CMOS BASED LENSLESS IMAGING SYSTEMS
AND SUPPORT CIRCUITS

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Changhyuk Lee
August 2014
While much progress has been made in various fields of study in past few decades, leading to better understanding of science as well as better quality of life, the role of optical sensing has grown among electrical, chemical, optical, and other physical signal modalities. As an example, fluorescent microscopy has become one of the most important methods in the modern biology. However, broader implementation of optical sensing has been limited due to the expensive and bulky optical and mechanical components of conventional optical sensor systems. To address such bottleneck, this dissertation presents several cost-effective, compact approaches of optical sensor arrays based on solid state devices that can replace the conventional components.

As an example, in chapter 2 we demonstrate a chip-scale ($<1 \text{ mm}^2$) sensor, the Planar Fourier Capture Array (PFCA), capable of imaging the far-field without any off-chip optics. The PFCA consists of an array of angle-sensitive pixels manufactured in a standard semiconductor process, each of which reports one component of a spatial two-dimensional (2D) Fourier transform of the local light field. Thus, the sensor directly captures 2D Fourier transforms of scenes. The effective resolution of our prototype is approximately 400 pixels. My work on this project [15] includes a circuit design and layout and the overall testing of the imaging system. In chapter 3 we present a fully integrated, Single Photon Avalanche Detector (SPAD) using only standard low-
voltage (1.8V) CMOS devices in a 0.18m process. The system requires one high-voltage AC signal which alternately reverse biases the SPADs into avalanche breakdown and then resets with a forward bias. The proposed self-quenching circuit intrinsically suppresses after-pulse effects, improving signal to noise ratio while still permitting fine time resolution. The required high-voltage AC signal can be generated by resonant structures and can be shared across arrays of SPADs [24]. An ideal light sensor to provide the precise incident intensity, location, and angle of incoming photons is shown in chapter 4. Single photon avalanche diodes (SPADs) provide such desired high (single photon) sensitivity with precise time information, and can be implemented at a pixel scale to form an array to extract spatial information. Furthermore, recent work has demonstrated photodiode-based structures (combined with micro-lenses and diffraction gratings) that are capable of encoding both spatial and angular information of the incident light. In this chapter, we describe the implementation of such grating structure on SPAD to realize a pixel-scale angle-sensitive single photon avalanche diode (A-SPAD) using a standard CMOS process. While the underlying SPAD structure provides the high sensitivity, the diffraction gratings consisting of two sets of metal layers offers the angle-sensitivity. Such unique combination of the SPAD and the diffraction gratings expand the sensing dimensions to pave a path towards a lens-less 3-D imaging and a light-field time-of-flight imaging. In chapter 5, we present a 72 × 60, angle-sensitive single photon avalanche diode (A-SPAD) array for lens-less 3-D fluorescent life time imaging. A-SPAD pixels are comprised of (1) a SPAD to resolve precise timing information, to reject high-powered UV stimulus, and to map the lifetimes of different fluorescent sources and (2) integrated diffraction gratings on top of the SPAD to extract incident angles of incoming light, enabling 3-D localization at
a micrometer scale. The chip presented in this work also integrates pixel-level counters as well as shared timing circuitry, and is implemented in conventional 180nm CMOS technology without any post-processing. Contact-based read-out from a revolving MEMS accelerometers is problematic therefore contactless (optical) read-out is preferred. The optical readout requires an image sensor to resolve nanometer-scale shifts of the MEMS image. Traditional imagers record on a rectangular grid which is not well-suited for efficiently imaging rotating objects due to the significant processing overhead required to translate Cartesian coordinates to angular position. Therefore, in chapter 6 we demonstrate a high-speed (1kfps), circular, CMOS imaging array for contact-less, optical measurement of rotating inertial sensors. The imager is designed for real-time optical readout and calibration of a MEMS accelerometer revolving at greater than 1000rpm. The imager uses a uniform circular arrangement of pixels to enable rapid imaging of rotational objects. Furthermore, each photodiode itself is circular to maintain uniform response throughout the entire revolution. Combining a high frame rate and a uniform response to motion, the imager can achieve sub-pixel resolution (25nm) of the displacement of micro scale features. In order to avoid fixed pattern noise arising from non-uniform routing within the array we implemented a new global shutter technique that is insensitive to parasitic capacitance. To ease integration with various MEMS platforms, the system has SPI control, on-chip bias generation, sub-array imaging, and digital data read-out. My work on this project [20] includes a circuit design and layout and some testing including, a FPGA based controller design of the imaging system. In the previous chapters, compact and cost effective imaging systems have been introduced. Those imaging systems show great potential for wireless implantable systems. A power rectifier for the implant provides a volt-
age DC power with a small inductor, for small volume, from a small AC voltage input. In the last chapter we demonstrate an inductively powered, orthogonal current-reuse multi-channel amplifier for power-efficient neural recording. The power rectifier uses the input swing as a self-synchronous charge pump, making it a fully passive, full-wave ladder rectifier. The rectifier supplies 10.37\(\mu\)W at 1.224V to the multi-channel amplifier, which includes bias generation. The prototype device is fabricated in a TSMC 65nm CMOS process, with an active area of 0.107mm\(^2\). The maximum measured power conversion efficiency (PCE) is 16.58% with a 184mV input amplitude. My work on this project\[25\] includes the rectifier design and overall testing to combine “orthogonal current-reuse neural amplifier” designed by Ben Johnson.
BIOGRAPHICAL SKETCH

Changhyuk Lee earned his Bachelor of Science degree in electrical engineering from Korea University, Seoul, Korea, in 2005. He is currently working towards his Ph.D. degree. Before starting his research at Cornell, he worked at Samsung, Chang-won, Korea, where he was responsible for gas turbine engine control system design from 2004 to 2006, and he also worked for Com2us, Seoul, Korea, developing mobile games from 2007 to 2008. His Ph.D. work focuses on the integration of non-conventional devices on CMOS for lens-less imaging and RF integrated circuits for wireless communication.

While in graduate school, he also initiated a collaboration with Sunwoo Lee a PhD candidate in Professor James Hones group and Michael Lekas a PhD candidate in Professor Kenneth Shepard’s group at Columbia, to use graphene, a sheet of carbon a single atom thick that has outstanding electrical properties and is the world’s strongest material, to develop smaller, more flexible, and more frequency-selective on-chip filters that could lead to improved wireless devices. With the collaboration he is a recipient of the Qualcomm Innovation Fellowship 2013-2014 for Graphene Resonator Based Mixer-First Receiver on CMOS for Digitally Controlled and Widely Tunable RF Interface.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Alyosha Molnar. I still remember his e-mail asking me for a phone interview with an opportunity to study with him, which has proven to be an incredibly wonderful experience in my life. He is truly a great mentor and excellent teacher. During the study, he always tried to inspire me to propose my own ideas and opinions and that encouragement and patience as an advisor led me to explore a wide range of knowledge and experience. I appreciate his contributions of time, ideas, and also fund raising to support my Ph.D. experience productive and stimulating. I would also like to acknowledge my committee members, professor Amit Lal and Ehsan Afshari. Unlike any other usual graduate study, my study has had numbers of great experiences to learn from my committees through various projects and classes. I also want to thank Christoph Studer for his advice and invaluable effort to evaluate my work.

I would also like to acknowledge my fellow students from the Molnar group: A significant portion of my work has been done through collaboration, PFCA, with Dr. Patrick Gill. His advice and expertise on a fast algorithms for signal processing and design of lens-less 3-D imagers have elevated my work to a higher level. Albert Wang for his help initiate a lens-less imaging using ASP and conversations and discussions with him that I enjoyed every bit of, Caroline Andrews for her collaboration in the LO overlap project and providing the simple LTI model of the passive mixer that later inspired my studies, Ben Johnson for helping many of my layouts, writings and discussing so many things that span from good to bad ideas which ended up contributing to many of my projects including ladder rectifier with stacked amplifier project and circular
imager, and Sriram Sivaramakrishnan for being the only one to discuss about ASP after Albert left and for being a catalyst in the circular imager testing to bring it to life. And also to the colleagues still in Molnar group: Hazal Yuksel for wonderful Korean food and joyful discussions, Dong Yang for the collaboration in duplexer project, Suren Jayasuriya for the humors in our office, Rose Agger for inviting me to a party which I could not make. Also, many thanks go to Sungyoon Park for initial SPAD research, Donggue Lee for working together on PFCA project, Yoonho Lee for sharing the memory of my early Ph.D studies, Jaeyoon Park for all the support and encouragements and so many advices that he has given me throughout my Ph.D study and still today.

I would particularly like to thank Sunwoo Lee from the James Hones Lab with whom I’ve been able to do a very interesting and intense research which has offered us a wonderful award. The Sonic MEMS group also have my deepest gratitude for their companionship, in particular, Jason Hoople, Po-Cheng Chen, Dr. Kwame Amponsah, Ved Gund, Sachin Nadig, and Justin Kuo. I would like to express my sincere gratitude to my friends who I met at Cornell, Hamidreza Aghasi, Waclaw Godycki, Avik Dutt, Jiyong Shin, Dr. Sung Gu Kang, Hanjong Paik, Jayhun Lee, Moon Kyung Kim, Dr. Eugene Hwang, Dr. Jungheyeon Hwang, and Jinsup Kim. I would especially like to thank Hyeseung Chung for sharing amazing memories.

Without the sincere support of my father, Dr. Han seon Lee, I would not be able to take this journey by the smallest chance. I thank my mother, Shinhee Park, for all her love and prayer. I thank my brother, CJ Lee, for his great support and care as the only small brother and also my best friend.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biographical Sketch</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>4</td>
</tr>
<tr>
<td>1.2 Overview</td>
<td>7</td>
</tr>
<tr>
<td>2 A microscale camera using direct Fourier-domain scene capture: PFCA</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Introduction to Lenseless Far-field imaging</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Theory: Fourier imaging</td>
<td>10</td>
</tr>
<tr>
<td>2.3 System Design</td>
<td>11</td>
</tr>
<tr>
<td>2.4 Measurement Results</td>
<td>15</td>
</tr>
<tr>
<td>2.5 Conclusions</td>
<td>16</td>
</tr>
<tr>
<td>3 Self-Quenching, Forward-bias-reset for Single Photon Avalanche Detectors</td>
<td>17</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>3.2 Self-Quenching Active Reset SPAD Design</td>
<td>19</td>
</tr>
<tr>
<td>3.2.1 SPAD design</td>
<td>19</td>
</tr>
<tr>
<td>3.2.2 Passive quenching (PQ)</td>
<td>20</td>
</tr>
<tr>
<td>3.2.3 Active quenching circuits (AQC)</td>
<td>21</td>
</tr>
<tr>
<td>3.2.4 Self quenching forward-bias reset (SQFBR)</td>
<td>22</td>
</tr>
<tr>
<td>3.3 Readout Circuit</td>
<td>23</td>
</tr>
<tr>
<td>3.4 Result &amp; Discussion</td>
<td>24</td>
</tr>
<tr>
<td>3.4.1 Fabrication process &amp; test setup</td>
<td>24</td>
</tr>
<tr>
<td>3.4.2 Dark Count Ratio</td>
<td>26</td>
</tr>
<tr>
<td>3.4.3 After-pulsing</td>
<td>26</td>
</tr>
<tr>
<td>3.4.4 Photon Count efficiency</td>
<td>28</td>
</tr>
<tr>
<td>3.4.5 Time-resolved measurements</td>
<td>29</td>
</tr>
<tr>
<td>3.4.6 Scalable and resonant drive</td>
<td>30</td>
</tr>
<tr>
<td>3.5 Conclusion</td>
<td>32</td>
</tr>
<tr>
<td>4 Angle Sensitive Single Photon Avalanche Diode</td>
<td>33</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>33</td>
</tr>
<tr>
<td>4.2 Angular Sensitivity</td>
<td>33</td>
</tr>
<tr>
<td>4.3 Temporal Sensitivity</td>
<td>35</td>
</tr>
<tr>
<td>4.4 A-SPAD Design</td>
<td>37</td>
</tr>
<tr>
<td>4.5 Measurement</td>
<td>40</td>
</tr>
<tr>
<td>4.6 Angular Resolution</td>
<td>43</td>
</tr>
<tr>
<td>4.7 Conclusion</td>
<td>45</td>
</tr>
</tbody>
</table>
5 An On-chip 72×60 Angle-Sensitive Single Photon Image Sensor Array for Lens-less Time-resolved 3-D Fluorescence Lifetime Imaging
5.1 Introduction .............................................. 46
5.2 Angle-sensitive Single Photon Avalanche Diode Design ............... 49
5.3 Large Scale SPAD Array Considerations: Desensitization ............. 60
5.4 Measurement ............................................. 62
5.5 Digital to Time Converter Static Performance .......................... 64
5.6 Conclusion ............................................... 71

6 Efficient Imaging: A High-Speed Polar-Symmetric Imager for Real-Time Calibration of Rotational Inertial Sensors
6.1 Introduction .............................................. 73
6.2 Theory ................................................... 75
6.3 System Design .............................................. 77
  6.3.1 Architecture ........................................... 77
  6.3.2 Circular Layout ....................................... 80
  6.3.3 Pixel Readout Circuitry ................................ 81
  6.3.4 Backend Readout Circuitry .............................. 82
6.4 Measurement Results .......................................... 82
6.5 Conclusion ............................................... 85

7 A Sub-threshold Voltage Ladder Rectifier for Orthogonal Current-reuse Neural Amplifier
7.1 Introduction .............................................. 88
  7.1.1 Energy Efficient Neural Amplifier .......................... 89
  7.1.2 Power Conversion Efficiency (PCE) and Passive Rectifier ........ 91
7.2 System Design .............................................. 92
  7.2.1 Ladder Charge Pump Design ............................... 92
  7.2.2 Ladder Rectifier Analysis ................................. 94
  7.2.3 Amplifier Design ....................................... 96
7.3 Measurement Results .......................................... 97
7.4 Conclusion ............................................... 98

Bibliography ................................................................ 101
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>System summary</td>
<td>57</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

1.1 High resolution trend and pixel scaling shrinkage [3]  

2.1 (Color online) Trade-off between imager size and image resolution. The PFCA fills in a gap left in the available resolution-size trade-off between a naked single photodiode and the smallest commercial focusing cameras [33].  

