Titre: Title:	Miniaturized Optical Probes for Near Infrared Spectroscopy
Auteur: Author:	Sreenil Saha
Date:	2018
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Saha, S. (2018). Miniaturized Optical Probes for Near Infrared Spectroscopy [Ph.D. thesis, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/3776/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/3776/
Directeurs de recherche: Advisors:	Mohamad Sawan, & Frédéric Lesage
Programme: Program:	génie électrique

UNIVERSITÉ DE MONTRÉAL

MINIATURIZED OPTICAL PROBES FOR NEAR INFRARED SPECTROSCOPY

SREENIL SAHA
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L’OBTENTION
DU DIPLÔME DE PHILOSOPHIAE DOCTOR
(GÉNIE ÉLECTRIQUE)
DÉCEMBRE 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

MINIATURIZED OPTICAL PROBES FOR NEAR INFRARED SPECTROSCOPY

présentée par : SAHA Sreenil
en vue de l'obtention du diplôme de : Philosophiae Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. LEBLOND Frédéric, Ph. D., président
M. SAWAN Mohamad, Ph. D., membre et directeur de recherche
M. LESAGE Frédéric, Ph. D., membre et codirecteur de recherche
M. AUDET Yves, Ph. D., membre
M. NABKI Frédéric, Ph.D., membre externe

DEDICATION

To my Parents

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my advisers Professor Mohamad Sawan and Professor Frédéric Lesage, for their understanding, encouragement, and mentoring throughout my PhD studies. I also like to thank my committee members Professor Frédéric Leblond, Professor Yves Audet and Professor Frédéric Nabki for accepting to evaluate this thesis.

I would also like to thank Dr. Sascha Weyers of Fraunhofer IMS, Professor Eduoardo Charbon, Claudio Bruschini and Samuel Burri of Advanced Quantum Architecture Laboratory (AQUA) in École polytechnique fédérale de Lausanne (EPFL) for providing technical support.

Special thanks to all my friends in Molecular and Optical Imaging Laboratory and Polystim Neurotechnologies Laboratory, especially Yuankang Lu, Maxime Abran, Parikshat Sirpal, Samuel Bélanger, Sami Hached, Leila Montazeri, Mohammad Ali, Antoine Letourneau, Md. Hasanuzzaman and Elie Bou Assi, who were always there for me and made my study in Polytechnique one of the best experiences of my life. I would also like to thank Frederick Kalinain of Innotime Technologies for his support with fabrication.

I would like to gratefully acknowledge the help of one of my very special colleague, Laurent Mouden for his technical support and training in LASEM and Biostim labs. This work would have not been possible without his help and support. I am grateful to all the staff of Electrical engineering department of Polytechnique Montreal, especially Marie-Yannick Laplante, Réjean Lepage and Jean Bouchard for all their support. I am also grateful for the financial support from the Canada Research Chair on Smart Medical Devices and Optical Vascular Imaging, NSERC, ReSMiQ and CMC Microsystems.

Last but not least, I am grateful to my parents and my friends for their unconditional love, continued support and motivation throughout my studies. They always encouraged me to continue my goals and to go to the end of my ambitions.

RÉSUMÉ

L’étude de la propagation de la lumière dans des milieux hautement diffus tels que les tissus biologiques (imagerie optique diffuse) est très attrayante, car elle offre la possibilité d’explorer de manière non invasive le milieu se trouvant profondément sous la surface, et de retrouver des informations sur l’absorption (liée à la composition chimique) et sur la diffusion (liée à la microstructure). Dans la gamme spectrale 600-1000 nm, également appelée gamme proche infrarouge (NIR en anglais), l'atténuation de la lumière par le tissu biologique (eau, lipides et hémoglobine) est relativement faible, ce qui permet une pénétration de plusieurs centimètres dans le tissu. En spectroscopie proche infrarouge (NIRS en anglais), de photons sont injectés dans les tissus et le signal émis portant des informations sur les constituants tissulaires est mesuré. La mesure de très faibles signaux dans la plage de longueurs d'ondes visibles et proche infrarouge avec une résolution temporelle de l'ordre de la picoseconde s'est révélée une technique efficace pour étudier des tissus biologiques en imagerie cérébrale fonctionnelle, en mammographie optique et en imagerie moléculaire, sans parler de l'imagerie de la durée de vie de fluorescence, la spectroscopie de corrélation de fluorescence, informations quantiques et bien d’autres. NIRS dans le domaine temporel (TD en anglais) utilise une source de lumière pulsée, généralement un laser fournissant des impulsions lumineuses d'une durée de quelques dizaines de picosecondes, ainsi qu'un appareil de détection avec une résolution temporelle inférieure à la nanoseconde. Le point essentiel de ces mesures est la nécessité d’augmenter la sensibilité pour de plus grandes profondeurs d’investigation, en particulier pour l’imagerie cérébrale fonctionnelle, où la peau, le crâne et le liquide céphalo-rachidien (LCR) masquent fortement le signal cérébral.

À ce jour, l'adoption plus large de ces techniques optique non invasives de surveillance est surtout entravée par les composants traditionnels volumineux, coûteux, complexes et fragiles qui ont un impact significatif sur le coût et la dimension de l’ensemble du système. Notre objectif est de développer une sonde NIRS compacte et miniaturisée, qui peut être directement mise en contact avec l'échantillon testé pour obtenir une haute efficacité de détection des photons diffusés, sans avoir recours à des fibres et des lentilles encombrantes pour l'injection et la collection de la lumière. Le système proposé est composé de deux parties: i) une unité d’émission de lumière pulsée et ii) un module de détection à photon unique qui peut être activé et désactivé rapidement. L'unité d'émission de lumière utilisera une source laser pulsée à plus de 80 MHz avec une largeur d'impulsion de picoseconde. Cette source laser sera intégrée dans la sonde avec l'unité de détection de lumière qui comprend des détecteurs à photon unique et d'autres circuits périphériques de commande. Le couplage de source / détecteur à petite distance, de préférence sur une seule puce, a le potentiel d’accélérer considérablement la méthode traditionnelle d’imagerie cérébrale portable.

ABSTRACT

The study of light propagation into highly diffusive media like biological tissues (Diffuse Optical Imaging) is highly appealing due to the possibility to explore the medium non-invasively, deep beneath the surface and to recover information both on absorption (related to chemical composition) and on scattering (related to microstructure). In the 600–1000 nm spectral range also known as near-infrared (NIR) range, light attenuation by the biological tissue constituents (i.e. water, lipid, and hemoglobin) is relatively low and allows for penetration through several centimeters of tissue. In near-infrared spectroscopy (NIRS), a light signal is injected into the tissues and the emitted signal carrying information on tissue constituents is measured. The measurement of very faint light signals in the visible and near-infrared wavelength range with picosecond timing resolution has proven to be an effective technique to study biological tissues in functional brain imaging, optical mammography and molecular imaging, not to mention fluorescence lifetime imaging, fluorescence correlation spectroscopy, quantum information and many others. Time Domain (TD) NIRS employs a pulsed light source, typically a laser providing light pulses with duration of a few tens of picoseconds, and a detection circuit with temporal resolution in the sub-nanosecond scale. The key point of these measurements is the need to increase the sensitivity to higher penetration depths of investigation, in particular for functional brain imaging, where skin, skull, and cerebrospinal fluid (CSF) heavily mask the brain signal.

To date, the widespread adoption of the non-invasive optical monitoring techniques is mainly hampered by the traditional bulky, expensive, complex and fragile components which significantly impact the overall cost and dimension of the system. Our goal is the development of a miniaturized compact NIRS probe, that can be directly put in contact with the sample under test to obtain high diffused photon harvesting efficiency without the need for cumbersome optical fibers and lenses for light injection and collection. The proposed system is composed of two parts namely; i) pulsed light emission unit and ii) gated single-photon detection module. The light emission unit will employ a laser source pulsed at over 80MHz with picosecond pulse width generator embedded into the probe along with the light detection unit which comprises single-photon detectors integrated with other peripheral control circuitry. Short distance source and detector pairing, most preferably on a single chip has the potential to greatly expedites the traditional method of portable brain imaging.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NIRS	Near Infrared Spectroscopy
NIR	Near Infrared
TD-NIRS	Time Domain Near Infrared Spectroscopy
DTOF	Distribution of Time of flight
SPAD	Single-Photon Avalanche Diode
ns-SDD	Null/Small Source-Detector Distance
TCSPC	Time-correlated single-photon counting
SiPM	Silicon Photomultiplier
SiP	System-in-Package
SoC	System-on-Chip
VCSEL	Vertical Cavity Surface Emitting Laser
LED	Light Emitting Diode
APD	Avalanche Photodiode
APP	After Pulsing Probablity
DCR	Dark Count Rate
PDP	Photon Detection Probablity
FWHM	Full Width at Half Maximum
PQC	Passive Quenching Circuit
AQC	Active Quenching Circuit
QE	Quantum Efficiency
CMOS	Complementary Metal Oxide Semiconductor
TG	Time-Gated
STI	Shallow trench isolation

LIST OF APPENDICES

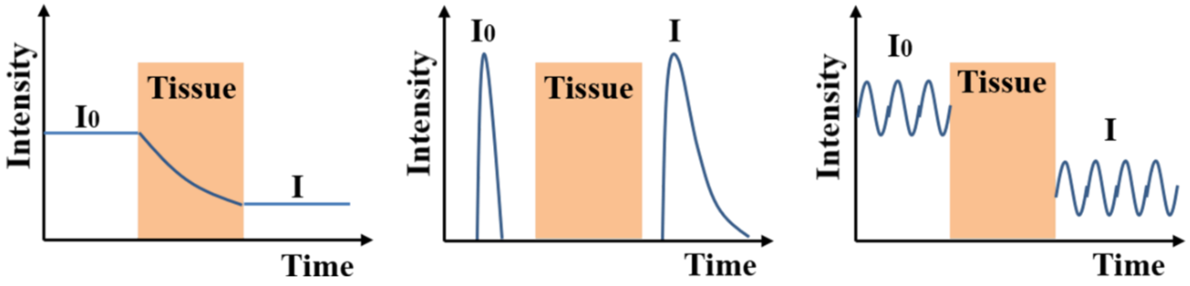
CHAPTER 1 INTRODUCTION

1.1. Background

The study of light propagation into highly scattering and diffusive medium is quite fascinating and has proven to be an effective method to explore the medium non-invasively and gather information relating to its chemical composition and microstructure [2-4]. Since the last decade researchers have shown an increasing interest in the near-infrared (NIR) spectral range (600 – 1000 nm) because the light attenuation by the biological tissue constituents (eg., water, lipid, and hemoglobin) in this wavelength range is relatively low which allows for the penetration through several centimeters of tissues (eg., muscle and brain cortex) [5-7]. The Near InfraRed spectroscopy (NIRS) technique involves the use of a point source to inject a light signal (few mW) into the media and the diffused photons or the emitted signal, carrying information on tissue constituents, are collected by a detector placed at a given distance from the source [8-10]. The analysis of the collected light, or the number of detected photons, which emulates the difference in the absorption spectra of oxygenated and deoxygenated hemoglobin, allows for the measurement and derivation of the hemodynamic parameters such as oxy-hemoglobin concentration (HbO_2), deoxy-hemoglobin concentration (HHb), total hemoglobin concentration ($tHb = HHb + HbO_2$) and blood oxygen saturation ($SO_2 = HbO_2 / tHb$). Thus, the photon migration phenomena, happening due to the complex interplay between the light absorption and light scattering, can be exploited for a wide variety of clinical applications such as functional brain imaging, molecular imaging, muscle oximetry, optical mammography, study of osteoporosis, fluorescence lifetime imaging (FLIM), fluorescence correlation spectroscopy (FCS), quantum information and many others.

A typical NIRS setup utilizes a light source such as laser diode or light-emitting diode (LED), having an emission wavelength spanning between 600 and 1000 nm, a detector, e.g. a photodiode, to detect the incoming light and flexible fiber optics for light delivery and collection from the tissues. Three different NIRS techniques can be employed to probe medium (Fig. 1.1):

(i) Continuous Wave (CW) NIRS is the simplest technique which requires the use of inexpensive laser diodes or even LEDs acting as a steady-state light source whose emission amplitude can be varied at a very low frequency (few kHz), and a highly sensitive detector such as an avalanche photodiode to detect the minute changes in the light intensity which



(a) (b) (c)

Figure 1.1. Schematic illustration of the three modalities, (a) CW irradiation permits detection of light extinction, (b) TD irradiation monitors photon's time of flight, and (c) FD reconstructions are made from attenuation and phase shifts.

reflect changes in the relative concentration of hemoglobin. But the disadvantage of CW systems is that they cannot differentiate between light absorption and scattering, hence HbO_2 and HHb concentrations cannot be computed in absolute terms.

(ii) Frequency domain (FD) NIRS is composed of a modulated light source which modulates the emitted light intensity (at frequency of the order of 100 MHz up to 1 GHz) and a detection system capable of demodulation to measure the intensity of the incoming light as well as the phase shift which corresponds to the time of flight. The advantage of FD NIRS systems is that they can be easily miniaturized, and the measurements can be used in various ways to compute absorption and scattering properties. As compared to CW-NIRS systems, FD is more expensive but less than the Time Domain (TD) systems. Its accuracy is also between TD and CW systems.

(iii) Time Domain (TD) NIRS employs a pulsed light source, typically a laser source emitting light pulses with a duration ranging from a few tens of picoseconds down to femtoseconds, and a detection module with sub-nanosecond temporal resolution, placed at a certain distance from the light injection point, to measure the photon distribution of time of flight (DTOF). The DTOF (also known as the temporal point spread function, TPSF) is essentially subjected to the same physical phenomenon which translate in delaying, broadening and attenuating. The delay is the result of the finite time taken by the injected photons to reach the detector, broadening is due to multiple scattering resulting in the different paths the photons undergo, attenuation happens because of absorption and diffusion that significantly decreases the number of detected photons. TD techniques can provide better sensitivity and penetration depth by relying on its ability to measure the photon DTOF and exploiting the timing information of scattered and re-emitted photons.

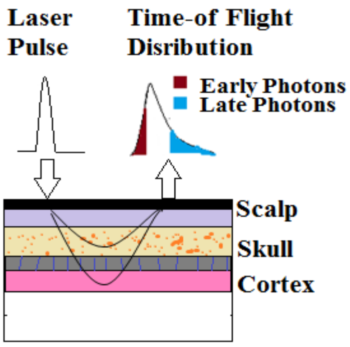


Figure 1.2. Showing the various layers of the brain, the late photons are mainly coming from the cortical region

Henceforth, TD-NIRS systems has an inherent advantage of providing a better accuracy and spatial resolution by yielding the highest amount of information about the migration of photons through tissues, clearly making it a better choice than the cheaper but inaccurate CW and FD systems. This thesis aims to demonstrate a standalone TD-NIRS system integrated with pulsed light emission and single photon detection units, both placed very close to each other preferably on a single substrate. The light emission unit comprises a pulsed laser source with pulse width of tens of picosecond and a repetition rate of over hundreds of MHz and the detection unit is based on a fast time-gated technique to detect the impinging photons in some specified time windows. The proposed module can be integrated in any experimental setup and is easily configurable by means of a simplified user interface for bidirectional communication with a remote computer.

1.2. Research Hypothesis

In TD NIRS increasing the source–detector distance [11] not only results in an increased delay and broadening of the DTOF but also decreases the number of detected photons, as observed in the case when the scattering increases. The absorption doesn’t have any substantial impact on the temporal position of the DTOF but affects both the signal intensity and the trailing edge (i.e., slope of the tail) of the DTOF. A major misconception in TD NIRS is that penetration depth is dependent on source-detector separation, like in CW NIRS. On the contrary, for an absorbing point-like inclusion embedded in a homogeneous medium, it was demonstrated that the approach based on null source-detector distance yields better spatial resolution and contrast as compared to that of large source-detector separations [12, 13]. The same observation holds true for absorption and scattering inclusions with finite dimensions or layered geometries which closely mimics some biological structures, such as head or muscle. An innovative and ground-breaking approach to TD NIRS reflectance measurements, based on null source–detector separation, exhibits some significant advantages over the classical TD NIRS approach (fixed source and detector at a large distance) by avoiding the inaccurate description of light propagation due to the photon diffusion at early time and short distance. One of the key requirement is to increase the sensitivity to higher penetration depths of investigation, in particular for functional brain imaging, where skin, skull, and cerebrospinal fluid (CSF) heavily mask the signal coming from cortical regions. Figure 1.2 presents the basics of TD NIRS, illustrating the effect of source detector distance, absorption, and reduced scattering in a diffusive medium, which is an oversimplification of the real geometry of a human head [14]. Generally, Time-correlated single-photon counting (TCSPC) is the used technique to measure fast light pulses at the single-photon level especially if coupled with silicon single-photon avalanche diodes (SPADs) [15].

Even though the null/small source-detector distance (ns-SDD) configuration is the one that has highest sensitivity to deep tissue, since detected photons propagate a shorter path, detector dynamic range limits its applicability. Even with a time-domain single-photon technique, photon pile-up with increasing illumination power limits sensitivity to late photons leading to long integration times. The detector gating technique [15-18] for TCSPC allows probing at depth with small source-detector distance by rejecting the large signal/early photons from the surface which, considering the limited dynamic range of detectors, otherwise restricts the sensitivity of the probe to the vicinity of its tip. In this work, we will exploit unique expertise in microelectronic design of integrated circuits to develop a novel integrated electronic and optical design, incorporating functionalities such as time-gated detection and pulsed-laser illumination within a single chip, a feat that remained out of reach until now. Miniaturization is possible as our probes will use detector and laser source side-by-side, leading to a very small source detector distance, which will not only allow interrogating deeper tissue volumes (cortical surface) with late time gates but also integrating early gates to monitor surface physiology (skin, bone) and remove its contributions to deep signals.

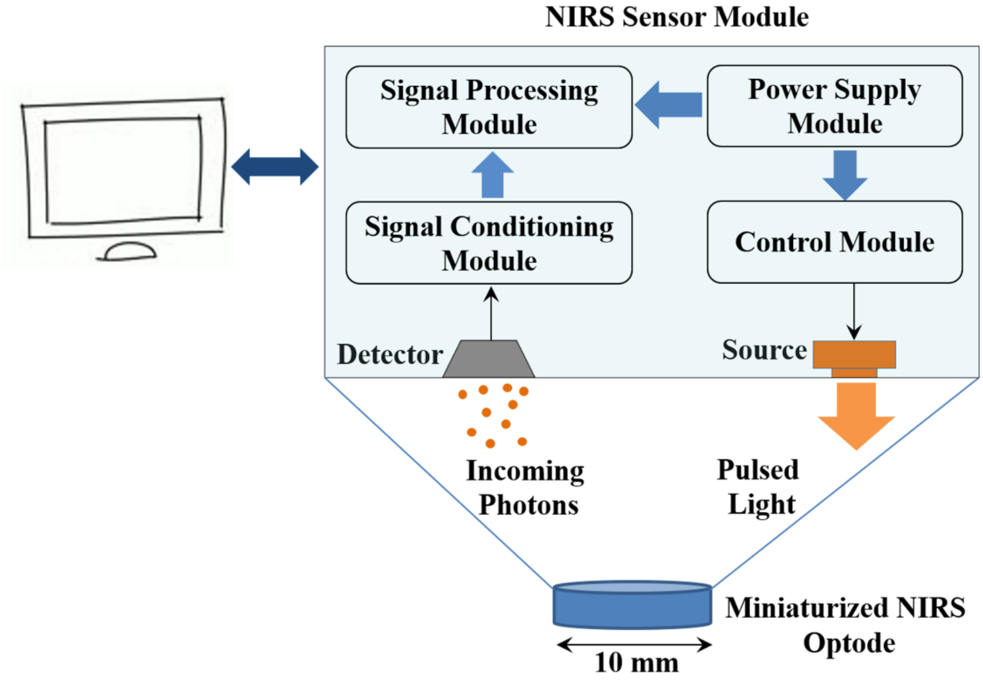


Figure 1.3. Block diagram of the Proposed NIRS probe

1.3. Research Objectives

With decreasing source detector distance, several technological issues should be taken into account and the most severe obstacle is the presence of early photons which increase at a much faster pace than the late photons and may eventually saturate the detection electronics, thus preventing the extraction of long-lived photons that carry useful information from deep structures. Hence, to fully exploit the advantages of the TD null source-detector distance approach, an efficient mechanism should be implemented to gate, or at least to reduce burst of early photons.

The long term goal for this research project is the development of a miniaturized optode device, as illustrated in Fig. 1.3, integrated with both laser source and single photon detector primarily for the purpose of portable brain imaging, having the capability of mapping and understanding the functioning of the human brain cortex and can also be effectively utilized in the applications such as optical mammography and fluorescence lifetime imaging. In order to achieve the long term goal as described earlier, the project has been divided into the following three specific objectives which are accomplished during this research:

Objective 1: Design a single photon counting setup for preliminary testing and getting accustomed with the overall optical experimental setup and performing time-gating of pulsed laser sources. For ease of testing and proof-of-concept, a compact PCB based standalone system, integrated with time-gated single photon counting module and a pulsed light emission unit, was developed using off the shelf discrete components. The module was made user-configurable via a Universal asynchronous receiver-transmitter (UART) interface and can be integrated in any optical setup.

> **Hypothesis 1:** The source-detector separation affects the investigation depth and should be reduced to increase the number of collected photons which has a positive effect on the signal-to-noise ratio and image contrast. On the other hand, the decrease in source-detector distance also increases the number of unwanted early arriving photons, so the time-gated detection scheme should be adopted to extract information from late photons.
>
> **Related Publication:**
> S. Saha, F. Lesage, and M. Sawan, "Time-resolved reflectance using short source-detector separation," in Circuits and Systems (ISCAS), 2016 IEEE International Symposium on, 2016, pp. 333-336: IEEE

Objective 2: Design a miniaturized system-in-package (SIP) optode equipped with pulsed laser source as light emitting unit and fast time-gated single photon counting module as light detection unit capable of detecting photons in specific time windows, thus enabling the measurement of very faint optical signals.

> **Hypothesis 2:** The novel time-gated detection approach has paved the way to the development of null-distance source detector optode that is easier-to-handle, convenient for cap design and could yield a simplified image reconstruction. The developed compact probe can be seen as a standalone time-resolved near-infrared spectroscopy system and has potential for widespread exploitation, especially for brain measurements where smaller optical probes eliminates the need for optical fibers for light injection and collection, making the measurement process more practical.

Related Publication:

S. Saha, Y. Lu, S. Weyers, M. Sawan, and F. Lesage, "Compact Fast Optode-based Probe for Single-Photon Counting Applications," IEEE Photonics Technology Letters, 2018.

S. Saha, F. Lesage, and M. Sawan, "High-voltage pulse generator with variable delay for ultrafast gating of single photon detector," in Circuits & Systems (LASCAS), 2016 IEEE 7th Latin American Symposium on, 2016, pp. 131-134: IEEE

S. Saha, Y. Lu, S. Weyers, F. Lesage, and M. Sawan, " Miniaturized Probe for Time-Domain Near-Infrared Spectroscopy," has been accepted for lecture presentation at the Biomedical Circuits and Systems Conference (BioCAS), 2018 IEEE, Cleveland, Ohio USA from October 17-19.

S. Saha, Y. Lu, S. Weyers, F. Lesage, and M. Sawan, “Compact Optical Probe for Time-resolved NIRS- Imaging,” under review in IEEE Transactions of Circuits and Systems I: Regular Paper.

Objective 3: Develop a compact fiber-free optical probe integrating silicon photomultiplier and light-pulsing unit, which can be directly put in contact with the sample under test without the need for cumbersome optical setup.

Hypothesis 3: The integrated circuit, assembled on a 2.5D silicon interposer platform of size 1 cm interfaced with other ancillary components such as capacitors and resistors, leads to significant optode-size reduction and the elimination of optical fibers and has the potential to change the landscape of Near Infrared Spectroscopy (NIRS) hardware commercially. The introduction of silicon photomultipliers (SiPM) would result into high diffused photon harvesting efficiency.

CHAPTER 2 LITERATURE REVIEW

This chapter first provides an overview about the various neuroimaging techniques, followed by an illustration of the the characteristics and working principle of the various building blocks/elements of the NIRS module. The various functional units presented in this chapter are viewed individually based on the literature and from the knowledge garnered during the course of this thesis work. To comprehend the basic features and adopt the best possible choice, a brief comparison between the various light emitters and single photon detectors is presented. Also, the physics behind the various photodetectors is briefly explained to understand the characterization procedures and parameters. This chapter also demonstrates how single-photon detectors can be fabricated in commercial deep-submicron complementary metal oxide semiconductor (CMOS) processes, for miniaturization with low dark noise, while comparing the various state-of-the-art structures implemented so far. The quench and recharge circuitry for the single photon detectors are briefly discussed.

Chapter organization: Section 2.1 briefly describes the various brain imaging modalities, including their strengths and limitations associated with the various techniques. Section 2.2 presents the comparison of the various light emitters. Section 2.3 presents the various multi and single photon detectors. Section 2.4 illustrates the passive and active techniques used to perform the quench and recharge operations and the SPAD fundamentals. Section 2.5 provides a short comparison of the various SPADs designed and fabricated in various standard CMOS process. Section 2.6 presents some of the recent state-of-the-art NIRS systems. Section 2.7 summarizes the chapter.

2.1. Current functional neuroimaging techniques

2.1.1. Electroencephalography

Electroencephalography (EEG) is the earliest and most utilized functional neuroimaging technique for the recording of electrical activity along the scalp produced by the firing of neurons within the brain. The abnormal neural activities are visualized by means of multiple electrodes placed on the scalp which continuously record voltage fluctuations resulting from ionic current within the neurons. EEG plays a vital role in epilepsy diagnosis [19] and in determining seizure types, and thereby in choosing anti-epileptic drug (AED) and predicting prognosis. EEG has also been validated as a reliable tool in the localization of the epileptic focus area [20]. Focal epileptiform abnormalities can be clearly distinguished from background on EEG. Extensive research has been conducted to assess the feasibility of monitoring EEG signals for automatic seizure onset detection. Furthermore, analysis of EEG signals both in the time and frequency domain enables the study of functional connectivity network in epilepsy [21].

EEG has some inherent disadvantages, for example, in scalp EEG the signal is attenuated (sometimes even cancelled) by soft tissues/bone and frequently degraded by muscle artifacts resulting in the decrease of sensitivity in the detection of epileptic events. A large area of the cortex has to be activated synchronously to generate enough potentials to be able to be detected by a scalp electrode [22]. Even invasive intracranial EEG is limited in both sampling time and area while increasing the risk for possible complications such as bleeding or infection. On major limitation of EEG monitoring is that it does not provide information on blood volume and oxygenation which could help in the evaluation of the metabolic impact of seizures.

2.1.2. Magnetoencephalography

Magnetoencephalography (MEG) is a noninvasively functional neuroimaging technique that can measure the weak magnetic fields resulting from the intracellular current flow within neurons at a high spatiotemporal resolution [23]. In contrary to scalp EEG, MEG signal is less affected by distortion from the skull and intervening soft tissue [24]. The main clinical use of MEG in epilepsy is to detect and localize the source of pathological activity in patients [25]. However, MEG often has poor sensitivity in recording paroxysmal activity within the mesial structures of the brain because of the rapidly decaying magnetic fields associated with medial discharges [26]. Furthermore, MEG is also not suitable for long-term recording, making it challenging or almost impossible to record seizures.

2.1.3. Positron Emission Tomography

Positron Emission Tomography, popularly known as PET, is an innovative and relatively new form of medical diagnostic imaging technique that has begun to be adapted in a clinical setting since the early 1990s and is used by physicians to examine biochemical changes occurring within a patient. A PET imaging can be used as a diagnostic imaging tool by the physicians to examine biochemical changes occurring within a patient. As the diseases usually affect the biochemistry of a patient and PET can be effectively utilized to detect these changes. A metabolic PET scan is used to diagnose, stage and follow-up on the treatments by examining a patient’s biochemistry. PET scans can detect diseases making it an effective diagnostic tool, thanks to its ability to study bodily functions through biochemical processes. PET imaging helps in the study of metabolic functions of a patient and hence can be used to guide biopsies and also other exploratory surgeries conducted to distinguish between benign (non-cancerous) and malignant (cancerous) tumors. It can determine how far a disease has spread and reduce the number of unnecessary surgeries performed due to incorrect diagnosis and staging data. It can also be effectively used for the diagnosis of early stages of certain neurological illnesses such as Alzheimer’s disease, epilepsy and other mental illnesses. For individuals afraid of getting infection from medical procedures, PET scan is perhaps the best option.

Regardless of its numerous advantages, PET scans also have its disadvantages. The radioactive component used during PET procedure may pose some risk. Even though the radioactive components used in PET imaging are not long lasting but there is only a limited amount of times a patient can undergo this procedure as it may not be suitable for some patients and may cause some complications if the patients are pregnant. Since it is a relatively new procedure, PET imaging is quite expensive compared to other forms of medical imaging. Also, PET scans are not offered in the majority of medical centers in the world. Another problem with PET imaging is that the sensitivity is so high that even if a patient is suffering from chemical imbalance due to diabetes and it may trigger false imaging results.

2.1.4. Magnetic Resonance Imaging

The physical principle of magnetic resonance imaging (MRI) is based on the directional magnetic field, or moment, associated with charged particles in motion. All atomic nuclei containing an odd number of protons and/or neutrons possess an intrinsic magnetic moment or precession. Because nuclei are charged particles, this precession produces a small magnetic moment. When a human body is placed in a large magnetic field, majority of hydrogen nuclei align parallel to the magnetic field.and rotate with a frequency proportional to the applied magnetic field strength. Next, an exposure to radio-frequency (RF) pulse at the same frequency as that of the precessing hydrogen nuclei results in non-invasive displacement of the nuclei.

Once the RF signal is removed, the nuclei realign themselves and return to equilibrium by emitting their own RF signal referred to as the free-induction decay (FID) response signal. The FID response signal is then measured by a conductive field coil placed around the object, then reconstructed to obtain 3D grey-scale MR images. Functional MRI (fMRI) has made it possible to examine active brain regions, apart from structural imaging of the brain, by relying on the blood oxygen level dependent (BOLD) contrast. The deoxygenated blood is paramagnetic (attracted to an external magnetic field) while the oxygenated blood is diamagnetic (repelled from an applied magnetic field) [27, 28]. Thus, the difference in magnetic susceptibility of deoxygenated and oxygenated blood helps in the derivation of the source of the BOLD contrast, which serves as a proxy measure of the neural activation elicited by a stimulus [29].

Many of the reasons that have resulted in the emergence of fMRI BOLD as a popular choice are it is non-invasive, repeatable, widely available and possesses superior spatial resolution. Furthermore, this technique provides high-resolution activation maps that is more accessible to non-expert audiences and allows for visually communicable results. However, fMRI also has a number of disadvantages such as contraindications to being in a magnetic field (e.g., metal in the body from a previous surgery), claustrophobia, restrictions on movement, the typical supine position of subjects, the requirement for compatible response equipment, and the noise produced by the scanner [30]. Furthermore, given the high costs required to maintain the equipment and employ technicians to operate the scanner, functional MRI studies can be quite expensive.

2.1.5. Functional Near Infrared spectroscopy

Functional Near Infrared spectroscopy (fNIRS) is a modern non-invasive optical imaging technique that can be used to measure changes in hemoglobin (Hb) concentrations in brain tissue. fNIRS relies on the different absorption properties of biological chromophores to measure the hemodynamic response to neural activity, unlike fMRI which relies on the paramagnetic properties of Hb [4, 28, 31]. Near infrared spectrum of the light (between 600 nm and 1000 nm) easily penetrates through biological tissues because only a few biological chromophores absorb light at these wavelengths. Changes in chromophore concentrations can be calculated according to the modified Beer-Lambert law. The absorption spectra of Hb vary in its oxygenation state and so the amount of NIR light transmitted through the tissue results in the measurement of oxygenated Hb (oxy-Hb) and deoxygenated Hb (deoxy-Hb) and information about the oxygenation-deoxygenation states can be derived. Functional NIRS has proven to be a promising technology as an independent functional neuroimaging method or for use in conjunction with MRI and fMRI and can be utilized across a wide range of areas. fNIRS has also been used to investigate the hemodynamic response during activities such as walking, exercise and other everyday tasks, which otherwise cannot be measured within a fMRI scanner. Commercial clinical fNIRS instruments were first available in 1989 but more recently advanced systems with multiple probes have enabled broader use in research and clinical practice.

Functional NIRS has a number of advantages in terms of its utility that make it an ideal choice for interrogating brain function. fNIRS is non-invasive, relatively inexpensive, and has a temporal resolution that is comparable to that of fMRI [32]. Recent technological advancements have resulted in development of portable modern fNIRS devices that are miniaturized, wireless and battery-operated which relaxes the need for strict restrictions on motion [33]. fNIRS has allowed for advancements in the study of neurocognitive processes in unconstrained environments, including studies outdoors and in various other ambulatory settings. Even though fNIRS signals are still sensitive to degradation due to motion artifacts but several several groups have elaborated on methods for real-time motion correction that makes fNIRS an attractive alternative, particularly for people where complete stillness can be extremely challenging (eg., infants). Furthermore, other significant advantages of fNIRS include insensitivity to common electrical or magnetic devices, such as hearing aids, pacemakers, or cochlear implants, and ease of integration with other neurocognitive applications. Recent technological advances have centered on the combined technological and functional uses of fNIRS and EEG or fMRI, with the goal of advancing brain-computer interface technologies.

The main limitations of fNIRS as a clinical neuroimaging tool relate to physical and technological constraints imposed by the device setup, for eg., the size of the measurable brain area strongly depends on the number of sources and detectors in the fNIRS setup. Another key limitation is that fNIRS is only able to detect NIR light that penetrates the first few centimeters of cortical tissue. fNIRS signals are also prone to be contaminated by physiology-based interference from cardiac pulsation, respiration, spontaneous changes in local vascular tone and a variety of spontaneous low frequency oscillations, hence a variety of processing methods are needed to be adopted to remove these frequencies from the data [34, 35]. Moreover, fNIRS signals also contain superficial scalp signals embedded in the, more desired and useful, ‘‘brain signal’’. An overview of strengths and limitations associated with fMRI, fNIRS and EEG is listed in Table 2.1.

Table 2.1. An overview of the strengths and limitations of fMRI, fNIRS and EEG

Modalities / Parameters	fMRI	fNIRS	EEG
Strengths	✓ Non-invasive ✓ Repeatable ✓ Widely available ✓ Superior spatial resolution ✓ Whole brain measurement (lateral surface and depth)	✓ Non-invasive ✓ Repeatable ✓ Comparable temporal resolution to fMRI ✓ Inexpensive ✓ Portable ✓ Less restriction on motion	✓ Invasive/Non-invasive ✓ Reliable tool in the localization of the epileptic focus area ✓ Focal epileptiform abnormalities can be clearly distinguished from background
Limitations	• Expensive • Strict restrictions of motion • Need for supine position • Noisy scanner • Physiological noise • Restrictions based on metal in the body	• Limited to frontal regions and surface analysis • Physiological noise (including superficial scalp signals) • Lacks anatomical information • Interpretation challenges related to multiple sources of vascular signal	• Signal attenuation by soft tissues/bone • Muscle artifacts results in the decrease of sensitivity • Large area of the cortex to be activated • Doesn’t provide information on blood volume and oxygenation

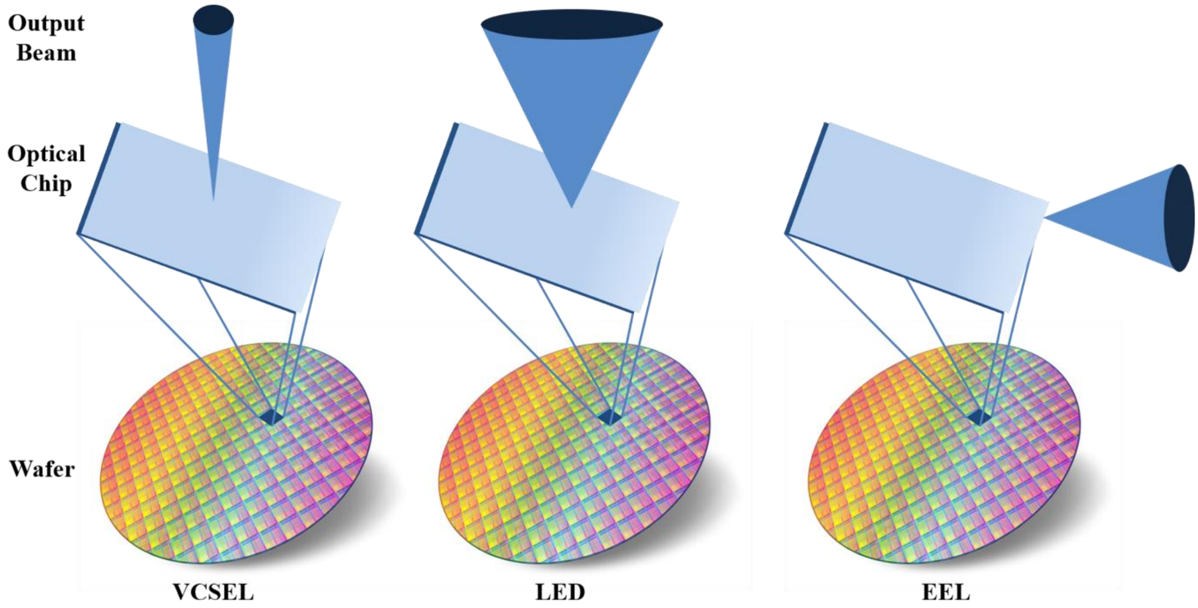


Figure 2.1. Comparison of the characteristics of three common semiconductor light emitters

2.2. Light Emitters

Over the years, semiconductor light emitters have emerged as ideal and vital light sources for a wide range of biomedical sensing/imaging applications and can be commonly classified to be of three types – vertical cavity surface emitting laser (VCSEL), light emitting diode (LED) and edge emitting laser (EEL). Among them VCSEL is a relatively recent entrant to the list of choices of semiconductor light-emitting devices. Figure 2.1 illustrates the basic differences in the characteristics of the three types of commonly used optical sources. All three types of semiconductor devices are fabricated by growing epitaxial III-V semiconductor layers (such as GaAs, AlGaAs, InP, or InGaAsP) on a GaAs or InP substrate. The wafers are processed through several photolithographic steps to define the individual devices, such as etching ridges, depositing and patterning metal contacts, or depositing and patterning dielectrics. A LED is created by incorporation a p-n junction in the epitaxially grown layers whereas laser typically requires a feedback mechanism which is implemented by means of mirrors sandwiching the gain region. In an EEL this is achieved by cleaving the wafer, and the cleaved edges form the parallel mirrors of the device which results in light emission from the cleaved edge of the wafer. Hence, EEL devices cannot be tested at the wafer level and the wafer has to be diced to test the individual lasers. In case of VCSEL, mirrors with sufficient reflectivity are grown or deposited on the wafers, sandwiching an active region which results in vertical light emission from the wafer. Both the VCSEL and LED typically emit light from the top surface of the wafer, and so they can be tested at the wafer level by probing bond pads accessible at the surface of the wafer.

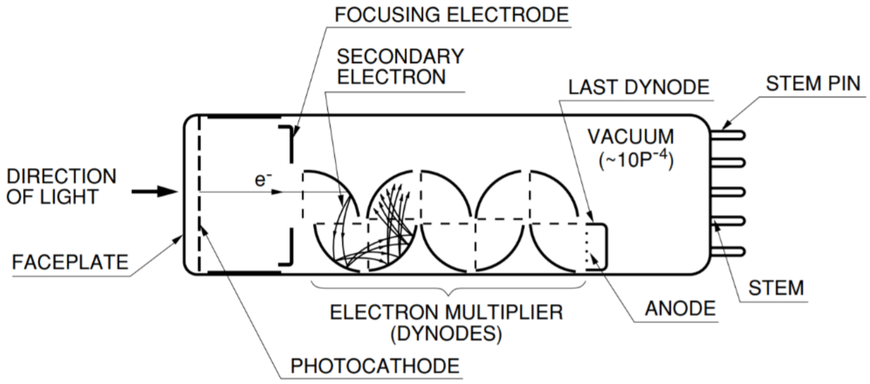


Figure 2.2. Construction of a Photomultiplier Tube (PMT) [36]

In addition to the manufacturing advantage of wafer scale testing, VCSEL also exhibits a lot of performance advantages as compared to the low cost LEDs or EELs which can be listed as follows:

1. Integrability: Semiconductor manufacturing and wafer integration of VCSELs is compatible with other optical devices such as detectors and other circuitry.
2. Scalability: VCSELs can be easily fabricated on a single chip into one and two dimensional arrays to scale optical output power matching some specific applications.
3. Packageability: VCSELs allow the use of low-cost packaging as that of LEDs or EELs to reduce cost, thus providing the option of replacing the LEDs in the existing applications. Also, VCSELs are compatible with chip-on-board technology and other custom packaging techniques that greatly simplifies system integration.
4. Low Power: Power consumption is extremely low which increases the efficiency, battery life and thermal design constraints in larger systems.
5. High efficient optical beam: It emits coherent, symmetrical and low divergent optical beam with a narrow spectral width.

Two-dimensional arrays of VCSELs can be flip-chip attached or wire-bonded to CMOS integrated circuits, thus providing major opportunities in both miniaturization and hybrid integration of multiple heterogenous chips. Optical interconnects based on CMOS/VCSEL technology have been widely proposed for high-performance switching and computing applications, and have been the focus of numerous research and development efforts. Optical interconnects based on VCSELs are already part of LAN standards, and parallel optical interconnect products are now impacting the marketplace. By using a relatively simple flip-chip bonding technique [37], it has been previously demonstrated the flip-chip bonding of a 2x10

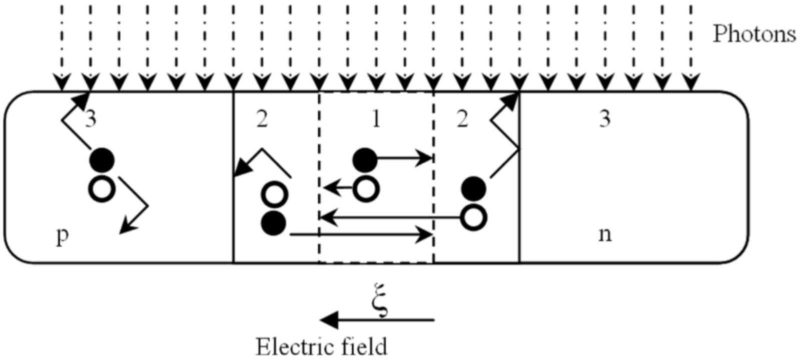


Figure 2.3. Cross-section of a p-n junction diode showing the various regions where electron-hole pairs can be generated along with their direction of flow [1]

array of 970-nm VCSELs to silicon VLSI circuits operating at 1Gb/s. The ion-implanted, index-guided VCSELs, were attached to the top-level metal on the CMOS chip and did not interfere with the operation or the layout of the underlying circuits. The direct integration of VCSELs and photodetectors to CMOS circuitry can be pursued as a key technology in the evolution of biomedical imaging.

2.3. Single and Multi-Photon Detectors

2.3.1. Photomultiplier Tube (PMT) and microchannel plates

A photomultiplier tube (PMT) is a vacuum tube consisting of an input window, a photocathode, focusing electrodes, an electron multiplier, and an anode. A PMT is based on the external photoelectric effect by which the electrons in the valence band of the material, upon absorbing the energy of the impinging photons, are emitted into the vacuum. Figure 2.2 shows the cross-section of a photomultiplier [36]. Light passes through the input window and the impinging photons excites the electrons in the photocathode. The photoelectrons emitted into the vacuum are accelerated and focussed on to the surface of the dynodes by the focussing electrodes. Considering that δ is the multiplication factor of each dynode, a PMT with n dynodes generates δ^n secondary electrons by means of secondary electron emission. The multiplied secondary electrons emitted by the last dynode are finally collected by the anode, which converts the electron cloud into an electrical signal. After their absorption the charges must be removed quickly from the anode surface for the prevention of surface charge effects. GaAsP photocathodes can achieve the peak quantum efficiency of 45% at a wavelength of 600 nm down to 30% at 900 nm.

Microchannel plate (MCP) PMT consists of an array of $10^4 – 10^7$ miniature electron multipliers, usually fabricated from a lead glass, oriented parallel to one another and fused together to form a thin disk. The channel diameters are typically in the range of 10 to 100 µm and have length to diameter ratios between 40 and 100. An electron, which enters the channel considered as a continuous dynode structure, hits the channel wall and produces the secondary electrons which are accelerated by the electric field. The accelerated secondary electrons hit the opposite wall and make additional secondary electrons. Such microchannel plates, used singly or in a cascade, allow electron multiplication factors of 10^4 -10^7 coupled with ultra-high time resolution (< 100 ps) and spatial resolution limited only by the channel dimensions and spacings. The dark current in PMTs is the small amount of current that flows in the PMT even without the presence of any photon. The quantum efficiency and dark current increases by increasing the supply voltage of PMTs. Nowadays, PMTs are one of the technologies of choice for single-photon detection, due to their maturity and wide availability. However, the disadvantages of this technology are their fragility, bulkiness, non-portability, non-scalability, and weak spectral response in near infrared (NIR) spectral range [38] .

2.3.2. PN Photodiode

A photodiode is a reverse biased p − n junction diode, having a depleted semiconductor region with a high electric field, that produces an electric current in response to the incoming photons. When a photon of ample energy strikes the diode, it generates electron-hole pair. This phenomenon is often referred to as inner photoelectric effect. If the absorption happens near to the depletion layer, then the electrons and holes are separated and transported in the opposite directions owing to the high electric field of the junction; the holes move towards the anode and the electrons towards the cathode resulting in the generation of photocurrent. Since the electric field is very high in the depletion layer, the photocarrier generation is mostly confined to this region. Figure 2.3 shows the cross-section of a p − n junction and possible locations of electron-hole pair generation. For high-speed operation the depletion region should be thin to reduce the transit time of the photo generated carriers. However, to increase the quantum efficiency (number of generated electron-hole pairs per incident photon), the depletion region should be sufficiently thick to be able to absorb the maximum amount of incident light. Therefore, there exists a trade-off between the speed of response and quantum efficiency. Furthermore, for a given semiconductor the optical absorption coefficient, denoted by α, is strongly dependent on

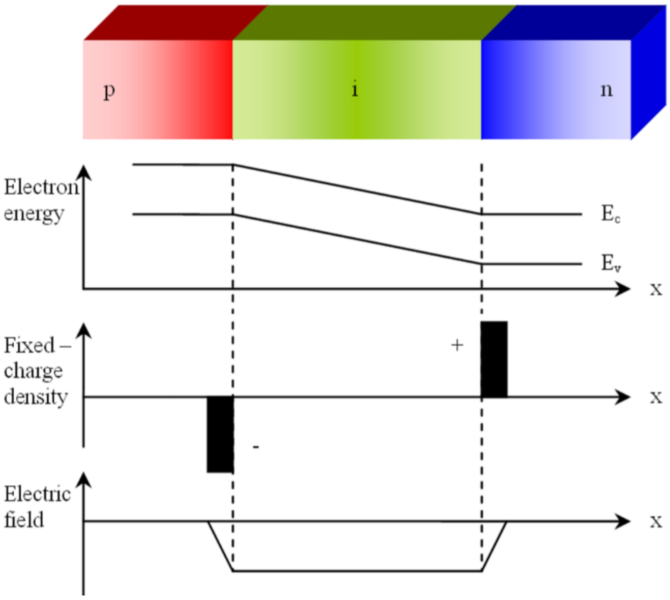


Figure 2.4. Cross-sectional view of a PIN diode along with the energy band diagram, charge distribution and electric field distribution.

the wavelength, so the wavelength range for which an appreciable photocurrent can be generated is limited. The photodiode has an I-V relation as:

$$I_d = I_s\left[\exp\left(\frac{eV}{kBT}\right) - 1\right] - I_p \tag{2.1}$$

The term I_p, the photocurrent, is proportional to the photon flux. Ge photodiodes, III-V ternary photodiodes (e.g., InGaAs) and III-V quaternary photodiodes (e.g., InGaAsP) have shown high quantum efficiencies.

