

Single-Photon Image Sensors in CMOS: Picosecond Resolution for Three-Dimensional Imaging

THÈSE N° 4161 (2008)

PRÉSENTÉE LE 29 JUILLET 2008

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE D'ARCHITECTURE DE PROCESSEURS

PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Cristiano L. NICLASS

ingénieur en microtechnique diplômé EPF
de nationalité suisse et originaire de Lully (FR)

acceptée sur proposition du jury:

Prof. G. De Micheli, président du jury
Prof. E. Charbon, directeur de thèse
Prof. R. S. Popovic, rapporteur
Prof. P. Seitz, rapporteur
Dr A. Rochas, rapporteur



TABLE OF CONTENTS

ABSTRACT	v
VERSION ABREGEED	vii
1. INTRODUCTION	1
1.1 GENERAL CLASSIFICATION OF OPTICAL 3D IMAGE SENSORS	2
1.2 SOLID-STATE 3D IMAGE SENSORS: STATE-OF-THE-ART	6
1.2.1 <i>TOF based on Phase of First Harmonic</i>	10
1.2.2 <i>Direct Time-of-Flight Principle</i>	15
1.2.3 <i>Indirect Time-of-Flight Principle</i>	17
1.2.4 <i>Time-of-Flight based on Pseudo-Noise Sequences</i>	21
1.2.5 <i>Time-of-Flight based on Time-Correlated Single-Photon Counting</i>	22
1.3 APPLICATIONS	24
1.3.1 <i>Automotive Applications</i>	24
1.3.2 <i>Human-Computer Interfaces</i>	26
1.3.3 <i>Machine Vision</i>	28
1.3.4 <i>Other Applications</i>	30
1.4 THESIS MOTIVATIONS AND CHALLENGES	32
1.5 ORGANIZATION OF THE THESIS	35
2. SINGLE-PHOTON AVALANCHE DIODES IN CMOS TECHNOLOGY	37
2.1 INTRODUCTION	37
2.2 PERFORMANCE PARAMETERS	40
2.2.1 <i>Photon Detection Probability</i>	40
2.2.2 <i>Dark Count Rate</i>	42
2.2.3 <i>Dead Time</i>	43
2.2.4 <i>Afterpulsing Probability</i>	44
2.2.5 <i>Timing Resolution</i>	45
2.2.6 <i>Crosstalk Probability</i>	46
2.3 DEVICE ENGINEERING	48
2.3.1 <i>State-of-the-Art</i>	48
2.3.2 <i>Deep-submicron CMOS Technology</i>	58
2.4 FRONT-END CIRCUITS	59
2.4.1 <i>Passive Quenching and Recharge based on Single MOS Transistors</i>	62
2.4.2 <i>Dual-Threshold Quenching and Recharge Circuit</i>	71
2.4.3 <i>Event-Driven Readout Circuit</i>	76
2.5 SPAD INTEGRATION IN 0.8-MICRON CMOS	80
2.5.1 <i>Photon Detection Probability</i>	83
2.5.2 <i>Dark Count Rate</i>	85
2.5.3 <i>Timing Resolution</i>	87

2.5.4	<i>Afterpulse Probability</i>	88
2.5.5	<i>Crosstalk Probability</i>	89
2.6	SPAD INTEGRATION IN 0.35-MICRON CMOS.....	91
2.6.1	<i>Photon Detection Probability</i>	92
2.6.2	<i>Dark Count Rate</i>	93
2.6.3	<i>Timing Resolution</i>	95
2.6.4	<i>Afterpulse Probability</i>	96
2.6.5	<i>Crosstalk Probability</i>	97
2.7	SPAD INTEGRATION IN 130-NM CMOS.....	98
2.7.1	<i>Photon Detection Probability</i>	102
2.7.2	<i>Dark Count Rate</i>	104
2.7.3	<i>Timing Resolution</i>	105
2.8	CONCLUSION	106
3.	OPTICAL TIME-OF-FLIGHT 3D IMAGING BASED ON TCSPC.....	111
3.1	INTRODUCTION	111
3.2	TCSPC PRINCIPLE AND ASSUMPTIONS	112
3.3	MODELING OF PERFORMANCE BASED ON TCSPC	117
3.3.1	<i>Analytical Model</i>	121
3.4	MODELING OF ENVIRONMENT CONDITIONS AND OPTICAL COMPONENTS	132
3.4.1	<i>Background Light</i>	132
3.4.2	<i>Illumination</i>	134
3.4.3	<i>Optical Band-pass Filter</i>	135
3.4.4	<i>Imaging Lens</i>	136
3.4.5	<i>Target Reflectivity</i>	136
3.4.6	<i>Single-Photon Image Sensor</i>	138
3.4.7	<i>Signal and Background Count Rates</i>	138
3.5	FULLY-INTEGRATED TCSPC 3D IMAGE SENSOR DESIGN	141
3.5.1	<i>Imager Architecture</i>	141
3.5.2	<i>Pixel Circuit</i>	142
3.5.3	<i>Time-to-Digital Converter</i>	144
3.5.4	<i>Readout Circuit</i>	149
3.5.5	<i>Imager Characterization</i>	150
3.6	EXPERIMENTAL RESULTS IN SHORT RANGE	154
3.7	EXPERIMENTAL RESULTS IN MEDIUM RANGE	165
3.7.1	<i>Illumination System</i>	165
3.7.2	<i>Range-finding Evaluation</i>	168
3.8	CONCLUSION	177
4.	OPTICAL TIME-OF-FLIGHT 3D IMAGING BASED ON SPSD.....	181
4.1	INTRODUCTION	181
4.2	SPSD PRINCIPLE AND ASSUMPTIONS	181
4.3	MODELING OF PERFORMANCE BASED ON SPSD	187
4.3.1	<i>Distance Resolution</i>	188
4.3.2	<i>Non-Linearity Errors</i>	189

4.3.3	<i>Dynamic Range</i>	192
4.4	FULLY-INTEGRATED SPSD 3D IMAGE SENSOR DESIGN	193
4.4.1	<i>Imager Architecture</i>	193
4.4.2	<i>Pixel Circuit</i>	193
4.4.3	<i>Readout Circuit</i>	196
4.5	3D CAMERA PROTOTYPE	196
4.5.1	<i>Image Sensor</i>	196
4.5.2	<i>Illumination System Design</i>	198
4.5.3	<i>Camera Optics</i>	198
4.6	EXPERIMENTAL RESULTS	198
4.6.1	<i>Single-Photon Image Sensor</i>	199
4.6.2	<i>Illumination Characterization</i>	200
4.6.3	<i>Range-Finding Performance</i>	203
4.7	CONCLUSION	208
5.	CONCLUSION	211
5.1	MAIN ACHIEVEMENTS	211
5.2	OUTLOOK FOR IMPROVEMENTS	214
APPENDIX A	217
REFERENCES	223
LIST OF ACRONYMS	237
LIST OF SYMBOLS		239
ACKNOWLEDGEMENTS		243
ABOUT THE AUTHOR		247
LIST OF PUBLICATIONS		248

ABSTRACT

Among advanced electronic designs, vision systems are perhaps those that have undergone the most dramatic technological revolutions of the last decades. Recently, a new family of vision systems has emerged aimed at three-dimensional (3D) imaging. Many applications exist or will soon come to existence that can use 3D image sensors and the image sensor community is responding with conventional and unconventional solutions.

This thesis presents a new concept for imaging and, in particular, for 3D imaging, based on single-photon detection. In our approach, single-photon detection is performed by a device known as single-photon avalanche diode (SPAD). Large arrays of SPADs were demonstrated for the first time in this thesis and time resolutions consistently in the picosecond range were achieved. It has thus become feasible to design solid-state 3D imagers with millimeter depth perception based on the time-of-flight principle.

Initially, large arrays of SPAD devices and associated circuitry have been investigated in a $0.8\mu\text{m}$ CMOS technology. Extremely low-noise detectors were implemented in this technology. In arrays of 64×48 pixels, the median dark count rate was 0.5Hz per square micrometer of active area. For these devices, the maximum photon detection probability was 26% at 550nm optical wavelength. Subsequently, a technology migration of SPADs towards deep-submicron CMOS has been successfully achieved. For the first time, SPADs in $0.35\mu\text{m}$ and $0.13\mu\text{m}$ CMOS technologies were designed and characterized. Outstanding photon detection probabilities up to 35.4% and 34.5% have been measured in devices manufactured in $0.35\mu\text{m}$ and $0.13\mu\text{m}$ CMOS, respectively. Median dark count rate has increased with technology scaling. The median dark count rate over an array of 128×128 pixels was $18\text{ Hz}/\mu\text{m}^2$ in $0.35\mu\text{m}$ CMOS, whereas it was $385\text{Hz}/\mu\text{m}^2$ in $0.13\mu\text{m}$ CMOS for a single device.

With the aim of enabling high-performance system-on-a-chip implementations using single-photon detectors, appropriate front-end and ancillary circuits utilized in the detection, measurement, and storage of time-of-flight evaluations have been introduced. For the first time, passive quenching circuits using a single MOS transistor were investigated. Based on the proposed biasing regime, the single MOS quenching circuit provides well defined dead time and leads to high pixel fill factor. Active recharge is achieved in this thesis by means of a new

| Abstract

dual-threshold quenching and recharge circuit. This circuit allows the implementation of a hold-off time using low in-pixel transistor count, thus improving afterpulsing probability with little or no penalty in fill factor.

3D image sensors based on a pulsed detection technique known as time-correlated single-photon counting (TCSPC) have been investigated. Moreover, a technique has also been investigated for range imaging using SPADs that operate with continuously modulated optical signals. This technique, known as single-photon synchronous detection (SPSD), was invented in the course of this thesis and it is introduced here for the first time.

In order to demonstrate the potential of TCSPC in solid-state 3D imaging, a complete theoretical and experimental investigation was conducted. An analytical model for the evaluation of ranging performance based on TCSPC was introduced. A fully-integrated TCSPC system for single-photon time-of-flight evaluation was implemented in CMOS, for the first time. Depth maps of 3D scenes were achieved at millimeter precisions using only 1mW of laser power and an integration time of 50ms. Thanks to SPAD technology, accurate distance measurements are now possible even with extremely low photon count rates of a few thousand per second. The maximum non-linearity in distance measurement was 9 millimeters over the full measurement range. Time-varying uncertainty (1σ) at the farthest distance was 5.2 millimeters.

The SPSD approach has been theoretically and experimentally investigated. The design of the first fully-parallel implementation of a single-photon image sensor in CMOS has been introduced. The sensor has enabled the acquisition of real-time 3D images, based on CMOS SPADs, for the first time. A 3D camera prototype was designed and built based on the SPSD image sensor with a field-of-view of 50°. Experimental results showed that the SPSD rangefinder is effective. Distance measurement performance was characterized with a maximum non-linearity error of 11cm within a range of a few meters. In the same range, the maximum repeatability error was 3.8cm.

Keywords: Solid-state 3D imaging, 3D image sensor, 3D camera, flash camera, time-of-flight, rangefinder, range imaging, depth sensor, single-photon detector, avalanche photodiode, SPAD, GAPD, APD, TCSPC, SPSD.

VERSION ABRÉGÉE

Parmi les designs électroniques avancés, les systèmes de vision sont peut-être ceux qui ont connu l'évolution technologique la plus spectaculaire durant les dernières décennies. Récemment, une nouvelle classe de systèmes de vision permettant l'imagerie tridimensionnelle (3D) est apparue. Un grand nombre d'applications actuelles ou futures reposent sur ces capteurs d'images 3D, auxquelles la communauté scientifique et l'industrie dans ce domaine répondent avec des solutions conventionnelles et non-conventionnelles. Cette thèse présente un nouveau concept d'imagerie, en particulier d'imagerie 3D, basée sur la détection de photons individuels. Dans notre approche, la détection de photons individuels est assurée par un type de photodiode à avalanche plus connu sous l'acronyme de SPAD (*single-photon avalanche diode*). Pour la première fois, de grandes matrices de SPADs ont été réalisées. La caractérisation de ces dispositifs a systématiquement montré des résolutions temporelles de l'ordre de quelques dizaines de picosecondes. La conception de capteurs d'images 3D intégrés avec une résolution en profondeur millimétrique est par conséquent devenue possible en utilisant la mesure du temps de vol de la lumière.

Initialement, la faisabilité de matrices de SPADs ainsi que les circuits d'interfaces associés ont été examinés dans un procédé de fabrication CMOS 0.8 μm . Des photodétecteurs à bruit électronique très faible ont été réalisés dans cette technologie. Dans des matrices de 64x48 pixels, le taux médian de fausses détections était de 0.5Hz par micromètre carré de surface active. Pour ces dispositifs, la probabilité de détection photonique maximale était de 26% à 550nm de longueur d'onde. Ces SPADs ont ensuite été transposés dans des technologies CMOS avancées avec succès. Pour la première fois, des SPADs ont été conçus, réalisés, et caractérisés dans des procédés CMOS 0.35 μm et 0.13 μm . La probabilité de détection photonique maximale atteint 35.4% et 34.5% dans les dispositifs fabriqué en CMOS 0.35 μm et 0.13 μm respectivement. Le taux de fausses détections a progressé avec l'évolution technologique. Le taux médian de fausses détections sur une matrice de 128x128 pixels était de 18Hz/ μm^2 en CMOS 0.35 μm , alors qu'il était de 385Hz/ μm^2 en 0.13 μm pour un seul dispositif. Afin de permettre la réalisation systèmes intégrés monolithiques performants à partir de photodétecteurs à photon individuels, des circuits d'interface et auxiliaires utilisés dans la détection, la mesure proprement dite, et la mémorisation de « temps-de-vol » des photons ont été introduits. Pour la

première fois, les circuits d'extinction d'avalanche à un seul transistor MOS ont été étudiés. A l'aide du mode de polarisation proposé, le circuit à un seul transistor MOS induit un temps mort de détection bien défini et un taux de surface active très élevé. De nouveaux circuits de recharge active sont proposés dans cette thèse, notamment par le biais d'un nouveau circuit d'interface à double-seuil. Ce circuit permet l'implémentation d'un temps de récupération à partir d'un nombre très faible de transistors, diminuant par conséquent la probabilité de redéclenchement parasite d'avalanche, sans pénaliser le taux de surface active. Des capteurs d'image 3D basés sur une technique de mesure optique pulsée, appelée comptage temporellement corrélé de photons individuels (TCSPC), ont été investigues. En outre, une nouvelle technique qui utilise des signaux optiques modulés continûment, appelée détection synchrone de photons individuels (SPSD), a été inventée et est introduite dans cette thèse pour la première fois.

Afin de démontrer le potentiel de la technique TCSPC dans l'imagerie 3D intégrée, une investigation théorique et expérimentale complète a été conduite. Un modèle analytique pour l'évaluation de la performance du mode TCSPC a été créé. Un système TCSPC complètement intégré a été réalisé pour la première fois. La mesure tridimensionnelle de grands volumes a été effectuée avec des précisions millimétriques. Une vérification quantitative de performance de la mesure en distance a été opérée avec une puissance optique de 1mW et avec un temps d'acquisition de 50 millisecondes. Grace à la technologie de détection photons individuels, des mesures de distance précises sont désormais possible, même dans des conditions où le taux de comptage des photons disponibles n'est que de quelques milliers de photons par seconde. La déviation maximale de la linéarité était de 9 millimètres sur toute la longueur de mesure. L'erreur de répétabilité (1σ) à la distance la plus éloignée était de 5.2 millimètres.

L'approche SPSD a été théoriquement et expérimentalement investiguée. La conception et la réalisation du premier capteur d'image à photons individuels fonctionnant en temps réel ont été présentées. Le capteur a permis la première validation de l'acquisition en temps réels d'images 3D, basée sur les SPADs en CMOS. Un prototype de caméra 3D SPSD a été implémenté avec un angle de vue de 50°. Des résultats expérimentaux ont montré que la technique SPSD est efficace. Dans une plage de quelques mètres, l'erreur de non-linéarité maximale était de 11 cm et l'erreur statistique maximale de 3.8 cm.

Mots clefs: Imagerie 3D intégrée, capteur d'image 3D, caméra 3D, temps-de-vol, télémètre optique, capteur optique de profondeur, détecteur à photon individuel, photodiode à avalanche, SPAD, GAPD, APD, TCSPC, SPSD.

1. INTRODUCTION

Originally based on photogrammetry, vision systems for capturing and rendering three-dimensional (3D) images have existed since the dawn of photography. Despite considerable technological progress, especially in rendering techniques, 3D vision has been relegated to research, particularly in robotics, and in relatively low-volume applications. Recently, the appearance of low-cost digital still and video cameras has reversed this trend and has accelerated the push to create low-cost and compact 3D native vision systems. Currently, a number of novel and potentially high volume applications requiring fast and precise depth map evaluation are being devised and the market share of software for capturing, processing, and rendering 3D images has grown exponentially. More recent applications include a number of automotive uses such as pedestrian detection, volumetric rear vision for parking assistance and automation, blind spot detection for safe lane change, lane departure warning, and adaptive cruise control, to name a few. Non-automotive applications include virtual human-computer interfaces providing immersive experience in gaming devices as well as in computer-aided design software, biometric access control, object and person monitoring, land and sea surveying, non-ionizing medical tomographic imagers, unmanned ground and aerial vehicles, etc. Another class of applications in which 3D vision may provide major contributions is in manufacturing environments. For example, for factory automation, quality inspection, alarm and emergency stops in environments where men and robots cooperate as a team, etc. Cost concerns have prompted the emergence of a new generation of solid-state depth imagers that can achieve the desired accuracy and speed in compact implementations.

This thesis explains the theory, design, modeling, and characterization of 3D image sensors based on large arrays of single-photon detectors manufactured in complementary metal-oxide semiconductor (CMOS) technologies. At pixel level, single-photon detection is enabled by a particular photodetector known as single-photon avalanche diode (SPAD). Unlike linear avalanche photodiodes (APDs), SPADs are biased above the breakdown voltage in the so-called Geiger mode. Thanks to their excellent time resolution in detecting arrival time of photons, SPADs enable the realization of high-performance solid-state 3D image sensors based on the time-of-flight principle, as investigated in this thesis. Appropriate front-end and ancillary circuits utilized in the detection,

measurement, and storage of time-of-flight evaluations are integrated on a single CMOS chip, thus leading to system-on-a-chip implementations.

3D image sensors based on a pulsed detection technique known as time-correlated single-photon counting (TCSPC) are investigated. Furthermore, in order to diversify range imaging schemes based on single-photon detectors to operate with continuously modulated optical signals, a detection technique called single-photon synchronous detection (SPSD) is introduced in this thesis for the first time.

In this introductory Chapter, the state-of-the-art in solid-state scannerless 3D image sensors is described. In order to position this type of sensors in the general context of all optical techniques used to acquire 3D images, a succinct classification of optical 3D image sensors is firstly described in Section 1.1. In the general classification, a critical review of the techniques that are not directly relevant to this thesis is summarized. In Section 1.2, the focus is given to scannerless solid-state 3D image sensing based on the time-of-flight principle, which is the most relevant method in this thesis. A detailed description of the state-of-the-art is therefore provided in that Section, following a narrower classification. The most important applications that may directly benefit from the results of this work are listed and described in Section 1.3. Motivations and challenges for this thesis are described in Section 1.4. However, since the description of operating principles of SPAD as well as its current state-of-the-art is presented in Chapter 2, the reader is directed to that Chapter for a deeper understanding of the challenges involved in this work. Finally, Section 1.5 describes the organization of the remaining Chapters of this dissertation.

1.1 GENERAL CLASSIFICATION OF OPTICAL 3D IMAGE SENSORS

The acquisition of three-dimensional images from arbitrary scenes is enabled by a relatively high number of optical techniques. Several authors have proposed different classifications for rangefinders depending on the physical principle involved. For example, in [1], Besl distinguishes six basic principles, namely (i) radar, (ii) triangulation, (iii) moiré technology, (iv) holographic interferometry, (v) focusing, and (vi) Fresnel diffraction. A different classification was proposed in [2] that takes advantages of only three pairs of criteria, that is, (a) active vs. passive, (b) monocular vs. multiple view, and (c) image base vs. direct sensing. Figure 1.1 shows the classification of optical range-finding techniques based on the proposed scheme. Note that stereoscopy

and photogrammetry may be assumed to be a passive implementation of the triangulation method. However, these techniques are intentionally distinguished from triangulation in the classification due to their passive nature and because they had been discovered many decades before the first implementations of active triangulation methods.

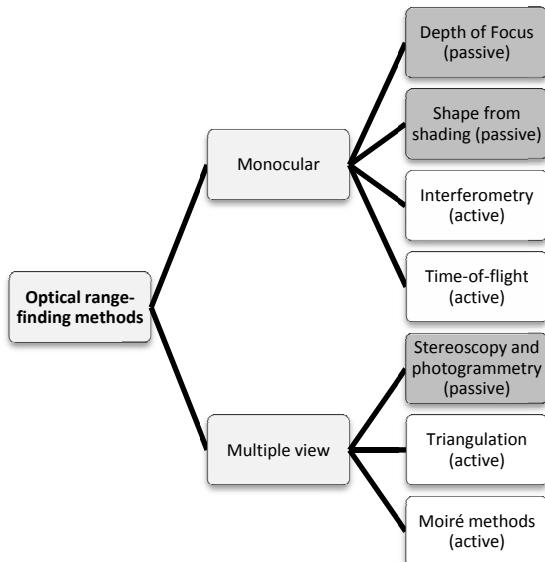


Figure 1.1. Classification of optical range-finding techniques according to the classification of Jarvis in [2].

When evaluating different range-finding techniques, some general criteria such as performance and constraints should be considered. Common performance criteria include accuracy, repeatability, resolution, acquisition time, measuring range, and field of view. Constraints may be cost, size, weight, robustness, operating condition, and eye-safety hazard level.

Passive methods provide an important advantage over active methods: non-invasive depth measurements. Since no active illumination is required, it also simplifies the sensor design, thereby reducing costs. However, passive methods cannot be used in featureless scenes, such as planar surfaces. This limitation relegates passive range-finding image sensors to some specific applications in which assumptions are made with respect to the objects in the scene.

