two p-well guard rings that have lower doping than the p⁺ layer. However, if the p-well regions get too close, the active area of the APD will be almost fully depleted. This will result in bypassing the p⁺/n-well junction and the APD will perform as a p-well/n-well diode, as shown in Figure 5.1 (c).

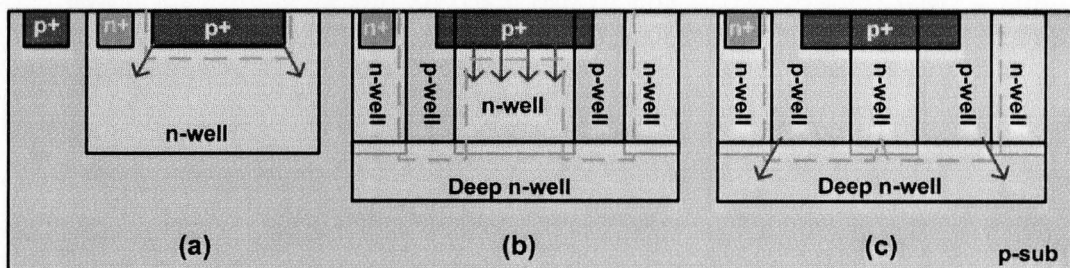


Figure 5.1: Possible diode layout cross-sectional view. (a) Typical pn-junction. (b) Typical APD layout and (c) Smaller size APD layout. The dashed lines show the depletion region.

The APD that was fabricated in this work using IBM's 130 nm CMOS technology is shown in Figure 5.2 (a) [86]. In this technology kit, when implanting an n-well within a p-well that is isolated by a deep n-well, the n-well will not go deep enough to touch the deep n-well. This allowed a modification to the standard APD shown in Figure 5.1 (b) that uses a p⁺/n-well diode. The diode implemented in this work uses an n⁺/p-well diode, which allowed for smaller diode implementation since the case of merging depletion layers previously explained in Figure 5.1 (c) no longer happens. The small size is important for parallel diode implementations that reduce the deadtime and will be explained later. Also, when comparing the n⁺/p-well diode (Figure 5.2 (a)) to the p⁺/n-well diode (Figure 5.1 (b)), the parasitic capacitance of the device is lower in the former. This is due to the n-well having a larger capacitance to the substrate, while the n⁺ layer

has a much smaller capacitance to p-well. Since in both cases the signal is sensed at the n-side, the design with a lower parasitic capacitance at the sense node (Figure 5.2 (a)) would be more desirable for increased speed of operation.

The APD layout top view is shown in Figure 5.2 (b), which has an active area of $10\ \mu\text{m} \times 10\ \mu\text{m}$. Due to the layout spacing requirements, the active area of the diode is only 41%. The APD design needs to be optimized and the effect of the spacing and width of the n-well guard-rings needs to be investigated. We expect that adjusting the separation between the active region and the shallow trench isolation (STI) may have an effect on the dark count rate. It is obvious that the extra layers added by the deep n-well layer results in a significant increase in overhead area. If arrays of APDs were to be implemented using the configuration shown in Figure 5.1 (b), each APD would require its own deep n-well isolation, whereas, if the n+/p-well implementation was used, the increase in overhead space would be minimal.

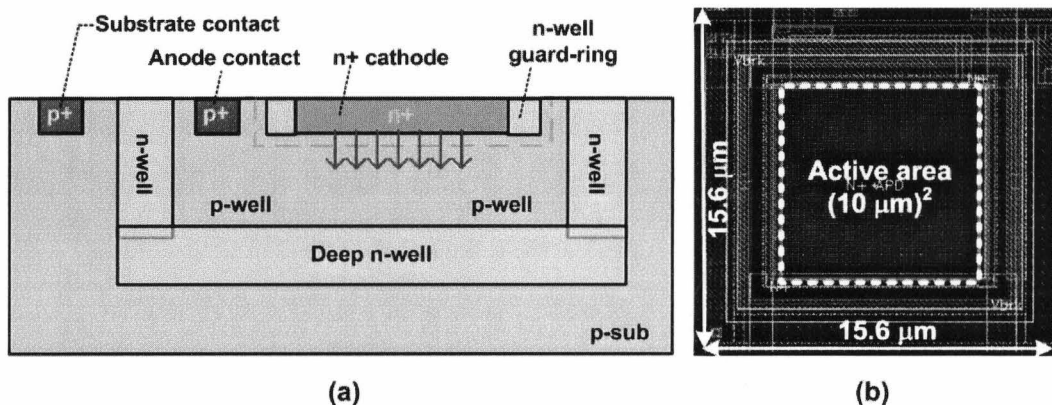


Figure 5.2: (a) APD layout cross-sectional view, showing how the avalanche area was confined only under the n+ region when using the n-well guard-ring [86] and (b) layout top view showing APD device dimensions.

Figure 5.3 shows the $I-V$ characteristics of the diodes with and without the guard ring. The breakdown voltage is slightly higher for the APD with the guard ring as compared to the photodiode without the n-well guard-ring. Also, the APD $I-V$ curve shows the multiplication region where the dark current is multiplied by a finite gain, at which point there is a linear relationship (saturation on the semi-log scale) with a resistance of roughly $100\ \text{M}\Omega$, acting as a microplasma [87]-[89].

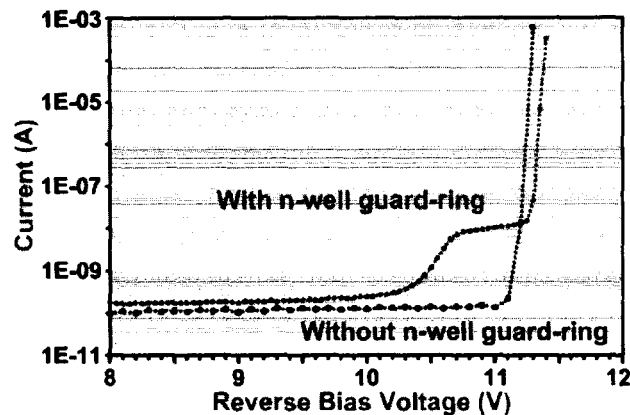


Figure 5.3: Measured I-V profiles of the APD with the guard-ring and the exact same structure without the guard-ring. The breakdown voltage of the APD is 11.3 V.

5.2. Geiger Mode Passive APD

Single-photon avalanche-photodiodes (SPADs) are APDs operating at a voltage that is slightly above the onset of the breakdown, so that the avalanche multiplication triggers a microplasma current. The microplasma current is in the range of several tens to hundreds of microamperes [87]-[89], producing a sharp pulse that can be counted. This regime of operation is known as the Geiger mode [90]-[92]. The extent of which the SPAD is biased above the breakdown onset is referred to as the excess bias (V_E). In Geiger mode, the breakdown mechanism of the SPAD is designed to be avalanche multiplication, where photogenerated carriers become multiplied by impact ionization in a high electric field in the depletion region, thus triggering a self-sustaining avalanche process which delivers a current pulse in the milliamperere range to the external circuit. The leading edge of the generated pulse marks very precisely the photon arrival time. The number of carriers generated as a result of the absorption of a single photon determines the optical gain of the device, which in the case of SPADs operating in Geiger mode is virtually infinite. The breakdown can also occur by tunneling, which reduces the optical gain. Therefore, the electric field profile within the depletion region needs to be properly designed (by means of doping profile) to avoid band-to-band tunneling effects. The self-sustaining microplasma current would continue flowing until the electric field is reduced below breakdown onset. The conduction process can be stopped using a quenching circuit, which can be passive or active. The simplest quenching circuit uses a large

resistor in series with the APD, as shown in Figure 5.4. The excess bias in this case can be defined as

$$V_E = V_R - V_A - V_{BD}, \quad (5.1)$$

where V_{BD} is the breakdown knee voltage of the APD and V_A is the negative voltage applied. As the breakdown process starts, whether due to an incident photon generating an electron-hole pair, or due to thermal generation, an avalanche current will start to flow through resistor R . This current will cause the voltage V_C to drop from its initial value (V_R) until the excess bias drops to nearly zero taking the device out of Geiger mode. After that, the voltage V_C will rise back to V_R according to the RC time constant of resistor R and the APD's capacitance plus any other parasitic capacitance.

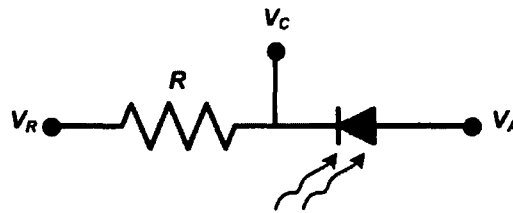


Figure 5.4: SPAD passive quenching circuit.

Figure 5.5 shows oscilloscope screen captures of the measured voltage at node V_C with and without applying light to the device at two different values of excess voltage. It can be seen that the amplitude of the SPAD pulses, in addition to the number of pulses with incident light, increases at higher excess biases. This will be explained in more detail later.

Figure 5.5 (a) and Figure 5.5 (c) also show that even without applying any light, there will be pulses due to thermally generated electron-hole pairs. These current pulses are indistinguishable from a pulse produced by the detection of a photon and therefore, result in a dark count rate (DCR). In many applications, the average dark rate can be measured and subtracted. However, the statistical variation in the dark rate cannot be subtracted, and it constitutes a noise source that determines the minimum detectable signal. Thus, the DCR is one of the main factors limiting the performance of silicon SPADs. The DCR includes primary and secondary pulses.

The primary dark pulses are due to carriers thermally generated in or near the APD junction. The carriers responsible for primary dark pulses can originate from Shockley-

Read-Hall (SRH) generation via impurities or from band-to-band generation in the depletion region [93]-[94]. Very high electric fields can also significantly increase the DCR, since the carrier emission probability from generation centers is strongly enhanced by field dependent effects, such as the Frenkel-Poole effect and the phonon-assisted tunneling effect [96]. These effects combine to increase the avalanche triggering probability in the depletion region.

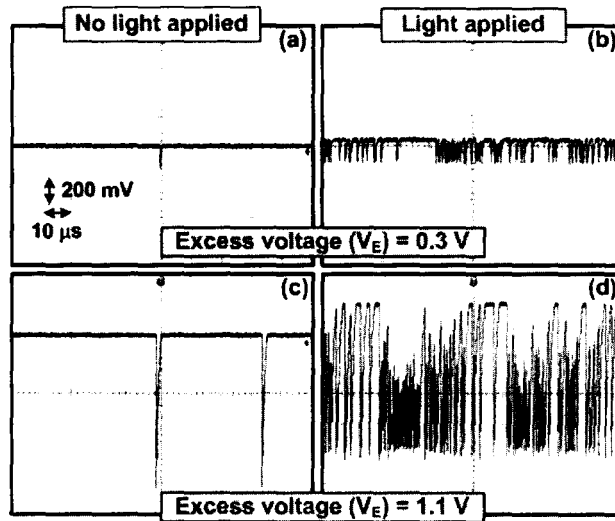


Figure 5.5: Oscilloscope screen capture showing the SPAD operation for two different excess biases with and without light. The figures are shown at the same time and voltage scales of 10 μ s and 200 mV per division, respectively.

Secondary dark pulses are afterpulsing effects which are defined as counts occurring at the end of the dead time that are caused by a previous photon detection event. Afterpulses are the result of secondary avalanches that were triggered by carriers being captured by deep trap levels in the junction depletion layer during the primary avalanche process and subsequently released with a statistically fluctuating delay [97]. Released carriers retrigger the avalanche, generating afterpulses that are correlated with the previous avalanche pulse. In contrast to the primary dark pulses, the secondary dark count rate increases with decreasing temperature, since at lower temperatures, the emission lifetime of a given trap gets longer. Accordingly, the probability that a carrier is emitted after the dead time (thus being able to trigger an avalanche) gets higher. The afterpulse probability is also extremely dependent on the excess bias. When the electric field increases, more carriers cross the depletion region, thus increasing the number of

trapped carriers, as well as increasing the probability of a released carrier triggering an avalanche. Consequently, afterpulsing limits the speed of SPADs, since a long deadtime has to be intentionally introduced to allow the trapped carriers to decay [93]-[94].

The passive quenching SPAD detector of Figure 5.4 operates in two modes, breakdown mode (quenching) and re-charge mode (reset), as shown with the equivalent circuits in Figure 5.6 (a) and Figure 5.6 (b), respectively. The voltage V_C is initially equal to V_R and during the quench mode V_C is given as

$$V_C(t) = RI_{BD}e^{-t/\tau} + V_R - RI_{BD}, \quad (5.2)$$

where I_{BD} is the breakdown current and τ is the circuit time constant given as

$$\tau = R(C_{APD} + C_{Probe}), \quad (5.3)$$

with C_{APD} and C_{Probe} being the capacitances of the APD and the oscilloscope's probe, respectively. During passive reset mode the voltage V_C has an initial value roughly equal to $V_R - V_E$ and is given as

$$V_C(t) = (-V_E)e^{-(t-t_0)/\tau} + V_R, \quad (5.4)$$

and t_0 is the time it takes the passive quenching to exit Geiger mode, which is given by

$$t_0 = -\tau \ln\left(1 - \frac{V_E}{RI_{BD}}\right). \quad (5.5)$$

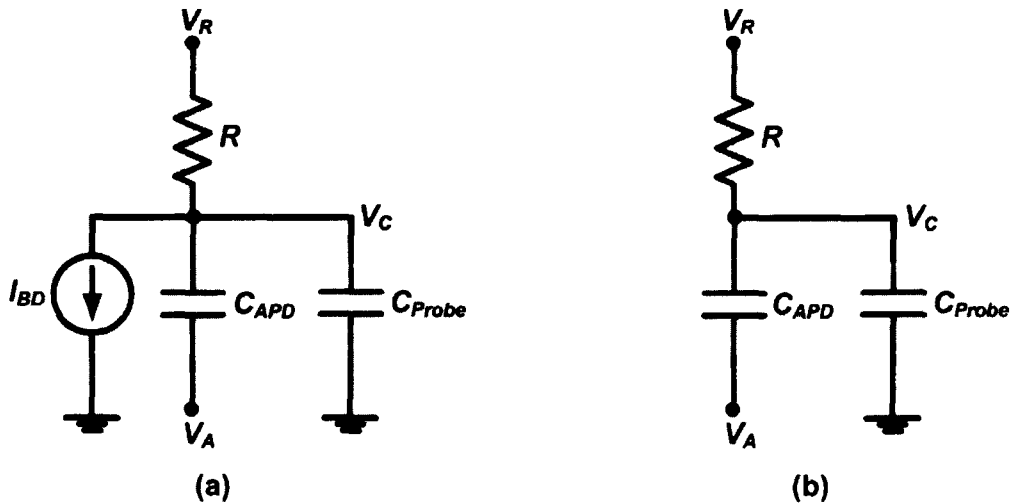


Figure 5.6: (a) Passive quench equivalent circuit in breakdown mode, and (b) passive reset equivalent circuit in charge mode of the circuit shown in Figure 5.4.