2.3.3. PIN Photodiode

A p − i − n photodiode is a special case of p − n junction photodiode because the depletion-region thickness or the lightly doped intrinsic layer sandwiched between p and n layers can be tailored to optimize the quantum efficiency and frequency response. Figure 2.4 shows the schematic representation of the p-i-n photodiode. Light absorption in the semiconductor results in the creation of electron-hole pairs. Because of the high electric field across the junction, the pairs produced in the depletion region or within a diffusion length will eventually be separated and drift across the junction in opposite directions leading to current flow in the external circuit. Figure 2.4 illustrates the energy band diagram, charge distribution, and electric field distribution for a reverse biased p − i − n diode. Using the intrinsic region has two advantages. 1. Increased depletion width: it can be translated in increasing the photon absorption depth which means

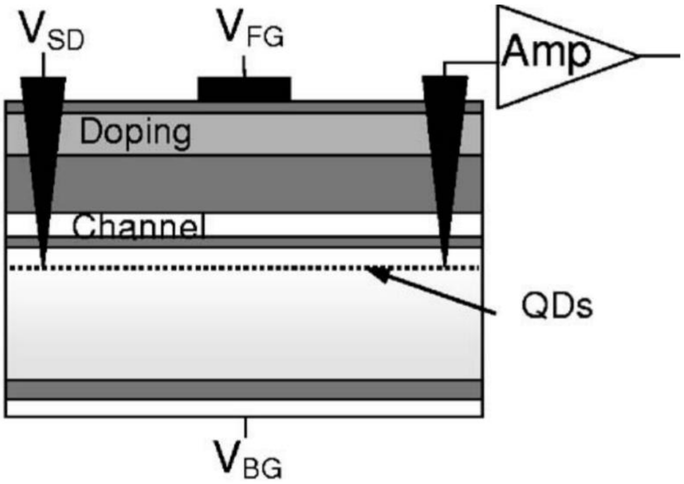


Figure 2.5. Schematic of a quantum dot photon detector [39]

more absorption in red and IR region of light. 2. Decreased junction capacitance, thereby reducing the RC delay constant. The transit time of carriers drifting across the depletion region, and the RC time response are the main sources of the response time of photodiode detectors. The response time of p − i − n photodiodes can be as low as a few hundreds of picoseconds. It should be noted that in the case of silicon, the maximum responsivity occurs on wavelengths which are substantially shorter than the bandgap wavelength. This is due to the indirect-gap nature of silicon where the photon absorption transitions typically take place from valence to conduction band states which are above the conduction band edge.

2.3.4. Quantum dot photon detector

Recently, the quantum dot based detector have been introduced for use as single-photon detector which consists of a transistor made of different layers of gallium arsenide and aluminum gallium arsenide. One of the layers consists entirely of quantum dots, which are highly sensitive to photons, placed at just nanometers separation from each other. A photon hitting the detector releases an electron trapped in the one of the dots which is picked up by a nearby highly sensitive channel. This causes a change in the resistance of the channel which is just a few nanometers above the quantum dot layer. Figure 2.5 shows the schematic of the detector. A photon detection probability of 50-86 % has been reported for GaAs, and AlGaAs types of these quantum dot optically gated field effect transistors. It is suggested that by avoiding avalanche process which can amplify both photo generated carriers and dark carriers, quantum dot photodetectors can have lower noise. Since this type of single photon counter is based upon a transistor, it is expected that the device can scale with technology and operate at low voltages (< 5 V).

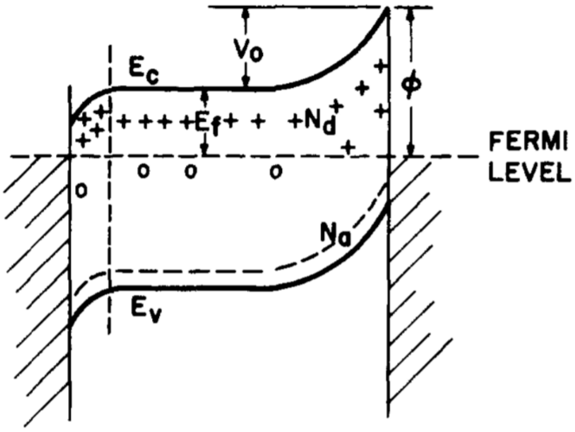


Figure 2.6. Energy band diagram of a Schottky-barrier photodiode

2.3.5. Superconducting Single Photon Detector

Single-photon detectors based on superconducting nanowires (SSPDs or SNSPDs) incorporate thin films of superconducting material shaped into meandering nanowires and have rapidly emerged as a highly promising photon-counting technology for infrared wavelengths. The superconducting single photon detectors generally use niobium-nitride (NbN) nanowires [40] as the medium of photon absorption. The resistance of the superconducting nanowire is zero, hence no voltage appears on the nanowire even in the presence of an electrical current. However, the absorption of a single photon at some point in the nanowire raises the temperature of that point, locally destroying the superconductivity. Thus, a finite resistivity appears in the nanowire which rapidly disappears as the nanowire is being cooled by the cryocooler and it turns superconducting again. This change of resistance creates a voltage pulse, clearly depicting the arrival time of a single photon at the superconducting nanowire. In general, the superconducting devices offer high quantum efficiency, low dark counts, low leakage current and excellent timing resolution but they have to be cooled to a very low temperature typically 2-4 K. The complexity and huge size of the cooling equipment limits the practical utilization of these devices.

2.3.6. Schottky barrier photodiode

Schottky barrier photodiode is simply a depletion layer photodiode in which the p-n junction is replaced by the metal-semiconductor heterojunctions. Schottky-barrier photodiode would result if the p or n type layer in the device is replaced by a thin transparent metal and is able to operate in two photo-detection modes: (i) Electron pair generation occurring from band to band or energy gap excitation in the semiconductor and (ii) Emission of carriers occurring from the metal to the semiconductor over the Schottky barrier, often referred to as internal photoemission.

The Schottky-barrier band diagram is shown in Fig. 2.6. There are many reasons for the popularity of these devices such as (i) the depletion region is at the surface resulting in decrease of surface recombination, (ii) low resistance of the metal causing less RC delay and larger bandwidth of operation and, (iii) n and p doped regions cannot be fabricated for every semiconductor material, so Schottky barrier devices proves an effective alternative for implementing photodetectors in these cases.

2.3.7. Avalanche Photodiode (APD)

An avalanche photodiode (APD) is operated at a very high reverse-bias voltage so that the resulting electric field across the junction exceeds the critical value necessary to trigger impact ionization [41]. An absorbed photon generates an electron-hole pair in the depletion layer of the APD and due to the strong electric field across the PN junction, the electrons drift towards the N+ side and the holes drift towards the P+ side. The higher the electric field strength, the higher will be the drift speed and when the electric field is high enough the carriers collide with the lattice structure and new electron-hole pairs are generated. This phenomenon is known as impact ionization and the generated electron-hole pairs create additional carriers, initiating a chain reaction of ionization referred to as avalanche multiplication process. The ability of an electron or a hole to trigger ionization is quantified by ionization coefficient and is represented by α_e and α_h for electrons and holes respectively. The ionization coefficients increase with electric field due to electron acceleration and decrease with temperature due to an increase in frequency of collision, thus reducing the probability of a carrier to gain enough energy for causing ionization. Two factors should be regarded simultaneously while designing an APD - maximizing photon detection probability which needs wider depletion regions and reducing the thickness of the multiplication region to minimize the possibility of localized uncontrolled avalanches (instabilities and microplasmas) being produced by electric field. A high uniform electric field should be designed in order to prevent the microplasmas.

2.3.8. Single Photon Avalanche Diode (SPAD)

Single-photon avalanche diodes (SPADs) are semiconductor devices that have been known for decades and have been exploited in several fields where single optical photons are to be detected: chemistry, physics, biology; laser ranging; optical time-domain reflectometry; single molecule detection; astronomy; and photon correlation techniques. In the last years an

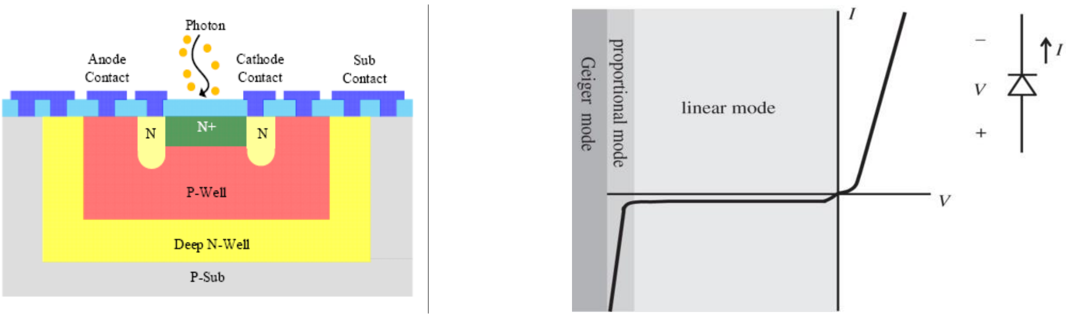
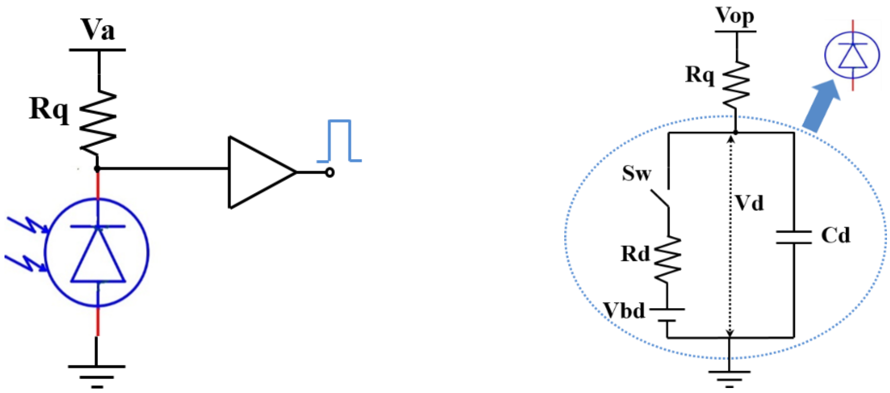


Figure 2.7. Simple cross section of a Single Photon Avalanche Diode and its I-V curve

increasing number of applications require the use of photodetectors able to acquire light signals down to single-photon level with high temporal resolution, extremely low noise and extended Dynamic Range (DR).

SPADs, like the Avalanche photodiode (APD), exploits the photon-triggered avalanche current of a reverse biased p-n junction to detect an incident radiation. The fundamental difference between SPADs and APDs is that SPADs are specifically designed to operate with a reverse bias voltage well above the breakdown voltage (on the contrary APDs operate at a bias lesser than the breakdown voltage). This kind of operation is also called Geiger-mode in literature, for the analogy with the Geiger counter. SPADs are semiconductor devices (Fig. 2.7) based on a p-n junction reversed biased at a voltage higher than the breakdown voltage by a so-called "excess bias". At this bias, the electric field is so high (higher than 3E5 V/cm) that a single charge carrier injected in the depletion layer can trigger a self-sustaining avalanche. The current rises swiftly (sub nanosecond rise-time) to a macroscopic steady level, in the milliampere range. If the primary carrier is photo-generated, the leading edge of the avalanche pulse marks (with picosecond time jitter) the arrival time of the detected photon. The current continues to flow until the avalanche is quenched by lowering the bias voltage down to or below breakdown voltage (V_B): the lower electric field is not able any more to accelerate the carriers to impact-ionize with lattice atoms, therefore current ceases. In order to be able to detect another photon, the bias voltage must be raised again above breakdown. Ignitions due to carrier generated by thermal or tunneling processes represent the detector noise and cause the so-called dark count rate (DCR). Moreover some carriers also get trapped during an avalanche and are released later when the SPAD is ready to detect a photon, thus causing spurious ignitions called afterpulses [42]. Suitable circuits, possibly tailored on fine modeling of the electrical behavior of such



(a)

(b)

Figure 2.8. (a) Basic Passive Quenching Circuit (PQC) schematic (b) Equivalent SPAD model

detectors, have to be employed together with the SPAD in order to achieve the best timing performance and lower the afterpulsing effect. Silicon SPAD detectors are becoming increasingly widespread in TCSPC setups due to their good photon detection efficiency (PDE) (e.g. about 50% at 550 nm, higher than 15% at 800 nm) and low timing jitter (few tens of picoseconds). In addition, a SPAD can be turned on and off by analog modulation of its biasing voltage (few volt) from below to above the breakdown voltage. A photon absorbed when the detector is biased below breakdown does not trigger a self-sustaining avalanche, hence is not detected by the read-out electronics. A SPAD requires a suitable circuit, which has to i) sense the leading edge of the avalanche current; ii) generate a standard output pulse synchronous with the avalanche build-up; iii) quench the avalanche by lowering the bias down to the breakdown voltage; iv) restore the photodiode to the operative level. This circuit is usually referred to as a quenching and processing circuit.

2.4. Metrology

2.4.1. Passive Quenching and Recharge Circuit

After the initiation of an avalanche, the avalanche current must be quenched to avoid damaging the diode. This can be achieved by lowering the biasing voltage of the SPAD below its breakdown voltage preventing the occurrence of the multiplication process. The simplest way to accomplish this is by using a Passive Quenching Circuit (PQC) [43] consisting of a high-value ballast resistor connected in series with the SPAD (Fig. 2.8). In quiescent conditions, the SPAD is reverse-biased to the desired voltage (above breakdown voltage) through the ballast resistor. When an avalanche is triggered in response to an impinging photon, the diode current

(I_d) swiftly rises to its peak value, given by the transient excess bias voltage ($V_{ex} = V_d - V_{bd}$) divided by the diode resistance (R_d) as below :

$$I_d\,(t) = \frac{V_d(t) - V_{bd}}{R_d} = \frac{V_{ex}(t)}{R_d} \tag{2.2}$$

where, V_d is the diode voltage and V_{bd} the breakdown voltage of the SPAD. The diode resistance R_d is given by the series of space–charge resistance of the avalanche junction and ohmic resistance of the neutral semiconductor crossed by the current. The value of R_d depends on the semiconductor device structure, which is typically lower than 500Ω for SPADs with a wide area and thick depletion layer and from a few hundred ohms to various kiloohms for devices with a small area and a thin junction [44]. The self-sustaining avalanche current develops a voltage drop across the high resistive load and the avalanche process simply quenches itself. The avalanche current discharges the parasitic capacitance at the SPAD cathode so that I_d exponentially drops and reaches the steady state value of I_{ds}.

$$I_{ds} = \frac{V_{OP} - V_{bd}}{R_d + R_q} \approx \frac{V_{EX}}{R_q} \quad (since\ R_d \ll R_q) \tag{2.3}$$

The quenching time constant or the exponential decay of cathode voltage is set by the terminal capacitance C_d and the resistances R_q and R_d in parallel as demonstrated by the following equation.

$$T_q = C_d \frac{R_d\,R_q}{R_d + R_q} \approx C_d R_d\ (since\ \ R_d \ll R_q) \tag{2.4}$$

As I_d declines, the diode voltage also reduces approaching V_{bd} which causes a decline in the number of carriers traversing the avalanche region. Since the avalanche process is statistical, so it may happen that after a random time none of the carriers crossing the high field region cause impact ionization. However, the probability of the occurrence of zero multiplied carriers rapidly increases when the diode current I_d falls below the latching current level I_q (Fig. 2.9). The value of I_q is not sharply defined, although a value of about 100µA has often been used. If the steady state current I_{ds} is set to a value much lower value than the latching current I_q, then the PQC works in the best possible way, quenching the avalanche within a well defined time with fairly small jitter. However, if I_{ds} approaches I_q, quenching occurs with a progressively longer delay and wider time jitter and above I_q, the avalanche is no longer quenched and a steady current flows just like that of normal diodes used as voltage references in electronic circuits which may cause permanent damage due to excessive heating. The value of the ballast resistor must be

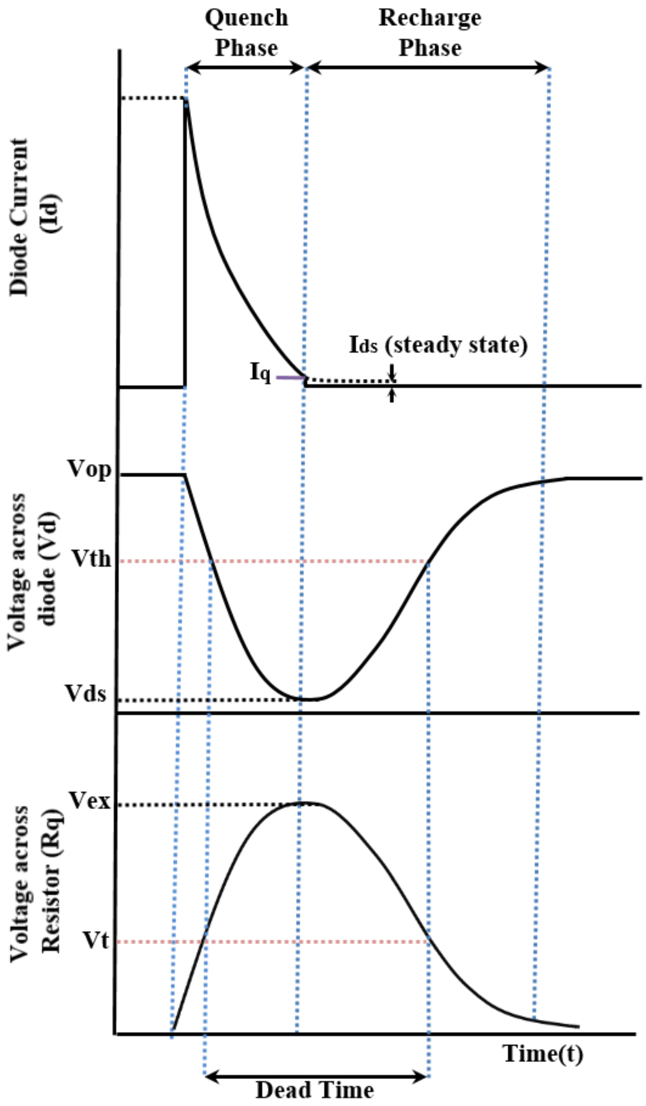


Figure 2.9. Timing diagram during passive quench and recharge operation

large enough to ensure that the asymptotic I_{ds} value is sufficiently lower than the quenching level I_q. A good rule of thumb suggests that I_{ds} should not exceed 20µA. The value of the quenching resistor R_q should be at least 50 kΩ for 1V of applied excess bias voltage V_{ex}. Henceforth, lower values of R_q should be avoided and the minimum recommended values to be employed range from 50 to 500 kΩ for thin-junction SPAD devices that work with V_{EX} from 1 to 10 V, and 200 kΩ to 2.5 MΩ for thick-junction SPADs that work with V_{EX} from 4 to 50 V. The avalanche process is self-sustaining above the latching current level Iq (<100µA) and self-quenching below it. Jitter of the quenching time with respect to the avalanche onset influences the value of I_q and results in the corresponding jitter of diode voltage V_q at which quenching occurs. Theoretically, V_q is slightly higher than V_{bd} as can be inferred from the equation below:

$$V_q = V_{bd} + I_q R_d \tag{2.5}$$

The total charge Q_{av} in the avalanche pulse is related to the steady state current I_{ds} and can be evaluated using the equation :

$$Q_{av} = \frac{V_{OP} - V_q}{C_d} \approx \frac{V_{EX}}{C_d} \approx I_f T_r \tag{2.6}$$

$$T_r = R_q C_d \tag{2.7}$$

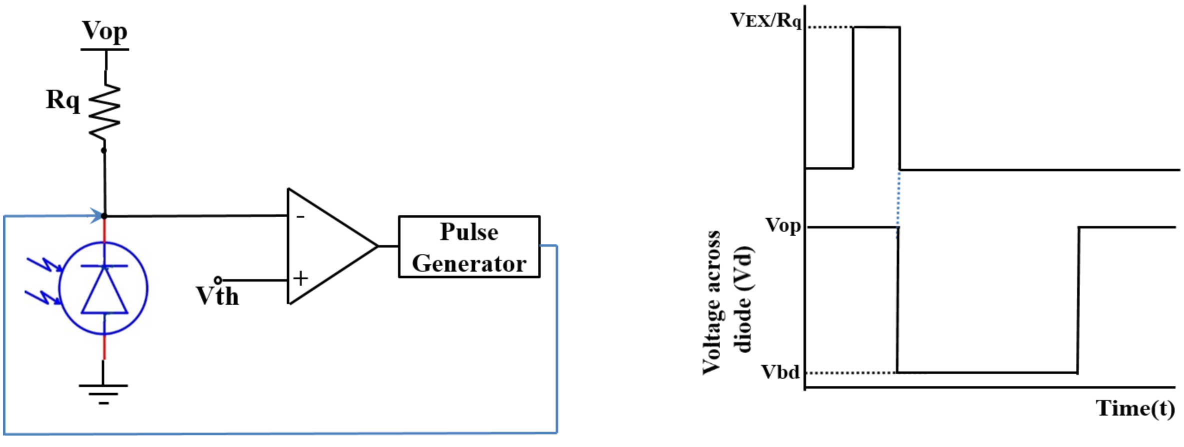
T_r corresponds to the characteristic time constant of the voltage recovery. An output pulse can be obtained from a PQC as shown in Fig 2.8 (a). The voltage-mode output generates a peak amplitude of :

$$V_u = (V_{OP} - V_{bd} - I_q R_d)\,\frac{R_d}{R_q + R_d} \approx V_{EX}\,\frac{R_d}{R_q} \tag{2.8}$$

One drawback of this configuration producing a voltage waveform from the fast avalanche current pulse is that the detector timing performance is not fully exploited owing to the time constant T_q, generated by the intrinsic low-pass filter, that acts on the fast current pulse to produce the voltage waveform. Such low-pass filtering has a detrimental effect on the photon timing accuracy but can be compensated by employing a very low threshold level in the subsequent fast timing electronics.

2.4.2. Active Quenching and Recharge Circuit

The drawbacks of passive quenching circuits such as slow voltage reset and undefined dead time are overcome by a different kind of quenching circuits called Active Quenching Circuits (AQCs) in which the avalanche is sensed through a low impedance and both quenching and reset are performed using active components. The operation of AQCs starts with the detection of the avalanche signal followed by the quenching process in which the bias voltage is lowered below breakdown for a predetermined time-period (also known as hold-off time) and then the device is brought back to its quiescent state (Fig. 2.10). Here, the quenching time is not affected by fluctuations due to the statistical nature of avalanche multiplication process. As compared



(a) (b)

Figure 2.10. (a) Basic schematic of a AQC (b) SPAD cathode voltage waveforms, connected to AQC

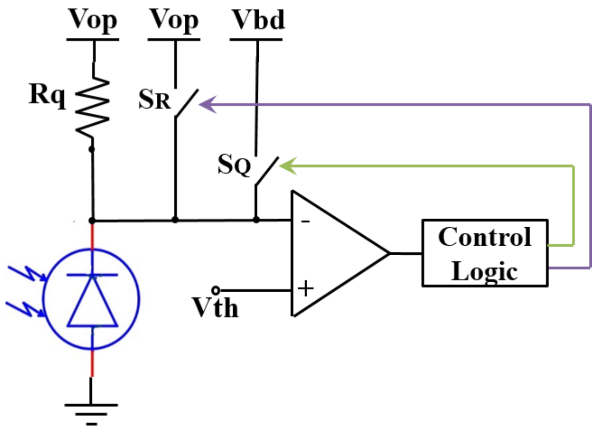


Figure 2.11. Basic schematic of a mixed active-passive quenching circuit

to PQC, the quenching transition time (T) is faster which determines the total avalanche charge as shown in the equation below:

$$Q = \frac{V_{EX}}{R_q} \; T \tag{2.9}$$

2.4.3. Mixed Active-Passive Quenching and Recharge Circuit

Mixed active-passive quenching overcomes the limitations of purely passive and active quenching by combining their respective advantages [45, 46]. Also, it is benefitted by the direct integration with the detector avoiding the use of discrete components which provides a well-

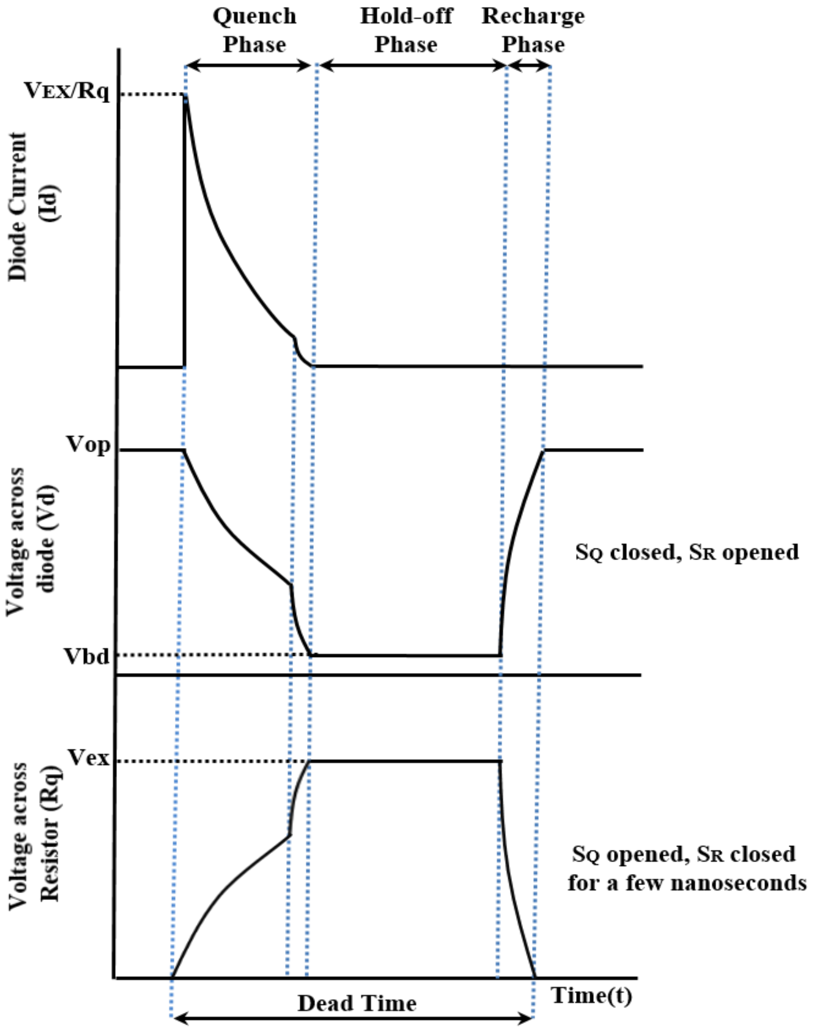


Figure 2.12. Timing diagram when using active quench and recharge operation

defined delay and speeds up the voltage defined dead time with much reduced parasitic capacitance. This also leads to a smaller avalanche charge which consequently reduces the intervention reset. In this kind of quenching circuits, the SPAD is connected both to a passive high impedance ballast resistor and the active quenching and reset circuitry as shown in Fig. 2.11 where the switches represent the active circuitry. At the advent of an avalanche process, the avalanche current flows through the ballast resistor initiating the passive quenching process as in PQCs. The active circuitry soon senses the avalanche and starts the active quenching by closing the switch S_Q during which the diode capacitance is discharged rapidly with a time constant $RS_Q \times Cd$, where RS_Q represents the nominal on-resistance of the switch S_Q. Using this technique, it is possible to keep the SPAD in the quench state for a definite period of time also known as hold-off time. In steady state condition the diode voltage is very close to V_{bd} and the diode capacitance Cd is fully discharged. For recharge and restoring the bias voltage, the switches S_Q is opened and S_R is closed in quick succession leading to the quick recharge of the SPAD with a time constant given by RS_R×Cd, where RS_R is the on-resistance of S_R. After a few nanoseconds, S_R is opened leading to the end of recharge phase and makes the SPAD ready for next detection. SPAD cathode voltage and diode current waveforms are shown in Fig. 2.12. During the quench and recharge phase, the SPAD is rendered inactive for the detection of any photon. Hence, the time-period starting from the initiation of the avalanche signal to the end of recharge phase is considered as dead-time (quenching time + hold-off time+ recharge time) which limits the measured photon count rate to 1/dead time. The dead time is not fixed in passive techniques which makes the situation more complicated. The count rate saturates in case of active techniques when the rate of the actual triggered avalanche pulses is higher than 1/dead-time whereas in case of passive techniques the count rate drops. The dark count rate has a significant dependence on the SPAD dead time which increases in situations where the SPAD triggers before it is fully recharged to the set threshold voltage for detection resulting in the next avalanche signal to go undetected. The effect of varying dead time on the count rate measurement can be modeled using the equation 2.10.

$$m = \frac{nexp(-nt_{r1})}{1+nt_q} \tag{2.10}$$

m- the measured count rate, n - Actual count rate, t_q - Quench time, t_{r1} - Time difference between the starting point of the recharge phase and the instant when the diode voltage reaches the set threshold. For the avalanche signal to be processed by the frontend electronics, the output of the current sense circuit or the voltage across the diode or the resistor Rq can be used with a set threshold.

2.4.4. Dark Noise

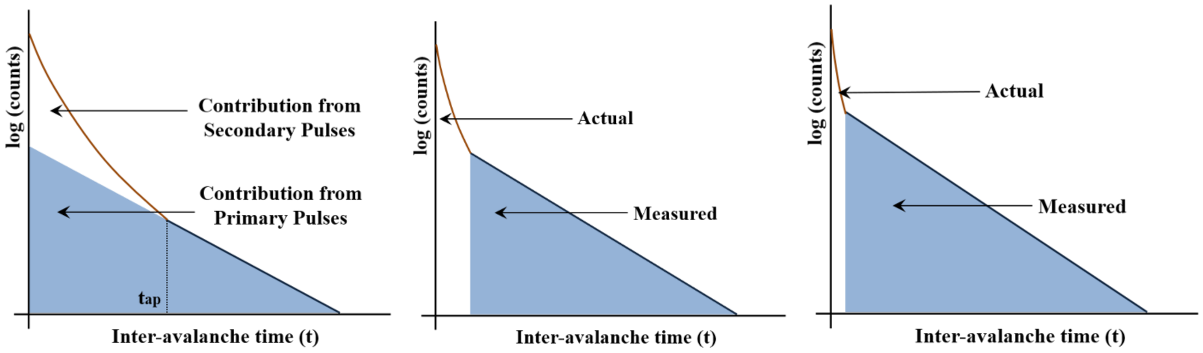
In single-photon counting mode, one of the first characterizations that needs to be performed is the measurement of dark noise. Generally, dark noise is measured as the dark count rate (DCR) which represents the spurious avalanche pulses trigged in dark conditions in the absence of any impinging photons. Dark count rate is composed of two components, namely the primary and secondary [44], and is defined as the average number of measured SPAD output pulses in one second under dark condition (Fig. 2.13a). Primary dark count rate is attributed by the natural processes such as thermal carrier generation or band-to-band tunneling or trap assisted processes. On the contrary, secondary pulses, also known as afterpulses, occur due to trapping of a carrier during an avalanche process and subsequent release after a certain time-period or possibly the quenching phase, leading to the possibility that the released carrier may trigger a false avalanche. The secondary pulses are measured as the afterpulsing probability, which is also in correlation with the primary pulses and depend on the lifetime of the trapped carriers. In some cases, afterpulses may also result due to an avalanche initiated by a photon. Generally, DCR is characterized in three stages - In the first stage the total DCR (primary and secondary pulses) is measured, second stage involves the characterization of secondary pulses and in the third stage the primary pulses are characterized in isolation [47]. The total DCR for a given SPAD at certain excess bias voltage and temperature, which includes both primary and secondary pulses, is measured under complete dark conditions by counting SPAD output pulses for a certain time-period (integration time) and then normalizing the measured counts to one second.

The best way to measure the secondary pulses or after pulsing probability is by utilizing the time correlated carrier counting technique [48] where the time-periods between two consecutive avalanche pulses are used to build a histogram [49]. In silicon the carrier life time in a trap is of the order of a few nanoseconds to microseconds and hence the exponential distribution observed after inter avalanche period of 20 µs represents the primary pulses. Since the release of trapped carriers leads to secondary pulses resulting in multi-exponential behavior, the secondary pulse count is then obtained by subtracting the primary pulse count from the total measured avalanche pulses. The afterpulsing probability can be defined as follow:

$$APP = \frac{Secondary\ Pulse\ Count}{Total\ Avalanche\ Count} \tag{2.11}$$

where, APP is Afterpulsing probability

The afterpulsing probability reduces with increase in dead time and could go undetected when the photon rate or the primary DCR rate is higher than 1/carrier life time in a trap. Afterpulsing also depends on the number of carriers that flow through a diode during quench process and hence an increase in the excess bias voltage increases the afterpulsing probability. Quenching techniques have a direct effect on the afterpulsing probability, for example, in case of passive quenching, the terminal capacitance (C_t) discharges through the diode whereas in active techniques the terminal capacitance (C_t) is mostly discharged through the low resistance



(a) (b) (c)

Figure 2.13. (a) Inter-avalanche time histogram highlighting contributions from primary and secondary pulses, (b) afterpulsing measurement using passive recharge, (c) afterpulsing measurement using active recharge

MOSFET switch (S_Q) as shown in Fig. 2.11 and only a small fraction of charge flows through the diode. Because of smaller charge flow through the diode, afterpulsing probability is expected to be low in active quenching techniques. Also, when using passive techniques, after the quench operation the diode is progressively recharged which results in the progressive increase of the SPAD triggering probability. This phenomenon can suppress some of the afterpulsing effects and hence one could overestimate the SPAD afterpulsing performance particularly in the case of passive quenching techniques (Fig. 12.13b). The event of suppression of afterpulsing effects doesn’t really occur in active quenching as the recharge time is much faster making it almost impossible for the SPAD to fire during the recharge time-period. However, even if an avalanche process is triggered during the active recharge time, the SPAD will fire immediately as soon as it comes out of the dead time interval. Even though negligible, this principle results in the increase of the counts in the first time bin in the inter-avalanche histogram (Fig 2.13c) and these are considered to be afterpulsing affects as it is almost impossible to find out whether the counts are observed due to some other influences. Henceforth, keeping in view all the advantages, active quench and recharge circuit would be the best choice for afterpulsing measurement. By compensating for the afterpulsing on the measured count rate (MCR), afterpulsing can be eliminated as per the formula below:

$$APCR = (1-APP) \times MCR \tag{2.12}$$

where, APCR is the afterpulsing compensated count rate and APP is afterpulsing probability Primary pulses are mainly attributed by tunneling process or Shockley Read Hall (SRH) process (traps). Although there is no accurate technique available, it is possible to study the major source of dark noise by operating with various temperatures and excess bias voltage. It was reported that the tunneling noise is highly dependent on bias voltage as compared to the noise generated by the SRH process. On the other hand, the dark noise generated by tunneling processes, is almost insensitive to temperature variations, whereas the noise that originates from the SRH process increases with temperature [50, 51]. When performing primary pulse measurements, it is vital to choose the SPAD deadtime very carefully so that the afterpulsing becomes negligible. Also, it is strongly recommended to utilize active quenching and recharge techniques for primary pulse characterization such that the deadtime is well defined and have precise count rate measurements.

2.4.5. Photon Detection Probability (PDP)

For SPADs, photon detection probability (PDP) represents the probability that a photon incident on its active area gets detected by the generation of a photocarrier pair contributing to detector current. The SPAD PDP is measured with the help of a reference diode, by illuminating both of them with the same light intensity of a known wavelength. A lamp may be used to emit a wide range of wavelengths or a specific wavelength that can then be selected by a monochromator. An integrating sphere is then used to diffuse and scatter the output light of the monochromator uniformly across the reference photodiode and the SPAD under test. The effective photons detected by the SPAD is measured by computing the difference between the SPAD output pulses rate and its DCR. The SPAD PDP is then evaluated using the formula below:

$$PDP = \frac{MCR-DCR}{PCR} \tag{2.13}$$

where, MCR is the measured count rate and PCR is the photon count rate incident on the active area, measured using the reference photodiode.

Theoretically, PDP can be defined as the product of quantum efficiency (QE) and avalanche triggering probability. QE is the ratio of the number of carriers reaching or generated in the depletion region, to the number of incident photons on the SPAD active area. QE strictly depends on the silicon absorption coefficient, henceforth making it a function of wavelength and the SPAD design. The avalanche triggering probability mainly depends on the excess bias voltage and is defined as the probability that a free carrier present in the depletion region can trigger an avalanche. For a given SPAD the higher the excess bias voltage (at field strengths

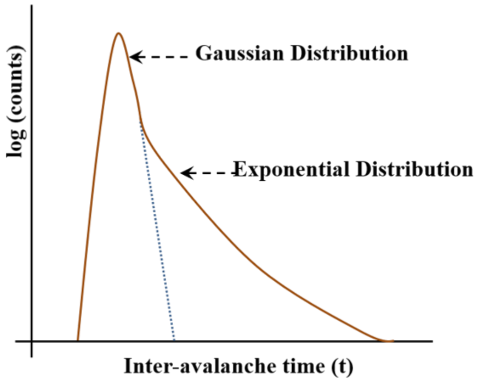


Figure 2.14. Statistical distribution of photo-response of the SPAD [47]

higher than 5×10^5 V/cm), the higher will be the avalanche triggering probability. However, at very high excess bias voltage, the avalanche triggering probability tends to saturate due to the saturation of the carrier ionization coefficient. PDP also depends on the SPAD size and the variations in PDP occurs across the periphery of the active area as a result of relative low field strength region due to the diffusion of guard region dopants.

2.4.6. Timing Jitter

Timing jitter for SPADs can be defined by the statistical distribution of the photon detection time, which is composed of two components namely the gaussian and an exponential term [52]. The photocarriers that diffuse to reach the multiplication region contribute to the exponential component whereas the gaussian component is introduced by the avalanche build-up dynamics (Fig 2.14). In CMOS SPADs, two stages are involved in the avalanche build up process. In the first stage the avalanche is initiated and grows locally resulting in a current flow influenced mostly by the avalanche build-up statistics. The avalanche build-up dynamics is dependent on the applied excess bias voltage. When the excess bias voltage increases, the timing jitter reduces and so lower becomes the statistical fluctuation in avalanche build-up. In the second stage the avalanche spreads laterally to the other regions through multiplication assisted diffusion process. At given bias conditions, the resulting current flow during the second stage strongly depends on the rate of spread of avalanche to other regions which is basically based on the spatial position of the initiation point of the avalanche, for example avalanche multiplication process can spread to all directions uniformly if initiated at the center of the active area rather than being initiated at the periphery.

Hence, in addition to the jitter introduced by the avalanche build-up statistics of the first stage, the statistics involving the avalanche initiation points across the active area introduces the jitter.

When the light intensity or DCR is very high, independent avalanche process are initiated at multiple points in the active area and the avalanche current spreads faster as compared to the case when it is initiated at a single point. In such cases, the statistical fluctuation of the current rise time also depends on the variance in the number of independently created avalanche initiation points. In addition to the gaussian component of the SPAD timing jitter, the exponential component is also affected by the multiple avalanche initiation points.

Generally, the timing jitter for SPADs is expressed by the full-width-at-half-maximum (FWHM) of the statistical distribution of the photo-response of the SPAD, which is measured using a pulsed laser source having the potential to emit picosecond light pulses. The SPAD is placed just in perpendicular to the light source and a histogram is generated based on the photon detection timings by means of a time-corelated-single-photon-counting (TCSPC) instrument. Also, for measurements of timing jitter, it is strongly recommended to use active quench and recharge circuitry. This is because active quenching reduces the impact of parasitic capacitances and afterpulsing effects. Furthermore, since the timing jitter depends on the excess bias voltage, one should avoid using passive recharge as there is a possibility that the SPAD can fire at a lower excess bias voltage.

2.5. Single Photon Avalanche Diode (SPAD) types

2.5.1. SPAD in custom process

Previously, silicon based SPADs were only designed in a dedicated process and can be categorized by at least two different forms of design:

1. Reach-Through Devices: Earlier the devices were based on reach-through structure [53, 54], for eg., n-p- Π-p+ double diffused reach-though APD [55] (Fig. 2.15) in which the p-type (boron) and n-type (phosphorous) diffusions were carried out in sequence and are so adjusted that when a reverse bias voltage is applied, the depletion layer of the diode just reaches-through to the low concentration Π region. The Π region of the device has typically about 5kΩ-cm resistivity and when peak electric field at the junction is almost equal to that required to cause avalanche breakdown, the depletion layer increases rapidly out to the p+ contact while the field throughout the device increases relatively slowly.
2. Planar Devices: Although reach through devices achieved high sensitivity they exhibited poor timing resolution and requires a very high biasing voltage, thereby limiting its usage to a great extent. For these reasons, researchers started focusing on planar devices such as the one shown in Fig. 2.16 in which the active junction [56] was formed by a shallow (0.3

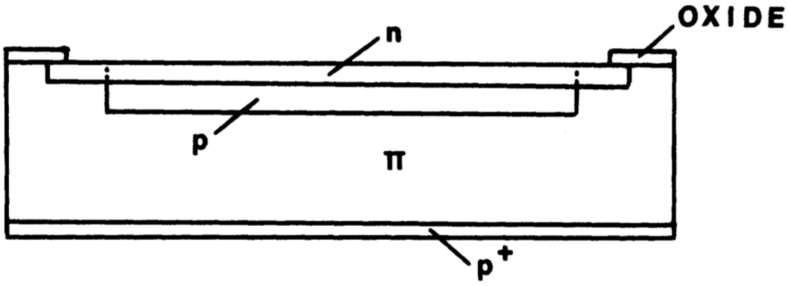


Figure 2.16. Cross-sectional view of a typical reach-through SPAD [1]

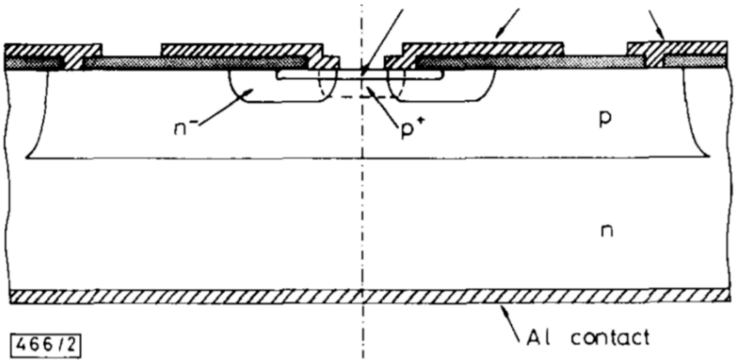


Figure 2.15. Cross-sectional view of a planar SPAD [1]

μm) n+ layer in a p-bulk substrate of 0.6 Ωcm resistivity. A surrounding guard ring, composed of deep diffused 5-8μm n-layer, avoids premature edge breakdown which has a typical breakdown voltage of 28 V. The contact to the anode, is provided by using the p-diffusion layer. The area spread over by the main junction is sensitive to impinging photos and so this region is referred to as the active area. The designs of silicon based planar type SPADs have significantly evolved over years, thanks to the breakthroughs in the fabrication process technologies [57-61]. The main advantage of the planar design as compared to reach-through structures is the requirement of low bias voltage and a better timing performance, however the major drawback in using the dedicated processes is its limitation in realizing large arrays.

2.5.2. SPAD in Standard CMOS Process

The need for single-photon detection array with precise timing circuitry and on-chip data processing has precipitated the development of planar SPADs in standard CMOS process. With the introduction of deep-submicron (DSM) CMOS, the progress got expedited and the realization of parallel arrays of SPADs became feasible. The state-of-the-art SPAD arrays, fabricated in CMOS technology, possess the capability to accurately time stamp individual incoming photons and have even surpassed the traditional PMTs in terms of their spatial resolution. The introduction of multichannel SPAD arrays has enabled the acquisition and mapping of several photoelectron timestamps for each scintillation photons to improve the timing estimation in PET systems [62]. Massive arrays of SPADs commonly known as silicon photomultipliers have been proven to be Magnetic resonance (MR) compatible which has enabled the possibility of realizing promising hybrid imaging modality [63, 64], combined with dual high sensitivity PET (positron emission tomography) and MRI (magnetic resonance imaging) systems. An improved wireless image sensor can be fabricated by integrating SPAD array sensors for the detection of fluorescence emitted by biological tissue to help monitor cancer cells during surgery [65]. The state-of-the-art SPAD technologies, has enabled the increase of the throughput of single-molecule fluorescence spectroscopy in solution using parallel arrays of SPADs [66, 67]. A SPAD array camera, with temporal resolution in the order of micro-seconds and zero read-out noise, offers great potential for super-resolution microscopy allowing for the detection of extremely weak signals at ultra-fast imaging speeds [68].

2.5.3. State-of-the-art SPADs fabricated in CMOS technology

The cross-section of the SPAD[69], fabricated in 180nm CMOS technology, using p+-nwell (NW) junction is presented in Fig. 2.17. In this design the main junction was engineered to be between p+ and deep nwell-2 (DNW) with the pwell acting as guard ring to protect the p+ edges from premature edge breakdown. The DNW and highly doped buried-N (BN) layers, placed beneath nwell, serve as photon collection region which is much wider than the other conventional p+-nwell devices. The BN layer also protects DNW from depletion when the substrate is reverse biased. The DNW layer has less dopant concentration near the junction which reduces tunneling noise and enhances the spectral response with wider depletion. In this configuration, the photo-generated carriers in the BN, DNW and nwell layers reach the multiplication region through the diffusion process. For a circular device with a diameter of 12µm, the circular SPAD has a breakdown voltage of 14.64V. The SPAD exhibits photon detection probability (PDP) greater than 40% from 440 to 580 nm at an excess bias voltage of 4V and a timing jitter of 95 ps FWHM and 141ps FWHM using 405 nm and 637nm lasers respectively. The device also features a low dark count rate (DCR) of 31 Hz and 1.8 kHz at an excess bias voltage of 1V and 4V respectively.