Interferometer-based depth sensors provide the highest absolute accuracy and resolution. However, their use is limited to extremely short ranges, typically not larger than a micrometer. Recent work in the field [3] has extended the measurement range up to 200 millimeters by means of multiple-wavelength interferometry. Nonetheless, this technique is unsuitable for most applications, where depth imaging of macroscopic objects is required.

Active triangulation methods are well suited to range-finding, profiling and 3D imaging over a range of a few meters. For longer ranges, the resolution, which is proportional to distance in this technique, tends to get too coarse. In laser-based systems, the power required for generating the structured lighting may be beyond the relevant eye-safety regulations due to its typical collimated form. The dynamic range of an active triangulation system is also limited by the depth-of-field for illuminating the projecting and viewing optics, the triangulation geometry and field-of-view. Moreover, in order to project light patterns on the scene, a mechanical mechanism is required, typically based on a scanner, unless LCD-based video projectors are utilized. Mechanical mechanisms are very sensitive to vibrations, thus making their utilization difficult in automotive and a number of other applications. When LCD projectors are utilized, the time required to generate a number of light patterns reduces severely the acquisition speed. Furthermore, all triangulation-based methods suffer from occlusion, i.e. the target surface must be visible from two viewpoints simultaneously.

Moiré methods have found a use in special cases where smooth surfaces must be measured at high resolution, e.g. in [4]. However, their use is usually restricted to environments with limited dynamic range. An example of typical application of this technique is in manufacturing for quality inspection of flat parts, where a well-controlled measurement environment is typically affordable.

Taking into account the limitations of previously mentioned techniques and the broad nature of the applications of interest in this thesis, the choice on the range-finding techniques reduces to a single class of devices, known as time-of-flight (TOF) rangefinders. The optical TOF technique is based on active illumination. The time taken by the light beam to travel from the source to objects in the scene, them back to a photodetector, is accurately measured. TOF is then converted to distance z as

$$z = \frac{c}{2} TOF, \quad (1.1)$$

where c is the speed of light. Despite its simplicity, optical TOF became practical only in end of the 60's [5] due to stringent speed requirements on the light source, photodetector, and associated circuitry.

The optical TOF technique may be further divided into sub-classes depending on its imaging capability. In Figure 1.2, the sub-classes of TOF-based range sensors are schematically shown. Until recently, TOF rangefinders were either limited to single-point measurements or they inevitably required a mechanical scanner device in linear sensors and imaging applications.

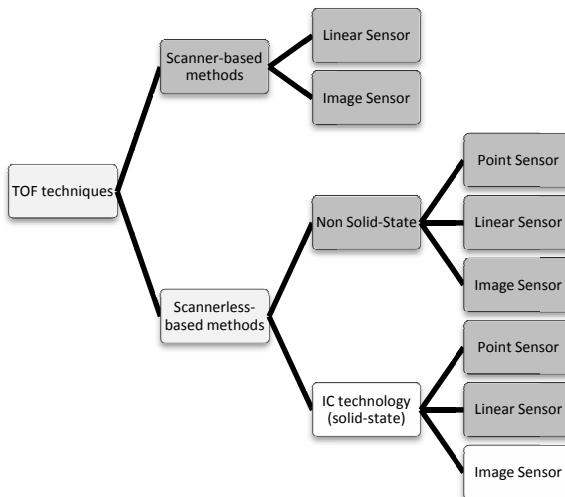


Figure 1.2. Classification of optical TOF range-finding techniques.

As previously mentioned, scanner-based depth sensors have limited use in a number of applications due to their sensitivity to mechanical vibrations. Moreover, mechanical devices are somewhat cumbersome and have limited potential in low-cost applications due to the required precision involved. Scanner-based TOF sensors utilize a collimated laser beam and therefore have restricted use due to its high level of eye-safety hazard. Furthermore, due to the scanning scheme, the acquisition time of a full image based on these devices is unacceptable in many applications. For example, a scanner-based sensor that spends 1 millisecond per point would require over 1 minute acquiring a single image of 256x256 pixels. Even when point acquisition times are reduced down to extreme low levels, e.g. to a few microseconds, scanner-based imagers cannot achieve acceptable frame-rates in real-time applications.

Scannerless 3D image sensors were originally implemented based on non solid-state devices [6-8]. Non solid-state implementations typically utilize an image intensifier based on microchannel plate (MCP) photomultipliers [9] to form an electro-optical modulator or shutter on top of a conventional CCD/CMOS imager. Among reported devices, the most utilized range-finding TOF principle was based on the measurement of the phase shift a sine or square experiences while travelling to and back from the scene. Although reported devices exhibited good ranging performance, the use of non solid-state electro-optical shutters reduced its application field to a small number of low-volume uses, especially in scientific and military applications. Indeed, the MCP device within the image intensifier operates in vacuum and requires an extremely high-voltage (HV) of up to several kilovolts to operate. These constraints enforce special sealing packaging and HV circuitry, which in turn adds cost and limits its utilization to low-volume applications only.

As suggested in Figure 1.2, thanks to the advances made in IC technology over the last decades, especially in Silicon, scannerless implementations of 3D image sensors followed a natural evolution towards all solid-state designs [10-12]. These implementations offer a number of inherent advantages: low-cost fabrication, short acquisition time thanks to parallelism, high lateral resolution, insensitivity to mechanical vibrations, standard packaging, etc.

Solid-state implementations of range imaging devices are of high relevance in this thesis. A detailed analysis and classification of existing techniques and implementations is reported in next Section.

1.2 SOLID-STATE 3D IMAGE SENSORS: STATE-OF-THE-ART

The implementation of solid-state TOF sensors capable of acquiring and delivering full images involves a number of important challenges in sensing technology. Indeed, unlike other types of sensors, these devices have not followed a direct and natural evolution from implementations based on discrete components toward fully-integrated solutions. The main challenge involved in those devices is the need for scalable pixel solutions that could be arranged in two-dimensional arrays. In particular, on an integrated circuit, the pixel solution needs to detect and demodulate the incident optical signal so as to determine TOF. These requirements, which enforce stringent speed constraints at the detector and associated circuitry, were not met in conventional image sensors.

The literature on single point devices and non-scalable small array solutions, based on discrete components, is extensive. Some relevant examples may be found in [13, 14]. Despite the important scientific contributions of these references to the art, especially the concept of using CMOS time-to-digital converters for TOF ranging, the first implementations of scalable solid-state TOF image sensors, reported in [10-12], were based on CCD technology. In [11], a linear sensor of 32 pixels was reported achieving a resolution of approximately 10cm at 1.5 meters, in dark conditions. A square-modulated optical wave was used to illuminate the full scene of interested. As the light beam traveled from the source to the targets in the scene, and back to the sensor, it experienced a phase shift proportional to TOF. The sensor of [11] extracted this phase shift pixel-wise by demodulating the incident optical signal. Although the sensor prototype was based on single line of 32 pixels, the CCD-demodulation structure used in the pixel was scalable. This structure, introduced in [10] and referred to as “lock-in CCD”, enabled low-noise detection and demodulation of very weak optical signals. In the lock-in pixel, demodulation occurs in the charge domain and over a high number of illumination cycles. Charge-to-voltage conversion is achieved only at the end of the acquisition time, thereby achieving superior performance. In [12], Lange and Seitz reported the first fully-integrated 3D image sensor featuring a two-dimensional (2D) array of pixels, based on the lock-in CCD structure. The sensor consisted of 64x25 pixels that operated independently and simultaneously so as to demodulate the phase from incident optical signals. Real-time depth images were achieved within a range of 7.5 meters and with a resolution of 5cm. The illumination consisted of a sinusoidal modulation using an array of light-emitting diodes (LEDs) at a frequency of 20MHz. The illumination module generated an average optical power of 900mW.

In order to analyze subsequent contributions to the art, it is convenient to classify existing methods as a function of the (i) type of modulation, the (ii) TOF demodulation principle, the (iii) IC technology involved, and finally the (iv) nature of photodetection.

Among all TOF image sensors, one can distinguish two groups of modulation used: methods using continuous wave (CW) modulation and pulsed methods. Continuous modulation encloses all types of modulation with a duty cycle close to 50%. Examples of CW modulation are sine and square waves. Pulsed modulation on the other hand stands for a waveform consisting of a train of short pulses. Pulsed waveforms thus typically have a very small duty cycle. Another distinction between CW and pulsed modulation may be identified by looking at the measuring quantities. In CW modulation, TOF is determined via a

phase shift measurement whereas in pulsed methods, TOF is typically determined as a time measurement.

Figure 1.3 shows a classification that covers the current state-of-the-art on optical solid-state 3D image sensors. In the proposed classification, the methods introduced during this thesis work were intentionally omitted.

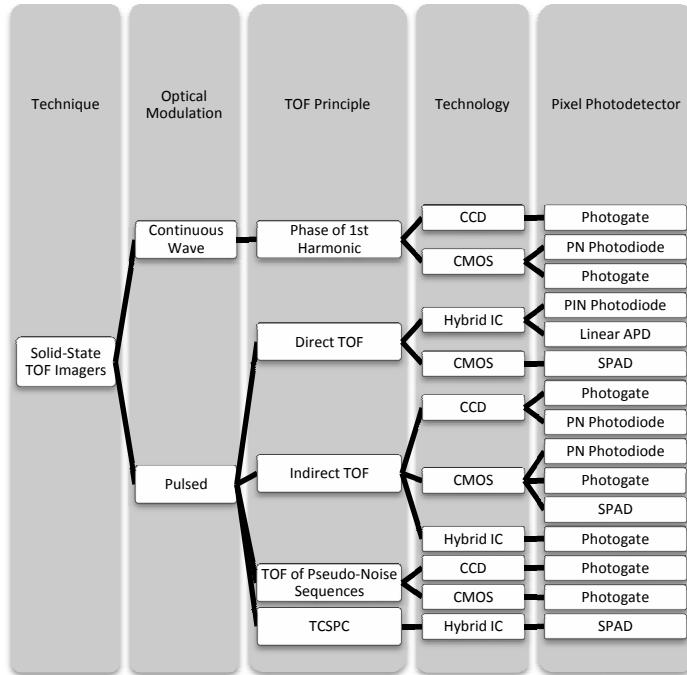


Figure 1.3. Classification of state-of-the-art optical TOF image sensors. The methods introduced during thesis work were intentionally omitted.

As can be seen in the Figure 1.3, the second criterion in the proposed classification is the TOF principle involved. TOF principles describe the basic measurement setup utilized to achieve range-finding. An example of TOF principle is the detection of the phase shift a modulated optical wave experiences when travelling to and back from the scene. The first principle shown involves for instance the detection of the phase corresponding to the first harmonic (or fundamental frequency) of the optical signal. These principles are described with more details in the next paragraphs.

Another important criterion, depicted in Figure 1.3, is the IC technology involved. Standard and widely available IC technologies are CMOS and CCD. However, a few solid-state 3D image sensors have been implemented in hybrid technologies, i.e. a custom integration process used for the photodetectors and usually a CMOS IC for the readout and signal processing. Since electrical interconnections are required pixel-wise between two integrated circuits, special packaging technology such as bump array bonding and miniature interchip vias have been introduced.

Finally, the last criterion is the photodetector device. The nature of the photodetector is of high importance since not all devices fulfill simultaneously the conditions in speed and efficiency required in a TOF imager. In particular, small p-n photodiodes in conventional CMOS do not provide charge-domain signal processing capability and therefore exhibit limited performance. The reason for this limitation is that small p-n photodiodes generate extremely small levels of photocurrent that cannot be directly processed. As an alternative, one may increase considerably pixel sizes and illumination power, the latter being unacceptable in most applications. Photogate devices, which have been originally introduced in CCD technology, enable charge-domain signal processing [10]. Interestingly, these devices may be conveniently fabricated in standard CMOS. As a result, CMOS photogates have also been rapidly used to demodulate the phase of incident optical signals for range-finding [15, 16]. Linear avalanche photodiodes (APD) and p-i-n photodiodes (PIN) are the detectors of choice in scanning-based rangefinders owing to their superior speed performance. However, in a scanning-based TOF rangefinder, the APD or PIN photodetector is a relatively large device and, when coupled to a highly collimated optical beam, it delivers significant photocurrent signals. Under this condition, effective signal demodulation is practical. However, concerning the design of PIN/APD photodetector arrays for range-finding, only highly optimized devices fabricated in custom technologies have been reported in the art [17]. These implementations therefore required hybrid IC packaging. Similarly, arrays of SPAD devices fabricated in custom technologies were effectively used in 3D image sensors, as reported in [18]. With the introduction of SPADs in CMOS technology [19, 20], these devices exhibited significant potential for fully-integrated 3D image sensors. The evaluation of CMOS SPADs for 3D imaging were thus first reported in [21-23], in the framework of this thesis. Following these contributions, a number of other researchers joined the investigation of CMOS SPADs for high-performance range-finding [24, 25] as discussed in the following paragraphs.

1.2.1 TOF BASED ON PHASE OF FIRST HARMONIC

In this method, a periodic optical signal, typically a sine or square wave is modulated at a given frequency and used to illuminate the entire scene. Range-finding is achieved by measuring the phase shift the optical wave experiences when it travels from the source to the targets in the scene and back to the sensor. Depending on the waveform used, the optical signal spectrum may contain more than a single harmonic. However, in this principle, the rangefinder determines the phase of only the first harmonic. An illustration of this method is shown in Figure 1.4. Back-reflected photons are imaged onto the array of pixels at the focal plane. In each pixel, the incident optical signal is detected and mixed with two electrical signals that are synchronous to the light source. One signal is in phase with respect to the light source and the second is in quadrature (I/Q). The mixed signals are then low-pass filtered by means of two integrators, thus leading to two components representing the real and imaginary part of the received optical signal. These components, in turn, allow one to determine the phase φ , amplitude A , and offset B of the back-reflected signals for every pixel [26]. Once the phase determined, the distance z to each point of the scene, is computed as

$$z = \frac{c}{2} \frac{\varphi}{2\pi f}, \quad (1.2)$$

where f is the fundamental frequency. This detection principle, known as homodyne or lock-in demodulation, may be alternatively implemented by a cross-correlation operation between the received optical signal and the electrical reference signal [12]. By evaluating the resulting correlated signal at four phases, e.g. 0, 90°, 180° and 270°, one may unambiguously determine φ , A , and B .

A third demodulation approach used to determine φ consists of synchronously sampling the received signal with three or more samples within one illumination cycle [10, 12]. In order to reduce the effects of noise, these samples may be averaged over a high number of cycles by means of pixel-level integrators. At the end of the integration time, the sample values are used to determine φ . For instance, assuming four sample values, namely $S_0, S_{90}, S_{180}, S_{270}$, the phase φ , amplitude A , and offset B are computed as

$$\varphi = \arctan \left(\frac{S_{270} - S_{90}}{S_0 - S_{180}} \right), \quad (1.3)$$

$$A = \frac{1}{2} \sqrt{(S_{270} - S_{90})^2 + (S_0 - S_{180})^2}, \text{ and} \quad (1.4)$$

$$B = \frac{1}{4}(S_0 + S_{90} + S_{270} + S_{180}). \quad (1.5)$$

The basic lock-in demodulation structure in CCD/CMOS typically utilizes a central photogate that is exposed to light. Under this photogate, electrons are photo-generated. Surrounding the central photogate, two or more transfer gates are used to synchronously transfer photogenerated electrons to two or more independent gates or diffusions, operating as integrators. These integrators were typically covered by metal layers not to be exposed to the incident optical signal. The lock-in structure allowed the synchronous sampling and averaging of the incident optical signal, thereby leading to depth information. As previously mentioned, the basic lock-in structure was reported firstly in [12, 27], based on CCD technologies.

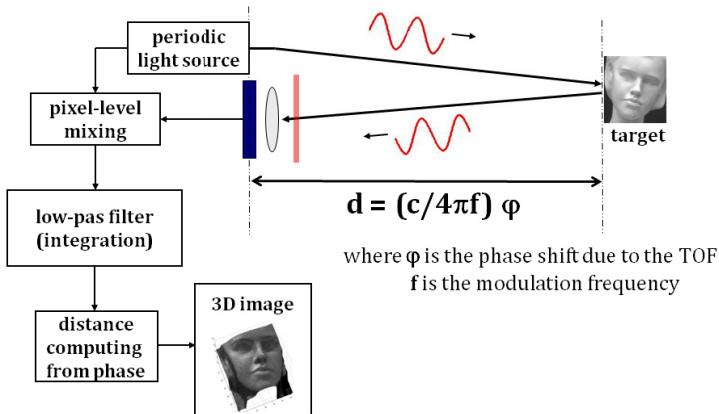


Figure 1.4. Range-finding principle based on sine or square continuous wave modulation. A pixel-level demodulation circuit extract the phase shift the optical beam experiences.

In CMOS technology, the first implementation of the photogate lock-in structure was reported in [15, 16]. In those papers, although a single pixel was tested, the authors also described the concept of single-chip 3D image sensors.

In [26], the implementation of the lock-in principle was not based on photogates. Instead, conventional p-n photodiodes (PN) in CMOS were used. The photocurrent in each pixel was mixed by synchronously modulating the

quantum efficiency of the photodiodes. Although quantitative evaluation of this method was not reported in the art, prototypes of 3D cameras based on this principle have been commercialized [28]. The use of CMOS technology enabled single-chip camera modules to be conceived.

Despite the interesting potential of lock-in pixels in CCD and CMOS technology, these devices exhibited a number of drawbacks in their first implementations. One of the most important drawbacks of these devices is their vulnerability to uncorrelated background illumination. On a typical optical setup, used in solid-state depth sensors, a narrow-band interference filter is used to block a large fraction of background light. Nonetheless, since background light illuminance may be extremely high, especially outdoors, the depth sensor should provide additional rejection to background light. Background light typically causes a common mode signal in all integrators. Although the mean value of the common mode signal may be disregarded, its noise component significantly impairs depth accuracy. More importantly, when the intensity of background light or of correlated optical signal is very high, it causes one or more integrators to saturate, thereby losing the signal contrast between them. This effect is very severe as it leads to incorrect depth information. In order to address these issues, a number of new designs have been proposed in the art [29-33]. In [29], a new pixel design based on CMOS photogates was proposed as an alternative to the original design, which utilized p-n photodiodes. The new pixel is based on a symmetric structure on which correlated optical signal appears as a differential signal. Uncorrelated background light generates common mode signals that are disregarded during pixel readout. No details were given by the authors if some mechanisms to prevent saturation effects were added. The device exhibits a typical resolution of a few centimeters within a range of a few meters. References [30-33] and [34] proposed promising solutions to suppress common-mode signals in CCD and CMOS implementations, respectively. In [30-33], different approaches were proposed to address the background light as well as signal saturation issues. The first approach, called pixel-wise integration, address the saturation effect by reducing the effective pixel integration time whenever one integrator reaches saturation. Since the CCD photogates, used as integrators, of all the pixels cannot be monitored in real time, the full integration time is divided into a number of sub-integration intervals. At the end of each sub-integration time, an analog saturation check is executed for each pixel, by means of a local comparator added into the pixel. If one or more integrators reach saturation, the pixel is kept disabled until the end of the integration time. As a result, the contrast between integrators is maintained, thereby preventing impaired depth measurements due signal saturation to occur. The approach appears to be effective in reducing saturation effects. It

increases the dynamic range of the sensor by 24dB, when compared to the conventional structure [33]. However, the solution slightly reduces the effective sensor exposure time and the fill factor due to added complexity. The second approach, called minimum-charge transfer, does not reduce the effective integration time. Alternatively, the approach relies on integrators with very large capacity. However, the capacity of the diffusion sensing nodes is kept low (unchanged) in order to maintain the conversion gain of the pixel's source follower high. In order to prevent the sensing nodes to overflow during readout, the charge transfer from integrators to sensing nodes is achieved progressively. A comparator was added in the pixel, so as to determine the time at which the voltages of all the sensing nodes change from their reset value. At that time, the charge transfer is stopped since additional charges would result in common mode signal only. At the same time, differential signals between integrators are available on diffusion nodes to be read out. Measurement results have demonstrated the effectiveness of this approach in increasing the saturation level by 15.5dB. While the first approach is very effective against saturation due to correlated signals, the second appeared to be more appropriate against the saturation caused by background light. In both cases, the price to be paid is a lower fill factor and added complexity in analog processing. In reference [34], a similar approach to reduce the effect of background light is used, however no implementation details were reported.

Another important drawback of conventional lock-in pixels was the limited demodulation efficiency due to slow carrier transit times in photogates. Typically, on surface-channel CCD and CMOS photogates the electric field below the photogate is planar and charge transport is dominated by diffusion effects. When an electron is generated as a result of photon absorption at a given time, ideally, this electron should be immediately collected in the corresponding integrator so as to generate a useful signal. However, due to slow diffusion processes involved in conventional photogates, the mean electron transit time ranges from a few nanoseconds to several tens of nanoseconds [12]. As a result, a non-negligible fraction of detected photons does not contribute to depth measurements. On the contrary, these photons generate a common mode signal that appears in all integrators simultaneously, similarly to the effect of background illumination. In [12], the demodulation contrast was characterized as a function of modulation frequency and of incident optical wavelength. As expected, signals with longer wavelength produce lower demodulation contrast due to their higher absorption depth. For these reasons, the demodulation efficiency reported for an 850-nm optical signal modulated at 1MHz was already as low as 20%.

Low demodulation contrast is a highly undesirable feature as it forces system designers to increase the illumination power budget, for a given performance objective. Naturally, a number of solutions have been investigated to address this limitation [33, 35-37]. In [35], Lange has proposed to use buried-channel CCDs (BCCDs) to improve the speed of charge transport. Below BCCD photogates with small area, electric fields exhibit smooth profiles which induce lateral drift effects. Transit times of electrons under photogates are significantly reduced when drift effects dominate. Although the BCCD solution works well in small pixels, it is not suitable for larger photogates.