2.2 (Color online) Angle-sensitive pixels. Incident light interacting with a grating composed of an upper metal layer produces an interference pattern at the depth of the second-layer grating. Light is either passed or blocked depending on the alignment of the interference pattern and the second layer. This alignment, in turn, is sensitive to changes in incident angle; the net effect is that the light passed by an ASP depends sinusoidally on incident angle.  

2.3 Decomposition of natural images into Fourier components. (a) All images can be expressed as a sum of 2D Fourier basis functions [e.g., (b)] by taking the sum over all values in (c) the basis-scaled image.  

2.4 (Color online) Selecting devices for the PFCA. Filled circles indicate manufacturable devices with maximal m for 520nm light; open circles indicate the suite of devices we manufactured.  

2.5 (Color online) Manufactured PFCA. Concentric rings of ASPs with increasingly high sinusoidal periodicity yield a complete Fourier description of the light intensity from the far field. Slowly varying orientation is evident from the enlarged section, where schematics show different metal layers in different colors.  

2.6 (Color online) Manufactured PFCA. Concentric rings of ASPs with increasingly high sinusoidal periodicity yield a complete Fourier description of the light intensity from the far field. Slowly varying orientation is evident from the enlarged section, where schematics show different metal layers in different colors.  

2.7 Image reconstructions. Using the basis functions obtained in the calibration phase (Fig. 2.6), (a) the image presented was (b) reconstructed up to the Nyquist limit of our array. No off-chip optics were used.  

3.1 Single Photon Detector. (cross section)  

3.2 Self-Quenching circuits without readout.  

3.3 Self-Quenching and reading out mechanism.  

3.4 Waveform vs. time (Note: (a) no avalanche, (b) avalanche in T₃, and (c) avalanche out of T₃)  

3.5 SPAD with SQC die photo  

3.6 Autocorrelation of output bits (SQFBR vs. synchronous PQ)  

ix
3.7 Detection probability vs. $V_{PP}$ and vs. Incident numbers of photon (Numbers for Number of Photon/ $T_S$) and DCR ................. 30
3.8 8 by 8 SPAD array with resonant drive .................................. 31

4.1 Cross-section of proposed A-SPAD structure ............................... 36
4.2 Die-photo of the A-SPAD unit prototype chip with two $\beta$s (7,15) 38
4.3 Balanced tri-phase ................................................................. 39
4.4 Time and angle independence measurement ................................ 40
4.5 Angular amplitude modulation and signal to noise ratio for two different illumination conditions and sample size (averaged frames) ................................................................. 41
4.6 FTDT wave simulation result (b) Conventional photo-diode ASP measurement (c) Angle sensitivity Measurement (N=8000) ........ 42

5.1 3-D imaging without lens and filter (a) conventional fluorescent imaging, (b) on-chip fluorescent imaging ................................. 47
5.2 A cross-section of the proposed A-SPAD structure .......................... 50
5.3 Pixel schematic of A-SPAD ....................................................... 51
5.4 Schematic of an in-pixel differential counter ................................ 52
5.5 Repeated time-resolved fluorescent imaging timing diagram ......... 53
5.6 Schematic of 10 bit asynchronous in-pixel ripple counter ............. 53
5.7 Statistical sampling with detection window and time histogram acquisition ................................................................. 54
5.8 Layout of two A-SPAD pixel ...................................................... 55
5.9 (a) System architecture of a 72 x 60 A-SPAD array and (b) a brief summary of the system ................................................................. 56
5.10 Schematic of high dynamic range, area efficient DTC .................... 58
5.11 Schematic of delay unit ............................................................. 59
5.12 (a) SPAD bias network impedance and Local decoupling capacitors (b) Simulated SPAD reverse bias voltage (red: avalanche pixel, gray: neighbor pixel) ..................................................... 60
5.13 Micro photograph of the A-SPAD FLIM IC .................................. 63
5.14 (a) DNL of the DTC (b) INL of the DTC ..................................... 65
5.15 On-chip FLIM measurement setup ........................................... 65
5.16 (a) Computing intensity & angle (b) sub-images for each A-SPAD type ................................................................. 67
5.17 Concept of lens-less, time-resolved 3-D FLIM using two different fluorophores with different lifetimes ................................. 68
5.18 3-D reconstruction with ASP .................................................... 69
5.19 Example snap-shot frames for reconstructing the fluorophores .. 70
5.20 Time histogram for lifetimes (intensity change) from the pre-localized fluorophores position ................................................................. 71
5.21 3-D reconstruction of two groups of fluorophores ...................... 72
CHAPTER 1
INTRODUCTION

Enhanced digital imaging capabilities have been driven in large part through development of better and CMOS image sensors, largely through reduced pixel size. The pixel miniaturizing trend in today’s semiconductor technology, however, has been slowing down as shown in Fig 1.1. This is not only due to the diminishing benefit of cost reduction but also the physical limit, i.e. cross-talk between pixels, that prevents an obvious performance enhancement from the device size scaling. [3]. Despite the effort to scale down the pixel size, a typical imaging system is as much limited by the requirement for a lens and often a filter as by the following homogenous array of photodiodes.

The lens particularly is hard to miniaturize and optical filters are expensive and wavelength specific. And yet, many sensing applications do not require the ultra-high resolution that comes from a precise lens, and many Megapixels of sub-micron pixels, we present an optical-component-less imaging systems and supporting circuits for ultra compact far-field imaging, contact-less nano-meter resolution optical ruler for MEMS accelerometer calibration, power rectifier for a wireless sensor nodes, and most importantly, lens and filter-less 3-D fluorescence lifetime microscope (FLIM). The majority of the effort here focuses on the last particular class of imaging system, which is lens-less FLIM imaging IC.

Fluorescence imaging devices can be categorized into four major types. 1) non-imaging fluorescent sensors, 2) fluorescent microscopy, 3) time resolved imaging, and 4) Multi-photon imaging. The first and also the simplest type
is for an application of chemical detection without a spatial resolution [17]. A typical example of this instrument type utilizes either micro-fluidic channel or optical fibers to limit the position of excitation and emission. They often have a strong focus on reducing the size due to needs of applications of interest to replace the bench-top instrumentation, and instead build a lab-on-chip solution. The first instrumentation reported in this category uses a fiber optic bundle, illuminated with a 532 nm Nd-YAG laser. A series of lens, filters and an avalanche photodiode (APD) followed by a readout electronics are used to detect fluorescence in electrophoresis. A second example of an instrument is capable of a duo wavelength fluorescence detection [5]. The device is to measure chlorophyll fluorescence and it includes a 635 nm laser diode, optical fibers, two PIN photodiodes, interference filters, and off-the-shelf readout system. The second, most popular option is fluorescence imaging microscopy, which incorporates a light source, multiple, movable focusing lenses, as well as filters and a CCD or CMOS
image sensor. Within fluorescent microscopy, confocal imaging provides better control over an imaging resolution, especially in the z-dimension, and therefore more popular than wide-field imaging. Despite the excellent performance, the price and physical volume of these systems is high. However, note that the main source of both volume and price is because of two critical components, optical filter and lens. The third type is to eliminate the optical filter. Even for non-imaging fluorescent sensors, which are relatively small, the volume of optical filters dominates overall system size. Moreover, if more than one fluorescence emission wavelength needs to be imaged the number of filters increases proportionally. The third type of system uses the time-domain instead of wavelength domain to distinguish between fluorescent signals. One example of this type reported is fluorescence lifetime imaging microscopy and 3D imaging [14]. Even such systems no longer require filters, the focusing optics are essential, which are large and expensive. As a result, this is still not adequate with implantable or portable devices. The fourth type of fluorescent detection scheme, multi-photon excitation, is the state of the art instrumentation, and is the bulkiest but also the highest performance. The fourth type utilizes multiple long wavelength excitation sources and scanning to deliver an accumulated excitation energy on target. This type of fluorescent sensing has various advantages: it enables 3D imaging with high signal-to-background and signal-to-noise ratio, and the eliminates needs for filters. However, scanning reduces the frame rate as it uses diffraction grating to multiplex the excitation illumination [29]. Another example shares the same principle but differs in its realization. This example, instead using mechanical scanning of two photon excitation, leverages structural illumination via compressive sampling algorithm to trade off the z-dimension of the third type with lens-less high resolution 2D fluorescent
However, the volume of the instrumentation still cannot be reduced significantly with the use of prism in exchange of focusing optics. And also note that the absorption filter is still needed to provide enough rejection. Based on these types, it is apparent that even with the promises that these types to resolve few challenges, in the end, there is no complete solution yet to eliminate needs for both lens and optical filters. The summary of comparison among the examples from these types are shown in Table 1.1.

### 1.1 Motivation

The use of fluorescent imaging has become widespread in neurobiology due to the increased S/N ratio that is produced as compared to stained tissues. The ability to genetically limit the expression of fluorescent proteins to anatomy of interest has also created an opportunity to visualize the widespread expression widely and deeply across the brain. The development of confocal and multi-photon imaging has helped to increase the z axis resolution but still requires expansive and expensive tools in order to do accomplish these tasks. An important direction that must be undertaken is the scaled down capability to detect
fluorescent emissions without the use of complex filters and lenses. So ideally, the FLIM system needs to be small enough to be portable/implantable, and inexpensive. These requirements are in direct contrast to the most widely used fluorescent microscopes. However, this does not mean that the conventional fluorescent microscope/imagers are inferior and should be replaced. In fact, the reason for the volume and cost is usually not only because of the high performance but the board functionality. Consequently, they also support simple imaging tasks while they are focused on complex high end tasks. In an engineering perspective, this often means the system is inefficient for the use of such an instrumentation for much simpler task. On the other hand an application specific image sensor array focused on a specific task does have a promising margin to make use of the redundancy. Angle sensitive Single Photon Avalanche Diode The capability of miniaturization would allow implantable electronics capable of detecting fluorescence from targeted tissue layers as well as deep tissue Ca imaging opportunities. Traditionally, the confirmation of implanted tetrodes used during awake-behaving rodent experiments requires rodent sacrifice and staining for electrode tracks after months of data collection has already taken place. Online fluorescence detection would increase the accuracy of electrode location to targeted tissue regions during implantation. Furthermore, two-photon imaging of the intact rodent brain is still limited to mm’s beneath the skull. Visualization of neural assembly population activity with Ca imaging could be possible with deep implantable electronics.

Single photon avalanche diodes, integrated in monolithic state-of-the art semiconductor technology have been shown a great potential to enable super sensitive imaging devices that can exceed the performance of conventional pho-
todiode based imaging. These sensors provide binary sensitivity, and therefore resolve arrival time of a few photons. The monolithic SPAD has been extensively studied lately to improve the light sensitivity while also expanding in application space including time-of-flight (TOF) imaging and scientific imaging. The TOF utilizes the sub-nanosecond accuracy of SPAD on detecting single or few photons, and the SPAD also provides enhanced sensitivity for such scientific imaging that often suffers from limited signal to noise ratio owing to the photon starving environment. However, previously demonstrated SPADs lack one important characteristic of an ideal light field image sensor - angle sensitivity. We demonstrate that SPADs high sensitivity can be best combined with metal gratings much like shown in recent works in angle sensitive pixel (ASP), to make a closer-to-ideal image sensor.

The metal gratings, which are parts of conventional CMOS foundry processes, modulate the signal intensity with respect to the incident angle due to the Talbot Effect. The transfer function of angle-to-amplitude modulation can simply be approximated as a sinusoidal function of the angle. In this manuscript we demonstrate the combination of the ASP and SPAD leads to sub-nanosecond range timing accuracy and angle sensitivity at the same time. In addition, due to the structural similarity between our angle sensitive single photon avalanche diode (A-SPAD) and the classical double slit experiment, we see particle-like behavior of light when incoming photons are significantly subdued, while wave-like behavior is observed either when the incoming photons are increased or sample size is increased, hence confirming wave-particle duality of a light in a nano-scale device level.

The extra modality, in coming angle and nano second scale time resolution
also led me to focus on other CMOS compatible lens-less imaging sensor array.

1.2 Overview

Overall this dissertation includes a discussion of a variety of novel CMOS image-sensor technologies including: (1) Lens-less far-field imaging based on Fourier domain imaging using ASPs, (2) A Low noise SPAD quenching method for enhanced SNR with a power and area efficient medium size SPAD array, (3) an angle sensitive single photon avalanche diode sensor that combines ASP and SPAD, and finally (4) A-SPAD fluorescent lifetime imaging array for on-chip fluorophore localization and identification. Furthermore, in the course of SPAD and ASP design we developed a number of computer-aided-design (CAD) techniques which we also used in the design of (5) a complete custom circular symmetric image sensor for MEMS sensor calibration. (6) Finally, because lensless imaging is likely to be well suited to untethered low power applications (such as medical implants) I will introduce an on-chip power rectifier that generates high DC power output for current re-use neural recording amplifier.
2.1 Introduction to Lenseless Far-field imaging

Electronic image capture permits digital photography, as well as a variety of automatic sensor \[4,37\] and robotic \[6,44\] applications. Far-field imaging is traditionally accomplished through focusing optics (lenses or mirrors), which map incoming light based on its incident angle to a sensor plane made up of photosensitive pixels. In general, image resolution is proportional to the $2^3$ power of instrument volume (see Fig. 2.1), but technology relying on focusing permits only a limited range of this trade-off. Moreover, focusing optics require precision manufacturing and occupy significant volume, limiting the degree to which even low resolution imaging systems can be miniaturized. To minimize cost and size, an ideal micro-camera would be manufacturable in a standard planar semiconductor process and be capable of imaging without any off-chip optics.