Figure 5.7 (a) shows a comparison between the measured and calculated voltage V_C as a function of time with a 2.2 V excess bias. The values used for the calculations are $R=50\text{ k}\Omega$, $C_{APD}=100\text{ fF}$, $C_{Probe}=15\text{ pF}$, $I_{BD}=1.8\text{ mA}$, $V_A=-8.2\text{ V}$ and $V_R=5.15\text{ V}$. Figure 5.7 (b) illustrates the waveforms of V_C for several excess biases shifted to have the same maximum value, that is, the vertical axis in Figure 5.7 (b) is V_C-V_R . The quenching time is hardly affected by the excess bias, because the avalanche process is very fast. However, the reset time depends greatly on the excess bias since it is passive. Note that the large reset time is due to the large capacitance of the oscilloscope's probe. Based on calculations, the actual passive reset time without the probe capacitance is around 50 ns. The peak-to-peak amplitude increases almost linearly with the excess bias, with amplitudes starting from the range of few millivolts up to volts. When designing an active quench and reset circuit, the minimum excess voltage used should be the minimum voltage detectable by the active quench circuit, otherwise the APD will passively quench and reset without generating a pulse in the active circuit. The photon-detection efficiency (PDE) was measured at room temperature at an excess bias of 10 mV, and the normalized values of PDE are shown in Figure 5.8 as a function of wavelength. The peak PDE is achieved at a wavelength of 570 nm.

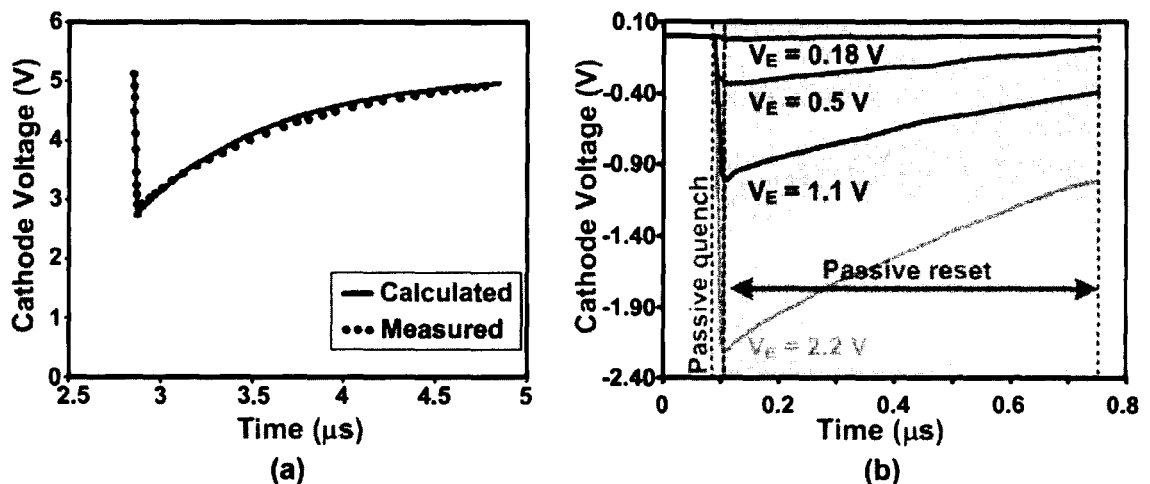


Figure 5.7: Measured cathode voltage as a function of time, showing (a) a comparison to the calculated waveform at an excess bias of 2.2 V and (b) for several excess bias values.

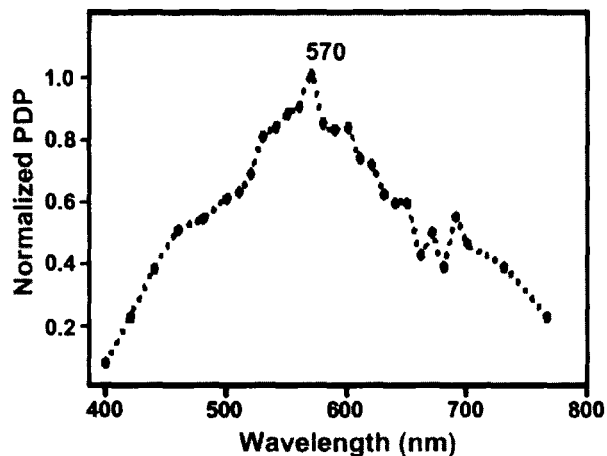


Figure 5.8: Normalized photon-detection efficiency (PDE) as a function of wavelength for an excess bias of 10 mV.

Figure 5.9 (a) shows the DCR and the count-rate with light applied (38 pW, 570 nm) as a function of excess bias. At the lowest excess bias that pulses were generated, the DCR was 8.7 Hz and it goes up to 218 kHz at high excess bias. The elevated excess bias increases the electric field as well as the depletion width, both of which increase the DCR. Although using the minimum excess bias results in a significantly low DCR (25 thousand times less than the high excess bias) and a very high SNR (400 times higher than the high excess bias), shown in Figure 5.9 (c), the PDE at the low bias is also 100 times lower than the PDE of the high excess bias. In addition, the peak-to-peak amplitude of the pulses in V_C at the low bias is only a few millivolts, which makes it very difficult to use these pulses for control of the active quench and reset circuits. A suitable operating range for the excess bias is from 0.2 V to 0.4 V, for which the SNR would be around 40, the DCR would be around 1 kHz, and the peak-to-peak signal amplitude would be greater than 200 mV. At this case, the minimum detectable optical intensity is around 174 nW/cm², assuming that the minimum detectable photon count rate is at least 10 times the DCR with a 2 % PDE. Figure 5.9 (a) also shows that the count rates tend to saturate at increasing excess bias and DCR begins to approach the count rate under illumination. These effects can be related with the high capacitance of the oscilloscope probe, which enlarges the SPAD deadtime at high excess bias, as explained earlier by Figure 5.7 (b). The peak PDE of the device that is obtained at a high bias could not be detected due to this limitation. The saturation was corrected assuming a paralyzable model [95], where the events that occur during the deadtime are not counted and have no effect on the operation of the device, such as prolonging the deadtime [95]. The corrected counts that take the saturation effect into account, were obtained using the following equation [95]

$$n = \frac{m}{1 - m\tau}, \tag{5.6}$$

where n is the corrected count, m is the measured count, and τ is the deadtime. This equation was applied to both the light and dark counts and the corrected counts are shown in Figure 5.9 (b). The figure also shows the calculated PDE as a function of excess bias, where the PDE is around 13 % at an excess bias of 2.2 V and a dark count of 100 kHz.

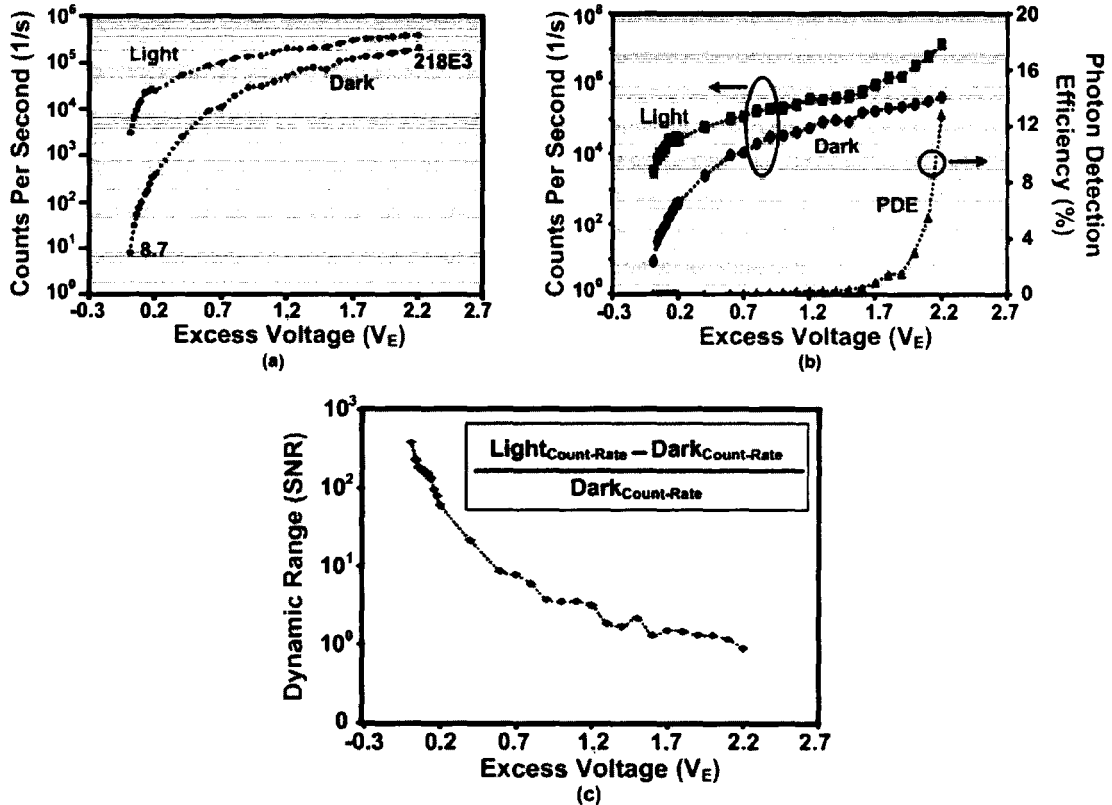


Figure 5.9: Measured dark and light count rates as a function of excess bias, (a) without saturation correction, and (b) with saturation correction. The power of the applied optical signal is 38 pW at a wavelength of 570 nm. (c) The dynamic range or signal-to-noise ratio (SNR) as a function of excess bias.

Table 5.1 shows a comparison to the state-of-the-art SPAD devices fabricated in several deep-submicron CMOS technologies. Note the comparable performances of the SPADs fabricated in these DSM technologies, while CMOS 0.13 μm allows for ultrahigh-speed operation of signal processing circuits.

Table 5.1: Summary of the various DSM SPADs available in the literature.

Ref.	Technology (CMOS)	DCR	Diameter	Deadtime	Breakdown Voltage
[85]	0.35 μm	2 kHz @ 3 V_E	20 μm	40 ns	24.0 V
[87]	0.18 μm	70 kHz @ 0.5 V_E	10 μm	30 ns	10.2 V
[94]	0.13 μm	100 kHz @ 1.7 V_E	10 μm	450 ns	9.4 V
[98]	0.18 μm	10 kHz @ 0.7 V_E	7 μm	---	11.0 V
[99]	0.35 μm	4 kHz @ 4 V_E	20 μm	500 ns	27.5 V
[100]	0.18 μm	100 kHz @ 0.5 V_E	10 μm	---	10.0 V
This work	0.13 μm	1 kHz @ 0.3 V_E	10 μm	50 ns	11.3 V

5.3. Geiger Mode APD with Active Quench and Reset

In order to reduce the deadtime of the SPAD and speed up the quench and reset phase, an active circuit can be used. Figure 5.10 (a) shows an example of an active quench and reset SPAD circuit and Figure 5.10 (b) shows the APD voltage waveform. As soon as a photon arrives and triggers an avalanche breakdown, a passive quenching phase starts, as explained in the previous section. After some delay that depends on how long it takes to reach the threshold voltage that the active quenching circuit can detect, the active quenching will start, and the APD will be discharged at a very high speed. After some delay, the active reset phase will start, which will pull up the APD voltage back to its initial value. The total time required by the circuit to quench and reset the APD corresponds to the deadtime. As previously mentioned, in order to avoid afterpulsing effects, a delay may be required. This delay may be inserted between the active quench and the active reset phases.

Although the active quench and reset circuit reduces the deadtime significantly, it has some drawbacks. The added electronics required for active quench and reset result in increasing the parasitic capacitance on the sense node. Also, the fill-factor of the pixel is significantly reduced and the pixel areas are increased. These problems have prevented actively quenched SPAD pixels from being used in large arrays. Typical fill-factors of

previous work ranged between 1% and 5% with pixels that have an area in the range of $100 \mu\text{m} \times 100 \mu\text{m}$.

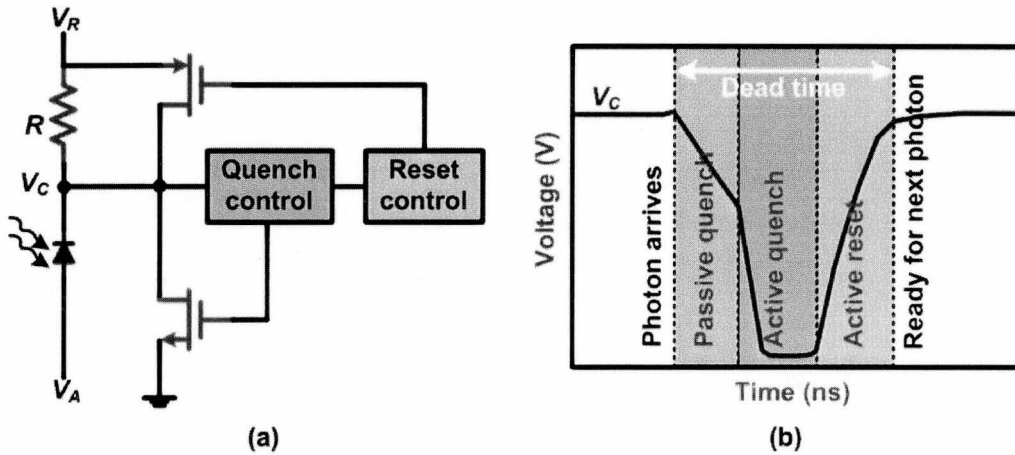


Figure 5.10: (a) Schematic diagram of a Geiger Mode APD with active quench and reset and (b) the APD voltage waveform during photon detection.

In this work, a number of novel ideas were combined in order to obtain a small size, high fill-factor, and high speed SPAD that can be suitable for large array integration. By successfully implementing an APD in a DSM 130 nm CMOS technology, smaller-size devices that can operate at higher speeds are possible. Also, the design was simplified using the minimum number of components possible in order to achieve a small layout area with high speed operation. Figure 5.11 shows the schematic diagram and corresponding pixel layout of the proposed high speed SPAD. The layout has a FF of 25% which is 5-25 times more than previous work, in addition to a smaller pixel size.