One approach investigated in [33, 36] was to dynamically modify the electric field profile below the photosensitive and demodulation photogates. The central photosensitive and adjacent transfer gates were replaced by a single photogate, in which an electrical current was injected. The photogate was fabricated using high-resistive poly-silicon gates and therefore, depending on the direction of the current injected, a voltage distribution on the photogate and thus in the channel was induced. By modulating the photogate current synchronously with the illumination source, homodyne demodulation was achieved by means of a single photogate. Experimental results, reported in [36], demonstrated good improvements. A maximum demodulation contrast of 25% was achieved using an 810nm illumination operated at 20MHz. Using an illumination at 630nm, the demodulation contrast was 55%. However, these results are far below the maximum theoretical contrast of 90%. While the solution may be suitable in some implementations, it may lead to excessive power (and heat) dissipation if a large array of pixels ought to be designed. Furthermore, the designed sensor leads to relatively poor range accuracy and repeatability. The maximum non-linearity error was approximately 20cm whereas the maximum 1σ repeatability error, measured at each distance, was 13cm. The measurement range was from 35cm to 3.5 meters. Büttgen *et al.* investigated another approach in [32, 38]. In order to mitigate the optimization trade-off between photogate size and carrier transit time, the authors decided to physically separate the light collection area from the demodulation site. A large photosensitive area was made using a number of narrow photogates. The voltages applied between these photogates were progressively (but statically) increased so as to generate a lateral field, thus making it possible to direct the transport of electrons to a desired location. At that location, two small transfer photogates were used to synchronously demodulate photo-electrons. Since the demodulation gates were very small, short transit time to integrators was expected. The estimated overall transit time of electrons from the top of the photosensitive area, including transit time in the demodulation photogates, was 4.17 nanoseconds. Based on this estimation, a -3dB cut-off demodulation frequency of 105MHz was estimated.

Unfortunately, effective demodulation contrast was not reported. Instead, the transit time from the top to the bottom of the photosensitive area was experimentally measured. The measured transit time was 0.5 nanoseconds, nonetheless, this value did not include the transit time within the demodulating gates. Despite the effectiveness of this approach, including the potential for relatively high fill factor, experimental results in range-finding was somewhat disappointing. A highly integrated sensor with 176x144 pixels and featuring over 25% of fill factor was reported. The sensor exhibited a maximum non-linearity error of approximately 30cm within a range of 3.5 meters, using a 20 MHz illumination source. Repeatability performance was characterized by a 1σ repeatability error of 2.5cm at 1 meter. The 1σ error increased quadratically with distance, due to illumination field-of-view, to 25cm at 3.5 meters. Repeatability performance may be mitigated by increasing acquisition time or by increasing illumination power. However, non-linearity errors do not improve with averaging and are more therefore difficult to deal with. The authors stated that systematic errors were caused by harmonic distortions in the optical illumination. Indeed, odd harmonic components cause aliasing errors in lock-in demodulation when four samples per illumination period is acquired [35].

1.2.2 DIRECT TIME-OF-FLIGHT PRINCIPLE

This principle, as the name suggests, involves the direct TOF measurement of optical pulses sent by a light source. Similarly to all solid-state imaging principles, the optical beam is used to illuminate all points in the scene simultaneously with a given field-of-view. In each pixel, a photodetector is used to detect incident optical pulses and to generate a trigger signal that is used to stop a very precise timer, thereby providing TOF. Despite its simplicity, the direct TOF measurement of light pulses was not extensively adopted in the art for solid-state 3D image sensors. Note that the TOF principle based on time-correlated single-photon counting (TCSPC) is distinguished from this principle, according to Figure 1.3. The TCSPC principle is described in another paragraph in this Section.

The direct TOF principle suffers from high sensitivity to background light when single-shot measurements are required. Otherwise, it is possible to average the mean TOF over a number of measurements. However, when averaging is performed, then TCSPC might achieve better background light rejection. Similarly to the TCSPC principle, direct TOF measurements require illumination sources with extremely fast rise and fall times. Furthermore, in order to precisely detect optical pulses, very fast photodetectors and front-end electronics are equally necessary.

Reference [17] describes the design of solid-state 3D image sensors based on arrays of PIN and linear APD photodiodes. The implementations are based on hybrid stacking of two integrated circuits. A custom processed IC with an array of photodiodes is bump bonded to a CMOS readout IC. The array of photodiodes may be highly optimized to achieve the required performance independently of the readout IC. For instance, an array of InGaAs PIN photodiodes is used for imaging at $1.5\mu\text{m}$ wavelength whereas HgCdTe PIN photodiodes are used in imaging near $3.8\mu\text{m}$. Arrays of linear mode APDs based on HgCdTe with broadband sensitivity ranging from $0.8\mu\text{m}$ to $4.5\mu\text{m}$ are also reported. On the readout IC, an array of front-end and time-to-amplitude converters (TAC) was implemented. The front-end circuit was based on low-noise and high-speed trans-impedance amplifiers used to generate trigger signals on pulse detections. As a result, each pixel was able to measure TOF independently and in parallel. An analog readout circuit provided three readout channels, one for range and two for intensities, operated at 1 megasamples per second each. Prototypes featuring 64×64 and 256×256 (range only) pixels were built. These prototypes are able to resolve resolutions of 50cm. The flexibility of independent optimizations of sensor array and readout circuitry leads to the design of rangefinder image sensors capable of operating at specific wavelengths. This is a very important advantage when the 3D image sensor needs to operate with high reliability in outdoor conditions. This feature cannot be achieved in silicon devices, which operate in the visible and very short near-infrared range only. Unlike silicon devices made in CCD/CMOS however, hybrid implementations typically involve high development and fabrication costs. Moreover, special stacking packaging is required, which may potentially reduce overall fabrication yield. These limitations may prevent 3D image sensors based on hybrid technology to be widely utilized in applications with high volume, at least in a very near future.

SPADs fulfill the conditions of high-speed and may be utilized in direct TOF measurements. However, SPAD-based sensors may be conveniently utilized in combination of the TCSPC principle as well. In references [24], the design of a linear arrays of SPADs fabricated in CMOS technology was evaluated as a rangefinder based on the direct TOF principle. The sensor, fabricated in a $0.8\mu\text{m}$ CMOS technology, features a line of 64 single-photon pixels. Each pixel also includes a local TAC and an analog averaging circuit on an area of $180\times 38\mu\text{m}^2$. Thanks to picosecond timing resolution of SPADs and optimized TAC circuit, the sensor exhibited a precision of $\pm 0.75\%$ ($\pm 3.75\text{cm}$) within a range of 5 meters. These measurements were based on a cooperative target and under controlled background light.

1.2.3 INDIRECT TIME-OF-FLIGHT PRINCIPLE

A number of solid-state 3D imagers have been introduced based on the indirect TOF measurement principle. In this principle, each pixel does not rely on the arrival time of rising edges to determine the TOF of optical pulses. On the contrary, the measurement is based on the gated integration of the optical pulses over time. As a result, the method mitigates the requirement on fast photodetectors and illumination sources.

The indirect TOF range-finding principle relies on a shutter mechanism, typically MOSFET switches although different implementations were also reported. Some research groups also call this principle “light wall” ranging technique. A light pulse of some tens of nanoseconds is emitted by a laser or LED within a field-of-view so as to illuminate the entire scene. In Figure 1.5 (a), an illustration of the illumination principle is shown. Synchronously to every pulse sent, the shutter is activated so as to integrate amplified incident photons on an image sensor. The shutter is then synchronously blocked after a very short period of time (also some tens of nanoseconds). As a result, depending on the distance of every point on the target, only a fraction of the light pulse arrives before the integration stops, as illustrated in Figure 1.5 (b) and (c). The method therefore allows the computation of TOF by measuring how much photocurrent has been integrated. However, since the amount of optical signal integrated on pixels also depends on the target’s reflectivity and that some background light is also measured, this principle requires at least two acquisitions to form one depth image.

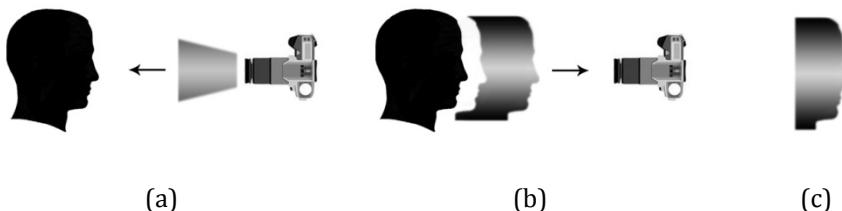


Figure 1.5. Principle of indirect time-of-flight, also known as “light wall” principle: (a) illumination pulse forming a wall of light travels along the field-of-view toward the scene, (b) reflected light pulse is time-delayed according to target’s 3D shape, and (c) truncated light pulse integrated on pixels due to shutter action.

In references [39, 40], the design of a CMOS prototype was first described. Jeremias, Mengel *et al.* have taken advantage of switched-capacitor (SC) circuit techniques to implement electrical shutters. A monolithic array of 2x32 pixels based on conventional p-n photodiodes and achieving a minimum shutter time of 30ns was introduced. The main limitation of the approach is the requirement of high-power illumination and large photodiodes to overcome noise levels. In particular, the photodiode current should be high enough to overcome the effects of the parasitics inherent to SC circuits, e.g. KT/C, charge injection, etc. The measured noise-equivalent illumination power was 5.2W/m², based on 260μm² photodiodes. The sensor achieves non-linearity errors below 5% within a range of approximately 1 meter, assuming 5 to 10 laser pulses. Another design was proposed in [41], by the same research group. An array of 4x64 pixels with an area of 130x300μm² was fabricated in a 0.5μm CMOS technology. This sensor improves the figure of noise equivalent illumination power to 2W/m², mostly due to the use of photodiodes with large area. Non-linearity errors were below 5% and 1σ error was 1cm assuming 100 laser pulses. Despite the performance improvement when compared to the previous generation, the main limitation of this design was the use of non-scalable pixels. Every pixel was individually connected to its corresponding readout circuit, outside the pixel array.

The same measurement principle was evaluated by another research group in [42-45]. Stoppa, Viarani *et al.* took advantage of a new full-differential pixel implementation [42] to increase significantly the dynamic range of the sensor. An array of 8x8 pixels was designed in CMOS with an area of 97x116 μm², described in [43]. A dynamic range of 96dB was expected, although no experimental results were reported. In [44], a single-pixel was redesigned and its ranging performance was characterized experimentally using a highly reflective target. The sensor exhibits a maximum non-linearity error of 15cm and 1σ error of 24cm, based on approximately 2000 optical pulses (64 frames of 31 pulses averaged). The measurement range was from 3 to 9 meters whereas the illumination power was 40mW. The same authors redesigned a small array of 16x16 squared pixels with a pitch of 80μm and fill factor of 20%. This image sensor demonstrates high background light rejection thanks to the differential topology utilized. Ranging performance was characterized by a maximum non-linearity error of 5% (2%) at 30 fps (3fps) within a range from 2 to 9 meters. The sensor also demonstrates outstanding noise-equivalent illumination power of 40mW/m², assuming 128 optical pulses with 200ns of duration.

Furthermore, Elkhalili *et al.* reported another generation of their design in [46], doubling the number of pixels with respect to [41]. However, the sensor

architecture is very similar to the previous generation, thus limiting the scalability of the pixel array. In this design, an array of 8x64 pixels was designed, each pixel individually connected to its corresponding readout circuit outside the pixel array. The pixel area, mostly occupied by a PN photodiode, is unchanged, i.e. $130 \times 300 \mu\text{m}^2$. Ranging performance is relatively similar to the previous generation. The noise equivalent illumination power is increased to 2.16W/m^2 , at the shortest shutter speed, while 1σ repeatability error is improved to 2.5mm when averaged over 100 pulses. The image sensor IC dissipates 3 Watts of power in operation.

Finally, Stoppa *et al.* also evaluated the indirect TOF principle in [25] using CMOS SPADs. An array of 1x64 pixels with an area of $38 \times 180 \mu\text{m}^2$ was integrated in a $0.8 \mu\text{m}$ CMOS. The sensor architecture is based on an analog counter and averaging circuit at pixel level to count photon events while shutter is on. Based on the indirect TOF principle, the sensor achieved a 1σ repeatability of 5cm at 2 meters of distance. Interestingly, the same design achieved 1.5cm of repeatability error at 2 meters based on direct TOF.

Another 3D image sensor using the indirect TOF and a light wall illumination was introduced in [47]. Yahav *et al.* developed a 3D camera based on a new type of electro-optical shutter, introduced in [48], which replaced the typical image intensifier-based shutter. The new solid-state electro-optical shutter provides sub-nanosecond gating times and operates based on a single and relatively low voltage. TOF demodulation is operated globally on a full image, which allowed the authors to use conventional off-the-shelf CCD image sensors. A key advantage of the approach is that, since no special circuit is required at pixel level, relatively high (lateral) resolution image sensors may be utilized, including conventional CMOS imagers. However, the implementation requires a hybrid packaging of the electro-optical IC on top of conventional image sensors. The commercial specifications of the 3D camera [49] are: a remarkable 640×480 (VGA) range format, depth range from 0.5 to 2.5 meters, a 60° field-of-view, and a maximum of 60fps of frame rate. The ranging performance reported was characterized by a precision of 2cm with 8-bit of resolution. Accuracy was not specified however. Despite the relatively high-power consumption of 12 Watts, especially for mobile applications, the camera technology exhibits relatively good performance. Regrettably, no technical information can be found in the literature concerning the actual contrast ratio between ON and OFF states as well as maximum transmission for the electro-optical shutter.

Another variation of the indirect TOF principle was recently introduced by Ushinaga, Kawahito *et al.* in [50, 51]. A 3D image sensor featuring 336×252

pixels was integrated in a $0.35\mu\text{m}$ CMOS technology. The sensor pixel was based on demodulating photogates, very similar to lock-in pixels described previously. The evaluation of TOF is based on a new demodulation scheme that utilizes two transfer photogates, driven synchronously with the pulsed light source, and two floating diffusion integrators. Unlike the lock-in pixels in CCD/CMOS, in this approach, the optical illumination is operated with very small duty cycles, typically 10%, but 1% was also suggested in the references. In every illumination cycle, at the moment an optical pulse with duration T_0 is emitted, the first transfer photogate is activated so as to integrate photogenerated charges in the first integrator. It stays active for a duration of T_0 exactly, then, it is turned off and the second photogated is activated, also for a duration of T_0 . Accumulated photogenerated charges in the two integrators followed a distribution depending on the arrival time of optical pulses. For instance, if the TOF to be measured is zero, then the first integrator detects the optical pulse entirely, thus leaving the second integrator empty. On the other hand, if the TOF was not zero, then a fraction of the optical pulse would be detected by the second integrator, thus changing the distribution of charges among the two integrators. If the TOF is exactly $T_0/2$, i.e. half of distance range, both integrators collect the same number of electrons. The method therefore allows one to compute TOF by reading out the voltages of the two integrators. Since the duty cycle was chosen very small, the two integrators are inactive during a very important fraction of the illumination period. During this time, only background illumination is collected. Two additional transfer gates were added into the pixel so as to drain these unused photogenerated electrons to the power supply. Ranging performance based on this approach was characterized by a maximum non-linearity of 3.3% within a range from 1.8 to 3.3 meters. This result was obtained after an offset compensation to improve the linearity of the sensor, which had been affected by a common offset voltage on both integrators. The performance evaluation reported of non-linearity error was limited to a small fraction of the measurement range only. At both ends of the range, higher non-linearity errors were clearly visible in the sensor characteristics, according to [51]. Despite the important contribution to the art, especially with respect to the number of TOF pixels operating simultaneously on single-chip solution, the proposed method suffers from a number of limitations. First, the range repeatability, approximately a few centimeters in [51, 52], depends on the distance even if the illumination irradiance on the target surface keeps constant. Practically, the signal irradiance on the target decreases with the square of the distance due to the field-of-view, thus affecting the sensor precision. However, in this implementation, the precision itself exhibited a systematic component depending on the distance. Second, the sensor linearity was impaired due to offset voltages that appeared on both integrators. This effect suggests that

higher non-linearity errors should be expected when the sensor is operated under strong background light, as background light leads to an offset voltage on both integrators. Finally, this approach utilizes two points to sample a square pulse that lasts for the duration between the two samples. The detection method is therefore violating the Nyquist-Shannon sampling theorem. As a result, the approach suffers from aliasing effects which impair distance performance, even in theoretically ideal conditions.

1.2.4 TIME-OF-FLIGHT BASED ON PSEUDO-NOISE SEQUENCES

In order to extend the application of 3D image sensors to environments in which two or more devices should operate, special modulation/demodulation techniques are necessary. Indeed, in the context of multiple-camera operation, 3D image sensors that are based on active illumination typically interfere with each other when their illumination signal spatially overlaps. A very useful analysis of channel multiple-access techniques, typically utilized in communication systems, applied to 3D image sensors was described in [33]. A complete classification of multiple-access techniques is beyond the scope of this state-of-the-art analysis. Instead, in this paragraph, a description of actual implementations in the context of imaging rangefinder is provided.

Among all possible modulation/demodulation configurations, the use of binary pseudo-noise sequences has important practical advantages. This type of modulation belongs to a class of multiple-access methods, called direct sequence code division multiple access (DS-CDMA) [33]. In this technique, a code sequence is used for intensity modulation. The use of orthogonal codes enables multiple-access to a single-channel, which in turn is typically based on the choice of a given illumination wavelength. The first actual implementation taking advantage of pseudo-noise sequences was reported in [16] and in references from the same authors there cited. Although a single pixel was designed, the concept of scannerless solid-state 3D image sensor based on PN sequences was analyzed. In the pixel, a central photogate, two transfer gates, and two integrators were utilized to synchronously demodulate incident optical pseudo-noise sequences. The pixel structure was the same as the lock-in CCD/CMOS pixel based on two integrators, utilized to measure the phase-shift corresponding to the first harmonic of sinusoidally modulated signals. The demodulation scheme assumes that each integrator corresponds to one data symbol. In [16], since binary pseudo-noise sequences were assumed, the first integrator was for instance assigned to “0” and the second one to “1”. By synchronously and exclusively activating the integrators using the same binary pseudo-noise sequence emitted by the illumination source, the lock-in pixel

enables to detect a cross-correlation signal between the emitted waveform and the received one. Correlation measurements of pseudo-noise sequences lead to useful properties. In particular, the cross-correlation between a pseudo-noise sequence and a sample of itself, but delayed in time and with different amplitude, exhibits a sharp peak at a time position corresponding to delay between them. This mechanism was therefore proposed in [16] to determine the TOF of an optically modulated pseudo-noise sequence. Another useful correlation property of pseudo-random sequences is that, when two different sequences are cross-correlated, the resulting waveform is a constant. On a 3D image sensor, when the illumination signal belonging to another device is simultaneously detected, it produces a constant contribution in the cross-correlation being measured. As a result, the optical power that belongs to other devices in the environment has the same effect as uncorrelated background light, thus enabling multiple cameras to operate simultaneously.

Although the concept was introduced in [16], Schwarte *et al.* have not reported actual evaluation of this principle in the context of multi-camera operation. A simple device test, under a very low frequency of 5kHz, was reported. Due to the large area of the pixel, i.e. $120 \times 120 \mu\text{m}^2$, relatively low demodulation efficiency was demonstrated.

In [33], theoretical modeling and simulation of distance performance based on pseudo-noise sequences was reported. Experimental validation of the model was also reported, though not complete. Measurements were carried out using a existing 3D camera [53], exhibiting 176x144 pixels. The experimental setup assumed a pseudo-random sequence optical modulation with length set to 2047 bits (chips) of 50ns. Based on these values, the period of the full sequence was $102.35 \mu\text{s}$ whereas the non-ambiguous distance range was 7.5 meters. The theoretical demodulation principle was demonstrated experimentally. However, non-linearity errors of approximately 5% were noticed. Büttgen also validated the theoretical model of repeatability error with experimental results. Based on this demodulation technique, the theory predicts that sensor precision depends relatively strongly on the TOF to be measured, which was demonstrated experimentally. Finally, the author showed that, under the effect of background light only, the precision of this demodulation technique was from 3 to 5 times worse than the approach based on sine wave modulation. However, in a multi-camera context, pseudo-noise sequences clearly outperformed when compared to the simple sine wave modulation.

1.2.5 TIME-OF-FLIGHT BASED ON TIME-CORRELATED SINGLE-PHOTON COUNTING

Thanks to the introduction of arrays of SPADs in custom IC technologies [54, 55], a relatively old detection technique known as time-correlated single-photon counting (TCSPC) [56, 57] became practical in solid-state 3D image sensors. In TCSPC, typically an illumination source with very short optical pulses is used to illuminate the scene of interest with a given field-of-view. For every light pulse sent, the light source also provides an electrical or optical reference signaling the moment the pulse was emitted. This trigger signal may be used as a START signal for all the pixels in the image sensor. SPADs at pixel level generate independent STOP signals upon photon detections, which are used in the TOF measurement. For each pixel, a high-speed timer is used to measure time interval between START and STOP and the resulting TOF is recorded in an individual histogram. Instead of measuring TOF based on a single back-reflected photon, in TCSPC, the detection principle typically involves a large number of detections. A very important difference between TCSPC and direct TOF is the use of histogram processing. The first solid-state 3D imagers taking advantage of TCSPC were introduced in [18, 58-61]. In [18, 58], Aull, Heinrichs, *et al.* described an array of 4x4 then 32x32 single-photon pixels made in silicon using a custom process, then mounted on top of a CMOS IC on which timing circuitry was implemented. A special hybrid process was developed, including special inter-chip vias, so as to allow the SPAD array chip to be mechanically and electrically connected to the CMOS readout IC. Once the hybrid integration was in place, conventional packaging was utilized, similar to single-chip solutions. The rangefinder sensor was based on a pulsed frequency-doubled Q-switched Nd:YAG laser. However, only preliminary range results have been reported. In [59, 60], the approach was evaluated experimentally. A highly optimized, laboratory-style, optical setup was designed. The sensor exhibited a range resolution of 15cm within a range of several hundred meters, with low-repetition rate, i.e. 10 kHz, of a 523nm light source. TCSPC, combined with an optimized array of focused laser beams, enabled the authors to acquire full range measurements and then scan the objects along the range by numerical processing. Qualitative evaluation of the approach, based on histogram processing, was achieved by enabling the 3D image sensor to capture depth images of objects hidden by the foliage of trees. Recently, in [61], Aull *et al.* have scaled the approach to an array of 64x64 pixels. A special process based on a stacking of three ICs was developed. The 3D integration allowed the authors to build interchip vias used to connect the SPAD array chip to two remaining readout ICs. Excessive power dissipation on the chip constrained the authors to acquire range images with a resolution of 30cm (2ns), instead of the expected 1.95cm (130ps). Although the approach demonstrates the potential of TCSPC for high-performance 3D imaging, it still suffers from limitations that prevent it to be used in high-volume applications. In particular, the effectiveness

of the 3D integration technology has still to be proven in other, non-imaging, large-scale integrated circuits.