Here we introduce such a device: the Planar Fourier Capture Array (PFCA). Our approach is to recover far-field information from an array of diverse angle-sensitive pixels (ASPs) \[45,48\]. ASPs are composed of a photodiode under two metal gratings formed by interconnected metal layers intrinsic to modern planar semiconductor processes (Fig. 2.2). The top grating generates an interference pattern at the depth of the second grating; the second grating blocks or passes
Figure 2.1: (Color online) Trade-off between imager size and image resolution. The PFCA fills in a gap left in the available resolution-size trade-off between a naked single photodiode and the smallest commercial focusing cameras [33].

Figure 2.2: (Color online) Angle-sensitive pixels. Incident light interacting with a grating composed of an upper metal layer produces an interference pattern at the depth of the second-layer grating. Light is either passed or blocked depending on the alignment of the interference pattern and the second layer. This alignment, in turn, is sensitive to changes in incident angle; the net effect is that the light passed by an ASP depends sinusoidally on incident angle.
light depending on the relative phase between the interference pattern and the second grating, which is in turn dependent on the incident angle of incoming light.

### 2.2 Theory: Fourier imaging

Individually, ASPs have a light sensitivity that varies sinusoidally with incident angle as

\[ R = I_0 (1 - m \cos(b \theta + \alpha)) F(\theta) (1 + \eta), \]  

(2.1)

where \( R \) is the readout of the ASP, \( I_0 \) is proportional to the light flux at the ASP, \( \theta \) is the incident angle along the sensitive axis, \( b \) is the angular sensitivity [designed to range from 7 to 39 see Eq. (2.2)], \( m \) is the modulation depth of the ASP [see Eq. (2.3)], \( \alpha \) is a designable phase offset, \( F(\theta) \) is a slowly varying aperture function, and \( \eta \) is multiplicative noise.

Both the grating pitch \( p \) and the depth between the gratings \( d \) influence \( b \), as follows:

\[ b = 2\pi \frac{d}{pn}, \]  

(2.2)

where \( n \) is the refractive index of the medium, in this case \( \text{SiO}_2 \). Modulation depth \( m \) is at a maximum when a self-image of the top grating is formed at the depth of the second grating; optimal depths \( d \) are therefore integer multiples of
Figure 2.3: Decomposition of natural images into Fourier components. (a) All images can be expressed as a sum of 2D Fourier basis functions [e.g., (b)] by taking the sum over all values in (c) the basis-scaled image

\[ m \text{ maximal when } d = a \left( \frac{p^2}{\lambda} \right); \quad \alpha \in \mathbb{I}. \]  \hspace{1cm} (2.3)

Additionally, the angle-sensitive axis is set by the orientation of the metal gratings. This oriented, sinusoidal response is similar to the individual components of a two-dimensional (2D) Fourier Transform, which is commonly used in image analysis \[10\].

2.3 System Design

In general, the complete 2D Fourier transform of a far-field image can be computed by taking inner products of the image with oriented sinusoids of dif-
different periodicity (Fig. 2.3). Therefore, an array of appropriately chosen ASPs can report Fourier-complete information about the far field. To accomplish Fourier completeness, we selected an ensemble of 18 combinations of $p$ and $d$ available in a 180nm complementary metal–oxide–semiconductor (CMOS) process \footnote{1} chosen to yield a collection of closely spaced values of $b$. Given minimum $p$ constraints and a discrete number of manufacturable $d$ values, there are a finite number of possible devices that satisfy Eq. (2.1). The solid circles in Fig. 2.4 plot $b$ and $p$ for devices satisfying Eq. (2.3) exactly for $\lambda = 520$nm (in vacuum) light. An array using only devices optimal for 520nm light would leave large gaps in the tiling of $b$. To better span these gaps, our ensemble of devices includes designs optimal for different wavelengths, but all fairly close to optimal devices for 520nm light (see red open circles of Fig. 2.4). We laid out $10\mu m \times 10\mu m$ devices in concentric rings around four central devices similar to those reported in \footnote{22} to capture low frequency information. Each design constitutes one ring of ASPs, with larger rings containing devices of higher $b$
Figure 2.5: (Color online) Manufactured PFCA. Concentric rings of ASPs with increasingly high sinusoidal periodicity yield a complete Fourier description of the light intensity from the far field. Slowly varying orientation is evident from the enlarged section, where schematics show different metal layers in different colors.
Figure 2.6: (Color online) Manufactured PFCA. Concentric rings of ASPs with increasingly high sinusoidal periodicity yield a complete Fourier description of the light intensity from the far field. Slowly varying orientation is evident from the enlarged section, where schematics show different metal layers in different colors.

(Fig. 2.5). The greater number of ASPs for devices with higher $b$ permits more grating orientations. This is desirable, since the number of independent Fourier components a given design must observe is proportional to $b$. Devices opposite each other have $\alpha$ values [see Eq. (2.1)] $90^\circ$ apart to capture magnitude and phase information of each Fourier component. We built two complementary arrays (with $\alpha$ values differing by $180^\circ$; otherwise identical) of $38 \times 38$ unique ASPs (2,888 sensors total) in an unmodified 180nm CMOS process (Fig. 2.5).

Fourier completeness relies on our measurements tiling Fourier space up to our Nyquist limit. As the range of allowed incident angles increases, so does the frequency resolution needed to cover all Fourier space. Specifically, the relationship between the maximum allowable eccentricity of incident light (relative to the normal to the device) $h$ and the maximum difference in $b$ between consec-
The largest gap between implemented devices (open circles of Fig. 2.4) gives a $\Delta b$ of 2.62, such that the ensemble transfer functions of either of the 1,444-sensor arrays yield overcomplete coverage of Fourier space for images with $h < 48.6^\circ$.

### 2.4 Measurement Results

To characterize the array, we presented a series of random calibration images [Fig. 2.6(a)] to the sensor using a square CRT area 20cm on a side, 22.86cm from the PFCA [2] for an $h$ of 31.7$^\circ$ at the squares corners. With this maximum eccentricity well under the limit of Fourier completeness [48.6$^\circ$; see Eq. (2.4)], the PFCAs outputs are somewhat redundant. We then used linear system identification tools [50] to reconstruct the kernel of each ASP [Fig. 2.6(c)]. Afterward, we presented various test images on the CRT [Fig. 2.7(a)], captured the ASP responses with an accumulation time of 16.7ms, and successfully reconstructed the images [Fig. 2.7(b)] up to the Nyquist limit of our sensor, set by the highest-$b$ design in our array ($b_{\text{max}} = 39$). Our effective resolution on the active square was approximately $20 \times 20$ pixels; the number of resolvable pixels scales with $b_{\text{max}}^2$. 

\[ h = \frac{180^\circ}{\sqrt{2}\Delta b}. \] (2.4)
Figure 2.7: Image reconstructions. Using the basis functions obtained in the calibration phase (Fig. 2.6), (a) the image presented was (b) reconstructed up to the Nyquist limit of our array. No off-chip optics were used.

2.5 Conclusions

We have demonstrated a PFCA: an ASP array that relates complete Fourier information (up to a maximum spatial frequency) about the far field without using focusing optics. The device is manufactured in an unmodified semiconductor process and requires no external optics or alignment. Its construction cost and resolution specifications fill gap between the smallest miniaturized cameras and single photodiodes (Fig. 2.1), making it a suitable choice for a large range of cost- and size-sensitive applications that cannot be served with focusing optical systems.
CHAPTER 3
SELF-QUENCHING, FORWARD-BIAS-RESET FOR SINGLE PHOTON AVALANCHE DETECTORS

3.1 Introduction

Many imaging applications require the ability to detect single photons with high time precision. For example, Fluorescent Lifetime Imaging Microscopy (FLIM) \cite{27} relies on being able to measure the precise time of photon emission from fluorescent molecules after a brief stimulation flash. Similarly, several 3D imaging cameras make use of a brief laser pulse and capture the time-of-arrival of reflected photons to measure distance \cite{31}. Single Photon Avalanche Detectors (SPADs) have been deployed in both of these applications to detect single photons with precise timing. An advantage of SPADs is that they can be implemented in silicon \cite{9}, deployed as arrays with on-board support circuits \cite{27,31}. Unfortunately, SPAD array designs generally require specialized, high-voltage transistors, increasing cost. Here we present a SPAD design that eliminates the need for high-voltage transistors, and so can be implemented in standard low voltage CMOS processes. SPADs are p-n junction diodes, reverse biased at a voltage higher than breakdown. At this bias, the electric field across the junction (called the multiplier region) is high enough to trigger avalanche breakdown from single carriers in the depletion region (Geiger mode). There are three main mechanisms that generate carriers in the multiplier region before avalanching: 1) Thermal generation, which generates a baseline dark count of avalanche events; 2) charges from previous avalanches trapped in defects \cite{9},
can be released, resulting in after-pulsing effects; and 3) incident light (photons) can generate carriers, which is the desired mode of operation. Various designs of monolithic SPADs have been implemented on silicon substrates [9,27]. These designs have been integrated with quenching circuits, which protect the SPADs from damage due to excessive avalanche currents. Quenching circuits can be either passive or active. Passive quenching (PQ) is simpler, while active quenching circuits (AQC) typically provide higher refresh rates and avoid after-pulsing effects by applying an adjustable hold-off-time to allow traps to clear. However, AQC have generally required high voltage transistors [53] in each pixel of a SPAD array. Here we present a new quenching technique, called Self-Quenching with Forward-Bias Reset (SQFBR) which provides synchronous reset with no high voltage devices. Additionally, SQFBR intrinsically clears stray charge to reduce after-pulsing. SQFBR can be synchronized to a light source, which is useful for applications such as FLIM and range-finding. This paper is organized as follows: Section II introduces the proposed SQFBR with SPAD structure and compares it with existing methods. Section III describes the readout circuit with low voltage CMOS devices. Section IV provides experimental results with discussion. Section V presents our conclusions.
3.2 Self-Quenching Active Reset SPAD Design

3.2.1 SPAD design

Our designed SPAD cross section is shown in Fig. 3.1. The multiplication region was built using an n-well to p+ (PMOS drain/source diffusion) junction. A surrounding p epitaxial region was included to provide a wider peripheral depletion region. This prevents avalanche multiplication in the lateral direction and at the surface, where imperfections would increase dark count and after-pulse effects [28,35]. The entire structure was built within a deep n-well on a p-type substrate. This additional n-well-p-sub junction provides three benefits: 1) a local sink for stray carriers in the n-well beneath the multiplier region; 2) isolation from substrate carriers and adjacent SPADs; and 3) the ability to dy-
namically set both terminals of the SPAD, enabling self-quenching and active reset without dedicated high-voltage transistors. SPADs can be permanently damaged by operating at avalanche voltages unless quenching is used to prevent excessive currents from developing. There are two main ways of quenching SPAD devices: PQ and AQC.

3.2.2 Passive quenching (PQ)

The simplest quenching circuit is a passive quenching circuit using a resistor \( R_Q \) in series with the SPAD (see Fig. 3.2a). The SPAD is biased beyond its breakdown voltage \( V_{BD} \) by excess bias voltage \( V_{EX} \). Before avalanching, the voltage across the SPAD is \( V_D = V_{BD} + V_{EX} \). When the avalanche occurs, avalanche current \( I_{AV} \) flows through the reverse-biased diode junction and generates a voltage drop across \( R_Q \). \( R_Q \) is chosen such that \( I_{AV} > V_{EX}/R_Q \), so that \( V_D \) falls below \( V_{BD} \) to stop the avalanche. After the avalanche, \( V_D \) returns to its original value, with a time constant of \( \tau = C_D \times R_Q \), where \( C_D \) is the effective capacitance of the SPAD. The main problem of passive quenching is that two
key parameters, refresh rate and hold-off time are tied to a single design parameter (resistance), which sets an exponential time constant. In order to stay close to or below the breakdown threshold long enough to clear traps and so avoid long sequences of after-pulses, this RC constant must be large. A large $R_Q$ in turn implies a significantly longer time to rise to the desired excess bias voltage, dramatically reducing the refresh rate. A better approach would be to hold the SPAD below threshold for a fixed time, and then switch quickly to the desired excess voltage. In order to provide this controlled hold-off time, as well as to synchronize Geiger-mode sensing with triggered light pulses, one can replace the bias across the SPAD with a pulsed (gated) voltage that periodically drops below $V_{BD}$ to reset the SPAD [9].