The circuit shown Figure 5.11 (a) operates as follows. Initially, the APD voltage (V_C) is held high at the reset value (V_R), which causes the output of the 1st inverter stage to be low, switching off transistor 2. At the same time, the output of the 2nd inverter stage is high, switching off transistor 3 and turning on transistor 1. As soon as a passive quench process is detected by the 1st inverter due to triggering an avalanche process, the output of the 1st inverter becomes a digital high, turning on transistor 2. Since initially transistor 1 was on, the pull down path is now on and the active quench starts. This process will continue until the output of the second inverter changes from high to low, which results in turning off transistor 1 and stopping the active quench, as well as turning on transistor

3 and starting the active reset. The circuit will remain in active reset node until the APD voltage (V_C) becomes high again, driving the output of the second inverter back to high. The time at which the circuit stays in the active quench mode is dependent on the delay of the second inverter stage, which was controlled by carefully selecting the sizes of the transistors in the circuit to achieve a delay that will maintain stable operation.

Since the high-speed SPAD proposed in this work can achieve a very low deadtime, afterpulsing maybe an issue. However, afterpulsing can be reduced or avoided by operating the circuit at a low voltage and using a low APD excess bias voltage. This was possible since the design was implemented in a DSM technology that has a nominal supply voltage of 1.2V. The inverter stages were biased at a supply of 1.5 V, while lower voltages close to 0.9 V could be applied to (V_R). By controlling terminal V_A , the APD can be operated with a low excess bias (as low as 200 mV) to achieve low afterpulsing values. By having a separate reset voltage level for the APD than for the digital inverters in the pixel, and by controlling the excess bias using (V_A), the APD's voltage can be varied to the optimal operating point required for a specific application. Using a low excess bias of 300 mV allows for a low afterpulsing probability due to reducing the avalanche current, as well as a low dark-count-rate of about 400 Hz (Figure 5.9 (a)), at the expense of a reduction in photon-detection-efficiency (PDE).

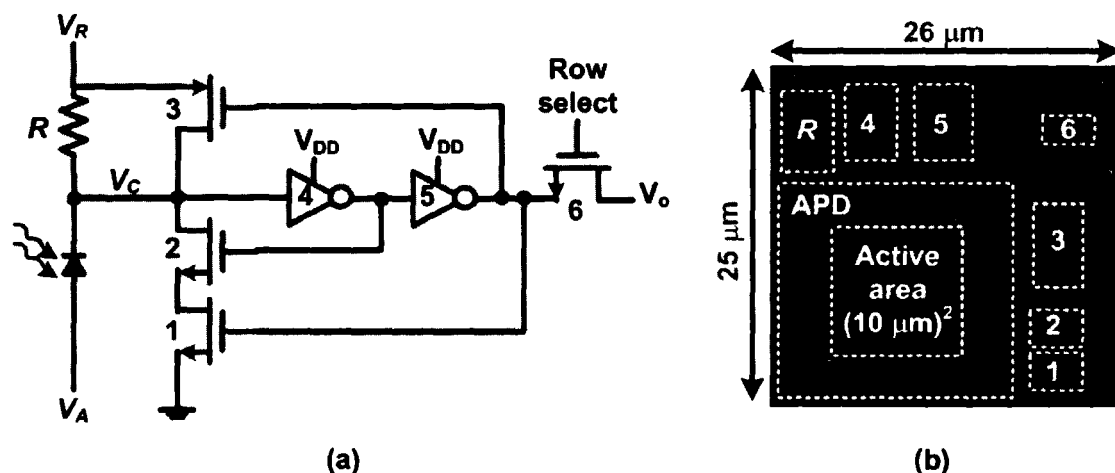


Figure 5.11: (a) Schematic diagram of a the proposed Geiger Mode APD with active quench and reset and (b) the layout screen capture in 130 nm CMOS technology with a 25% fill-factor.

In order to accurately simulate and design the circuit proposed in Figure 5.11, the Geiger mode operation of the APD was modeled into the SPICE circuit simulator based

on the measurement results of the passive quench and reset APD, as shown in Figure 5.12. The clock source V_{ph} simulates the incoming photon that activates transistor M_4 , allowing for a constant current discharge of the pre-charged APD's capacitance (C_d). The switch, together with the inductor (enclosed in dashed lines), are a relay that switches on and off depending on the voltage V_C .

As in the passive SPAD, breakdown is triggered by an incoming photon or thermally generated electron-hole pairs, and it continues until the voltage across the APD drops below the breakdown voltage. This voltage is almost equal to the excess bias (V_E). The settings of the relay are set to open the switch when $V_C = V_R - V_E$. The switch is closed during the reset phase and stays closed until the next photon arrives and causes a discharge. The simulation results of the low deadtime SPAD pixel are shown in Figure 5.13. As can be seen from the figure, the APD passive quench behavior has been modeled, which was extracted from measurements, and the pixel achieves an active quench and reset in a very low dead time of 250-300 ps, which is 20 to 100 times lower than previous work, as shown in Table 5.2. Since the pixel provides very high speed digital pulses, care was taken to ensure that the row select transistor and the second inverter stage can drive the column load capacitance.

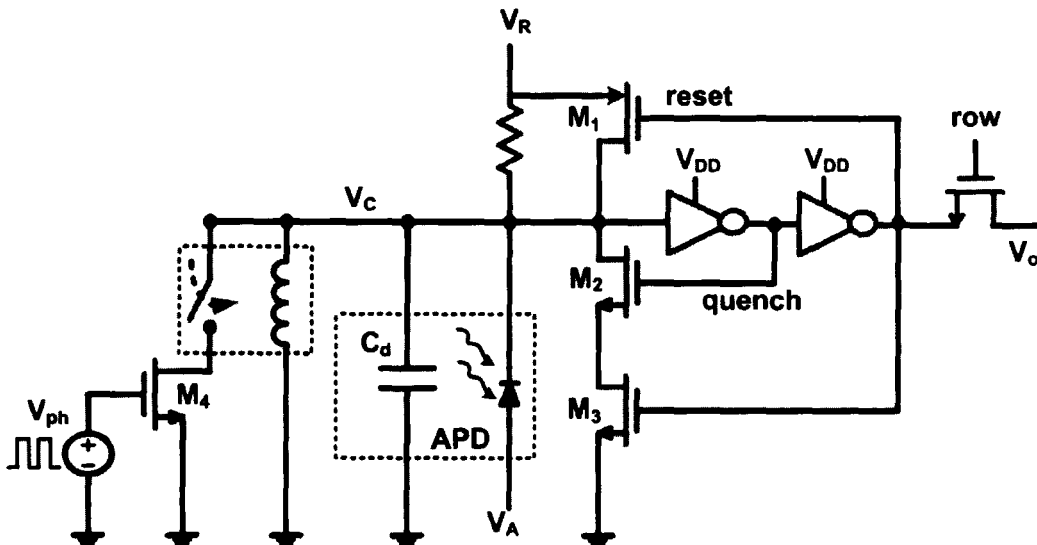


Figure 5.12: Active quench and reset SPAD circuit showing the SPICE simulation equivalent circuit model.

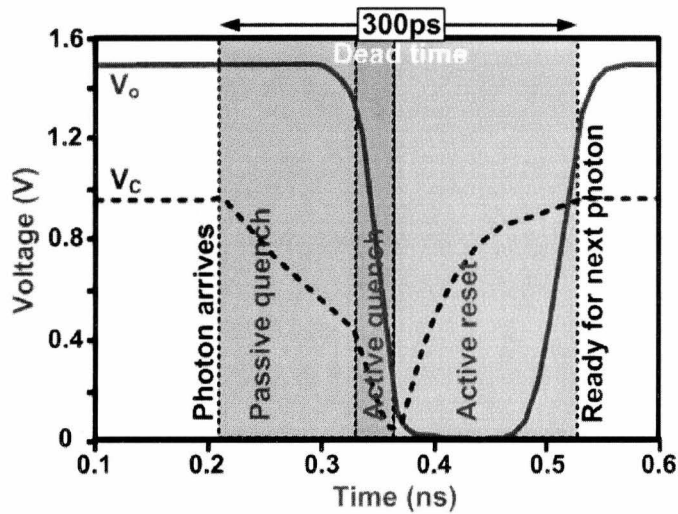


Figure 5.13: Simulation results of the active quench and reset SPAD circuit show in Figure 5.12.

A prototype imager was implemented, as shown in Figure 5.14, in a standard 130 nm CMOS technology from IBM. The microchip measures an area of 2 mm × 2 mm with an array of 2304 pixels. The pixels are multiplexed to a 10-bit high-speed on-chip counter that is in charge of obtaining the event counts over a period of time that is controlled by the FPGA. Having an on-chip counter was necessary since the digital pulses have a very short duration and will be severely distorted by the parasitics of the package.

Table 5.2: Summary of the various DSM SPADs with active quench and reset available in the literature.

Ref.	Technology	Dead Time	Count rate	Pixel Area
[84]	0.8 μm HV CMOS	34 ns	30 Mcps	n/a
[85]	0.35 μm HV CMOS	40 ns	20 Mcps	(100 μm × 50 μm)
[87]	0.18 μm CMOS	30 ns	30 Mcps	(120 μm × 80 μm)
[101]	0.8 μm HV CMOS	30 ns	30 Mcps	n/a
[102]	0.8 μm HV CMOS	50 ns	20 Mcps	SPAD off-chip
[103]	0.8 μm HV CMOS	30 ns	40 Mcps	SPAD off-chip
[104]	1 μm HV CMOS	36 ns	25 Mcps	SPAD off-chip
[105]	0.18 μm CMOS	5 ns	200 Mcps	(180 μm × 96 μm)
[106]	0.8 μm HV CMOS	< 10 ns	100 Mcps	n/a
[107]	0.7 μm HV CMOS	32 ns	30 Mcps	n/a
This work	0.13 μm CMOS	250 ps	4 Gcps	(26 μm × 25 μm)

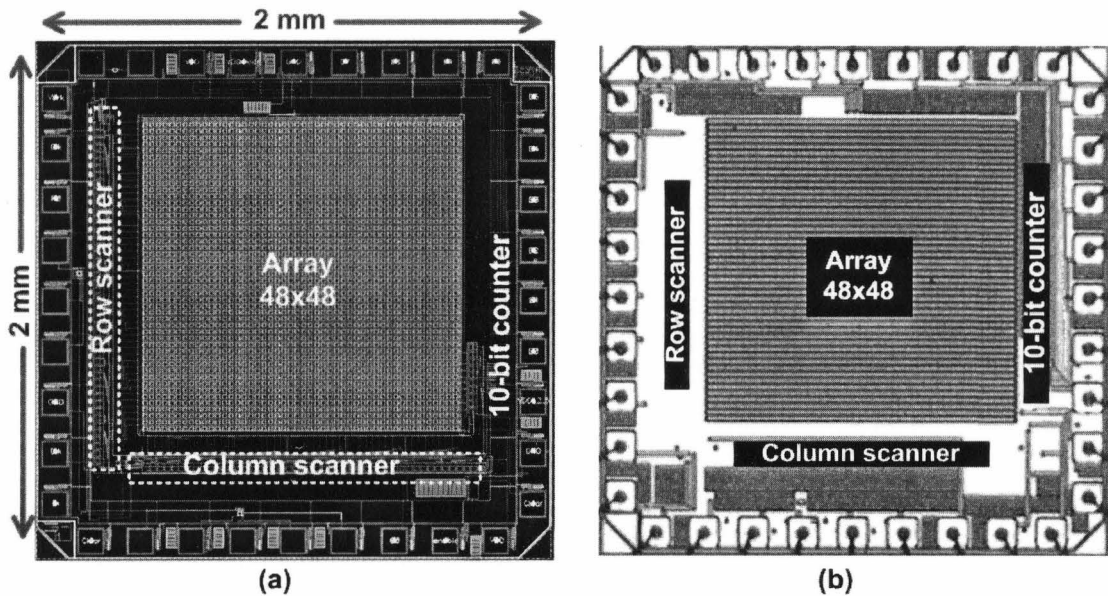


Figure 5.14: Active quench SPAD imager (a) layout screen capture and (b) chip photomicrograph.

5.4. Deadtime Reduction Technique

The deadtime reduction technique mentioned in the previous section suffers from a trade-off with sensitivity. This is because the circuit must be operated at a low excess bias to avoid afterpulsing, which reduces the PDE. This section proposes a novel deadtime reduction technique that trades off fill-factor rather than PDE. In order to reduce the deadtime without sacrificing PDE, the technique shown in Figure 5.15 can be used. By using two or more APDs per pixel, the deadtime will be reduced by two or more, respectively. Figure 5.15 (a) shows the layout cross section of the multiple APDs that share a common well. This allows for maintaining almost the same fill-factor with a slight reduction due to adding some active devices. Counting of both APDs is done simultaneously, using a common counter and a logic decision circuit that either increments the counter by one or two, as shown in Figure 5.15 (b). Using multiple APDs per-pixel can allow for a significant reduction in deadtime at almost no expense, other than a slight reduction in FF, as long as the counter is outside of the pixel since the logic complexity increases exponentially with the number of APDs per pixel. If in-pixel counting is needed for high speed simultaneous array counting, then a novel compact solution is required, which is explained as part of the design shown in Chapter 6.

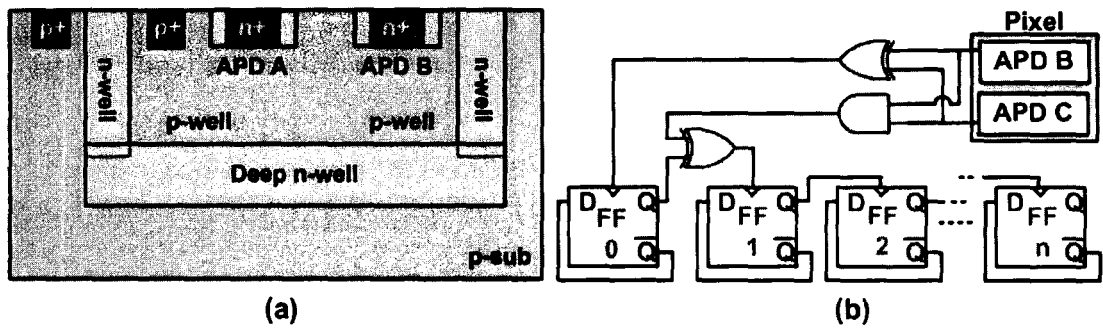


Figure 5.15: (a) The layout cross-sectional view of the multiple APDs, and (b) the modification to the counter for simultaneous counting.

Chapter 6

TIME-DOMAIN SINGLE-PHOTON IMAGER

As explained in Chapter 5, avalanche photodiodes used in Geiger mode as single-photon counters have become very attractive imaging tools. Single-photon imaging can be used in very low-light level applications such as night shots, military and surveillance imagers, quantum computing, and biomedical imaging such as chemiluminescence, autofluorescence, and fluorescence lifetime imaging. However, the count-rate of the single-photon detector has a lower limit at the dark-count-rate and an upper limit at the deadtime, where any incoming photons at rates out of this range cannot be detected. This was improved in the previous chapter using a low deadtime SPAD design. Furthermore, a typical avalanche-based single-photon detector cannot offer the high dynamic range that is needed for many biomedical and surveillance applications. In this chapter, a novel imager is presented that includes the low deadtime single-photon detector explained in the previous chapter, in a novel time-domain single-photon imager (TDSPI) using mainstream deep-submicron CMOS technology. The imager offers high dynamic range and sensitivity, while maintaining high-speed operation and is of low cost. This chapter starts with an introduction to time-domain imaging followed by how time-domain imaging can be applied to single-photon counters. The proposed pixel design is discussed next, after which, a novel in-pixel counting technique is shown in order to simplify the pixel design. Finally, the TDSPI implementation is shown.