1.3 APPLICATIONS

The field of applications involving depth map evaluation and acquisition is vast. Furthermore, measurement ranges vary from a few nanometers to several tens of kilometers. Naturally, each 3D imaging method or technology suits best a subset of applications. In this Section, a selective list of applications that may directly benefit from solid-state 3D image sensors based on TOF is presented. Importance is given to applications that are achievable with the techniques and technology introduced in this thesis. In particular, applications requiring fast, accurate and robust depth map evaluation in the range of a meter to several tens of meters are potential candidates. Since TOF ranging involves an active illumination, typically an array of light-emitting diodes (LEDs) or laser diodes (LDs), candidate applications should require a certain level of measurement reliability. Otherwise, simpler and potentially cheaper ranging techniques would suffice. Reliability, in this context, is related to performance parameters such as accuracy, repeatability, and compatibility with target properties. Another important criterion is cost. Scientific, military, and a few other classes of applications do not involve high volumes, thus cost is typically irrelevant. On the other hand, for industrial applications, especially those involving consumer, cost is a very important parameter. In view of the single-chip feasibility of SPAD image sensors in CMOS technology, target applications are those that demand low-cost implementations.

1.3.1 AUTOMOTIVE APPLICATIONS

This class of applications is notably one of the most demanding in terms of reliability, robustness, cost, and environmental constraints. While performance parameters may be relaxed in some cases, depending on the actual use, operating constraints, such as temperature range and background light levels, impose significant challenges to TOF rangefinders. For example, the grade for full temperature range in automotive applications requires reliable operation from -40°C to 125°C. Background-light rejection to cope with 100klux of sun light is mandatory. As will be discussed in Chapter 3, TCSPC-based TOF makes SPAD image sensors a unique technique that overcomes such background-light rejection constraints.

Environment sensing on cars is expected to considerably increase road safety in the future [62-64]. A number of possible utilizations for 3D image sensors in this context are illustrated in Figure 1.6.

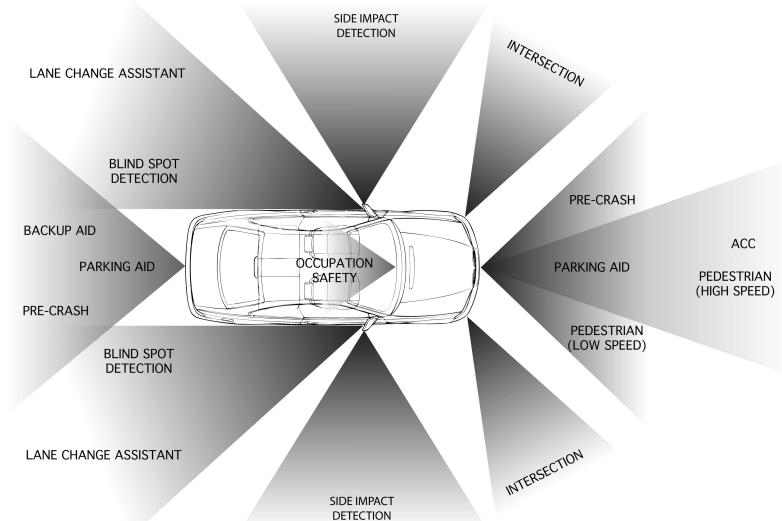


Figure 1.6. Road safety applications for depth sensors.

One of the most important uses of a 3D camera, installed in the front of a car, is the protection of vulnerable road users, such as pedestrians, bicycles, motorcycles, etc. Frontal detection scheme also enables collision mitigation and, possibly, adaptive cruise control (ACC). These systems would allow autonomous or semi-autonomous braking, which significantly reduces the kinetic energy dissipated on impact. Furthermore, collision mitigation would involve the intervention of active structure components such as controllable bumpers, crash boxes, motor hoods, and safety belt pre-tensioners. These components would further reduce the peak deceleration suffered by the parties involved in a collision. Front detection is completed with a number of additional measurement views. Front tilted and sideways ranging sensors would monitor the risk of accidents that occur typically when vehicles cross intersections. Long-range rear sensors could be utilized to assist drivers in lane change maneuvers. Short-range front and rear detectors could also be utilized for back up aid, autonomous or semi-autonomous parking-aid. Finally, at the interior of the car, range sensors could be useful to monitor the position of occupants in pre-crash situations. Such information could be used to optimize air-bag

deployment. As a result, safe airbag systems may be designed relying on contactless technology.

Automotive systems represent an important potential for the large-scale commercialization of solid-state range sensors. However, alternative techniques exist. For example, millimeter wavelength radio detection and ranging (RADAR) systems are currently under investigation. Radio-frequency (RF) of approximately 24 GHz is a typical candidate for short-range detection whereas millimeter wave in 77-79 GHz range are better adapted for a range of several tens of meters [63]. The advantages of RADARs over IR depth sensor include higher immunity to weather conditions, such as fog, rain and snow, in specific spectral ranges. However, while these systems perform well to detect targets with high reflectivity in the corresponding spectral range, such as vehicles, they are challenged by the low reflectivity of some vulnerable road users like pedestrians. In this particular application segment, visible-light and IR depth sensors are likely to dominate over millimeter-wave RADARs.

1.3.2 HUMAN-COMPUTER INTERFACES

An application field that is likely to experience a remarkable development over the next years is related to how human beings interact with computers and computer-based systems. Since the introduction of the modern computer mouse, almost 30 years ago, the basic human-computer interface has not evolved significantly. However, the subject has been investigated very early in the personal computing era by a number of research institutions and individuals [65]. Recently, with the current state of technology and the appearance of new user interface paradigms, especially in multi-touch [66-71], the theme is expected to get a second wind [65, 72].

3D image sensors enable fast acquisition and processing of real-time depth images. Based on this functionality, it is possible to detect and recognize user gestures in the three-dimensional space. In this context, some examples of applications that can take advantage of this type of user interface are described. In entertainment, a high number of different applications exist. The most promising candidates are applications that can benefit from an immersive user experience and user convenience. Video-game consoles, home entertainment devices, and videoconferencing are a few examples of this trend.

Video Gaming

The video-game industry originally focused on improving gaming quality by increasing processing power and image fidelity. Recently, with the successful

introduction of a new type of remote control that captures the movements of arms and body, and that can mimic real physical activities and sports [69], another trend appears to just have started. 3D cameras are expected to increase user experience in gaming by going a step further. For example, in [47, 73], a number of gaming applications that do not require any type of remote control are proposed. In those examples, users stand in front of a TV set, and a 3D camera [49] is utilized to detect user's body gestures, thereby replacing the usual remote control. The same approach has been envisioned by the remaining TOF 3D camera providers currently available on the market [28, 74, 75].

Interactive Home Entertainment Devices

The typical TV set has also been experiencing a major paradigm shift recently thanks to the availability of high-resolution and large-format display technology as well as the ubiquity of broad-band internet. A modern TV set is typically connected to, or integrates, a device to store a complete library of media files. These so-called media centers are becoming very popular and they represent a high potential for novel user interfaces. In particular, user interfaces that are based on simple user gestures have an important potential. Furthermore, it is likely that these media platforms will enable on-line commerce by adding new functionality. In [72], an example of such application is described. A user utilized a 3D camera to record her body metrics by taking a number of snapshots. Once her data were captured, the user was able to purchase apparel on line. The new functionality, in that example, was the possibility for the user to watch a virtual model of herself on which she was able to try a desired number of clothes. Other examples taking advantage of virtual models exist and have been reported, especially in entertainment. Finally, assuming that additional electronic devices and appliances at home interact with each other, the media center is likely to play a central role in a near future with respect to user interface.

Videoconferencing

In TV studios, frequent activities, such as weather forecast reporting, involves the suppression of the background in the scene behind a presenter. Typically, in those activities, the presenter stands in front of a panel colored with a specific color, e.g. blue. Background suppression is therefore performed by eliminating all the pixels that recorded the specific color. This technique is known as chroma key. Although the approach is effective, it enforces the recording procedure to take place at a specific location in the studio. In references [52, 76], the authors propose the utilization of depth cameras to overcome that limitation. A high-resolution 3D image sensor was coupled with a studio camera

so as to provide color and depth images at video rate. As a result, background suppression may be performed by applying a threshold in distance, i.e. depth keying. If the presenter is standing at approximately 2 meters from the camera, a threshold of 3 meters would allow the suppression of any scene portion that is located beyond 3 meters. Based on the proposed setup, since depth information is used to discriminate the presenter from the background, the presenter can be physically in any location, including outdoors. This functionality therefore enables virtual images to be superimposed and subtracted in a high number of situations, especially when involving reporters that are in the field.

In the same context, background suppression and replacement has a high potential in consumer applications such as home and professional videoconferencing. In many situations, users are concerned with privacy and have some reluctance to use videoconferencing. Depth sensing video camera with the possibility of background replacement would be valued. Some consumer electronics suppliers have proposed this feature based on conventional video cameras. However, the approach requires the background to be recorded prior to the utilization when the user is not in the scene. Moreover, the scene needs to be static. When a person unintentionally enters the field of view of the camera, the background suppression algorithm may fail. Depth-based background replacement on the other hand behaves reliably. Once image classification is performed based on depth keying to separate foreground from background, additional features may be added into the videoconferencing. For instance, the actual image of the user could be replaced by an avatar of the user him/herself or of any character. Furthermore, since 3D cameras can record facial expressions, head movements, and body gestures, any chosen avatar may represent the actions of the person being recorded.

1.3.3 *MACHINE VISION*

In this area, 3D imaging is expected to add important contributions. Depth sensing, combined with grayscale or color imaging, enables more reliable imaging processing. Computer vision may be more reliable than human vision to process repetitive tasks. However, this advantage is very often limited to very specific tasks in controlled environments. Depth sensing and imaging can add additional decision criteria in tasks such as image classification, scene feature extraction, pattern recognition, etc. Reliability is therefore expected to be highly improved thanks to 3D imaging in computer vision, which, in turn, may extend its application to uncontrolled environments.

Robotics

An important field that may benefit from reliable computer vision is autonomous and fixed robotics. Research in this field is very active, especially in autonomous robotics, and a number of research groups have already focused on solid-state 3D imaging [77-83].

Typically, mobile robots utilize conventional image sensors for creating maps, self localization, and collision avoidance. Also, most of systems use a horizontal laser scanner which offer clear advantages [77]. However, laser scanners also have some drawbacks. For instance, the field-of-view in these sensors is reduced to a single plane parallel to the floor, typically at knee height. In indoor environments, typical examples of obstacles are furniture (e.g. tables, chairs, cupboards, etc.), wall features (e.g. decorations, signalization, etc), floor features (e.g. steps), etc. In this context, difficulties have been reported for collision avoidance and localization due to non-unique situations [77]. In order to reliably prevent collisions, a high number of additional sensors, e.g. based on ultrasound, are utilized around the robot. Nevertheless, these additional sensors have a limited range of view and only produce one data at time; hence there is a need to have a plurality of them so as produce denser data. Solid-state 3D image sensors, which provide wide field-of-view, are therefore excellent candidates in autonomous robots. Furthermore, additional advantages are compactness, speed, and low-power consumption.

Safety in Manufacturing

In modern production processes, flexibility is a major concern. In this context, protective devices used in the workplace are required to support flexibility as much as possible [84]. The development of vision-based protective devices for safeguarding of machinery will allow the creation of new concepts, especially in the application of robots. However, the main challenge involved is the design of imaging systems that have the same reliability as existing protective devices. In addition, sufficient knowledge about safety-related aspects relevant to testing and utilization of such devices is necessary. Environment conditions, such as ambient light, shadows, dazzling, etc., are hard to cope with in vision-based protective devices based on conventional video cameras. Range image sensors are able to monitor 3D protection zones and provide more flexibility than laser scanners, which are currently being used to replace conventional devices such as light curtains and other fixed light beam devices. Moreover, TOF-based 3D image sensors are expected to be more robust against environment interferences. The application of range imaging to vision-based protective devices is therefore very promising, even if many detailed challenges are still

not met [85]. The main advantages of range imaging-based devices compared to existing approaches are: (i) possibility of monitoring larger areas with one device, (ii) expected low-cost, (iii) possibility of setting up multiple detection zones individually, (iv) possibility of switchover between multiple detection zones, (v) differentiation between man and material such as production goods allowing automatic muting, and (vi) possibility of visually check the monitored area and to record shut-down scenarios.

Cooperation between man and robots is an impressive example of a possible future use of range imaging vision-based protective devices. In particular, the use of cobots in manual handling activities and in complex manufacturing and assembly processes is increasing [86]. A cobot is a robot for direct physical interaction with a human operator, within a shared workspace. 3D vision-based protective devices are capable of automatically distinguish the worker from the cobot, and from the products being manufactured. As a result, these devices are likely to ensure a good teaming cooperation in human-machine systems, thus increasing operator's safety.

Finally, standardization of visual-based protective devices has already started due to the great interest of manufacturers and users. As a result, the first IEC technical report on the subject has been recently released [87].

1.3.4 OTHER APPLICATIONS

An exhaustive analysis of candidate applications for TOF-based 3D image sensors as well as for the monolithically integration of large arrays of SPADs in CMOS is beyond the scope of this dissertation. Furthermore, as briefly mentioned, the subject that relates to the capture and retendering of 3D images is extensive. This Section is concluded with two additional application areas that are likely to be enhanced with the introduction of SPAD-based 3D image sensor technology.

Biometrics

Currently, most of the face recognition systems are based on the processing of conventional intensity images of the face [88]. Typically, they can achieve good performance when the face images are taken under controlled indoor lighting conditions. However, since human face is a 3D elastic surface, the conventional image projection of a face is very sensitive to such changes as pose, illumination, and expression. Illumination variation is a challenging problem for face recognition because it changes the relative grayscale distribution of face images. Since intensity images reproduce the light reflection of the face surface at a

single observation angle, variations in lighting conditions will result in different images. Changes in illumination can have a substantial effect on the robustness and accuracy of face processing and recognition procedures. For instance, when changing illumination direction, the appearances of a face image can be highly affected and sometimes render images of the same face may appear dramatically different. The variation of pose results from the change of viewpoint or the head rotation. Although many facial recognition products claim good recognition rates with the frontal views of faces, the system performance drop significantly when pose variations are present between the face in the database and that in input images. The main reason for this is that pose changes can bring a strong variation to the facial images due to the relative camera-face pose, and some facial features may become partially or wholly occluded. In order to overcome these limitations, in the context of multi-modal biometrics [88, 89], the combination of 3D images with conventional intensity imaging is expected to improve accuracy in biometric recognition systems [88].

Another example of the application of multi-modal biometrics may be found in [90], where biometric recognition systems combine 3D images with infrared thermal pictures. By adding infrared thermal images, not only the accuracy of the system is improved but it also leads to additional user verifications. In particular, it enables the verification of a potential medical condition, especially in public environments such as in airports [90].

Applications in Research

In addition to 3D image sensors, this thesis focuses on the implementation of large arrays of SPADs and associated circuitry, implemented in CMOS. With the availability of these sensors, a number of scientific applications involving time-correlated imaging are enabled or highly improved, especially with respect to acquisition time. In a conventional camera a scene is projected onto the image plane, where a multi-pixel sensor captures its intensity. In a time-correlated or time-resolved camera each pixel evaluates the time-of-arrival of impinging photons, in addition to light intensity. Time-resolved optical imaging has many uses in other disciplines such as physics, molecular biology, and medical sciences, just to name a few. In particular, deep sub-nanosecond timing resolution, in combination with high sensitivity, is becoming increasingly important in a number of imaging methods, usually based on TCSPC. Time-resolved imagers can be employed to characterize the relaxation of specific molecules from an excited to a ground state, for example, by measuring the lifetime of photons emitted by the molecule. Specific applications in this domain range from fluorescence-based imaging, such as fluorescence lifetime imaging

microscopy (FLIM) [91], and fluorescence correlation spectroscopy (FCS) [92]. Other techniques used in biology and physics can also take advantage of time-resolved imaging. Such techniques are, for example, voltage sensitive dye (VSD) based imaging [93, 94], particle image velocimetry (PIV) [95], instantaneous gas imaging [96], etc.

Conveniently, potential implementations of time-correlated image sensors, based on CMOS SPADs, share a significant number of challenges involved in a fully-integrated SPAD-based 3D image sensor. Since these challenges are addressed in this thesis, future time-correlated image sensors may take advantage of a number of techniques and methods introduced here.

1.4 THESIS MOTIVATIONS AND CHALLENGES

At the beginning of this thesis, the largest array of SPADs reported in the art was a 32x32 array [18] based on a custom fabrication process. However, the approach chosen by the authors was to build a chip stack, comprising the SPAD-based image sensor chip and a CMOS circuitry chip, using a flip-chip bonding technique. As a result, every pixel on the sensor chip was interconnected to a front-end and readout circuit on the CMOS chip. This solution allowed the authors to mitigate the challenge of implementing in-pixel circuit so as to allow parallel imaging. Indeed, on the CMOS integrated circuit, the total silicon area underneath the pixels was available to implement circuitry to store the captured image and to provide means for readout off-chip. Alternatively, when it comes to realizing a true single-chip image sensor, the required in-pixel circuitry occupies silicon area and it therefore restricts the area available for the SPAD. In fact, the SPAD principle of operation worsens the situation and imposes an additional challenge. In order to clarify this statement, a comparison with conventional CMOS and CCD image sensors is convenient. On a CMOS/CCD conventional image sensor, photo-generated current is typically stored on the photodiode capacitance. As a result, only readout circuitry is required at the pixel level. This mode of operation is similar for p-n photodiodes and CCD photogates. In CMOS APS, the capacitance also includes the parasitic capacitance of the in-pixel readout transistors. The situation of a SPAD-based pixel is very different. Since the parasitic capacitance of a SPAD is fully discharged upon photon detection, it is necessary to add extra circuitry to store photo-generated information. For instance, it might be important to add a counter so as to store the information of how many photons the pixel has detected in a given time interval. Since the number of transistors required to implement a counter is by far higher than the number of transistors required at

pixel level for a CMOS APS image sensor, the design of single-chip image sensors based on SPADs involves very important challenges. In the approach of Aull *et al.*, this challenge is clearly mitigated. However, the approach involves more than a single chip solution and it strongly depends on the successful realization of extremely-high-yield interchip vias. Indeed, since each SPAD pixel in [18, 61] relies on a interchip via, it is very important that the via fabrication yield approach 100%. Furthermore, the parasitic capacitance introduced by the interchip via may dominate the SPAD capacitance in small devices. As described in Chapter 2, the parasitic capacitance reduction is a key achievement towards successful SPAD-based image sensor.

The work of my former colleagues on the realization of the first CMOS SPAD array [97] was a very important contribution to the art. Indeed, at that stage, clarification of device performance, especially with respect to crosstalk probability was reported. However, many challenges still persisted towards the realization of the first high-resolution fully-integrated image sensor based on CMOS SPADs. In particular, statistical evaluation of noise performance was incomplete due to the relatively low number of devices fabricated. In order to achieve a statistically relevant assessment of noise performance, it is necessary to realize and test a large number of devices on a single-chip. The size of the SPAD array in [97] consisted of only 8x4 SPAD pixels. Furthermore, the readout scheme implemented by Rochas *et al.* was based on simple multiplexing paradigm, thus leaving the challenge of scalable readout schemes unaddressed. Indeed, readout circuitry based on multiplexing does not allow parallel image acquisition for SPADs as it lacks storing elements to simultaneously save photo-generated information from all pixels. However, the quality of the work previously achieved by Rochas, Popovic, Besse, *et al.* on the design of SPADs in CMOS technology was a solid base for this thesis. Their personal encouragement and subsequent contributions have been highly appreciated.

In this thesis, the investigation of large arrays of SPADs was carried out for the first time. A general concern in this investigation is the utilization of scalable front-end and readout circuits in pixels that can take advantage of the progresses of CMOS technology. Furthermore, since SPAD signals are inherently of digital nature (see Chapter 2), this investigation is expected to lead to image sensors that exhibit high rejection of electrical noise, thus improving overall robustness.

This thesis is motivated by the need for a new generation of solid-state depth imagers that can achieve the accuracy, precision, and speed desired in a number of applications. Furthermore, cost, robustness, and a compact implementation are major concerns. CMOS system-on-a-chip realization is therefore of high-

importance. Millimeter accuracy and precision has been demonstrated in 3D imaging based on a single SPAD and scanner mechanisms. The application of the so-called TCSPC technique to all the pixels of a CMOS image sensor chip is therefore investigated in this thesis, which is expected to highly enhance the art in solid-state 3D imaging. However, in order to achieve this goal, important challenges at device, circuit, and system level need to be addressed.

Although the main focus of this thesis is not device research, it is essential to conceive and realize SPADs in advanced CMOS technology. At beginning of this investigation, the only CMOS SPADs available were based on an outdated $0.8\mu\text{m}$ CMOS technology. While SPAD performance was remarkable in that technology, that CMOS process only featured two metallic layers for interconnection and relatively low-speed transistors. In order to overcome these limitations, the evaluation of SPADs in newer CMOS processes is an imperative step towards the integration of complex digital circuitry on a single integrated circuit. As described in Chapter 2, the design and realization of SPADs on a new technology is very challenging. Edge breakdown prevention and compatibility between the high voltage requirement for the SPAD and the low supply voltage of the electronic circuitry are the main issues to be overcome in any technology migration.