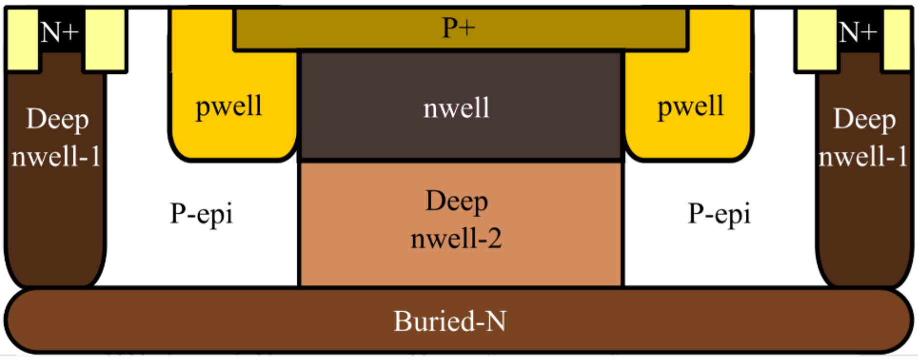


Figure 2.17. p+ /deep nwell SPAD fabricated in 180nm CMOS technology [69]

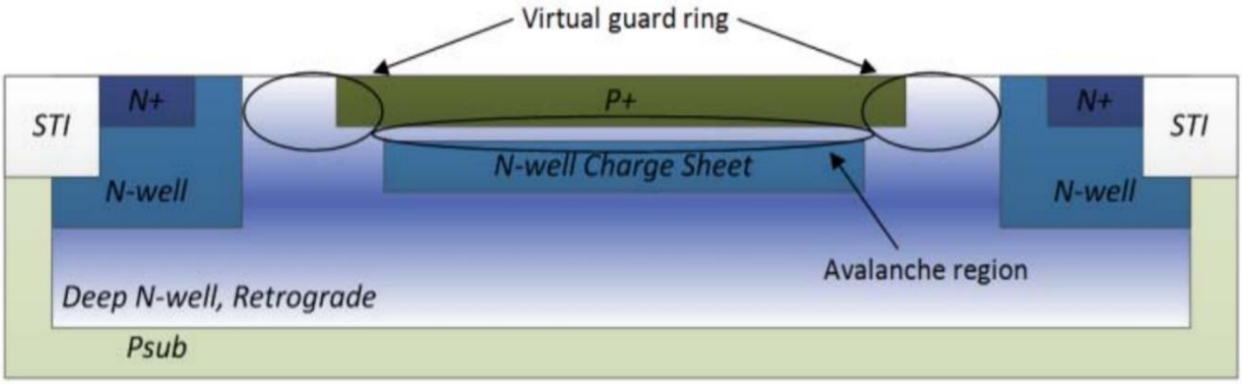


Figure 2.18. Schematic cross section of the device with a retrograde doping profile in 180nm CMOS technology [70]

Another SPAD device[70] (Fig. 2.18) fabricated in a low voltage commercial 180nm CMOS image sensor technology exhibits good characteristics and is based on a triple well process to limit the substrate noise effects. The device cross section shows a 10-μm diameter n-type charge sheet formed by ion implantation which is centered below the 12μm diameter circular P+ region. The avalanche multiplication occurs in this charge sheet, thus behaving as an active region. The deep n-well is implanted with a retrograde doping profile which causes the doping to be effectively P− around the P+ region periphery creating a virtual diffused guard ring. The photon detection efficiency is more than 40% from 420 to 550 nm and DCR is less than 100 Hz at room temperature for excess voltage of 3.3 V.

A SPAD structure[71], implemented in a 130-nm imaging process, is depicted in Fig. 2.19. The structure employs a PWELL anode with P contact. The cathode terminal connection is formed by NWELL with N contact, down to retrograde DEEP NWELL. A novel guard ring structure compatible with standard CMOS technologies was incorporated where the implantation of either PWELL or NWELL above DEEP NWELL was avoided by use of an IMPLANT STOP

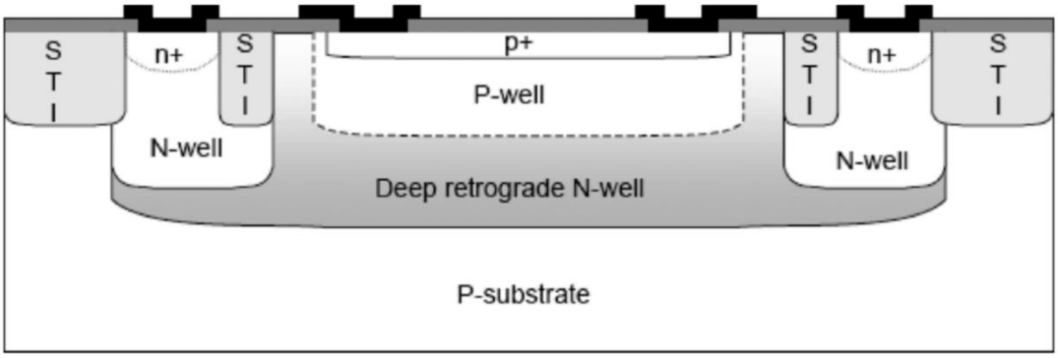


Figure 2.19. SPAD Cross section implemented in 130nm technology [71]

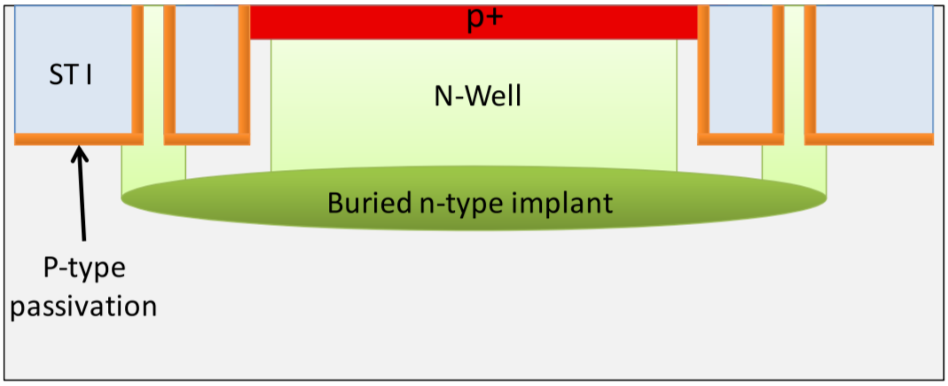


Figure 2.20. Cross-section of the circular STI-bound SPAD 130nm CMOS technology [72]

layer resulting in a deep retrograde NWELL with low surface doping, increasing with depth. The device with an active area of 50µm exhibits a photon detection efficiency (PDE) peak of 28% at 500 nm and a dark count rate of 25 Hz at 20°C. Another circular shaped STI-bound SPAD structure[72] was implemented in 130nm technology as shown in Fig. 2.20. The main sources of DCR in SPADs are either the thermally generated free carriers entering the depletion region of the diode or tunneling induced avalanche breakdown. In this SPAD structure glove-like p-type structure surrounds the STI interface reducing the probability of minority carriers entering the multiplication region and triggering dark counts. DCR is further decreased by reducing n-well doping of the SPADs which diminishes tunneling phenomenon. At the STI interface the doping level is high resulting in a very short mean free path of the minority carriers and the probability of these carriers entering the active region of the SPAD reduces significantly. To reduce the probability of edge breakdown the electric field between the guard ring and the n-well has to be minimized. Hence, the p-type doping concentration is reduced with increasing distance from the STI. The buried n-type implant acts as an ohmic contact to the n-well. The SPAD displays a PDP of up to 36% and a timing jitter of 125 ps.

A CMOS back-side illumination-compatible SPAD[73] (Fig. 2.21) was reported in 90-nm imaging technology. The multiplication region is the junction between the deep n-well (DNW) and the thin p-type epitaxy layer on a p-substrate. P-well, placed on top of the active SPAD region, is biased at the same potential as that of the cathode terminal and used to collect short-wavelength photoelectrons. A virtual guard ring is formed by using the natural retrograde implantation profile of the DNW creating a sharp doping gradient at the planar junction with the p-epitaxy and avoiding the formation of p-well. Diffusion of the implant and spreading causes a low doping gradient at the sides which ensures a high breakdown voltage and effective guard ring operation. The device achieves a PDE of greater than 20% at 850 nm and low DCR of 100 Hz with 51 ps FWHM timing resolution and a low after-pulsing probability of 0.375%. Another SPAD[74] implemented in 90nm technology was developed with an octagonal multiplication region as illustrated in Fig. 2.22. A guard ring using a combination of Shallow

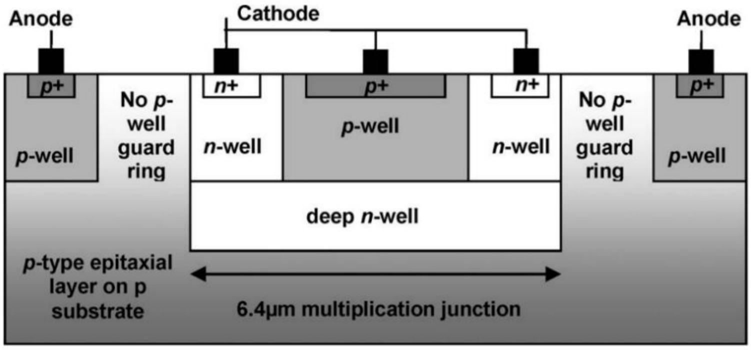


Figure 2.21. CMOS back-side illumination-compatible SPAD reported in 90-nm imaging technology [73]

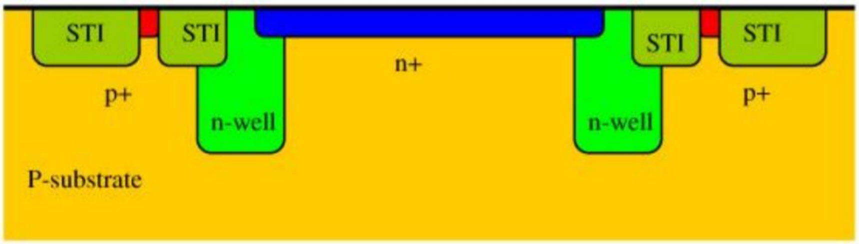


Figure 2.22. CMOS SPAD developed with an octagonal multiplication region in 90nm technology [74]

Trench Isolation (STI) and n-well was created for the prevention of premature edge breakdown. At room temperature the device exhibits a dark count rate of 8.1 kHz, a maximum PDP of 9% and timing jitter of 398ps at a wavelength of 637nm at an excess bias voltage of 0.13V. Similarly, a SPAD[75] was implemented in CMOS 350nm technology (Fig. 2.23) which comprises a circular p+ active region over n-well. Premature edge breakdown was prevented by implanted p-well guard rings. Another structure[76], shown in Fig. 2.24, was designed in the same technology with a deep low-doped n-well implant and a p+ shallow implant to form the p-n junction. Premature edge breakdown was prevented by the implantation of a p-type guard-ring reducing the peripheral electric field surrounding high doped p+ active area of the SPAD. The high-field region in the active area has a breakdown voltage of 26 V and the device yields a PDE above 50% at wavelength of 420 nm, DCR below 50 cps at room temperature and maximum timing response (FWHM) of 77 ps and 120 ps for 10 µm and 30 µm active areas, respectively.

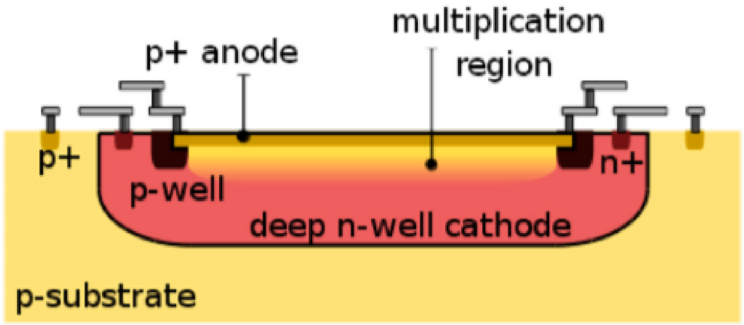


Figure 2.23. Cross-section of the SPAD in 0.35 µm CMOS technology [75]

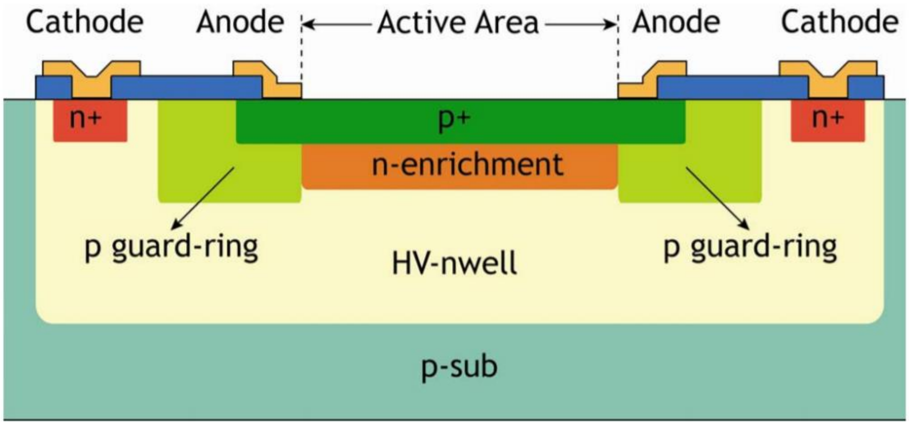


Figure 2.24. Cross-section of the SPAD in 0.35µm CMOS technology, diameter of the active area are 10 µm-30 µm [76]

The image sensor[77] (Fig. 2.25) was implemented in the 3D two-tier Tezzaron's FaStack process, which is based on a CMOS 130 nm technology. This process enables the integration of SPADs in the top tier with a minimum amount of circuitry for its operation, while the rest of the circuits based on standard CMOS are placed on the bottom tier. The high-field multiplication region in the employed SPAD structure consists of a circular n-type low doped drain (NLDD)/p-well junction which is surrounded by the n-well/p-well junction having a higher breakdown voltage and thus acts as the guard ring to prevent premature edge breakdown in the SPAD. In this technology two standard CMOS wafers are bonded face-to-face utilizing copper to copper bonds called direct bond interfaces (DBIs), having a diameter of 0.6µm and 4µm pitch. After wafer-to-wafer bonding, the substrate of the top tier is thinned to 4.2µm. The TSVs, with approximately 1.2µm in diameter, are then exposed and metallization is added to create the bond pads. Due to its deep multiplication region, the PDP is more than 12% between 650 nm and 800 nm wavelengths at an excess bias voltage of 1.5V.

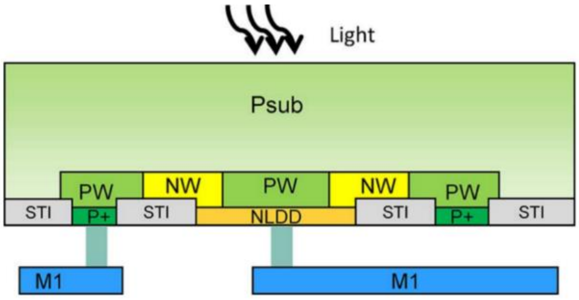


Figure 2.25. SPAD structure cross section - sensor was implemented in the 3D two-tier Tezzaron's FaStack [77]

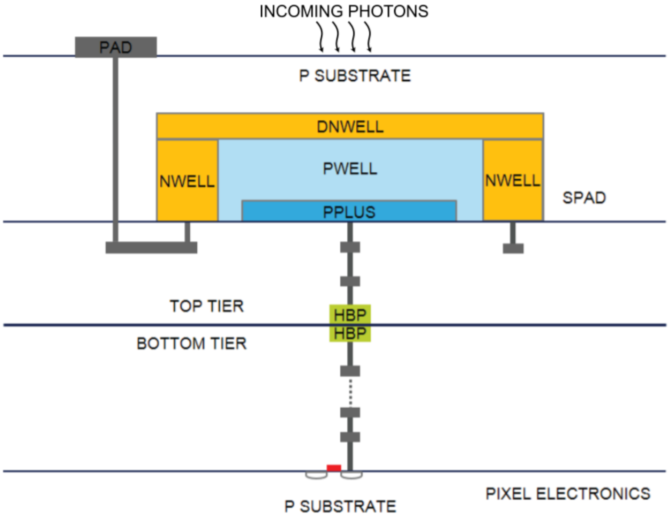


Figure 2.26. p-well/deep n-well SPAD in 65/40 nm 3D IC CMOS technology [78]

Table 2.2. Specifications of the various SPADs designed in CMOS technologies. All the values are reported at room temperature

No.	**Tech**	**Timing jitter**	**DCR @ 25ºC**	**Vbd**	**Active Area**	**PDP @ 800nm**
[70]	180nm	-	≈130Hz @ Vex=3.3V	21.4V	12µm (dia.)	≈ 12% (Vex=4V)
[79]	180nm	165 ps @790nm	≈5kHz @ 750ns dead time, Vex=4V	19.7V	10µm (dia.)	≈ 20% (Vex=4V)
[80]	130nm	90 ps @654nm	≈18Hz @ Vex=2V	20V	8µm (dia.)	≈ 20% (Vex=4V)
[81]	130nm	198 ps	≈1000Hz @ Vex=2V	12.13V	5 µm (dia.)	≈ 6% (Vex=2V)
[82]	130nm	-	≈5kHz @ Vex=2V	13.7V	12.3 µm (dia.)	≈ 20% (Vex=3V)
[83]	130nm	-	≈110Hz @ 30ns dead time,Vex=1.8V	-	8 µm (dia.)	-
[84]	90nm	107ps @410nm	≈250Hz	10.3V	2 µm (dia.)	≈ 3% (Vex=0.13V)
[85]	350nm	70 ps @ 470 nm	≈580Hz @Vex=3V	18.6V	15 µm (dia.)	≈ 4% (Vex=2.5V)
[76]	350nm	71.3ps @ 390 nm	≈43Hz @Vex=6V	26.5V	10 µm (dia.)	≈ 5% (Vex=6V)
[86]	65nm	235 ps	≈500Hz @Vex=250mV	9.1V	64 µm^2	≈ 0.8% (Vex=250mV)
[78]	65nm	107 ps	≈200 @Vex=1.4V	16.5V	336 µm^2	≈ 10% (Vex=1.4V)

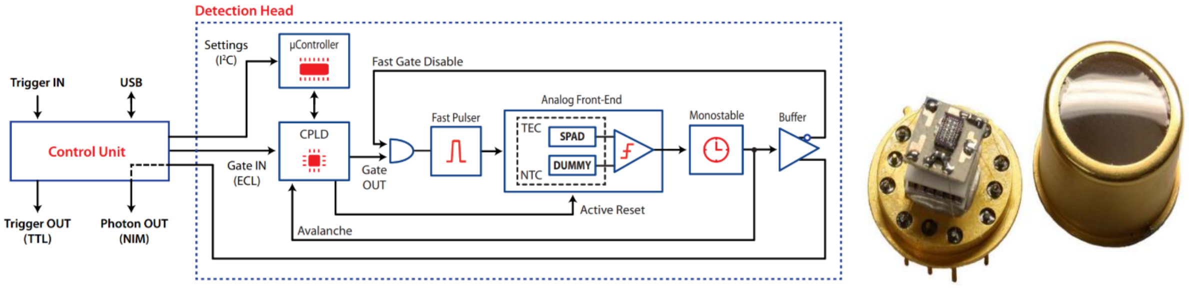
A SPAD based complete pixel fabricated in a backside-illuminated (BSI) 3D IC technology was presented in [78] (Fig. 2.26). The chip stack comprises a photodetection and data processing tiers designed in standard 65nm BSI image sensor and 40nm CMOS process technology respectively, raising the prospect of smaller pixel pitches at the cost of increasingly complex electronics. Connections are made between the two tiers using hybrid bonding pads (HBP). Much smaller pixel pitches can be achieved utilizing the resulting 3D IC technology which gives the 65 nm BSI imaging technology an extra edge over the other less scaled CMOS or custom technologies. When operated at 4.4-V excess bias, the design exhibits a timing jitter of 95ps FWHM, maximum PDE of 21.9% at 660 nm and 0.08% after pulsing probability with a dead time of 8 ns. Table 2.2 compares the main characteristics of SPADs in different CMOS technologies.

2.6. RECENT NEAR INFRARED SPECTROSCOPY (NIRS) MODULES

To detect weak and fast light signals, several detection methods were proposed with adequate sensitivity and temporal resolution in the sub nanosecond range. Earlier this approach required complex and bulky system and setups that limited its effective utilization and applications to the laboratory, for example, molecular imaging of small animals by optical projection tomography, characterization of photonic glasses etc. Typically, as a result of the limitations associated with the size of TD-NIRS light sources and detectors, the light pulses are conveyed to the sample (e.g., phantoms or biological tissues) and delivered to the detection unit by means of complex optical setup which involves coupling to optical fibers or bundles. Although, the technique provides the inherent advantage of electrically isolating the measurement site from the device, but it makes the measurement process very inconvenient and cumbersome. Scaling down near infrared (NIR) measurement setups could help create low-cost, portable and fully non-invasive devices for applications such as optical mammography and functional brain imaging. Recent developments and growth in photonics components for light sources (e.g., VCSELs) and detectors (e.g., SPADs) has enabled the scaling down of both size and cost. In the following sections a complete and updated description of some of the recent TD-NIRS systems are presented.

2.6.1. Time-Gated Single Photon Counting Modules

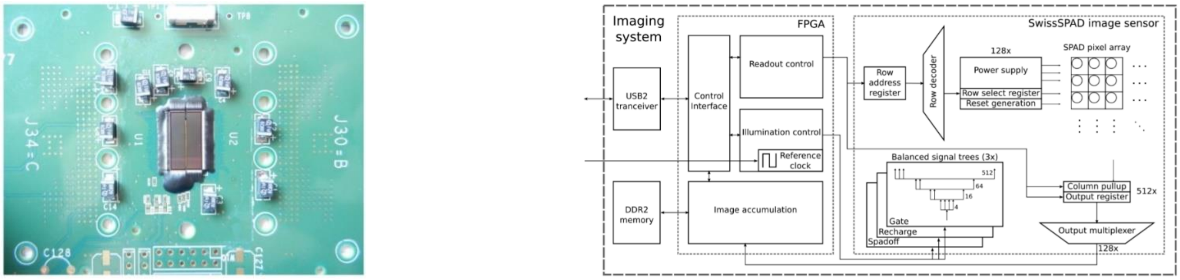
Ref [87] presented the design and characterization of a complete single-photon counting module capable of time-gating a SPAD, as shown in Fig 2.27. This instrument can operate SPADs of different active area diameters ranging from 20 μm to 200 μm and is capable of turning ON and OFF a SPAD detector with gating transitions down to 110 ps. The gate duration can also be adjusted between 2 ns and 500 ns, with a repetition rate up to 80 MHz. The experimental characterization of the module highlights its very flat temporal response, with a time resolution of the order of 30 ps. The module (Fig. 2.27) is also capable of operating in free-running mode and is fully user-configurable via a PC interface that can be integrated in any experimental setup. In order to reject undesired strong light pulses preceding or following the signal of interest, fast time-gating capability is extremely beneficial which avoids overshadowing of the signal of interest. This also allows to increase the dynamic range of optical acquisitions up to 7 decades. The detection head (Fig. 2.27b) dimensions are 60 mm × 60 mm × 120 mm, thus displaying a small form-factor. The detector parameters such as excess bias voltage, hold-off time, temperature are user-programmable, ensuring flexibility and ease-of-use in most of the applications. The circuits and methods for fast-gating a SPAD were demonstrated in [88], achieving a wide dynamic range optical measurement in the visible and near-infrared spectrum. For effectively gating a silicon SPAD, a mixed-signal amplifier was designed using a chain of fast ECL logic and wideband microwave monolithic integrated circuits (MMICs). The fast pulse generator generates gating pulses with transition times (turn-ON and turn-OFF) below 200 ps and repetition rates up to 50 MHz. The gate window duration can be adjusted from 10 ns down to just few hundreds of picoseconds. To discriminate the avalanche pulses from spurious capacitive feed-through spikes (caused by sub-nanosecond transition times of the gating pulses), differential front-end electronics was designed reaching a low timing resolution below 30 ps. The module (Fig. 2.28) is user configurable in terms of all operating conditions such as excess bias voltage, gate-ON duration and DC bias of the SPAD and can be utilized in many applications providing a wide dynamic range of the measurement by rejecting the strong interfering light signal in the time-domain.



(a) (b)

Figure 2.27. (a) Block diagram of the fast-gating single-photon counting module [87] which includes the detection head and the control unit provides all the timing and trigger inputs/outputs, the USB link, (b) Windowed TO-8 package assembly, containing both active and “dummy” SPADs glued on a ceramic chip carrier.

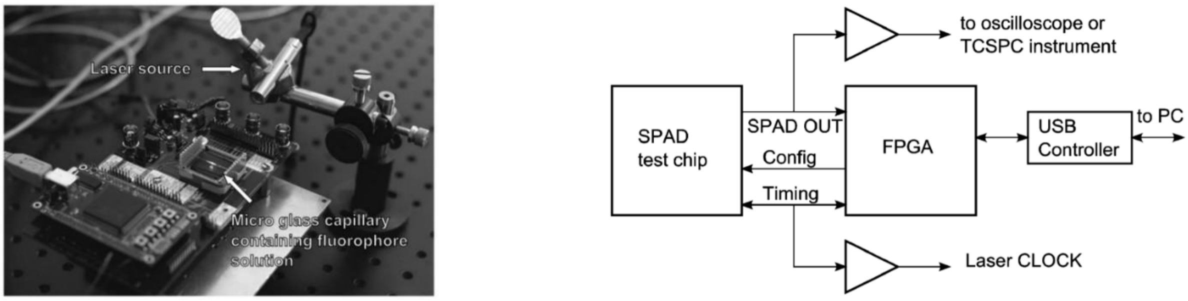
Figure 2.28. Fast-gated module [88] consisting of two boards (left) and its housing (right)



(a) (b)

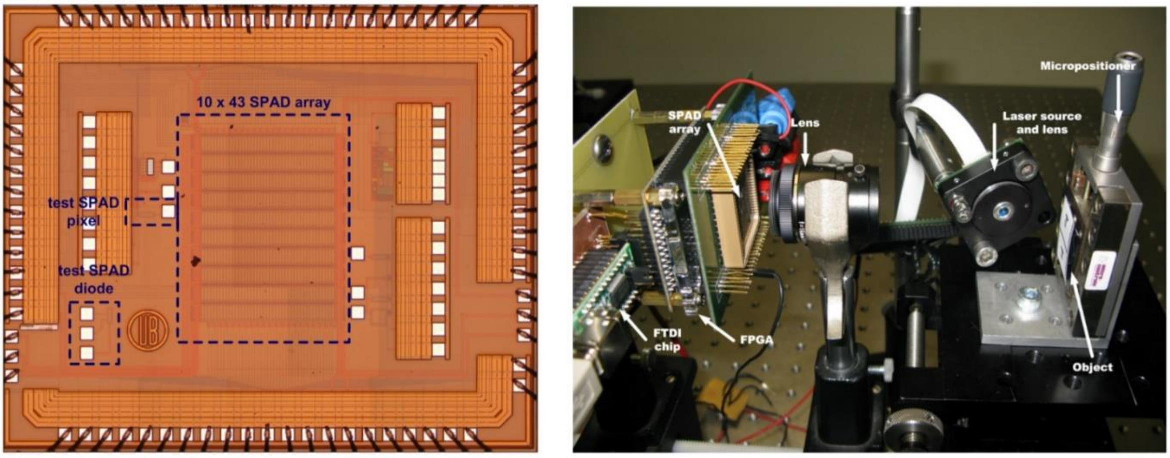
Figure 2.29. (a) Chip carrier PCB with two chips bonded side by side for simultaneous operation at doubled resolution. The gap is less than 6 pixels wide [89]. (b) A block-level representation of the imaging system. The part on the right depicts the interior of the SPAD chip built around the central array of 128 x 512 pixels. An FPGA, depicted on the left, is used to generate the control signals and receive the data generated by the pixels.

2.6.2. CMOS-based SPAD sensors



(a) (b)

Figure 2.30. Experimental setup showing the chip [90] with the peripheral circuits (b) Block diagram of the chip with interfacing board



(a) (b)

Figure 2.31. (a) Micrograph of the fabricated 10×43 SPAD array [91], (b) 2D imaging set-up

SwissSPAD [89] was a large format, highly sensitive SPAD based image sensor fabricated in a high-voltage CMOS process, having a resolution of 512×128 pixels and a pitch of 24μm. The fill-factor could be increased to 30% with the use of micro-lenses to concentrate incoming light on the sensitive areas of the chip. Fast ard precise gating signals, having duration as small as 4 ns, controlled the exposure windows for application such as time-resolved imaging, fluorescence lifetime measurements, fluorescence correlation spectroscopy and generation of true random numbers. The uniformity of the gate edges location is ~140 ps (FWHM) over the whole array, while in-pixel digital counting enables frame rates as high as 156 kfps.

Another SPAD–based pixel array [90] was used for the fluorescence lifetime measurements. Each pixel was composed of a single photon detector combined with an active quenching circuit and a 17-bit digital events counter to measure the number of events generated within a user-defined observation window. The digital control phases, required by the pixel array with a full programmability of the main timing synchronisms, was provided by an on-chip master logic. The pixel exhibits an average dark count rate of 3 kcps and a dynamic range of over 120dB in time uncorrelated operation. The sensor was characterized using an FPGA-based control board and a USB interface for data acquisition.

An optical imager consisting of a 10×43 array of single-photon avalanche pixels[91] was developed which showcased a fill-factor of 67%. The array exhibits an average dark count rate of 67 kHz at an excess bias voltage of 1V. The array is operated in time-gated mode, minimizing the electrical crosstalk that appears between SPAD sensors sharing the same well to increase the fill-factor. With a time-gate of 4 ns, the probability of detecting the sensor noise goes down to 10^{-4} noise counts per frame even though the sensor area is quite large (20µm×100µm). Hence, time-gated technique is advantageous in improving the imager signal-to-noise ratio, dynamic range, contrast and spatial resolution. This sensor had been specifically designed and fabricated for particle tracking at future linear colliders but could also be utilized in biomedical imaging, Raman and near infrared spectroscopy, 3D cameras, space and HEP experiments. A 32×32 pixel SPAD[92] based image sensor, featuring 25-µm pitch and 20.8% fill factor, was fabricated in a high voltage 0.35-µm CMOS technology for time-gated fluorescence lifetime detection. The chip includes a phase-locked loop circuit for gating window generation, with a maximum repetition rate of 40 MHz or 80 MHz if an external delay generator is used. Another single-photon, time-gated, 160×120 pixel imager[85] was designed and implemented in a standard high-voltage 0.35µm CMOS technology. The sensor was validated with FLIM measurements, consuming power of around 20.6 mW in analog mode and 156.7 mW in digital mode. The compact pixel achieves a 21% fill-factor in a 15 µm pitch with a delay locked loop to stabilize the internal waveform generation for reliable timing performance. The analog counting scheme behavior is independent from the SPAD characteristics and the voltage step can be tuned using external reference voltages. Time-gate windows can be programmed down To 750 ps width with at a maximum repetition frequency of 50 MHz. A self-referenced analog-to-digital converter allows for fast and reliable digital readout up to 486 fps, addressing the nonuniformity and noise issues.

A smart SPAD based pixel[93] was designed and fabricated in cost-effective 0.35µm CMOS technology for time-of-flight (TOF) and time-correlated single photon counting (TCSPC) applications. Each pixel of the sensor contains a large photoactive area SPAD, an analog Variable Load Quenching Circuit (VLQC) for fast avalanche sensing and quenching, pulse shaping electronics with a well defined output, and a 10 bit TDC to measure the photon arrival time with 312ps resolution and 320-ns full-scale range (FSR). An on-chip 10 bit memory latch stores the data during the readout and a frequency doubler and global Delay Looked Loop (DLL) feeds all TDCs with common signals. The SPAD with an active area of 30µm exhibits a photon detection efficiency of about 50% at 410 nm and down to 5% at 800 nm with very low noise of 12 counts per second (cps) at room temperature and 5V excess bias voltage. The single-photon imager array can be utilized for 3-D depth ranging in safety and security applications and 2-D fluorescence lifetime decays in biomedical imaging.

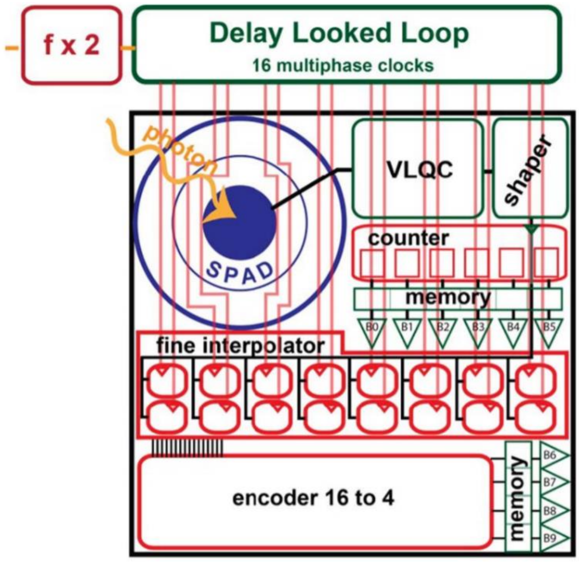
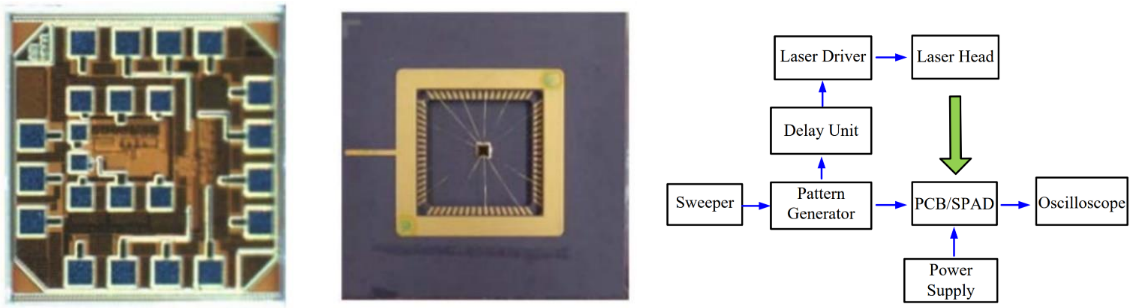


Figure 2.32. Block diagram of the SPAD based smart pixel [93], showing the main components: SPAD detector, quenching circuit (VLQC), pulse shaper, coarse counter, fine interpolator, 16-to-4 encoder, memories and 10 bit output buffers.



(a) (b)

Figure 2.33. Micrograph of the fabricated chip [94]; (b) A typical experimental Setup of the illumination measurement

A time-gated (TG) SPAD[95] was designed and fabricated in a standard high voltage 0.35 μm CMOS technology for Raman spectroscopy. The sub-ns adjustable time gating window (300 ps) was used to suppress the fluorescence background typical of Raman spectra and also minimizes the dark count rate (DCR) so as to increase the signal to noise ratio of the signal. This is a clear advantage compared with continuous-time Raman spectroscopy devices. Another TG-SPAD[94] was developed for the detection of the Raman signal suppressing the strong background fluorescence. The on-chip pulse generator produces a fixed gate window of 3.5 ns, which offers a minimum detection windows of about 200ps. The temporal resolution of the SPAD was reported to be about 60ps when tested with a very short laser pulse. A delay unit is used to adjust the position of the gate window with respect to the laser pulse. For synchronization the laser beam is split into two paths by a beam splitter – one path is used to illuminate the sample, and other path is sensed by a high-speed photodiode to trigger the TG-SPAD and oscilloscope.

An all-digital, time-gated CMOS SPAD imager[96] was presented which comprises an array of 128×128 SPAD pixels capable of detecting single photons integrated with time gating circuitry and a 1-bit memory. The time-gate is controlled by a combination of an on-chip 600ps delay line and an off-chip FPGA-based 200ps delay line. The digital data can be read out through an acquisition board for further processing. The sensor allows on-chip CMOS based platform for filter-less fluorescence detection and fluorescence lifetime imaging microscopy (FLIM).

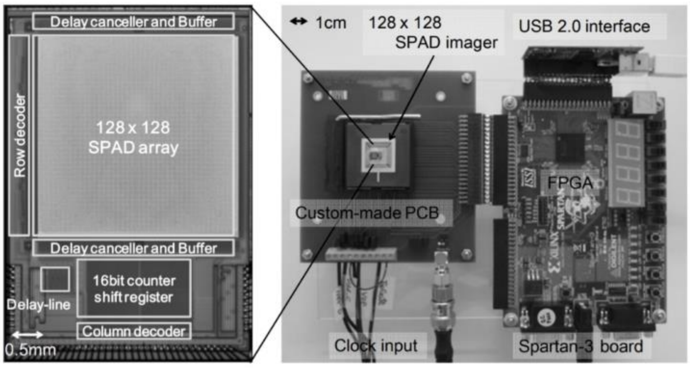
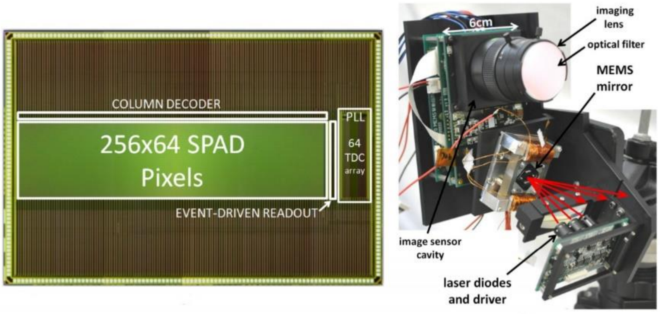


Figure 2.34. Micrograph of the sensor chip and photograph of the detection system [96]



(a) (b)

Figure 2.35. (a) Photomicrograph of the CMOS single-photon image sensor(8x5mm^2) [97], (b) Experimental setup of the overall LIDAR system.

An optical time-of-flight 256x64-pixel image sensor[97], fabricated in a high-voltage 0.18µm CMOS technology, was introduced taking advantage of a MEMS-based laser scanning device which comprises fast column synchronization decoders, an event-driven readout circuit, an array of 64 row-level high-throughput time-to-digital converters, and a 16Gbit/s global readout circuit. For circular SPADs of 10.5µm in diameter, the median dark count rate was 750 counts/s. The sensor was tested under 2klux of background light and had a repeatability error of 13.5cm throughout the distance range of 20 meters. A high resolution Time Correlated Single Photon Counting (TCSPC) 256 × 256 array image sensor[98], based on sample and hold Time to Amplitude Converter (TAC) pixels and a global ramp voltage, was presented. The array demonstrated a pixel pitch of 8 µm, fill factor of 19.63 % and time jitter of 368 ps at 10 fps employing an off-chip 14-bit differential Analogue to Digital Converter (ADC). A bidimensional array [99] based on single-photon avalanche diodes, fabricated with the standard 0.35µm HV-CMOS process, for triggered imaging systems was presented. The detector exhibits a photon detection efficiency of 1.4% and dynamic range of 15 bits with gated ‘on’ periods of 10 ns and an excess bias voltage of 1.0V.

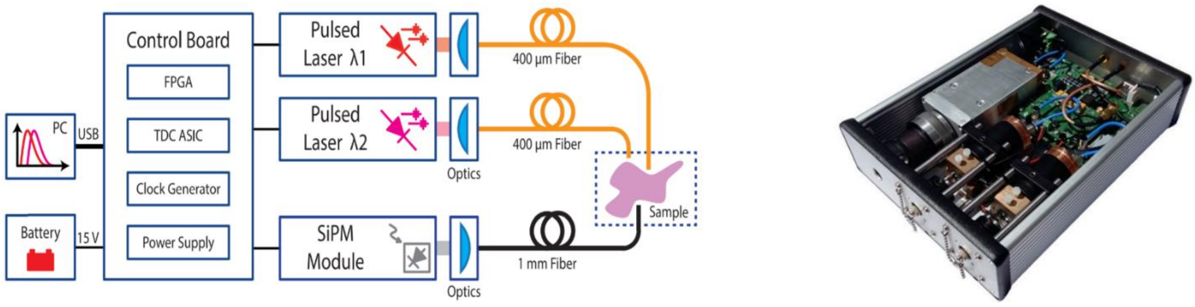
A CMOS 64×64 SPAD array[100] fabricated in a 0.18µm standard CMOS technology with in-pixel 11b time-to-digital converter (TDC) was developed for time-resolved imaging, in particular 3D imaging. The sensor features a pixel pitch of 64µm with a fill factor of 3.5% and time resolution of about 145ps at 9µW normalized power consumption per TDC. An innovative pixel architecture [101] was introduced which included four separate SPADs with independent active time-gating and quenching circuit, a shared time-to-digital converter (TDC) with 50-ps resolution, four independent photon counters, and multiple operation modes. A 23 bit memory cell stores both the photon timing and photon counting acquisitions by each SPAD element. A array of 32×32 SPADs was fabricated in 0.18 µm high voltage CMOS technology with a fill-factor of 9.6% and overall active area of 1 mm^2. A short comparison of the various state-of-the art SPAD sensors developed in CMOS 350nm technology is presented in Table 2.3.

Table 2.3. Comparison of the state-of-the-art SPAD imagers fabricated in CMOS 350 nm technology

Parameter	**[102]**	**[103]**	**[75]**	**[104]**	**[105]**
Chip Size	24.7 × 1.2 mm^2	12.3 ×3.1 mm^2	9.6 ×4.8 mm^2	2.95 × 4.3 mm^2	3.5 × 3.5 mm^2
Breakdown Voltage	19.6V	19.6V	26.1V	18.9V	24V
PDE@465 nm	0.3-9.6% (Ve = 3V)	20 %	30 %	25 %	43%
Temporal Resolution	250ps	-	120 ps	-	-
Array Format	1024 ×8	512 ×8	64 ×32	2 × (4) × 128	32 ×32
Technology	350 nm	350 nm	350 nm	350 nm	350 nm
Pixel Pitch	24 µm	24 µm	150 µm	-	100 µm
Fill Factor	4.9 -44.3%	5%	3.14 %	23 %	3.14 %
DCR	80 -5.7k (Ve = 3V)	200 cps	100 cps	~19k	4 KHz
Minimum Gate Window	0.7 ns	3.8 ns	-	100 ps	-
Gating Frequency	1 to 950 Hz	156 KHz	-		-

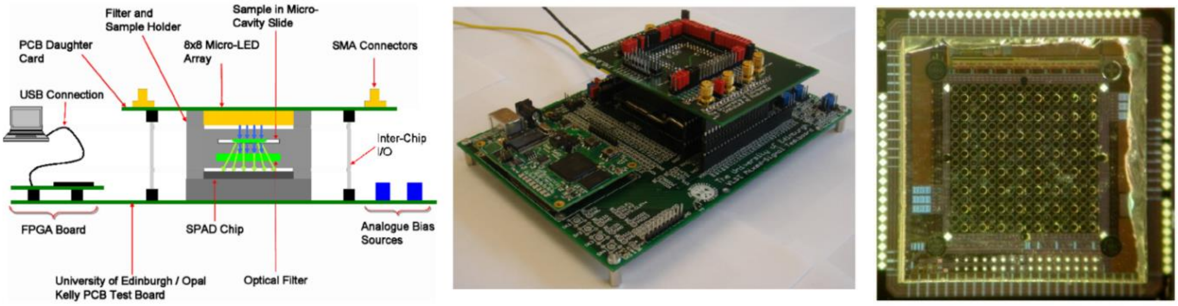
2.6.3. Compact NIRS instrument employing light sources

A compact, and low power consumption instrument[106] was designed for time-domain near-infrared spectroscopy which employed two custom-designed pulsed diode lasers, a single-photon detection module and a custom time-to-digital converter with 10 ps time resolution. The average optical power of the laser diodes is higher than 2 mW at 40 MHz repetition frequency and operational wavelengths of 830 and 670 nm. The single photon detection module was based on a 1 mm^2 active area silicon photomultiplier. The system experimental characterization shows an instrument response function (FWHM) of 240 ps and 280 ps for the 830 nm and 670 nm channels when operating at 40 MHz repetition rate with measurement stability less than ±1% over several hours of operation. The instrument can be interfaced with a remote PC for real-time data acquisition and displaying of the acquired DTOF curves. The instrument is housed into a compact aluminum case of size 200×160×50 mm^3, ensuring portability and ease of operation.



(a) (b)

Figure 2.36. (a) Block diagram of the complete system [106] consisting of two pulsed laser sources (at 670 nm and 830 nm), one 1 mm^2 active area SiPM based single-photon detection module and a TDC-based TCSPC system to reconstruct the optical waveform and send it to a portable PC for data analysis, (b) Picture of the instrument enclosed in an aluminum box of size 200 × 160 × 50 mm3, which hosts three fiber ports on front side and other connectors (power supply, USB, and trigger IN/OUT) on the rear panel.



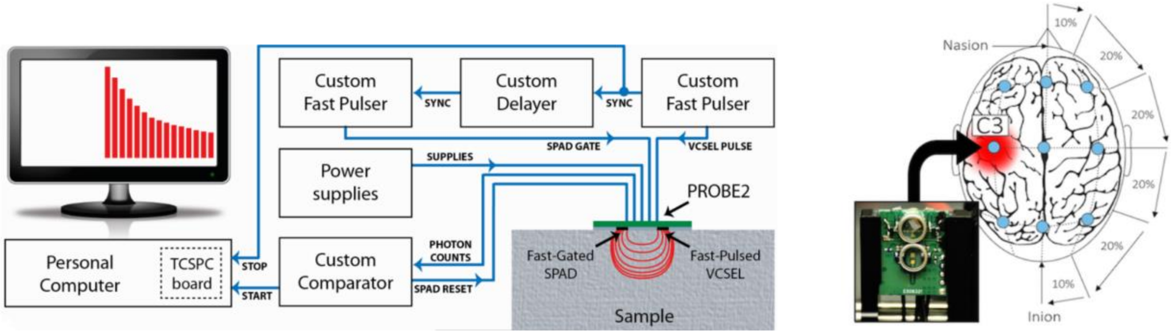
(a) (b)

Figure 2.37. (a) Complete two-chip micro-system [107]. The PCB daughter card is physically supported by the filter and sample holder and stacked header pins, (b) AlInGaN micro-LED array bump-bonded to an 8 × 8 CMOS driver array.

A CMOS-based micro-system[107] was introduced for time-resolved fluorescence lifetime analysis. It consisted of a 16×4 array of SPADs fabricated in 0.35µm high-voltage CMOS technology with a pixel size of 100µm × 200 µm incorporated with time-gated photon counting and signal processing circuitry. In-pixel 9-bit ripple up-down counters allows programmable scanning of time resolved events over a 48 ns range with a 408 ps resolution, which is interfaced with a FPGA and connected to a PC where the photon count histograms can be displayed. The sensor is also integrated with an 8 × 8 array of 72 µm diameter AlInGaN blue micro-pixellated light-emitting diodes (micro-LEDs), fabricated from “standard” InGaN/GaN quantum well blue LED wafers, which are bump-bonded to an equivalent array of LED driver circuits realized in a standard low-voltage 0.35 µm CMOS technology. The wavelength spectra of the CMOS driven blue micro-LED device peaks at a wavelength of 450 nm. Each element in the LED array can be addressed individually and capable of emitting optical pulses with a maximum DC optical output power of 550 µW and 300 ps in width (FWHM). Similar device[108] was designed by bonding a 4×16 array of AlInGaN UV micro-pixellated light-emitting diodes to a CMOS chip.