6.1. Introduction

Time-domain imaging using an active-pixel sensor (APS) is a typical technique used to enhance the dynamic range of a detector [108], [109]. In time-domain imaging, the output voltage of the APS is compared to a reference voltage and the time required for the

output voltage of the APS to drop below the reference (threshold) voltage is recorded. A strong optical signal will have a shorter time than a weak one, and the dynamic range is the ratio of maximum time to the minimum time. However, the main drawback of time-domain imagers is the trade-off between frame-rate and dynamic range. Figure 6.1 (a) shows an example of the calculated outputs of an APS for four different optical powers, corresponding to photocurrents of 1 nA, 1 pA, 100 fA, and 10 fA, which give a dynamic range of 100 dB. Neglecting the noise sources, the following relationship, which was previously presented in Chapter 2, describes the waveform of the photodiode voltage v_d of a typical APS

$$v_d(t) = V_{DD} - \left(\frac{i_{ph} + i_{dark}}{C_d} \right) t, \tag{6.1}$$

where V_{DD} is the supply voltage, i_{ph} is the generated photocurrent, i_{dark} is the dark current of the photodiode, t is the pixel integration time and C_d is the capacitance of the photodiode.

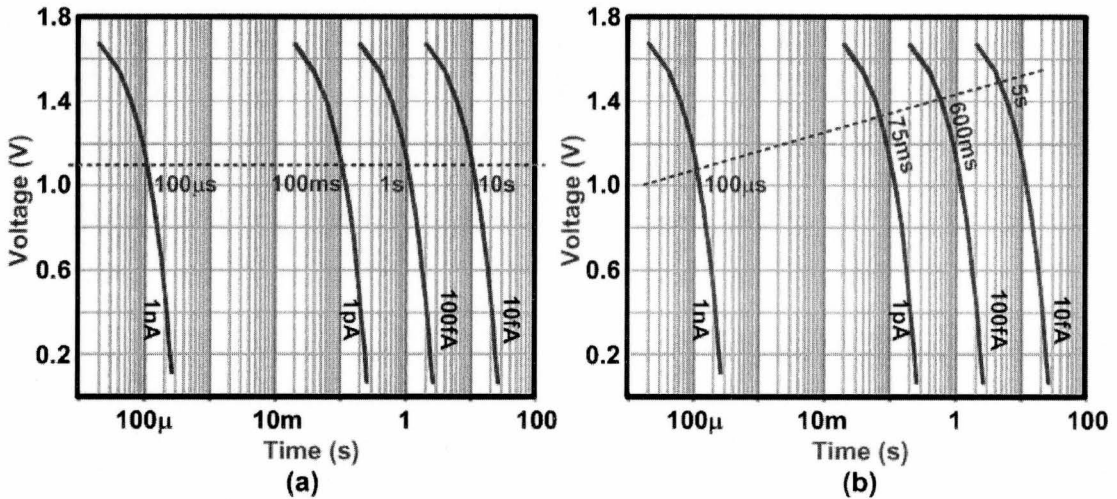


Figure 6.1: Calculated APS output voltage for four different optical signals showing the generated photocurrents and the corresponding times required to drop below the (a) constant threshold voltage, and (b) ramp threshold voltage. The threshold voltages are shown as the dashed lines.

Assuming a V_{DD} of 1.8 V and a C_d of 150 fF, the pixel conversion rate ($1/t$), with a constant reference voltage of half the supply, would depend on the integration time t required for the weakest light signal, giving a rate that is $1/t = 1/13.5s = 0.074$ pixel/s for a dynamic range of 100 dB. If a variable reference voltage is used, such as a ramp

threshold voltage, for weaker signals, the pixel rate can almost be doubled, as shown in Figure 6.1 (b). However, even with the doubled pixel rate, this technique seems inadequate to high speed applications.

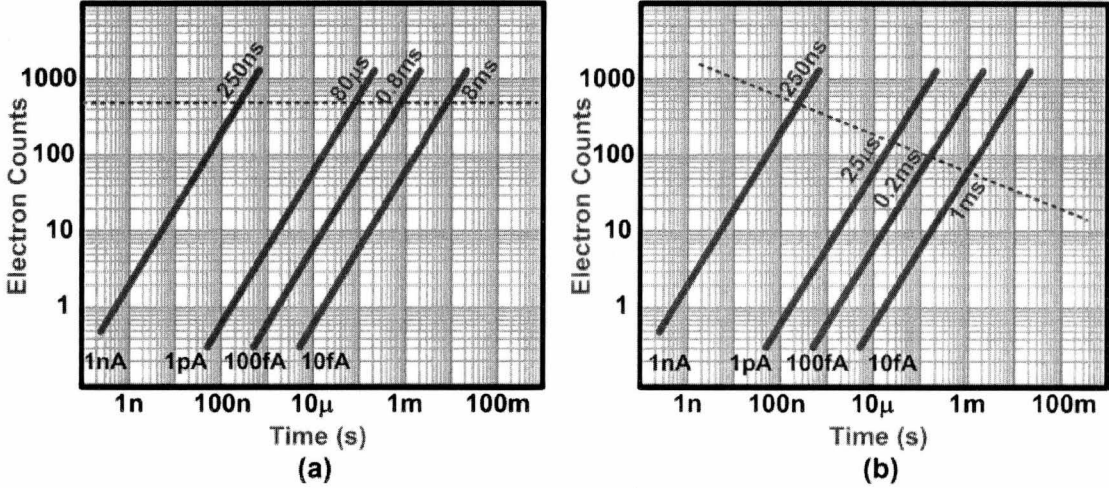


Figure 6.2: The electron count equivalent of the generated photocurrents from Figure 6.1 and the corresponding times required to drop below the (a) constant threshold count, and (b) a variable threshold count. The threshold counts are shown as the dashed lines.

6.2. Time-Domain Single-Photon Imaging

Figure 6.2 shows the electron count equivalent of Figure 6.1, which can be calculated using

$$n(t) = \left(\frac{i_{ph} + i_{dark}}{e} \right) t, \quad (6.2)$$

where n is the electron count during the integration time t , and e is the electron charge. Such electron counts can be generated using single photon counting Geiger mode avalanche photodiodes (APDs). When using a constant count threshold of 512, and where the pixel rate is $1/8\text{ms} = 125$ pixels/s for the same dynamic range of 100 dB. This shows a speed improvement of three to four orders of magnitude for both the constant and variable threshold methods. The conversion for the electron count of the 1 nA photocurrent was limited by the 250 ps deadtime of the SPAD. For the 1 pA photocurrent shown in Figure 6.2 (a), 82 µs are required to reach a threshold count of 512. This can be converted back to Figure 6.1 (a) using equation (6.1) for a current of 1 pA and an integration time of 82 µs, to get a voltage drop of 547 µV. This voltage is below the

detection limit of a simple in-pixel circuit. On the other hand, counting 512 events in a single photon counter, which counts pulses in digital format, is quite straight forward. This shows how the time-domain single-photon imaging (TDSPI) technique utilizes the high sensitivity and speed of photon or electron counting methods to achieve high dynamic range, while maintaining a reasonable frame-rate that is at least 3 orders of magnitude higher than the conventional high dynamic range techniques.

6.3. Pixel Design

The pixel rate relates to the frame rate based on the number of pixels and the readout time. The most basic implementation of the TDSPI would be to use an array of single-photon detectors that are connected through multiplexing to one common counter for the entire array, as presented in the previous chapter in Figure 5.14. In this case, rather than counting for a fixed integration time, the counting will continue until a selected threshold is reached. Using this technique however, based on the calculations shown in the previous section, with a 100 dB dynamic range and a 125 pixel/s pixel rate, would mean that for an array of 1000 pixels, 8 seconds are needed for array conversion, which is a very long time for such a small array. Therefore, in order to maintain a high frame rate, simultaneous in-pixel counting and threshold detection are needed

Figure 6.3 shows a block diagram of the time-domain single photon counting (TDSPC) pixel that allows for simultaneous pixel counting and conversion for high speed operation. Each pixel contains a low deadtime active quench and reset SPAD, an in-pixel counter for threshold detection, a threshold selection multiplexer and an SRAM. The time, from the output of 9-bit counter, which is connected to all pixels, is stored in the pixel SRAM once the threshold count is reached. The timer is clocked with a fixed rate from an external clock provided by the FPGA that interfaces to the imager for control and data processing. As the count of the timer increases, the threshold selection logic will make a decision to reduce the count threshold by controlling the threshold multiplexer. Initially, the threshold multiplexer monitors the most significant bit of the counter and as the count threshold is reduced, the multiplexer moves from the most significant bit to the bit before, and so on.

The block diagram shown in Figure 6.3 is simplified and does not show details such as logic elements that control the counter and stop the count once the threshold is reached. Using this pixel, a frame is completed once the maximum time is reached, or once all pixels are done, which can be obtained by ANDing the write enable signals from all pixels in the array. After the frame conversion is complete, the readout of the SRAM values can start using conventional array access and SRAM array pre-charge.

Due to the complexity of the pixel shown in Figure 6.3, this design would require more than 200 transistors per pixel, mainly due to the use of in-pixel counters. Also, if a deadtime reduction technique using multiple APDs per pixel technique were to be used, the in-pixel counter design would be even more complex. This makes the TDSPC pixel too complex to implement using current mainstream technologies. Perhaps, using a specialized imaging technology kit could help implement APDs in smaller scale technology nodes so that the TDSPC pixel can be implemented with a reasonable fill-factor. The following section proposes a novel technique to simplify the TDSPC pixel design and make the implementation more feasible for current mainstream technology nodes.

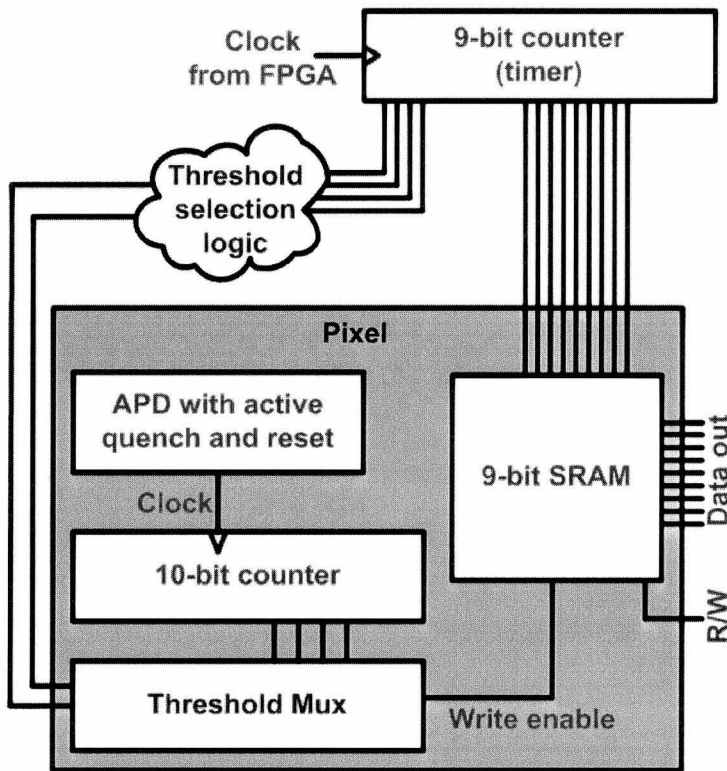


Figure 6.3: Block diagram of the TDSPC pixel.

6.4. Analog Counter

Since the in-pixel counter and threshold detector are the components with the most number of transistors, a novel in-pixel counting and threshold detection technique was proposed in this work in order to reduce the number of transistors, which is shown in Figure 6.4 (a). The digital pulses from the APD, which have a fixed width equal to the deadtime of the active quench and reset SPAD circuit, control a switched current source that charges a capacitor (C_H). Figure 6.4 (b) shows the waveform operation of the analog counter. With every applied SPAD pulse, the capacitor voltage increases according to

$$V_c(n) = \frac{nI\tau_{SPAD}}{C_H} + V_0, \quad (6.3)$$

where V_c is the voltage across the capacitor C_H and is a function of n , I is the current of the pulse-controlled current source, τ_{SPAD} is the deadtime of the SPAD circuit, and V_0 is the initial voltage of the capacitor prior to counting. Then, V_c is compared to a reference voltage V_{Ref} that acts as the count threshold. The comparison is performed using a high speed comparator. Once the capacitor voltage exceeds the threshold voltage, the Done signal goes low, clearing the capacitor voltage to V_0 and restarting the count.

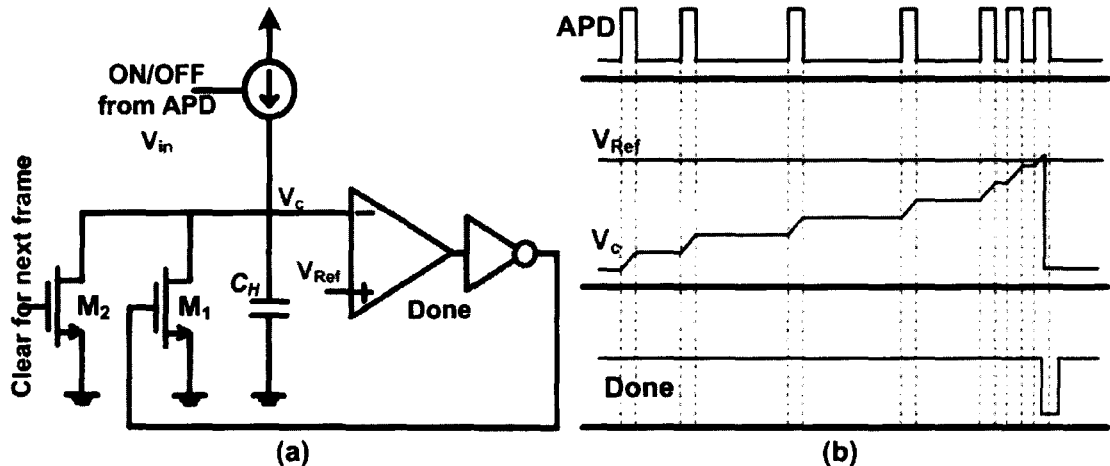


Figure 6.4: (a) Schematic diagram of the designed analog counter, and (b) waveform operation of the analog counter.