### 3.2.3 Active quenching circuits (AQC)

Active quenching circuits sense and feed back avalanche events to quench them (see Fig 2b). AQC{s have two advantages over PQs. 1) AQC{s have shorter recharging time constants when resetting SPADs into Geiger mode after an avalanche event. 2) Hold-off-time, the time delay between quenching and recharging, can be optimized in AQC{s to reduce the probability of after-pulsing. In the case of synchronously reset SPADs (synchronized to a stimulating pulse of light, such as in FLIM or time-of-flight), the hold-off-time and resetting are built into the time between cycles of the pulsed light source.
3.2.4 Self quenching forward-bias reset (SQFBR)

Our method dynamically drives SPADs into Geiger-mode, but still provides high quenching impedance. This is done without requiring dedicated high-voltage transistors for each SPAD (Fig. 3.2C). The key idea is to reset the SPAD by forward biasing it through its anode and charging up the floating n-well to a voltage of $V_C = V_{MAX} - V_T$, where $V_{MAX}$ is the peak voltage of the $V_{PP}$ waveform on the anode and $V_T$ is the forward bias diode voltage (Fig. 3.4). This charge is held by parasitic capacitances from readout circuitry and the n-well in which the SPAD is embedded. The anode is then driven to a negative voltage, $V_{MIN}$, and the SPAD is reverse biased with $V_D$, where $C_P$ is the parasitic capacitance associated with the n-well and readout circuits.

$$V_D \approx \frac{C_P \cdot (V_{max} - V_{min})}{C_P + C_D}$$  \hspace{1cm} (3.1)

Provided the peak-to-peak signal on the anode is large enough, this reverse bias will drive the SPAD beyond breakdown. If triggered, the avalanche process then discharges $C_P$ sufficiently to automatically quench the process (Fig. 3.3). SQFBR has the advantage of exposing only the SPAD, its n-well and AC coupling capacitor to high voltages, eliminating the need for high-voltage transistors. Furthermore, the n-well and AC coupling capacitor only see voltages in the range of $V_{MAX} - V_T > V_C > V_{MIN} + V_{BD}$, which, if $V_{MIN}$ is sufficiently negative and $C_P$ is large enough, can result in voltages of much smaller magnitude than is across the SPAD. Nonetheless, for reliability, $C_C$ and the n-well to bulk diode were carefully designed to prevent unwanted breakdown mechanisms. This ap-
proach still requires a single high-voltage signal to drive the anode. However, that signal can simultaneously drive the anodes of many SPADs, making this approach scalable to large arrays of SPADs.

### 3.3 Readout Circuit

Our readout circuit is shown in Fig. 3.3. A capacitor, $C$, couples the SPADs cathode to the sense node ($V_{SENS}$). When $V_{CTRL}$ is low (Fig. 3.3A), transistor $M_1$ pre-charges the sense node to $V_{DD}$. To detect arriving photons, $V_{CTRL}$ is set high so that the sense node is floating. An avalanche event is shown in Fig. 3.3B. The avalanche current pulls charge from $C$, lowering $V_{SENS}$. The latch changes state when $V_{SENS}$ falls below the switching threshold of the input inverter. To allow larger values of $V_{EX}$, the sensing node is diode protected against large voltages. With two control signals, $V_{CTRL}$ and $V_{CLK}$, one can control the sampling window ($T_S$). When $V_{CTRL}$ is low, $M_1$ is strong enough to hold $V_{SENS}$ at $V_{DD}$ even when the SPAD avalanches. Raising $V_{CTRL}$ sets the beginning of the sampling window. The other signal, $V_{CLK}$, clocks the D-flip-flop, which captures the state of the circuit and sets the end of the sampling window. A timing diagram of the important signals is shown in Fig. 3.4. $V_{PP}$ in Fig. 3.4 is sinusoidal, but similar function is achieved using square pulses, which were used when comparing to previous work. All FETs are standard 1.8V transistors in triple wells to isolate them from SPAD avalanche current noise and cross talk. The coupling capacitor is a comparatively large in area, and doubled as a light shield for the active circuitry.
3.4 Result & Discussion

3.4.1 Fabrication process & test setup

SPADs with SQFBR and associated digital readouts were manufactured in standard, mixed-signal 180 nm CMOS (see Fig. 3.5). Various stand-alone test structures with different size SPADs were fabricated, as well as a small array. $V_{pp}$ drove the anode with a sine or square wave with peak-to-peak voltages
Figure 3.4: Waveform vs. time (Note: (a) no avalanche, (b) avalanche in $T_S$, and (c) avalanche out of $T_S$)

ranging from 11.8V to 13.8V. SPADs were tested in the dark and under various levels of illumination (0-300 photons per SPAD/µsec). Tests were performed at room temperature.
3.4.2 Dark Count Ratio

Conventionally the dark count rate (DCR) is described with dimensions of Hz to represent counts per second in a dark environment [53]. For a synchronous design, only one event is possible per cycle, so we redefine the DCR as the probability of a thermal avalanche event during a given cycle:

\[ DCR = \frac{\text{# of Dark Counts}}{\text{# of Periods}} \]  

(3.2)

The DCR of the fabricated SPAD was 0.035 at \( V_{EX} = 1.6V \), over a 30ns sample window and square-wave \( V_{pp} \). This result is comparable to other, similar SPAD structures [28, 35, 53] once the length of the sample window is accounted for.

3.4.3 After-pulsing

Trapped charges in the depletion region can cause after-pulsing [9]. Traditional AQC's must implement a hold-off time, during which trapped charges are passively released. SQFBR, however, drives the SPAD into forward bias during the reset phase, actively removing these charges. We confirmed this benefit experimentally by comparing the probability of an after-pulse in SQFBR to the same SPAD structure with synchronous PQ. Synchronous PQ was performed by pulsing the bias across the SPAD and resistor with the same \( V_{EX} \) and pulse width as in the SQFBR. We characterized after-pulsing by computing the autocorrela-
Figure 3.5: SPAD with SQC die photo
tion of avalanche events across periods of $V_{pp}$. Counts due to thermal effects, and photo-generated events should result in a random bit sequence which has close to zero autocorrelation across successive periods ($n \neq 0$). However, since after-pulses depend on the recent occurrence of an avalanche, they will generate events with non-zero auto-correlation. Fig. 3.6 shows the autocorrelations for synchronous PQ and SQFBR. It is clear that the probability of an after-pulse is all but eliminated by using SQFBR.

3.4.4 Photon Count efficiency

SPADS with SQFBR were also tested under various levels of illumination, with various amplitudes of 3MHz AC voltage on the anode ($V_{pp}$). Dark count was measured for various peak-to-peak values of $V_{pp}$ as shown in Fig. 3.7. A green LED with dc current supply was then applied, and a commercial ambient light sensor (TEM6010FX01) was used to calibrate its intensity. The probability of a photo-generated avalanche event under these various conditions was measured in order to estimate the Photon detection efficiency (PDE), shown in Fig. 3.7.

The PDE is the ratio between the number of incoming photons over the active area of the SPAD and the number of output digital events, [53]. Note that, in order to estimate PDE (%) we subtracted the dark count from the measured count for each $V_{pp}$. The PDE of the SPAD under SQFBR is slightly higher than other SPADs built in similar processes [53] (perhaps due to reduced masking from after-pulsing), and competitive with other previous works built in CMOS.
3.4.5 Time-resolved measurements

As shown in Fig. 3.4, $V_{CTRL}$ and $V_{CLK}$ form a sampling window relative to $V_{PP}$. If these signals are also synchronized to a pulsed light stimulus, the relative timing of the pulse can be extracted by scanning the sampling window, and building up a histogram of when avalanche events occur. To measure the accuracy of these time measurements a green LED was used to generate short
We then set $V_{CTRL}$ to switch when $V_{EX} > 0$ and swept the phase of $V_{CLK}$ to build up a cumulative histogram of photon arrival times. Shifts of less than 0.7 ns in the relative timing of the LED pulse to $V_{PP}$ were easily resolved.

### 3.4.6 Scalable and resonant drive

One benefit of the SQFBR approach is that $V_{PP}$ can be shared across multiple SPADs in an array. To confirm this scalability, we built an 8 by 8 SPAD
array with a single anode voltage shared across all 64 SPADs (Fill factor: 19.4%).
When driven by $V_{pp}$ signals similar to those used for single SPADs, we observed independent avalanche events from SPADs in the array, and these events were light-responsive, indicating SPAD function. To eliminate the need for even a single high-voltage active circuit, we explored the idea of replacing the high-voltage external signal source with a lower voltage source in series with an external inductor, 22uH, resonated with a capacitor (see Fig. 3.8). This structure provided passive amplification to transform a 3V input (which could be generated using I/O transistors) into the 12.6V signal required to drive the SPADS at an optimal signal to dark count rate. This resonant drive generated outputs similar to directly driven devices.
3.5 Conclusion

We have presented a method by which SPADs can be made to self-quench and synchronously reset without using high-voltage transistors. This approach resets the SPADs by forward biasing them, which provides added benefits in after-pulse rate.
4.1 Introduction

In order to extract as complete information as possible about incident photons, an ideal sensor should be small and parallelizable to form an array for high spatial resolution, encode angular information about incident light for ray-tracing and 3-D reconstruction, resolve arrival times down to a sub-nanosecond range, and yet be able to detect even a small number of photons.

4.2 Angular Sensitivity

Recently, angle sensitive pixel (ASP) arrays have been reported \[48\] to enable light field imaging, lens-less far field imaging, computational refocusing and range finding, and on-chip 3D localization \[45, 48\]. Such versatile functionalities of ASPs are mainly based on sensors where two stacked diffraction gratings are included in each pixel which generates angular sensitivity modelled by the equation:

\[
I_{\text{out}} = I_0 (1 + m \cos(\beta \theta + \alpha)) \times F(\theta)
\]  

(4.1)
Incident light striking the top grating generates a diffraction pattern, a self-image at the Talbot depth \[38\] with a lateral offset that depends on the incident angle of the light. The bottom grating, acting as an analyzer, selectively blocks or passes the incoming diffraction patterns based on their lateral offsets, effectively modulating the light intensity based on the incident angles of the incoming light. Quantitatively, Equation (4.1) can be used to approximate the output of an ASP \[48\] where \(I_o\) is incident light intensity, \(\theta\) is incident angle, \(m\) is the modulation depth, \(\alpha\) is sampling phase, and \(\beta\) is angular sensitivity. The aperture function, \(F(\theta)\), accounts for effects such as surface reflections, which can hamper the angular sensitivity of the ASP. In order to resolve the ambiguity in incident angles and local intensities, previous work \[45, 48\] have implemented sets of 4 ASPs, where each set is responsible for one of the quadrature different phase (\(\alpha = 0^\circ, 180^\circ, 90^\circ,\) and \(270^\circ\)). Such scheme is akin to differential, quadrature modulation often used in communication systems to determine phase and amplitude of carrier signals. It is also important to note that such ASPs are built in a standard CMOS manufacturing flow with the gratings consisting of the already-existing CMOS metal layers, allowing for low-cost, scalable arrays of ASPs.

Although the ASP arrays enable novel applications such as far field lens-less imaging also known as Planar Fourier Capture Array (PFCA as described in Chapter 2) \[15\] and on-chip localization of fluorescent sources \[45\], the lack of focusing optics can degrade their signal-to-background ratios. Moreover, previous study on ASPs use conventional photodiodes to transduce photons into charged carriers, by which the temporal resolution for intensity modulation is limited. Nonetheless, in many applications, the temporal resolution on a
nanosecond scale is crucial for real-time data acquisition and such fine temporal resolution can lead to the suppression of background interferences. For instance, a fluorescent imaging requires significant rejection of short wavelength stimulus lights in order to detect longer-wavelength emissions, which is often accomplished through an optical band pass filter. However, a sub-nanosecond resolution time-gating can also enable the rejection of the high energy stimulus in the time domain without using expensive and complex optical filters, in addition to capturing more accurate real-time information.

### 4.3 Temporal Sensitivity

On the other hand, SPADs are known for their high sensitivity as they can detect single photon arrivals. This also leads to an extremely fine temporal resolution of photon arrival times down to sub-nanosecond scale. SPADs on CMOS are emerging as favorable low-cost options for time of flight (TOF) or fluorescence lifetime imaging (FLIM) owing to their scalability, integration in conventional monolithic semiconductor technology, and therefore co-integration with high computation power [13]. SPADs amplify a single or few electron-hole pairs generated by incident photons in a p-n junction. An electric field above the intrinsic breakdown voltage across the p-n junction can trigger an impact ionization which leads to an avalanche breakdown current. In order to prevent a permanent damage on the junctions of SPAD pixels, a quenching circuit is placed to lower the electric field across the junction after the avalanche [7, 12, 24, 32]. One common implementation of such quenching circuit converts the current pulse
Figure 4.1: Cross-section of proposed A-SPAD structure.

to voltage by putting a resistor in series with the SPAD, which is often implemented by a p-type MOSFET in triode region for its advantages over a passive resistor faster reset time, asymmetric resistance, and smaller footprint [41]. It should be noted that the output of a SPAD is inherently binary and probabilistic. Though such discreteness is needed to achieve an ultra-fine temporal resolution, applications exist where characterizing the probability that a photon arrivals in a given time window, relative to a stimulus, is more appropriate than measuring the presence or absence of a single event. By repeating the stimulus and measurement many times and averaging, such stochastic data can be obtained, with mean-square error inversely proportional to the number of measurements.
4.4 A-SPAD Design

In order to achieve characteristics close to an ideal light sensor, it is necessary to combine the superior sensitivity, time resolution and inherent photon counting sensitivity of CMOS compatible SPAD device with the high angular sensitivity of ASPs. Here, we propose an angle sensitive SPAD on a conventional CMOS platform by integrating ASP gratings on top of silicon SPADs with optimized readout circuitry. Fig. 4.1 shows the cross section of the angle sensitive single photon avalanche diode (A-SPAD). Two metal gratings are used for ASP where the upper grating consisting of Metal 5 generates the diffraction pattern that shifts based on the incident angle and the lower grating consisting of Metal 3, analyzer grating near the fractional Talbot depth (1/2 Zt), generates an interference patterns onto the SPAD surface. A highly doped p-type diffusion layer and a lightly doped n-type well forms a p-n junction for SPAD multiplication process. A p-well buffer is formed between an active optical N-well area and a cathode electrode n-well around p+ anode diffusion area to provide a guard-ring to minimize unwanted lateral breakdown due to near-surface defect [12, 24, 32]. All the active circuitries are optically isolated by metal layers to prevent any photovoltaic interference, and noise sensitive circuits including comparator are electrically isolated by independent p-well to minimize a charge injection through the substrate.