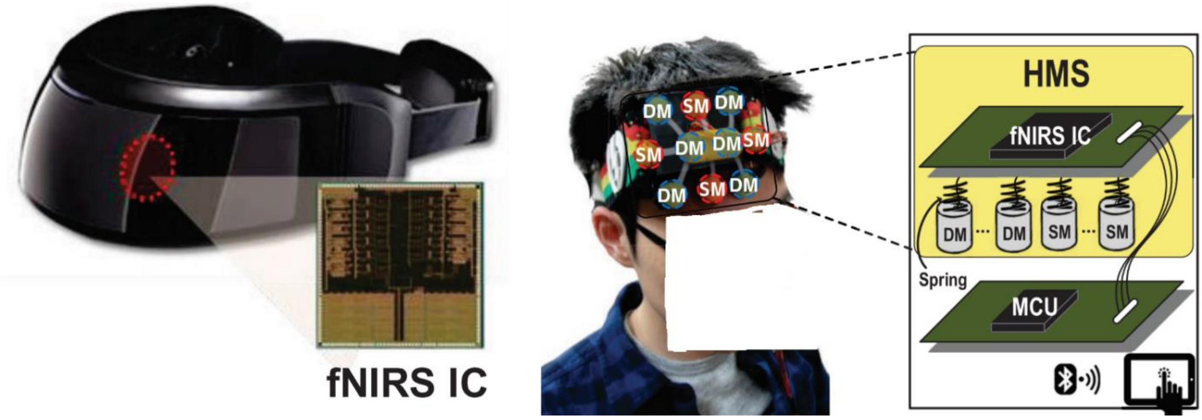
2.6.4. Compact optical probes

In [109], a source-detector pair was directly embedded into the probe composed of a pulsed Vertical-Cavity Surface-Emitting Laser (VCSEL) to allow parallelization for dense coverage and a time-gated SPAD to gather the timing information of the incoming photons. The SPAD, integrated with the front-end circuitry for detector bias, was fabricated in a high voltage CMOS process which can be gated ON and OFF in less than 200 ps and features an active area diameter of 200 µm with a photon detection efficiency of 16% at 690 nm and around 5% at 800 nm. The source is based on a VCSEL die with a dedicated pulsing electronic circuit which emits light pulses up to 900 µW average power with a repetition rate of tens of MHz and 250 ps FWHM at 690 nm. The probe size is 20 mm x 25 mm with the pulsed VCSEL and time- gated SPAD mounted at a relative separation of 5 mm. It also requires some external electronics such as a fast pulse generator (220 ps, 7 V pulses, 40 MHz repetition rate) to operate the VCSEL and provide the global synchronization signal, a second pulse generator (6 ns, 6.5 V) to gate the SPAD, a passive delayer to shift in time the gate window, a comparator to pick up the SPAD avalanche signal and restore the detector operating condition after each photon detection and a TCSPC board.



(a) (b)

Figure 2.38. (a) Scheme of the experimental setup employed to operate the probe [109] based on small (5 mm) source-detector separation. (b) The probe (shown in photo) was positioned over the sensorimotor area of the left hemisphere demonstrating its effective utilization.



(a) (b)

Figure 2.39. (a) Picture of fNIRS IC-based portable fNIRS system [110], (b) Architecture of HMS-based fNIRS system

A custom tranceiver IC (5 × 5mm^2) [110], fabricated in 0.18 μm 1P6M CMOS process and located in the head-mounted subassembly (HMS), was designed to implement a portable brain imaging system based on functional near-infrared spectroscopy (fNIRS). The HMS is connected to source modules (SMs) and detector modules (DMs) while communicating with an external microcontroller. fNIRS IC drives the light emitting devices in SM which consists of dual VCSEL diodes having wavelengths of 780 and 850 nm. The module uses silicon photodiode (Si-PD) for light detection and performs optimum filtering, quantization and serialization in the receiver chain (DM) that also includes an off-chip transimpedance amplifier (TIA) with a minimum detectable light power of 400 fW. The emission power of 780 nm and 850 nm lasers can be controlled between 2.5–9.7 mW and 3–10 mW, respectively. Similarly, the gain of the receiver can also be tuned from 12 to 80 dB. The SMs and DMs are attached to a metal spring to ensure tight contact with the forehead of a subject. The measured results are wirelessly transmitted to a tablet device for real-time visualization through Bluetooth communication.

2.7. Summary

In this chapter we discussed briefly about the various brain imaging modalities. The various building blocks of NIRS system are described such as light emitters and detectors along with a comparison of their working and basic characteristics. We discussed about the various fundamental properties of single photon detectors and their figures of merit such as dark noise, spectral response and timing jitter. We presented an extensive comparison of the various SPAD structures implemented in standard CMOS technology. In addition, a brief study of the various SPAD quench and recharge circuits were also presented. This is followed by a survey of state-of-the-art and next generation NIRS systems that have been used for research and clinical studies in the last six years (2012–2018). Most of them considered to be a breakthrough as compared to classical laboratory systems which were typically based on massive accessories such as bulky gas lasers. They are compact to the level of being transportable out of the lab and some of them can operate at two or more wavelengths with the facility of parallel acquisition of several channels. Below are the main conclusions:

1. VCSELs are ideal for sensor applications, such as Biomedical Fluorescence Sensing, Label-free sensing, Displacement sensing, in terms of cost of fabrication, power output, ease of fabrication, beam profile and sensitivity.
2. Several detectors have been reported in literature, many among them have good single-photon sensitivity and timing resolution but SPADs are the most suitable because they are not prone to be damaged by strong light pulses, small and can be easily integrated into CMOS technology.
3. SPADs can be characterized in terms of dark noise composed of primary and secondary pulse/afterpulses, single-photon sensitivity measured as the PDP at various wavelengths and timing jitter which is the statistical distribution of the photo-response.
4. SPAD requires the quench and recharge circuitry for operation. Among the various techniques, passive technique is very simple as compared to active technique but not very effective for SPAD characterization. Mixed quench and recharge circuitry, a hybrid of active and passive techniques, is quite complex but well suited for measurements and any applications owing to its tight control on measurement dead time.
5. The Gated-mode technique is highly appealing for applications where the photon arrival is known as it can be used to significantly reduce dark count and helps in improving the signal-to-noise ratio of the measurements by masking the useful photon signal.
6. The reported TD systems provided proof of principle results after conducting clinical studies and are based on novel approaches to implement advanced technological solutions aiming at improving performances of TD-NIRS devices.

In the next chapter, we will discuss the theoretical background, Monte Carlo simulation for preliminary validation of concept and methodology. A simplified photon detection scheme is developed for experimental validation of proposed scheme.

CHAPTER 3 THEORY AND METHODOLOGY

In Optical Time-Resolved Reflectance, a pair of injecting and collecting optical fibers are placed at a fixed source-detector separation from each other typically in the range of 20 to 40mm limited by the detector dynamic range. To increase the sensitivity to higher penetration depth of investigation, the source and the detector separation should be small. We first showed with simulation results that short source-detector separations results in the detection of a higher number of photons coming from a greater depth. However, at these shorter distances the number of early arriving photons also increase (mainly coming from the skull and scalp regions in brain imaging) which is a constraint. To reject the early arriving photons we need a gated detector to enable detections at specified time windows. We introduced the proposed time-gated technique for the detection of photons in specified time windows. A simplified photon detection scheme is developed to validate the concept and the results were confirmed in an elaborate experimental study. The dependency of the photon counts on the gate window and the source-detector separation is analyzed. We conclude that placing the laser source and the detector quite close to each other is an option to consider for the design of optodes so as to improve the image quality in various biomedical appications.

3.1. Introduction

The study of light propagation into highly diffusive media like biological tissues is highly fascinating because we can explore the medium non-invasively and recover information both on absorption and on scattering. In the Near-Infrared spectral range, light attenuation by the biological tissue constituents is relatively low and allows for penetration through several centimeters of tissue. The difference in the absorption spectra of oxygenated and deoxygenated hemoglobin allows the separate measurement of the concentration of these two species (O_2Hb and HHb, respectively). In reflectance optical spectroscopy measurements, the diffusive medium is illuminated from a point source and diffused photons are collected at a given distance from the source and thus the information on tissue constituents is measured. The study of very faint light signals in the near-infrared wavelength range is an effective technique to study biological tissues in functional brain imaging, optical mammography and molecular imaging, fluorescence lifetime imaging, fluorescence correlation spectroscopy, quantum information and many others. Time-resolved reflectance measurements at null source-detector separation yields better spatial resolution and contrast as compared to the classical approach which employs a separation of 20 to 40 mm [1]. Practical implementation and performance of the approach based on the use of small but not null source-detector separation was also discussed [2,3,5]. It has been shown in a previous paper [2] that a time-resolved setup combining Time-Correlated Single Photon Counting (TCSPC) and fast-gated Single Photon Avalanche Diodes (SPADs) is able to exploit the late photons for reflectance measurements at short source-detector separations (5 to 15 mm). The time-gated acquisition to suppress early photons [3] demonstrates the advantages of using short source-detector distances.

In this chapter we have first introduced the proposed time-gated detection and simulation is conducted to demonstrate that short source-detector separations results in the detection of higher number of photons coming from a greater depth. We then confirm these results in an experimental study with a setup involving a laser source and a fast-gated single-photon detection module using a simplified detection scheme. We have also shown that the count rate of the detected photons is dependent on the gate-ON time of the gated detector module as well as the source-detector separation. Our goal is to consider the design of optodes integrating both source and detector so as to improve the image quality in various biomedical applications.

3.2. Proposed Gated Detector

The detector (in this case SPAD) can be biased above breakdown voltage during a very short period of time, commonly known as gate which is adjustable in width and frequency and can be periodically repeated, with respect to the trigger frequency (external or internal trigger). The elegance of this technique is that the detector is only sensitive during the gates (Fig. 3.1) which results in the reduction of dark noise. The gated mode is used for applications where the photon arrival is known as the impinging photon won't be detected if the gate signal is not applied or a deadtime is initiated after an earlier detection. When an avalanche occurs within the gate due to the detection of a photon or a dark count, a pulse of adjustable width is produced at the output of the detection chain or electronics. At the same time, the quenching circuitry is triggered which closes the gate and a dead time is applied, resulting in one or more blanked pulses.

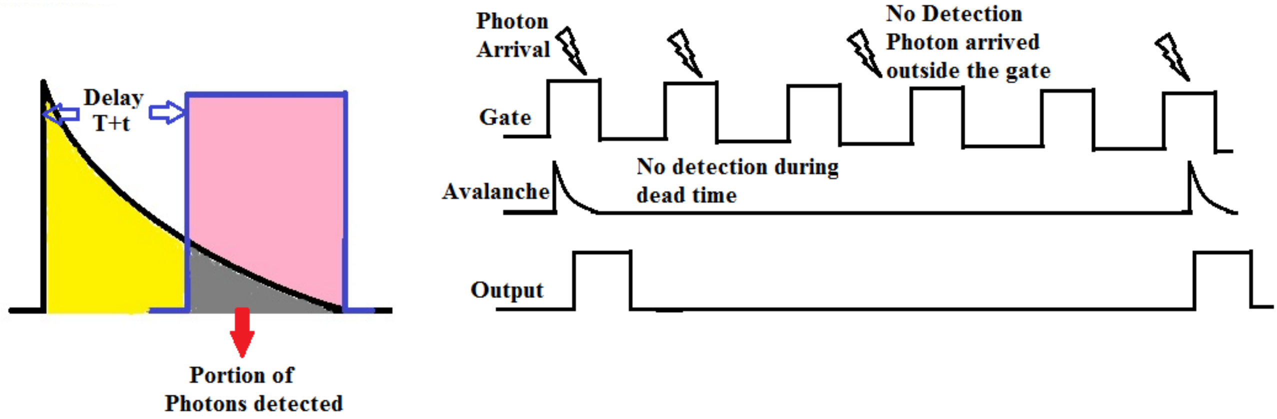


Figure 3.1. Proposed Gated Mode Description

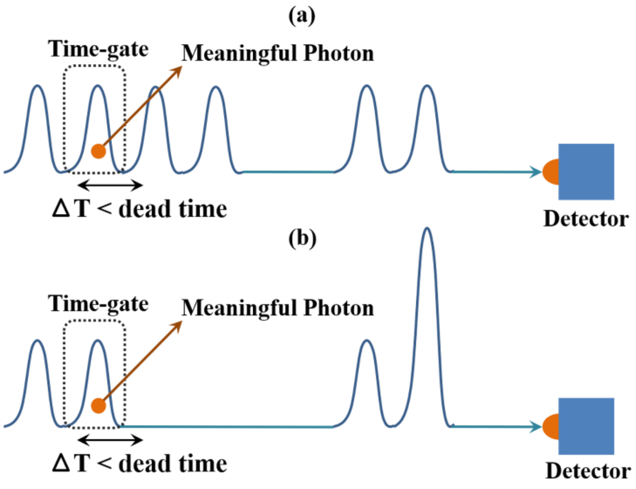


Figure 3.2. (a) Applying a gate ensures that the detector is active when the photon of interest arrives, even if the detector is constantly being illuminated. (b) Applying a gate ensures that the detector is not blinded by the preceding strong pulse

Gated detector [111] posseses or exhibits some significant advantages in the following two situations illustrated below:

(a) Detection of photon concealed in continuous faint light: Figure 3.2 (a) shows the scenario in which the detector is continuously illuminated by very faint signal in which the average time interval between two subsequent photons is smaller than the deadtime of the detector. A free running detector will thus be unable to detect the photons of interest hidden in the beam and will be saturated or completely blinded in this situation which is very predominant in applications such as quantum-cloning and faithful swapping experiments.

(b) Detection of photon arriving after a strong pulse: When a strong light pulse impinges on the detector just before the arrival of the photon of interest, a free running detector will not be in a state to detect the weak signal as it will be either in the dead time period or the weak signal will be overshadowed by the dark noise and afterpulsing effects. This situation (Fig. 3.2b) is very common in applications such as optical time-domain reflectometry, fluorescence spectroscopy and quantum-memory experiments, all of them requiring strong preparation pulses.

Hence, detector gating is an effective technique to keep the detector disabled during intense photon fluxes (e.g. caused by a reflection of the main laser excitation or by "early" photons back diffused by superficial layers of the tissue under investigation), while enabling its single photon sensitivity only during desired gate-ON time windows, when the signal is expected. This gating feature can be exploited to drastically reduce the effect of DCR and afterpulsing. It also helps in improving the signal-to-noise ratio of the measurements by masking the useful, but faint, photon signal. Moreover, it allows to acquire photons only in short and well-defined time slots whilst rejecting all the others by performing a time-dependent windowing of the optical signal under investigation. This feature is very useful in time-resolved spectroscopic investigations of biological tissues at small source–detector separation, where the "early" photons scattered by superficial layers like the skin outnumber the few "late" photons traveling through deeper layers of the sample (i.e. suffering higher absorption probability), like in the brain, by several orders of magnitude. It enables the detection of those few "late" photons while rejecting the burst of "early" arriving photons. It has already been proved that, in time-resolved diffuse optical spectroscopy for a reflectance measurement with very less or null source–detector separation, the temporal separation between the "late" photons and the "early" photons is in the order of less than 1 ns. Such experiments become feasible only with a fast-gated detector module with transition times below 1 ns to guarantee the rejection of the "early" photons and the correct reconstruction of the "late" photons' diffusion curve.

3.3. Monte Carlo Simulation of Photon Propagation Through Diffusive Medium

In a semi-infinite homogeneous medium, when the source and detector are separated by a distance of ρ, the number of photons per unit area and time emerging from tissue after having reached at least depth z is given by [1]:

$$N(\rho,t) = p(z,t)\, R(\rho,t), \tag{3.1}$$

where, $R(\rho,\ t) = k\, t^{-5/2}\, exp(-\mu_a\, \upsilon t)\ exp\left(-\frac{\rho^2}{4\,D\upsilon t}\right) S(D,s_0,t)$

R(ρ,t) is the time-resolved reflectance i.e., the total number of photons per unit area and time emerging from tissue at distance ρ, and p(z,t)is the time-resolved probability that photons penetrate at a certain depth before being detected at the tissue surface. k is a constant, S(D, S0, t) is a dipole term taking into account boundary conditions and thickness S0 of the probed medium, $D = (3\mu_s^{'})^{-1}$ is the diffusion coefficient, $\mu_s^{'}$ is the reduced scattering coefficient, μ_a is the absorption coefficient, and v is the speed of light in the medium. Since p(z, t) is independent of the source-detector separation ρ and of μ_a , we can write that the relative percentage ratio in photons, at a given time, due to the use of a finite source-detector separation ρ as compared to ρ = 0 is :

$$G(\rho,t)\ = \frac{N\ (\rho,t)}{N(0,t)}\ = \frac{p(z,t)\ R(\rho,t)}{p(z,t)\ R(0,t)}\ = \frac{R\ (\rho,t)}{R\ (0,t)}\ = exp[-\ (\rho/\rho_0)^2] \tag{3.2}$$

where, ρ_0 is $\sqrt{4}$Dvt. We observe a significant loss of photons when we perform Time-resolved reflectance measurements with values for ρ in the range 20 – 40 mm as compared to that in the case of a null source detector separation. Also depth information is not degraded when we reduce the source-detector distance because p(z, t) does not depend on ρ. We have tested the proposed approach (ρ = 0, Null Distance Source-Detector) against the classical one (ρ>>0) by means of Monte Carlo simuations. The latter deals with the transport of an infinitely narrow photon beam perpendicularly incident on a multi-layered tissue. The medium is defined as a cube with cube-shaped voxels, typically a 100x100x100 voxel cube. Each voxel is assigned an integer value which identifies a particular type of tissue with unique optical properties of μ_a, μ_s and g. We considered a simple adult head model consisting of four different homogenous media including a surface, CSF, Gray-matter and White-matter layers. The surface layer imitates the scalp and skull. The thickness of each layer is 10mm, 2mm, 4mm and semi-infinite respectively (Fig. 3.3). In the simulations the source and the detector were placed on the surface separated by a certain distance. A large amount of photons (10^6 to 10^7) were injected into the medium and a certain percentage of the injected photons are detected. The diffused photons probe a volume corresponding to a “banana shape”. The index value depicts the total number of photons that are coming from the depth (x=14mm) divided by the total number of detected photons. The depth is chosen in the gray mater. The index values are shown for different positions and depth both for the classical method (source-detector distance, ρ=40) and the proposed method (Null

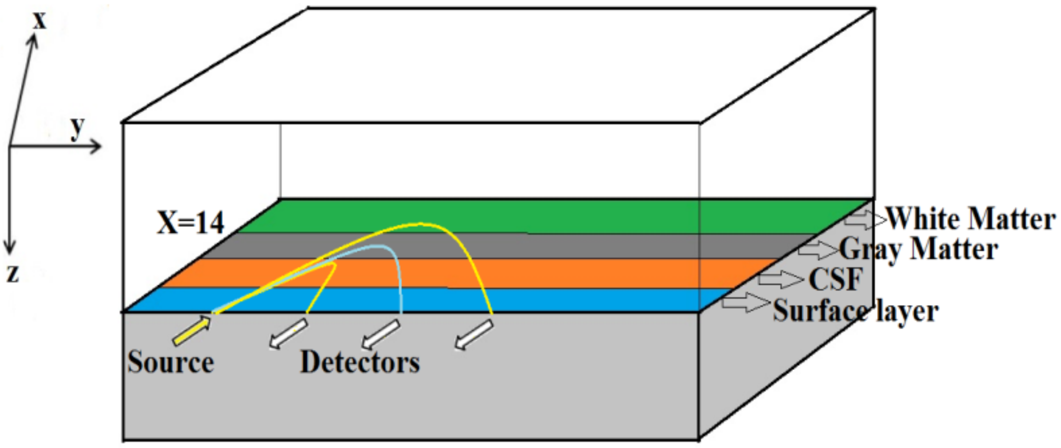
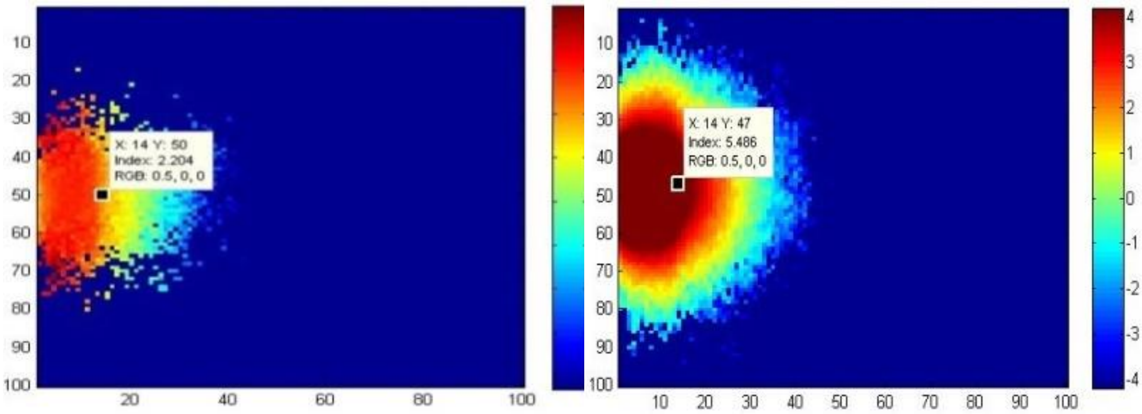
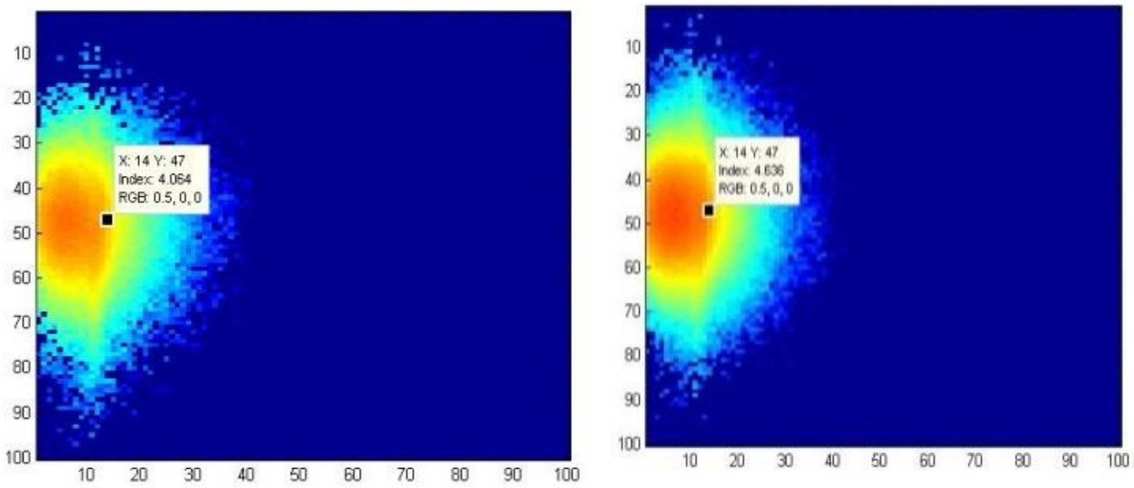


Figure 3.3. Illustration of the sensitivity of different source-detector separations, modeled and simulated by Monte Carlo, using fluorescent light propagation in the defined multilayers diffusive medium

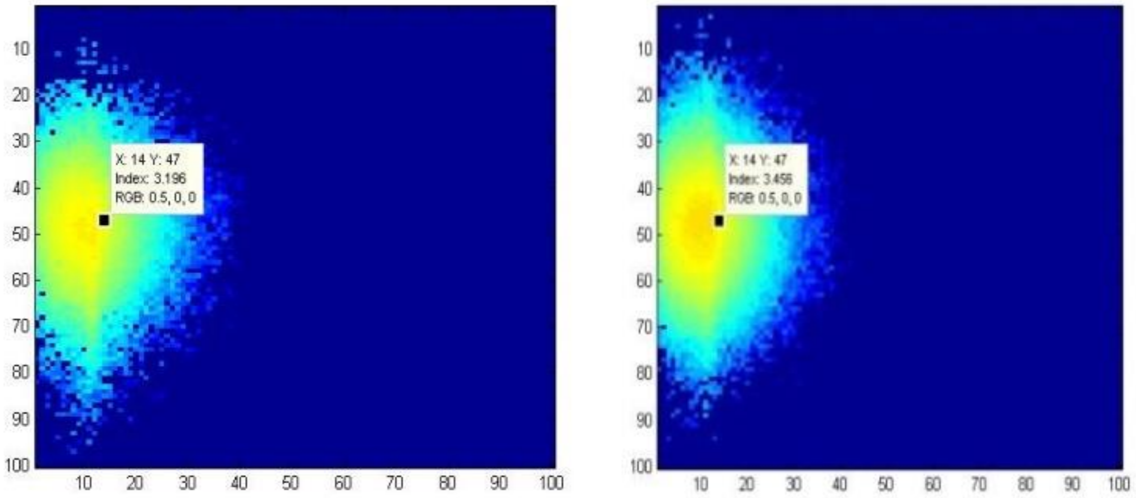
source detector distance, ρ=0). Sensitivity profiles are shown in Fig. 3.4. Reducing source-detector separation results in a better localization of photons in the proposed approach as compared to the classical one and also provides better spatial resolution. As seen in Fig. 3.4, when the source-detector is separated by 40mm, the index value reduces to a great extent indicating a significant loss of photons as compared to the null source-detector separation. The most severe problem as observed in Fig. 3.4 is the presence of early photons at zero source-detector distance, that increases upon decreasing ρ at a much faster rate as compared to the late photons. One possible approach is to use a detection system operated with a time gate with rise time in the range of 100-200 ps and delayed appropiately with respect to the initial burst of photons. This enables the detection of only the late photons while rejecting the early photons that are coming from the surface of the brain and so do not contain any relevant information. To support the experimental feasibility of the proposed gating approach, for short source-detector separation, the detector is gated from 0.4ns to 1.5ns i.e., the detector is turned off for the first 0.4ns. So, we can observe the decrease in the number of detected early photons. As the time gate shifts from 400ps to 600ps, the number of detected photons (both early and late) further decreases as is inferred from the index values. The major advantage of gating is to avoid the detection of the early burst of photons coming from the surface area due to which the detector will be completely blinded and henceforth will not be able to detect the late photons. One more inference, which can be drawn after seeing the index values, is that the number of detected photons is independent on the source-detector distance.



(a) ρ = 40mm ρ = 1mm



(b) ρ = 40mm ρ = 1mm



(c) ρ = 40mm ρ = 1mm

Figure 3.4. Comparison of the Sensitivity profiles for the classical approach (ρ = 40mm) and for the novel approach (ρ = 1mm) (a) no Time gating, (b) detecting photons after 400ps and (c) 600ps

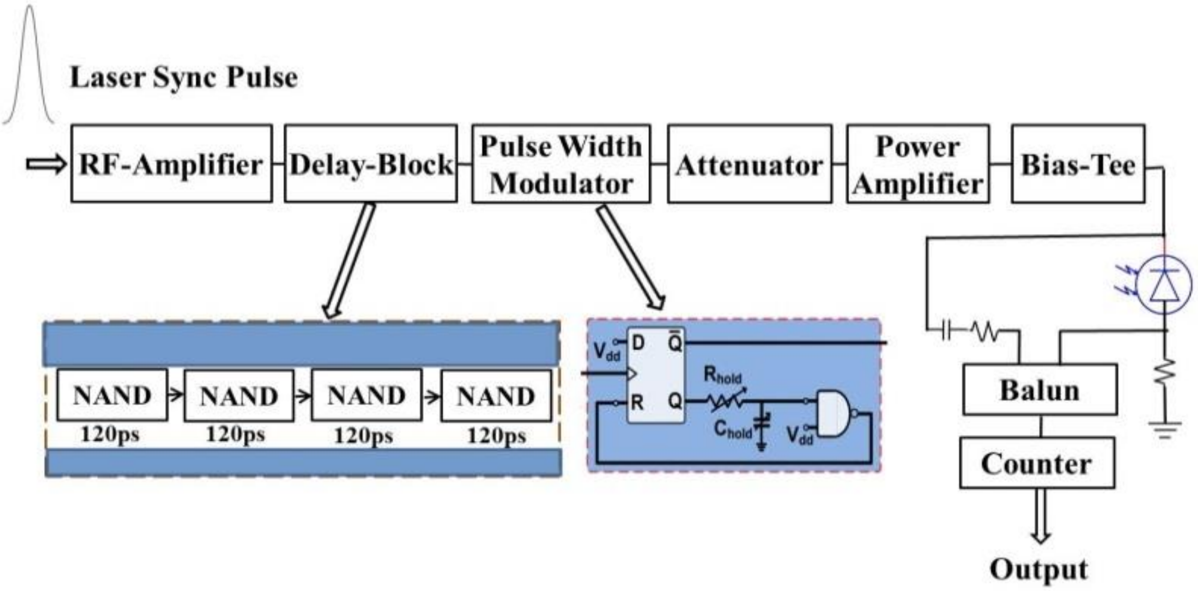


Figure 3.5. Complete block diagram of proposed Gated Single Photon Detection Unit

3.4. Proposed Single Photon Detection Unit

The laser source deliver a synchronization electrical signal once it injects a light pulse into the diffusive medium. So, it denotes the time instant when the photons are emitted from the laser source. This acts as a reference signal for gating to perform a time filtering of incoming photons by acquiring photons only in short and well-defined ime slots whilst rejecting all the others. The laser synchronisation signal which is acting as the “Trigger-In” signal is first fed to the RF amplifier, which has a gain of about 25dB. The 70mV Trigger signal is amplified to 2.5V. The delay block accepts signal from 0 to 3V and provides a Common Mode Logic (CML) level. A series of four logic gates is used and each gate is providing a delay of 150ps. They have CML outputs which provide matching internal 50 ohms termination and produce 400 mV output swings when externally receiver terminated with a 50 resistor to supply voltage. In this way it is possible to get a square wave whose rising edge is in phase with the trigger signal. After the delay block the signal is fed to the Pulse Width Modulator which consists of a D flip-flop followed by a configuration of the resistor and capacitor as shown in Fig. 3.5. The time duration of the pulse can be tuned by varying the values of the resistor and the capacitor based on he application and trade-off among SPAD parameters. The signal is fed to the clock input of the D-Flip Flop whose other input is set to logic “1” in order to form a stage that toggles at each clock pulse. The overall amplication factor required to gate a SPAD with up to 9V excess bias starting from a CML signal with 400mV of logic swing is approximately 20 (>30dB). No commercially available wideband amplifier has such gain and features. So, a minimum excess bias of 2 to 5V is chosen to have a Photon Detection Efficiency(PDE) of 5 to 10% and an amplifier with a gain of 18dB is chosen. The appropiate high power RF amplifier is chosen from Minicircuits (ZHL-2) and is installed in the setup after an attenuator. The variable attenuator reduces the amplitude of the signal and thus can be used to adjust the peak-to-peak amplitude of the gate signal. The function of the bias-tee is to simultaneously allow a DC bias voltage and an RF signal to be applied to the SPAD for varying its bias voltage . The purpose of the inductor is to prevent the RF signal from entering the DC path and the purpose of the capacitor is to block the DC signal from entering the RF path. The inductor and capacitor should be chosen such that the upper cut-off of the low pass DC path is lower than the lower cut-off frequency of the high pass RF path. For the processing of the avalanche pulses a new PCB is designed with a new auxiliary line, simulating the capacitance and resistance of the detector has to be introduced to produce equal spikes which are then subtracted from the signal using a balun (Minicircuits RF transformer, TCM4-1W). The avalanche will be easily discriminated from the residual spikes after the balun which will be followed by a counter (NXP DataAquisition Unit).

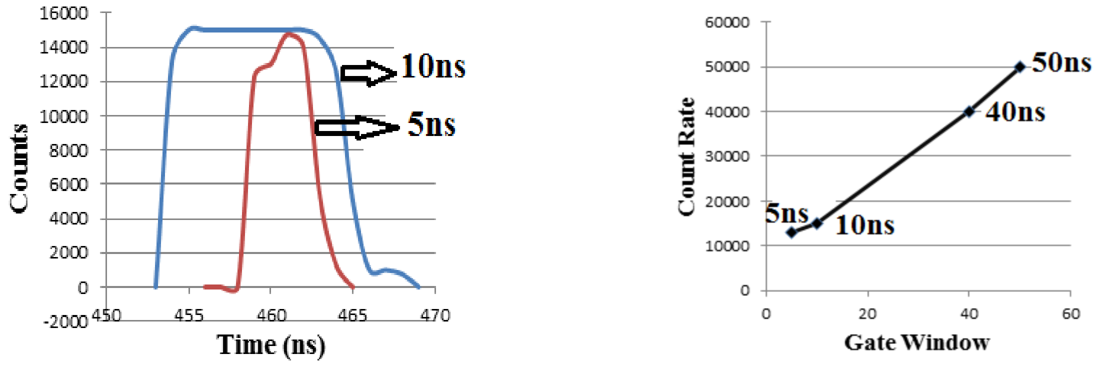
3.5. Experimental Results

The laser source has a wavelength of 780 nm and the pulse duration is 180ps. The laser power is attenuated to a very low level. An electrical pulse, generated by the laser source, acts as a synchronization signal for the gated single photon detection module. The laser pulsing is fixed at a certain time and then the gating pulse is swept across by applying a variable delay. The number of detections (avalanches occurring during these gates) are recorded versus the delay between the rising edge of the gate and the laser pulse. In our work we used a commercially available single photon detector of Excelitas - C30921SH. An external delay generator is also used to have a tunable delay over a wider range.

Fig. 3.6 shows the count rate distribution for different gate-ON time periods. The count rate can be expressed as follows:

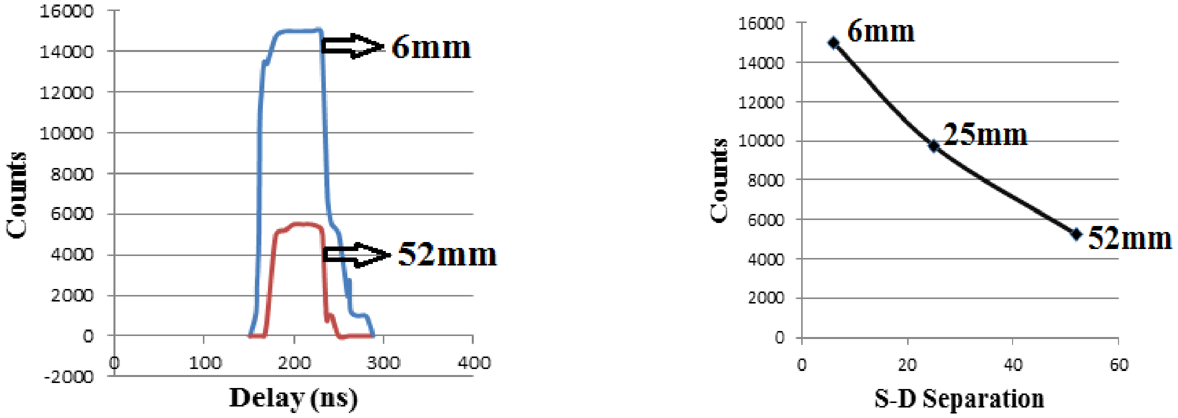
$$R_T = \frac{Nt}{f\ tg\ t} \tag{3.3}$$

where, N_t is the total count rate which comprises of both the dark count rate and photon count



(a) (b)

Figure 3.6. (a) Count rate distribution for 5ns and 10ns gate-ON time window, (b) Dependence of Photon count rate on the Gate window



(a) (b)

Figure 3.7. (a) Count rate when SPAD and Laser source are separated by 6mm and 52mm (b) Count rate vs Source-Detector distance

rate, t_g is the gate-ON time , f is the gating frequency and t is the integration time. It is understood from equation (3.3) that count rate is linearly dependent on the gate width. The total count rate is a function of the excess bias voltage applied to the detector as follows :

$$N_t = \frac{f\, tg\, t\, (Vr - Vb)}{Rs} \tag{3.4}$$

It can be inferred from equation (3.4) that total count rate is directly proportional to the gate-ON time (t_g), the gating frequency (f) and the excess bias ($V_r - V_b$) where R_s is the quenching resistor. The variance of count rate at different Gate-On time windows is shown in Fig. 3.6 and it can be concluded that the count rate increases almost linearly with the increase of the gate-ON time windows. The photon count rate also depends on the separation between the source and the detector. An experimental setup was created and the photon counts are measured at different separations between the source and the detector. Both the laser source and the gated single photon detection module are placed on the same side of the homogenous scattering medium. A highly scattering medium (Intralipid, 1 wt% aqueous solution) was used as a mimic of biological tissue. Light is scattered in all directions reproducing realistic in-vivo conditions. This is done to demonstrate that increasing the source-detector distance decreases the number of detected diffused photons. Fig. 3.7 shows that increasing the separation between the source and detector has a negative impact on the number of detected photons. It also shows the variance of the count rate versus the source-detector distance. Short source detector separation increases the total number of detected photons but it also increases the detection of the early photons, coming from the surface, which shadows the late photons in depth imaging. In brain imaging the early photons come from the surface area of the brain (skull, CSF) which do not carry any information and our goal is to detect only the late photons coming from the cortical regions. The solution is the use of a gated detector which allows to acquire photons only in short and well-defined time slots while rejecting all the others by performing a time-dependent windowing of the optical signal under investigation. This feature is mostly used in time-resolved spectroscopic investigations of biological tissues at small source–detector separation.

3.6. Conclusion

The approach to imaging in highly diffusive media based on time-resolved reflectance at short source-detector separation has improvements over the classical configuration ($\rho >> 0$) which has been demonstrated through the Monte Carlo simulations. The experimental results also validate an increased number of collected photons at short source-detector separation. To increase the investigation depth and the signal-to-noise ratio, the source-detector separation can be reduced so as to increase the number of collected photons and also image contrast and spatial resolution. The decrease in the source-detector distance also increases the number of early arriving photons. So the detector has to be gated to reject the early unwanted photons. The experimental results also show the feasibility of incorporating both the laser source and the detector close to each other most preferably on a single chip which will be a great improvement over the traditional method of brain imaging. In the next chapters we have demonstrated the miniaturized optical probes integrating time-gated detection module and pulsed-light emission units.

CHAPTER 4 COMPACT FAST OPTODE-BASED PROBE FOR SINGLE-PHOTON COUNTING APPLICATIONS

Most of the NIRS systems introduced in the last chapter make the use of conventional electronics which significantly increases the size of each optode and leads to a hedgehog cap that cannot comfortably be used in a clinical context where the subject frequently lies on a bed. Herein, we have exploited our unique expertise in micro-electronic design of integrated chips to develop a novel integrated CMOS electronic and optical design to integrate gated detection and pulsed-laser illumination within a single chip, a feat that remained out of reach until now. We have demonstrated the electrical and optical characterization results of the developed optical probe. The following sections are the reproduction of a published article in IEEE Photonics Technology Letters.

- S. Saha, Y. Lu, S. Weyers, M. Sawan, and F. Lesage, "Compact Fast Optode-Based Probe for Single-Photon Counting Applications," *IEEE Photonics Technology Letters,* vol. 30, no. 17, pp. 1515-1518, 2018.

Abstract—We present a compact standalone optode probe integrated with a pulsed Vertical Cavity Surface Emitting Laser (VCSEL) and fast-gated single-photon counting module. The single-photon counting module is fabricated in 0.35 μm HV CMOS technology and coupled to a silicon-photon avalanche diode (SPAD). The integrated 10-μm diameter SPAD has photon detection efficiency of 5 % at wavelength of 850nm and very low dark count rate of 60 counts per second (cps) at room temperature. A compact and very high modulation rate top-emitting GaAs-based VCSEL chip is used to deliver pulsed pico-second light. High intensity short-time reflections, occurring due to the close proximity of the light source and detector, are avoided by fast-gating of the detector which enables the detection of photons in some specified time-windows. A full characterization of the module is done without any cooling, testing both its electric and optical properties. The compactness of the module paves the way for its utilization in a wide variety of single-photon counting applications.

Index Terms— Fast-Gated Detector, Vertical Cavity Surface Emitting Laser (VCSEL), Single Photon Avalanche Diode (SPAD), Near-Infrared Spectroscopy (NIRS).

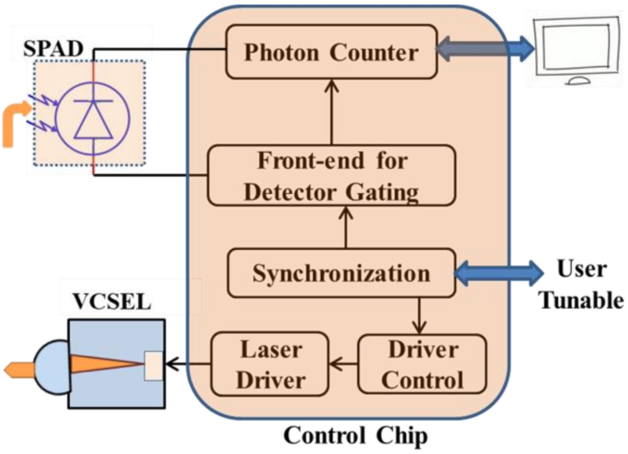


Figure 4.1. Simplified block diagram of the proposed optical probe

4.1. Introduction

There is an increasing trend to scale down optoelectronic setups to create low-cost, portable and fully non-invasive devices. In particular, the study of very faint light signals in the near-infrared wavelength range has shown to be an effective technique for clinical diagnostics and monitoring in portable brain imaging [109, 110], fluorescence measurements[106, 108] as well as other applications such as time-resolved laser Raman spectroscopy and laser-induced breakdown spectroscopy[102]. The accelerated development of photonic components for light sources (e.g., Vertical Cavity surface Emitting Lasers - VCSELs) and detectors (e.g., Single Photon Avalanche Diodes - SPADs) has enabled the scaling down of both size and cost. Several detectors have been reported in literature, for example, hybrid detectors, photomultiplier tubes (PMT) and semiconductor-devices like SPADs [75, 76, 95]. Many among them have good single-photon sensitivity and timing resolution but SPADs are the most suitable because they are not prone to be damaged by strong light pulses, small and can be easily integrated into CMOS technology. The CMOS process offers a cost-effective platform for integration of both digital and analog circuitry. The attachment of VCSEL's and SPAD with CMOS circuitry can lead to a compact optode that is easy-to-handle. Our target application for this design is brain measurements where smaller optical probes are more convenient to build caps with very small optodes, devoid of optical fibers making the imaging process more practical.

The recently developed fast-gated SPAD modules [87] have good timing resolution but they use many discrete electronics and cooling, thus increasing the dimension of the module to a great extent. More recently a miniaturized and compact non-gated single-photon counter was introduced [112]. However, until now, modules presented in literature, did not integrate a light

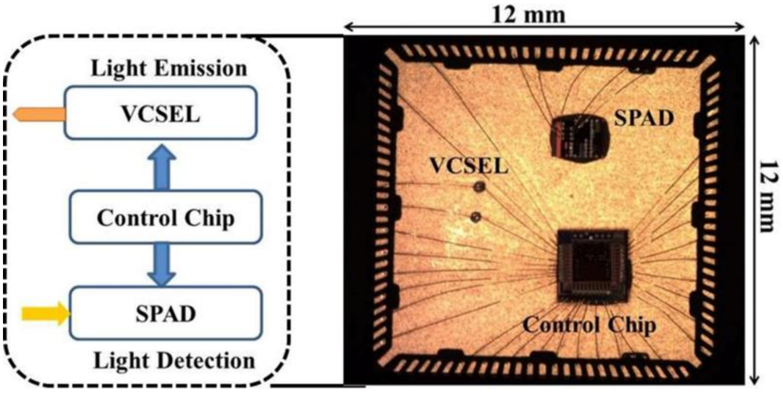


Figure 4.2. Photograph of the Proposed Optical Probe

source and have been developed for research purposes only, lacking scalability and hence cannot be exploited as wearable devices for personal health monitoring. In this paper, we present an optode probe integrating both light emission module and fast-gated single photon counter. A control chip administers and provides the synchronization signal for both the light emitting and detection units. VCSEL, SPAD and Control chips are placed side-by-side in a common package and interconnected at the package level, via wire bonding, often referred to as system-in-package (SIP). Our proof-of-concept SIP provides major opportunities in both miniaturization and hybrid integration of multiple chips, resulting in a compact probe that has potential for widespread exploitation.

4.2. Architecture

Figure 4.1 shows the simplified block diagram of our proposed optical probe. The novelty of our work lies in the integration of a gated photon counter on chip so that a simple output can be generated. The photograph of the prototype optical probe is shown in Fig. 4.2, which houses the CMOS driver chip, two VCSELs and one SPAD chip. In brain-imaging, such integrated optode makes the system very compact and easy-to-handle, relaxing the need for complex source-detector pairings and positioning. With diffuse optical tomography, the close proximity of the source and detector significantly increases the unwanted so-called early photons that are reflected back from surface regions (e.g. skin, scalp). Here we use a detection system operated with a time gate delayed appropiately with respect to the initial burst of photons. This procedure avoids the detection of the early burst of photons which can completely blind the detector and henceforth not be able to detect the meaningful late photons that probe deeper into the tissues. The integrated system consists of two blocks – a Fast-Time Gated Detection Module, and a Light Emitting Module. The light emitting module includes a VCSEL driver wire-bonded to

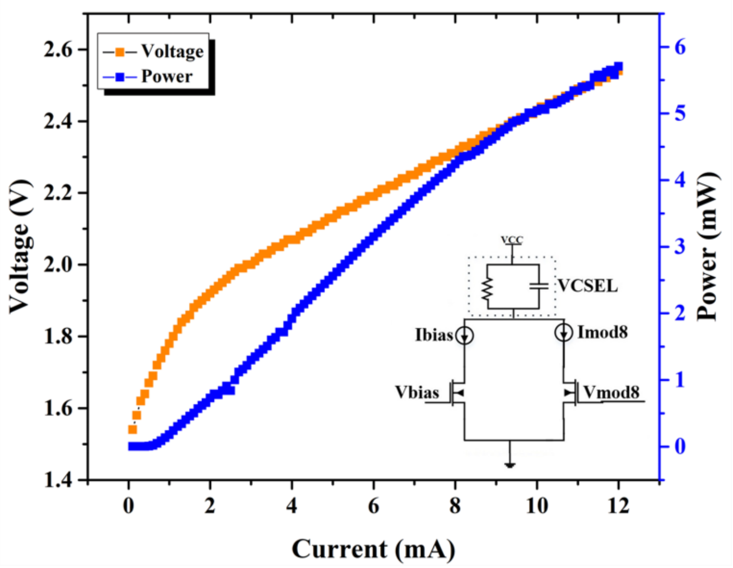


Figure 4.3. Light-Current-Voltage (LIV) module exhibits an optical to electrical conversion efficiency of 5mW/V. Inset is the schematic of the VCSEL driver

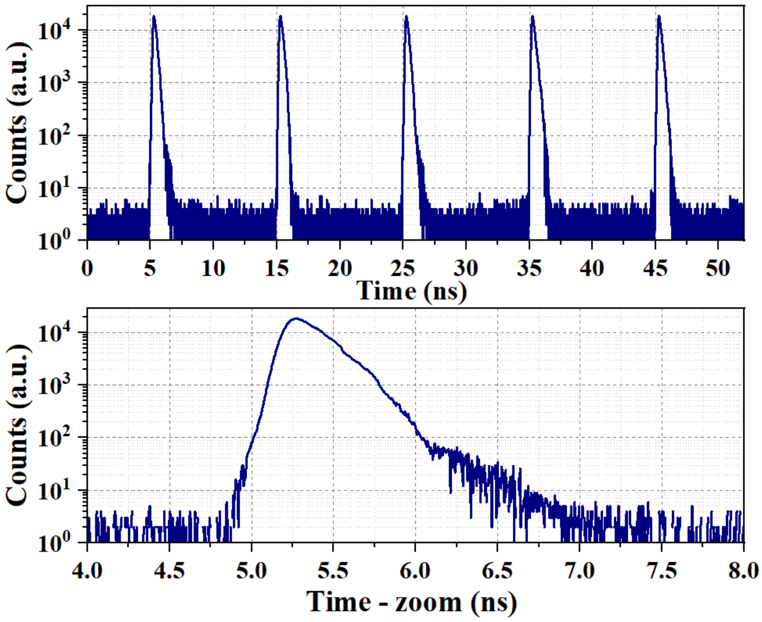


Figure 4.4. The measured optical waveforms, below shows the zoomed section

VCSEL chip (Fig. 4.1). The fast-gated single photon counter includes a control driver wire-bonded to SPAD chip. The Control chip was fabricated in AMS 0.35-μm high-voltage CMOS technology and the SPAD detector was fabricated in standard 0.35-μm CMOS technology (Fraunhofer IMS, Germany). A synchronization block generated pulses of frequency ranging from 2 MHz to 150 MHz that act as a sync signal for both modules. A high-voltage driver,

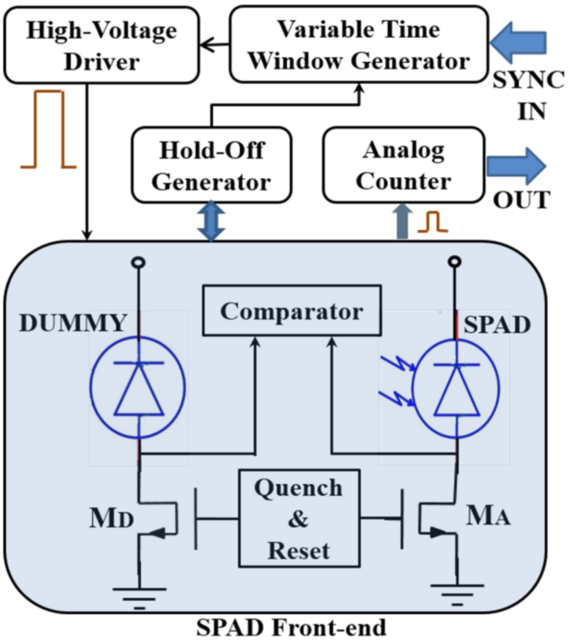


Figure 4.5. Simplified block diagram of the Fast-Gated Detector Module with the integrated counter

comprising of a HEMT transistor acting as a common-source switching stage, provides suitable biasing voltage to the detector.