As previously mentioned, a number of challenges to be addressed in this thesis are related to SPAD front-end and readout circuits. In addition, state-of the-art time-evaluation circuits needs to be designed and characterized on the same CMOS integrated circuit. These circuits need to achieve time-to-digital conversion with accuracy and precision of the order of the SPAD timing resolution, typically within a few tens of picoseconds. Although realizations of CMOS time-to-digital converters were reported in the art, typically these circuits were not optimized for silicon utilization area and measurement throughput. In most cases, an entire integrated circuit was devoted to a single time-to-digital converter (TDC). The most complex circuit in the art was able to provide 8 measurement channels with conversion rates within a few megasamples per second. In this thesis, in order to take advantage of TCSPC, much larger arrays of TDCs have been implemented. Furthermore, these TDCs are capable of providing superior measurement throughputs so as to prevent saturation due to background light. Despite the required specifications in terms of number of channels and measurement rate, the TDCs need to provide similar performance achieved in the art, when single devices with slow conversion times are reported.

System level challenges include the evaluation and design of extremely fast illumination sources to be used in the proposed TCSPC-based TOF method.

Furthermore, the feasibility study of the approach, which is based on extremely fast light sources, especially with respect to eye-safety concerns, is also an important milestone.

1.5 ORGANIZATION OF THE THESIS

In Chapter 2, the performance parameters of SPADs are first described. In particular, the section is intended to define some useful parameters utilized in the remaining Chapters of this thesis as well as in the review of the literature for SPADs, which is also presented. A number of new front-end and readout circuits especially conceived for SPADs are then introduced. Chapter 2 describes the design of SPADs in three different CMOS technologies. The devices are then characterized and discussed based on experimental results.

In Chapter 3, the principle of solid-state optical TOF 3D image sensors based on TCSPC technique is described. Operating assumptions in TCSPC are also discussed. A theoretical model for the evaluation of ranging performance is introduced. The model aims at validating design parameters in the context of reliable TOF evaluations, thus helping in the design of device, circuit, and system components. Moreover, the design of the first fully-integrated TCSPC image sensor in CMOS is reported in detail. Experimental evaluation of this sensor in a short range scenario, using an extremely low illumination power budget is reported. Finally, the experimental evaluation of a CMOS SPAD imager as a TCSPC rangefinder in a distance range of up to 70 meters is reported. In this distance range, a number of design, operation and safety concerns with respect to the illumination system exist. These issues are discussed and addressed in the same Chapter.

In Chapter 4, a new demodulation technique for solid-state range finding is introduced. The technique is called single-photon synchronous detection (SPSD) and it aims at decreasing the complexity in circuit implementation for single-photon pixels, with respect to the TCSPC approach. In the Chapter, the design and characterization of the first design based on SPSD is described. The sensor is also evaluated experimentally with respect to its range-finding performance.

This thesis is concluded in Chapter 5. The main achievements realized in the thesis are summarized. Finally, an outlook for future improvements is given.

2. SINGLE-PHOTON AVALANCHE DIODES IN CMOS TECHNOLOGY

2.1 INTRODUCTION

A p-n junction reverse biased above its breakdown voltage (V_{BD}) enables single-photon detection [98]. Reverse biasing a p-n junction beyond its breakdown voltage typically intrigues individuals who are familiar with the current-voltage characteristics of a typical p-n junction. The explanation of this principle requires a study of temporal evolution of the avalanche phenomenon. Let us assume a p-n junction in complete darkness. If one suddenly applies a reversed bias voltage above V_{BD} , there will be no breakdown current until a carrier enters the diode's depletion region. During this short period of time, a very high electric field exists within the p-n junction forming the avalanche's multiplication region. With the diode in this state, let us suppose now that a photon is absorbed in the depletion region. The photon absorption generates a primary electron-hole pair inside the high electric field region which in turn forces the electron to accelerate towards the cathode and the hole towards the anode. Possibly, one of those carriers acquires enough kinetic energy to generate secondary electron-hole pairs by impact ionization. In addition, this phenomenon may occur on primary and secondary carriers, which leads to a chain of impact ionizations, thus causing the diode's depletion capacitance to be rapidly discharged. As a result, a sharp current pulse is generated and may be easily measured. This effect is commonly known as avalanche multiplication.

Typical p-n junctions, e.g. as those used in conventional CMOS image sensors [99], are not compatible with this mode of operation since they do not support high voltage biasing without producing an instantaneous breakdown current. The main reason for this behavior is that the peak electric field is not planar in those devices over most of the junction area so as to form a large multiplication region. On the contrary, it is concentrated in a small area, typically in the diode's periphery, thus causing the junction to prematurely break down. As the diode breaks down prematurely, the electric field in the planar or multiplication area

does not reach the required value to provide sufficient impact ionization rate, thus preventing the device from operating in Geiger mode. Premature breakdown is prevented by utilizing specific avalanche photodiode structures instead of a simple p-n junction. Details on avalanche device engineering are provided in Section 2.3.

It is interesting to clearly state the difference between the specific single-photon avalanche diode (SPAD) term and the broad avalanche photodiode (APD) designation, which comprises different operating modes.

The term APD is commonly referred to devices that are reverse-biased slightly below V_{BD} and that operate in analog linear mode. This mode of operation implies linear conversion from optical power to electrical current and it is commonly applied to conventional photodiodes. However, since the reverse bias of a linear mode APD is near V_{BD} , it is sufficiently high to enable avalanche multiplication. The multiplication results in internal current gain, thus generally improving overall signal-to-noise ratio in optical detection. Nonetheless, feasible gains in linear mode are typically lower than 200, which are challenged by electronic noise. Nonetheless, some research groups have investigated the performance of low gain APDs in photon counting applications based on sub-Geiger mode [100, 101]. In this mode, the APD is used in conjunction with a low-noise charge-sensitive amplifier with a noise-equivalent charge of a few hundred electrons and a comparator with a threshold accurately set to a few times above the amplifier noise level. The APD is then capable of counting those photon-generated pulses for which the gain exceeds the threshold level.

Linear-mode APDs can respond to light modulated at frequencies as high as a few GHz. As a result, they are generally used in applications involving simultaneously high-frequency and low-intensity light signals. For example, applications for APDs include optical fiber communication, spatial light transmission, general low-light-level detection, laser radar, and biomedical devices [102]. In addition, silicon APD-based image sensors implemented in standard fabrication processes have been investigated to provide high sensitivity in low-light imaging applications [103]. Nevertheless, excess noise generated by the statistical fluctuation of finite multiplication gain, unsatisfactory sensitivity, high power consumption, and large in-pixel circuits still prevent APD image sensors to favorably compete with state-of-the-art CMOS and CCD imagers. Lately, the sub-Geiger mode has been revisited [104, 105]. In particular, reference [104] proposes a AlInAs/GaInAs APD that may be integrated with a low-noise ultra-fast readout circuit, thus potentially enabling single-photon counting rates up to the gigahertz range. Unfortunately, the approach relies on a high-cost non-standard fabrication process. In addition,

linear mode requires analog circuits at pixel level that are difficult to miniaturize and typically consume high quiescent electrical power [103]. As a result, the approach is unlikely to enable large arrays of pixels, the focus of this thesis.

The term SPAD is used to designate an avalanche photodiode that is biased far above V_{BD} and that operates in the so-called Geiger mode. In the literature, SPADs have also been referred to as triggered avalanche detectors or Geiger-mode avalanche photodiodes. In Geiger mode, the magnitude of the electric field in the multiplication region is of the order of a few hundreds of kV/cm, which engenders a sufficient internal gain to operate without any additional electronic amplification circuit. These devices therefore may be triggered by a single photon and generate a signal with a few Volts of amplitude. The photon-generated signal exhibits timing properties that are of high relevance in the framework of this thesis.

Unless otherwise specified, the term APD will be utilized in this work to specify linear-mode APDs whereas Geiger-mode avalanche photodiodes will be referred to as SPADs.

Physically, the structure of an APD may be very similar to a SPAD. In many cases, an APD designed to operate in linear mode may be properly biased to successfully operate as a SPAD. This technique has been utilized in practice for many years, especially in scientific applications, mostly due to limited availability of specifically designed SPADs. Only recently, commercial available SPAD modules have appeared [106-108] and may be ordered off-the-shelf.

In Section 2.2 the performance parameters of SPADs are described. In particular, the section is intended to define some useful parameters utilized in more detail in Section 2.3. A succinct review of SPAD literature is presented in chronological order. Priority is given to detectors that are relevant to this thesis. In Sections 2.5, 2.6, and 2.7, actual implementations of SPADs in three different CMOS technologies are presented and discussed with the support of experimental results. Finally, in Section 2.8, this Chapter is concluded with a comparison between the devices reported in this thesis. An outlook on the future of SPAD technology in CMOS is also given.

2.2 PERFORMANCE PARAMETERS

Similarly to any optical detector, a SPAD presents non idealities that may have an important impact on the single-photon applications. In this Section, the parameters that characterize SPADs are described. These parameters are then discussed in the context of device design in Section 2.3.

2.2.1 PHOTON DETECTION PROBABILITY

At the risk of stating the obvious, the most important property of a photodiode is given by its ability to detect optical power. Typically, linear mode photodiodes are characterized by a responsivity given in Ampere per Watt. In other words, it expresses the photodiode sensitivity by giving how much photocurrent is generated as a function of incident optical power. Since photodiode response strongly depends on the spectral contents of the input signal, most of photodetector suppliers provide data specifying the responsivity as a function of optical wavelength in the spectral range of interest.

Alternatively, the optical sensitivity of a detector may be expressed by means of the so-called quantum efficiency (QE). This sensitivity measure exploits a more fundamental discrete notion in which one wants to specify measurable quantities in terms of elementary particles: electron/hole for current and photon for optical power. Simply stated, QE indicates the mean percentage of photons hitting the photodiode's active area that will produce an electron-hole pair.

APDs, biased just below V_{BD} , operate in linear mode and have a finite multiplication gain. Due to this inherent internal gain, optical sensitivity in APDs is enhanced, as a single photon may generate tens of electron-hole pairs. The mean avalanche gain is therefore an additional parameter for APDs. Accordingly, the sensitivity of APDs is typically reported as responsivity with unity gain or simply as QE.

As briefly mentioned in the previous Section, APDs suffer from an excess noise source due to the statistical fluctuation of a finite multiplication gain. SPADs, on the other hand, are not concerned with these gain fluctuations since only photons that give rise to a very high number of carriers are counted as a logic signal. Nevertheless, the statistical variation of the avalanche buildup is translated onto a detection probability. Indeed, when a photon is absorbed in the SPAD's depletion region, an electron-hole pair is generated in the electric field. Both the electron and the hole are accelerated and may generate

secondary carriers that in turn are also accelerated. Eventually, the multiplication chain leads to a full junction discharge which triggers detection. Nonetheless, there is still a probability that the multiplication chain extinguishes by itself, especially in the beginning of the process. In this case, the photon absorption is simply missed. It is therefore important to rely on the notion of *photon detection probability* (PDP) as sensitivity measure in single-photon detectors. PDP is simply defined as the probability of obtaining a successful detection when a photon hits the SPAD's active area. Conveniently, the definition of PDP also relies on fundamental discrete quantities.

Similarly to responsivity and QE, PDP depends strongly on photon wavelength. Furthermore, besides the non-ideal internal amplification mechanism of SPADs, there are a number of loss mechanisms that may affect PDP. For instance, for a photon to generate a primary electron-hole pair, it needs to be absorbed ideally in the depletion region of the SPAD. However, there is a probability that it is absorbed perhaps too deep in the chip substrate or that it is reflected back even before reaching the semiconductor material. Assuming a silicon SPAD, the losses an incident optical signal may face prior to generating a successful detection are mainly due to (a) parasitic reflections, (b) ineffective absorption, and (c) non-ideal breakdown initiation probability.

Parasitic reflections of incident photons occur at the interface between air and the passivation layer of the integrated circuit due to the mismatch in refractive index. In addition, possible multi-reflections may also arise within the sub-layers forming the passivation and oxide on top of the integrated circuit. Another reflection loss occurs at the interface of the oxide layer and the silicon substrate. These sources of light loss are commonly present in any image sensor and may be addressed, to some extension, by means of antireflection coatings.

Ineffective absorptions are represented by light absorption within the passivation and oxide layers due to non-ideal transparency. Moreover, absorption of photons at the silicon surface is inefficient. Photogenerated carriers in the vicinity of the surface have extremely short life times and therefore are unable to diffuse to the avalanche multiplication region. Finally, absorption of incident photons beyond the multiplication region at a depth distance larger than the diffusion length does not lead to successful detection.

Assuming that a primary carrier is generated in the depletion region, an avalanche breakdown occurs with a probability that depends on the SPAD structure, internal electric field, and carrier type. This probability, defined as breakdown initiation probability, is obviously not ideal and results in a lower

overall photon detection probability. Furthermore, breakdown initiation probability is lower for holes than for electrons.

A detailed description and analytical calculation of PDP for SPADs may be found in [20, 55]. Furthermore, some insights on how to optimize detection efficiency are reported in Section 2.3.

2.2.2 DARK COUNT RATE

Unlike linear mode APDs, in SPADs, the signal amplitude does not provide intensity information since a current pulse has the same amplitude whether it had been triggered by a single or multiple photons. Intensity information is obtained by counting the pulses during a certain period of time or by measuring the mean time interval between successive pulses. The same principle may be used to evaluate noise. Thermally or tunneling generated carriers within the p-n junction, which generate dark current in linear mode, can trigger avalanche pulses [109]. In Geiger mode, they are indistinguishable from regular photon-triggered pulses and they produce spurious pulses at a frequency known as *dark count rate* (DCR). DCR is a very important parameter since it defines the noise level in the dark, thus limiting the detection dynamic range from the low end. Since the mean value of DCR is systematic and may thus be subtracted from every pixel in some applications, the ultimate noise contribution arises from the time-varying component of DCR.

Analytical evaluation of DCR may be described by calculating the rate of carriers in the quasi-neutral and depletion regions of the p-n junction that attain the SPAD's multiplication region. In silicon, under strong reverse bias and at usual temperatures, the generation rate in the depletion region dominates DCR due to diffusion current in quasi-neutral regions. The cause is the small value of the intrinsic concentration n_i of silicon [110]. The generation rate of carriers in the depletion region is given by (a) thermal generation and (b) generation due to tunneling effect.

Thermal generation-recombination, firstly described by the model of Shockley-Read-Hall [111], occurs whenever the thermal-equilibrium condition of a structure is disturbed, that is, $p \cdot n \neq n_i$, where p and n are the local hole and electron densities, respectively. Mechanisms exist to restore the semiconductor structure to equilibrium. In silicon, assuming usual temperature ranges, these processes involve defects in the semiconductor that act as donors or acceptors. These defects have energy levels close to the center of the band gap of silicon and are referred to as generation-recombination (G-R) centers. The generation of an electron-hole pair, that is, the transition of an electron from the valence

band to the conduction band, involves the emission of a hole to the G-R center level, followed by the emission of an electron from the G-R center to the conduction band. The recombination process is the opposite, that is, the capture of an electron from the conduction band followed by the capture of the hole from the G-R center to the valance band. Based on these processes, it is possible to relate the resulting generation rate in the depletion region as a function of important parameters such as temperature, G-R center density and energy levels [110, 111]. As its name suggests, thermal generation rate depends strongly on temperature.

Electron-hole pair generation due to tunneling effects occurs in SPADs when the doping concentration of both p and n sides of the junction are very high. The depletion region becomes extremely narrow and there is a significant probability that band-to-band generation occurs, even at low temperatures. Tunneling effect in a p-n junction depends strongly on the electric field [112], which is particularly high in a SPAD. As a result, an additional generation rate component should be added to the existing SRH theory.

Furthermore, tunneling may also enhance thermal generation by favoring additional electronic transitions such as emission of hole to a G-R center, followed by a tunneling transition to the conduction band. In reference [113] a model is proposed to introduce the effect of tunneling in the SRH theory.

A detailed calculation of DCR based on generation rates in the depletion region is reported in [20]. DCR typically depends strongly on temperature and on the SPAD bias voltage. DCR may be improved by reducing the temperature of the SPAD chip by means, for instance, of a thermoelectric (TE) cooler device. Nonetheless, as it is discussed hereafter and in Section 2.3, this technique has some limitations, especially in structures that are prone to tunneling effects.

2.2.3 DEAD TIME

As previously mentioned, SPADs are biased above V_{BD} . The overdrive voltage with respect to V_{BD} is known as excess bias voltage or simply excess voltage (V_E).

When a Geiger discharge occurs, a front-end circuit senses the avalanche current so as to indicate an event detection. In addition, the same circuit additionally lowers the SPAD bias voltage to stop the avalanche. This function, known as avalanche quenching, is followed by a recharge phase that brings the SPAD to the same bias condition that it had prior to the Geiger event. As a result, the SPAD recovers its photon detection ability some time after each Geiger

event. The time it stays under a lower bias condition, thus unable to detect photons, is defined as *dead time* (DT).

Quenching and recharge circuits are discussed in detail in Section 2.4. Quenching circuits based on a simple resistor or equivalent component are referred to as passive quenching (PQ) circuits. On the other hand, quenching circuits that react on the SPAD biasing by means of active devices are defined as active quenching (AQ) circuits. Similarly, recharge circuits may be defined as passive recharge (PR) or active recharge (AR) depending on the actual implementation.

Dead time includes quenching and recharge times. Depending on the type of front-end circuit, the recharge phase may exhibit weakly defined edges. In the context of this thesis, dead time is precisely defined from the moment a photon that triggers an event is absorbed until the moment the SPAD recovers 90% of its nominal V_E . This definition, which was adopted in [20], may be changed to be more restrictive in applications requiring extremely high timing accuracy.

Dead time should be as low as possible as it restricts the maximal rate of photon detections. In imaging applications, it limits therefore the dynamic range of the image sensor. Even though dead time is not a direct property of SPAD, it cannot be arbitrarily reduced by means of fast circuits. As it is discussed in Section 2.2.4, dead time has an impact on afterpulsing probability. As a result, the lower limit for dead time is very often given by the SPAD cleanliness.

2.2.4 AFTERPULSING PROBABILITY

In Section 2.2.2, defects in the semiconductor crystal, due to impurities added intentionally or involuntarily during fabrication, were described to engender G-R centers with energy levels in the middle of the forbidden band. In fact, impurities in silicon may engender many sorts of trapping centers having energy levels in the forbidden band ranging from the vicinity of valence up to the conduction band [110].

In a SPAD when a Geiger discharge occurs, a very large number of carriers traverse the junction and some of them may be captured by trapping centers. When a captured electron or hole is reemitted in the multiplication region after dead time from a trap, there is a probability that another Geiger event is triggered [109, 114]. Since the later event is correlated in time with the original Geiger discharge, it is known as afterpulse. Afterpulses lead consequently to another type of unwanted count.

Afterpulsing performance in a SPAD is characterized by the probability of generating, in average, an afterpulse after a photon-triggered or original dark event, namely *afterpulsing probability*.

Afterpulsing probability depends on the density of deep-level trapping centers, that is, centers that have energy levels near the center of the bandgap. Traps with energy levels close to edges of the forbidden bands have short lifetimes, that is, they tend to reemit carriers shortly after the capture process. As a result, they marginally contribute to afterpulsing since carriers are likely to be reemitted during dead time. Deep-level traps, on the contrary, tend to reemit carriers after the dead time and as consequence, they typically induce afterpulses. One may note that, from an energetic stand point, deep-level traps are indistinguishable from G-R centers. Nonetheless, a distinction is commonly made: a trap is a center whose probability of reemitting the carrier is much higher than the probability of inducing a recombination with a carrier of opposite polarity.

In Section 2.3, afterpulsing performance is discussed in the context of device and circuit design.

2.2.5 TIMING RESOLUTION

Another important feature of SPADs is the ability of accurately detecting the arrival time of photons [98]. The statistical fluctuation of the time interval between the arrival of photon at the SPAD and the leading edge of the output pulse is defined as the timing jitter or *timing resolution*. In this definition, the term resolution actually stands for the statistical spreading instead of actual resolution. This designation, which may lead to confusion in information theory, is commonly used in the literature and it is therefore retained in this thesis. Timing resolution is typically reported as the full width at half maximum (FWHM) of the photon arrival time distribution.

In a SPAD, timing resolution typically exhibits two timing components: (a) a narrow distribution and (b) a low-amplitude tail given by one or more exponential decay components.

The narrow distribution around the actual photon arrival time may be approximated by a Gaussian distribution. It is mostly related to photons that are absorbed in the SPAD's depletion region. The lower theoretical limit for timing resolution is given by the carrier transit delay from the absorption point in the depletion layer up to the multiplication region. At saturation velocity, it takes approximately 10ps per micrometer of depth. In addition, the fluctuation of the

avalanche buildup also broadens timing resolution. This contribution increases with ionization coefficients of electrons and holes, which in turn may be enhanced by increasing the SPAD bias voltage. As a result, high V_E is desirable to achieve good timing resolution. The avalanche buildup propagates laterally over the whole junction area. In large SPADs, the propagation time of the avalanche is significant and it may carry large timing fluctuations. This propagation results in a timing resolution that depends on the photon absorption location, the SPAD's center leading to the fastest timing response [115]. Since the energy involved in the breakdown discharge may be high in large devices due to parasitic capacitance, the lateral propagation of the avalanche is enhanced by photons emitted by hot carriers [116]. These photons are absorbed at different locations in the SPAD depletion region and initiate secondary avalanches, thus accelerating the lateral propagation of the discharge. This mechanism, when present, additionally reduces the dependence of timing resolution on the photon absorption location.