Unlike conventional ASP sensor designs, we have replaced the quadrature phase ($\alpha = 0^\circ, 180^\circ, 90^\circ,$ and $270^\circ$) sampling with a balanced tri-phase ($\alpha = 180^\circ, 60^\circ, -60^\circ$) sampling to maximize the sensor density. While the reconstruction of the incident angle and intensity is formulated as a two di-
Figure 4.2: Die-photo of the A-SPAD unit prototype chip with two $\beta$s (7,15)
Figure 4.3: Balanced tri-phase

- Balanced tri-phase dimensions a complex number, and therefore resolvable with three phases of the conventional quadrature sampling ($\alpha = 0^\circ, 90^\circ, 180^\circ$), balanced tri-phase provides an easier intensity detection through averaging the three phase sampling outputs (as shown in Fig. 4.3), enabling a simpler angle information extraction. Fig. 4.2 shows a micro photograph of the foundry-fabricated prototype A-SPAD unit cell, consisting of six A-SPAD pixels for three different phases ($\alpha = 180^\circ, -60^\circ, 60^\circ$) for both $x$ and $y$ direction, where a single A-SPAD pixel occupies the area of $35 \mu m \times 35 \mu m$. The readout circuitry is consisting of quenching pseudo resistor, AC coupling capacitor which also provides optical shield over non optical area, and a comparator. Two different $\beta$ values (7, 15) are implemented by using different analyzer grating depth using different metal layers such as Metal 3 or Metal 1, respectively.
4.5 Measurement

Fig. 4.4 (a) shows a measurement setup to verify the sub-nano second temporal resolution while measuring amplitude and angular information simultaneously. Two LEDs are driven with the relative time offset of 4 ns using digital delay generator (SRS Inc. DG645) and angular offset of $\pm 20^\circ$ degrees, for A-SPADs to detect. A sampling window to acquire a time histogram of average counts is shifted with 139 psec step size, which defines the temporal resolution. Figure 4.4 (b) shows a resulting amplitude response (blue curve) acquired by averaging the photon counts in all three phase outputs. Unlike conventional SPAD, (red curve) shows how the three phase outputs can be used to reconstruct the angle information.
The photon detection efficiency of the SPAD is measured at VEX = 1.2 V bias with an input light wavelength of 550 nm. A broad-band light (Oriel Apex illuminator: Newport) is first passed through a visible light band-pass filter (Melles Griot Inc.) before going through Metallic ND filter (FBS-ND10, Newport) in series with a collimator. An optical power meter (PM203, Thorlab Inc.) is used to determine an average optical power on the sensor position before the actual A-SPAD measurement. The dark counts were also measured by examining the sensor outputs while the light source is turned off. The A-SPAD resulted in the quantum efficiency is 18.7% with dark count rate of 404 Hz.

Due to the binary nature of SPAD, the signal to noise ratio (SNR) of both in-
incident light intensity and angle is a function of the detector SNR and the sample size. Input signal strength can be modulated by pre-calibrated LED light pulse with varying peak intensity while the sample size can be varied by increasing the LED pulse repetitions. Fig. 4.5 shows the change in sensor response due to the incident light intensity modulation as well as the changes in peak intensity and the number of sample size. The sub-figures (a) through (b) show an angle-sweep response under different optical powers. The range for the light peak intensities are chosen to be less than 5% of the onset of sensor non-linearity to avoid the photon pile-ups issue [24,32]. It should be noted that the sample size plays much more crucial role than intensity in optimizing the device output SNR. Although the angle sensitivity can be preserved even with low sample size, this result indicates that there is a minimum sample size to meet certain
angular resolution. This strong dependency on the sample size is not surprising as SPADs are extremely sensitive, in fact, down to a single or few photon level the photon detection using A-SPAD, therefore, can be viewed much like the classical double-slit experiment where the particle-like characteristic of photons dominates for low sample size. In addition, once the sample size becomes sufficiently large, the averaged A-SPAD sensor output converges to the conventional photo-diode based sensor output, in accordance with the law of large number.

Fig. 4.6 shows (a) FTDT simulation based on light as a wave and (b) macroscopic response recorded from a photo diode, on which (c) an averaged output of A-SPAD output converges. This agreement together with the particle-like behavior mentioned above demonstrates the waveparticle duality of the light in nanometer scale optical gratings.

4.6 Angular Resolution

The minimum angular resolution can be derived from the SNR of the binary output, which is a function of various parameters such as sensitivity, average number of incident photons, false count, and sample size. In reality, the maximum SNR is mainly limited by SPADs detection efficiency and dark-count rate, and hence, higher quantum efficiency and lower false count rate are desirable. In addition, increasing the number of sample size and optimizing the sensitivity through improving photon detection probability while minimizing dark-count and furthermore increasing the sample size can further improve the SNR, however, at the cost of lower frame rate and pixel size.
An ideal photon counter should have sensitivity down to the single photon level. However, currently available photon counters, including monolithic integrated SPAD used in this work, are limited by the jitter during the time recording and the convolution between averaged output counts and fixed sampling window, leading to less-than-desirable SNR. Moreover, the sensor can produce a false count from tunneling and thermal charge generation. An incoming photon can also be neglected due to the finite quantum efficiency of the device. Such aforementioned limitations coupled with non-linear distortion from photon pile-up [18] can result in the reduction of sensors finite photon counting ability with altered the averaged count and amplitude. Since the readout mechanism of A-SPAD is based on the modulation of the incident angle and the amplitude of the incoming photons, a finite SNR or an amplitude distortion can lead to a detection error. Therefore, the limits on SNR and amplitude distortion define an angular resolution of the system. Interestingly, the periodicity of the angular response of ASP leads to a non-uniform noise distribution, which but is a function of the incident angle. Equation (4.2) quantifies such noise sensitivity. The sensitivity, by definition, is an inverse of a derivative of the ASP response from Eq. (4.1), and a corresponding angular noise can be calculated by multiplying amplitude noise ($N_A$) and the sensitivity as shown in Eq. 3. However, since a unit A-SPAD is consisting of three phases

$$S(\theta) = \left| \frac{d\theta}{dR(\theta)} \right| = \left| \frac{1}{\beta I_0 m \sin(\beta \theta + \alpha) \times F'(\theta)} \right|$$

$$\sigma_A = \frac{I_0 m}{SNR}$$

$$\sigma_\theta = \frac{d\theta}{dR(\theta)} \times \frac{I_0 m}{SNR} = \left| \frac{1}{\beta \times SNR \times \sin(\beta \theta + \alpha) \times F'(\theta)} \right|$$
per each axis, the overall angular resolution for a given SNR is factor of $\sqrt{3}$ higher than a single SPAD pixel.

4.7 Conclusion

In conclusion, we present a light sensor that combines the merits of SPAD and ASP. The demonstrated sensor indeed exhibit high sensitivity to incoming photon, a characteristic of SPAD, as well as high angular sensitivity, a characteristic of ASP. This combinatory device extends the dimension of light sensing to a local photon arrival angle. Such extension in sensing ability provides not only a local intensity probe with sub-nanosecond temporal precision, but also a method to mitigate an existing limitation of 2D time of flight imaging based on its angular sensitivity.
CHAPTER 5
AN ON-CHIP 72×60 ANGLE-SENSITIVE SINGLE PHOTON IMAGE SENSOR ARRAY FOR LENS-LESS TIME-RESOLVED 3-D FLUORESCENCE LIFETIME IMAGING

5.1 Introduction

Fluorescent imaging has been one of the major enablers of modern biology, owing to its unique capability in resolving multiple markers (fluorophores) in a single sample in three dimensions. However, its broader implementation has been hindered by the lack of low-cost and portable fluorescent imaging system. High precision scanning/focusing optics are required to resolve fluorophores in three dimensions, and optical filters are needed to separate fluorescent emission from shorter-wavelength stimulation. Both of these capabilities have been extremely difficult to implement as a low-cost compact system. Recent progress on lens-less imaging avoids expensive, bulky microscope optics and expands the field of view, but such efforts require bright-field imaging with structured illumination \[16\] which are not suitable for fluorescent imaging.

Fig. 5.1 (a), depicts a conventional fluorescent imaging microscope with bulky optical components such as color filter and objective lens. Unfortunately, due to its size and cost, its application space is often limited to laboratories. In this work, we propose a cost-effective alternative as depicted in Fig. 5.1(b). Such system not only lowers the manufacturing cost but also miniaturizes the imaging system, while minimizing the performance trade-offs. Its application
space, hence, is far broader compared to the conventional system, exhibiting great promises for numerous applications in modern biology and medicine.

The most expensive and bulky components in the aforementioned conventional fluorescent imaging setup are stimulus light source, lens, and optical filters. Because the stimulus light source provides excitation energy, it cannot be eliminated; however excitation energy can often be supplied using an LED or a solid-state laser source at low cost. Filters can be eliminated by projecting the stimulus perpendicular to the sensor surface to avoid the sensor saturation. Nonetheless, it should be noted that the scattered stimulus light from the sample, even when the stimulus is aligned to miss the sensor array, can saturate the sensor due to the finite sensor dynamic range. In this work, we use SPAD
sensor arrays, with which various methods for time-resolved fluorescent imaging have been reported to drive SPAD sensitive (ON) only during the time of interest to avoid the saturation. Fluorophores spatial resolution can also be improved despite the scattering by incorporating additional information such as the time constant for the fluorophore emission, fluorescence lifetime ($\tau$).

As real-time imaging of fluorophores allows optical in-vivo probing of intracellular calcium in living animals, a fluorescent imaging can be used to create a recording of a neuronal activity in glial cells within neuronal circuits, promising its great potential in neurophysiology. This imaging technique in particular is known as fluorescent lifetime imaging microscopy (FLIM), where fluorophores are distinguished by their lifetimes ($\tau$s) rather than by their wavelengths (color). A previous work has demonstrated a FLIM with arrays of SPADs in a standard CMOS with standard microscope optics [43]. A lens-less FLIM with SPAD arrays has also been demonstrated [13], but it didn't have the ability to resolve in 3-D. Unfortunately, due to the diffusive nature inherent in on-chip imaging, spatial resolution based only on the intensity map without focusing optics is inevitably inferior to that of conventional systems. However, light-field imaging based on computational optics has shown promising results to mitigate such drawbacks by utilizing computational re-focusing [47], lens-less far-field imaging [15] and on-chip imaging [47]. Angle Sensitive Pixel (ASP) in specific uses a set of two gratings to modulate the local light intensity to create an incident angle dependent diffraction patterns. Moreover, the ASP structure is compatible with the standard CMOS process and it does not require any post-processing steps that are often required for adding micro-lens [26].
This work combines the area-efficient time-gated SPADs with the CMOS-compatible integrated optical structure, ASP, to replace expensive and wavelength-dependent optical filter turrets. By utilizing the time-resolved SPAD imaging, this structure eliminates the conventional focusing optics while adding two extra dimensionality incident angles ($\theta$, $\phi$), often referred to as light field imaging [47], to not only enable fine 2-D on-chip imaging but also expand its spatial resolution from 2-D to 3-D. We present a $72 \times 60$, angle-sensitive SPAD (A-SPAD) array fabricated in a conventional 180nm CMOS process. The pixels temporally reject high-powered UV stimulus [24] while mapping the lifetimes of different fluorescent sources in 3-D.

5.2 Angle-sensitive Single Photon Avalanche Diode Design

Fig. 5.2 shows a cross-section of an A-SPAD structure. The SPAD is formed by the junction between the P+ implant and the N-Well where the P-epi layer is used as a guard ring [24, 43]. The two sets of metal gratings are implemented using the conventional CMOS metal stacks, arranged to modulate light based on incident angle [46] to provide the angle sensitivity. As the incoming light strikes the top grating, an incident angle dependent diffraction pattern is generated. Then, the second grating, analyzer grating, selectively blocks or passes the diffraction pattern. The response of an angle-sensitive pixel to light of intensity $I_0$ and incident angle $\theta$ is proportional to $I_0 \cdot (1 + m \cos(\beta \theta + \alpha))$, where $m$, $\beta$ and $\alpha$ are design parameters dictated by the geometry of the gratings. In order to provide the maximum 3-D information, our design uses two angular
frequencies ($\beta = 8, 15$), three phases ($\alpha = -180^\circ, -60^\circ, 60^\circ$) and two orientations of A-SPAD to capture the incident angle to reconfigure the light field image. In addition, the local intensity change over a very short period of time resolution (sub nano-second) provides another extra information to achieve the lens-less imaging.

Fig. 5.3 shows the circuitry of a single pixel which contains an A-SPAD with time-gated passive quenching (PQ) using transistor $M_1$, an AC-coupled comparator to detect photon arrival, and a 10-bit ripple counter. For the time-gated PQ in FLIM, every SPADs anode is connected to a shared signal, VDRIVE, which drives the SPADs beyond their breakdown voltage ($V_{BD}$) into the Geiger mode (ON). The $V_{DRIVE}$ is synchronized to the stimulus light source such that the SPAD is not sensitive until UV excitation, to provide replacement for wavelength spe-
Specific filters. A detection of a photon results in a reverse bias drop across the SPAD which then couples through a MIM capacitor ($C_C$) to the $V_{SENSE}$ node, where the $V_{SENSE}$ has high impedance.