4.3. Optical Probe Description

4.3.1. Light Emission Module

The VCSEL is a 100-μm top-emitting GaAs-based laser with a very high-modulation rate. The designed VCSEL driver converts CMOS-level input signals to current mode output signals [113, 114]. An integrated signal generator can be used to generate the electrical signal inside or can also process off-chip signal. Figure 4.3 (inset) shows the schematic of the VCSEL driver circuit, working on a simple principle that a MOS transistor in saturation acts as a current source. It is based on a two-transistor model which had their source shorted together : one transistor is designed to provide a bias current to the VCSEL, while the other is used to supply the modulation current. The gate of one transistor is tied to an analog voltage to set the VCSEL bias current while the gate of the other is connected to a CMOS-level digital input signal that controls the modulation current. The drain terminals of the NMOS transistors are connected to the cathode of the VCSEL, while its anode is connected to a single power supply for the circuit. Both NMOS transistors have a drawn gate length of 0.5 µm and a gate width of 30 µm, which

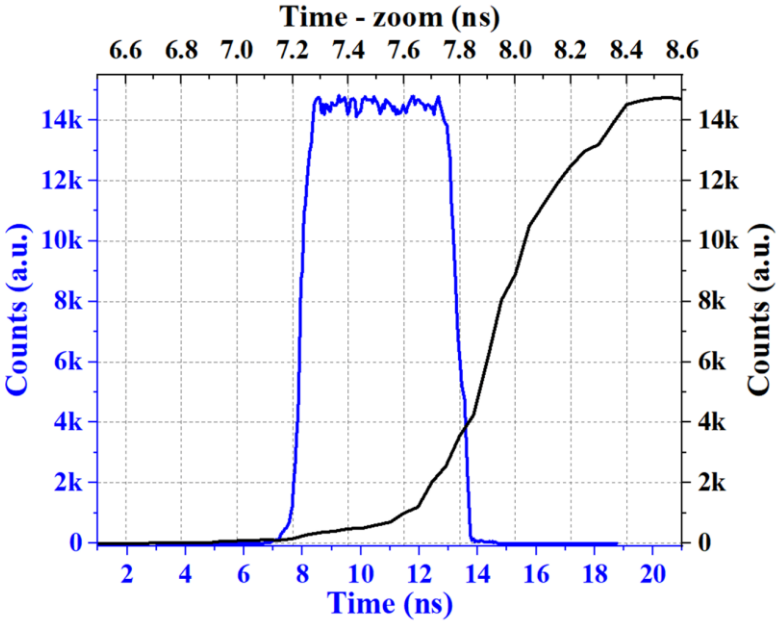


Figure 4.6. Photon counts distribution within a 5 ns gate-ON time window.

were wide enough to allow biasing and modulation currents ranging from 0 to 8-mA. The proposed technique allows for easy integration of VCSEL with CMOS circuitry. An optimized monostable multivibrator, generating pulses with width varying from 500 ps to 4 ns with very sharp transitions, was used to drive the transistors ensuring high-speed operation.

To validate the system, we connected a VCSEL (VI integrated systems, 850nm) and fixed the biasing voltage (Vbias) while varying the source (VCC). The measured peak output power of the VCSEL was 5.8 mW. We first measured the I-V characteristics (Fig. 4.3) and observed the linearity of the optical power with the current. We then aimed at characterizing the operation in pulse mode. The gate modulation voltage varied from 1.0 to 3.0 V with a pulse width of less than 1 ns. The resulting modulated light was loosely focused on a fast silicon single-photon detector, whose output was coupled to a Time Corelated Single Photon Counting (TCSPC) module (Picoquant). Figure 4.4 shows the resulting detected signal from the input-modulated VCSEL with a repetition rate of 100 MHz and FWHM of 350 ps.

4.3.2. Fast Time-Gated Detection Block

The module can be operated in free-running or time-gated mode. The functionality of the time-gated detection unit is to turn-on the SPAD during specific time windows. The module can operate with off-chip sync pulses or the same can also be generated internally by the synchronization block. The variable time window generator modulates the duty cycle of the

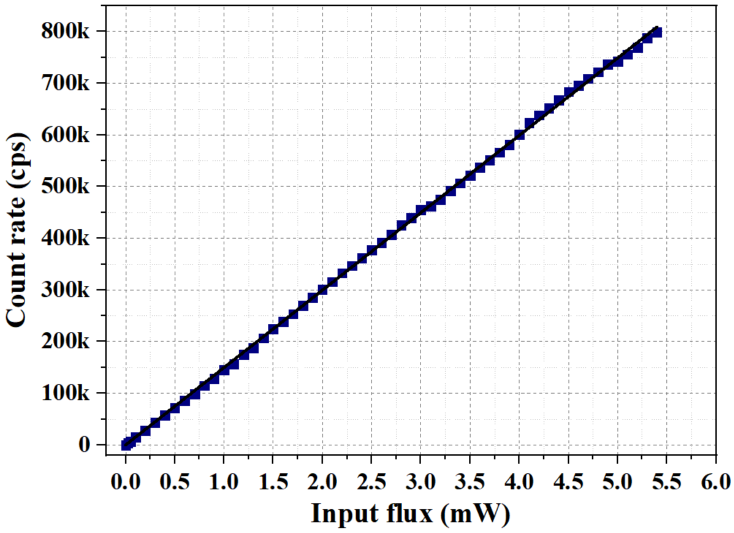


Figure 4.7. Dependence of the SPAD count rate on the input photon flux

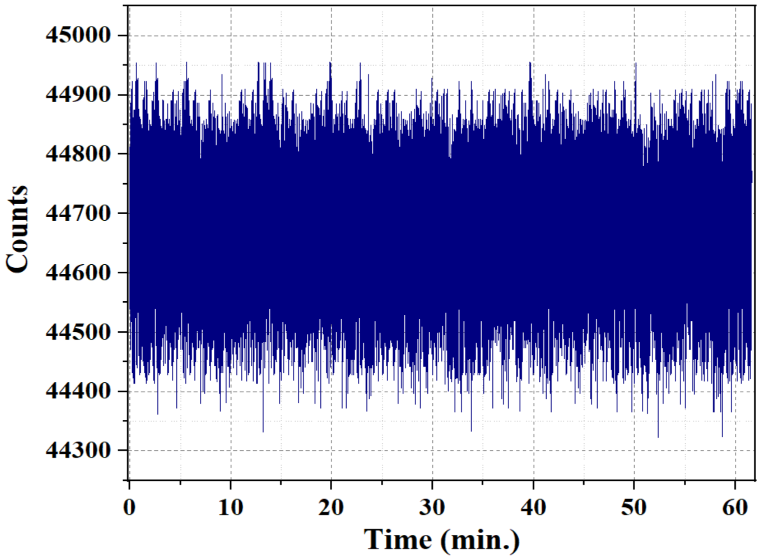


Figure 4.8. Stability of the module count-rate over a time range of 1 hour.

gating pulse to provide a suitable ON-time to the detector ranging from from 1 to 50 ns. A high-voltage driver was designed to adjust and provide a suitable excess-bias voltage to the SPAD (2 to 9 volts) which has a breakdown voltage of 26.5V. The avalanche signal is sensed by the SPAD Front-end consisting of an active SPAD and a dummy one, operating in passive-quenching active-reset mode. The differential topology rejects the spurious noise and a clean rectangular pulse is generated at the output of the comparator. The quench and reset block provides the necessary signals to restore the initial operating conditions of the SPAD and get it ready for another detection. After the detection of a photon, the hold-off generator turns off the SPAD for a specific hold-off time, in the range of 5 to 25 ns. The analog counter counts the photon pulses within an observation window. The small area SPADs (10, 20 and 30 µm) were

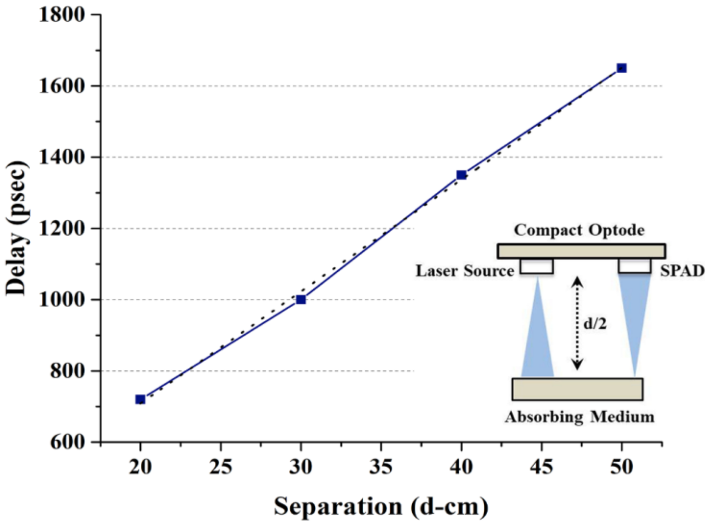


Figure 4.9. Time-of-Flight experiment with the compact module

fabricated in Fraunhofer IMS [76]. The ability of the module to provide a good time-filtering of incoming photons can be demonstrated by Fig. 4.6 which shows the photon count distribution for a 5 ns gate-ON time window. The sharp transition times of the gate window (20–80% rise time of 500 ps with a smooth edge shape) is instrumental in the prompt rejection of the unwanted photons outside the window.

Figure 4.7 shows the relationship between incoming photon flux and measured count rate with the fast-gated SPAD detector. The integration time is 50 ms and was counted for 30 sec at each observation point. It is also observed that the SPAD has extremely low dark count rates (DCR) with a maximum of 60 cps. The slight deviation in linearity might be due to the fluctuation in the optical power of the laser source. Figure 4.8 shows the stability of the count-rate of the single photon counting module over a time-period of 60 min. The photon rate was kept at a very low level using a continuous-wave laser source with the module running at ambient temperature without any cooling system. The fluctuation of less than 1% might be due to poissonian noise and slight instability in the laser output power or ambient conditions. These values highlight the stability of the module without any cooling system. Temperature control using a two-stage Thermo Electric Cooler (TEC) is a major drawback for the miniaturization of the time-gated module to be integrated in an optode with the light source. Due to measured characteristics, our data suggests these SPADs can be used without any cooling modules and are most suitable for the development of on-chip source and detector, which can be used for photon migration measurements.

4.4. Application Example

In order to check the accuracy of the gating circuit, Time-of-Flight experiments were conducted with the integrated VCSEL module. An absorbing medium was placed on top of the optode and the separation ‘d/2’ was varied as shown in Fig. 4.9 (inset). The internal synchronization block provided the triggering signal for both the VCSEL module and the Fast-Gated Detection module. The frequency of the trigger signal was fixed at 5 MHz, which acted as a ‘Sync IN’ pulse for the detection module. The ‘Sync IN’ pulse was then applied to an internal variable time window generator to create a detection time-window of 5 ns. The rising edge of the trigger signal denotes the time instant at which the light pulse is emitted by the VCSEL module. This light pulse is expected to be detected after a time period of (c/d), where c is the speed of light. This time period corresponds to the round-trip time taken by the photons to reach the absorbing/reflecting surface and return back to the optode. The detection window is delayed to observe at which time period the SPAD starts detecting the maximum count. This time-delay is basically the round-trip time taken by the photons. Figure 4.9 shows the linearity of the delay versus the separation.

This experiment showcases the potential of application of the proposed module in Diffuse Optical Tomography. The integration of the pulsed laser source significantly relaxes the need of control over positioning. The internal pulse generator easily synchronizes both light emission and detection frame. The module delivers a digital pulse after each photon detection, which can be easily processed by an external data acquisition unit. It also possesses the functionality of counting the photons internally and delivering an analog output voltage proportional to the number of counts in the observation window, thus eliminating the need of an external counter. The module is user-configurable in terms of its output laser power, photon detection and counting time-frame.

4.5. Conclusion

We implemented and characterized a compact standalone optode probe composed of a VCSEL and fast-gated single-photon detector. The module can be integrated in many experimental setups and is highly robust in terms of its configurability of operating parameters. In recent literature, the cooling system presents a limiting factor in the complete integration of a fast-gated single photon counting module on a single chip. Henceforth, a full characterization of the module, was done without any cooling system, testing both its electric and optical properties. Another limiting factor is placing the source and detector in close proximity to each other which leads to large short-time reflections resulting in detector saturation. This limitation was addressed by fast time-gating of the single photon avalanche diode allowing for the detection of photons in some specified time frame. The novel approach paves the way to the utilization of this module as a cost-effective alternative to bulkier and expensive commercially available systems. It can be effectively used in a wide variety of time-correlated single-photon counting applications, such as fluorescence lifetime imaging (FLIM), functional near-infrared spectroscopy (fNIRS), and fluorescence correlation spectroscopy (FCS).

4.6. Acknowledgement

Authors acknowledge support from NSERC, CIHR, Canada Research Chairs on Smart Medical Devices and Optical Vascular Imaging, and CMC Microsystems for design and simulation tools.

CHAPTER 5 COMPACT OPTICAL PROBE FOR TIME-RESOLVED NIRS-IMAGING

The integrated micro-electronic optode, that has been introduced in the last chapter, was developed to work for the essential required functionality of nsSDD-TCSPC implementation: detection gating of early photons and fine control of the pulsed laser. Moreover, laser and detector were combined in a single package demonstrating integration feasibility. Miniaturization is possible as the probe uses detector and laser source side-by-side, leading to a very small SDD, which will facilitate clinical adoption in many applications and change the landscape of fNIRS hardware commercially due to significant optode-size reduction and the elimination of optical fibers. Herein, we have described in detail the architecture of the proposed system and have showcased a more elaborate characterization to exhibit its suitability in real-life applications. The following sections are the reproduction of an article currently under revision in IEEE Transactions on Circuit and systems - I.

- S. Saha, Y. Lu, S. Weyers, F. Lesage, and M. Sawan, "*Compact Optical Probe for Time-resolved NIRS- Imaging*," under review in IEEE Transactions of Circuits and Systems I: Regular Papers.

Abstract— We present the design of a standalone optical probe integrated with a Time-Gated Single Photon Detection module and Pulsed Light Emission unit. The detection module is capable of detecting photons in specific time windows, thus enabling the measurement of very faint optical signals. The pulse repetition frequency extends up to 100 MHz with a gate-ON time range up to 35nsec. The module is user-configurable via a PC interface and can be integrated in any optical setup. The pulsed light emission unit uses a 50 GHz GaAs Vertical Cavity Surface Emitting Laser diode. Experimental characterization was conducted for all the building blocks of the module to ensure its proper functioning. The module was also employed in a real life application, heart rate monitoring by illuminating the index finger of the volunteer. The novelty of the proposed miniaturized optode lies in the development of a simplified interface for portable Brain-computer interfaces and can be effectively used in applications such as Near-Infrared Spectroscopy and functional brain imaging.

Index Terms— Time-Gated (TG) Detector, Vertical Cavity Surface Emitting Laser (VCSEL), Single Photon Avalanche Diode (SPAD), Near-Infrared Spectroscopy (NIRS).

5.1. Introduction

The measurement of very faint light signals propagating through highly diffusive medium like biological tissues has proven to be an effective technique to investigate the medium non-invasively and gather information both on chemical composition and tissue function [115]. In the Near-Infrared wavelength range, light attenuation by the biological tissue constituents is relatively low and allows for penetration through several centimeters of tissue. In reflectance optical spectroscopy measurements, the diffusive medium is illuminated from a point source and diffused photons are collected at a given distance from the source to avoid detector saturation when tissue constituents are measured. The analysis of the time of flight distribution of detected photons can yield measurements of the concentrations of oxygenated (O_2Hb) and deoxygenated hemoglobin (HHb). Time Domain (TD) Functional Near-Infrared Spectroscopy (fNIRS) has been used to study biological tissues in optical mammography, fluorescence molecular imaging, brain monitoring and many others [12, 15, 16, 116]. In the case of reflectance-only measurements, the main challenge is depth sensitivity. In diffuse optical tomography (DOT) instruments, depth sensitivity is addressed by using larger source-detector distances so that the collected photons probe deeper layers of tissue. However, the use of small source-detector distances based on recent development in time-domain technology [9, 13, 117] enables the use of optical probes that are easier-to-handle, making the measurement more practical. For brain measurements, smaller optical probes increase convenience of cap design and could yield a simplified image reconstruction.

Practical implementation and performance of the approach based on the so-called null source-detector separation was demonstrated using a fast-gated Single Photon Avalanche Diode (SPAD), achieving a dynamic range of 8 decades, almost double to that of the conventional non-gated detection [5, 90, 116, 118]. SPADs are semiconductor p-n junction devices, reversed biased at a voltage higher than the breakdown voltage, which can trigger a self-sustaining avalanche pulse even if a single charge carrier is injected into the depletion region. The avalanche pulse is then processed by the read-out electronics, marking the arrival time of the photon. One major advantage of using SPAD is that, due to its low capacitance, it can be turned ON during a very short period by biasing it above the breakdown voltage with an applied voltage pulse, making it suitable for fast time-gated operation. When the voltage pulses are not applied, the SPAD remains in its off condition. This also prevents charge accumulation and reduces the dark count rate (DCR). Since, the SPAD is only sensitive when voltage gates are applied, a time-gated mode is highly appealing for applications where the photon arrival is known. For reflectance measurements at short source-detector separations (5 to 15 mm), fast time-gated SPADs [17, 111] can be exploited to collect the so-called meaningful late photons while rejecting the burst of unwanted early arriving photons.

A large time-gated module was built that demonstrated 110 psec transition time, 30 psec temporal resolution, and up to 80 MHz gate repetition rate[87]. Recently, a miniaturized and compact non-gated single-photon counter was also reported, displaying low dark count-rate and good timing resolution without time-gating capability[112]. A compact probe based on a fast silicon photomultiplier (SiPM) also equipped with laser diodes was developed for TD-fNIRS[119]. A complete instrument based on pulsed laser diodes and SiPM was designed for multichannel and multi-wavelength TD NIRS but its size remains too large to be put into small sockets for wearable applications such as brain monitoring [106] . All these modules used discrete electronics and none of the systems were fully CMOS integrated. The id110 module [120], developed by ID Quantique SA (Switzerland), has higher gate repetition frequency but has poorer time resolution and lower ON-time duration range. Other devices like Intensified Charge-Coupled Device (I-CCD) cameras are very expensive and bulky. They are also prone to be damaged by strong light pulses.

Scaling down Time Resolved Diffuse Optical Spectroscopy (TR-DOS) setups could help create low-cost, portable and fully non-invasive devices for optical mammography and functional brain imaging. Recent developments and growth in photonics components for light sources (e.g., Vertical Cavity Surface Emitting Lasers - VCSELs) and detectors (e.g., SPADs) has enabled the scaling down of both size and cost. We present in this paper the design and characterization of a complete and stand-alone optical probe integrated with a fast-gated single photon detection and pulsed light emission modules. The fast-gated single photon detection module is capable of time-gating the SPAD detector with repetition rates up to 160 MHz and gate-ON width extending to 35 nsec. The fast-gated detector enables the detection of photons in a specified delayed time-window, thus avoiding the detection of early photons that may be due to short-time reflections. The pulsed light emission module comprises a top-emitting GaAs-based VCSEL chip. The module is configurable in terms of operating parameters including

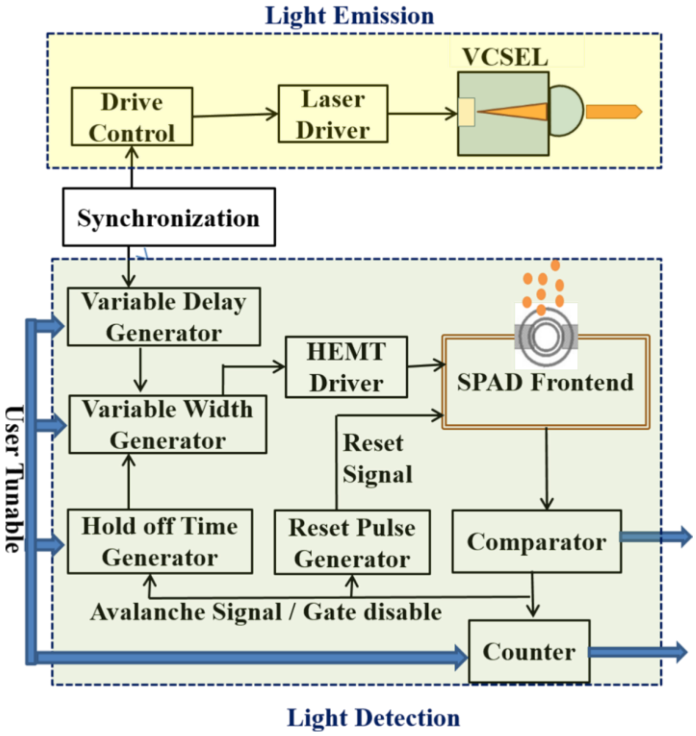


Figure 5.1. Block Diagram of the Compact Optical Probe

detector biasing voltage, hold-off time, modulation rate of the light source and photon counting window. The miniaturized version in the form of an integrated optode, relaxes the need of control over source-detector positioning but significantly increases the unwanted early arriving photons. This limitation is overcome by exploiting the time-gated single photon counting module enabling the exclusive measurement of late photons in the presence of overwhelming amount of early light. SPAD cooling is the main limitation for the scaling down of the time-gated reflectance setups. Hence, to evaluate the performance of our system, all the experiments are performed at room temperature.

The remaining of this manuscript includes the description of the proposed module in Section 5.2. We analyze and describe in Sections 5.3 and 5.4. the various building blocks of the system. Section 5.5 reports the electrical and optical characterization and some of the target applications. Section 5.6 concludes the paper.

5.2. System Description

Figure 5.1 shows a simplified block diagram of the developed compact optical probe. The originality lies in the integration of the pulsed light emission and gated light detection modules in a single package. The light emission module comprises a VCSEL diode wire-bonded to a

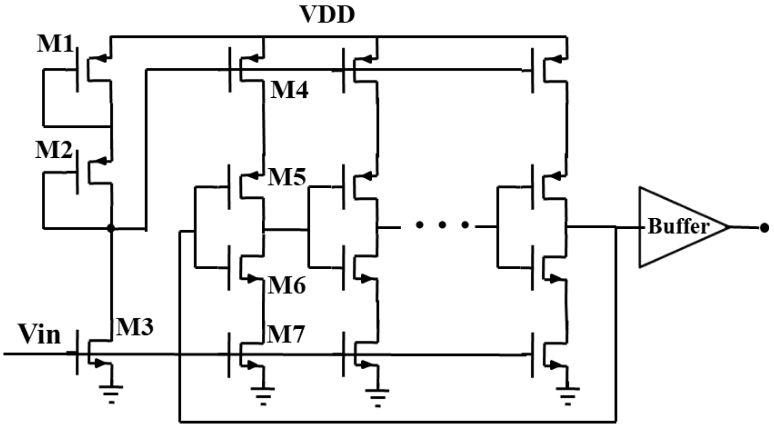


Figure 5.2. Schematics of the Synchronization Block

custom CMOS VCSEL driver which is integrated in the control chip. The light detection module includes a SPAD as detector, also controlled by the control chip. An internal signal generator was designed to generate synchronization pulses of frequency ranging between 1 MHz and 160 MHz, which can be used to drive both the light emission and detection units. The control chip is also capable to process off-chip synchronization (Sync) signal generated by an external laser source. The Sync signal denotes the time at which the light pulse is emitted by the VCSEL which is then delayed by an integrated variable delay generator to activate the SPAD at a specific delayed time window[121, 122]. A variable width generator controls the width of the gating pulse thus providing a user-tunable ON-time for the detector. To provide a suitable biasing voltage to the detector, a high-voltage driver and level shifter were integrated in the design. After the detection of a photon, the avalanche signal is processed by the SPAD front-end which is composed of an active and a dummy detector along with other peripheral components and circuitry. The hold-off time generator turns off the detector for a certain time after each detection and the reset pulse generator is invoked to restore the biasing voltage of the detector. In response to the avalanche pulse, a comparator generates a rectangular pulse which can be processed externally but is also fed to an integrated analog counter. The latter functions like a digital counter and delivers an analog voltage corresponding to the number of photon counts in the gated observation window. One advantage of using the analog counter is the simplification of the number of outputs for further interface. The VCSEL is a 100-μm top-emitting GaAs-based laser while the SPAD chip was fabricated in standard 0.35-μm CMOS technology. The control chip was fabricated in high-voltage 0.35-μm CMOS technology. The optical probe was integrated by housing two VCSEL chips, the CMOS driver chip and SPAD chip, in the same QFN package of dimensions 12 x 12mm. This leads to a miniaturization of size and allows for the hybrid integration of multiple chips.

5.3. Time-Gated Single Photon Detection module

5.3.1. Synchronization Block

A current-starved Voltage Controlled Oscillator (VCO), shown in Fig. 5.2, was implemented and used as a synchronization unit. MOSFETs M4 and M7 act as current sources which supply current to the inverter composed by M5 and M6. The drain currents of transistors M1, M2 and M3 are the same and are set by the input control voltage Vin. The currents in these MOSFETs are mirrored in each inverter/current source stage. The input resistance of the VCO is practically infinite and the input capacitance is small compared to the capacitances present in the loop filter. Its operation is similar to that of the ring oscillator. The VCO provides the synchronization signal for the Light Emission and Light Detection units (Fig. 5.1). The output frequency of the VCO can be tuned between 1 MHz and 160 MHz.

5.3.2. Time Window Selection for Photon Detection

The starting point of the Time Window is selected with reference to the input synchronization pulse. The synchronization pulse denotes the time instant when the light pulse is emitted by the module. This SYNC pulse, either generated by the internal signal generator or taken off-chip, is then delayed turning on the SPAD at some specific time window. The amount of delay is user tunable and is determined by 6-bit digital input applied to a variable delay block. The designed delay block, shown in Fig. 5.3, comprises a series of multiplexers with delay buffer connected to one of its inputs. The variable delay block has an intrinsic propagation delay of 1 nsec and step delay of 90 psec. The delay can be configured by changing the digital inputs according to (5.1) :

$$Delay \ = \ 1 \ nsec+(90 \ psec * digital \ value \ input) \tag{5.1}$$

where, 1 nsec and 90 pec are the intrinsic propagation delay and step delay respectively.
The parameters of the proposed delay block are elaborated in Table 5.1. In order to select the width of the delayed gating window, a pulse width modulator was designed, thus providing a suitable ON-time to the detector. The output of the delay generator signal is fed into the clock input of a D-FlipFlop (D-FF) and through a second variable delay block to its reset input[123]. The pulse width is made user-tunable by choosing the digital inputs of the delay block. The

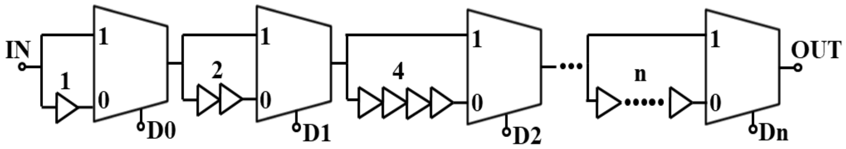


Figure 5.3. Designed Variable Delay Block

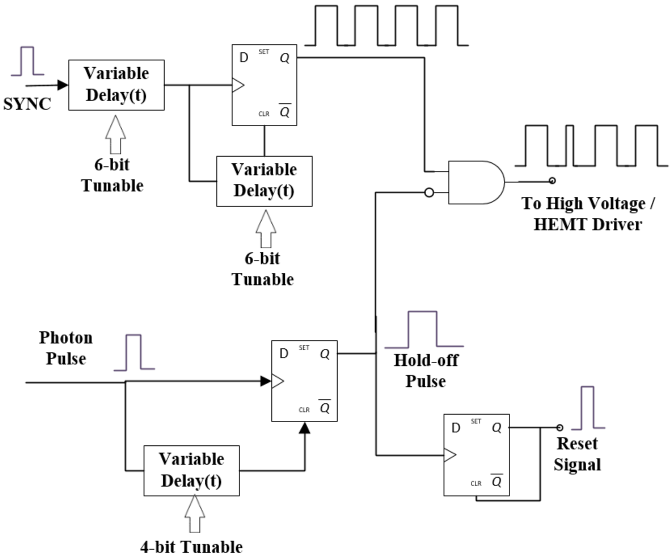


Figure 5.4. Proposed Hold-off Time and Reset Pulse Generator

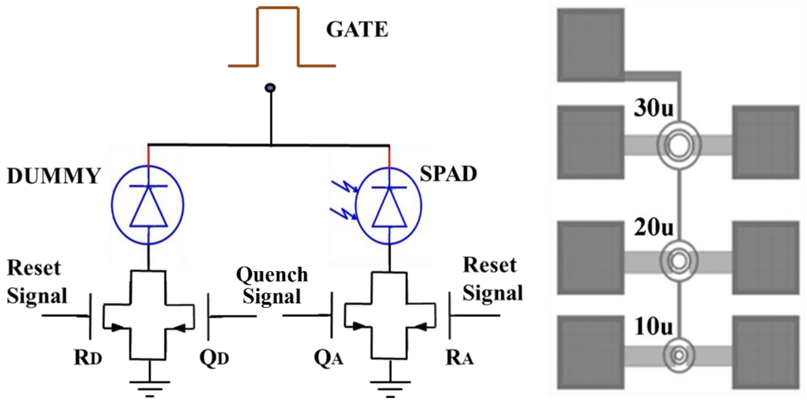
width of the time-window can be chosen from 2.5-35 nsec determining the time frame during which the single-photon detector remains active and ready for photon detection.

5.3.3. Variable Hold-off and Reset Pulse Generator

After the detection of a photon, a hold-off pulse is generated to turn-off the detector for a certain time by cutting off the gating pulse being applied to the detector. The hold-off time is

Table 5.1. Parameters of Delay Block Architectures

Parameters	**Digital Inputs (DI)**	**Step Delay (psec)**	**Maximum Delay (nsec)**
Variable Delay	6	29	22
ON-time	6	80	35
Hold-Off	4	1000	50



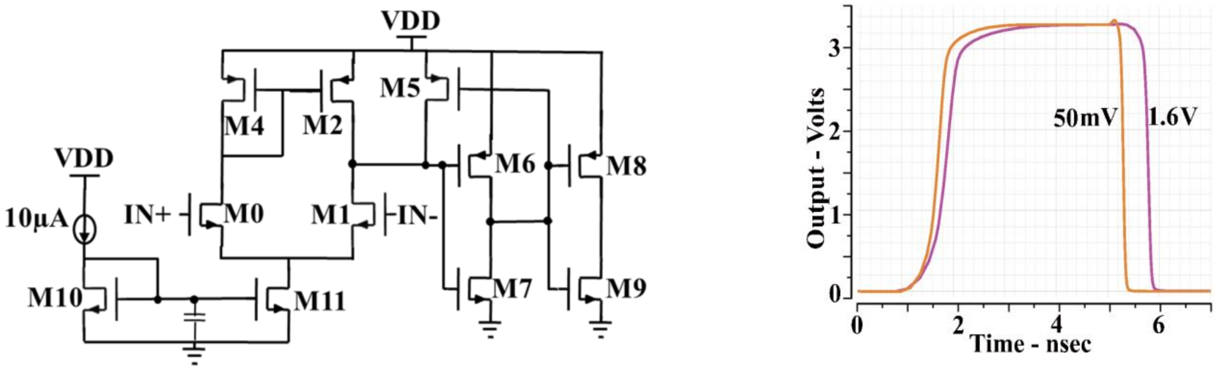
(a) (b)

Figure 5.5. SPAD frontend electronics (a) Simplified block diagram, (b) SPAD structures [95]

also user tunable from 1 to 50 nsec. The generated photon pulse is fed into the clock input of the D-FF and through a variable delay block to its reset input. By varying the digital inputs of the delay block, the user can control the width of the hold-off pulse during which the gate pulses are cut-off and thus disable the detector during the entire hold-off period. A Reset pulse, of a few nanoseconds, is also generated at the same time which goes to the SPAD front-end to restore its biasing conditions. The designed hold-off time and reset pulse generator are shown in Fig. 5.4.

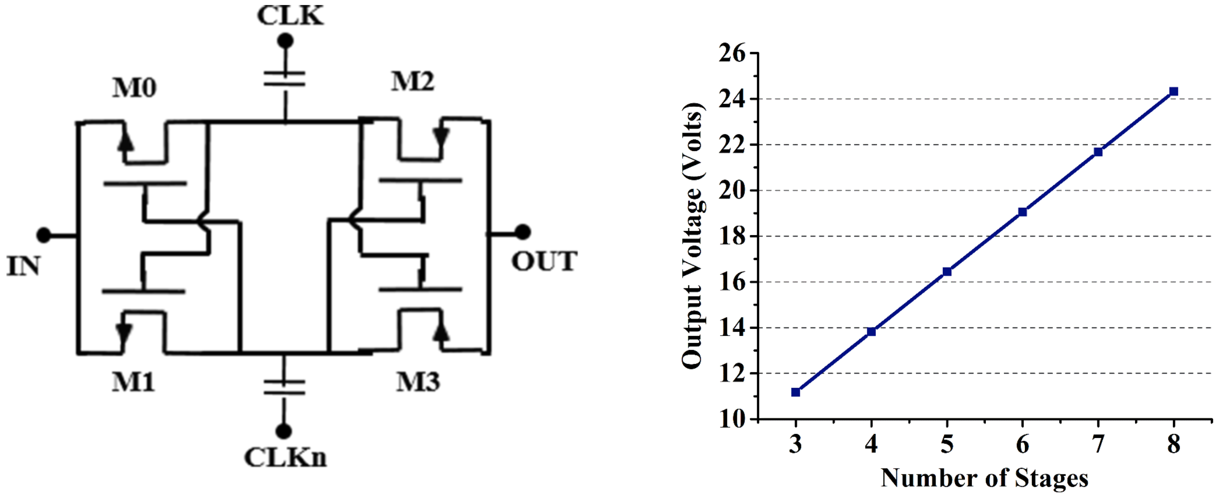
5.3.4. SPAD Front-end Electronics

Figure 5.5 shows the simplified block diagram of the SPAD frontend electronics [46], which comprises an active SPAD and a dummy one. The active SPAD is biased just below the breakdown voltage and is only sensitive for photon detection, or switched-on when gating pulses are applied to its cathode. The dummy is always biased below breakdown voltage. The differential topology is adopted to cancel out strong spurious noises appearing at the anode terminals of the SPADs. The spurious voltage spikes appear as a common- mode signal at the comparator inputs and is discriminated from the avalanche signal. As a result, in response to an avalanche signal at one of its input, the comparator produces a short pulse at its output which is then fed to a monostable circuit. The latter generates a rectangular pulse with a fixed time duration for each avalanche signal. The SPAD is operated in the so-called passive-quenching active-reset regime. MOSFETs Q_A and Q_D are used as quenching resistors, having value in the order of 100kΩ. On the contrary, MOSFETs R_A and R_D are involved in restoring the bias voltage at the anodes of the SPADs, thus acting as reset transistors. During



(a) (b)

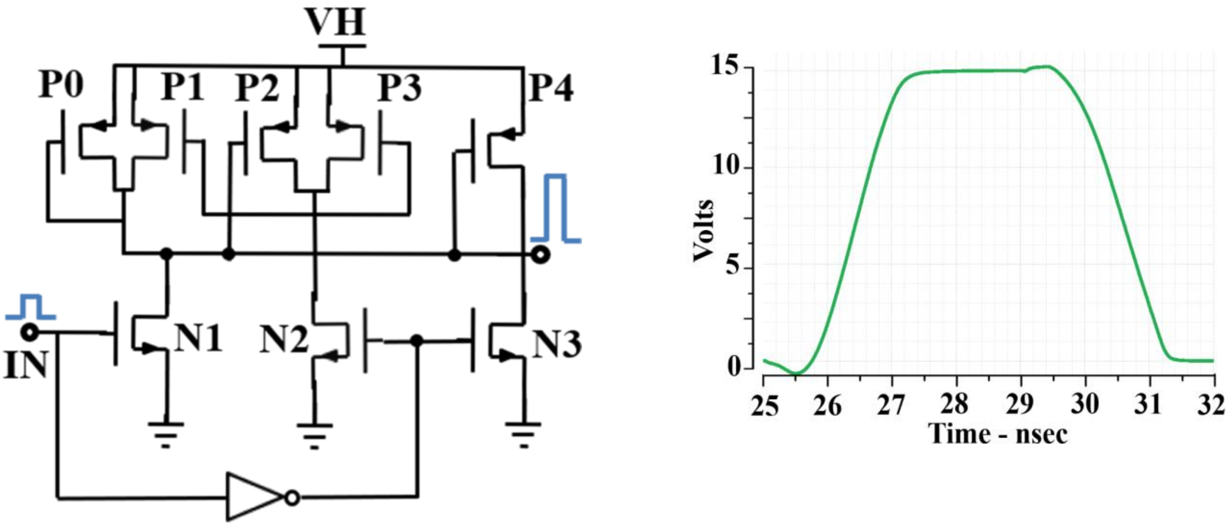
Figure 5.6. Integrated Comparator (a) Simplified Schematics, (b) Simulated comparator output with different threshold voltages



(a) (b)

Figure 5.7. Two-phase charge pump (a) Schematics of single stage, (b) Relation between the derived output voltage and the number of stages

the reset phase, R_A and R_D are turned ON for a few nanoseconds. Figure 5.6(a) shows the schematic of the integrated comparator whose input stage is a differential pair which is biased with a current of around 100µA[124]. The avalanche signal is connected to the non-inverting input of the comparator, while the inverting input is connected to the dummy detector which defines the threshold. In response to an avalanche pulse, the output of the differential pair goes up, close to the positive voltage rail. The output of the differential pair is fed to the inverter composed of transistors M6 and M7, whose output goes to low logic level, turning-on the transistor M5. The latter provides hysteresis by drawing current from the differential pair and holding up its output. The gate length of transistor M5 is kept high to have a high on-resistance.



(a) (b)

Figure 5.8. High-voltage level shifter to provide a suitable biasing voltage to the SPAD (a) Schematics of the integrated level shifter, (b) Ouput voltage pulse of the high-voltage level shifter

Because of the high on-resistance only a fraction of the bias current of the differential pair flows through M5. After few nanoseconds the output of the differential pair returns to the initial condition. The first inverter drives a second one made of M8 and M9, whose output swings to the positive rail. When the input at the non-inverting terminal goes down, the output of the differential pair also goes down. At the same time, the output of the first inverter goes high, thus opening transistor M5, and the output of the comparator is pulled down. The width of the output pulse is proportional to the amplitude of the input signal. A buffer is used at the output to fix the width of the photon pulses to 4 nsec. An external comparator was also used to separately analyze the performance. The external comparator is an ultrafast SiGe ECL comparator (Analog Devices Inc. ADCMP572) with 8 GHz equivalent input rise-time bandwidth, and 40 psec output rise/fall times. The small area SPADs (10µm, 20µm and 30µm) were fabricated in Fraunhofer Research Institute[76]. These SPADs were suitable for fast-gating operation and have extremely low dark count rate. The fabricated devices are appropriate for low light level imaging and for applications that require picoseconds timing resolution and detection of fast signals. The photon detection efficiency reaches its peak around 420 nm with a maximum value of almost 50% for the 10 µm diameter SPAD, and above 50% for the larger area SPADs, decreasing its value with longer wavelengths. The timing responses are characterized by the absence of slow tails with best timing performances of 77 psec for the

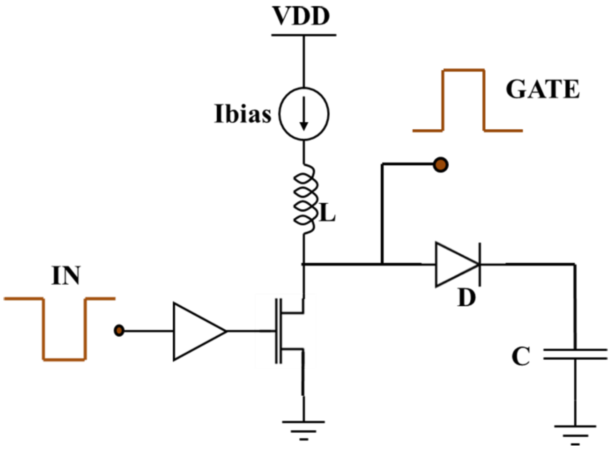


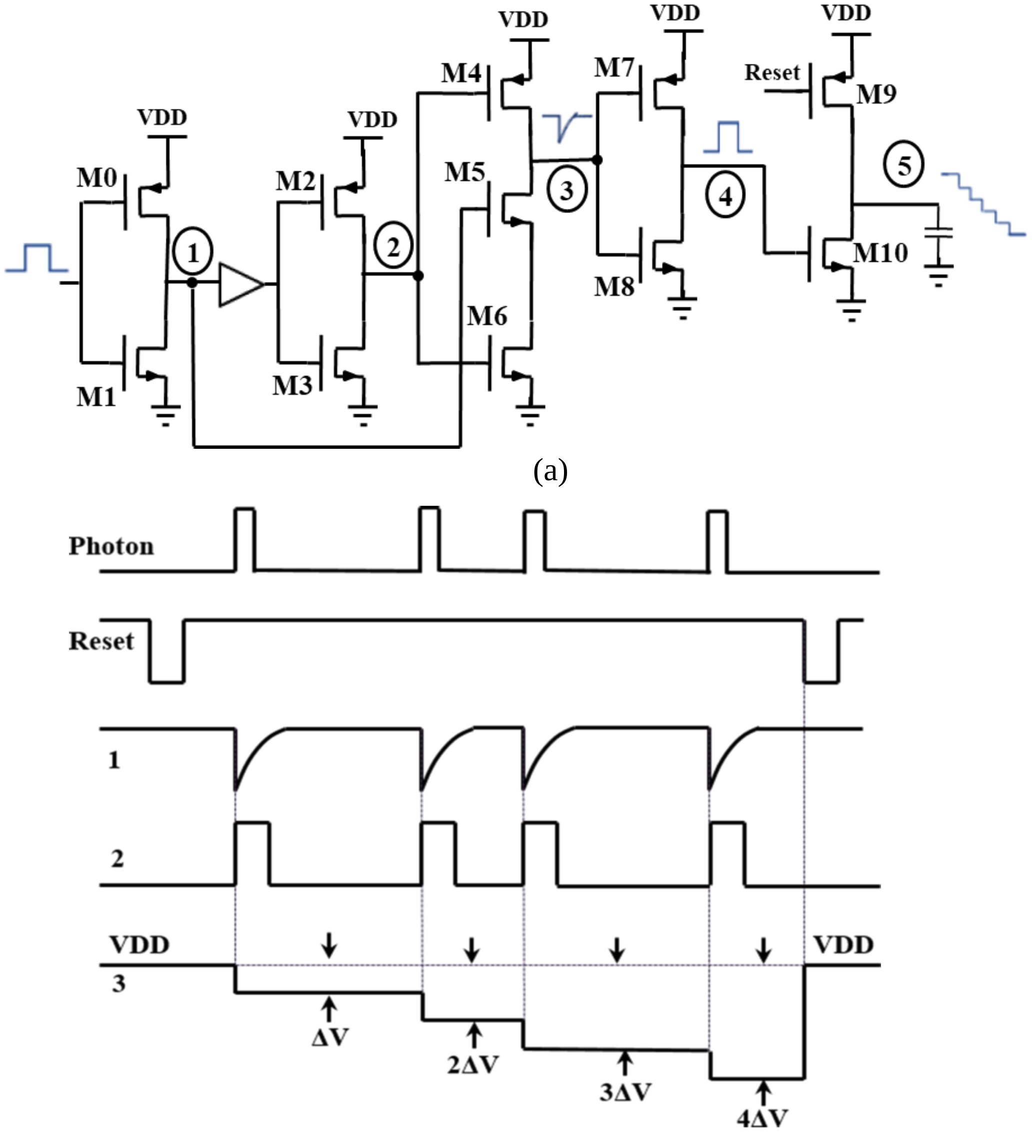
Figure 5.9. High Voltage pulser to have sub-nanoseconds transition times

10 µm diameter SPAD and around 120 psec for the other structures. The very low dark counting rate and the absence of the diffusion tail is extremely advantageous for the detection of optical signals having very fast decays (less than 100 psec) or very faint peaks following the main one. In other words, the SPADs work well with applications requiring very high dynamic range.

5.3.5. High Electron Mobility Transistor based Driver

Figure 5.7(a) shows the schematic of a single stage of a Pelliconi charge pump [125, 126], which is used to boost the standard low input supply voltage (3.3 V) to produce +6V to 15V output voltage. The functionality of the charge pump is based on a simple clocking technique requiring a 2-phase non- overlapping clock generator. The rise time (10% to 90%) from 0 to 8 V is 500 ns for the five stage charge pump. Figure 5.7(b) shows the relation between the derived output voltage and the number of interconnected stages. The output voltage ripple is observed to be around 250 mV with a 10 pF capacitive load and could be further minimized with an additional capacitive load at the expense of a longer rise time. Figure 5.8(a) illustrates the schematic of a high-voltage level shifter [126, 127] whose supply voltage (VH) is derived from the integrated charge pump. The circuit can produce 0/20V (maximum) transitions at its output from 0/3.3V on its input and can be divided into two parts – a level - up shifter circuit composed of transistors P0, P1, P2, P3, N1 and N2 and an output buffer composed of P4 and N3. The voltage pulse at the output of the level shifter is shown in Fig. 5.8(b) and the transition times is estimated to be around 1 nsec., which is its main limitation.