The slow tail typically present in timing resolution is due to incident photons that are absorbed in quasi-neutral regions of the p-n junction [117]. Primary carriers in those regions need to diffuse to the SPAD depletion region in order to trigger an avalanche event. The longer the distance from the origin of the carriers to the depletion region, the longer is the mean value and the fluctuation of the diffusion delay. Since diffusion is a slow process, the typical timing resolution tail is orders of magnitude longer than the narrow peak induced by the effects described in the last paragraph. In addition, since the photon absorption point is statistically distributed in depth, the response tail does not exhibit a single exponential decay component, unless a proper design limits the quasi-neutral regions thickness (see section 2.3). As a result, the typical tail response may be only fitted by the sum of many exponential decay components and it will depend on the incident photon wavelength.

Finally, front-end circuits add additional Gaussian electronic jitter that is eventually convolved with the SPAD's timing resolution. Modeling and simulation of SPAD timing resolution are reported, for instance, in references [117, 118].

2.2.6 CROSSTALK PROBABILITY

Similarly to the CMOS and CCD image sensing technologies, it is desirable to take advantage of large arrays of SPADs implemented on a single integrated circuit for imaging applications. Usually, image sensors suffer from an unwanted effect known as crosstalk. Assuming an imaging lens that produce a sharp image

on the sensor plane, the photons absorbed through the area of a pixel are expected to contribute only to the signal of that pixel. In practice, a fraction of those photons are detected by neighboring pixels, thus inducing the so-called crosstalk effect. Crosstalk reduces the effective spatial resolution of an image sensor, leading to blur images. Crosstalk effects are similarly present in array implementations of SPADs. One may distinguish two mechanisms affecting the performance of a SPAD-based image sensor: (a) electrical crosstalk and (b) optical crosstalk.

Electrical crosstalk is due to photons absorbed deeply under the pixel area. Photogenerated carriers in the quasi-neutral region below the p-n junction might diffuse laterally and trigger detection on a neighbor pixel. Since the mean penetration depth of photons depends strongly on the photon wavelength, photons in the red and near-infrared ranges tend to induce more crosstalk than short-wavelength photons. Fortunately, electrical crosstalk may be effectively addressed by proper design techniques, according to Section 2.3.

Optical crosstalk in SPAD arrays is based on an effect that is not present in conventional image sensor owing to their operating mode. As briefly mentioned in Section 2.2.5, during an avalanche some photons may be emitted from a SPAD, thus leading to an electro-luminescence phenomenon [119]. These photons may be detected by neighboring pixels in an array of SPADs and cause unwanted counts [120]. In [55], Kindt reviewed the physical mechanisms of electro-luminescence in SPADs and reported experimental results of crosstalk in small arrays. The author has adopted a solution in which isolation trenches were etched around each SPAD and covered with metal so as to prevent that photons emitted from one pixel reach neighbor pixels. Optical crosstalk is addressed in this thesis by significantly reducing electro-luminescence in SPADs by taking advantage of CMOS technology. This solution was adopted and verified by Rochas *et al.* at low pixel densities [97].

Crosstalk performance is characterized by the *crosstalk probability*, which is defined as the probability that a Geiger event is produced as a consequence of a photon absorption in a neighbor pixel. In order to simplify performance evaluation, crosstalk probability includes both electrical and optical effects. Note that electro-luminescence in SPADs does not depend on incident photon wavelength. As a result, optical crosstalk behaves differently than electrical crosstalk when SPADs are tested under various incident light wavelengths. Furthermore, optical crosstalk may be conveniently evaluated by measuring signal correlation between two adjacent pixels, whereas electrical crosstalk usually requires more sophisticated optical testing.

2.3 DEVICE ENGINEERING

In this Section, a discussion on the design and fabrication of SPADs in CMOS technology is described. A review on state-of-the-art of silicon SPADs is reported. In this analysis, device structures of greatest interest towards the integration of large arrays of SPADs in CMOS technology are reviewed. Performance data is reported assuming room temperature, otherwise the temperature is specified.

2.3.1 STATE-OF-THE-ART

Even though the potential of SPADs in the detection of optical power down to a single photon became practical in early 1980s [98], single-photon detection by means of avalanching p-n junctions was observed very early in the microelectronics era [109, 121-123]. Haitz *et al.* made very important contributions in the comprehension of avalanche photodiodes. In particular, they have described the physical phenomena in the generation of current pulses in Geiger mode, and the mechanisms leading to dark counts and to afterpulses. An accurate electrical model of SPADs was also proposed in [123] and in spite of the advances made in the technology over more than 40 years after that work, the proposed model may still be used nowadays to simulate fundamental electrical behavior in Geiger mode.

Evidently, silicon technology was too premature at that time and it was therefore very difficult to implement p-n junctions free of the so-called microplasmas. Microplasmas are material defects such as dislocations and metal precipitates in the p-n junction that induce non-homogeneities in the electric field. As a result, the breakdown current in a p-n junction containing a microplasma concentrates in a small spot instead of flowing uniformly through the whole junction area [124]. Microplasmas may be easily identified by inspecting the reverse biased p-n junction under a microscope owing to electroluminescence effect of hot-carriers.

The solution found by Haitz *et al.* to build defect-free p-n junctions was to fabricate a large number of small devices on a silicon wafer and to choose one that possibly featured a planar avalanche breakdown over the whole junction area. Furthermore, it had been predicted in [125] that in a defect-free p-n junction, the most probable location for field concentration was at the junction periphery. As a result, Haitz *et al.* surrounded the p-n junction with a diffusion guard-ring. Figure 2.1 shows a cross-section of the device implemented by Haitz with the intention of studying avalanche multiplication effects. Since the device

was not intended to be used as a photodetector, the authors did not optimize the structure to achieve the performance desired in current SPAD implementations. The device consisted of a circular highly doped n-type cathode in a p-type silicon substrate. An n-type guard-ring was diffused to surround the circular cathode with a deeper n-type layer having a lower doping concentration. Since the doping concentration was lower in the guard-ring, the peripheral p_{SUB}-n_{GR} junction exhibited a wider depletion region which, in turn, reduced its internal electrical field. As a result, the highest electric field was formed as a disk in the central region of the device, thus leading to a planar multiplication region.

Probably unknown at the time Haitz *et al.* implemented this device, the n-type diffusion guard-ring also provided an additional function. Deep-level traps, caused typically by metal contamination during device processing, increase dark count rate and afterpulsing probability in SPADs (see Section 2.2.4). In fact, it has been demonstrated later that phosphorous diffusion has efficient gettering properties. In SPADs, gettering occurs when a given location outside the high field region favors the formation of metal precipitates, thus keeping metal atoms distant from the high field region. As a result, the n-type guard-ring has a beneficial effect in dark count rate and afterpulsing performance. In [55] and [126], review of gettering techniques for SPADs and of the physical mechanisms of gettering in silicon are presented respectively.

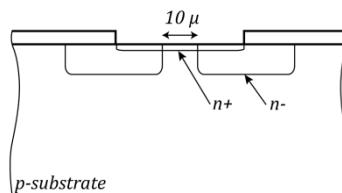


Figure 2.1. Cross-section illustration of the avalanche photodiode introduced by Haitz *et al.* Diffusion guard rings have been used to prevent edge breakdown.

The device of Haitz exhibited very important characteristics: firstly, it was based on shallow and highly doped cathode that leads to a low voltage device, i.e. a few tens of Volts; and secondly, it was fabricated using a silicon planar technology. These characteristics were already in line with a possible integration using nowadays CMOS technology. A low-voltage operation is very important to increase reliability, to reduce power consumption, and it favors the use of CMOS compatible transistors.

A different APD structure was proposed by McIntyre *et al.* [127-129], namely reach-through structure. Figure 2.2 shows a cross-section of the device and an idealized internal electric field in the center of the device as a function of depth. Edge breakdown is prevented by enhancing the electric field in the center of the APD by means of a p-type implantation. Under a sufficiently high reverse bias, the APD is characterized by an electric field that reaches the p+ layer at the bottom of the device through the high resistivity π region. As a result, a wide drift region is established which forms the photosensitive region. The device is exposed to the incident optical power via the back-side of the chip. Compared to the shallow structure of Haitz *et al.*, the reach-through structure exhibits some advantages. Assuming that the π layer thickness is larger than the light absorption length, the reach-through structure ensures that all the photogenerated carriers reaching the multiplication region, at the top p-n junction, are electrons, thus reducing multiplication noise in silicon devices. Indeed, the fluctuation of the avalanche gain in silicon devices, where ionization coefficient for electrons is higher than that for holes, is minimized when carrier injection into the multiplication region is limited to electrons only [110]. Another important advantage of reach-through APDs is that they exhibit high detection efficiency owing to their wide absorption region. Moreover, detection efficiency in those devices does not depend on incident optical wavelength, at least over a large spectrum range [128].

These advantages motivated the development of the first commercially available SPAD module to be based on a reach-through structure [106]. Nevertheless, reach-through devices have important drawbacks when compared to the devices of Haitz *et al.*: Firstly, the wide depletion region of reach-through SPADs forces their operating reverse bias to be extremely high, typically of the order of hundreds Volts. Each photon detection involves an electrical discharge whose peak power may be as high as 10 Watt [130], thus dissipating a large amount of heat. As a result, those devices need special packaging, typically based on thermo-electric coolers, so as to maintain the operating temperature controlled. They typically present poor reliability under strong illumination [20] and are therefore not suitable for a number of applications, especially in mass production. Fabrication yield and consequently cost are additional concerns as device fabrication requires a dedicated process based on special ultra-pure and ultra-high-resistivity silicon wafers and a non-planar technology [130]. These requirements, combined with high voltage operation, prevent co-integration between front-end circuits and reach-through SPADs on a single integrated circuit.

Single-photon detection with silicon SPADs gained interest in the scientific community when Cova *et al.* demonstrated the potential of SPADs for optical time-correlated applications with picosecond resolution [98]. In that work, a SPAD based on the structure of Haitz *et al.* was utilized. In fact, the main breakthrough was not the photodetector device but the use of a new ancillary circuit: the active quenching circuit. This circuit, which had been introduced by the same authors in [131], made SPADs a very effective replacement for photo-multiplier tubes (PMTs) and micro-channel plates (MCPs), so far the detector of choice in optical time-correlated applications.

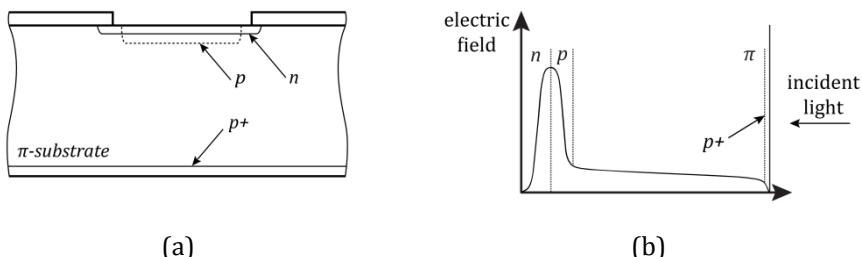


Figure 2.2. (a) Cross-section of a reach-through APD proposed by McIntyre *et al.* (b) Idealized internal electric field as a function of depth.

Acknowledging the benefits of the shallow structure of Haitz *et al.* over the reach-through SPAD, Cova *et al.* have selectively improved timing resolution performance of silicon SPADs by proposing advances in the shallow device structure [132-134]. Figure 2.3 depicts the cross-section of device structures introduced by the authors.

In reference [132], Ghioni, Cova *et al.* introduced a SPAD that takes advantage of an epitaxial p-epistrate grown on a n-type wafer, shown in Figure 2.3 (a). This SPAD structure was very similar to the device of Haitz. Nonetheless, the introduction of a carefully grown p-epistrate resulted in better performance. First, the purity of the p-epistrate was higher than that of the original wafer. As a result, good noise performance was obtained from reasonable-quality silicon wafers. Secondly, since the thickness of the epilayer was a design parameter, it became possible to reduce the effective thickness of the quasi-neutral region below the SPAD, thus reducing the contribution in timing resolution due to diffusion time of minority carriers in the p-epistrate. Nonetheless, the later optimization was not perfect since the epilayer thickness could not be reduced arbitrarily. Edge-breakdown was prevented by means of n-type diffusion guard-rings. As a result, the epilayer thickness had to be large enough to accommodate

the guard-ring such that it is still isolated from the n-substrate. In fact, the thickness of the p-type layer between the guard-ring and the n-substrate had to be large as it added an extra parasitic series resistance between the SPAD anode and its terminal contact. The authors slightly mitigated this constraint by reducing the thickness of the diffusion guard-ring. In order to effectively prevent edge-breakdown with a shallower guard-ring, they locally enhanced the electric field in the center region of the SPAD by adding a p+ implantation, similarly to the device from McIntyre *et al.* In addition, the gettering action of phosphorus diffusion was retained in this structure as the authors added a deep phosphorus implantation around each device. Cova *et al.* have implemented a number of devices based on this structure, whose breakdown voltage were typically a few tens of Volts. In [135], the authors focused on timing resolution performance. They fabricated a device based on highly doped layers, thus leading to a breakdown voltage of 13V. This device demonstrated an excellent timing resolution of 20ps, measured at a low temperature of -65°C and with V_E of 6V. It also exhibited a reduced diffusion tail when compared to the SPAD of reference [98]. Nonetheless, the excessive optimization in timing resolution via highly doped implantations leads to high DCR, exceeding 100kHz with V_E of 6V.

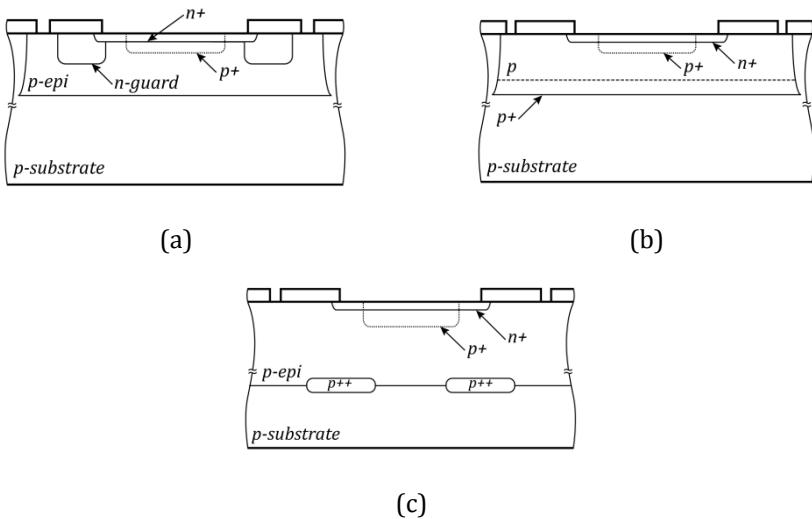


Figure 2.3. Device cross-section of : (a) epitaxial SPAD, (b) double-epitaxial SPAD, and (c) dual-junction epitaxial SPAD.

In Figure 2.3 (b), the cross-section of another device structure, introduced by Lacaita, Cova *et al.*, is shown [133]. This structure, called double epitaxial SPAD, solved most of the problems existing in the previous generation. First, the n-diffusion guard-ring was completely removed. As a substitute, the authors improved the n+/p+ enrichment in the p-epistrate, thus preventing edge breakdown. This modification also improved the field homogeneity in the multiplication region. Previously, the effective doping concentration of the anode was lower at the periphery of the device due to the poorly defined diffusion guard-ring. As a result, the electric field was not constant over the device area, thus leading to a non-uniform photon detection probability [130]. Secondly, the parasitic series resistance at the anode was diminished by utilizing two epitaxial layers. On top of the p-substrate, a highly doped p+ was grown, followed by the growth of a lightly doped p-epistrate. As a result, the resistance from the center of the device to the surrounding anode p+ contact was fairly reduced. This structure enabled the designers to theoretically freely adapt the quasi-neutral layer thickness below the junction, i.e. the low doped p-epistrate, so as to optimize timing resolution. The effective quasi-neutral region responsible for the diffusion tail in timing resolution was limited to a few microns in thickness. As a result, typical diffusion tail comprising many exponential decay components was virtually reduced to a single component. Double-epitaxial SPADs featuring excellent timing resolution have been reported by the authors for several years [133, 136-139]. In the original work [133], the timing resolution was 45ps. Afterwards, this device structure was retained by the authors to implement large area SPAD equally featuring excellent timing performance. In [136] and [138], SPADs with 100 μ m and 200 μ m in diameter are presented, respectively. Both devices exhibited 35ps FWHM of timing resolution. Remaining performance parameters based on this device structure are outstanding as well. Photon detection probability up to 45% was reported at V_E of 5V. Low dark count rates, typically a few hundred Hertz and one kilohertz at V_E of 5V for devices of 20 μ m and 50 μ m, respectively, were reported.

In order to ultimately remove the residual diffusion tail in timing resolution, Lacaita, Cova *et al.* introduced a highly sophisticated device structure [134], depicted in Figure 2.3 (c). This device, namely dual-junction SPAD, was based on a complex and potentially low yield fabrication process [130]. The basic idea was to utilize the n-substrate to form a junction with the low-doped p-epistrate below the SPAD to collect all the carriers that would result in diffusion time. The p-epistrate thickness and second junction reverse bias was adjusted to allow the quasi-neutral layer between the n+/p+ and p-epistrate/n-substrate junctions to be fully depleted. In order to implement a sufficient wide second

junction, a ring-shaped p+ layer was implanted on the n-substrate prior to the growth of the p-epistate. Similarly to the previous structure, the function of the buried p+ layer was to reduce the anode contact resistance. The original device, with 8 μm in diameter, showed 45ps FWHM in timing resolution without the presence of the typical diffusing tail. In [140], the same structure was revisited exhibiting 35ps FWHM in timing resolution. Nonetheless, this device structure was not viable in large devices due to the complex process involved. As a result, the authors turned to the previous structure to implement large devices and, more recently, small arrays of SPADs [139, 141].

A number of similar silicon SPADs based on a p-epistate and on central field enhancement have been proposed by other researchers. Kindt *et al.* have investigated monolithic integration of arrays of silicon SPAD for the first time [55, 142]. The final device structure retained by the authors is shown in Figure 2.4 (a). It is relatively similar to the double epitaxial SPAD of Cova *et al.*, except that Kindt has utilized a p-substrate wafer. Based on this structure, the authors have implemented small arrays of SPADs for characterization purposes. Individual detection and noise performance of their device was similar to those also reported by Cova et al [55]. Nonetheless, Kindt had reported high values of crosstalk probability between pixels. Afterwards, the authors introduced isolation trenches between pixels, covered with metal, so as to reduce optical crosstalk effects. This approach resulted in better crosstalk performance. However, the remaining crosstalk probability of 1% and enlarged pixel size due to isolation trenches discouraged the authors to continue the approach.

Aull *et al.*, some years after Kindt, revisited the silicon SPAD array approach [18, 61]. However, the authors considered hybrid integration of a stack of two and more silicon wafers. On the top wafer, an array of back-illuminated SPADs was implemented. Special process steps were developed to allow the realization of interchip vias. These vias were then used to connect each SPAD pixel to a measuring circuit, built in CMOS technology. The SPAD structure, shown in Figure 2.4 (b), was based on a variation of the double-epitaxial device. The p+ layer, used to enhance the field in the SPAD active region, was implanted deeper in the p-epistate. The device region area was composed of an n+/p-epi/p+/p-epi/p+substrate structure. The lower p-epilayer is where the photons are absorbed when the device is illuminated from the p+substrate side. Reverse biasing at the proper operating voltage establishes a modest electric field in this photon absorption layer that causes photogenerated electrons to drift up into the upper p-epilayer. The field in the upper p-epilayer is much stronger, responsible for avalanche multiplication. Surprisingly, the breakdown voltage of the device was reasonably small, approximately 20V. PDP as high as 60%, DCR

below 1kHz and timing resolution of 150ps FWHM were reported. Moreover, crosstalk probability assumed to be lower than 1% in arrays of 32x32, and recently 64x64 SPADs, was reported. Crosstalk performance was significantly superior when compared to the arrays without isolation trenches, built by Kindt. A possible explanation for this behavior is a potential lower parasitic capacitance between SPADs and ancillary circuitry due to the compact inter-chip via, in the devices of Aull. When the parasitic capacitance is small, a reduced amount of energy is dissipated during a Geiger discharge. As a result, the number of photons emitted by internal electro-luminescence effects is reduced, thus decreasing optical crosstalk. This characteristic essentially enables monolithic implementation of arrays of SPADs and it has been of most importance in this thesis.

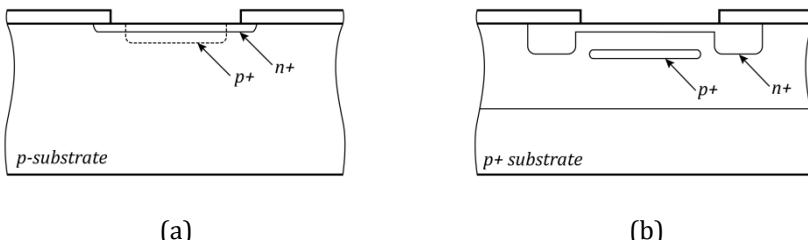


Figure 2.4. Cross-section of the devices implemented (a) by Kindt *et al.* and (b) by Aull *et al.*

An important breakthrough in silicon single-photon detection was achieved when Rochas, Popovic, *et al.* successfully integrated SPADs and front-end circuits on the same integrated circuit. In particular, the device and circuit design was based on commercially available 0.8 μ m CMOS technologies [19, 20, 97, 143-146]. Monolithic integration of SPADs and circuits leads to important performance improvements. As described in Section 2.2, a number of performance parameters strongly depend on the parasitic capacitance seen by the SPAD. For instance, afterpulsing probability depends on the density of deep-level traps in the multiplication region and on the number of carriers generated during an avalanche, which populate those traps. Trap density depends mostly on the cleanliness of the CMOS process, which cannot be modified by design. However, the number of carriers generated during an avalanche may be reduced to a minimum by decreasing the parasitic capacitance introduced by the front-end circuit. In aforementioned hybrid approaches (e.g. by McIntyre *et al.*, Cova *et al.*, Kindt *et al.*, and Aull *et al.*), in which the SPAD and front-end circuit are not co-integrated on the same IC, extra parasitic capacitance of the

order of a few picofarad are added to the total capacitance seen by the SPAD. As a result, the number of carriers generated in an avalanche in the hybrid approach is at least one order of magnitude larger than the fully integrated one. As briefly mentioned, another negative effect of large parasitic capacitance is crosstalk probability. Low parasitic capacitance reduces the total energy dissipated at each photon detection, thus reducing crosstalk probability, power consumption, and keeping device temperature low at high count rates. In particular, these improvements are more evident for small SPADs. The junction capacitance of SPAD with $10\mu\text{m}$ in diameter may be as low as a few tens of femtofarads, especially when a proper biasing is chosen (see Section 2.4).