Despite the strict area limitation of the in-pixel circuits in order to maximize the fill factor, a differential comparator, not an inverter, is utilized for pulse detection. This also allows an independent control of the detection threshold ($V_{TH}$) and therefore provides an extra degree of freedom to control excessive drive voltage ($V_{EX}$), which defines the overall sensor sensitivity as well as the binary amplitude. That extra degree of freedom also compensates for the PVT variations across the chip. Fig. 5.4 shows the schematic for the comparator. Though a minimum area should be used for the comparator, it still has to make fast decisions while tolerating the variation across the large-scale imaging array. For instance, the lower limit of the input pair device size is designed to suppress the threshold voltage variation to a negligible amount ($\sigma < 4.76mV$). Furthermore, a positive feedback (latch) load with reset switch with a high gain,
Figure 5.4: Schematic of an in-pixel differential counter

which reduces the timing uncertainty, by enabling the decision process.

In order to create a FLIM time histogram, we repeat the laser excitation and measurement cycles, counting and averaging the photon detection during the detection window. By shifting the detection window relative to the laser pulse, the resulting photon count creates an average sum of light intensity for the time window, where the count accumulation within the detection window can be interpreted as an integration of probability density function (See Fig. 5.6).
The detection window is formed with two global timing references. The start of detection window is applied to the gate of p-FET (M1) reset switch. When it is high $V_{SENSE}$ node has a high impedance, therefore stores the AC coupled avalanche induced voltage drop from the SPAD. The end of the window is defined by the comparator reset (comparison). Therefore the combination of two global clock edges defines the detection window.

To increase the frame rate and to provide sufficient SNR, a 10 bit asynchronous ripple counter integrates output of the comparator over repeated mul-

Figure 5.5: Repeated time-resolved fluorescent imaging timing diagram

Figure 5.6: Schematic of 10 bit asynchronous in-pixel ripple counter.
multiple stimulus (Fig. 5.5). The counter (See Fig. 5.7) consists of tri-state inverters with cascoded asynchronous reset switches and output buffers. The fixed window accumulation method significantly reduces the overall data rate compared to a time interval counter (TDC). Especially for moderate or low SNR imaging environment acquiring an accurate time-histogram, the measurement can demand a large number of averaging. The high on-to-off chip data bandwidth has been well known for a large scale SPAD array. A typical fluorescent lifetime timing histogram estimation can be acquired with 500 average samples [11]. Moreover, an average activity should be controlled to be less than 1 ~ 5% to avoid a non-linearity due to a photon pile-up [18] and a local saturation/desensitization. In fact, simple multiplication estimates over 50,000 measurements for a single histogram. Unfortunately, the ASP gratings exacerbates the situation as approximately 80% of optical power reflects from the ASP structure. Thus 100,000 samples of averaging would be needed. If ev-
ery measurement of a TDC is to be exported off-chip with a large scale array, the data bandwidth requirement can be impractical without use of on-chip data compression [11].

To solve this issue, an on-chip data process such as a decimation or compression for a data reduction is necessary [11]. However, the possibility of such a data compression or decimation implies an inherent redundancy in data. A pixel level averaging, on the other hand, can be a simpler yet efficient alternative for A-SPAD array with lower signal and extra information.

A layout of two A-SPAD pixels are shown in Fig. 5.8. The reported highest fill factor of SPAD pixels implementing either TDC or counter with sufficient bits is possible through the simple design that is dominated by the 10bit ripple counter and SPADs.
Figure 5.9: (a) System architecture of a 72 x 60 A-SPAD array and (b) a brief summary of the system.
Fig. 5.9 shows the architecture of the sensor system. The system synchronizes VDRIVE, an off-chip UV stimulus to internal delay lines which in turn control the detection window. Other supporting circuits around the A-SPAD imaging core include a column and row decoder for pixel access, data MUX for a readout, SPI/ROM for serial communication to off-chip FPGA, three delay-lines, and a system state machine.

This simple system architecture allows an overall robustness; during three-month of continuous testing the system reliably operated without a system reset. The median dark count probability was 1.96% ($\sigma = 0.47\%$) for a 100ns window. The photon detection probability was measured 2.72% at 540nm with $V_{EX} = 1.2V$. This is degradation compared to other reported SPADs due to light blocked by the ASP gratings. Including a local comparator and 10bit ripple counter, the pixel pitch is 35m x 35m. The system achieves a fill-factor of 14.4% across the active area of the array, corresponding to 9.6% including support circuits and pads. The system consumes 83.8mW in a dark condition.
The TDC-less measurement utilizing a global timing reference is similar to a ramp generator of a slope ADC. The open-loop delay-line utilizes a ring oscillator instead of a long delay line. The schematic of the digital-to-time converter (DTC) for the detection window reference is shown in Fig. 5.10. A 10-bit time resolution delay line provides synchronous and digitally controlled pulses to form a detection window within a small area. For synchronous open-loop operation, single reference clock is provided by off-chip. Both positive edge and negative edge of the reference clock are extracted by an edge detector and each initiates a positive edge ring oscillator or a negative edge ring oscillator. Each delay line has a coarse control (5-MSBs) that selects a number of cycles around the oscillator that consists of 16 differential delay units. The intermediate nodes between delay units are connected to thermometer-coded multiplexer to provide a fine control (LSBs). By sharing the delay line a monotonic code to delay correspondence is realized. The delay unit is a current mode differential XOR logic gate. We use a current mode logic unit delay for finer LSB (<80 psec) and XOR gate is used to reset the oscillator to a known state (See Fig. 5.11).

Figure 5.10: Schematic of high dynamic range, area efficient DTC.

A data readout of SPAD in an array can be classified into two types: a single-
bit type with a single bit memory or a shared off-pixel TDC or a multi-bit type that has an in-pixel TDC or an accumulator. The first type is efficient for a small scale array but drastically decreases the frame rate as a number of pixel scales. The second type, however, is well suited for a scalable array yet still requires high data bandwidth as explained earlier. In the rolling shutter mode, accumulation occurs synchronously with the global detection window. The number of accumulation of each pixel can be controlled by the time delay between memory reset and read. Unlike conventional imager the A-SPAD FLIM system acquires additional dimension, time, to measure the fluorescent lifetime by recording the histogram. The histogram, as mentioned earlier in Fig. 5.6, can be acquired by sliding the detection window and accumulation. Because consecutive frames accumulate with sliding detection windows, two detection windows are necessary to ensure continuous readout and accumulation.

We use a rolling shutter scheme with two global detection windows in conjunction with an in-pixel clock pointer, a single bit SRAM, to select the on-going
Figure 5.12: (a) SPAD bias network impedance and Local decoupling capacitors (b) Simulated SPAD reverse bias voltage (red: avalanche pixel, gray: neighbor pixel).

detection window. A pixel readout is followed by a pixel reset: counter memory reset.

5.3 Large Scale SPAD Array Considerations: Desensitization

A sensitivity, or a detection efficiency, of a photon detector (SPAD) is one of the most important performance factors. Numerous studies report enhanced the quantum efficiency or/and higher fill factor. However, the sensitivity measurement is often performed in a device level measurement and underestimates
various challenges that can occur in a large scale imaging array. An avalanche current pulse of SPAD induces a binary response that contains a sharp edge. In standard CMOS process, high voltage bias can have a limited access to passive devices. The SPADs reverse bias to put it under the Geiger mode (ON) has to be supplied from off-chip. However, in reality all interconnect including wire-bond ($L_{bond}$) and on-chip interconnect introduces a finite impedance, especially high impedance and resonance due to the high frequency component of the avalanche current. Despite the use of a sufficient ground and $V_{drive}$ interpolation of wire-bond (6 pairs) to stabilize the reverse bias, local and global fluctuation occurs with a neighboring avalanche activity, and this can modulate the array sensitivity which will degrade image quality.

To evaluate the effect of the internal bias line impedance to an avalanche count of a single pixel in the array, a simple lumped pixel model with simplified SPAD model is used, Fig. 5.12(a). Integrated SPAD arrays can experience transient drops in bias voltage when one or more SPADs are triggered in a short time window. The combined avalanche current through finite bond-wire impedance reduces VEX, limiting the number of photons that can be detected in a short period of time. The reverse bias (VEX) variation of the SPAD with an avalanche at 0.5 nsec is depicted in Fig. 5.12(b). More importantly the effect of local voltage drop from the current couples to an adjacent SPAD reverse bias. Since the change in reverse bias corresponds to a time varying sensitivity, it can cause distortion or saturation and as a result it desensitizes the neighboring pixels. To mitigate this effect, aside from six bond-wires in parallel carry SPAD bias, in-pixel AC decoupling capacitors ($C_{AC} = 375 fF/SPAD$) absorb a portion of the avalanche current. Thus the amplitude of the VEX drop reduces by a factor of
2.5 as shown in [5.12](b). The local bias decoupling capacitor and AC coupling capacitor that couples voltage drop to $V_{S\text{ESE}}$ node also serve as an optical shield of the active circuit area to prevent any undesirable photo voltaic effect, especially leakage of tri-stated in-pixel counter value.

### 5.4 Measurement

We fabricated the proposed A-SPAD FLIM IC in a standard 180-nm CMOS process, and a micro-photograph of the chip is shown in Fig. [5.13](b). The imaging IC was serially interfaced with a FPGA (Altera: Cyclone II) for SPI control, each of 20 LVDS ($2 \times 10$ bit) channels are connected to a high speed DAQ (NI PXIe-6555) and the time histogram is generated by computer software.

By rotating the sensor surface relative to the collimated beam of light generated by a green LED ($\lambda = 540$nm), we measured the average count of different phases of the three A-SPAD designs to changes in incident angle. Fig. [4.6](c) shows measured angle sensitivity curves for three of the A-SPADs with $\beta = 8, 15$ and phases $\alpha = 60^\circ, 180^\circ, -60^\circ$. The responses are very similar to the FDTD simulation result, wave based prediction, of ASP responses. Two different $\beta$s, low periodicity (small inter-grating spacing) and high periodicity (large inter-grating spacing), are shown in Fig. [4.6](c). The acquired angle sensitive responses contain the combination of local intensity and incident angle information.
The result of the A-SPAD pixel average count (Fig. 4.6 (c)) for different incident angles are shown and compared to the FTDT simulation result (Fig. 4.6 (a-b)). The FTDT result comes from a classic wave characteristic and shows an agreement to the accumulated sum of a single/few photon average. This confirms the particle and wave duality and also verifies an extra dimension of such a type of sensor that is capable of detecting 5 dimensions: X, Y, φ, θ, and time.
5.5 Digital to Time Converter Static Performance

Each of the three global DTCs is independently characterized by a 20MHz reference signal which is applied to the input of a function generator (Tektronix AFG3102). The input signal and delayed output signal from the DTC are both applied to Universal Counter/Timer, 350MHz (Agilent Technologies 53230A) to characterize the performance.

In order to measure the DTC linearity we swept the 10bit delay code from a FPGA to test mode I/O pads. Two off-chip regulators are used, first, for lower noise power supply and, second, for the lower logic level regulation of the current mode logic biasing. The tail current and p-FET load voltage bias were adjusted to have LSB delay of 72 psec with $V_{\text{LOW}} = 0.2$V. A 13.2 psec RMS jitter was measured (the average over the full dynamic range) while consuming 24.1mW from 1.8V supply voltage. We collected and averaged 10 measurements for each digital delay code. The results of the DTC linearity measurements are shown in Fig. 5.14. Fig. 5.14(a) shows the DNL and (b) shows the INL of the DTC. The maximum measured DNL is less than 0.54 LSB and INL is less than 1.27 LSB. In the figure, the INL has a jump in the middle of output codes. This matches with the result of a simulation in which parasitic capacitance induced by the layout routing presents. Therefore, there is a possibility that the jump is responsible for this inference.

The measurement setup used to evaluate the proposed A-SPAD array for on-chip fluorescence lifetime 3-D image is shown in Fig. 5.15. The A-SPAD sensor array was tested with a UV ($\lambda = 385$nm) stimulus source (DeltaDiode,
Figure 5.14: (a) DNL of the DTC (b) INL of the DTC

Figure 5.15: On-chip FLIM measurement setup.
HORIBA Scientific) with an average optical power of 5mW/mm². Two clusters of fluorophores, one of Muscimol Bodipy TMR-X conjugate and the other of Fluoresbrite YG microspheres, were positioned using a micro-manipulator and transparent wire (Sutter Instrument) and were illuminated by the UV source from the side. The fluorophores were stimulated by 50ps wide pulses at a rate of 2MHz while the 100ns detection window was stepped in 139ps segments.

For the localization of the two fluorophores, we placed two irregularly shaped fluorescent voxels suspended to a transparent wire with a sphere of diameter approximately 100 um.

Fig. 5.17 shows the detailed cartoon of the fluorescence lifetime imaging setup. To extract the angular response of the ASPs, we subtracted complementary (differential) phase outputs. Otherwise, we added them to extract the amplitude (intensity) response as shown in Fig. 5.16(a). Fig. 5.16(b) shows the array-wide angular response in a raw sub-image output of three different A-SPADs formats.

Localization was performed using the difference-and-normalize approach. Since overall intensity information is normalized, all of our results rely exclusively on the measured angular information. Suppressing intensity information provides better insensitivity to measurement artifacts such as fixed pattern noise. Furthermore, the fact that ignoring intensity information does not cause degradation our ability to localize sources indicates that there is more information in angular information than in intensity information. 3-D Reconstruction was performed by extracting the sparse set of source locations and intensities.
Figure 5.16: (a) Computing intensity & angle (b) sub-images for each A-SPAD type.
Figure 5.17: Concept of lens-less, time-resolved 3-D FLIM using two different fluorophores with different lifetimes.
Figure 5.18: 3-D reconstruction with ASP

(X) that best explain the A-SPAD outputs (Y). When an output of the array is y and a projected location is x, the expected output of the array for a given mesh position \( A_{ij} \) is definitive, so it can be calculated or calibrated. However, due to the mismatch between the dimension of two vectors, x and y, response matrix A is not square and, therefore, not invertible. Fortunately, localizing fluorophore in 3-D can be formulated as a quadratic program \( \| y - A\hat{x} \| \) when the amount of fluorophore clusters are limited to a low number.