The high-voltage pulse generator is one of the most critical part of the system. The amplitude of the high-voltage pulse must be tunable over a wide voltage range (from 2.5 to 12 V) and its



(a)

(b)

Figure 5.10. Proposed Analog Counter (a) Simplified Schematics, (b) Timing Diagram showing the voltage levels at the various nodes

20% - 80% transition times should be minimized (sub nanosec.). The Photon Detection Efficiency (PDE) of the SPAD is directly proportional to the excess bias voltage and to achieve a uniform SPAD response for the whole detector-ON time, the amplitude of the high-voltage pulse must be flat without any ripples or oscillations. To meet all these requirements, one solution was the use of an off-chip High Electron Mobility Transistor (HEMT) with 6 GHz cut-off frequency. Figure 5.9 shows the high-voltage pulser circuit implemented with HEMT transistor acting as a common-source switching stage. The pulser converts the 3.3 V unipolar input pulse into a unipolar high voltage pulse with the required specifications. A HEMT driver

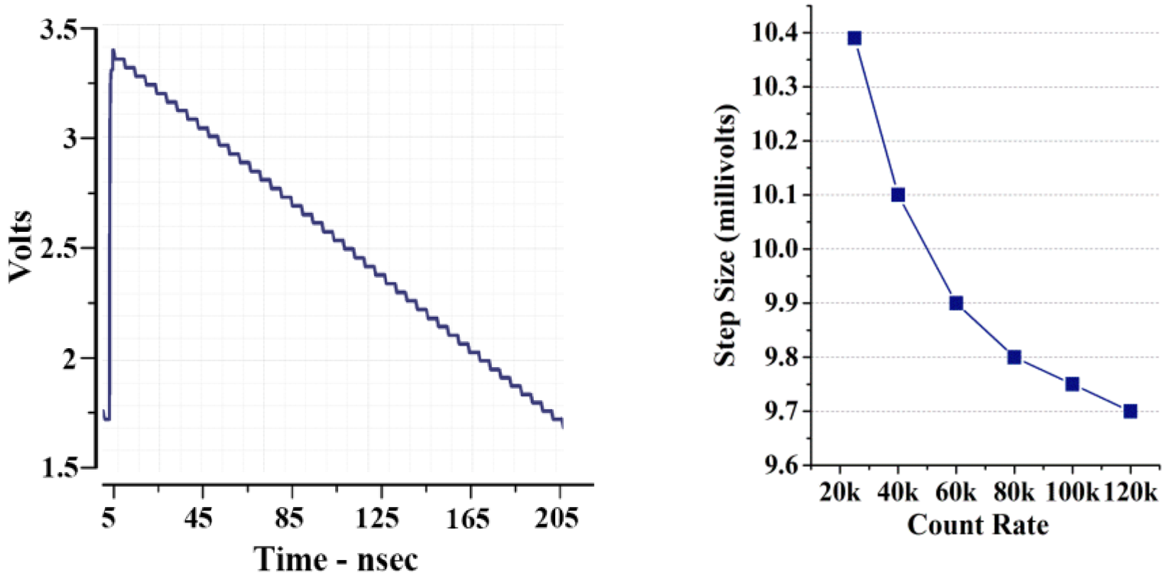
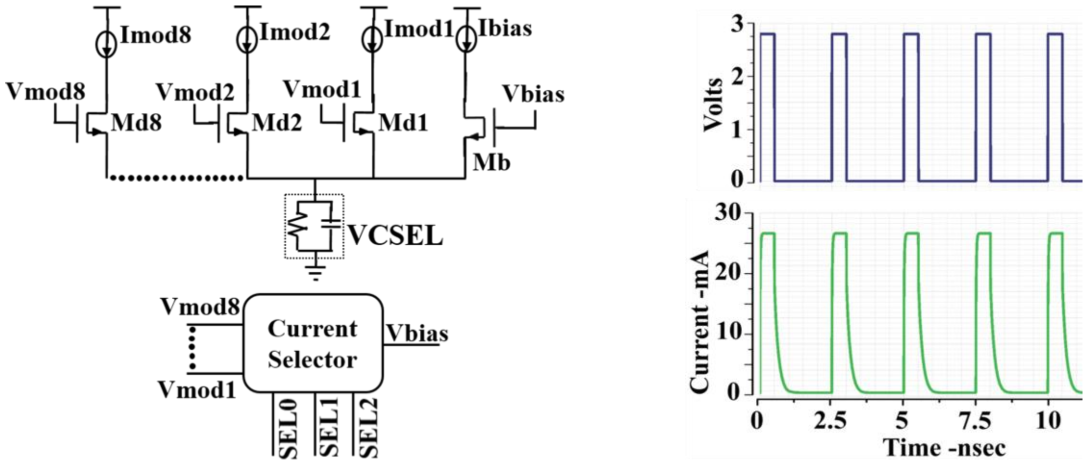


Figure 5.11. Simulations results of the analog counter (a) Transient simulations of the output voltage, (b) Output voltage step size vs the Count Rate

is integrated in the circuit, achieving low jitter and extremely low transition times.

5.3.6. Analog Counter

The function of the analog counter[92] is to deliver a voltage proportional to the number of photon counts in an observation window. The schematic of the proposed analog counter is shown in Fig. 5.10(a), depicting the signal at the most relevant nodes. The photon detection is indicated by a short rectangular pulse which is negated by the inverter composed of transistors M0 and M1. The output of this inverter is applied to M5. At node 1 the delay buffer provides a small delay to the negated avalanche pulse, thus a slightly delayed pulse is created at node 2. The NAND gate, composed of the transistors M5 and M6, results in the generation of a short negative pulse at node 3. The width of the negative pulse, determined by the delay provided by the buffer at node 1, is of the order of few hundred picoseconds. This negative pulse generates a short rectangular pulse at node 4, which turns on the transistor M10 and allows for the extraction of a small charge packet from the output capacitor. As a result, a small voltage step ΔV appears at the analog counter output. This voltage step depends upon the charge extracted from the output capacitor, which can be adjusted by varying the amount of delay of the buffer connected at node 1. The proposed circuit can count photon arrival events only when the Reset signal is in high state. When the Reset signal is set to low state, the transistor M9 is turned on, which recharges the output capacitor to VDD and a new count cycle is initiated.



(a) (b)

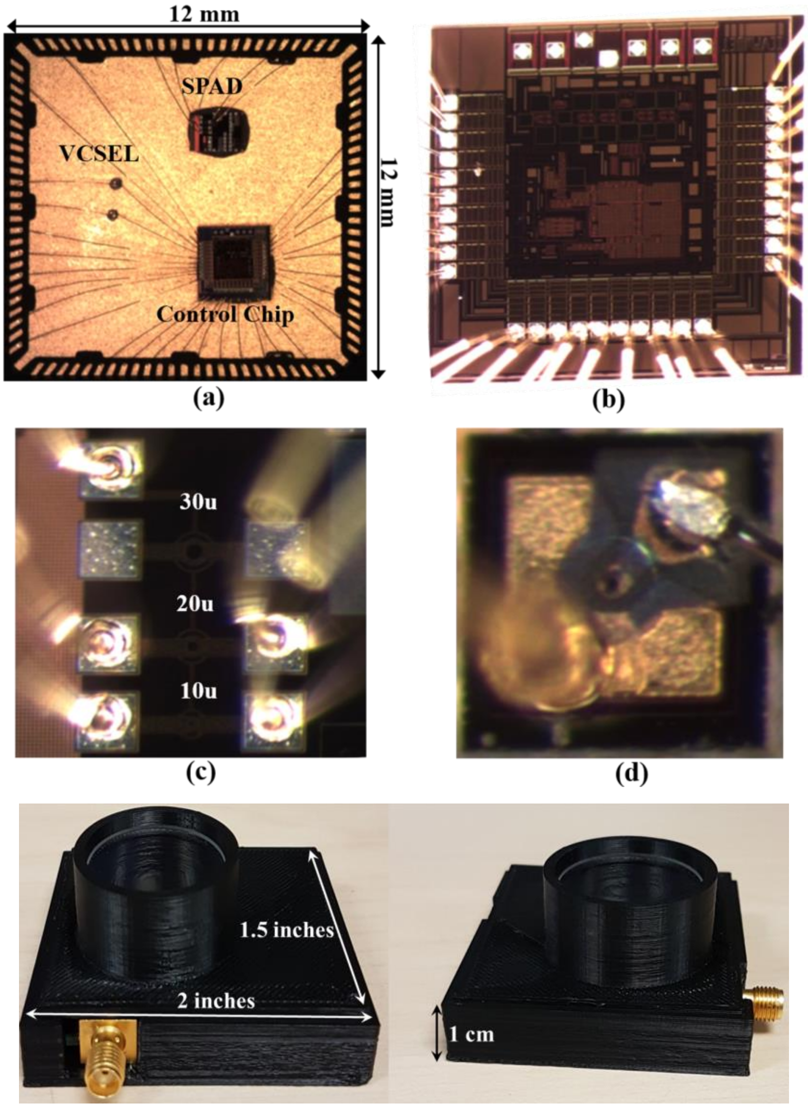
Figure 5.12. Proposed VCSEL driver Circuit (a) Simplified Schematics, (b) Graph showing the simulations of the VCSEL driver circuit, top - CMOS-level input voltage pulses, bottom - the current mode signals

The Reset cycle is user-tunable and can be varied from 100 nsec to 10 msec. The voltage level at significant nodes of the proposed analog counter is shown in Fig. 5.10(b). In order to better understand and adjust the circuit characteristics, a set of simulations, such as transient variations were performed to analyze the dynamic operation of the circuit. The pulse width at node 4 of the proposed analog counter was set to 300 psec. In order to have a wide dynamic range for photon counting, the voltage step size should be small but large enough to be easily discriminated from the electronic background noise. Figure 5.11(a) displays the output voltage of the analog counter. The step size is set to 10mV and a maximum count rate of 190 Mcounts/sec is achieved using the proposed circuit. The channel length of the transistor M10 has been carefully designed, thus achieving a compromize between the voltage step size and linearity of the whole discharge ramp. The linearity of the step size at different count rates is shown in Fig. 5.11(b). The output is read out via a unity gain ampifier which acts as an isolation and prevents the parasitic capacitance from loading the output of the analog counter.

5.4. Pulsed Light Emission Unit

5.4.1. VCSEL Driver

The VCSEL is a 100-μm top-emitting GaAs-based laser having an emission wavelength of 850nm, with very high-modulation rate, sharp optical rise and fall times. Figure 5.12(a) shows



(e)

Figure 5.13. Proposed System-In-Package (SiP) Optical probe (a) QFN package of dimensions 12 x 12 mm2, (b) Control Chip, (c) SPAD structures of 10u, 20u and 30u active areas, (d) VCSEL chip, (e) Overall module, housed in a 2 x 1.5-inches enclosure

the architecture of the VCSEL driver [113, 114], which provides digital control of the pulse duration, modulation and bias currents. The proposed programmable current-mode driver sources current directly into the anode pad of a VCSEL and ensures that it emits constant power over a wide range of conditions. The current selector is composed of a Digital to analog converter (DAC) and a high- speed buffer. The transistor Mb is designed to supply bias current to the VCSEL. The NMOS source transistor Mb supplies bias current to the VCSEL upon the application of tuning voltage, Vbias. Other source transistors Md1– Md8 are used to modulate the current flowing through the VCSEL. Md1-Md8 have their drain terminals electrically

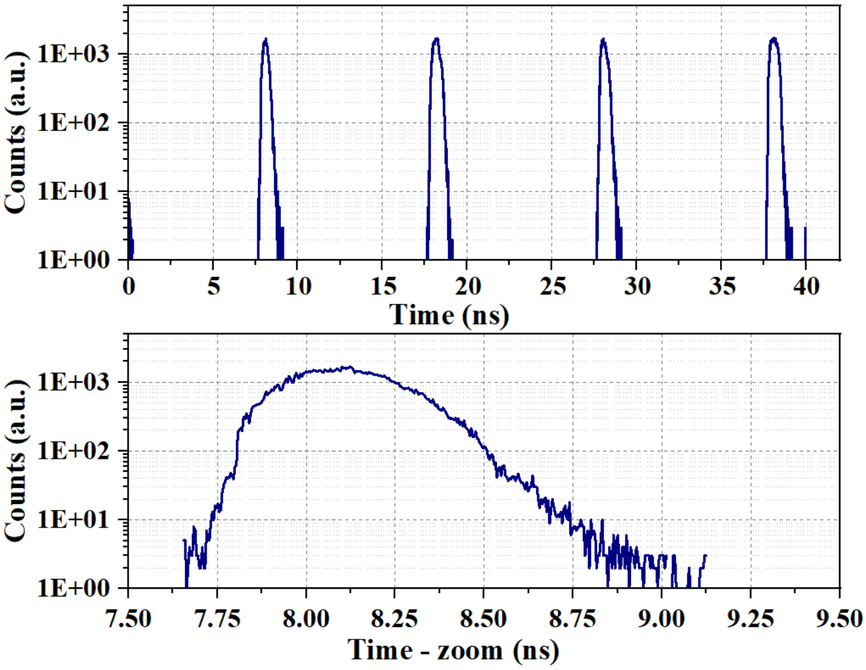


Figure 5.14. The measured optical waveforms. The driving input voltage amplitude varies between 1.0 and 3.0 V.

connected to a current source, their source terminals are shorted together and connected to the anode of the VCSEL. As demonstrated in Fig. 5.12(b), applying a sharp voltage pulse to their gate modulates the current which results in the modulation of the optical output. The driver was properly designed to supply bias and modulation current in the range of 0 to 12 mA and is optimized for driving wide variety of commercially available VCSEL devices, having variable resistance (ranging from 10 to 250 ohms) and significant capacitance (typically upto 2 pF).

5.4.2. Monostable Multivibrator

A high-speed operation is ensured by a buffer, generating electrical pulses with width varying from 500 psec to 2 nsec and transition times less than 100 psec.

5.5. Experimental Results

The proposed optical probe, as shown in Fig. 5.13, was implemented in one package integrating multiple heterogeneous chips – GaAs-based VCSEL chip, SPAD chip fabricated in CMOS 0.35μm technology and Control chip fabricated in CMOS 0.35μm High-Voltage technology. Experimental characterization was carried out for the various building blocks to analyze the performance of the module. The operation of the pulsed light emission unit of the module was

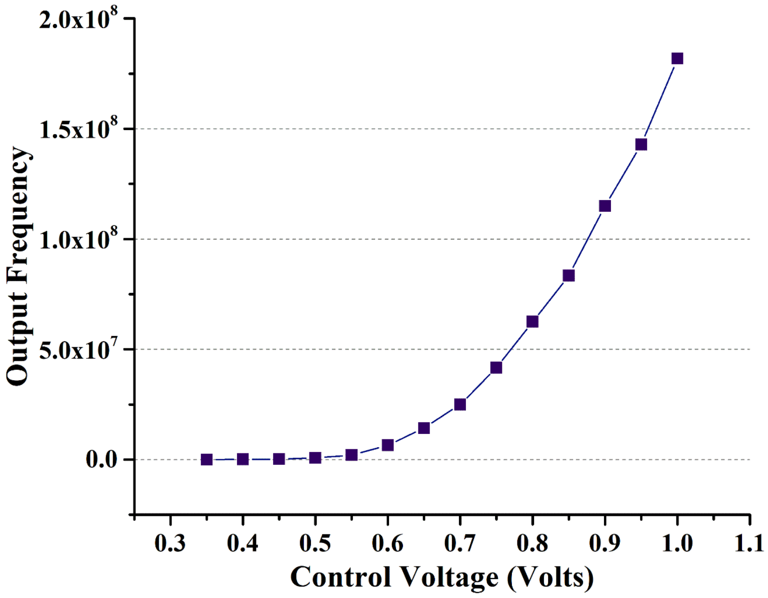
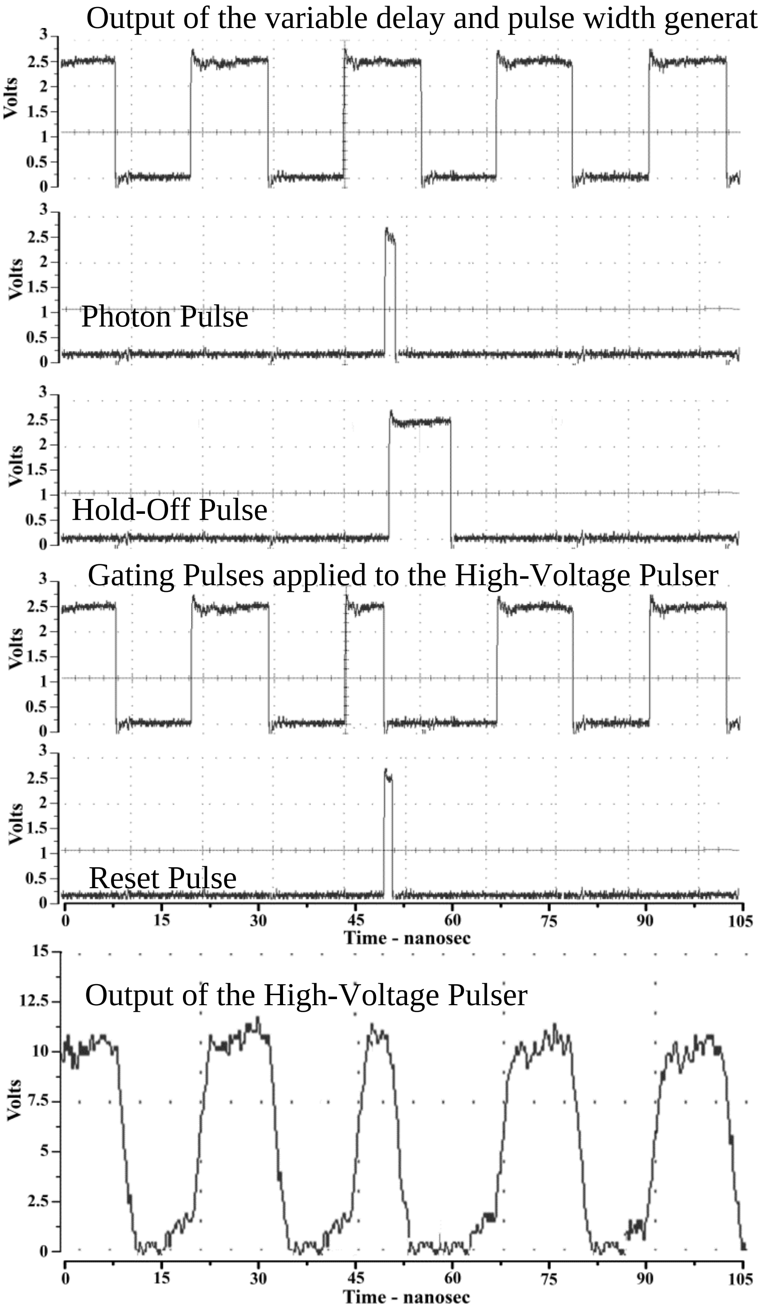


Figure 5.15. Output frequency as a function of the control voltage

characterized by applying a suitable bias voltage (Vbias) and varying the modulation voltage (Vmod) which results in the regulation of current flow to the VCSEL. The VCSEL (VI integrated systems, 850nm) has contacts on the top-surface and is connected to the driving circuit using wire bonds. The gate modulation voltage varied from 1.0 to 3.0 V with a pulse width of less than 1 nsec. The resulting modulated light was focused on a fast silicon single-photon detector, whose output was coupled to a Time Corelated Single Photon Counting (TCSPC) module (Picoquant). The resulting detected signal from the input-modulated VCSEL is shown in Fig. 5.14 with a repetition rate of 100 MHz and full width at half maximum (FWHM) of 350 psec. The integrated synchronization block was capable of generating gating pulses of frequency ranging from 1 to 160 MHz. The module is also capable of processing an off-chip trigger signal fed to the “SYNC” input, in case an external laser source is used for experimentation. While turning off the detector gating signal the module will operate in free-running mode, thus is also able to detect photons at any time. Figure 5.15 shows the plot of the output frequency of the synchronization block as a function of applied control voltage. At lower control voltages, the output frequency is not linear, but above 25 MHz the output frequency starts varying more linearly with the control voltage.

The synchronization signal, either generated inside or taken off-chip, goes to the variable delay block followed by a variable pulse width generator which modulates the width of this pulse to


Output of the variable delay and pulse width generator

Photon Pulse

Hold-Off Pulse

Gating Pulses applied to the High-Voltage Pulser

Reset Pulse

Output of the High-Voltage Pulser

Figure 5.16. Voltage waveforms during the module operation

provide a variable ON time to the detector. These two blocks, variable delay unit and variable pulse width generator, are utilized to turn-on the single photon detector at specific time windows. The step delay of the variable delay block is 90 psec and can provide a maximum delay of 22 nsec to the Sync signal, thus controlling the starting point of activation of the detector. The width of the Sync signal can be modulated in the range of 2.5 to 35 nsec, which determines the time-period for which the SPAD is biased above the breakdown voltage and is ready for photon detection. The output of the variable delay and pulse width generator is applied

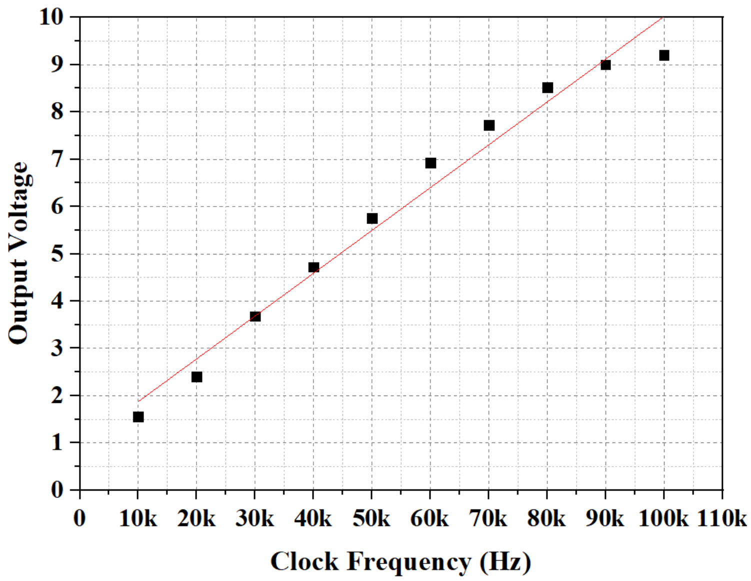


Figure 5.17. Output voltage of the 2-phase charge pump vs clock frequency

to the High-Voltage pulser which acts as level-up shifter, providing a suitable excess bias voltage to the SPAD. The amplitude of the high-voltage pulse can be tuned from 2.5 to 15V, as demonstrated in Fig. 5.16 where the amplitude was configured to be around 10V. As soon as the incoming photon is detected, the SPAD front-end generates a rectangular pulse,nwhich triggers the Hold-off and Reset time generator. In Fig. 5.16, the signal designated as photon pulse, marking the arrival of the photon, initiates the Hold-off pulse and the Reset Pulse. The hold-off pulse width determines the hold-off time and is configurable in the range of 1 to 50 nsec. In Fig. 5.16 the high-voltage pulse is turned off as soon as the hold-off time is initiated, thus deactivating the SPAD. The Reset pulse, of the duration of 2 nsec, is applied to the SPAD front-end to restore the biasing conditions. Measured signals corresponding to the variable delay, hold-off pulse, SPAD gating pulses and high voltage pulser are graphed confirming the functionality of the integrated control chip. To provide a suitable biasing voltage to the detector the 3.3V pulses are applied to the high-voltage pulser, where its output transition (rise and fall times) vary from 250 to 800 psec depending on its supply voltage with characteristics tuned to drive a single photon detector. The high-voltage pulser was tested with an external trigger signal at frequency of 80 MHz provided by an external laser source and also with the inbuilt synchronization unit. Waveforms were acquired using a 1GHz equivalent-bandwidth digital sampling scope (Tektronix 11801A). The active SPAD being used was a 10 µm active-area diameter CMOS SPAD, with a breakdown voltage of 26.5 V. The dummy detector was integrated on the same die and was always biased below the breakdown voltage. Due to

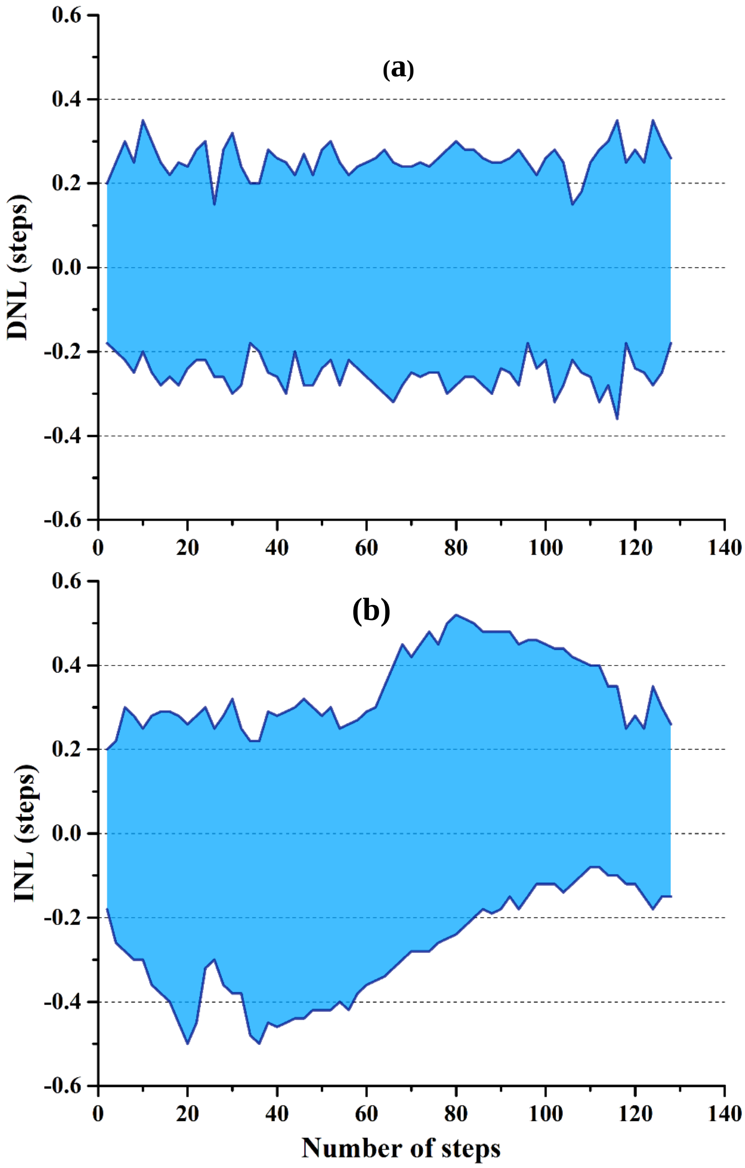


Figure 5.18. (a) Differential Non Linearity of the analog counter, (b) Integral Non Linearity of the analog counter

shallower depletion region the Peak Photon Detection efficiency (PDE) is above 50% at 450 nm and decreased to nearly 2 % at longer wavelengths (around 850 nm). The Dark Count Rate (DCR) was measured at different excess bias voltages. It was observed that the SPAD have extremely low dark count rates with a maximum of 60. Even at higher temperatures the dark count rate remained low (nearly 200) for the 10 μm active area SPAD. Due to these low dark count rates the SPAD could be used without any cooling.

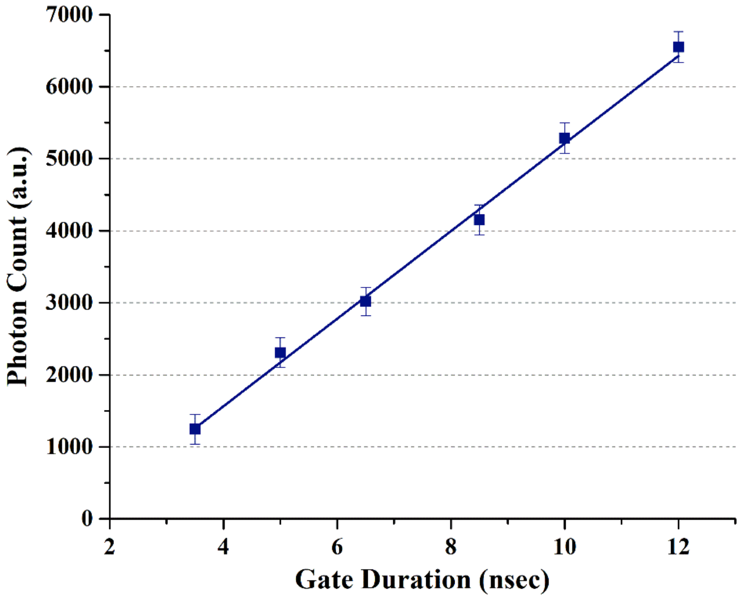


Figure 5.19. Photon Count vs Gate Duration

The biasing voltage for the SPAD can also be derived from the integrated charge pump. A set of measurements has been performed by changing the input voltage, varying the clock frequency and applying a variable current sink to the output. Measurement results are shown in Fig. 5.17 that features the relation between the clock frequency and the output voltage of the implemented 5-stage charge pump circuit. The average output voltage ripple was observed to be 250 mV but could be reduced if an additional capacitive load is added to the output. The maximum output current is 1 mA.

A characterization of the analog counter was carried out and the non-linearity was found to be less than 4%. The output voltage value of the analog counter was acquired for a wide range of input photon pulses spanning from 0 to 128. From this data, differential non-linearity (DNL) was calculated. The envelope of DNL is depicted in Fig. 5.18(a) with values of DNL are found to be within ±0.4 LSB , thus ensuring good uniformity and sensitivity. Figure 5.18(b) shows the INL envelopes whose value doesn’t exceed 1 LSB.

Figure 5.19 shows the dependency of the photon count rate on the gate duration. The count rate can be expressed as follows:

$$R_T = \frac{Nt}{f\,tg\,t} \tag{5.2}$$

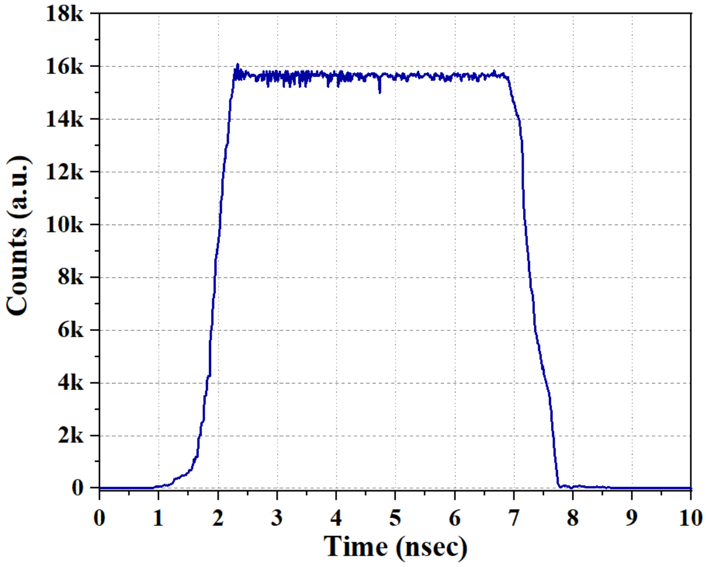


Figure 5.20. Photon counts distribution within a 5 ns gate-ON time window, when the SPAD is gated at 40 MHz with an excess bias voltage of 5 V.

where, N_t is the photon count rate, t_g is the gate duration, f is the gating frequency and t is the integration time. It is understood from equation (5.2) that photon count rate is linearly dependent on the gate-ON time. We performed a characterization of the time-gated detection module by illuminating the detector with attenuated ambient light and acquiring the temporal distribution of the incoming photons. The photon detection efficiency of the SPAD is proportional to the applied excess bias voltage making it necessary for the gating pulse to be flat and without any ringing which can be directly observed in the count distribution. In order to activate the SPAD in very short period of time and perform an efficient rejection of unwanted photons, the time-gating technique relies on the fast transition times of the gate window. Figure 5.20 shows the measured photon count distribution during a gate-ON time window of 5 nsec at an excess bias voltage of 5V. As observed using the 10µm active diameter silicon SPAD, the shape of the time-gating window is flat with clean and fast rise and fall transitions. The 20–80% rise time is found to be about 500 ps with a smooth edge shape without any relevant ringing or overshoot that would otherwise have an adverse effect on the reconstruction of an optical waveform. Almost no counts were recorded outside of the gate-ON time displaying a monotonic behavior and validating the response uniformity of the module within the ON-time window. The instrument response function (IRF) of the entire module is shown in Fig. 5.21 and the overall FWHM is found to be 350 ps.

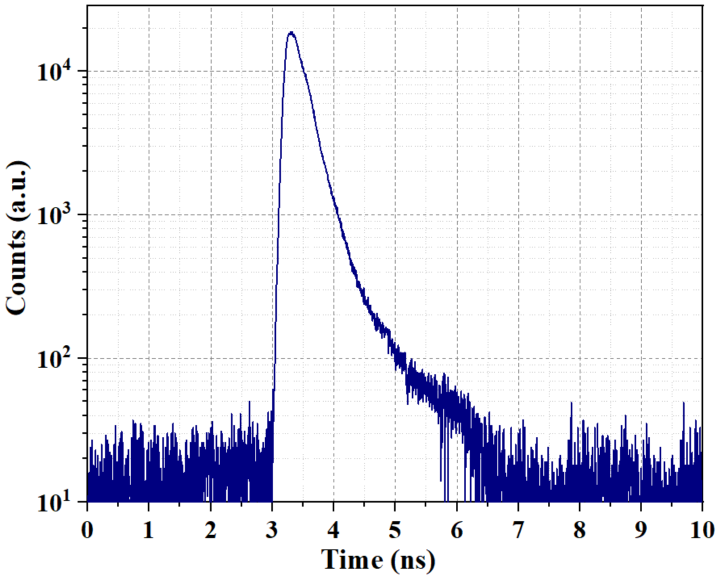


Figure 5.21. Time-response of the time-gated module to the integrated pulsed VCSEL source, with 350 ps pulse-width

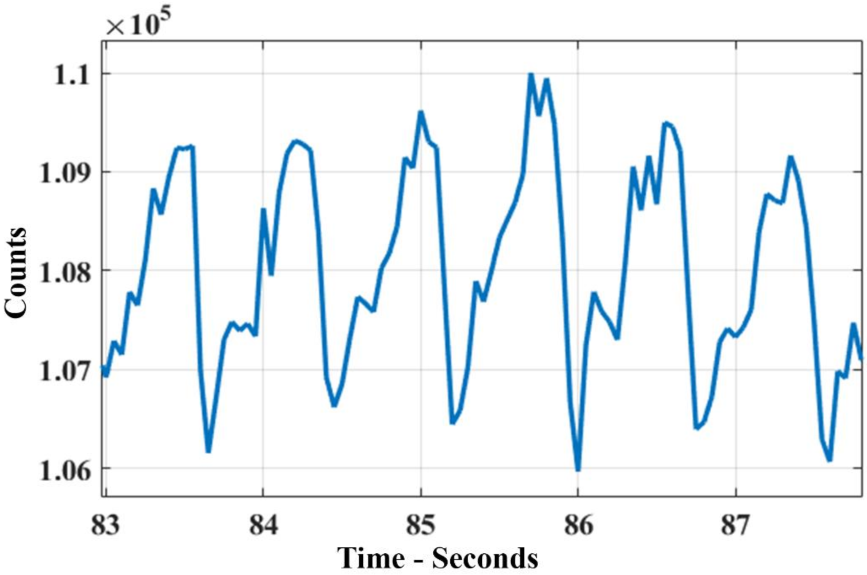


Figure 5.22. Count rate plot during heart-beat measurements by illuminating the index finger of the volunteer

In order to assess the performance of the module in a real-life application, we devised a simple setup to measure the heart rate of a healthy volunteer. A laser source was used to illuminate the index finger of the volunteer and the detector was placed on the other side of the finger. The count rate was acquired over 60 consecutive 50 ms time intervals. The data was collected by a PC via a USB link. A user interface was implemented in C++ to acquire the data. Figure 5.22 shows the acquired curve in which the variation of the photon counts is predominantly happening due to the change in the total blood volume in finger blood vessels and exhibits a periodicity of approximately 80 beats per minute (bpm), i.e. the heartbeat of a volunteer

Table 5.2. Comparison of the basic features of the various state-of-the-art NIRS systems

	Detector	**Light Source**	**CMOS Technology**	**Application**
This work	TG-SPAD	VCSEL	0.35µm HV	TD-NIRS
[108]	SPAD	µLEDs	0.18µm	TD-NIRS
[102]	TG-SPAD	No	0.35µm HV	LIBS
[106]	SiPM	VCSEL	Discrete	TD-NIRS
[109]	SiPM	VCSEL	Discrete	TD-NIRS
[110]	Si-PD	VCSEL	0.18µm	fNIRS

(measured independently). It is observed that even when the measured variation in the photon count rate is less than 1 % of the average detected signal, the pulses are clearly visible thus ensuring the feasibility of the usage of this module in various medical applications.

5.6. Conclusion

We presented the design of an integrated standalone near-infrared spectroscopy system which comprises a fast-gated single-photon counting and a pulsed laser source. All the operating parameters of the module (e.g., hold-off time, reset time, gate-on time) are completely user-tunable for an easy integration in numerous experimental setup. One limiting factor of placing the light source and detector in close proximity to each other is the massive short-time reflection that would otherwise result in the detector saturation. This problem was addressed by fast gating of a single photon avalanche diode to detect photons in some specified time-windows. The novel approach has paved the way to the development of on-chip null-distance source detector making it possible to devise a single optode significantly relaxing the need for control over positioning source and detectors while maintaining performance in Near-Infrared Spectroscopy. The cooling system is a limiting factor in the complete integration of a fast-gated single photon counting module on a single chip. Henceforth, a full characterization of all the building blocks of the module was done without any cooling system, testing both its electric and optical properties. A practical application of this device was then demonstrated by means of a setup that measures the heartbeat pulses, showcasing the prospective use of our module in various personal healthcare and medical equipment.

5.7. Acknowledgement

Authors acknowledge support from NSERC, CIHR, Canada Research Chairs on Smart Medical Devices and Optical Vascular Imaging, and CMC Microsystems for design and simulation tools

CHAPTER 6 MINIATURIZED SIPM-BASED NIRS PROBE INTEGRATED WITH PULSED LASER SOURCE

In this chapter we have presented the design of a miniaturized Silicon Photomultiplier (SiPM) based compact probe for Near Infrared Spectroscopy (NIRS), that can be directly put in contact with the sample under test to obtain high diffused photon harvesting efficiency without the need for cumbersome optical fibers and lenses for light injection and collection. The proposed system is composed of two parts namely, i) pulsed light emission unit and ii) single-photon detection module. The size of the optical probe was reduced to 10mm by integrating multiple heterogenous chips from various foundry on a silicon interposer platform ('2.5D' stacking).

6.1. Background

Nowadays, the widespread adoption of the TD technique is mainly hampered by the traditional bulky, expensive, complex and fragile components which significantly impact the overall cost and dimension of the system. Over the past few decades, significant research has been conducted to scale down the photonics components (e.g., picosecond pulsed lasers, time-resolved single-photon detectors) and opto-electronic setups but most of the commercially available state-of-the-art TD instruments [128, 129] are still large rack-based systems. The main barrier impeding the widespread clinical adoption is still represented by the use of expensive, bulky and fragile photomultiplier tubes (PMTs), which appears to be the main bottleneck for applications requiring a large number of detection points. The utmost need for low-cost, robust, portable and compact detectors has led to the recent development and growth in microelectronic devices

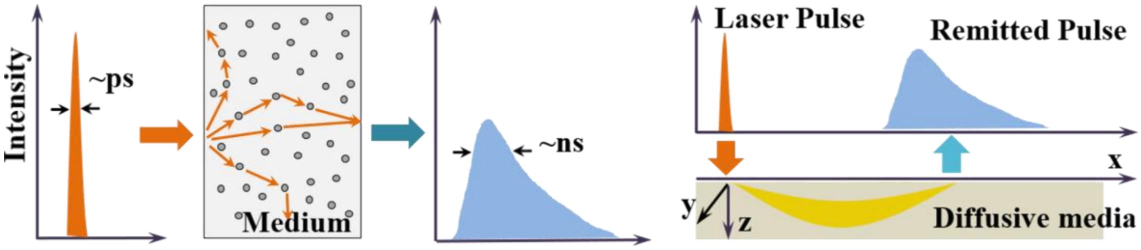


Figure 6.1. The geometry of TD NIRS measurements also depicts the region where photon paths are more likely to occur (so called "banana shape")

such as single-photon avalanche diodes (SPADs). SPADs can be easily integrated into complementary metal oxide semiconductor (CMOS) technology, exhibits good single-photon sensitivity, timing resolution and are not prone to be damaged by strong light pulses. However, their performance is greatly limited and often unsatisfactory owing to the small active area (few tens of micrometers in active diameter), which affects the overall collection efficiency of the diffused light. Recently, silicon photomultipliers (SiPMs), featuring wide active areas up to few square millimeters, have proved to be a rugged and inexpensive alternative. A SiPM consists of an array of hundreds (or thousands) of microcells connected in parallel with very high fill factor, resulting in a common global anode and a common global cathode connection. Each microcell includes a SPAD with its passive quenching resistor. The device is suitable for both single-photon and multi-photon application as the output is an analog voltage whose amplitude is proportional to the number of microcells which has been triggered by the impinging photon. The wide active area of the device maximizes the photon collection efficiency, thus permitting and enabling the detection of very low light levels with great efficiency. Also, characteristics such as high quantum efficiency (e.g., > 15% at 690 nm) and broad spectral coverage (350 - 1000 nm) make SiPM detectors ideal candidates for TD-NIRS systems. Other important aspects that make these detectors highly promising and versatile are that they are practically insensitive to electromagnetic fields and often don’t require any cooling systems.

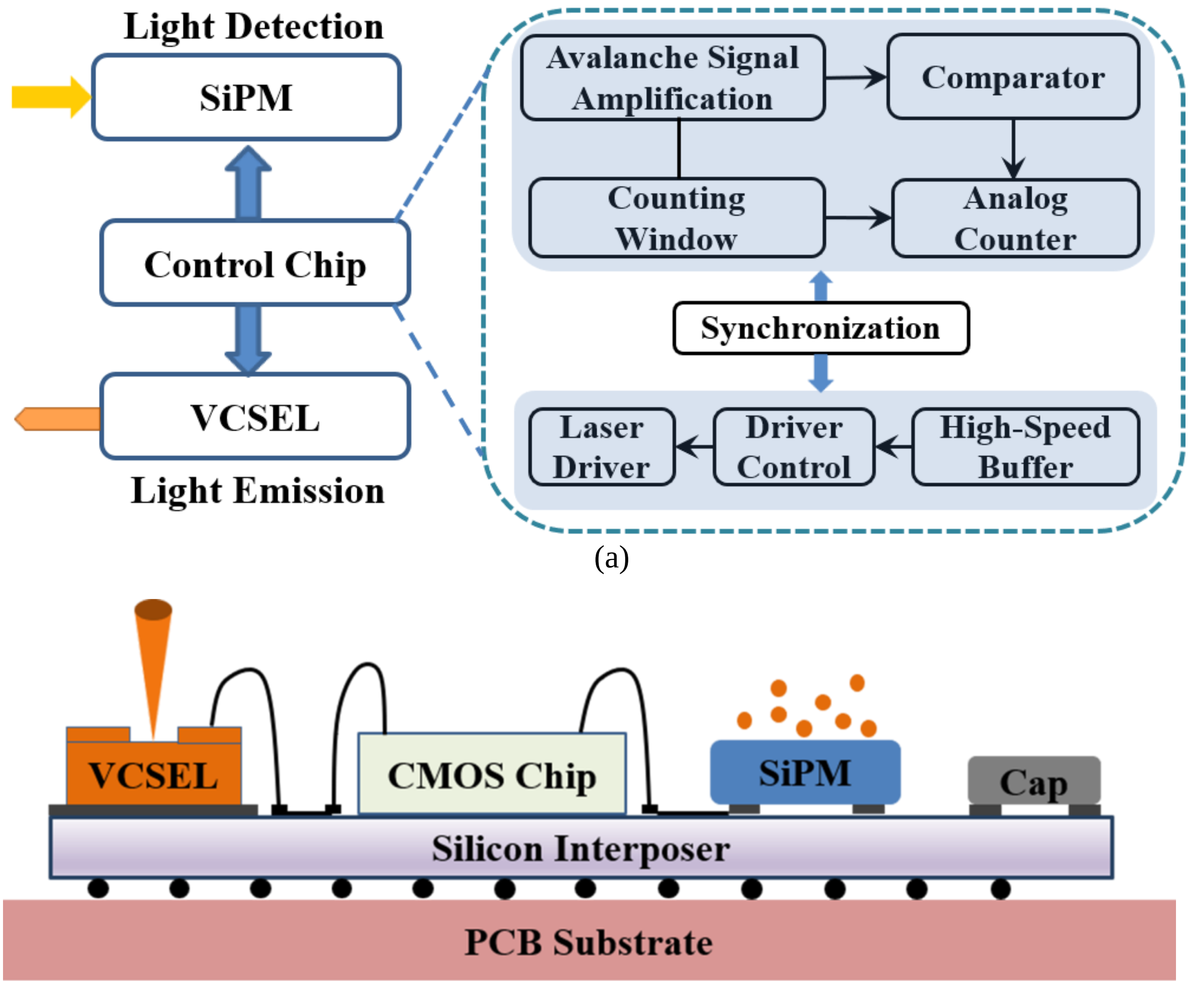
During the last few years, TD-diffuse optics (DO) systems [130-132] have started to benefit from the advantages brought by the adoption of SiPMs which hold the best features of PMTs and those of solid-state detectors. Various custom-made probes [119, 133] were designed based on commercially available SiPMs, with high throughput and high sensitivity performance, thus leading to the development of new class of instruments which features better characteristics in terms of dimension and complexity. To the best of our knowledge, only few complete TD-DO systems, based on SiPMs, are commercially available. Among the recently developed ones, tNIRS-1 [134] is a two channel time-resolved spectroscopy measurement system, developed using pulsed laser diodes at three wavelengths, SiPM for light detection and time-digital converter for time-of-flight measurements but it exhibits a broad Instrument Response Function (IRF) of 1.5 ns Full-Width at Half Maximum (FWHM). More recently, a complete instrument, including two custom-designed pulsed diode lasers (operating at 830 and 670 nm wavelength), a single-photon detection module (based on a 1 mm^2 active area SiPM packaged with a thermoelectric cooler), and a custom time measurement electronics with 10 ps time resolution, was developed for TD-NIRS [106] which displays an IRF narrower than 300 ps. Another complete detection chain [135], based on 8 commercially-available SiPMs, was proposed to improve the signal level in multi-wavelength time-domain optical mammography and could be easily adapted to various other DO applications. However, none of these systems are fully CMOS-integrated and used discrete electronics, which effects the scalability. All these systems still rely on optical fibers to inject and collect light from human tissues and the size remains too large to be put into small sockets for wearable applications such as brain monitoring.