The deadtime reduction technique, previously shown in Figure 5.15, can be implemented using the proposed analog counter, as shown in Figure 6.5. With every

additional SPAD, the only device increase required is an additional switched current source, which requires only four transistors, compared to 38 transistors for the logic increase of the digital counter. In addition, the increase in devices is a linear function of the number of SPADs used, rather than an exponential function as in the digital counter case. This makes the deadtime reduction technique using multiple detectors very attractive for use with an in-pixel counter, however, if the counter was outside of the pixel, the benefit in saving area will not be significant.

Although in theory, an analog counter can provide an infinite count based on equation (6.3), using a small voltage step size would require a very accurate comparator that is not easy to design for high-speed operation with a small area in the pixel. To avoid this issue, the count value can be limited to a range of 10 to 30 and two counters can be cascaded to obtain a count that is the multiplication of both stages. In this way, the accuracy is preserved by a reasonable trade-off only with pixel area.

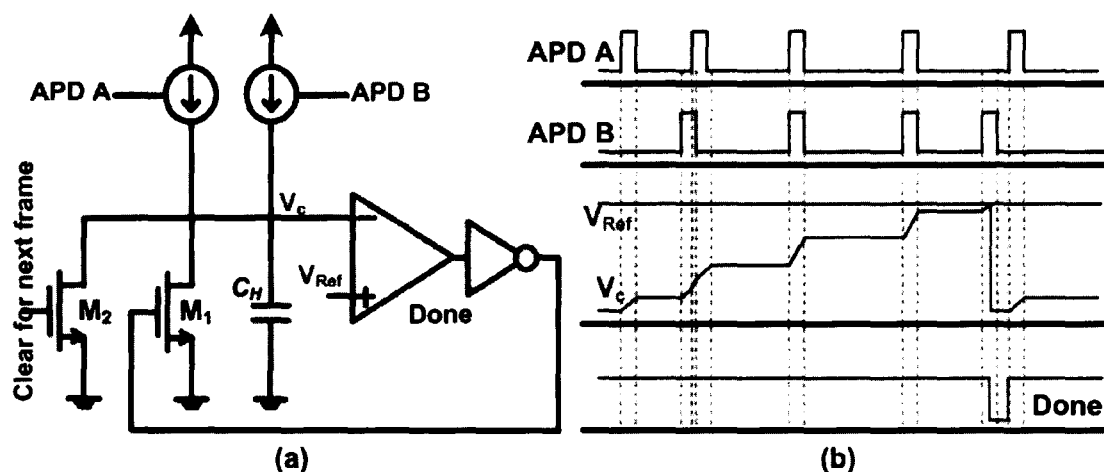


Figure 6.5: (a) Analog counter using multiple SPADs, and (b) the waveform operation of the multiple SPAD analog counter.

Although the analog counter was designed to be used for in-pixel counting in the time-domain single-photon counting technique proposed here, it can also be used for simultaneous pixel counting in a standard single-photon counting array. Figure 6.6 shows an example of the proposed novel pixel. The pixel operates as follows. First the analog counters are cleared, after which, they start counting the pulses that are coming from the SPAD for a fixed duration of time that can be controlled by the FPGA or an on-chip timer. Having an in-pixel counter allowed for integrating the pixels in the entire array

simultaneously. Simultaneous pixel counting is important in order to capture time correlated events between the pixels of an image. After the integration is over, the analog count needs to be converted to a digital value. And to avoid capacitor leakage and data loss, the ADC conversion process is done in parallel for all pixels. This is done using a ramp integrating ADC circuit, which was explained previously in Chapter 2 and Chapter 3. An out-of-pixel digital counter will count the time required for an analog ramp generated signal to reach the value of the in-pixel analog counter. The comparison between the ramp generated value and the analog count is done using an in-pixel opamp comparator. The novel pixel shown in Figure 6.6 converts from digital to analog and back to digital again. It can be argued that using a digital counter in-pixel would be a much easier approach. However, the in-pixel digital counter will have a much lower fill-factor and a much higher power consumption.

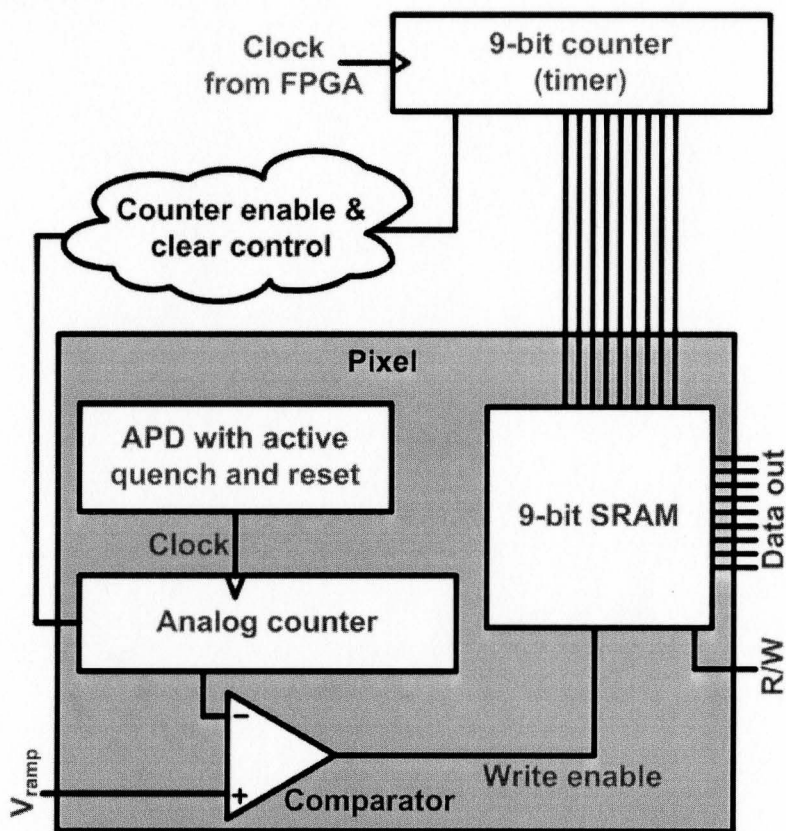


Figure 6.6: Block diagram of a SPAD pixel that can achieve simultaneous pixel counting as well as simultaneous pixel analog-to-digital conversion.

6.5. SPTD Imager

The imager was implemented based on the block diagram shown previously in Figure 6.3 using the analog counter. The following subsections discuss the design of the pixel and the array.

6.5.1. Analog counter design

Figure 6.7 shows the complete two-stage analog counter schematic diagram. The ideal current sources that were previously shown in Figure 6.4 (a) have been implemented as current mirrors. Transistors M_4 , M_6 and M_3 form the first current mirror that mirrors the current between transistors M_4 and M_6 to the branch between M_3 and the capacitor with a current multiplication that is equal to the ratio of the size of transistor M_3 to the size of transistor M_4 . Similarly, transistors M_8 , M_{10} and M_7 form the second current mirror. The current mirrors are switched on once the pull down transistors (M_6 , M_{10}) are on and transistors M_5 and M_9 are off. Transistors M_5 and M_9 were added to ensure complete shutdown of transistors M_3 and M_7 .

The first counter counts the pulses coming from the SPAD, which have a pulse width τ_{SPAD} that is equal to the SPAD deadtime, resetting itself once the threshold V_{Ref1} is reached. The second counter counts the reset pulses of the first counter until the threshold voltage (V_{Ref2}) is reached and, the write enable signal is generated to the SRAM to latch the time from the global digital timer. The pulse width that clocks the second counter is slightly longer than the SPAD deadtime due to the limited bandwidth of the opamp in the first analog counter. For this reason, a lower current ratio was used in the current mirror of the second counter. The count threshold of each counter can be reduced or adjusted dynamically by varying the reference voltages of the comparators. At reference voltages of 1.5 V applied to both comparators, the first counter has a count of 21, while the second counter has a count of 26, which gives a total count of 546 for the two-stage in-pixel analog counter. The capacitors C_H were designed using a MOS capacitor, since it offers a small layout size of $4 \mu\text{m} \times 3 \mu\text{m}$ with a capacitance of around 50 fF. The MOS capacitor had a thick-oxide in order to reduce its leakage.

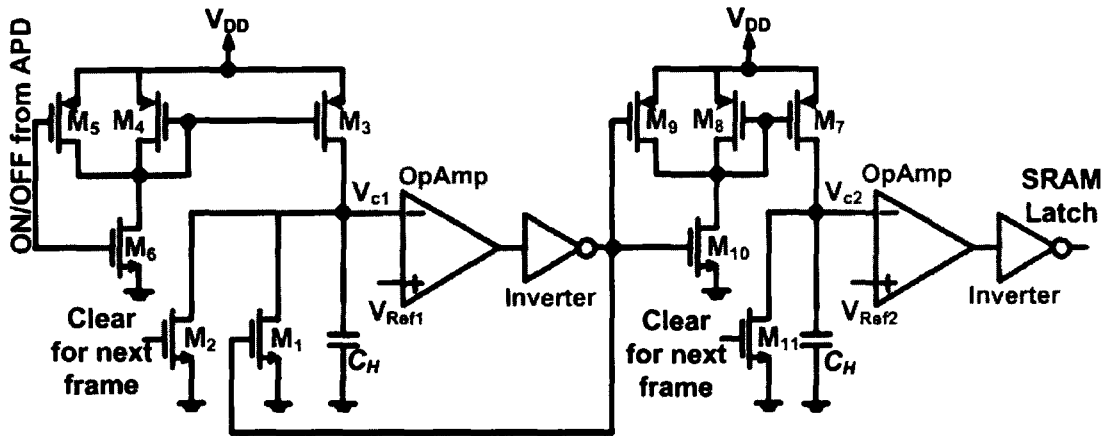


Figure 6.7: Schematic diagram of the cascaded analog counter.

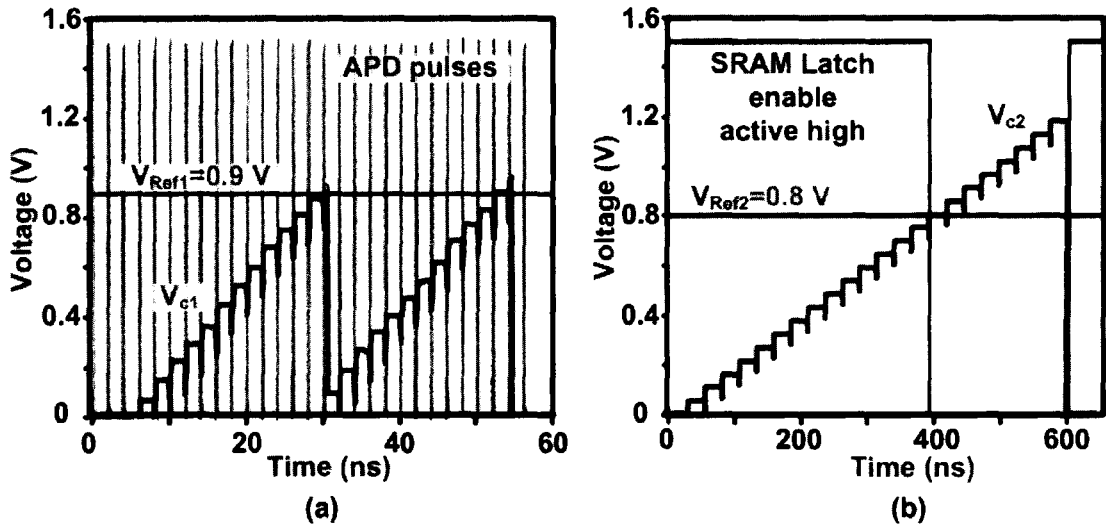


Figure 6.8: The Cadence simulation results of (a) the first counter, and (b) the second counter.

The SPICE simulation results of the first and second analog counters are shown in Figure 6.8 (a) and (b), respectively. For this circuit, SPICE models and simulations are reliable since the circuit operates at frequencies that are much lower than the cut-off frequency of the transistors used and no major parasitic effects would be present. With the selected reference voltages shown in the figure, the first counter has a count of 12 and the second has a count of 15, for a total count of 180. The circuit was designed to keep the writing in the SRAM enabled until the count threshold is reached, after which, the writing is disabled, thus, the counters can continue operating. In this way, the time is continuously being written to the SRAM until the threshold is reached, when the last time

before the threshold will be saved. The second counter will continue to count or even saturate without consequences. Both counters are cleared at the start of the acquisition of a new frame by a global signal that turns on transistors M_2 and M_{11} for a short time in the range of 10 ns.

A high-speed, small-size operational amplifier was designed to be used as an open loop comparator for the analog counter. Usually, in an opamp, capacitive compensation is needed to ensure that the amplifier is stable with no oscillations. Since the compensation capacitor is usually one of the largest components of the opamp, it was removed to reduce the layout area of the circuit. This was possible since the amplifier will be used in open loop. Figure 6.9 shows the designed high-speed comparator that consists of a differential stage (1st stage), an amplifying stage (2nd stage) and an output amplifier (3rd stage). In order to maintain high-speed operation, the gain was obtained from three stages rather than one or two, where each stage provides a small amount of gain in order to maintain the bandwidth.

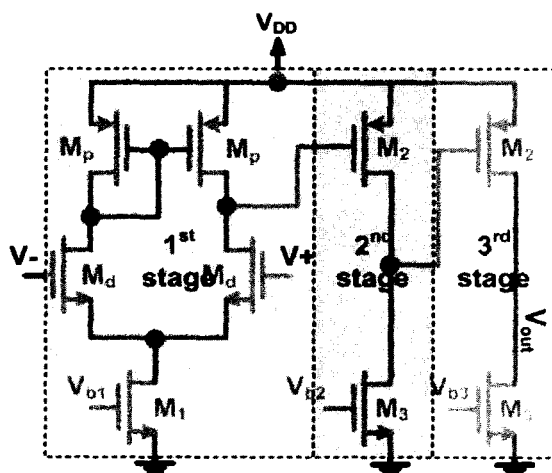


Figure 6.9: The schematic diagram of the high-speed comparator.

The simulation results of the high-speed comparator are shown in Figure 6.10. With a small signal differential input of 1 mV, the opamp provides a gain of 10 from the differential stage. The second stage provides a gain of 3 and the third stage provides a gain of almost 2. The total gain of the amplifier is around 60 with a bandwidth of 520 MHz. The low gain will result in a common-mode error, however, the high bandwidth is necessary for the analog counter design, since the deadtime of the SPAD is very low. Figure 6.10 (b) shows the large signal simulation of the amplifier's response to a fast

rising input pulse. The output has a rise and fall time that is less than 500 ps. The layout of the opamp measures an area that is less than $15 \mu\text{m} \times 15 \mu\text{m}$.