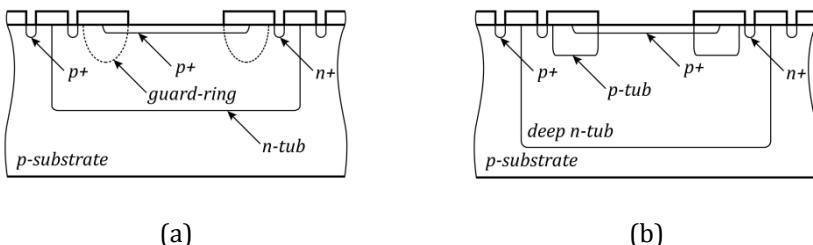


Figure 2.5. Cross-section of SPAD from Rochas *et al.*: (a) guard-ring based on lateral diffusion of two n-wells, (b) guard-ring based on p-well.

The first challenge solved by Rochas was to identify a device structure based on layers that are available in the $0.8\mu\text{m}$ CMOS technology utilized. A conventional n-well CMOS process is based on a p-substrate used as the bulk for NMOS transistors, and features an n-well layer for the bulk of PMOS transistors, n+ and p+ layers for the source and drain of NMOS and PMOS transistors, respectively. Alternatively, conventional p-well CMOS is based on an n-substrate, thus featuring p-well for the bulk of NMOS transistors. Moreover, in twin-well CMOS technologies, both p-well and n-well are available. As a result, the device structure of Haitz *et al.* (Figure 2.1) was a feasible solution in CMOS, i.e. by using the p-substrate as anode, n+ as cathode and n-well as guard-ring. Nonetheless, on a complete integrated circuit, comprised of sensor and circuits, the p-substrate must be maintained to the lowest potential in case of n-well CMOS (n-substrate to the highest potential for p-well CMOS, accordingly). On a typical n-well CMOS, the lowest potential is typically 0V. Consequently, with the device of Haitz, the on-chip avalanche sensing circuit should be implemented at the cathode terminal, which in turn should be biased above breakdown voltage. In standard CMOS, transistors cannot handle such high voltage, consequently, the n+/p-substrate with n-well guard-ring structure would require special front-

end circuits. Indeed, a possible solution to detect the avalanche current using low-voltage transistors would require a sensing circuit that decouples high voltages of the SPADs from low voltages of the front-end. However, as described above, it is very important to keep the parasitic capacitance seen by the SPAD as low as possible. Furthermore, the structure of Haitz had a slow diffusion tail that is unwanted in timing-correlated applications. Rochas *et al.* proposed a different configuration, shown in Figure 2.5 (a), in which a p+/n-well active junction was isolated from p-substrate. Consequently, both anode and cathode could be biased independently of the p-substrate. This technique allowed the authors for instance to bias the p+ anode at a negative voltage equal to V_{BD} , and to bias the cathode at V_E , thus applying a sufficient total bias to operate in Geiger mode. Importantly, at the cathode, a voltage pulse with an amplitude of approximately V_E (typically 5V in their work), was generated at each photon detection, thus allowing the authors to take advantage of low voltage transistors. In order to implement a low doped p-type guard-ring, the authors have introduced an innovative design technique. Indeed, without the availability of a p-type layer with doping concentration stronger than the n-well doping concentration, and with the right implantation depth, the guard-ring design was a challenge. Rochas *et al.* utilized the lateral diffusion between two n-wells, spaced by a distance smaller than the spacing required by the technology's design rule [143]. As a result, the two n-wells were completely overlapped in depth, near the bottom of the n-well, but not overlapped underneath the surface. Since the background p-substrate doping was not overcompensated underneath the surface, the resulting structure featured an artificial low doped and shallow p-type guard-ring, according to Figure 2.5 (a). The lateral diffusion guard-ring structure was effective. Devices with 10 μm in diameter exhibited DCR of the order of 50Hz, excellent timing resolution of 50ps FWHM, and maximum photon detection probability of 25% at V_E of 2.5V. Owing to the full integration of a SPAD and front-end circuits, the total parasitic capacitance seen by the SPAD was 68pF [20]. As a result, Rochas was able to utilize a passive quenching, based on a single resistor, at detection rates up to 20MHz. Indeed, the dead time, comprising quenching and recharge, was 32ns only. Under this condition, negligible afterpulsing probability was obtained [20]. Rochas *et al.* have also evaluated the first array of SPADs fully integrated in CMOS technology [97] based on this device structure. The array was based on 8x4 SPADs with a pitch of 75 μm . At this pitch, crosstalk probability was negligible, mostly due to the small parasitic capacitance of 68pF.

With the increased demand for ICs in control circuits, involving sensors and actuators, special CMOS technologies providing high-voltage devices have emerged [147]. These CMOS technologies require particular fabrication

procedures that are in line with typical requirements for SPAD fabrication [130]. For instance, they provide low-doped and high quality epilayer and enforce improved process cleanliness to prevent contamination. Moreover, additional implant layers are available in those processes that may be effectively utilized to implement SPADs. In Figure 2.5 (b), a device structure based on a high-voltage CMOS technology is shown. This device structure, proposed by Rochas *et al.* in [19, 145], is basically a redesign of the previous structure, whereby the lateral diffusion guard-ring was replaced by a p-well implanted guard-ring. The authors took advantage of an existing p-well layer, typically used as the bulk of isolated NMOS transistors. That is, transistors whose bulk is isolated from the p-substrate by means of a deep n-well. This structure in a $0.8\mu\text{m}$ CMOS exhibited a DCR of few tens of Hertz and afterpulsing probability of 8% in devices up to $10\mu\text{m}$ in diameter at V_E of 5V [20]. The integrated high-voltage front-end circuit also allowed the authors to increase V_E up to 10V. As a result, higher photon detection probability was achieved. The maximal detection probability was 35% at a wavelength of 550nm. Very recently, Rochas *et al.* have reported large SPADs based on the same technology and similar performance with device diameters up to $50\mu\text{m}$ [146].

2.3.2 DEEP-SUBMICRON CMOS TECHNOLOGY

The design of avalanche photodiodes in deep-submicron CMOS technologies involves additional challenges. As previously mentioned, in order to operate in the so-called Geiger mode, a SPAD requires a design configuration that supports a planar and uniform multiplication region extending laterally and possibly vertically underneath the area of the SPAD as much as possible. Even though this requirement is mandatory to allow the creation of a reasonably large photosensitive or active area, it is not sufficient in general. For example, reference [148] has recently reported the design of a SPAD fabricated in $0.18\mu\text{m}$ CMOS that implements a planar multiplication region, according to simulations, but exhibits high DCR levels. DCR was approximately 1MHz, unacceptable in most applications.

Noise performance becomes a major issue for SPADs in deep-submicron CMOS technologies. It is therefore very important to keep a strict notion of noise performance when assessing potential design structures. The main sources of noise in SPADs are more significant in deep-submicron due to (i) higher doping levels, (ii) reduced annealing steps, and (iii) the presence of shallow-trench isolation (STI).

Higher doping levels increase the effects of tunneling-induced dark counts and increase the parasitic capacitance. The increase of parasitic capacitance increases the number of carriers involved in an avalanche discharge and thus worsens afterpulsing probability.

Driven by miniaturization, state-of-the-art fabrication processes reduce the strength and duration of annealing and drive-in diffusion steps, if any, to a minimum. The lack of effective annealing steps increases the concentration of impurities that introduce carrier recombination-generation and trapping centers in the active volume, thus worsening both thermally-generated dark counts and afterpulsing effects.

At and below the $0.25\mu\text{m}$ mark, standard CMOS processes feature STI compulsorily. It is known that STI may dramatically increase the density of deep-level traps and generation centers at its interface [149, 150]. It is therefore important to restrict the multiplication region of the SPAD in a volume with the smallest possible density of traps. When a STI is close to or in contact with the multiplication region of a SPAD, such as in [148], one may expect high dark count rates and afterpulsing probability.

Very often designers do not have enough flexibility to change or adapt a process parameter in order to better fit the SPAD requirements in CMOS technology. In order to address the issues described above, designers are left with a number of design layers, models, and rules. Consequently, when considering deep-submicron CMOS technology in this thesis, the main goal was to improve the overall performance in a given application. In this context, the benefit of implementing large arrays of single-photon detectors in time-correlated applications usually mitigates the time budget constraint in image acquisition. When compared to a sequential-scanning paradigm, such as on a scanning microscope, the total acquisition time in an array-based setup may be several orders of magnitude smaller. Additional acquisition time might be useful in those cases to mitigate the effects of higher noise in SPADs implemented in deep-submicron CMOS.

2.4 FRONT-END CIRCUITS

As briefly mentioned in Section 2.2.3, the SPAD operating principle requires special front-end circuits involving the quenching of avalanche current and the reset of SPAD bias voltage on each photon detection. In the literature, those functions are respectively addressed by quenching and recharge circuits. A

review on these circuits was reported in [151]. Nonetheless, with the intention of implementing large arrays of SPADs and circuits fully-integrated on a single IC, additional considerations have emerged recently. Furthermore, circuits and functioning modes that are claimed to poorly perform in [151] are currently practical solutions in the context of this thesis. Consequently, an array aware study on front-end circuits is presented in this Section.

In its simplest form, quenching is implemented passively, typically with a single resistor. We therefore define it as passive quenching circuit (PQ). In devices exhibiting large parasitic capacitance, i.e. large-area SPADs or devices involving hybrid circuits, the avalanche quenching action may involve a prolonged time when it is based on passive quenching. In addition, it is not convenient to allow the total capacitance to be completely discharged through the SPAD junction as it increases the number of carriers generated in an avalanche. With the introduction of the active quenching circuit (AQ) in [98, 131], these drawbacks were addressed. An active quenching circuit typically monitors the SPAD voltage or current so as to identify a potential avalanche start-up, typically by comparing it to a threshold. When an avalanche initiation is identified, the active quenching circuit quickly discharges the total capacitance seen by the SPAD through a different path, e. g. a transistor, thus quenching the avalanche. Ideally, the main goal is to force most of the charges to flow through the secondary path so as to minimize the number of carriers generated in the SPAD. As a result, fast sensing and feedback circuits are required. For instance, the total time a state-of-the-art active quenching circuit requires to detect an avalanche on a off-chip SPAD and to fully discharge its capacitance is typically 25ns [152]. During the first 25ns, the off-chip SPAD is involved in the process of discharging the total capacitance and it contributes by generating carriers in the multiplication region. It is therefore very important to either reduce the circuit response time or to simply decrease the parasitic capacitance. The latter solution was adopted by Rochas when small SPADs and front-end circuits were fully integrated on the same integrated circuit. In [143], the total parasitic capacitance was estimated by the authors to be 68fF. With such low parasitic capacitance, a full discharge requires approximately 3ns based on passive quenching [20]. Consequently, it is often unnecessary to actively quench the avalanche. Furthermore, it is rather difficult to implement active quenching circuits, even on-chip, with response times significantly smaller than 3ns, which would justify the added complexity. As a result, passive quenching in small fully-integrated SPADs is more appropriate. Nonetheless, Rochas opted to integrate a polysilicon resistor as quenching device. In view of increasing array size in imaging applications, pixel miniaturization is of greatest interest. As a result, sensors reported in this thesis are typically based on passive quenching by

means of a MOS transistor, as described in Section 2.4.1. It is however important to note that for large SPADs, i.e. diameters larger than $20\mu\text{m}$, active quenching circuits are highly recommended even when SPADs and front-end circuits are monolithically integrated on a same chip. In those cases, the SPAD junction capacitance becomes significant and so does the quenching time, thus justifying the use of active means.

Following the avalanche quenching, the second function of a SPAD front-end circuit is to recharge (reset) the SPAD bias voltage to the nominal value. Similarly to quenching, this action may be accomplished based on passive or active devices. We therefore distinguish passive recharge (PR) and active recharge (AR), respectively. A passive recharge may be achieved through the same device used in passive quenching. Consequently, both functions may be implemented by means of a single device: a resistor or a MOS transistor according to Section 2.4.1. However, similarly to passive quenching, large parasitic capacitance induces unfavorable effects in passive recharge circuits. Since the recharge time scales proportionally to its parasitic capacitance, the maximum count rate achievable by a SPAD may be extremely low in passive recharge when large devices or hybrid circuits are utilized. As explained in Chapter 3, high-count-rate capability is extremely important in applications based on TCSPC. Passive recharge of SPADs with large capacitance also leads to inaccurate timing resolution [151]. This unwanted effect may be explained as follows. During a passive recharge, the SPAD bias voltage follows an exponential curve towards the nominal bias. Since the SPAD progressively recovers its detection capability with its bias voltage, it is possible that a photon is detected when the bias voltage is significantly smaller than the nominal voltage. In such cases, the signal crossing of a detection threshold occurs earlier, thus causing an impaired timing response. Repetitive occurrence of this effect results in a significant broadening of timing resolution. Once more, the benefit of a full integration of front-end circuits and small SPADs on a single integrated circuit mitigates these unwanted effects. For instance, a fully-integrated passive quenching and recharge circuit is reported in [143], whereby total a quenching and recharge time, i.e. dead time, is achieved in 32ns. Based on this result, photon counting rates up to 10MHz may be measured with little impact on timing resolution. As a result, passive quenching and recharged based on a single MOS transistor may be an area-effective solution in large arrays of SPAD pixels, developed in Section 2.4.1.

In standard deep-submicron CMOS technology, afterpulsing effects become an important issue due to the reduced annealing steps, possible metal contamination, and the presence of STI. In such cases, it is advantageous to

implement a hold-off time so as to allow trapped carriers in the multiplication region due to an avalanche to be released before the SPAD recovers avalanche multiplication capability. Based on passive recharge, a hold-off time is implemented by increasing the quenching/recharge resistance value. Nonetheless, large recharge times limit timing resolution. Consequently, passive recharge typically leads to an afterpulsing/timing resolution trade-off. A possible solution is the use of an active recharge circuit. Active recharge circuits were introduced to reduce long recharge time of picofarad-range capacitances [151]. Its operating principle may be simply described as follows: once the avalanche is quenched (passively or actively), a secondary path with low resistance is utilized, typically a MOS switch, to quickly recharge the SPAD bias to the nominal voltage. As a result, subsequent photon detections occur under nominal biasing, thus resulting in high timing resolution.

Active recharge may be advantageously performed after a well-defined hold-off time to mitigate afterpulsing effects. Furthermore, it is necessary to prevent metastability effects during the recharge action. Indeed, there is a non-zero probability that a photon or afterpulse triggers an avalanche at the moment the SPAD is exposed to the low resistance path. As a result, reliable active recharge circuits typically require a level of complexity that in turn results in significant silicon area, thus reducing fill factor in arrays. In Section 2.4.2, a passive quenching/active recharge circuit is introduced. This circuit requires a minimum number of active devices and no passive components at pixel level. It enables a well-defined hold-off time, thus reducing afterpulsing effects without impacting timing resolution. It is important to note that, even though active recharge addresses the afterpulsing/timing resolution trade-off, the maximum practical detection rate is still limited by afterpulsing performance. The reduction of dead-time down to a few nanoseconds by means of active recharge is straightforward in state-of-the-art CMOS technologies. However, as the designer opts to reduce dead-time for a given SPAD, he or she might experience higher afterpulsing probability. In addition, since the release of carriers from deep-level traps obeys an exponential decay, afterpulsing probability for a given SPAD is expected to increase exponentially when dead-time is reduced.

2.4.1 PASSIVE QUENCHING AND RECHARGE BASED ON SINGLE MOS TRANSISTORS

One of the most important requirements a pixel-level circuit must fulfill is compactness. Typically, in APS CMOS image sensors, the number of transistors at pixel level rarely exceeds four. Furthermore, very often the circuit is only based on minimum size transistors. This choice is justified by the importance of

pixel fill factor over other design parameters. Fundamentally, image sensor design involves overall optimization instead of single-device performance. In a SPAD-based image sensor, the same principle is applicable. It is therefore important to evaluate the potential of quenching/recharge circuits based on a single MOS transistor. As described in the previous Section, in fully integrated pixels based on small SPADs, passive quenching may provide suitable performance. Nonetheless, a resistor of a few hundred kilo-ohm occupies a silicon area that is too large at pixel level to be successfully utilized.

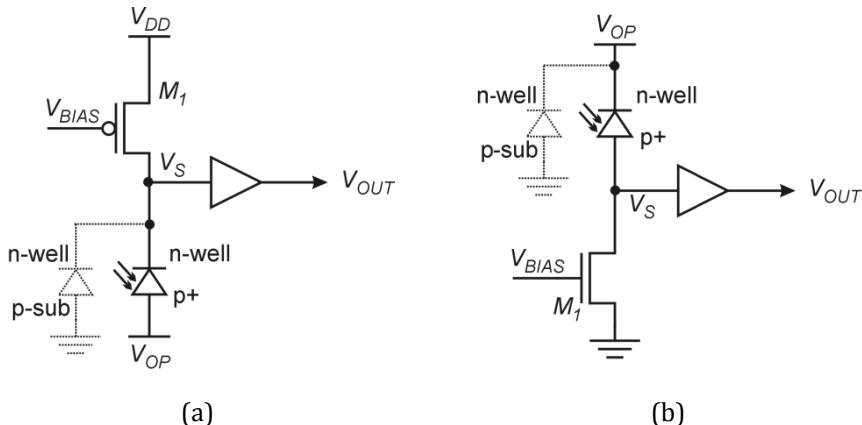


Figure 2.6. Single-MOS quenching and recharge circuit: (a) PMOS based circuit and (b) NMOS based circuit.

Typically in CMOS technology, the SPAD configuration of Figure 2.5, whereby the p+/n-well avalanche junction is surrounded by an n-well/p-substrate junction, is advantageous. For instance, the second junction conveniently enables the front-end circuit to operate at low voltage and it allows the p-substrate of the IC to be kept at the lowest potential. Additional motivations for p+/n-well/p-substrate SPAD is described in Section 2.3.1. Based on this structure, it is possible to implement two circuit configurations based on a single MOS transistor. In Figure 2.6 (a), a simplified circuit based on a PMOS quenching/recharge transistor is shown. This circuit is similar to the basic front-end described by Rochas *et al.*, apart from the fact that the quenching resistor was replaced by a PMOS transistor, M_1 . In the picture, the n-well/p-substrate diode is also shown. V_{DD} is the circuit power supply whereas V_{OP} is a negative bias voltage such that

$$V_E = V_{DD} - V_{SD,1} - V_{OP} - V_{BD}, \quad (2.1)$$

where $V_{SD,1}$ is the source to drain voltage of M₁, and V_{BD} is breakdown voltage of the SPAD. V_{BIAS} is a bias voltage such that M₁ exhibits large impedance at its drain terminal. When the SPAD is inactive, the current $I_{SD,1}$ through M₁ is zero, thus leading to $V_{SD,1} = 0$. Under this condition, V_E becomes

$$V_{E,MAX} = V_{DD} - V_{OP} - V_{BD}. \quad (2.2)$$

Typically, $V_{E,MAX}$ may be conveniently chosen to match V_{DD} . In this case, V_{OP} is set for a given SPAD such that

$$V_{OP} = -V_{BD}. \quad (2.3)$$

Equation (2.1) reduces to

$$V_E = V_S = V_{DD} - V_{SD,1}. \quad (2.4)$$

In the circuit of Figure 2.6 (a), the voltage at sensing node V_S is converted to a digital pulse at V_{OUT} by means of a comparator or digital gate with fixed threshold V_{TH} . A typical photon-detection waveform of the voltage at notes V_S and V_{OUT} as a function of time is shown in Figure 2.7 (a). Since the impedance seen by the drain of M₁ is much larger than the SPAD series resistance R_D , the avalanche current generated in the SPAD produces a voltage drop across M₁, i.e. $V_{DS,1}$, of approximately V_{DD} , thus forcing V_E in equations (2.1) and (2.4) to converge towards 0V. As a result, the current flowing in the SPAD becomes significantly smaller than the so-called SPAD *latching current* level, i.e. the current under which the avalanche multiplication is no longer self-sustainable. This quenches the avalanche at time T_Q . The SPAD latching current (I_Q), typically of the order of 50μA, specifies a limit on the number of carriers that should be present in the multiplication region at any point in time to sustain avalanche multiplication. Indeed, when the number of carriers in the high-field region decreases under a certain level, there is a non-zero probability that all the carriers drift out of the multiplication region even at a positive avalanche ionization rate. As soon as the SPAD junction is fully depleted of carriers, the avalanche is quenched.

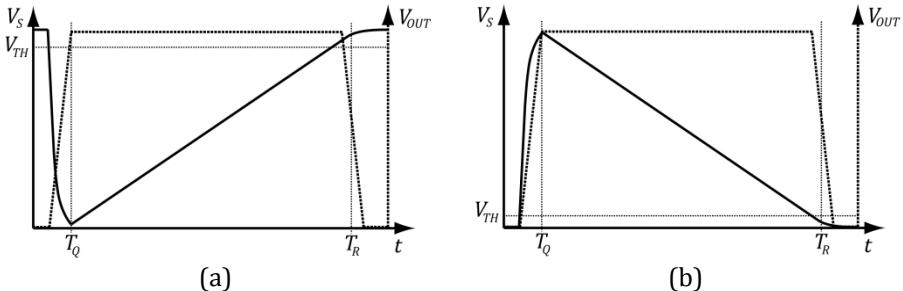


Figure 2.7. Voltage waveforms at sensing node V_S and at output node V_{OUT} for (a) PMOS and (b) NMOS circuits, respectively.