Our goal is the development of a miniaturized SiPM based compact probe for TD-NIRS, that can be directly put in contact with the sample under test to obtain high diffused photon harvesting efficiency without the need for cumbersome optical fibers and lenses for light injection and collection. In this paper we present the design of a fiber-free probe, integrated with a source-detector pair, to be placed directly in contact with the sample. The light emission unit is embedded into the probe in the form of a pulsed vertical-cavity surface-emitting laser (VCSEL) and the light detection unit comprises a SiPM along with other CMOS integrated control circuitry. Compared to our previous work with time-gated SPAD [136], here the novelty is on the demonstration of the direct integration of a miniaturized pulsed laser source with the SiPM leading to the development of a unique compact optode based on null-distance approach which represents the first step towards state-of-art compact TD-NIRS diagnostic instruments for wearable real-time monitoring applications.

The remaining of this chapter is organized as follows: Section 6.2 includes the description of the proposed module. We analyze and describe in Section 6.3 the single photon detection unit and the pulsed light emission unit in Section 6.4. Section 6.5 reports the electrical and optical characterization and some of the target applications. Section 6.6 concludes the chapter and Section 6.7 briefly explains the general assessment

6.2. System Description

Figure 6.2 (a) illustrates the simplified block diagram of the designed compact optical probe. The proposed system is composed of two parts namely, i) pulsed light emission unit and ii) single-photon detection module. All the fundamental building blocks were integrated in a control chip which is fabricated in CMOS 0.35µm technology. The single-photon detection module is based on a 1 mm^2 SiPM (C30742-11), by Hamamatsu Photonics with a detection efficiency >10% at a wavelength of 690 nm. The SiPM is interfaced with an avalanche amplification stage to amplify the low amplitude of the avalanche signal triggered in response to the incoming photon. The comparator and multivibrator generate a rectangular pulse with a fixed width which can be processed externally but is also fed to an integrated counter. Gated photon counter has been implemented to count the impinging photon arriving during well-defined counting window synchronized with the excitation laser pulse, the duration of which can be programmed down to few nanoseconds. The pulsed light emission unit is composed of a laser driver along with other peripheral circuits that are designed in the control chip and directly wire-bonded to the VCSEL diode (VI integrated systems). The emission wavelength of the laser diodes is 850nm and can be pulsed at a repetition rate of over 100 MHz. An internal signal generator was designed to generate synchronization pulses of frequency ranging between 1 MHz and 160 MHz, which can be used to drive both the pulsed light emission and detection units. While miniaturized probe-hosted source [137] and also monolithic SPAD array with on-chip timing circuits have been already well demonstrated in the literature, this module boasts the integration of miniaturized pulsed light emitting source and detector array on a 2.5D silicon interposer platform (Fig 6.2(b)). To reduce the size of the optical probe to 10mm^2, we designed and fabricated a silicon interposer platform (“2.5D stacking) [138-140] with through silicon via (TSV) structure to integrate multiple chips from various foundry technologies. A 2.5D-IC system has two dies on silicon interposer including the control chip and the VCSEL die. It also consists of the SiPM device with a ceramic package along with other passive components such as capacitors and resistors. The use of an interposer provides reduced parasitic effects, and the possibility of adding closely spaced ancillary components such as decoupling capacitors. The interconnections between the dies and other components on the interposer are done using metal redistribution layers (RDL) with a minimum trace width of 5µm and 6 µm trace thickness, which enables high density routing and small spacing between dies and components, resulting in a compact and high-performance device. The TSV technique connects top side and bottom side of silicon interposer, while the connection to the substrate is done utilizing micro-bumps attached to the bottom side of the interposer.



(a)

(b)

Figure 6.2. Compact Optical Probe, (a) Simplified Block Diagram, (b) System level integration on a silicon interposer platform

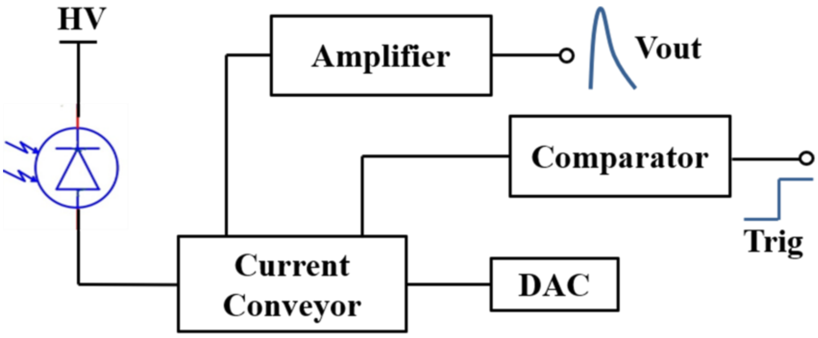


Figure 6.3. Scheme of the current mode approach where the SiPM current can be mirrored for both amplitude and timing measurement purpose

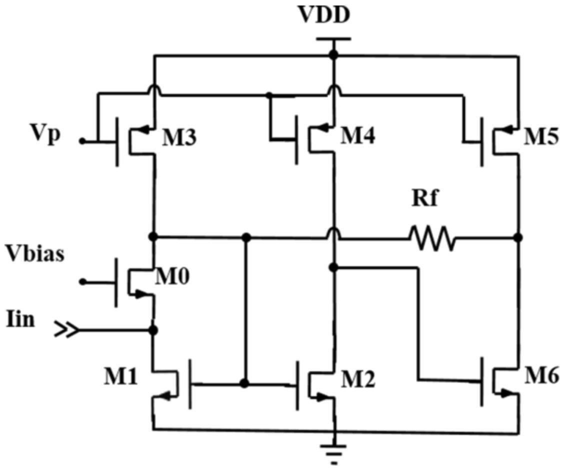


Figure 6.4. Simplified schematic of the Transimpedence Amplifier 1

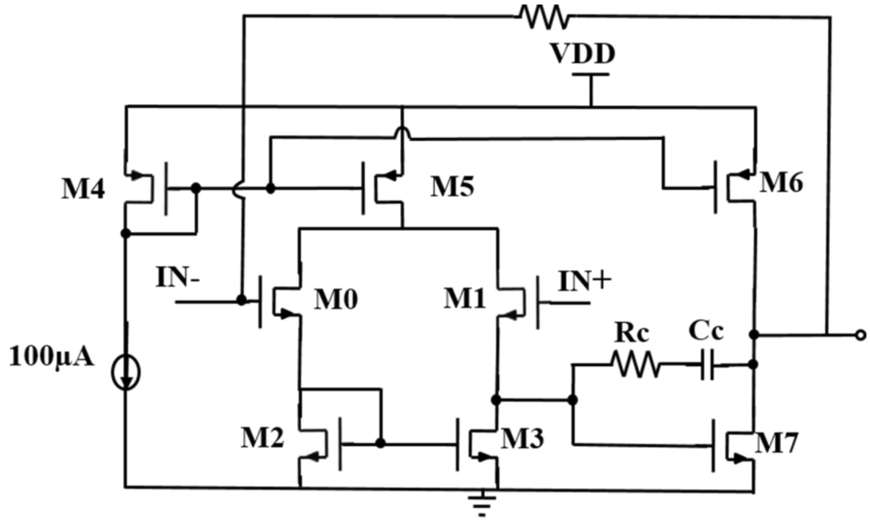


Figure 6.5. Simplified schematic of the Transimpedence Amplifier 2

6.3. Single Photon Detection module

The detection approach utilized in this paper is the current mode signal processing [141] as sketched in Fig. 6.3, which features a current conveyor, amplifier and a comparator. The current conveyor reads the SiPM current at an ideally null impedance, which acts as a current buffer biased by a digital-to-analog convertor (DAC). The amplifier receives the replica of the current and amplifies the weak signal to the output. The current comparator acts as the triggering stage which receives the very fast SiPM current and generates a pulse to alert the data acquisition phase. The current treatment of the signal ensures a large bandwidth, owing to the low input impedance seen by the detector capacitance. The underlying criterion is to scale the signal to suit the voltage range and obtain a wide dynamic range by exploiting the whole SiPM signal with its excellent timing resolution.

Figure 6.4 shows the schematics of the first implemented amplifier [142]. The input stage devices (M0, M1) are biased at the boundary of strong and weak inversion to achieve high transconductance per unit current and low noise while minimizing size and parasitic capacitance. The differential pair suppresses noise from the power supply and the ground, which is not possible with single-ended topologies. The Miller compensation leg (Rc, Cc) keeps the op-amp second pole well beyond the closed-loop bandwidth for good phase margin. A source follower stage is also used at the output to lower the op-amp output impedance enforcing accurate feedback. The closed loop gain is expressed as follows:

$$Z_{CL} = R_f \cdot \left(\frac{1}{1+s\,R_f\,C_f}\right) \cdot \frac{F.\,A_{OL}}{1+F.\,A_{OL}} \tag{6.1}$$

where, A_{OL} is the open loop gain, F is the feedback factor expressed as $Z_i/\left(Z_i + Z_f\right)$. The bandwidth is determined by the smaller of the two poles:

$$f_p = \frac{1}{2\pi\,R_f\,C_f} \tag{6.2}$$

$$f_i = \sqrt{f_c \cdot \frac{1}{2\pi\,(R_f \,||\, R_i)\,(C_f + C_i)}} \tag{6.3}$$

The DC bandwidth gain is R_f and the phase margin is roughly 60 deg. The second amplifier circuit replaces the traditional voltage amplifier by a current amplifier and uses the traditional transimpedance structure of a gain stage shunted by a feedback resistor [143]. The simplified schematic of the circuit is shown in Fig. 6.5. Since the impedence looking into the source of M_0 is much lower than the impedence looking into the drain of M_1, the input current is directed up through M_0 and into the node A and thus the injected charge will adjust the gate voltage of $M_1$with an input voltage of (V_{bias}–V_{GS0}). The gate-source voltage of M_2 is same as M_1 and so the input current is duplicated at the output. The gain of this structure is as follows :

$$\frac{V_{out}}{I_{in}} = \frac{R_f\,A_i - R_{in}}{1+A_i} \approx R_f \tag{6.4}$$

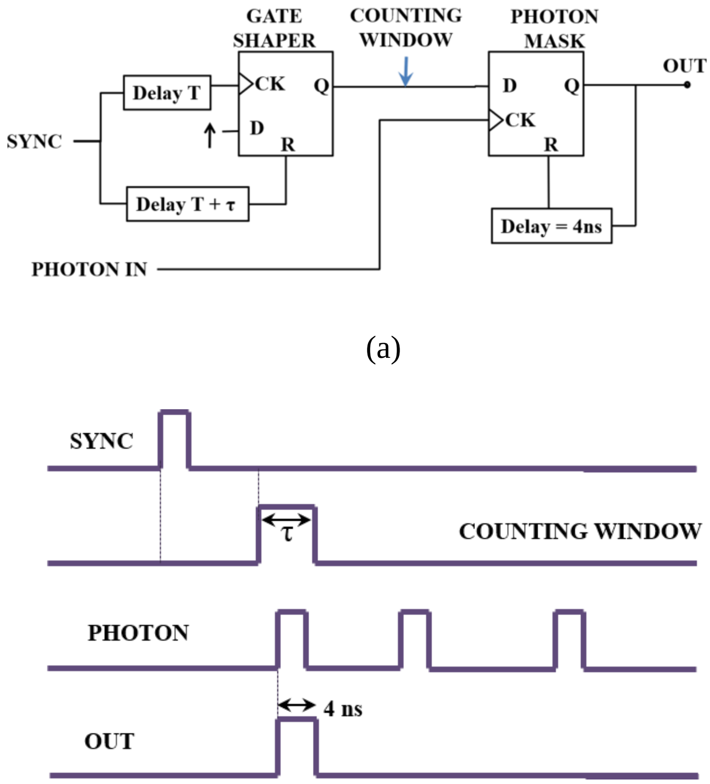
where R_{in} is the amplifier input resistance which is typically low. The simulated frequency response is shown in Fig. 6.6. The achieved gain bandwidth product of the two structures is suitable for this application but can be further increased significantly by tuning the shapes of the transistors.

6.3.1. Generation of the Gated Counting Window

Figure 6.6(a) shows the schematic of the gated window [123] block whose function is to integrate the photons in a certain position of the optical waveform. The first D-type flip-flop acts as a gate shaper where the delayed SYNC signal is fed to the CLOCK and RESET inputs of the D-type flip-flop while its D input is kept at logic level “1”. The output of the flip-flop is initially at logic level “0” and switches to “1” at the rising edge of the SYNC signal. It switches back to “0” at the rising edge of the RESET signal. To compensate for the intrinsic delay of the component, delay units are used both before the CLOCK and RESET inputs. The Q output of the gate shaper determines the the width of the counting window which is set by the time duration τ, tunable from 25 ns down to 1ns in steps of 90 ps. The output of the gate shaper is connected to the D input of the second flip-flop, whilst the photon pulses are applied to its CLOCK input. Only the CLOCK transitions falling within the counting window results in a transition of the Q output (Fig. 6.6(b)). Thus, the second flip-flop acts as the photon selection mask whose function is to selectively mask only those photon pulses that are falling within the counting window (i.e., when D is set to logic level ‘‘1’’) while ignoring all other photon pulses. The Q output of the second flip-flop is connected to the RESET input through a delay line of 4 ns which sets the output pulse to a fixed duration of 4 ns.

6.3.2. 2.5D Silicon Interposer Design

The interposer consists of a 250µm thick silicon substrate with three customizable metal redistribution layers (RDLs) among which two are topside and one is bottom side. The top and bottom side metal layers can be used to route electrical signals between various heterogenous dices which can be mounted on both the sides of the interposer for ultra-compact designs. Openings can be developed in the top and bottom side passivation layer to create custom sized pads with recommended standard pitch of 150 µm. Various chips as well as off-the shelf surface mounted devices can be either flip-chipped or wire-bonded on the interposer platform. The top layer signals and the bottom layer metal are interconnected using through silicon via (TSV) array. The TSV shape is cylindrical with a constant diameter and pitch of 25 µm and 500 µm respectively. The back side of the interposer terminates in a standard ball grid contact array (BGA) while connecting to the TSVs. The standard recommended BGA diameter is 125µm with a 1000µm pitch between openings. The cross-sectional view of the interposer platform



(a)

(b)

Figure 6.6. (a) Gate Shaper, (b) Voltage signal at the various nodes

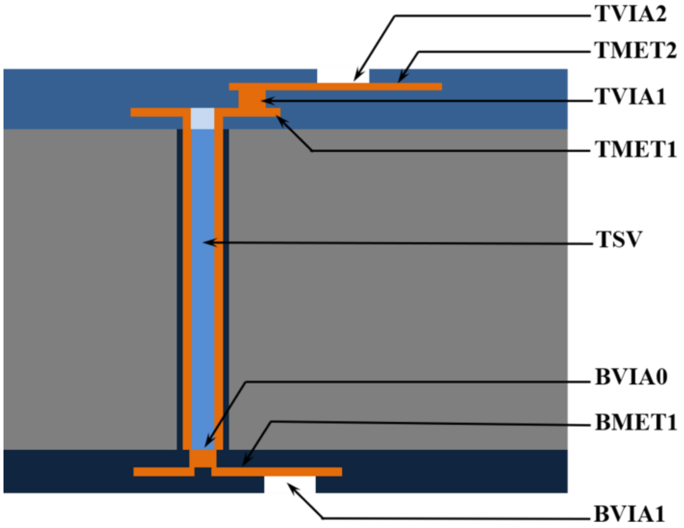
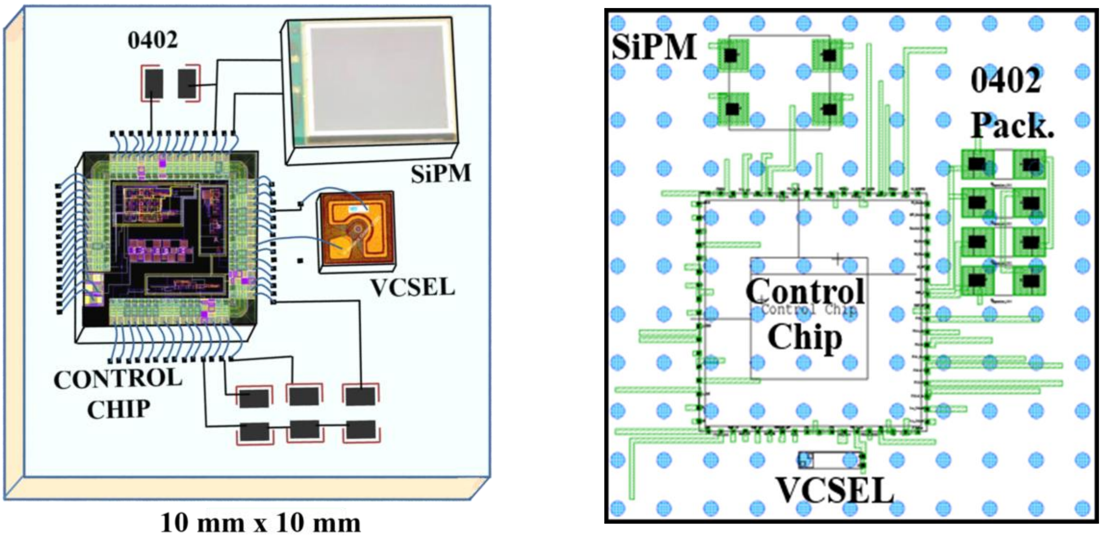


Figure 6.7. Cross-sectional view of the Silicon Interposer Platform with the layer definitions



(a) (b)

Figure 6.8. Silicon Interposer Platform (module 3) (a) Bare dice directly bonded to the silicon interposer along with other ancillary components for a high density multitechnology System-in-Package design, (b) Simplified layout of the proposed platform using Tanner EDA tool

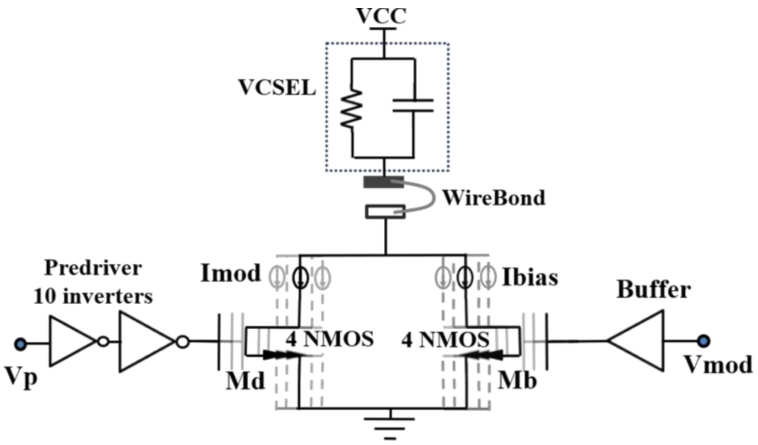


Figure 6.9. Proposed VCSEL driver circuit

along with the metal layers is shown in Fig 6.7. The top layer metal, used for creating custom pads and routing the signals, are represented by TMET1 and TMET2. The small openings in the top passivation layer is meant for the bond pads (either wire-bond or flip-chip). TMET2 and TMET1 can be connected by TVIA1. BMET1 denotes the metal for the redistribution layer on bottom side and for connecting to the TSV holes are created in the field oxide on the bottom side which in designated as BVIA0. The openings in the bottom side passivation layer (BVIA1) can be used to create either bond pads for mounting of components on the bottom side or BGA pad for connection to the PCB-substrate.

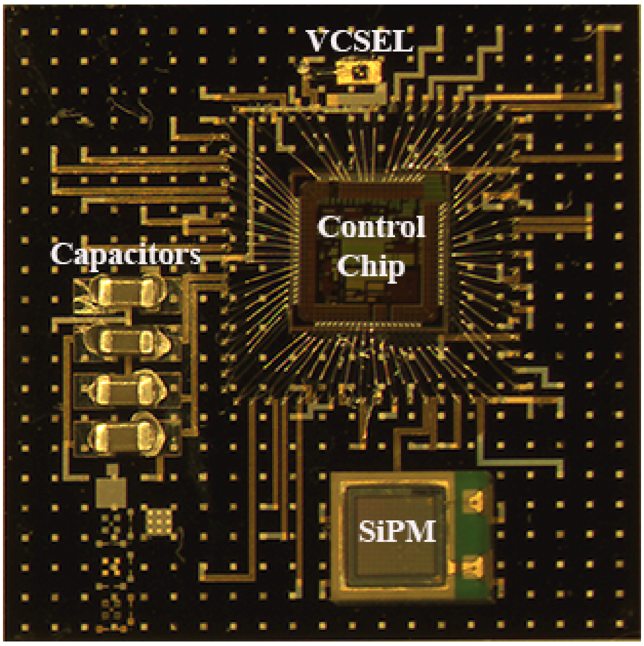
Heterogeneous semiconductor dice, such as microelectronic chips fabricated using different CMOS foundry processes (e.g., 350nm, 130nm, 65nm processes) or chips from different functional domains (e.g. microelectronic, MEMS, microfluidic, photonic), can be closely mounted on interposer platform that are otherwise too expensive or incompatible for monolithic integration. Integration of multiple chips using a platform-based approach leads to a cost-effective and compact design with significant reduction in form factor that is amenable for experimental setup and for integration with other sub-systems. This also offers researchers the unique advantage of rapid prototyping of system-on-a-chip (SoC) prior to full custom implementation. One more advantage of 2.5D integration on a common substrate is significant reduction of parasitic elements such as resistance, inductance or capacitance and better thermal expansion coefficient matching as compared to PCB-based integration. Figure 6.8(a) shows the pictorial representation of the proposed system-in-package (SiP) design based on the 2.5D silicon interposer platform, which consists of a GaAs VCSEL diode chip for pulsed emission of light, an opto-semiconductor (SiPM) with outstanding photon counting capability, control chip designed in CMOS 0.35µm technology and other ancillary components such as bypass capacitors and resistors. The layout of the platform, designed using Tanner EDA software tools, is shown in Fig. 6.8(b). The size of the designed interposer is 10 x 10 mm^2 which consists of both flip-chip and wire-bond interconnections, for.eg., the control chip and VCSEL are wire-bonded to the platform whereas the SiPM is flip-chip bonded.

6.4. Pulsed Light Emission Unit

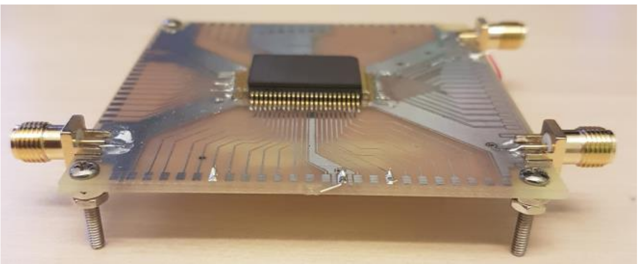
Figure 6.9 shows the schematic for the VCSEL driver circuit, based on a simple two-transistor model [113, 114] operating on a simple principle that a MOS transistor in saturation acts as a current source. NMOS transistors was chosen since NMOS has a higher mobility than PMOS transistors. Rather than choosing complex circuit topologies, this architecture was chosen and carefully designed to source output current, ensuring simple and robust operation. The main advantage of the selected structure is that only a high-speed switch-ON is required to generate a pulse current with a pre-determined time-period and relatively high amplitude. The proposed architecture can produce short optical pulses when driven with a current pulse with a peak amplitude of few milliamperes and width less than 1 ns. The laser diode used here is a 100-μm top-emitting GaAs-based laser having an emission wavelength of 850nm, with very high-modulation rate, sharp optical rise and fall times. The modulation and biasing currents are supplied by means of switches realized with 4 parallel high-voltage NMOS transistors. Thus, the circuit consists of two sets of transistors: MDs with their gates connected to a CMOS-level digital input signal to supply the modulation current, and MBs sets the bias current depending upon the analog voltage applied to their gate terminals. The transistors MDs and MBs act as switches which are turned-ON upon the application of control signal at their gate terminals. The effective W/L of the 4-parallel modulation NMOS is 6400μm/0.5μm and that of the biasing NMOS is 6400μm/0.5μm. The pre-driver consists of a chain of 10 inverters and the effective size of the NMOS transistor in the last inverter is 5600 μm/0.5 μm. The programmable current-mode driver provides digital control of the pulse duration, modulation and bias currents, sinking current directly from the cathode pad of a VCSEL and ensures that it emits constant power over a wide range of conditions. The driver was precisely designed to supply bias and modulation current in the range of 0 to 12 mA and is optimized for driving wide variety of commercially available VCSEL diodes having significant capacitance upto 10 pF and variable resistance ranging from few tens to hundreds of ohms.

6.5. Experimental Results

The proposed optical probe, as shown in Fig. 6.10(a), was implemented in one 2.5D interposer platform integrating multiple heterogeneous chips – GaAs-based VCSEL chip, SiPM and Control chip fabricated in CMOS 0.35μm technology. At first the various chips were tested by bonding them into separate ceramic packages and integrating them onto a PCB substrate (Fig. 6.10(b)). Experimental characterization was carried out for the various building blocks to analyze the performance of the module. The operation of the pulsed light emission unit was characterized by applying a suitable bias voltage (Vb) and varying the gate modulation voltage (Vp) with a pulse width of around 5ns which results in the regulation of current flow to the VCSEL. The VCSEL (VI integrated systems, 850nm) has contacts on the top-surface and is connected to the driving circuit using wire bonds. The resulting modulated light was loosely focused on a fast silicon single-photon detector, whose output was coupled to a Time Corelated Single Photon Counting (TCSPC) module (Picoquant). Figure 6.11 shows the resulting detected signal from the input-modulated VCSEL with a repetition rate of 100 MHz and FWHM of 425 ps. The measured peak output power of the VCSEL was 5.8 mW. We measured the I-V characteristics (Fig. 6.12(a)) and observed the linearity of the optical power with the applied



(a)



(b)

Figure 6.10. Proposed Optical probe, (a) Miniaturized version with the control chip, SiPM, VCSEL chip and peripheral capacitors integrated on 2.5D silicon interposer platform having a size of 1 x 1 cm^2, (b) PCB Module – control chip, SiPM and VCSEL are enclosed in separate packages and soldered onto a PCB

voltage. The slight deviation in linearity might be due to the fluctuation in the optical power of the laser source. Stability of the system over time is one the key requirement in NIRS measurements where any drift in the optical power of the laser sources or detection efficiency of the SiPM module may adversely affect the measurement results. Hence, the system has been extensively tested and verified by recording the optical responses for a period of one hour or more, with each distribution of time of flight (DTOF) curve collected with an integration time of 1 second. Recorded temporal widths of each recorded curve present a very stable behavior having values in the order of 400–430 ps with the centroid variations of just few picoseconds.

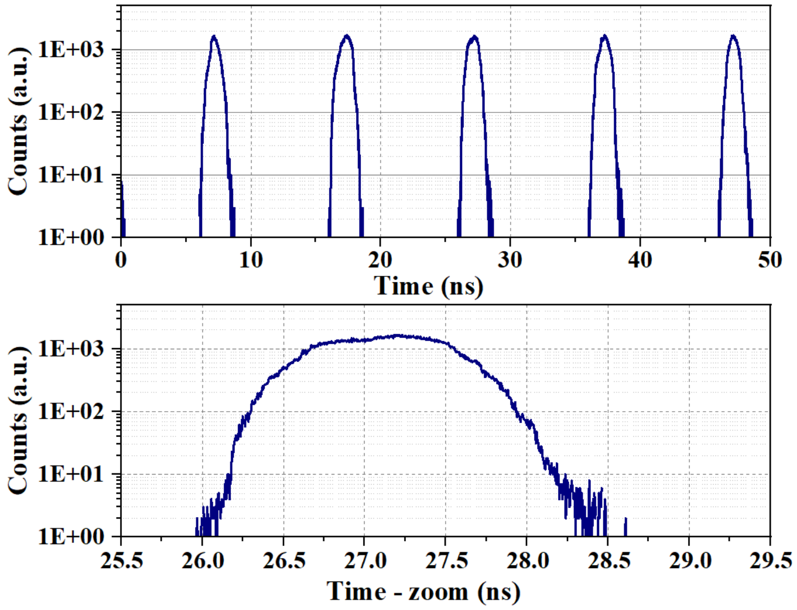
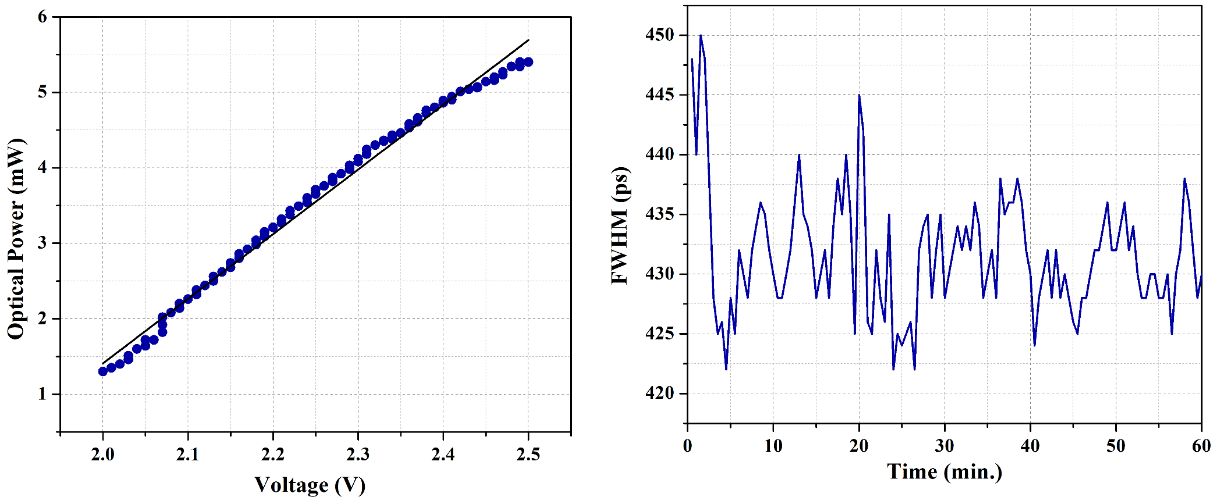


Figure 6.12. The measured optical waveforms. The driving input voltage amplitude varies between 1.0 and 3.0 V



(a) (b)

Figure 6.11. Stability analysis (a) linearity of the emitted optical power, (b) Variation of the temporal widths of the recorded curves

Figure 6.12 (b) shows the variation of FWHM over a period of 1 hour and it also observed that there is very nominal drift in the centroid position of each curve which is occurring either due to the laser driving circuits or the time-measurement electronics. The integrated synchronization block is capable of generating SYNC pulses of frequency ranging from 1 to 160 MHz. Figure 6.13 shows the plot of the output frequency of the synchronization block as as a function of applied control voltage, where the output frequency varies almost linearly with the control voltage. The module is also capable of processing an off-chip trigger signal for

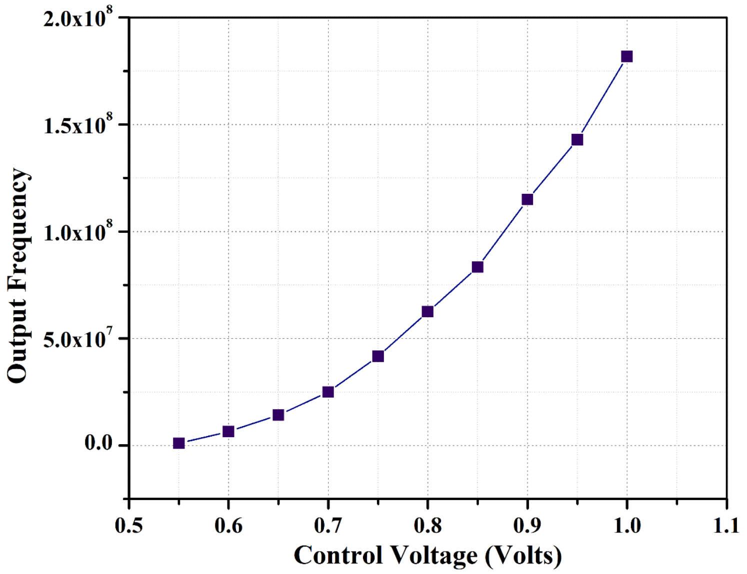


Figure 6.13. Relation between the output frequency as a function of the control voltage

synchronization when an external laser source is used for experimentation. Variations of count rates over time account for potential drifts of the detection module and remain below ±1% for the detection module as shown in Fig. 6.14 which demonstrates the stability of the count-rate of the single-photon counting module over a time- period of 60 min. The photon rate was kept at a very low level using a continuous-wave laser source with the module running at ambient temperature without any cooling system. The fluctuation of less than 1% might be due to poissonian noise and slight instability in the laser output power or ambient conditions. These values highlight the stability of the module without any cooling system. Temperature control using a two- stage Thermo Electric Cooler (TEC) is a major drawback for the miniaturization of the time-gated module to be integrated in an optode with the light source. The measured IRF of the module is shown in Fig. 6.15 which has a FWHM of 1ns in response to an optical pulse having a width of 5ns. In order to check the accuracy of the gating circuit, Time-of-Flight experiments were conducted with the integrated VCSEL module. An absorbing medium was placed on top of the optode and the separation ‘d’ was varied. The peak of the light pulse is expected to be detected after a time-period of (c/d), where c is the speed of light. This time-period corresponds to the round-trip time taken by the photons to reach the detection module when the SiPM unit starts detecting the maximum count. Figure 6.16 shows the linearity of the delay versus the separation

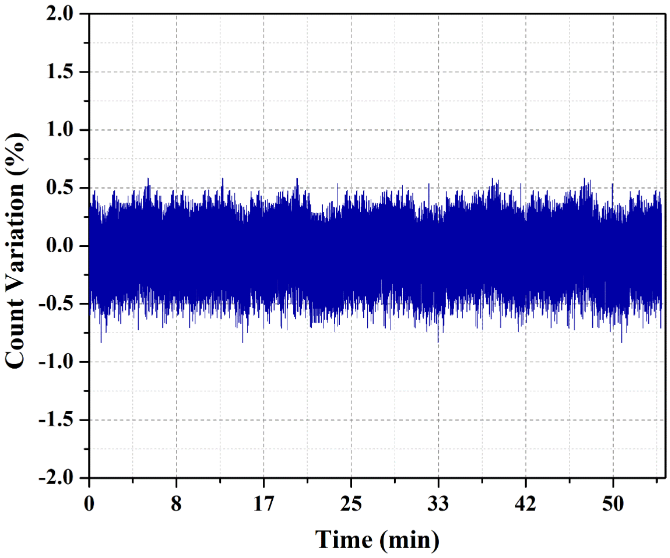


Figure 6.15. Stability of the module count-rate over a time range of 1 hour

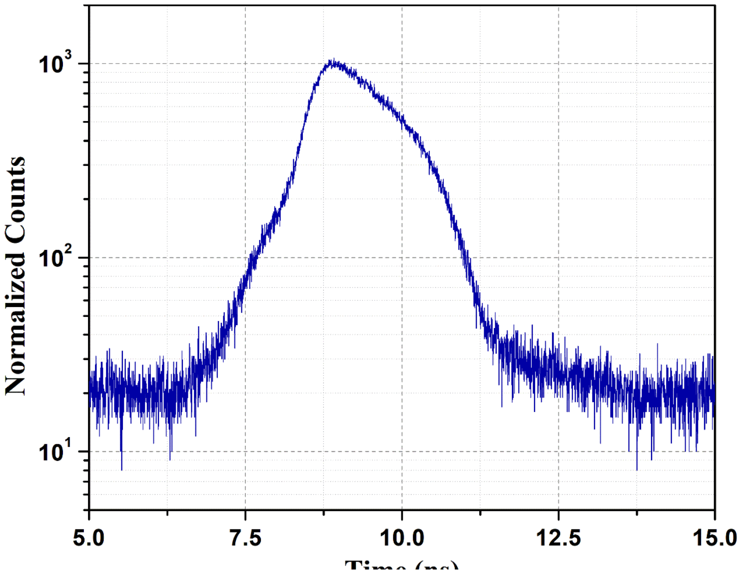


Figure 6.14. Measured IRF of the module with an emitted laser pulse width of 5ns

6.6. Conclusion

We presented a compact miniaturized time-domain near-infrared spectroscopy device, specifically designed for portability and ease of operation. The system is based on state-of-art components, including a pulsed diode laser operating at 850nm and a 1 mm^2 active area SiPM detection module. The system characterization demonstrated an instrument response function of 1 ns when operated with a synchronization signal of width 5ns and an average light power of 2 mW while running at 100 MHz repetition rate. The real-time acquisition and reconstruction

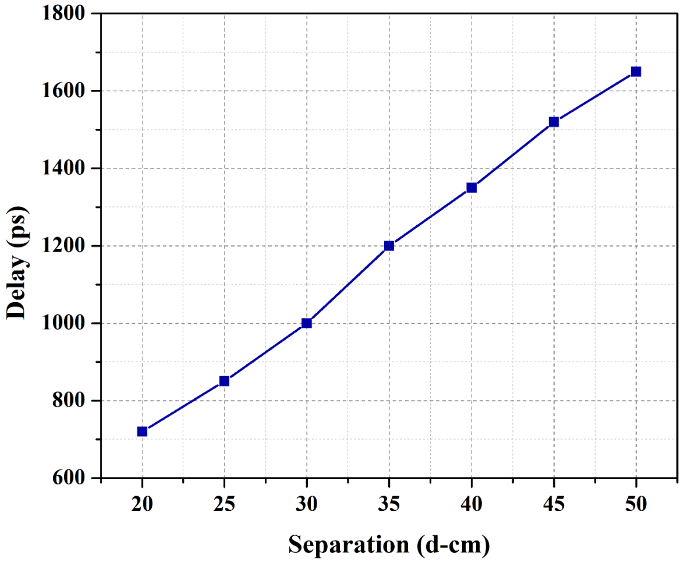


Figure 6.16. Time-of-Flight experiment with the compact module

of DTOF curves can be downloaded to a remote PC for data processing. The experimental validation shows the good linearity of the proposed system and exhibits a measurement stability of ±1% for operation time longer than an hour. The cooling system is a limiting factor in the complete integration of a fast-gated single photon counting module on a single chip. Henceforth, a full characterization of all the building blocks of the module was done without any cooling system, testing both its electric and optical properties. A practical application of this device was then demonstrated by means of a setup that measures the time of flight. The compact size, flexibility, and customization possibilities can be considered as the first step towards portable multichannel and multi-wavelength TD-NIRS diagnostic tools for wearable healthcare applications.

6.7. Assessment

The silicon interposer platform has surface scratching and TSV recessing issues. TSV resistance is generally very high from 10 - 100 kilo ohms range, which is affecting the functioning of the various control blocks especially the pulsed light emission operation and photon count read. This problem is being addressed now as the new batch of silicon interposer wafers employ a different TSV fill strategy leading to an improved or lower TSV resistance. Alternatively, we are also ordering separate organic interposer platform from Kyocera corporation which exhibits features such as Z0 matching and low insertion loss.

CHAPTER 7 GENERAL DISCUSSION AND RECOMMENDATIONS

Functional near infrared spectroscopy (fNIRS) is an emerging brain imaging technique that has been used for many applications in research such as the study of neurovascular coupling, the characterization of brain activity during specific tasks/sensori stimuli and more recently brain computer interface applications. However, its penetration in clinical practice has been quite limited. Among difficulties for translation of fNIRS to clinical context are inherent complexities in headgear installation and image reconstruction. Indeed, the technique requires a time-consuming montage optimization step with careful positioning and assignment of distinct source and detector pairs about 3 to 5 cm apart. However, as source and detectors get closer on the scalp, the sensitivity to signal from superficial tissue gets larger and overwhelms signal originating from the cortex. Due to finite dynamic range in silicon/PMT detectors, and exponential decay of sensitivity, sampling deep tissue is limited by noise and detector dynamic range. Even though the zero source-detector distance configuration is the one that has highest sensitivity to deep tissue, since detected photons propagate a shorter path, detector dynamic range limits its applicability. Recent progress in the development of very fast gated detectors (picoseconds) have shown that it is possible to exploit a time-domain photon-counting approach to quantify changes in the number of photons scattering back to nearly where they came from, by the design of a single optode which incorporates source and detectors very close to each other within the assembly. Even with a time-domain single-photon technique, photon pile-up in early gates with increasing illumination power limits sensitivity to late photons leading to long integration times. Saturation of the detector by shallow photons is eliminated by fast gating (turning it off when light is pulsed) enabling the choice of a detection window that is more sensitive to photons having propagated deep in cortical tissue. Thus, when the word optode is used, it means one pulsed light source and one detector separated by about 2 mm. The combination of large (hundreds, fewer initially) numbers of such optodes could then yield a completely novel fNIRS system architecture where each optode monitors hemodynamic changes locally under its position facilitating the creation of images and hemodynamic activation maps in real time. Furthermore, by moving in time the gating window, one can increase sensitivity to surface tissue and create regressors for systemic physiology originating in the skin. Our aim with this project is to develop a prototype optode for fNIRS imaging, integrating CMOS based time-gated source-detector pair. Until now this technology was demonstrated using off-the-shelf photon counters that are expensive and do not scale to an application with large numbers of optodes. To solve this issue and focussing on ease of use, significant progress has been made in promoting this approach by integrating detector-gating and light-pulsing electronics in a single 2x2 mm^2 CMOS (complementary metal-oxide semiconductor) chip. We developed an integrated chip [126, 144] with a fast pulse-laser driver and synchronized detection-gated circuit. With the perspective of simplifying the output, photon counts are accumulated in a counter on-chip and transmitted as an analog signal mimicking the signal one would get from normal linear analog photodetectors. Our approach will facilitate clinical adoption in many applications and change the landscape of fNIRS hardware commercially due to significant optode-size reduction and the elimination of optical fibers. Our unique-in-the-world developments in hardware, algorithms and their applications to the continuous monitoring setting will enable to pursue these techniques in the epilepsy unit settings.

7.1. Research Contributions

1. Beyond EEG, there are currently no non-invasive alternatives to monitor brain status over long periods, as putting a subject in the MRI for extended periods (days) is not feasible and costly. A novel integrated CMOS electronic and optical design was developed, incorporating gated detection and pulsed-laser illumination within a single chip, a feat that remained out of reach until now. Laser source and detector were combined in a single package demonstrating integration feasibility. Miniaturization is possible as the probes use detector and laser source placed side-by-side, leading to a very small SDD, which will not only allow interrogating deeper tissue volumes (cortical surface) with late time gates but also integrate early gates to monitor surface physiology (skin, bone) and remove its contributions to deep signals. A completely miniaturized system-in-package optode will enable very high-density recordings which have been demonstrated to yield fMRI-quality reconstructions using bulky complex systems. Furthermore, applications to mobile settings will be facilitated promoting brain imaging to more ecological settings if needed. Our work has implemented many technological breakthroughs: 1) micro-electronic integrated time-gated detection with gated single photon avalanche photodiode (SPAD) detectors will increase sensitivity and simplify the montage on the head with each optode measuring directly under its position. 2) Integration of a surface-tissue measures at every position will improve physiological confound removal 3) reduction in the size of the optode (1x1x0.3

cm) will improve comfort and facilitate the design of recording caps. 4) Simplification of the interface to signals (counter on CMOS) will reduce overall cost. With preliminary results in microelectronic design and in the proposed clinical application, we believe that we are uniquely positioned to carry the proposed work and anticipate translational benefits for the Canadian industry.

2. The proposed system features an integration of an optical interface to optimize both light collection on the SPAD area and light emission from the vertical-cavity surface-emitting laser (VSCEL). Combining the package with auxiliary parts on a common silicon interposer platform delivers a packaged optode for gated TD-fNIRS. Instead of time-correlated-single-photon-counting (TCSPC) and with the perspective of simplifying the output, photons will be counted on the chip and transmitted as an analog signal (mimicking the signal one would get from normal linear analog photodetectors) to provide a summary measure that can be relayed to the external microcontroller without the requirement of a fast FPGA or external photon counter. The design of the probe incorporates synchronized single-photon detection circuit composed of silicon photomultiplier to maximize light harvesting capability and light-pulsing electronics. Overall, based on current design, the size of the CMOS integrated circuit is 2.5x2.5 mm^2. The integrated circuit is assembled on a 2.5D silicon interposer platform of size 1 cm and interfaced with other ancillary components such as capacitors and resistors. The laser diode and SiPM module are connected to the control chip by wire-bonding and conductive epoxy respectively. The interposer was then soldered on a Printed Circuit Board (PCB) for experimental characterization. The optical face of the package will then be sealed with optically transparent epoxy forming a fully encapsulated optode. Our approach will facilitate clinical adoption in many applications and change the landscape of Near Infrared Spectroscopy (NIRS) hardware commercially due to significant optode-size reduction and the elimination of optical fibers.

7.2. Recommendations and Future Works

We have introduced the design of a CMOS based dual-wavelength integrated gated source-detector optode for fNIRS imaging. The detector, based on single photon avalanche diodes (SPADs), features an innovative pixel architecture with independent active time-gating, quenching and reset circuitry. Key features include the ability to operate in multiple operation modes such as the option of a counting-only mode as well as timing mode for capturing photons in a specified time-window. Recently, time-resolved detection of faint light signals in the near infrared region of the spectrum has proven to be an effective tool for many applications such as food assessment, breast cancer analysis, medical imaging, pharmaceutical analysis, LIDAR (Light Detection and Ranging), Time Correlated Single Photon Counting (TCSPC) and many others. Several detectors have been reported in literature, many among them have good single-photon sensitivity and timing resolution in the near-infrared spectrum but Single-photon avalanche diodes (SPADs) are the most suitable because they are not prone to be damaged by strong light pulses, smaller in size, low bias voltage and can be easily manufactured in a standard complementary metal–oxide semiconductor (CMOS) process. SPADs are solid-state detectors that have been known for decades and have been increasingly exploited in several fields where light signals are to be acquired down to single-photon level with better timing performance, extremely low noise and extended dynamic range (DR). Due to the breakthroughs in the fabrication process technology, the designs of SPADs have significantly evolved over the recent years and the realization of parallel large arrays of SPADs became feasible. The state-of-the-art monolithic SPAD array, fabricated in CMOS technology, has precipitated the integration of high-performance precise on-chip timing circuitry such as time-to-digital converters and other data processing units, thus enabling the capability to accurately time stamp individual incoming photons. Various CMOS imagers, capable for both photon-counting and photon-timing operations, have been presented in literature for the increase of the throughput of single-molecule fluorescence spectroscopy, acquisition and mapping of several photoelectron timestamps for each scintillation photons in PET systems, detection of fluorescence emitted by biological tissue to monitor cancer cells during surgery and super-resolution microscopy allowing for the detection of extremely weak signals. None of these imagers are capable of time-gating an SPAD array with ON and OFF transition times down to few hundreds of picoseconds, at repetition rates up to 80 MHz, allowing for the sharp temporal filtering of impinging photons to detect the signal of interest while rejecting the unwanted burst of photons. Until now this technology was demonstrated using off-the-shelf photon counters that are expensive and do not scale to an application with large numbers of optodes.

Recent progress in the development of very fast gated detectors (picoseconds) have shown that it is possible to exploit a time-domain photon-counting approach to quantify changes in the number of photons scattering back to nearly where they came from, by the design of a single optode which incorporates source and detectors very close to each other within the assembly.

Design of such optode could then yield a completely novel functional near infrared system (fNIRS) architecture where it monitors hemodynamic changes locally under its position facilitating the creation of images and hemodynamic activation maps in real time. The null/small source-detector distance (ns-SDD) configuration is the one that has highest sensitivity to deep tissue, since detected photons propagate a shorter path, detector dynamic range limits its applicability. Even with a time-domain single-photon technique, photon pile-up with increasing illumination power limits sensitivity to late photons leading to long integration times.