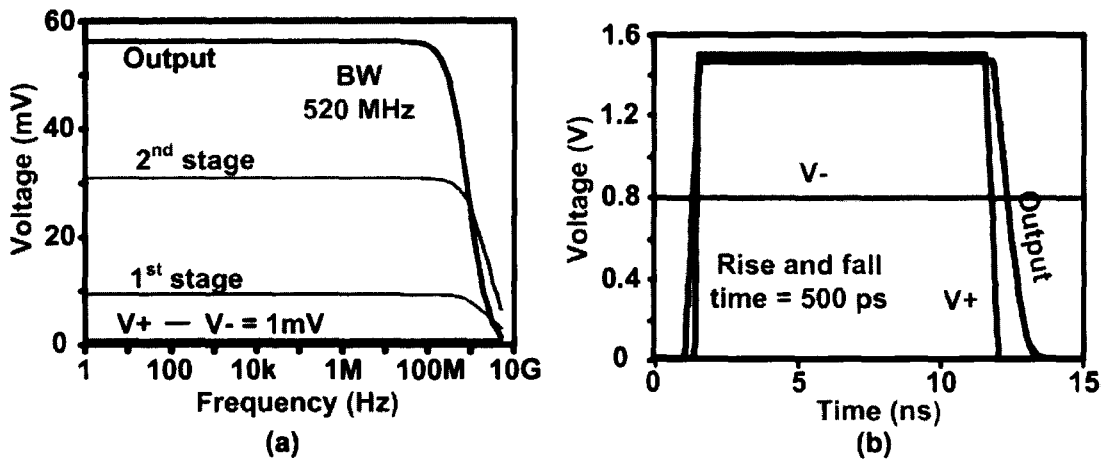


Figure 6.10: Cadence simulation results of the high-speed comparator. (a) Frequency response showing the gain and bandwidth, and (b) time domain rise and fall time simulations.

6.5.2. SRAM design

Figure 6.11 shows the schematic diagram of a standard six transistor static random access memory cell (6T-SRAM). The cell consists of two cross coupled inverters that act as a latch and two access transistors. When the write signal is high, it turns on the access transistors applying the data and the data bar signals to the memory latch. Once the write signal goes low, the data is latched for as long as power is applied to the circuit. Although this topology is quite simple, it requires both the data and its complement, which would add complexity to the pixel design.

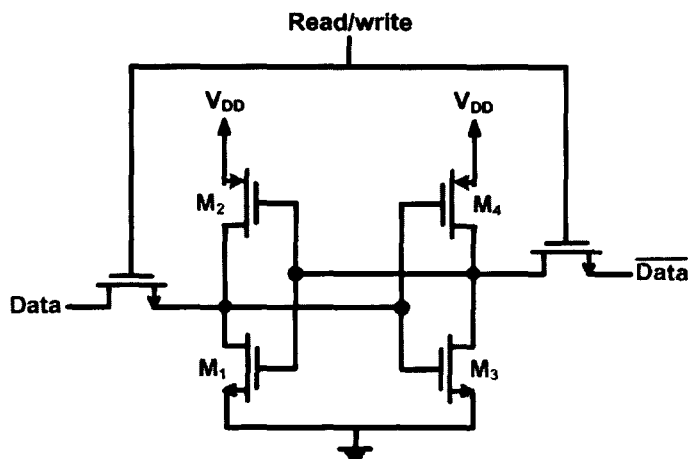


Figure 6.11: Schematic of a standard 6T-SRAM memory cell.

The SRAM topology used in this work was a dual-port 6T-SRAM, which is shown in Figure 6.12 [110]. This topology has the advantage of offering simultaneous read and write access to the cell, in addition to requiring only the data and not the complement of the data as well. When writing, if a digital high is applied to the input, it would turn off transistor M_4 and turn on transistor M_3 , which will turn off transistors M_2 and M_6 . When reading, the output line needs to be pre-charged to a digital high, and since transistor M_6 is turned off, the output line will remain high, indicating a digital high was latched. If a digital low was written into the cell, transistor M_3 will be turned off, while transistor M_4 will be turned on, which will turn on transistor M_6 that will drive the output line to a digital low. This topology was found most suitable for the in-pixel implementation.

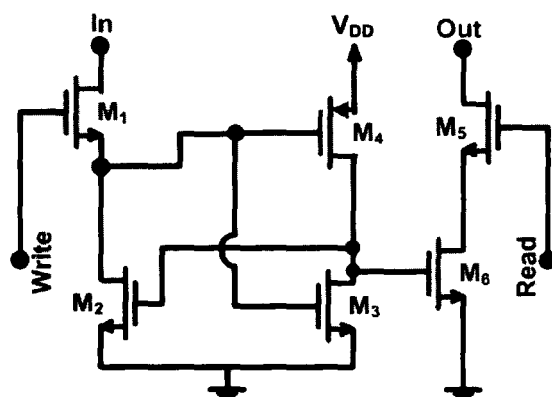


Figure 6.12: Schematic of the dual-port 6T-SRAM [110].

6.5.3. Array design

Figure 6.13 shows the layout of a time-domain single-photon counting (TDSPC) pixel that was implemented using a mainstream 130 nm CMOS process from IBM. The pixel occupies an area of $50 \mu\text{m} \times 50 \mu\text{m}$ and has a fill-factor of 4%. The pixel contains all the components necessary for the implementation of a TDSPC imager that can achieve simultaneous counting of all pixels in parallel. The layout of the opamp measures an area that is less than $15 \mu\text{m} \times 15 \mu\text{m}$ and the layout of each analog counter measures an area that is less than $15 \mu\text{m} \times 30 \mu\text{m}$. The area occupied by this pixel design is less than most state-of-the-art SPAD designs that were shown in Chapter 5, although the pixel contains counters, whereas the previous designs only contain active quench and reset circuitry. This pixel area reduction was mainly due to the compact analog counter and SPAD

designs, in addition to using a deep-submicron technology. Potential issues with low (4%) fill-factor can be addressed by using microlenslet arrays or fiber couplers. Initially, the SPAD alone had a FF of 25%, which was reduced to 4% in the TDSPC pixel. The effect of FF reduction can be calculated in terms of reduction in resolution. The resolution of an imager is obtained from the maximum modulation transfer function (MTF_{max}), which is equal to $1/(2p)$, where p is the pixel pitch. For a square pixel, the pitch and FF can be related by

$$p = \sqrt{AA/FF}, \quad (6.4)$$

where AA is the active area. Finally, the imager resolution can be related to the FF by

$$MTF_{max} = \sqrt{FF/4AA}. \quad (6.5)$$

The reduction in FF causes a square root reduction in resolution, meaning that reducing the FF of the pixel from 25% to 4%, which is a factor of 6.25, will result in a drop of 2.5 times in resolution.

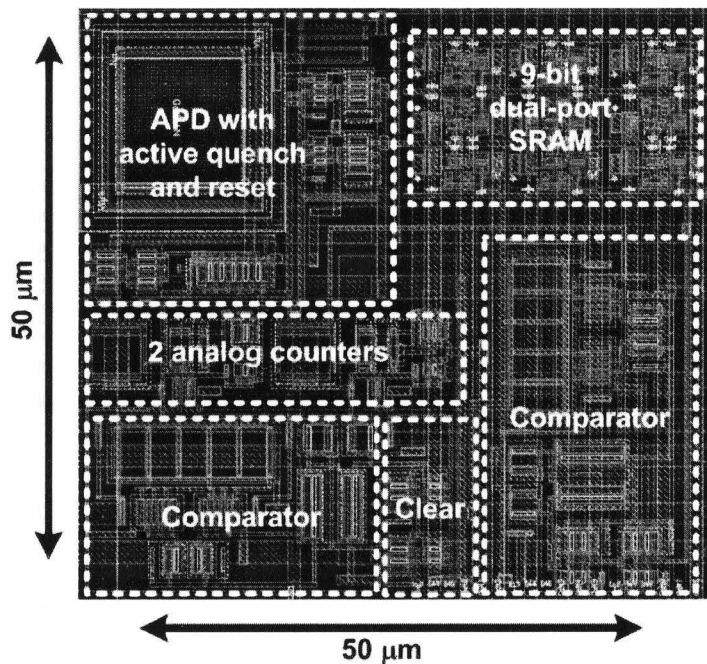


Figure 6.13: Layout of a TDSPC pixel with in-pixel analog counting and SRAM.

The pixel was used in a TDSPI array of 24×16 pixels, shown in Figure 6.14, in a complete camera-on-a-chip. The imager uses a single timer that is common to all pixels and row and column select circuitry, in addition to the SRAM pre-charge circuitry. When

a row is selected, the SRAM outputs are placed on a 9-bit column bus, that is then multiplexed by column selection onto a 9-bit row bus. The final output of the array is a 9-bit value, which is the value stored in the SRAM of the selected pixel.

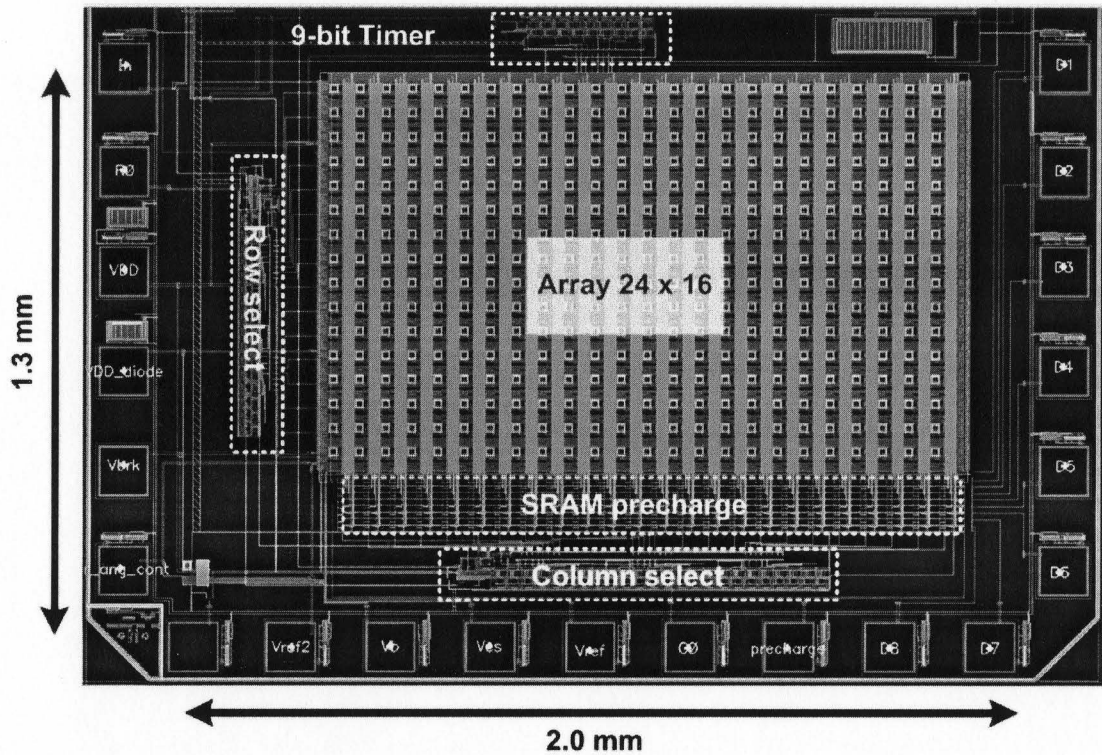


Figure 6.14: The layout of the 24x16 pixel TDSPI with timer, readout and array access circuits.

Chapter 7

CONCLUSIONS AND FUTURE WORK

In this thesis, the design of low-light level and high-speed complete camera-on-a-chip imagers that are fully integrated in mainstream CMOS technology were studied. The following section discusses some concluding remarks about the work presented in Chapters 3-6, which is followed by recommendations for future work.

7.1. Summary and Conclusions

A platform for testing CMOS imagers was designed and tested in Chapter 3, using a fully integrated 256-pixel CMOS camera-on-a-chip fabricated in a standard CMOS 0.18 μm technology. The imager uses pixels that have a 60% fill-factor, in an array that occupies an area of $646 \mu\text{m} \times 390 \mu\text{m}$. All digital and analog blocks, including the row and column controllers, multiplexing circuits, sample-and-hold circuit and dual-slope analog-to-digital converter (ADC), have been implemented on-chip. The imager was tested by controlling it with an Altera FPGA board. When clocking the ADC at a frequency of 1 MHz, images were obtained at about 60 frames/s. The n+/p-sub photodiode used has a peak responsivity at a wavelength of 680 nm and the APS has a maximum SNR of 35 dB.

A careful review of high-speed imagers in Chapter 4, led to the conclusion that ultrahigh-speed imaging was not achievable using conventional digital array access and analog-to-digital data conversion techniques. The frame-rate of the standard 3T-APS imager that was presented in Chapter 3 was increased in Chapter 4, using an ultrahigh speed APS design that can take 8 frames at a rate of over a billion frames/s. This was achieved by separating the acquisition phase from the conversion and readout phase, which maintains an unaffected frame-rate, but postpones a portion of the delay between the frames to after the acquisition of a number of images. The design was fabricated in a

standard 130 nm technology from IBM and an array of 1024-pixels was fabricated with a pixel fill-factor of 9%. The imager shows extremely high frame-rates, however, the actual design is limited by two factors. The first limitation is due to the capacitive leakage of the small in-pixel memory elements, which was measured to be -78 V/s. This showed that the size of the array was limited to 32×32 pixels if a pixel-by-pixel readout method was used. In order to increase the array size, faster readout methods, or high-speed analog-to-digital converters are needed. The second design limitation for the ultrahigh-speed imager is the large light power required. To cause a voltage drop of 1 V within an integration time of 400 ps, an incident light power of roughly 1.3 W is required. Using an APD with a gain of 100 would reduce the power requirement to 13 mW, making the ultrahigh-speed image design more practical.

In this work, the sensitivity of CMOS imagers was increased by using single-photon avalanche photodetectors (SPAD). The APDs were operated in Geiger-mode and the deadtime of the SPADs, which limits the rate at which photons can be counted, was reduced. This was done by successfully fabricating an avalanche photodiode in a deep-submicron CMOS technology that can offer a small pixel size with high-speed transistors. The designed APD has a measured breakdown voltage of 11.3 V. When used in a passive Geiger-mode operation, the APD has a deadtime of 50 ns and a DCR of 1 kHz at an excess bias of 0.3 V. The active quench and reset SPAD circuit was designed in a way to reduce the number of devices as much as possible, where the achieved deadtime was around 40 times lower than previous work. In order to reduce afterpulsing effects, the circuit was also designed to operate under very low biasing conditions with separate bias nodes for the APD and the logic elements, allowing for minimum excess bias. The designed circuit has a deadtime in the range of 200-300 ps and can operate for bias conditions as low as 0.5 V. The pixel also has a higher fill-factor of 25% compared to 1-5% in previous work. The deadtime was further reduced using a novel, multi-detector per-pixel technique that allowed for reducing the deadtime at the expense of a small reduction in fill-factor, which is acceptable since the fill-factor was increased significantly compared to previous work.