Additionally, the combination of ASP and fine time resolution, a difference in a lifetime of the fluorophore enables the sparser solution and, in turn, easier reconstruction. For each time step (detection window) the dominant spare solution in 3-D location is estimated by applying a threshold for the back correlation. The intensity of the two voxels at the peak response (See Fig. 5.19)
Figure 5.19: Example snap-shot frames for reconstructing the fluorophores.

(b) and (d)) was then extracted and plotted versus time, showing two distinct temporal signatures (Fig. 5.20). These temporal responses correspond to the expected exponential decay of the two fluorophores, but they are saturated due to array-wide/local desensitization. The array sensitivity was adjusted to meet both reconstruction and lifetime estimation SNR requirements for the Muscimol Bodipy TMR-X conjugate (Red voxel) due to a difference in the stimulus absorption efficiency. Unfortunately, the minimum power level requirement for the TMR-X is excessive for the Fluoresbrite YG microspheres (Green voxel), resulting in saturating the array. This can be alleviated by simply adding an additional excitation source where each of fluorophore types can be more efficiently stimulated.

With the combination of lifetime signature and the pre-localized 3-D position, the system can localize distinct fluorescent sources in 3-D without lenses or filters and distinguish between these sources based on lifetime, as shown in Fig. 5.21.
5.6 Conclusion

Apart from the powerful conventional fluorescent microscopy, there still is a significant need for low cost and compact fluorescent imagers that can potentially become implantable devices. Heading towards the goal, here we have built a first reported 3-D on-chip fluorescence lifetime imaging system without the use of any lenses nor filter. To eliminate two critical optical components that are crucial to the conventional FLIM system, focusing optics and wavelength selective optical filters are replaced with CMOS compatible optical structure (ASP) and circuit techniques. We propose to combine SPADs high sensitivity...
Figure 5.21: 3-D reconstruction of two groups of fluorophores with metal gratings much like the recent works in angle sensitive pixel (ASP), to make an image sensor closer to the ideal. The simple system architecture is optimized for robustness and high fill-factor with uniform sensitivity. The integrated image sensor array contains 4320 A-SPADs with in-pixel circuits to record the photon detection. The LSB of the global time reference, DTC, is 71 psec, the average over the full 10bit dynamic range where the measured RMS jitter is 13.2psec, while consuming an average power of 24mW. The total power consumption of the whole IC is 83.8mW and other performance highlights are shown in Table 5.1.
CHAPTER 6

EFFICIENT IMAGING: A HIGH-SPEED POLAR-SYMMETRIC IMAGER FOR REAL-TIME CALIBRATION OF ROTATIONAL INERTIAL SENSORS

This section presents a high-speed (> 1fps), circular, CMOS imaging array for contact-less, optical measurement of rotating inertial sensors (published in [20]). The imager is designed for real-time optical readout and calibration of a MEMS accelerometer revolving at greater than 1000rpm. The imager uses a uniform circular arrangement of pixels to enable rapid imaging of rotational objects. Furthermore, each photodiode itself is circular to maintain uniform response throughout the entire revolution. Combining a high frame rate and a uniform response to motion, the imager can achieve sub-pixel resolution (25nm) of the displacement of microscale features. In order to avoid fixed pattern noise arising from non-uniform routing within the array we implemented a new global shutter technique that is insensitive to parasitic capacitance. To ease integration with various MEMS platforms, the system has SPI control, on-chip bias generation, sub-array imaging, and digital data read-out.

6.1 Introduction

MEMS accelerometers are a favorable alternative to GPS for autonomous navigation in GPS-denied environments. However, bias and scale-factor drift remain major obstacles for precise and long-term position tracking using MEMS accelerometers [42]. One possible solution to these time-varying, non-deterministic errors is to calibrate the inertial sensor in real-time by measuring
the displacement of a revolving accelerometer in two opposite phases. Comparing these readings differentially means that the signal from the applied force should add while the drift and offset components cancel.

Contact-based read-out from a revolving sensor is problematic, favoring contactless (optical) readout. This can be implemented by probing the accelerometer with an overhead illumination and imaging the generated diffraction pattern (Fig. 6.1). Ideally this system is contained within a small volume (<1000mm$^3$) and all the processing for calibration is performed on-chip. Such a setup requires an image sensor to track a quickly revolving diffraction pattern and sense nanometer shifts within the pattern. Traditional imagers record on a rectangular grid which is not well-suited for efficiently imaging rotating objects. The number of pixels on a rectangular grid required for tracking the object is proportional to $D^2$, where $D$ is the diameter of the circular path. Since the relevant features are along a particular circular contour of the inertial sensor, many interior and corner pixels are superfluous. This both limits the maximum frame rate and increases the output data rate. Another difficulty with using traditional imagers for real-time calibration is that the computation of angle from Cartesian coordinates requires the implementation of an inverse tangent function. This processing overhead is significant considering it must be done in real-time and on-chip.

Therefore, this work presents an ASIC imaging sensor that efficiently captures the diffraction pattern generated by a rotating inertial sensor as part of a low-power calibration system. For more efficient spatial sampling and simpler processing, this work uses an array of pixels uniformly distributed along polar
Figure 6.1: System for continuous accelerometer calibration coordinates to directly extract the angular position.

6.2 Theory

A circular geometry reduces the pixel count to be proportional to the circumference of the rotation path which, scales linearly with diameter. For instance,
imaging the circumference of a 1mm diameter ring with 5\(\mu\)m-pitch pixels with an 8-pixel wide circular array requires only 1000 pixels, while a rectangular array would require 40,000 pixels.

Since the pixels are uniformly distributed in a circular array, pixel coordinates directly provide a measure of angular position. Therefore the angular location of an imaged spot distributed across several pixels can be measured to within \(\frac{2\pi}{N}\) radians by simply considering the position of the center (brightest) pixel, where \(N\) is the number of pixels along the circumference. A much more precise estimate of angle can be computed by considering the outputs of all the pixels that span the imaged spot and interpolating. Fig. 6.2 illustrates the process of achieving sub-pixel resolution of edges as they revolve around the array. Ideally the edge can be computed with an arbitrary precision far less than the pixel size; however, the resolution will be dictated by the noise of the system.
and its dynamic range ($DR$). The spatial resolution ($\sigma_x$) is then given by:

$$\sigma_x = \frac{d}{\sqrt{n} \cdot DR} = \frac{d}{\sqrt{n} \cdot SNR}$$  \hspace{1cm} (6.1)

where $d$ is the pitch of the pixels and $n$ is the number of pixels that image the edge. Given an imager radius $r$, spatial resolution can be converted to an angular resolution by $\sigma_\theta = \frac{\sigma_x}{r}$.

The precise computation of angle also requires the response of individual pixels to be independent of location in the array. This is achieved by using disc shaped photodiodes for each pixel. Additionally, photodiode shot noise is shape-dependent and is reduced by using a circular diode. This is due to an evenly distributed electric field which reduces stress-induced leakage current \[36], [24].

6.3 System Design

6.3.1 Architecture

Fig. 6.3 shows the top-level functional diagram of the proposed system. The sensor array is organized into 4 concentric circular bands of pixels. Each band
Figure 6.3: Top-level architecture of sensor system
is 8 pixels wide with the circumferential pixel count increasing from 320 pixels for the innermost ring (R0) to 512 pixels for the outermost ring (R3). The region inside the innermost ring of photo-detectors can be hollowed out to facilitate on-axis illumination of the rotating inertial sensor.

The top-level control circuitry includes the addressing and timing controls, serial programming interface (SPI), and bias, reference, and supply voltage generation. Addressing logic selects one of four pixel bands for readout. Within each band the pixels are addressed on a polar coordinate system via angular and radial selection controls. The address controls are designed to permit highly flexible readout from any desired sector of the array. By varying the range of valid addresses operation can be limited to a particular angular window. This mode of operation is intended to do allow for extremely high frame rate operation while reducing data-handling requirements. The position and extent of the angular window can be programmed in real time via the SPI to permit object tracking. Bias and reference levels are generated on-chip to reduce the number of I/O pads and off-chip components required.

A custom place and route program was implemented to efficiently layout the circular rings, as described in Section B. Given the circular symmetry of the system, the array was assembled as 4 identical quadrants each with dedicated quad-level back-end circuits and shared top-level circuits. Every pixel output is amplified by a programmable gain amplifier (PGA) and is digitized by a SAR ADC. The ADC outputs are time-interleaved by a serializer and output on a single pad (Section C).
6.3.2 Circular Layout

Since the sensor array is intended to work with various MEMS platforms with ongoing revisions, the array design process had to be flexible to cater to rapid redesign. Custom CAD automation, similar to [51], was required for polar coordinate place and route of pixels and readout, since both design tools and CMOS fabrication process are inherently optimized for Cartesian coordinate design. We developed a custom software tool to generate circular photodiodes for various radii and to automate the placing of photodiodes, pixel readouts, image core drivers, and decoders. Furthermore, the custom placement software placed pins to guide the Cadence Virtuoso Chip Assembly Router (VCAR).

Fig. 6.4(a) shows the completed layout of two pixel rings and a row of readout circuits. The CMOS process used for the proposed sensor array has 6 metal layers, where the first two layers were used for local routing (photodiode to pixel readout), metal 3 for optical shielding to avoid optical mismatch between pixels, while layers 4 and 5 were used for global routing (addressing and global shutter operations). The top metal layer (metal 6) was not used for imaging core layout due to thickness and minimum width constraints. In order to maximize
fill-factor, we separated photo-diode and pixel readout circuits into separate bands to share the readout circuits by two neighboring bands.

6.3.3 Pixel Readout Circuitry

Fig. 6.4(b) shows the signal pathway from the photodiode to the output. In order to limit motion artifacts, the sensor uses a global shutter scheme for photodetection. In this approach, all photopixels are reset via transistor M1 simultaneously at the start of the frame. Photocurrent is integrated during the frame period after which the accumulated charge is transferred to a storage capacitor (C1) by turning on a shutter transistor (M2). Also note that the shutter transistor also selects the diode ring connected to the readout circuitry. The capacitor, implemented with an NFET, forms a part of the pixel readout circuitry and is placed in a separate band from the diodes. Routing from photodiode to the readout circuitry varies from pixel to pixel, resulting in fixed pattern noise from the mismatch between the diode capacitance and the storage capacitor ratio. To address this problem we implemented a second global shutter scheme that is insensitive to the capacitance on the photodiode node. In this mode, the reset switch (M1) is unused. We hold the transfer switch (M2) at a voltage $V_{\text{shutter}} < V_{\text{DD}}$. This charges up the diode to a reset voltage equal to $V_{\text{shutter}} - V_T$. The charge required to restore the diode to the reset level is supplied by the storage capacitor. This simultaneously reads the photo-charge onto the storage node and resets the photodiode.
6.3.4 Backend Readout Circuitry

Subsequently the storage cap is read out through a source follower (M5). To remove offsets and low frequency noise from the source follower, the array implements correlated-double sampling with the switched-capacitor PGA. On the first phase, the pixel signal is sampled onto the input capacitor and then on the second phase the feedback switch is opened and the reset level is sampled onto the input capacitor, resulting in an output voltage of \( \frac{C_a}{C_f} \cdot (V_{\text{pixel}} - V_{\text{reset}}) + V_{\text{ref}} \). The amplified signal is then digitized by a 10-bit SAR ADC. Note that the DC output level of the PGA, set by \( V_{\text{ref}} \), is adjusted to be close to the high reference of the ADC to maximize the dynamic range since \( V_{\text{pixel}} \leq V_{\text{reset}} \). The ADC uses a 5b/5b split capacitor array to reduce area and input capacitance. Outputs from all eight ADCs are serialized and then transmitted off-chip through a single pad.

6.4 Measurement Results

We implemented the image sensor in a 180nm CMOS process. A die photo of the implemented system is shown in Fig. 6.5, where (1) is the PGA and ADC, (2) is the space for on-axis illumination, (3) is the concentric pixel array rings and (4) is the digital control and bias generation. The inset is the layout of readout circuitry and circular photodiodes. Each photodiode is 7\( \mu \)m in diameter. A single band is 80\( \mu \)m-wide with the inner diameter ranging from 1.08mm for the innermost ring to 1.8mm for the outermost ring. As can be seen in the inset of Fig. 6.5, alternating rings of diodes were angularly staggered to achieve a
closely packed arrangement. The fill factor calculated from the layout is 33.5%.

The central disc within the array core was hollowed out using an LPKF PCB prototyping laser. A 100\(\mu\)m clearance width was used during the drilling process to avoid damage to the imaging array. LED back-illumination can be observed through the cavity in the die photo of Fig. 6.5.
The chip test setup used a 3.45V supply for the on-board regulators, an external current reference for bias generation, and clock signals for the digital addressing controls. At 2MHz clock frequency, corresponding to 16fps for full array readout, the chip consumed 15.5mW. Digitized images were acquired in 8 channel parallel readout mode.

An iPad with Retina display with a maximum brightness of $455\text{cd/m}^2$ approximately 1m above the array was used to generate test inputs for the sensor array. A 20mm, f1.8 Sigma lens was used to focus images from the screen on to the array. A magnification of 30:1 was used to project the screen on to the sensor’s active area, meaning 1 pixel on the display corresponded to 3$\mu$m on the sensor. Prior to testing, the display and optics were aligned to the sensor by centering a cross-hair alignment mark. Fig. 6.6(b) shows the captured responses of all four array sub-bands to the alignment mark. The circular distribution of pixels reduces the pixel count required to image a 1.5mm diameter path by a factor of 20 compared to a rectangular grid.