We have proposed a novel system integrating a pair of two wavelength laser sources sensitive to oxygenated hemoglobin (HbO) and deoxygenated or reduced hemoglobin (HbR) and one time-gated SPAD array detector separated by less then 2 mm. Saturation of the detector by shallow photons is eliminated by fast gating (turning it off when light is pulsed at frequency over 100 MHz) enabling the choice of a detection window that is more sensitive to photons having propagated deep in cortical tissue. The detector gating technique for TCSPC allows probing at depth with small source-detector distance by rejecting the large signal from the surface which, considering the limited dynamic range of detectors, otherwise restricts the sensitivity of the probe to the vicinity of its tip. Furthermore, by moving in time the gating window, one can increase sensitivity to surface tissue and create regressors for systemic physiology originating in the skin.

The block diagram of the system is illustrated in Fig. 7.1. The sensor, consisting of an array of 16 × 16 square SPAD detectors with rounded corners, was developed in a 0.35 µm high-voltage CMOS technology. The size of each pixel is 24 µm × 24 µm having 100 µm pitch and integrated with fast readout interface electronics to achieve sub-nanosecond temporal resolution. The full sensor chip, has dimensions of 2.5 × 2.5 mm^2. The integrated phase locked loop (PLL) provides the trigger signal for laser pulsing and time-gated single photon detection unit. The time-gate timing is synchronized with respect to the laser pulse by means of the on- chip delay line, whose output serves as the control signal for the frontend of each SPAD pixel and is propagated through balanced binary trees for the minimization of the skew between the gating signals. Finally, gating pulses, with a user-tunable time period, are generated by the pulse width modulators.

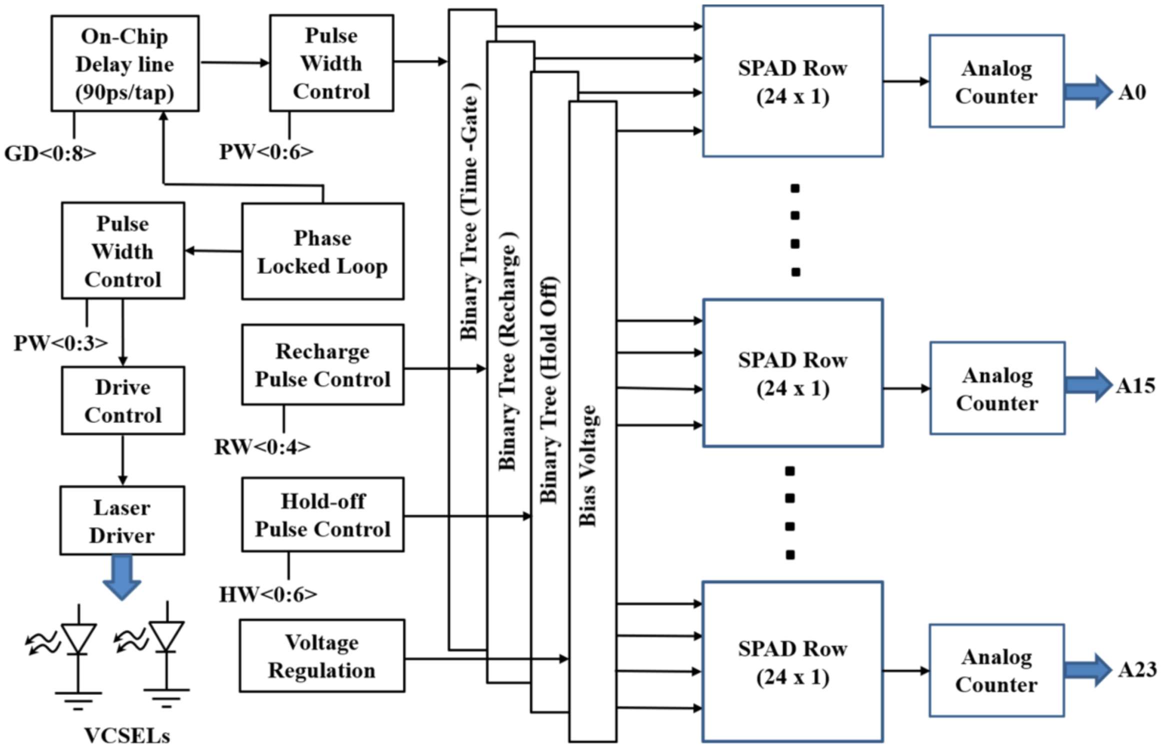


Figure 7.1. System block diagram showing an optical trigger module, pulse width control units, on-chip delay lines, binary trees, pulse generators (PGs), and 32 groups of SPAD array

SPAD detectors are only active and able to detect incoming photon when it is reverse biased above the junction breakdown voltage and thus it is possible to turn ON and OFF the SPAD by lowering the applied bias voltage below breakdown. The difference between the applied voltage and the breakdown voltage is referred to as excess bias voltage (V_{EX}) which can be applied during a short period of time. The detector is only sensitive during this time-period referred to as gate that can be adjustable and periodically repeated in synchrony with the pulsed light source. The frontend circuitry is properly designed to accommodate an estimated SPAD capacitance of 120 fF including the post-layout parasitics while providing prompt avalanche quenching and fast detector reset (280 ps anode fall-time, 20–80% transition). The simplified schematics of the SPAD frontend is shown in Fig. 7.2. The gating operation is performed by means of the PMOS transistor (M_G), which pulls the SPAD anode voltage up to an intermediate supply voltage above the diode breakdown voltage (V_{BD}), essentially the excess bias of the SPAD. Simulated time-gated operation is shown in Fig. 7.3, where the blue waveform indicates the scenario when a photon is detected by the SPAD and the orange one demonstrates the condition of no-photon detection. The circuit makes use of thick oxide transistors for time-gating, avalanche sensing and quenching, and low voltage transistors for other control units to

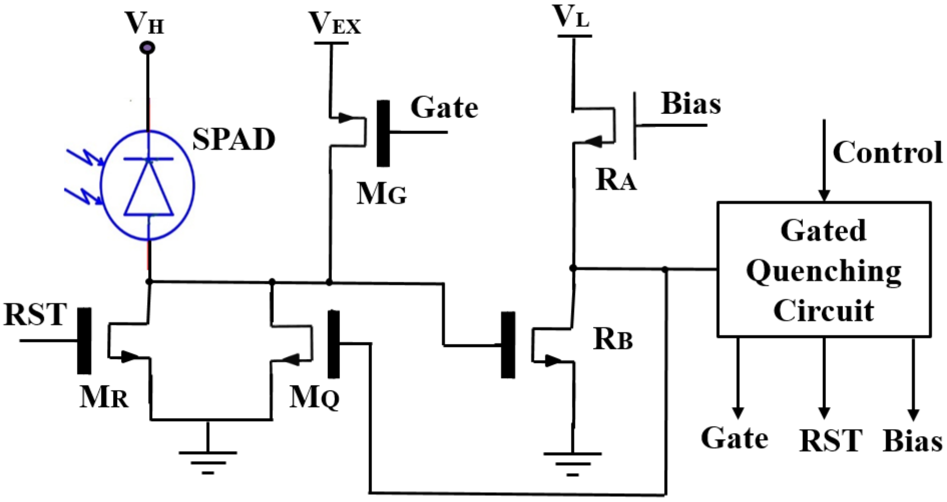


Figure 7.2. SPAD gating frontend circuit

reduce area occupation. M_R serves the purpose of reset transistor to quickly bring the SPAD above breakdown voltage and is activated for about 1 nanosecond while M_Q remains ON while the SPAD is armed. M_Q acts as a dynamic resistor providing the initial passive quenching of the avalanche and makes sure that the anode voltage does not drift during the time SPAD is active.

Once a photon is detected, hold-off time generator is activated to enforce the hold-off time, selectable between 0 to 3 gate windows to ensure that the re-arming of the SPAD occurs in synchronous with the gate signal. The SPAD front-end can also be configured to be operating in the free-running mode, where the SPAD is always kept above breakdown voltage making it feasible to detect photon at any time rather than a specified time-gate. In free-running mode, the gate signal is still required to activate the hold-off time generator. The layout of two pixels are shown in Fig. 7.4, with a pitch of 24 µm and a fill factor of 40%. The SPAD comprises of a reverse biased p+ deep n-well junction diode, which has been designed carefully and optimized to achieve a high fill factor with a common shared well for the cathodes. The goal is to achieve low noise, low jitter, and high photosensitivity, while maximizing the utility by utilizing minimal on-chip integrated circuitry for biasing and quenching photon-induced avalanches. The absorbed or incoming photons are converted into electronic digital pulses which are then fed to the processing logic. In order to obtain a much wider photon-sensitive

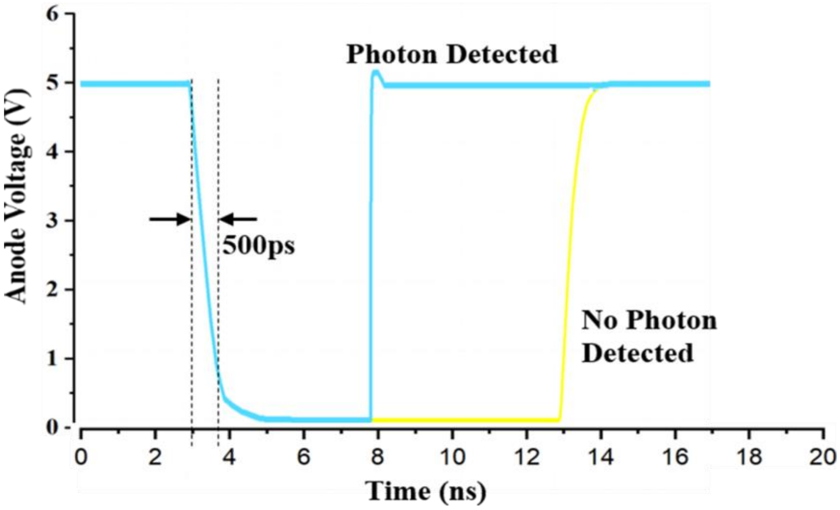


Figure 7.3. SPAD anode voltage waveforms during photon detection and no-detection

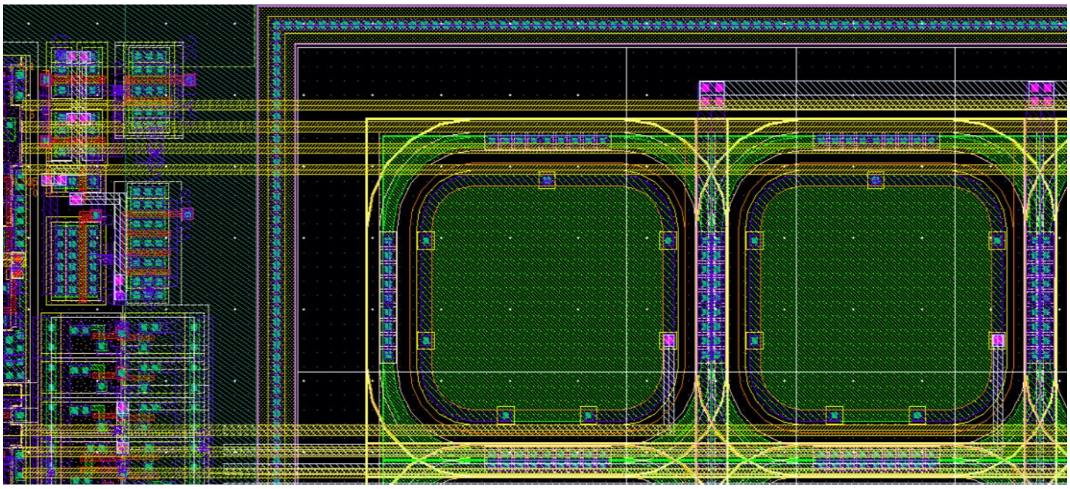


Figure 7.4. Layout of two pixels with pitch 24μm and fill-factor of 40%

area, we are also proposing to develop high-performance back-illuminated three-dimensional stacked single-photon avalanche diode (SPAD) array[78, 145], which will be implemented in 45-nm CMOS technology. The SPADs will be developed to achieve a higher photon detection efficiency over a wider spectrum, owing to the inherent wide depletion region which may also result in low tunneling noise and better timing jitter. With the optimized 3-D stacked 45-nm CMOS technology for back-illuminated image sensors, the SPAD can achieve a very dark count rate of 55.4 cps/μm^2 and a photon detection probability of over 10% in the 600–920 nm wavelength range. The timing-jitter can also be minimized (below 100 ps full width at half-maximum) with negligible exponential diffusion tail at room temperature.

CHAPTER 8 CONCLUSION

Recent development in time-domain technology and growth in photonics components for light sources and detectors has enabled the scaling down of size leading to the development of smaller optical probes [106, 119, 137] that are easier-to-handle, as in the case of brain measurements where smaller optical probes can be easily incorporated in the cap simplifying the process of image reconstruction. In this paper we first presented the design of a time-gated single-photon detection module on a printed circuit board (PCB) using discrete components and equipped with a pulsed light source. For the detection of faint light signal, Single Photon Avalanche Diode (SPAD) is utilized, thanks to its high photon detection efficiency (PDE), low noise, high timing resolution and good reliability. By modulating the biasing voltage of the SPAD from below to above its breakdown threshold, time-gates are generated, enabling the detection of impinging photons in a specific time-period. As a first step towards total integration, a control custom chip of dimensions 2.5mm x 2.5mm was fabricated in 0.35µm CMOS technology, reproducing and emulating the functionality of PCB based time-gated circuits. We adopted system-in-package (SIP) approach in which the various chips (two VCSEL diodes, one control chip and one SPAD chip) are placed side-by-side in a common package and interconnected at the package level. Both the GaAs-based VCSEL and SPADs are driven by the wire-bonded control chip. Although, SPADs exhibits good single-photon sensitivity, timing resolution and can be easily integrated into CMOS technology but its performance is greatly limited due to the small active area (few tens of micrometers in active diameter), which affects the overall collection efficiency of the diffused light. For obtaining high diffused photon harvesting efficiency, we developed a miniaturized SiPM based compact probe that can be directly put in contact with the sample under test for applications such as TD-NIRS eliminating the need for complex optical setup involving optical fibers and lenses for wearable real-time monitoring applications. The proposed unique compact optode represents the first step towards miniaturized state-of-art compact TD-NIRS diagnostic instruments and consists of an embedded light emission unit in the form of a pulsed VCSEL diode and the light detection unit which comprises a SiPM along with other CMOS integrated control circuitry. To further scale down the size of the optical probe, 2.5D silicon interposer platform was developed in a compact physical form, integrating the various heterogeneous semiconductor chips which were fabricated in different foundry processes. Furthermore, we have proposed a novel system integrating time-gated SPAD array detector and a pair of two wavelength laser sources sensitive to oxygenated hemoglobin (HbO) and deoxygenated or reduced hemoglobin (HbR). Each of the SPAD element is fast gated (can be turned on and off when light is pulsed at frequency over 100 MHz) enabling the choice of a detection window that is more sensitive to photons having propagated deep inside the tissues. Until now this technology was demonstrated using off-the-shelf photon counters that are expensive and do not scale to an application with large numbers of optodes.

BIBLIOGRAPHY

[1] M. A. Karami, "Deep-submicron CMOS single photon detectors and quantum effects," 2011.

[2] A. Torricelli *et al.*, "Time domain functional NIRS imaging for human brain mapping," *Neuroimage,* vol. 85, pp. 28-50, 2014.

[3] S. T. Hess, S. Huang, A. A. Heikal, and W. W. Webb, "Biological and chemical applications of fluorescence correlation spectroscopy: a review," *Biochemistry,* vol. 41, no. 3, pp. 697-705, 2002.

[4] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *Neuroimage,* vol. 63, no. 2, pp. 921-935, 2012.

[5] D. Contini *et al.*, "Effects of time-gated detection in diffuse optical imaging at short source-detector separation," *Journal of Physics D: Applied Physics,* vol. 48, no. 4, p. 045401, 2015.

[6] S. Del Bianco, F. Martelli, and G. Zaccanti, "Penetration depth of light re-emitted by a diffusive medium: theoretical and experimental investigation," *Physics in Medicine & Biology,* vol. 47, no. 23, p. 4131, 2002.

[7] A. Pifferi, A. Farina, A. Torricelli, G. Quarto, R. Cubeddu, and P. Taroni, "Time-domain broadband near infrared spectroscopy of the female breast: a focused review from basic principles to future perspectives," *Journal of Near Infrared Spectroscopy,* vol. 20, no. 1, pp. 223-235, 2012.

[8] M. Pagliazzi *et al.*, "Time domain diffuse correlation spectroscopy with a high coherence pulsed source: in vivo and phantom results," *Biomedical optics express,* vol. 8, no. 11, pp. 5311-5325, 2017.

[9] E. Alerstam *et al.*, "Single-fiber diffuse optical time-of-flight spectroscopy," *Optics letters,* vol. 37, no. 14, pp. 2877-2879, 2012.

[10] K. Vishwanath, B. Pogue, and M.-A. Mycek, "Quantitative fluorescence lifetime spectroscopy in turbid media: comparison of theoretical, experimental and computational methods," *Physics in Medicine & Biology,* vol. 47, no. 18, p. 3387, 2002.

[11] A. Pifferi *et al.*, "Time-resolved diffuse reflectance using small source-detector separation and fast single-photon gating," *Physical review letters*, vol. 100, no. 13, p. 138101, 2008.

[12] A. Puszka *et al.*, "Spatial resolution in depth for time-resolved diffuse optical tomography using short source-detector separations," *Biomedical optics express,* vol. 6, no. 1, pp. 1-10, 2015.

[13] A. Torricelli *et al.*, "Time-resolved reflectance at null source-detector separation: improving contrast and resolution in diffuse optical imaging," *Physical review letters,* vol. 95, no. 7, p. 078101, 2005.

[14] J. Steinbrink, H. Wabnitz, H. Obrig, A. Villringer, and H. Rinneberg, "Determining changes in NIR absorption using a layered model of the human head," *Physics in Medicine & Biology*, vol. 46, no. 3, p. 879, 2001.

[15] A. Puszka *et al.*, "Time-resolved diffuse optical tomography using fast-gated single-photon avalanche diodes," *Biomedical optics express,* vol. 4, no. 8, pp. 1351-1365, 2013.

[16] L. Di Sieno *et al.*, "Diffuse optics using a dual window fast-gated counter," *Applied optics,* vol. 53, no. 31, pp. 7394-7401, 2014.

[17] A. Tosi *et al.*, "Fast-gated single-photon counting technique widens dynamic range and speeds up acquisition time in time-resolved measurements," *Optics express,* vol. 19, no. 11, pp. 10735-10746, 2011.

[18] L. Di Sieno *et al.*, "Functional near-infrared spectroscopy at small source-detector distance by means of high dynamic-range fast-gated SPAD acquisitions: first in-vivo measurements," in *European Conference on Biomedical Optics*, 2013, p. 880402: Optical Society of America.

[19] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, M. Richardson, and V. Megalooikonomou, "Seizure detection using EEG and ECG signals for computer-based monitoring, analysis and management of epileptic patients," *Expert Systems with Applications,* vol. 42, no. 6, pp. 3227-3233, 2015.

[20] C. Plummer, A. S. Harvey, and M. Cook, "EEG source localization in focal epilepsy: where are we now?," *Epilepsia,* vol. 49, no. 2, pp. 201-218, 2008.

[21] S. Ramgopal *et al.*, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy & behavior,* vol. 37, pp. 291-307, 2014.

[22] S. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry,* vol. 76, no. suppl 2, pp. ii2-ii7, 2005.

[23] K. Peng, "Study of the Hemodynamic Response to Interictal Epileptiform Discharges in Human Epilepsy Using Functional Near Infrared Spectroscopy," École Polytechnique de Montréal, 2016.

[24] A. Ray and S. M. Bowyer, "Clinical applications of magnetoencephalography in epilepsy," *Annals of Indian Academy of Neurology,* vol. 13, no. 1, p. 14, 2010.

[25] B. Krishnan *et al.*, "Epileptic focus localization based on resting state interictal MEG recordings is feasible irrespective of the presence or absence of spikes," *Clinical Neurophysiology,* vol. 126, no. 4, pp. 667-674, 2015.

[26] H. Shigeto *et al.*, "Feasibility and limitations of magnetoencephalographic detection of epileptic discharges: simultaneous recording of magnetic fields and electrocorticography," *Neurological research,* vol. 24, no. 6, pp. 531-536, 2002.

[27] S. H. Faro, F. B. Mohamed, M. Law, and J. T. Ulmer, *Functional neuroradiology: principles and clinical applications*. Springer Science & Business Media, 2011.

[28] V. Scarapicchia, C. Brown, C. Mayo, and J. R. Gawryluk, "Functional magnetic resonance imaging and functional near-infrared spectroscopy: insights from combined recording studies," *Frontiers in human neuroscience,* vol. 11, p. 419, 2017.

[29] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature,* vol. 412, no. 6843, p. 150, 2001.

[30] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature,* vol. 453, no. 7197, p. 869, 2008.

[31] T. Huppert, J. Barker, B. Schmidt, S. Walls, and A. Ghuman, "Comparison of group-level, source localized activity for simultaneous functional near-infrared spectroscopy-magnetoencephalography and simultaneous fNIRS-fMRI during parametric median nerve stimulation," *Neurophotonics,* vol. 4, no. 1, p. 015001, 2017.

[32] V. Quaresima and M. Ferrari, "Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review," *Organizational Research Methods,* p. 1094428116658959, 2016.

[33] R. McKendrick, R. Mehta, H. Ayaz, M. Scheldrup, and R. Parasuraman, "Prefrontal hemodynamics of physical activity and environmental complexity during cognitive work," *Human factors,* vol. 59, no. 1, pp. 147-162, 2017.

[34] M. Izzetoglu, A. Devaraj, S. Bunce, and B. Onaral, "Motion artifact cancellation in NIR spectroscopy using Wiener filtering," *IEEE Transactions on Biomedical Engineering,* vol. 52, no. 5, pp. 934-938, 2005.

[35] M. Strait and M. Scheutz, "What we can and cannot (yet) do with functional near infrared spectroscopy," *Frontiers in neuroscience,* vol. 8, p. 117, 2014.

[36] P. Tubes, "Basics and Applications," *Hamamatsu Photonics KK, Iwata City,* pp. 438-0193, 2006.

[37] A. V. Krishnamoorthy *et al.*, "16 x 16 vcsel array flip-chip bonded to cmos vlsi circuit," *IEEE Photonics Technology Letters,* vol. 12, no. 8, pp. 1073-1075, 2000.

[38] H. Finkelstein, "Shallow-trench-isolation bounded single-photon avalanche diodes in commercial deep submicron CMOS technologies," UC San Diego, 2007.

[39] B. Kardynał, S. Hees, A. Shields, C. Nicoll, I. Farrer, and D. Ritchie, "Photon number resolving detector based on a quantum dot field effect transistor," *Applied physics letters,* vol. 90, no. 18, p. 181114, 2007.

[40] A. J. Kerman *et al.*, "Superconducting nanowire photon-counting detectors for optical communications," *Lincoln Laboratory Journal,* vol. 16, no. 1, p. 217, 2006.

[41] S. M. Sze and K. K. Ng, *Physics of semiconductor devices*. John wiley & sons, 2006.

[42] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sensors and Actuators A: Physical,* vol. 140, no. 1, pp. 103-112, 2007.

[43] F. Zappa, A. Tosi, A. Dalla Mora, and S. Tisa, "SPICE modeling of single photon avalanche diodes," *Sensors and Actuators A: Physical,* vol. 153, no. 2, pp. 197-204, 2009.

[44] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Applied optics,* vol. 35, no. 12, pp. 1956-1976, 1996.

[45] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in quenching circuits for single photon avalanche diodes," *IEEE Transactions on nuclear science,* vol. 57, no. 6, pp. 3815-3826, 2010.

[46] F. Zappa, A. Lotito, A. Giudice, S. Cova, and M. Ghioni, "Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors," *IEEE Journal of solid-state circuits,* vol. 38, no. 7, pp. 1298-1301, 2003.

[47] C. Veerappan, "Single-Photon Avalanche Diodes for Cancer Diagnosis," 2016.

[48] C. Veerappan, "Data acquisition system design for a 160x128 single-photon image sensor with on-pixel 55 ps time-to-digital converter," 2010.

[49] M. W. Fishburn, *Fundamentals of CMOS single-photon avalanche diodes*. fishburn, 2012.

[50] G. Hurkx, D. Klaassen, and M. Knuvers, "A new recombination model for device simulation including tunneling," *IEEE Transactions on electron devices,* vol. 39, no. 2, pp. 331-338, 1992.

[51] A. Rochas, "Single photon avalanche diodes in CMOS technology," Citeseer2003.

[52] A. Spinelli and A. L. Lacaita, "Physics and numerical simulation of single photon avalanche diodes," *IEEE Transactions on Electron Devices,* vol. 44, no. 11, pp. 1931-1943, 1997.

[53] R. McIntyre, P. Webb, and H. Dautet, "A short-wavelength selective reach-through avalanche photodiode," *IEEE Transactions on Nuclear Science,* vol. 43, no. 3, pp. 1341-1346, 1996.

[54] R. H. Haitz, "Mechanisms contributing to the noise pulse rate of avalanche diodes," *Journal of Applied Physics,* vol. 36, no. 10, pp. 3123-3131, 1965.

[55] P. Webb and A. Jones, "Large area reach-through avalanche diodes for radiation monitoring," *IEEE Transactions on Nuclear Science*, vol. 21, no. 1, pp. 151-158, 1974.

[56] M. Ghioni, S. Cova, A. Lacaita, and G. Ripamonti, "New silicon epitaxial avalanche diode for single-photon timing at room temperature," *Electronics letters,* vol. 24, no. 24, pp. 1476-1477, 1988.

[57] A. Lacaita, M. Ghioni, and S. Cova, "Double epitaxy improves single-photon avalanche diode performance," *Electronics letters,* vol. 25, no. 13, pp. 841-843, 1989.

[58] M. Ghioni, A. Gulinatti, P. Maccagnani, I. Rech, and S. Cova, "Planar silicon SPADs with 200-µm diameter and 35-ps photon timing resolution," in *Advanced Photon Counting Techniques*, 2006, vol. 6372, p. 63720R: International Society for Optics and Photonics.

[59] A. Lacaita, S. Cova, M. Ghioni, and F. Zappa, "Single-photon avalanche diode with ultrafast pulse response free from slow tails," *IEEE electron device letters,* vol. 14, no. 7, pp. 360-362, 1993.

[60] E. Sciacca *et al.*, "Silicon planar technology for single-photon optical detectors," *IEEE Transactions on Electron Devices,* vol. 50, no. 4, pp. 918-925, 2003.

[61] B. Aull *et al.*, "Laser radar imager based on 3D integration of Geiger-mode avalanche photodiodes with two SOI timing circuit layers," in *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International*, 2006, pp. 1179-1188: IEEE.

[62] E. Venialgo, S. Mandai, T. Gong, D. R. Schaart, and E. Charbon, "Time estimation with multichannel digital silicon photomultipliers," *Physics in Medicine & Biology,* vol. 60, no. 6, p. 2435, 2015.

[63] J. Wehner *et al.*, "MR-compatibility assessment of the first preclinical PET-MRI insert equipped with digital silicon photomultipliers," *Physics in Medicine & Biology,* vol. 60, no. 6, p. 2231, 2015.

[64] C. Bruschini *et al.*, "SPADnet: Embedded coincidence in a smart sensor network for PET applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 734, pp. 122-126, 2014.

[65] M. A. Al-Rawhani, J. Beeley, and D. R. Cumming, "Wireless fluorescence capsule for endoscopy using single photon-based detection," *Scientific reports,* vol. 5, p. 18591, 2015.

[66] X. Michalet *et al.*, "Silicon photon-counting avalanche diodes for single-molecule fluorescence spectroscopy," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 20, no. 6, pp. 248-267, 2014.

[67] D. U. Li *et al.*, "Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm," *Journal of biomedical optics,* vol. 16, no. 9, p. 096012, 2011.

[68] V. Krishnaswami, C. J. Van Noorden, E. M. Manders, and R. A. Hoebe, "Towards digital photon counting cameras for single-molecule optical nanoscopy," *Optical Nanoscopy,* vol. 3, no. 1, p. 1, 2014.

[69] C. Veerappan and E. Charbon, "CMOS SPAD based on photo-carrier diffusion achieving PDP> 40% from 440 to 580 nm at 4 V excess bias," *IEEE Photon. Technol. Lett.,* vol. 27, no. 23, pp. 2445-2448, 2015.

[70] T. Leitner *et al.*, "Measurements and simulations of low dark count rate single photon avalanche diode device in a low voltage 180-nm CMOS image sensor technology," *IEEE Transactions on Electron Devices,* vol. 60, no. 6, pp. 1982-1988, 2013.

[71] J. A. Richardson, L. A. Grant, and R. K. Henderson, "Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology," *IEEE Photonics Technology Letters,* vol. 21, no. 14, pp. 1020-1022, 2009.

[72] M. Gersbach *et al.*, "A low-noise single-photon detector implemented in a 130 nm CMOS imaging process," *Solid-State Electronics,* vol. 53, no. 7, pp. 803-808, 2009.

[73] E. A. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm," *IEEE Electron Device Letters,* vol. 33, no. 5, pp. 694-696, 2012.

[74] M. A. Karami, M. Gersbach, H.-J. Yoon, and E. Charbon, "A new single-photon avalanche diode in 90nm standard CMOS technology," *Optics express,* vol. 18, no. 21, pp. 22158-22166, 2010.

[75] I. M. Antolovic, S. Burri, R. A. Hoebe, Y. Maruyama, C. Bruschini, and E. Charbon, "Photon-counting arrays for time-resolved imaging," *Sensors,* vol. 16, no. 7, p. 1005, 2016.

[76] D. Bronzi *et al.*, "Low-noise and large-area CMOS SPADs with timing response free from slow tails," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, 2012, pp. 230-233: IEEE.

[77] J. M. Pavia, M. Scandini, S. Lindner, M. Wolf, and E. Charbon, "A 1× 400 backside-illuminated SPAD sensor with 49.7 ps resolution, 30 pJ/sample TDCs fabricated in 3D CMOS technology for near-infrared optical tomography," *IEEE Journal of Solid-State Circuits,* vol. 50, no. 10, pp. 2406-2418, 2015.

[78] S. Lindner, S. Pellegrini, Y. Henrion, B. Rae, M. Wolf, and E. Charbon, "A High-PDE, Backside-Illuminated SPAD in 65/40-nm 3D IC CMOS Pixel With Cascoded Passive Quenching and Active Recharge," *IEEE Electron Device Letters,* vol. 38, no. 11, pp. 1547-1550, 2017.

[79] S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, "A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology," *Optics express,* vol. 20, no. 6, pp. 5849-5857, 2012.

[80] E. A. Webster, L. A. Grant, and R. K. Henderson, "A high-performance single-photon avalanche diode in 130-nm CMOS imaging technology," *IEEE Electron Device Letters,* vol. 33, no. 11, pp. 1589-1591, 2012.

[81] R. M. Field, S. Realov, and K. L. Shepard, "A 100 fps, time-correlated single-photon-counting-based fluorescence-lifetime imager in 130 nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 49, no. 4, pp. 867-880, 2014.

[82] R. K. Henderson, E. A. Webster, and L. A. Grant, "A dual-junction single-photon avalanche diode in 130-nm CMOS technology," *IEEE Electron Device Letters,* vol. 34, no. 3, pp. 429-431, 2013.

[83] A. Eisele *et al.*, "185 MHz count rate, 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130 nm CMOS technology," in *Proc. Int. Image Sensor Workshop,* 2011, pp. 278-280.

[84] R. K. Henderson, E. A. Webster, R. Walker, J. A. Richardson, and L. A. Grant, "A 3× 3, 5μm pitch, 3-transistor single photon avalanche diode array with integrated 11V bias generation in 90nm CMOS technology," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, 2010, pp. 14.2. 1-14.2. 4: IEEE.

[85] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, "A 160x120 Pixel Analog-Counting Single-Photon Imager With Time-Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging," *IEEE JOURNAL OF SOLID-STATE CIRCUITS,* vol. 51, no. 1, pp. 155-167, 2016.

[86] E. Charbon, H.-J. Yoon, and Y. Maruyama, "A Geiger mode APD fabricated in standard 65nm CMOS technology," in *Electron Devices Meeting (IEDM), 2013 IEEE International*, 2013, pp. 27.5. 1-27.5. 4: IEEE.

[87] M. Buttafava, G. Boso, A. Ruggeri, A. Dalla Mora, and A. Tosi, "Time-gated single-photon detection module with 110 ps transition time and up to 80 MHz repetition rate," *Review of scientific instruments,* vol. 85, no. 8, p. 083114, 2014.

[88] G. Boso, A. Dalla Mora, A. Della Frera, and A. Tosi, "Fast-gating of single-photon avalanche diodes with 200ps transitions and 30ps timing jitter," *Sensors and Actuators A: Physical,* vol. 191, pp. 61-67, 2013.

[89] S. Burri, Y. Maruyama, X. Michalet, F. Regazzoni, C. Bruschini, and E. Charbon, "Architecture and applications of a high resolution gated SPAD image sensor," *Optics express,* vol. 22, no. 14, pp. 17573-17589, 2014.

[90] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, "Single-photon avalanche diode CMOS sensor for time-resolved fluorescence measurements," *IEEE Sensors Journal,* vol. 9, no. 9, pp. 1084-1090, 2009.

[91] E. Vilella, O. Alonso, A. Montiel, A. Vilà, and A. Diéguez, "A low-noise time-gated single-photon detector in a HV-CMOS technology for triggered imaging," *Sensors and Actuators A: Physical,* vol. 201, pp. 342-351, 2013.

[92] L. Pancheri, N. Massari, and D. Stoppa, "SPAD image sensor with analog counting pixel for time-resolved fluorescence detection," *IEEE Transactions on Electron Devices,* vol. 60, no. 10, pp. 3442-3449, 2013.

[93] F. Villa *et al.*, "SPAD smart pixel for time-of-flight and time-correlated single-photon counting measurements," *IEEE Photonics Journal*, vol. 4, no. 3, pp. 795-804, 2012.

[94] Z. Li and M. J. Deen, "Towards a portable Raman spectrometer using a concave grating and a time-gated CMOS SPAD," *Optics express*, vol. 22, no. 15, pp. 18736-18747, 2014.

[95] I. Nissinen *et al.*, "A sub-ns time-gated CMOS single photon avalanche diode detector for Raman spectroscopy," in *Solid-State Device Research Conference (ESSDERC), 2011 Proceedings of the European*, 2011, pp. 375-378: IEEE.

[96] Y. Maruyama and E. Charbon, "An all-digital, time-gated 128X128 spad array for on-chip, filter-less fluorescence detection," in *Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS), 2011 16th International*, 2011, pp. 1180-1183: IEEE.

[97] C. Niclass *et al.*, "Design and characterization of a 256x64-pixel single-photon imager in CMOS for a MEMS-based laser scanning time-of-flight sensor," *Optics Express*, vol. 20, no. 11, pp. 11863-11881, 2012.

[98] L. Parmesan *et al.*, "A 256× 256 SPAD array with in-pixel time to amplitude conversion for fluorescence lifetime imaging microscopy," in *International Image Sensor Workshop, Vaals, Netherlands, Memory*, 2015, vol. 900, p. M5.

[99] E. Vilella and A. Diéguez, "A gated single-photon avalanche diode array fabricated in a conventional CMOS process for triggered systems," *Sensors and Actuators A: Physical*, vol. 186, pp. 163-168, 2012.

[100] I. Vornicu, R. Carmona-Galán, and Á. Rodríguez-Vázquez, "A CMOS 0.18 μm 64× 64 single photon image sensor with in-pixel 11b time-to-digital converter," in *Semiconductor Conference (CAS), 2014 International*, 2014, pp. 131-134: IEEE.

[101] D. Portaluppi, E. Conca, and F. Villa, "32× 32 CMOS SPAD Imager for Gated Imaging, Photon Timing, and Photon Coincidence," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 2, pp. 1-6, 2018.

[102] Y. Maruyama, J. Blacksberg, and E. Charbon, "A 1024 X 8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 179-189, 2014.

[103] D. Bronzi *et al.*, "100 000 frames/s 64× 32 single-photon detector array for 2-D imaging and 3-D ranging," *IEEE journal of selected topics in quantum electronics,* vol. 20, no. 6, pp. 354-363, 2014.

[104] I. Nissinen, J. Nissinen, P. Keränen, A.-K. Länsman, J. Holma, and J. Kostamovaara, "A $2\times (4)\times 128$ Multitime-Gated SPAD Line Detector for Pulsed Raman Spectroscopy," *IEEE Sensors Journal,* vol. 15, no. 3, pp. 1358-1365, 2015.

[105] F. Guerrieri, S. Tisa, A. Tosi, and F. Zappa, "Two-dimensional SPAD imaging camera for photon counting," *IEEE Photonics Journal,* vol. 2, no. 5, pp. 759-774, 2010.

[106] M. Buttafava *et al.*, "A Compact Two-Wavelength Time-Domain NIRS System Based on SiPM and Pulsed Diode Lasers," *IEEE Photonics Journal,* vol. 9, no. 1, pp. 1-14, 2017.

[107] B. R. Rae *et al.*, "A CMOS time-resolved fluorescence lifetime analysis micro-system," *Sensors,* vol. 9, no. 11, pp. 9255-9274, 2009.

[108] B. Rae *et al.*, "CMOS driven micro-pixel LEDs integrated with single photon avalanche diodes for time resolved fluorescence measurements," *Journal of Physics D: Applied Physics,* vol. 41, no. 9, p. 094011, 2008.

[109] A. Dalla Mora *et al.*, "Towards next-generation time-domain diffuse optics for extreme depth penetration and sensitivity," *Biomedical optics express,* vol. 6, no. 5, pp. 1749-1760, 2015.

[110] J.-K. Choi, J.-M. Kim, G. Hwang, J. Yang, M.-G. Choi, and H.-M. Bae, "Time-divided spread-spectrum code-based 400 fW-detectable multichannel fNIRS IC for portable functional brain imaging," *IEEE Journal of Solid-State Circuits,* vol. 51, no. 2, pp. 484-495, 2016.

[111] T. Lunghi, E. Pomarico, C. Barreiro, D. Stucki, B. Sanguinetti, and H. Zbinden, "Advantages of gated silicon single-photon detectors," *Applied optics,* vol. 51, no. 35, pp. 8455-8459, 2012.

[112] G. Boso, M. Buttafava, F. Villa, and A. Tosi, "Low-cost and compact single-photon counter based on a CMOS SPAD smart pixel," *IEEE Photonics Technology Letters,* vol. 27, no. 23, pp. 2504-2507, 2015.

[113] D. Mathine, R. Droopad, and G. Maracas, "A vertical-cavity surface-emitting laser appliqued to a 0.8-µm NMOS driver," *IEEE Photonics Technology Letters*, vol. 9, no. 7, pp. 869-871, 1997.

[114] A. Krishnamoorthy *et al.*, "Vertical-cavity surface-emitting lasers flip-chip bonded to gigabit-per-second CMOS circuits," *IEEE Photonics Technology Letters*, vol. 11, no. 1, pp. 128-130, 1999.

[115] A. Yodh and B. Chance, "Spectroscopy and imaging with diffusing light," *Physics Today*, vol. 48, no. 3, pp. 34-41, 1995.

[116] R. H. Hadfield, "Single-photon detectors for optical quantum information applications," *Nature photonics*, vol. 3, no. 12, pp. 696-705, 2009.

[117] S. Saha, F. Lesage, and M. Sawan, "Time-resolved reflectance using short source-detector separation," in *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*, 2016, pp. 333-336: IEEE.

[118] A. Dalla Mora *et al.*, "Fast-gated single-photon avalanche diode for wide dynamic range near infrared spectroscopy," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 4, pp. 1023-1030, 2010.

[119] R. Re, E. Martinenghi, A. Dalla Mora, D. Contini, A. Pifferi, and A. Torricelli, "Probe-hosted silicon photomultipliers for time-domain functional near-infrared spectroscopy: Phantom and in vivo tests," *Neurophotonics*, vol. 3, no. 4, pp. 045004-045004, 2016.

[120] *ID Quantique ID100 Datasheet. [Online].* , Available: http:// www.idquantique.com/photon-counting/photon-counting-modules/id100-silicon-apd-single-photon-detector.html , accessed Aug. 6, 2017.

[121] M. Maymandi-Nejad and M. Sachdev, "A digitally programmable delay element: design and analysis," *IEEE transactions on very large scale integration (VLSI) systems*, vol. 11, no. 5, pp. 871-878, 2003.

[122] S. Nitta *et al.*, "A variable delay generator for DESKEW IC using ECL gate array," in *Symp. VLSI Circuits, Oiso, Japan*, 1991, pp. 55-56.

[123] G. Boso, A. Tosi, A. D. Mora, and F. Zappa, "High-throughput gated photon counter with two detection windows programmable down to 70 ps width," *Review of Scientific Instruments*, vol. 85, no. 1, p. 013107, 2014.

[124] P. Carniti, M. De Matteis, A. Giachero, C. Gotti, M. Maino, and G. Pessina, "CLARO-CMOS, a very low power ASIC for fast photon counting with pixellated photodetectors," *Journal of Instrumentation,* vol. 7, no. 11, p. P11026, 2012.

[125] R. Pelliconi, D. Iezzi, A. Baroni, M. Pasotti, and P. L. Rolandi, "Power efficient charge pump in deep submicron standard CMOS technology," *IEEE Journal of Solid-State Circuits,* vol. 38, no. 6, pp. 1068-1071, 2003.

[126] S. Saha, F. Lesage, and M. Sawan, "High-voltage pulse generator with variable delay for ultrafast gating of single photon detector," in *Circuits & Systems (LASCAS), 2016 IEEE 7th Latin American Symposium on*, 2016, pp. 131-134: IEEE.

[127] R. Chebli and M. Sawan, "Fully integrated high-voltage front-end interface for ultrasonic sensing applications," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 54, no. 1, pp. 179-190, 2007.

[128] R. J. Cooper *et al.*, "MONSTIR II: a 32-channel, multispectral, time-resolved optical tomography system for neonatal brain imaging," *Review of Scientific Instruments,* vol. 85, no. 5, p. 053105, 2014.

[129] R. Re *et al.*, "Multi-channel medical device for time domain functional near infrared spectroscopy based on wavelength space multiplexing," *Biomedical optics express,* vol. 4, no. 10, pp. 2231-2246, 2013.

[130] L. Di Sieno *et al.*, "Time-domain diffuse optical tomography using silicon photomultipliers: feasibility study," *Journal of biomedical optics,* vol. 21, no. 11, p. 116002, 2016.

[131] E. Martinenghi *et al.*, "Spectrally resolved single-photon timing of silicon photomultipliers for time-domain diffuse spectroscopy," *IEEE Photonics Journal,* vol. 7, no. 4, pp. 1-12, 2015.

[132] A. Dalla Mora *et al.*, "Fast silicon photomultiplier improves signal harvesting and reduces complexity in time-domain diffuse optics," *Optics Express,* vol. 23, no. 11, pp. 13937-13946, 2015.

[133] A. Farina *et al.*, "Time-Domain Functional Diffuse Optical Tomography System Based on Fiber-Free Silicon Photomultipliers," *Applied Sciences,* vol. 7, no. 12, p. 1235, 2017.

[134] S.-i. Fujisaka *et al.*, "A clinical tissue oximeter using NIR time-resolved spectroscopy," in *Oxygen Transport to Tissue XXXVII*: Springer, 2016, pp. 427-433.

[135] E. Ferocino, E. Martinenghi, A. Dalla Mora, A. Pifferi, R. Cubeddu, and P. Taroni, "High throughput detection chain for time domain optical mammography," *Biomedical optics express,* vol. 9, no. 2, pp. 755-770, 2018.

[136] S. Saha, Y. Lu, S. Weyers, M. Sawan, and F. Lesage, "Compact Fast Optode-based Probe for Single-Photon Counting Applications," *IEEE Photonics Technology Letters,* 2018.

[137] L. Di Sieno *et al.*, "Miniaturized pulsed laser source for time-domain diffuse optics routes to wearable devices," *Journal of biomedical optics,* vol. 22, no. 8, p. 085004, 2017.

[138] K. Oi *et al.*, "Development of new 2.5 D package with novel integrated organic interposer substrate with ultra-fine wiring and high density bumps," in *Electronic Components and Technology Conference (ECTC), 2014 IEEE 64th*, 2014, pp. 348-353: IEEE.

[139] X. Zhang *et al.*, "Heterogeneous 2.5 D integration on through silicon interposer," *Applied Physics Reviews,* vol. 2, no. 2, p. 021308, 2015.

[140] V. Sundaram *et al.*, "Low cost, high performance, and high reliability 2.5 D silicon interposer," in *Electronic Components and Technology Conference (ECTC), 2013 IEEE 63rd*, 2013, pp. 342-347: IEEE.

[141] P. TRIGILIO, "Development of an ASIC for SiPM readout in SPECT applications," 2016.

[142] K. Chen, H.-S. Lee, A. P. Chandrakasan, and C. G. Sodini, "Ultrasonic imaging transceiver design for CMUT: A three-level 30-Vpp pulse-shaping pulser with improved efficiency and a noise-optimized receiver," *IEEE Journal of Solid-State Circuits,* vol. 48, no. 11, pp. 2734-2745, 2013.

[143] K. Phang, "CMOS optical preamplifier design using graphical circuit analysis," *University of Toronto,* 2001.

[144] S. Saha, Y. Lu, S. Weyers, M. Sawan, and F. Lesage, "Compact Fast Optode-Based Probe for Single-Photon Counting Applications," *IEEE Photonics Technology Letters,* vol. 30, no. 17, pp. 1515-1518, 2018.

[145] M.-J. Lee *et al.*, "High-Performance Back-Illuminated Three-Dimensional Stacked Single-Photon Avalanche Diode Implemented in 45-nm CMOS Technology," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 24, no. 6, pp. 1-9, 2018.

APPENDIX A – CHIP PACKAGING EQUIPMENTS

CMOS chip wirebonding, packaging, and microwell fabrication are carried out in LASEM laboratory in Polytechnique Montreal. The equipments are listed as follows.

1) **Wirebonder**: To wirebond the loose dies to the packages, a wedge wirebonder machine was used. The wirebonder uses aluminum wires of 17.5 µm diameter with the minimum pad pitch of 80 µm.

2) **Oxford Lasers, A-SERIES**: This laser is a micromachining system for etching, microdrilling, and microcutting different materials such as glass, metal, PMMA, and polymers. The laser power is 50 mW with 355 nm wavelength.

3) **Flip-chip bonder**: The Femto Flip-chip bonder is a hybrid module assembly system used in manufacturing operations for precision placement of electronics and microelectronics components. The resolution is 0.5 µm, and the bonding force varies between 0.2 and 10 N.

4) **Precision spin coater:** The spin coater is used for preparing the PDMS layers for microwell fabrication. The spinning speed of the machine varies between 0 to 6000 rpm. Different thicknesses can be obtained by changing the spinning speed and its duration.

5) **Standard lab oven:** To cure the PDMS samples, a standard oven with the temperature ranging from 50 to 225 degrees was used.