Finally, the dynamic range of CMOS imagers was addressed using the conventional time-domain technique. However, unlike previous work, where the frame-rate would be

very low in order to achieve a reasonable dynamic range, high-speed operation was maintained. This was done using single-photon counting in time-domain, where the high sensitivity of single-photon counters was utilized to achieve count threshold detection in time-domain. The pixel can achieve a dynamic range of 100 dB in an 8 ms integration time, which is 3 orders of magnitude faster than conventional APS-based time-domain imagers. The speed was improved further using a scalable novel pixel design that can allow for simultaneous pixel counting and threshold detection. The pixel design however, was too complicated to be implemented using the technology node that was used in this work while maintaining a reasonable fill-factor. A novel in-pixel counting and threshold detection technique was implemented to reduce the complexity of the pixel using analog counters. The final pixel design had a fill-factor of 4% and contained a 9-bit dual-port SRAM, an active quench and reset SPAD and two cascaded analog counters and threshold detectors. The pixel was implemented in a complete camera-on-a-chip TDSPi with a 384-pixel array that could achieve high-dynamic range due to using time-domain, high-sensitivity due to using single-photon counting, and maintain high-speed due to low deadtime and digital event based counting.

7.2. Future Work

There are a number of areas of improvement that can be recommended for the work presented in this thesis. In addition to testing and implementing the imager designs using smaller scale CMOS technologies, the following discusses a number of ideas for future work.

Chapter 4 presented the ultrahigh-speed imager design that can capture 8 sequential frames with subnanosecond inter-frame delays. It was mentioned that the imager suffers from an array size restriction due to the capacitor leakage. The design can be implemented using a different capacitor in a different technology kit, or using metal-insulator-metal (MIM) capacitors that offer lower leakage at the expense of larger area. This will allow for increasing the resolution of the imager. The 3D packaging solution that uses optical fibers should also be investigated in order to allow for an increased number of frames by including the memory elements in an array outside of the line scan imager, as previously discussed in Chapter 4. This will allow for the MIM capacitors to

be used with no affect on the pixel fill-factor. The high-speed imager design also requires having high-speed clocking circuitry that operates in the gigahertz range, requiring the use of analog and RF clock sources, rather than digital ones to reduce jitter and clock skew. Also, to reduce the large optical illumination requirements, the integration of linear gain avalanche photodiodes, within the ultrahigh-speed pixel design, should be considered. Finally, when capturing images at such high frame-rates, a suitably high-speed analog-to-digital convertor needs to be designed and optimized for gigahertz conversion speeds. These designs need to be improved specifically for the operation of the proposed imager.

The avalanche photodiode devices tested were not optimized in terms of layout dimensions. In order to reduce the afterpulsing probability and dark count, physics-based models of the SPAD's breakdown behavior and probability are necessary to obtain the optimal layout. The layout needs to take into account the different noise sources and optimize the design based on noise contributions from the periphery of the device. Also, since the periphery wells are the major contributors to noise, the effect of biasing these diodes needs to be further investigated. The optimal APD bias voltage that minimizes the dark-count rate while maximizing the photon-detection probability needs to be obtained. This can be done by finding an analytical physics-based expression that relates the photon-detection probability and the dark-count-rate to the excess bias. The excess bias will also have an effect on the afterpulsing probability, as previously mentioned in Chapter 5, which is why an optimal excess bias operating point should be found. Once modeling of the SPAD is done, taking into account the location of major defect traps that contribute to afterpulsing, and also taking the trap lifetime into account, an optimal layout in mainstream technology can be achieved. Spacing out the different wells and guard layers of the APD and optimizing their dimensions can help in reducing afterpulsing. Currently, there is no clear guideline as to how to lay out an APD device with optimal dimensions in regards to improving speed and reducing noise. Designing a device that contains multiple smaller APDs that add up to the same area to be connected in parallel may be the best solution. This is possible using the APD layout discussed in Chapter 5, since a number of parallel devices can be implemented within the same well. The

performance may be enhanced by averaging multiple detectors simultaneously, in which case, the optimal number of detectors needs to be modeled.

Finally, since all the designs in this thesis use DSM mainstream CMOS technologies, which have multiple SiO₂ passivation and metal layers, the external quantum efficiency of the image sensors is reduced by around 60%. For example, the design presented in Chapter 3 has 6 metal layers, whereas, the designs presented in Chapters 4-6 have 8 metal layers. Even though the photosensitive part of the pixel is not covered by a metal layer, the passivation layers do exist and cause optical diffraction and cross-talk from one layer to the other, degrading the optical efficiency. Post processing techniques that can allow for the removal of some of these layers may be necessary. Also, it may be possible to use the pad mask layer, which indicates to the foundry to leave a glass opening over the pads to allow for a bond wire contact, to have the upper most passivation layers etched off from on top of the photosensitive elements.

References

- [1] N. Faramarzpour, M. M. El-Desouki, S. Shirani, Q. Fang, and M. J. Deen “CMOS imaging for biomedical applications,” *IEEE Potentials Magazine*, vol. 27, (3), pp. 31-36, May-June 2008.
- [2] A. El Gamal, H. Eltoukhy and K. Salama, “CMOS Sensors for Optical Molecular Imaging,” *CMOS Biotechnology*, Eds. H. Lee, R. M. Westervelt and D. Ham., Springer US, Book Series on Integrated Circuits and Systems, part IV, pp. 331-379, 2007.
- [3] Fluorescence Lifetime Imaging of NADH for Cancer Detection. Online:
<http://nimmi.bme.duke.edu/flim.html>
- [4] J. Siegel, D. S. Elson, S. E. D. Webb, K. C. B. Lee, A. Vlandas, G. L. Gambaruto, S. Le´ ve` que-Fort, M. J. Lever, P. J. Tadrous, G. W. H. Stamp, A. L. Wallace, A. Sandison, T. F. Watson, F. Alvarez, and P. M. W. French, “Studying biological tissue with fluorescence lifetime imaging: microscopy, endoscopy, and complex decay profiles,” *Journal of Applied Optics*, vol. 42, (16), 2995-2003, June 2003.
- [5] M. Kfoury, O. Marinov, P. Quevedo, N. Faramarzpour, S. Shirani, L. W. C. Liu, Q. Fang and M. J. Deen, “Towards a miniaturized wireless fluorescence-based diagnostic imaging system,” *IEEE J. Selected Topics Quantum Electronics*, vol. 14, no. 1, pp. 226-234, Jan./Feb. 2008.
- [6] M. M. El-Desouki, *Design of Integrated Power Amplifier Circuits for Biotelemetry Applications*. Master Thesis, Dept. Electrical and Computer Eng., McMaster University, Hamilton, Ontario, Jan. 2006.
- [7] G. Mone, “How it works: the endoscope camera in a pill,” *Popular Science Magazine*, March 13, 2008. Online: <http://www.popsci.com/node/19963>
- [8] N. Faramarzpour, M. M. El-Desouki, M. J. Deen, S. Shirani and Q. Fang, “CMOS photodetector systems for low-level light applications,” *Journal of Materials Science: Materials in Electronics*, vol. 20, pp. 87-93, Jan. 2009.
- [9] Concepts in Digital Imaging Technology, online:
<http://micro.magnet.fsu.edu/primer/digitalimaging/concepts/photomultipliers.html>
- [10] E. R. Fossum, Micron Technology, Inc., online: <http://www.ericfossum.com/>
- [11] W. S. Boyle and G. E. Smith, “Charge coupled semiconductor devices,” *Bell Syst. Tech. J.*, vol. 49, pp. 587-593, 1970.
- [12] A. El Gamal and H. Eltoukhy, “CMOS image sensors,” *IEEE Circuits and Devices Magazine*, vol. 21, (3), pp. 6-20, 2005.

- [13] E. R. Fossum, "CMOS image sensors: electronic camera-on-a-chip," *IEEE Transactions on Electronic Devices*, vol. 44, (10), pp. 1689-1698, 1997.
- [14] Bigas, M., E. Cabruja, J. Forest, and J. Salvi, "Review of CMOS image sensors," *Microelectronics Journal*, vol. 37, pp. 433-451, 2006.
- [15] Introduction to Digital Imaging in Microscopy, online:
<http://learn.hamamatsu.com/articles/microscopyimaging.html>
- [16] ITTFacts on Digital imaging, online: <http://www.itfacts.biz/index.php?id=P5517>.
- [17] ITTFacts on Digital imaging, online: <http://www.itfacts.biz/index.php?id=P7579>
- [18] ITTFacts on Digital imaging, online: <http://www.itfacts.biz/index.php?id=P1827>
- [19] CMOS Image Sensor and Module Pricing (May 2005),
online: <http://www.the-infoshop.com/study/s2d32523-cmos.html>
- [20] ITTFacts on Digital imaging, online: <http://www.itfacts.biz/index.php?id=P1827>
- [21] M. Mori, M. Katsuno, S. Kasuga, T. Murata, and T. Yamaguchi, "A 1/4in 2M pixel CMOS image sensor with 1.75transistors/pixel," *ISSCC Technical Digest*, vol. 47, pp. 110–111, 2004.
- [22] H. Takahashi, M. Kinoshita, K. Morita, T. Shirai, T. Sato, T. Kimura, H. Yuzurihara, S. Inoue, "A 3.9 μm pixel pitch VGA format 10b digital image sensor with 1.5-transistor/pixel," *IEEE Journal of Solid-State Circuits*, vol. 39, (12), pp. 2417-2425, 2004.
- [23] S. Parikh, Glenn Gulak, and P. Chow "A CMOS image sensor for DNA microarrays," *IEEE Custom Integrated Circuits Conference (CICC) 2007*, pp. 821-824, September 16-19, 2007.
- [24] F. Lustenberger, M. Lehmann, L. Cavalier, N. Blanc, W. Heppner, J. Ernst, St. Gick, H. Bloss, "A colour 3200fps high-speed CMOS imager for endoscopy in bio-medical applications," *European Solid-State Circuits Conference - ESSCIRC 2004*, pp. 415-418, September 21-23, 2004.
- [25] A.J.P. Theuwissen, "CMOS image sensors: state-of-the-art," *Solid-State Electronics*, vol. 52, pp. 1401–1406, 2008.
- [26] K. Salama, *CMOS Luminescence Detection Lab-on-Chip: Modeling, Design, and Characterization*. PhD Thesis, Dept. Electrical Computer Eng., Stanford University, California, USA, Sept. 2005.
- [27] C.L. Tisse, F. Guichard, and F. Cao, "Does resolution really increase image quality," *Proc. of SPIE*, vol. 6817, pp. 68170Q-68170Q-12, 2008.
- [28] F. Xiao, J. Farrell, P. Catrysse, and B. Wandell, "Mobile imaging: the big challenge of the small pixel," *Proc. of SPIE*, vol. 7250, pp. 72500K-72500K-9, 2009.
- [29] J.Farrell, F.Xiao and S.Kavusi, "Resolution and light sensitivity tradeoff with pixel size," *Proc. of SPIE Conference in Digital Photography II*, vol. 6069, pp. 211-218, 2006.
- [30] A. El Gamal, D. Yang, and B. Fowler, "Pixel level processing – why? what?, and how?," *Proc. SPIE*, vol. 3650 pp. 2-13, 1999.
- [31] P. Catrysse and B. Wandell, "Roadmap for CMOS image sensors: Moore meets Planck and Sommerfeld," *Proceedings of the SPIE*, vol. 5678, pp. 1-13, 2005.

References

- [32] N. Faramarzpour, *CMOS Photodetectors for Low-Light-Level Imaging Applications*. PhD Thesis, Dept. Electrical and Computer Eng., McMaster University, Hamilton, Ontario, April 2008.
- [33] S. M. Sze, *Physics of Semiconductor Devices*, 3rd Ed. New Jersey, John Wiley & Sons, Inc., 2007.
- [34] D. A. Neamen, *Semiconductor Physics and Devices*, 3rd Ed. New York: McGraw Hill, 2003.
- [35] Jun Ohta, *Smart CMOS Image Sensors and Applications*, Taylor & Francis Group, 2008.
- [36] J. Singh, *Semiconductor Optoelectronics*, New Jersey, Practice Hall, 2000.
- [37] S. O. Kasap. *Optoelectronics and Photonics*, New Jersey, Practice Hall, 2001.
- [38] P. Noble, "Self-scanned silicon image detector arrays," *IEEE Transactions on Electron Devices*, vol. 15, pp. 202-209, 1968.
- [39] N. Faramarzpour, M.J. Deen, S. Shirani, Q. Fang, L.W.C. Liu, F. Campos, and J.W. Swart, "CMOS based active pixel for low-light-level detection: analysis and measurements," *IEEE Transactions on Electron Devices*, vol. 54, (12), pp. 3229-3237, Dec. 2007.
- [40] K. Salama and A. El Gamal, "Analysis of active pixel sensor readout circuit," *IEEE Transactions on Circuits and Systems I*, vol. 50, (7), pp. 941-944, July 2003.
- [41] X. Wang, W. Wong, and R. Hornsey, "A high dynamic range CMOS image sensor with inpixel light-to-frequency conversion," *IEEE Transactions on Electron Devices*, vol. 53, (12), pp. 2988-2992, Dec. 2006.
- [42] S. Kleinfelder, S. Lim, X. Liu, and A. El Gamal, "A 10,000 frames/s CMOS digital pixel sensor," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 2049-2059, 2001.
- [43] 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, (1), pp. 92-101, Jan. 2001.
- [44] S. Majumder, M. M. El-Desouki, O. Marinov, and M. J. Deen, "Random telegraph signal noise in CMOS imagers and its impact on image quality," *The ECS 217th The Electrochemical Society Meeting*, Vancouver, Canada, April 25-30, 2010, (Abstract).
- [45] D. Pogany, S. Ababou, G. Guillot, X. Hugon, B. Vilotitch, and C. Lenoble, "Study of RTS noise and excess currents in lattice-mismatched InP/InGaAs/InP photodetector arrays," *Solid-State Electronics*, vol. 38, (1), pp. 37-49, 1995.
- [46] I. H. Hopkins and G. R. Hopkinson, "Further measurements of random telegraph signals in proton-irradiated CCDs," *IEEE Transactions on Nuclear Science*, vol. 42, pp. 2074-2081, Dec. 1995.
- [47] M. Ahoovie, and S. Kleinfelder, "Monolithic sensors for charged-particle imaging using per-pixel correlated double sampling," *IEEE Nuclear Science Symposium (NSSMIC)*, vol. 1, pp. 54-58, 2006.
- [48] W. Shockley and W.T. Read, "Statistics of the recombination of holes and electrons," *Physics Review*, vol. 87, pp. 835-843, Sept. 1952.
- [49] S. Majumder, "A study on noise and dynamic range in CMOS pixel," *Semiconductor Device Theory and Modeling course project report*, McMaster University, Jan. 2010.