To measure the angular precision of the sensor, we imaged two fixed illumination spots, one of which had a slightly displaced edge. For each test image 55 consecutive frames were acquired. Fig. 6.7(a) shows the test image with the shifted edge highlighted. The spot spanned a 240$\mu$m arc on the sensor covering a 7x20 band of pixels. The measured pixel outputs along a single arc, shown in Fig. 6.7(b), show a linear pixel shift along the angular coordinate corresponding to the angle shift in the input. The weighted average of the angular coordinates of the illuminated pixels was computed to determine the precise position of the illumination. The calculated angle has a standard deviation of 32$\mu\text{rad}$ across the
55 test frames. This translates to a spatial precision of 25nm for a ring radius of 780\(\mu m\).

A comparison of low light level images taken using conventional global shutter and our diode capacitance insensitive global shutter is shown in Fig. 6.8. Fixed pattern noise was reduced by a factor of two with the capacitance insensitive global shutter enabled.

### 6.5 Conclusion

We demonstrated a new sensor architecture that uses photopixels uniformly distributed along polar coordinates to directly extract angular position. We measured a sub-pixel angular resolution of 32\(\mu rad\), corresponding to a 25nm
Figure 6.7: Angular position of spot after edge shift ($\sigma_\theta = 32\mu rad$).

Figure 6.8: Conventional (left) and capacitance-insensitive (right) reset.
spatial resolution. The system also utilized a novel global shutter technique that reduced fixed pattern noise by a factor of two. In summary, this work presented a fully-integrated CMOS imaging array for optical rotation measurement of MEMS inertial sensors.
7.1 Introduction

In order to meet an ever-increasing interest in the neural circuits has led IC designers to create a variety of RF circuits to accommodate the needs of neuroscientists. In particular, IC designers have focused their research efforts on designing systems to monitor and record neural activity from the nervous system [19], [49], [30], [21]. Though implanted monitoring systems that use direct wired connections for power and data have been widely used for their simplicity and reliability, these wired systems have practical limitations, such as potential infection or restricted movement of the subject. Naturally, the shortcoming of the fore-mentioned wired systems had led to an intensive research to realize wireless monitoring systems. Recent advances in energy scavenging using electromagnetic fields or radiative RF power for medical devices have shown promise for such wireless monitoring device

Inductively coupled power transfer has been widely used [23][40][52], especially at low frequencies (< 20MHz) where the tissue absorption of field energy is negligible [34]. Nonetheless, previous systems have lacked the ability to convert low voltage inputs (<200mV) on the inductor. Furthermore, for bio-medical monitoring device applications, it is desirable to minimize the overall device size, which, for inductively powered devices, is limited by the size of coil used.
However, input power to the implants rectifier is limited by the size of the inductor, creating a trade-off between available power and the device dimensions. Moreover, while extremely low power circuits have been demonstrated with reduced inductor size, such circuits still require hundreds of millivolts of DC voltage to function, and typical rectifiers themselves require even greater AC voltages to be able to efficiently generate the needed DC voltages. Since output voltage is also a function of coil inductance (and so, coil size), there is a need for implantable power systems which can make use of low-voltage, low-power ac signals. In this paper, we demonstrate a novel rectifier consisting of a full-wave rectifier, and cascaded self synchronous charge-pump stages to provide high voltage DC output from a small inductor which can power an energy efficient current-reuse amplifier for neural recording (Figure 7.1).

7.1.1 Energy Efficient Neural Amplifier

Since front-end amplifiers typically dominate the power budget of neural recording systems, minimizing the tradeoff between power consumption and noise in amplifiers is critical for wirelessly powered systems \[52\]. Noise efficiency factor (NEF) is a popular metric that quantifies the trade-off between noise, bias current, and bandwidth in amplifiers with a first-order roll-off and is given by:

$$NEF = V_{rms, in} \sqrt{\frac{2 \cdot I_{total}}{\pi \cdot V_T \cdot 4kT \cdot BW}}$$

(7.1)
Figure 7.1: Wireless power transfer configuration

where $I_{\text{total}}$ is the current consumed by the amplifier, BW is the 3-dB bandwidth in hertz, and $V_{\text{rms, in}}$ is the input-referred noise voltage. Effectively, NEF is the ratio of the input-referred amplifier noise to that of an ideal BJT common emitter amplifier biased with the same supply current. Practical amplifiers always have NEF > 1 since they consist of more than one noise source. Most practical amplifiers have a differential topology, implying an NEF of at least 1.4. However, many designs that achieve a low NEF (<3.0) do so with a high supply voltage (2.8V) [49]. More recent work [30] tends to focus on reducing the supply voltage (as low as 0.5V) to be more amenable to wireless power transfer. In such low voltage designs, however the NEF is higher (5.3) and is very sensitive to distortion (2% THD at 200$\mu$V$_{\text{rms}}$). Unfortunately, many of the traditional energy efficient amplifier techniques, such as low-power folded cascodes with strong inversion biasing, are not practical at such low supply voltages. Our previous work [21] introduced an orthogonal current-reuse amplifier that achieved an extremely low NEF (1.64) with good dynamic range (1% THD at 16.5mV$_{p-p}$). Achieving this performance, however, requires a moderately high supply voltage of 1.2V, which is impractical for most rectifiers with $V_{TH}$ input
Therefore, the proposed ladder rectifier nicely complements this amplifier.

### 7.1.2 Power Conversion Efficiency (PCE) and Passive Rectifier

The PCE of a rectifier is the ratio between its DC power output and AC input power, given by the following equation:

\[
\frac{P_{\text{DC\text{OUT}}}}{P_{\text{AC\text{IN}}}} = \frac{P_{\text{DC\text{OUT}}}}{P_{\text{DC\text{OUT}}} + P_{\text{LOSS}}}. \tag{7.2}
\]

The PCE for a passive rectifier, especially a charge pump type rectifier, is limited by the threshold voltage of diodes or MOS transistor equivalent and the relative voltage amplitude across the rectifier. Figure 7.2 shows two passive rectifier structures (a) a simple bridge rectifier, and (b) a Self-Vth-Cancellation rectifier (SVC). The SVC circuit of (b) uses its own DC output to minimize the loss coming from the threshold [23], however the minimum input voltage amplitude is higher than desirable (>500mV). In this work, we show an energy efficient neural amplifier powered by a deep sub-threshold turn-on voltage (<200 mV\text{\text{\tiny A}}) passive rectifier with high voltage optimized operation.
7.2 System Design

7.2.1 Ladder Charge Pump Design

Figure 7.3 (a) shows a unit rectifier used in this work. It is composed of a transistor and a capacitor between its source/drain and gate. For a sub-threshold input voltage the input impedance is very high, conducting a very small current through the switches. This, in fact, is favorable for voltage matching at the input terminal of the coil. When a positive voltage is applied to the input, sub-threshold current causes charge to accumulate on the capacitor. This causes an increase in $V_{GS}$ with every input cycle as $V_{DC}$ is driven more negative. This example half-wave rectifier can be cascaded to create even larger DC voltages. However, the input voltage swing on the subsequent units will be reduced due to a capacitive divider, lowering the conversion efficiency.
The proposed ladder rectifier can be separated into two sub-pieces where P-type and N-type MOS transistors can build DC voltage on complementary phase to double the rectified voltage and lower the number of stages cascaded from the input. Figure 7.4 shows the cascaded multi-state ladder rectifier. The Cross Coupled Rectifier (CCR) serves as a full wave rectifier and two Ladder Charge-pumps (LCP) on each sides – LCPn and LCPp – are used to generate the targeted DC output level (> 1.2V) for the amplifier. Each side of charge-pump has complementary switching devices, P/N type CMOS transistor, for pull-up and pull-down, respectively.
Figure 7.4: (a) A schematic of Multi-stage Ladder CMOS rectifier circuit, (b) Simulation of internal rectifier nodes from start-up to steady-state.

### 7.2.2 Ladder Rectifier Analysis

To understand the rectifier in a full system, two different temporal conditions need be analyzed: start-up and steady-state (constant power supply). For the start-up analysis, when input power is applied, all the switches operate under the sub-threshold condition. However, the sub-threshold current flows through the capacitor across the gate and source, thus accumulating charges. Conse-
quently, the voltages build up and the currents increase as this process repeats. Note that the gate-source capacitance serves two critical functions for this rectifier. First, it accumulates enough voltage to compensate for the threshold voltage drop in switching devices, and second, it provides the charge to transfer to higher voltage stages in LCPs. For the steady-state analysis, the conservation of charge provides a governing equation for the output and input power and can be derived as,

$$P_{OUT} = I_{OUT} \times V_{OUT}, \quad P_{IN} = I_{IN_{total}} \times V_{IN}. \quad (7.3)$$

The output current and charge for a single stage is given by

$$I_{OUT} = \frac{Q_{OUT}}{T_{period}} \quad (7.4)$$

$$Q_{OUT} = Q_{IN} - Q_{LEAK},$$

while the total input current and charge is given by

$$I_{IN_{total}} = \frac{Q_{IN_{total}}}{T_{period}}$$

$$Q_{IN_{total}} = Q_{IN} \times N_{stages} \quad (7.5)$$

$$= N_{stages} \times (Q_{OUT} + Q_{LEAK}).$$

The total output voltage is

$$V_{OUT} \approx N_{stages} \times (V_{IN} - V_{LOSS}) \quad (7.6)$$

$$V_{LOSS} \approx I_{IN} \times R_{switch},$$

therefore the input power is related to the output power by
\[ P_{IN} = I_{IN_{\text{total}}} \times V_{IN} = \frac{N_{\text{stages}} \times (Q_{OUT} + Q_{\text{LEAK}}) \times V_{IN}}{T_{\text{period}}} = P_{OUT} + (I_{\text{LEAK}} \times (V_{OUT} + N_{\text{stages}} \times V_{\text{LOSS}})) + (N_{\text{stages}} \times I_{OUT} \times V_{\text{LOSS}}). \] (7.7)

Thus, the trade off is that as the switches are wider, \( R_{\text{switch}} \) gets smaller, making \( V_{\text{LOSS}} \) lower but \( I_{\text{LEAK}} \) higher.

Figure 7.4 also shows the Cadence Spectre simulated internal nodes voltage with respect to the negative DC supply port. A 400kHz 250mV amplitude sine-wave input is applied, and all internal nodes are initially grounded \( t < 0 \).

7.2.3 Amplifier Design

Orthogonal current-reuse ([21]) employs input stacking to increase the overall \( g_m \) for a given bias current of a differential amplifier proportional to the number of layers of stacked differential pairs (Figure 7.5). Since each differential pair only requires one \( V_{DS} (=150\text{mV}) \) of voltage headroom, the increase in \( g_m \) comes at a minimal cost in voltage. The bias current is split at every layer of the stack. Note that the overall \( g_m \) of every channel is equivalent since the inputs are biased in sub-threshold \( (g_m = \frac{I_{\text{bias}}}{V_T}) \). The effective width of every channel is identical to ensure similar operating points and noise performance. Every output current of the stack is a unique, linear combination of the input voltages. To extract the
output currents of the stack, the amplifier uses 16 low current folded cascode branches to mirror each output current. Each cascode branch is then replicated four times. Reconstructing the output voltage for differential input $V_1$, for example, is performed by summing currents $i_{0-7}$ for $+V_{out1}$ in an output load and summing currents $-i_{0-7}$ for $-V_{out1}$. Ideally the contributions from the other inputs ($V_{2-4}$) will cancel, though a small amount of crosstalk will be present due to mismatch in the input stack and recombination mirrors. Since every channel is independent, they can be wrapped in capacitive feedback typical of neural amplifiers [49].

7.3 Measurement Results

Figure 7.6 (a) shows the maximum PCE of 16.58% with 184mV input voltage amplitude at 400kHz and 200kΩ load. On the other hand, (b) shows the output voltage when the ladder rectifier is connected to a load. The output DC voltage drops as load increases (decreasing resistance), and the optimal input voltage amplitude of a given load can be derived by calculating the point when $dI_{LOAD}/dV_{IN}$ is minimum. This observation is in agreement with the charge balance analysis.

In order to demonstrate the capabilities of this rectifier powering the amplifier, we performed two different tests. A synthesized (pre-acquired from an olfactory bulb slice) extracellular neural waveform was applied using an arbitrary waveform function of a function generator to one of four channels of the ampli-
fier. The comparison of the input and the output of the amplifier is depicted on Figure 7.7. As a second test, we applied two different sine-waves from the same function generator to two of four channels to verify the orthogonality of separate channels, and is shown in Figure 7.8. With an input amplitude of 265mV, the rectifier supplied 8.4175µA at 1.224V to the amplifier; roughly 2.5µW per channel including the bias current generation of the amplifier.

7.4 Conclusion

We have presented a energy-efficient neural recording system consisting of a ladder rectifier powered stacked amplifier. The ladder rectifier acheives an efficiency of 16.58% with a 184mV input voltage amplitude. The proposed rectifier could also be used as a long-range wake-up rectifier that eliminates a requirement of battery and pair it with a high efficiency active rectifier.
Figure 7.6: (a) Measured PCE, and (b) $V_{OUT\, DC}$ as a function of $V_{IN}$
Figure 7.7: (Top) Stacked amplifier recording of synthesized neural signal and (Bottom) the original neural signal.

Figure 7.8: Simultaneous waveform capture of two channels of the stacked amplifier powered by the ladder rectifier.
BIBLIOGRAPHY

[1] p is constrained by the minimum manufacturable linewidth of the metal, and d must correspond to one of four available inter-layer depths of the five metal layers suitable for making gratings.

[2] As $23 \mathrm{cm} \times 570 \mu \mathrm{m}$ (the PFCA’s size), images presented are in the far-field regime where the light field at each point in the PFCA is essentially identical.