- [50] K.K. Hung, P.K. Ko, C. Hu, and Y.C. Cheng, "A Unified for the flicker noise in metal-oxide-semiconductor field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 37, pp. 654–665, Mar. 1990.
- [51] J. P. Bentley, *Principles of Measurement Systems*. Third edition, Longman Group Limited, United Kingdom, 1995.
- [52] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*. Second edition, Oxford Press, New York, 2002.
- [53] R. J. Baker, *CMOS Circuit Design, Layout and Simulation*. Second edition, IEEE Press, NJ, 2005.
- [54] B. Razavi, "Design of sample-and-hold amplifiers for high-speed low-voltage A/D converters," *IEEE Custom Integrated Circuits Conference*, pp. 59-66, May 1997.
- [55] I. Shcherback and O. Yadid-Pecht, "CMOS APS MTF Modeling", in "CMOS Imagers: From Photo-transduction to Image Processing", O. Yadid-Pecht and R. Etienne-Cummings, 1 ed, pp. 53-72, Kluwer, 2004.
- [56] M. M. El-Desouki, M. Jamal Deen, Q. Fang, L. Liu, F. Tse and D. Armstrong. "CMOS image sensors for high speed applications," *Sensors*, vol. 9, (1), pp. 430-444, 2009.
- [57] T. Huang, S. Sorgenfrei, K. L. Shepard, P. Gong, and R. Levicky, "A CMOS array sensor for sub-800-ps time-resolved fluorescence detection," *IEEE Custom Integrated Circuits Conference (CICC)*, pp. 829-832, 2007.
- [58] G. E. Hogan, K. J. Adams, K. R. Alrick, J. F. Amann, J. G. Boissevain, L. M. Crow, *et al* "Proton radiography," *Proceedings 1999 IEEE Particle Accelerator Conference*, vol. 2, pp. 579-583, 1999.
- [59] A. I. Krymski, D. Van Blerkom, A. Andersson, N. Block, B. Mansoorian, and E. R. Fossum, "A high speed, 500 frames/s, 1024 × 1024 CMOS active pixel sensor," *Proc. 1999 Symposium on VLSI Circuits*, pp. 137-138, 1999.
- [60] A. I. Krymski, N. E. Bock, N. Tu, D. V. Blerkom, and E. R. Fossum, "A high-speed, 240-frames/s, 4.1-Mpixel CMOS sensor," *IEEE Transaction on Electron Devices*, vol. 50, pp. 130-135, 2003.
- [61] Y. Nishikawa, S. Kawahito, M. Furuta, and T. Tamura, "A high-speed CMOS image sensor with on-chip parallel image compression circuits," *IEEE Custom Integrated Circuits conference CICC*, pp. 833-836, 2007.
- [62] M. Furuta, T. Inoue, Y. Nishikawa, and S. Kawahito, "A 3500fps high-apeed CMOS image sensor with 12b column-parallel cyclic A/D converters," *Technical Digest on VLSI Circuits*, pp. 21-22, 2007.
- [63] J. L. Trepanier, M. Sawan, and Y. A. Udet, "A new CMOS architecture for wide dynamic range image sensing," *Canadian Conference on Electrical and Computer Engineering - CCECE*, vol. 1-3, pp. 323-326, 2003.
- [64] R. Ghannoum, and M. Sawan, "A 90nm CMOS multimode image sensor intended for a visual cortical stimulator," *Int. Conference on Microelectronics ICM*, pp. 197-182, 2007.

References

- [65] J. Dubois, D. Ginhac, and M. Paindavoine, "A single-chip 10000 frames/s CMOS sensor with in-situ 2D programmable image processing," *International Workshop on Computer Architecture for Machine Perception and Sensing*, pp. 124-129, 2006.
- [66] N. Stevanovic, M. Hillegrand, B. J. Hostica, and A. Teuner, "A CMOS image sensor for high-speed imaging," *Proceedings IEEE International Solid-State Circuits Conference*, pp. 104-105, 2000.
- [67] S. Lauxtermann, P. Schwider, P. Seitz, H. Bloss, J. Ernst, H. Firla, "A high speed CMOS imager acquiring 5000 frames/sec," *IEDM Technical Digest*, pp. 875-878, 1999.
- [68] T. Sugiyama, S. Yoshimura, R. Suzuki, and H. Sumi, "A 1/4-inch QVGA color imaging and 3-D sensing CMOS sensor with analog frame memory," *IEEE International Solid-State Circuits Conference*, pp. 434-435, 2002.
- [69] T. G. Etoh, D. Poggemann, G. Kreider, H. Mutoh, A. J. P. Theuwissen, A. Ruckelshausen, Y. Kondo, H. Maruno, K. Takubo, H. Soya, K. Takehara, T. Okinaka, and Y. Takano, "An image sensor which captures 100 consecutive frames at 1,000,000 frames/s," *IEEE Transactions on Electron Devices*, vol. 50, pp. 144-151, 2003.
- [70] S. Kleinfelder, Y. Chen, K. Kwiatkowski, and A. Shah, "High-speed CMOS image sensor circuits with in-situ frame storage," *IEEE Transactions on Nuclear Science*, vol. 51, pp. 1648-1656, 2004.
- [71] N. Faramarzipour, M. J. Deen, S. Shirani, Q. Fang, L. W. C. Liu, F. Campos, and J. W. Swart, "CMOS based active pixel for low-light-level detection: analysis and measurements," *IEEE Transactions on Electron Devices*, vol. 54, (12), pp. 3229-3237, Dec. 2007.
- [72] B. J. Hosticka, W. Brockherde, A. Bussmann, T. Heimann, R. Jeremias, A. Kemna, C. Nitta, and O. Schrey, "CMOS imaging for automotive application," *IEEE Transactions on Electron Devices*, vol. 50, pp. 173-183, 2003.
- [73] A. Simoni, G. Torelli, F. Maloberti, A. Sartori, S. Plevridis, and A. Birbas, "A single-chip optical sensor with analog memory for motion detection," *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 800-806, 1995.
- [74] S.-W. Han, S.-J. Kim, J. Choi, C.-K. Kim, E. Yoon, "A high dynamic range CMOS image sensor with in-pixel floating-node analog memory for pixel level integration time control," *VLSI Circuits Symposium Digest*, pp. 31-32, 2006.
- [75] Y. Ni, and X. L. Yan, "CMOS active differential imaging device with single in-pixel analog memory," *IEEE European Solid-State Circuits Conference*, pp. 359-362, 2002.
- [76] A. Aslam-Siddiqi, W. Brockherde, M. Schanz, and B. J. Hosticka, "A 128-pixel CMOS image sensor with integrated analog nonvolatile memory," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 1497-1501, 1998.
- [77] G. Chapinal, and S. Bota, "A 128 x 128 CMOS image sensor with analog memory for synchronous image capture," *IEEE Sensors Journal*, vol. 2, pp. 120-127, 2002.
- [78] S. A. J. Radovanovic, and B. Nauta, "A 3-Gb/s optical detector in standard CMOS for 850-nm optical communication," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 1706-1717, 2005.

- [79] M. M. El-Desouki, M. Jamal Deen, Q. Fang and F. Tse. "Ultrahigh speed CMOS camera-on-a-chip for biomedical applications," *European Conferences on Biomedical Optics (ECBO)*, Munich, Germany, 3-pages, June 2009.
- [80] A. Rochas, M. Gösch, A. Serov, P. A. Besse, R. S. Popovic, T. Lasser and R. Rigler "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," *IEEE Photonics Technology Letters*, vol. 15, pp. 963, Jul. 2003.
- [81] F. Zappa , A. Gulinatti , P. Maccagnani , S. Tisa and S. Cova "SPADA: Single-photon avalanche diode arrays," *IEEE Photonics Technology Letters*, vol. 17, pp. 657, Mar. 2005.
- [82] V.H. Dhulla, G. Gudkov, D. Gavrilov, A. Stepukhovich, A. Tsupryk, O. Kosobokova, A. Borodin, B. Gorbovitski and V. Gorfinkel, "Single Photon Counting Module Based on Large Area APD and Novel Logic Circuit for Quench and Reset Pulse Generation," *IEEE Journal of Selected Topics in Quantum Electronics*, vol.13, (4), pp. 926-933, July-Aug. 2007
- [83] A. Lacaita, M. Ghioni, and S. Cova, "Double epitaxy improves single-photon avalanche diode performance," *Electronics Letters*, vol. 25, (13), pp. 841-843, 22 June 1989.
- [84] S. Tisa , A. Tosi and F. Zappa "Fully-integrated CMOS single photon counter," *Optics Express*, vol. 15, pp. 2873-2887, Mar. 2007.
- [85] S. Tisa, F. Guerrieri and F. Zappa "Variable-load quenching circuit for single-photon avalanche diodes," *Optics Express*, vol. 16, pp. 2232-2244, Feb. 2007.
- [86] M. M. El-Desouki, M. Jamal Deen, and Q. Fang. "A CMOS single-photon avalanche photodetector camera-on-a-chip for biomedical applications," *14th Canadian Semiconductor Technology Conference (CSTC09)*, Hamilton, Ontario, Canada, Aug. 2009.
- [87] N. Faramarzpour, M. J. Deen, S. Shirani, and Q. Fang, "Fully integrated single photon avalanche diode detector in standard CMOS 0.18- μm technology," *IEEE Transactions on Electron Devices*, vol.55, (3), pp.760-767, March 2008.
- [88] O. Marinov, M. J. Deen and J. A. J. Tejada, "Theory and Physical Explanation of the Microplasma Fluctuation and Noise in Avalanche Breakdown of the Silicon Diodes," *Journal of Applied Physics*, vol. 101, 297-302, March 2007.
- [89] O. Marinov and M.J. Deen, "Physical Model for Low Frequency Noise in Avalanche Breakdown of pn Junctions", *Fluctuations and Noise Letters*, Vol. 4, (2), pp. L287-L296 (June 2004).
- [90] A. Goetzberger, B. McDonald, R.H. Haitz, and R.M. Scarlett, "Avalanche effects in silicon p-n junctions. II. Structurally perfect junctions," *Journal of Applied Physics*, vol. 34, pp. 1591-1600, 1963.
- [91] A. Spinelli, and A. L. Lacaita, "Physics and numerical simulation of single photon avalanche diodes," *IEEE Transactions on Electron Devices*, vol. 44, (11), pp. 1931-1943, Nov 1997.
- [92] W. O. Oldham , R. R. Samuelson and P. Antognetti "Triggering phenomena in avalanche diodes," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 1056-1060, Sep. 1972.

References

- [93] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa and S. Cova “Progress in silicon single-photon avalanche diodes,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 13, pp. 852-862, Jul./Aug. 2007.
- [94] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, “A single photon avalanche diode implemented in 130-nm CMOS technology,” *IEEE Journal of Selected Topics in Quantum Electronics*, vol.13, (4), pp.863-869, July-Aug. 2007.
- [95] L. Neri, S. Tudisco, F. Musumeci, A. Scordino, G. Fallica, M. Mazzillo, M.; Zimbone, “Dead time causes and correction method for single photon avalanche diode devices,” *Review of Scientific Instruments*, vol. 81, pp. 086102-1-086102-3, 2010.
- [96] G. Vincent, A. Chantre, and D. Bois, “Electric field effect on the thermal emission of traps in semiconductor junctions,” *Journal of Applied Physics*, vol. 50, pp. 5484-5487, 1979.
- [97] S. Cova, A. Lacaita and G. Ripamonti “Trapping phenomena in avalanche photodiodes on nanosecond scale,” *IEEE Electron Device Letters*, vol. 12, pp. 685-687, Dec. 1991.
- [98] H. Finkelstein , M. J. Hsu and S. C. Esener “STI-bounded single-photon avalanche diode in a deep-submicrometer CMOS technology,” *IEEE Electron Device Letters*, vol. 27, pp. 887-889, Nov.2006.
- [99] D. Stoppa, D. Mosconi, L. Pancheri, L. Gonzo, “Single-photon avalanche diode CMOS sensor for time-resolved fluorescence measurements,” *IEEE Sensors Journal*, vol.9, (9), pp. 1084-1090, Sep. 2009.
- [100] M. A. Marwick, and A. G. Andreou, “Single photon avalanche photodetector with integrated quenching fabricated in TSMC 0.18 μm 1.8 V CMOS process,” *Electronics Letters*, vol. 44, (10), pp. 643-644, May 8 2008.
- [101] F. Zappa, S. Tisa, A. Gulinatti, A. Gallivanoni, and S. Cova, “Monolithic CMOS detector module for photon counting and picosecond timing,” *Proceeding of the 34th European Solid-State Device Research Conference ESSDERC 2004*, pp. 341-344, 21-23 Sept. 2004.
- [102] F. Zappa, A. Lotito, A. C. Giudice, S. Cova, and M. Ghioni “Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors,” *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 1298-1301, Jul. 2003.
- [103] A. Gallivanoni, I. Rech, D. Resnati, M. Ghioni, and S. Cova, “Monolithic active quenching and picosecond timing circuit suitable for large-area single-photon avalanche diodes,” *Optics Express*, vol. 14, pp. 5021-5030, 2006.
- [104] F. Zappa, M. Ghioni, S. Cova, C. Samori, and A. C. Giudice, “An integrated active-quenching circuit for single-photon avalanche diodes,” *IEEE Transactions on Instrumentation and Measurement*, vol. 49, (6), pp.1167-1175, Dec 2000.
- [105] H. Finkelstein, M. J. Hsu, S. Zlatanovic, and S. Esener, “Performance trade-offs in single-photon avalanche diode miniaturization,” *Rev. Sci. Instrum.*, vol. 78, p. 103103-103107, 2007.

- [106] A. Rochas, P.-A. Besse, and R. S. Popovic, "Actively recharged single photon counting avalanche photodiode integrated in an industrial CMOS process," *Sensors and Actuators A: Physical*, vol. 110 (1-3), pp. 124-129, February 2004.
- [107] L. Pancheri, and D. Stoppa, "Low-Noise CMOS single-photon avalanche diodes with 32 ns dead time," *The 37th European Solid State Device Research Conference ESSDERC 2007*, pp. 362-365, 11-13 Sept. 2007.
- [108] F. S. Campos, O. Marinov, N. Faramarzpour, F. Saffih, M. J. Deen, and J. W. Swart. "A multisampling time-domain CMOS imager with synchronous readout circuit," *Analog Integrated Circuits Signal Processing*, vol. 57, pp. 151-159, 2008.
- [109] A. Kitchen, A. Bermak, and A. Bouzerdoum, "A digital pixel sensor array with programmable dynamic range," *IEEE Transactions on Electron Devices*, vol. 52, pp. 2591-2601, 2005.
- [110] B. Hong, and I. Shi, "Dual port SRAM cell," U.S. Patent 7,116,605 B2, Oct. 3, 2006.