The waveform of V_S as a function of time after T_Q depends on the biasing and width-to-length (W/L) aspect-ratio of M_1 . An interesting mode of operation for M_1 would be to function as a mirrored current source. This mode of operation leads to superior quenching and recharge performance over a MOS transistor operated as a resistor, i.e. operated out of saturation for any value of V_{SD} . In Figure 2.8 (a), the current-to-voltage (I-V) characteristic of a PMOS-based current source is shown. In order to simplify the description, an idealized current source with infinite output resistance is assumed. An important design parameter is the drain current I_0 , which is chosen to be significantly smaller than the SPAD latching current I_Q , i.e.

$$I_0 \ll I_Q. \quad (2.5)$$

For instance, I_0 of $2\mu\text{A}$ may be suitable for shallow SPADs so as to ensure proper quenching independently of $V_{SD,1}$. It is important to note that, assuming a planar multiplication region, I_Q does not depend on SPAD area.

Another important design parameter for M_1 is its *source-to-drain saturation voltage* ($V_{SD,SAT,1}$). Practically, it is possible to size M_1 such that its $V_{SD,SAT}$ is very small compared to $V_{E,MAX}$, for instance,

$$V_{SD,SAT,1} \leq \frac{V_{E,MAX}}{33}. \quad (2.6)$$

As a result, following the SPAD current quenching, at time T_Q , in Figure 2.7 (a), M_1 linearly recharges the SPAD anode toward $V_{DD} - V_{SD,SAT,1}$ with a slope of C_S/I_0 , where C_S is the total capacitance at node V_S . When $V_{SD,1}$ reaches and

subsequently becomes smaller than $V_{SD,SAT,1}$ at time T_R , M_1 leaves its saturation region and V_S is recharged exponentially toward V_{DD} . However, if relation (2.6) is fulfilled, at time T_R , V_E is already recharged up to 97% of $V_{E,MAX}$. It is therefore very important to properly size and bias M_1 to reduce the effect of unwanted exponential recharge, typical in passive quenching based on a resistor.

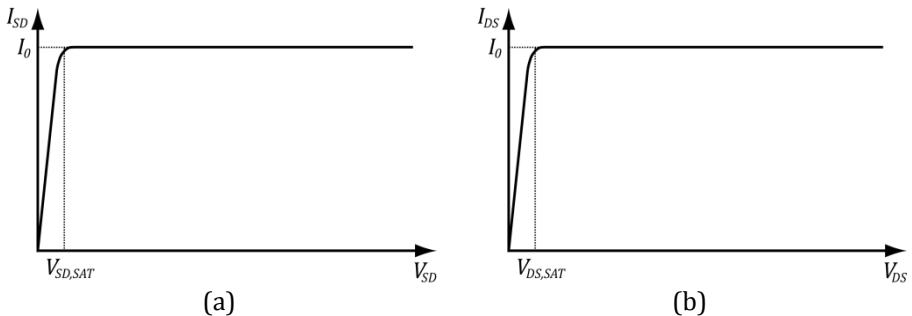


Figure 2.8. I-V characteristic of MOS mirrored current sources based on (a) PMOS and (b) NMOS transistors, respectively.

The circuit of Figure 2.6 (a) has an important drawback compared to the NMOS-based circuit, shown in Figure 2.6 (b). As can be seen in Figure 2.6 (a), the cathode of the inactive n-well/p-substrate diode is connected to node V_S . As a result, it increases the total capacitance C_S , which is discharged by the SPAD on photon detection. In the circuit of Figure 2.6 (b), the n-well/p-substrate diode is fully isolated from V_S , thus reducing C_S . Since the area of the inactive diode is larger than the area of the avalanching junction, it contributes to approximately half of the 68fF-parasitic capacitance reported by Rochas *et al.* In SPAD-based image sensors reported in next Chapters, the utilization of the configuration shown in Figure 2.6 (b) leads to parasitic capacitances lower than 25fF for similar SPADs. Furthermore, the configuration of Figure 2.6 (b) leads to smaller transistor sizes for M_1 so as to fulfill equation (2.6). It also helps achieving lower V_{TH} values in the subsequent sensing circuit, which in turn improves timing resolution.

Figure 2.7 (b) shows the waveform of V_S and V_{OUT} as a function of time for the NMOS-based configuration, whereas in Figure 2.8 (b), the I-V characteristic of a NMOS current source is plotted. Equations (2.7) and (2.8) replace equations (2.1) and (1.1), respectively. Note that V_{DD} is no longer necessary in the NMOS-based configuration. V_{OP} , which is a positive voltage in this case, must be high enough to account for V_{BD} and $V_{E,MAX}$.

$$V_E = V_{OP} - V_{BD} - V_{DS,1}, \quad (2.7)$$

$$V_{E,MAX} = V_{OP} - V_{BD}. \quad (2.8)$$

The design condition on M_1 , stated in relation (2.6), remains valid, except that $V_{SD,SAT,1}$ is replaced by $V_{DS,SAT,1}$, i.e.

$$V_{DS,SAT,1} \leq \frac{V_{E,MAX}}{33}. \quad (2.9)$$

It is relatively straightforward to generate a common V_{BIAS} for many pixels in an array based on the basic current mirror topology. In Figure 2.9 an example of a possible bias generation circuit is shown.

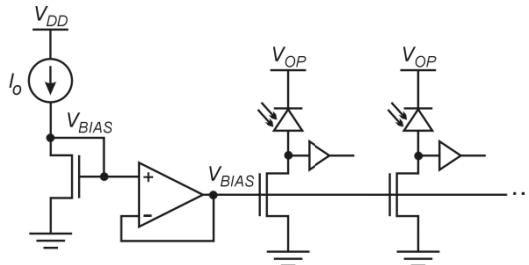


Figure 2.9. Example of V_{BIAS} generation and distribution in SPAD arrays.

In order to quantitatively evaluate the quenching and recharge performance of the NMOS-based circuit of Figure 2.6 (b), one may describe the evolution of V_S as a function of time in three intervals of interest: (i) quenching ($t \leq T_Q$), (ii) linear recharge ($T_Q < t < T_R$), and Ohmic recharge ($t \geq T_R$). Assuming that the avalanche is triggered at time $t=0$ and that M_1 acts as idealized current source, i.e. we neglect channel length modulation effects and we assume a perfectly Ohmic behavior for M_1 in the triode region, then V_S may be expressed as

$$V_S(t) \cong \begin{cases} V_{E,MAX} \left[1 - \exp \left(-\frac{t}{R_D C_S} \right) \right] - R_D I_0 & \text{when } t \leq T_Q \\ V_{E,MAX} - R_D I_0 - \frac{I_0}{C_S} (t - T_Q) & \text{when } T_Q < t < T_R, \\ V_{DS,SAT,1} \exp \left(-\frac{t-T_R}{R_{DS,1} C_S} \right) & \text{when } t \geq T_R \end{cases} \quad (2.10)$$

where R_D is the series resistance of the SPAD, according to its electrical model shown in Figure 2.10, and $R_{DS,1}$ is drain to source resistance of M_1 in the triode region. In Equation (2.10) a number of approximations were made. For instance, V_S does not actually reaches $V_{E,MAX} - R_D I_0$ at time T_Q given that avalanche quenching is a statistical process. Nonetheless, from a circuit designer standpoint, Equation (2.10) is sufficiently accurate. It is also important to note that small fluctuations of T_Q do not limit timing resolution given that the sensing circuit is triggered at the initial phase of the avalanche, i.e. V_{TH} is chosen as small as possible. Furthermore, when condition of (2.5) is fulfilled, these fluctuations are negligible. T_Q and T_R may be approximated as

$$T_Q \cong 5 \cdot R_D C_S \text{ and} \quad (2.11)$$

$$\begin{aligned} T_R &= T_Q + \frac{C_S}{I_0} (V_{E,MAX} - R_D I_0 - V_{DS,SAT,1}) \\ &\cong T_Q + \frac{C_S}{I_0} (V_{E,MAX} - V_{DS,SAT,1}). \end{aligned} \quad (2.12)$$

Assuming R_D of $2\text{k}\Omega$, C_S of 25fF , $V_{E,MAX}$ of 3.3V , $V_{DS,SAT,1}$ of 100mV , and I_0 of $2\mu\text{A}$, then T_Q is 250ps and T_R becomes 40ns . As a result of small capacitances C_S in fully-integrated pixels, quenching time T_Q becomes extremely short and it is nearly negligible compared to T_R . In these devices, T_Q causes active quenching circuits to become impractical and often unnecessary. In addition, T_R may be conveniently used to define *dead time*, which would result in 40.25ns based on these parameters. It is possible to further increase I_0 so as to reduce dead time. However, as mentioned previously, it is very important to allow enough time for trapped carriers to be released before V_E reaches $V_{E,MAX}$ so as to reduce afterpulsing probability.

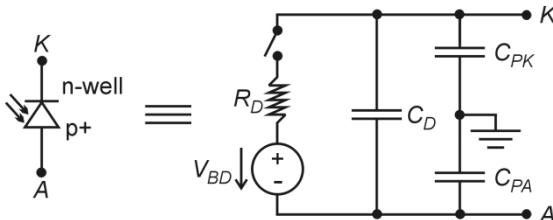


Figure 2.10. Electrical model of a SPAD. C_{PK} and C_{PA} are the $p+/p$ -sub and n-well/p-substrate parasitic capacitances, respectively. Note that C_{PK} may be neglected in the devices proposed in this thesis.

A detailed waveform of V_S for the passive NMOS-based circuit is shown in Figure 2.11, based on the parameters assumed in the last paragraph. In the picture, the equivalent waveform corresponding to the passive quenching based on a resistor is also shown for comparison purposes. The value of the quenching resistor was chosen to cause the same V_S value at time T_R with respect to the NMOS-based circuit.

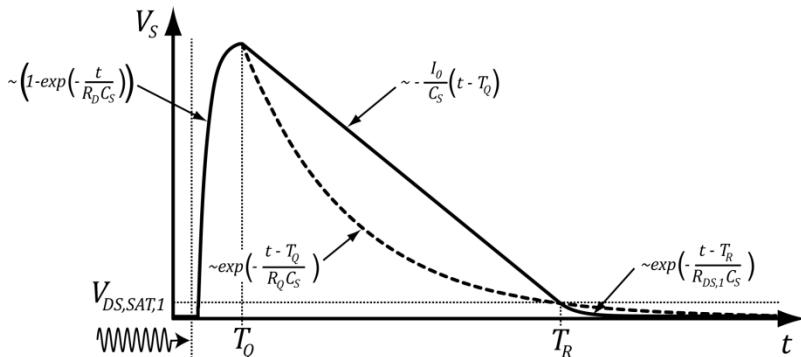


Figure 2.11. Detailed photon-detection waveform at sensing node V_S for NMOS-based quenching/recharge circuit.

As can be seen in Figure 2.11, the MOS-based passive recharge is superior to the resistor-based one with respect to afterpulsing performance. In order to reduce the probability of afterpulsing, it is important to keep V_E as low as possible (thus to keep V_S as high as possible) during dead time. Assuming a fixed dead-time, the MOS-based approach ensures that V_E is lower at any time from detection to recharge, with respect to the resistor-based circuit. Furthermore, in the passive quenching circuit based on a resistor, a residual V_S voltage persists beyond the dead-time, i.e. later than T_R . This voltage residue is responsible for limiting timing resolution under high photon-counting rate operation. Based on the proposed biasing scheme for the passive MOS quenching, the V_S voltage residue is considerably smaller.

The single-NMOS based quenching and recharge circuit was designed using a high-voltage 0.35 μ m CMOS technology [153]. The sensing circuit of Figure 2.6 (b) was implemented as a simple inverter. Its threshold V_{TH} was designed to be 750mV. A I_0 current of 1 μ A was also chosen whereas C_S was estimated to be 20fF. Figure 2.12 shows simulation results based on a transient analysis. When a photon detection is simulated at 10ns, V_S is raised extremely fast towards $V_{E,MAX}$, thus triggering a sharp high-to-low transition on V_{OUT} , shown in Figure

2.12 (top). The quench duration is extremely short, within 250ps. At the bottom of Figure 2.12, the transient response of $I_{DS,1}$ is also shown. Upon photon detection, M₁ leaves its cutoff region and $I_{DS,1}$ converges to I_0 . Note that the current peak of $I_{DS,1}$ around 10ns is only due to the parasitic capacitance corresponding to the drain of M₁. The integral of this sharp peak over time, i.e. electric charge, is much smaller than the total integral of $I_{DS,1}$. Thus, the drain capacitance of M₁ is negligible with respect to C_s . V_S follows its predicted linear decrease as a function of time towards ground. When it crosses V_{TH} , approximately at 45ns, it triggers the inverter back to its inactive (high) state. Note that a parasitic coupling between V_{OUT} and V_S is noticeable. Nonetheless, the circuit operating principle is not significantly altered. A few nanoseconds later, V_S crosses $V_{DS,SAT,1}$, approximately at 58ns, thus forcing M₁ to leave its saturation region. From this point in time, the SPAD recharge is completed following an exponential behavior, although it is almost fully-recharged.

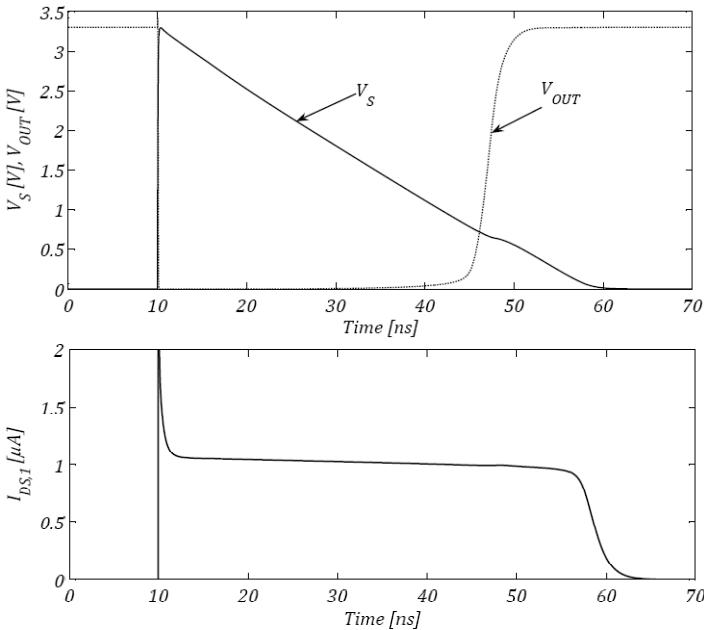


Figure 2.12. Simulation results for the circuit of Figure 2.6 (b), whereby a CMOS inverter was assumed for the sensing circuit. V_S and V_{OUT} (top), and $I_{DS,1}$ (bottom) as a function of time. I_0 of 1 μ A and V_{TH} of 750mV were chosen. C_s was assumed to be 20fF.

The basic single-NMOS quenching and active recharge may be therefore effectively used in pixels where miniaturization is highly relevant. Despite its simplicity, it reliably performs the required functions. In applications requiring high-performance, additional improvements may be achieved by means of active recharge circuits, involving nonetheless more devices. An example of circuit taking advantage of active recharge is presented in next Section.

2.4.2 DUAL-THRESHOLD QUENCHING AND RECHARGE CIRCUIT

As previously described, the dynamic behavior of SPAD signals on photon detection involves two phases: (i) quenching and (ii) recharge. Based on passive quenching by means of a single MOS transistor, quenching times of the order of a few hundred picoseconds are feasible in fully-integrated image sensors. It is therefore very difficult to enhance quenching by active means. When it is possible, very often it does not justify the added complexity and silicon area utilization within the pixel. As a result, one may advantageously preserve passive quenching and focus on the active recharge phase.

In SPAD-based image sensors, the utilization of active recharge circuits is not justified by the need for short dead times. According to equations (2.12) and (2.5), it is straightforward to obtain dead time of the order of 20ns. On the contrary, it is desirable to achieve a hold-off time whereby the V_E is kept low over a certain time so as to prevent afterpulses. Passive recharge does not provide this feature.

Delayed recharge of the SPAD bias by means of one or more additional active devices enables the realization of a hold-off time, preferably covering most of the dead time. Several circuits of this type have been reported in the literature for hybrid implementations. The approaches were reviewed in [151] and in [145] for fully-integrated SPADs. However, these circuits require a number of passive and/or active devices to generate an accurate hold-off delay to trigger the SPAD recharge phase after avalanche quenching. Typically, active recharge circuits rely on monostable circuits to provide hold-off time [151]. Alternatively in references [55, 145], a local oscillator is started upon photon detection and it is used to provide a fixed delay with respect to the quenching phase, i.e. hold-off time. The local oscillator also provides an additional function: if at the end of the first recharge period the SPAD is not fully recharged, for instance because of an afterpulse or a subsequent photon, the oscillator is used to generate a second hold-off time followed by another recharge phase. This cycle may be repeated as many times as needed, thereby assuring proper recharge. However, circuits that involve a typical monostable or local oscillator are nowadays far too large to be

implemented at pixel level. In this Section, a new circuit topology is introduced whereby the own SPAD capacitance is used as a passive component to generate the desired hold-off time. Consequently, this circuit involves a reduced number of small active devices.

Figure 2.13 shows the proposed front-end circuit, which is referred to as a dual-threshold quenching and recharge circuit. The SPAD cathode is biased with a positive bias V_{OP} above V_{BD} in order to ensure the device to operate in Geiger mode. Its anode is connected to a passive quenching NMOS transistor that operates as a current source I_o , as described in previous Section. The total parasitic capacitance seen by the SPAD anode is C_S . The operation of this circuit relies on two distinct voltage discrimination thresholds, V_{TH1} and V_{TH2} , used to detect transitions at sensing node V_S . An inverter is used to generate an output signal V_{OUT} and its discrimination threshold V_{TH1} is chosen as low as possible. A time-accurate and robust digital signal is therefore generated upon photon detection. V_{OUT} is fed back onto an input of a NOR gate, whereas the other input is connected to V_S . The NOR gate is designed with discrimination threshold V_{TH2} , higher than V_{TH1} . Its output, V_{RCH} , drives a NMOS transistor, M_2 , in order to provide active recharge.

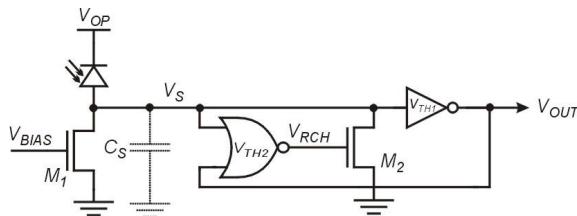


Figure 2.13. Dual-threshold passive quenching and active recharge circuit.

An example of signal waveforms of nodes V_S , V_{OUT} , and V_{RCH} are illustrated in Figure 2.14. On photon detection, V_S is rapidly increased to reach a voltage close to the SPAD excess bias V_E . Since C_S is of the order of 35 fF, the rise time of V_S is approximately a few tens of picoseconds. On its low-to-high transition, V_S crosses V_{TH1} , which makes V_{OUT} to make a high-to-low transition, thus reproducing precisely the arrival time of a photon. Since V_S also crosses V_{TH2} immediately after V_{TH1} on its rising edge, V_{RCH} does not change, thus keeping M_2 off during the quenching phase. Note that the feedback delay through the output inverter helps insuring that M_2 does not turn on the rising edge of V_S . After avalanche quenching, transistor M_1 reduces V_S towards V_{TH2} with a slope of I_o/C_S . As V_S goes across V_{TH2} , the NOR gate toggles and initiates an active

recharge. A hold-off time T_{OFF} may be set by setting the value of I_0 and V_{TH2} , according to Equation (2.13). In recharge phase, M_2 rapidly reduces V_S towards ground. Nonetheless, when V_S crosses V_{TH1} , V_{OUT} switches to $V_{E,MAX}$ which in turn makes the NOR gate to stop the active recharge. Assuming V_{TH1} is very low, the feedback delay from the moment it is crossed by V_S until V_{RCH} is deasserted allows the SPAD to be fully discharged by M_2 . The deassertion of V_{RCH} completes the detection cycle.

$$T_{OFF} = \frac{C_s(V_{E,MAX} - V_{TH2})}{I_0} \quad (2.13)$$

Note that a photon or an afterpulse may trigger an avalanche when M_2 is still active. In order to ensure avalanche quenching in this condition, M_2 is sized to provide a maximum drain-to-source current lower than I_Q , i.e. $I_{DS,2} < I_Q$ for any $V_{GS,2}$ and $V_{DS,2}$. As a result, V_S is pulled to $V_{E,MAX}$ when an avalanche is triggered even when M_2 is active, thus retriggering another hold-off and recharge cycle.

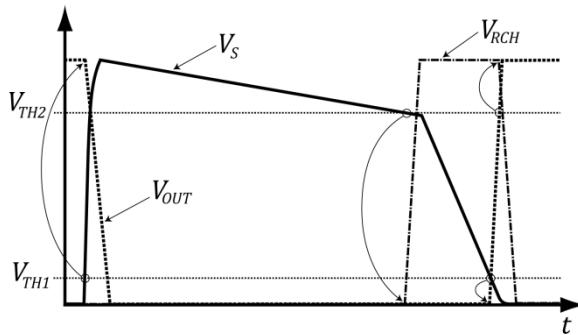


Figure 2.14. Signal waveforms of nodes V_S , V_{OUT} and V_{RCH} .

A transistor-level implementation of the dual-threshold quenching and recharge circuit is shown in Figure 2.15. The circuit consists of 8 MOS transistors, of which 5 are NMOS and 3 are PMOS. Most transistors are sized at or near minimum sizes, except for M_1 and M_2 where a small width-to-length (W/L) ratio may be necessary. The parasitic p-sub/n-well diode and parasitic capacitance C_s are explicitly shown in the circuit. Note that since the voltage across the parasitic diode stays constant upon photon detection, its junction capacitance does not contribute in C_s . The circuit does not dissipate any static power as a result of all the transistors being operated digitally, with the exception of M_1 which is cut-off during two detection cycles.

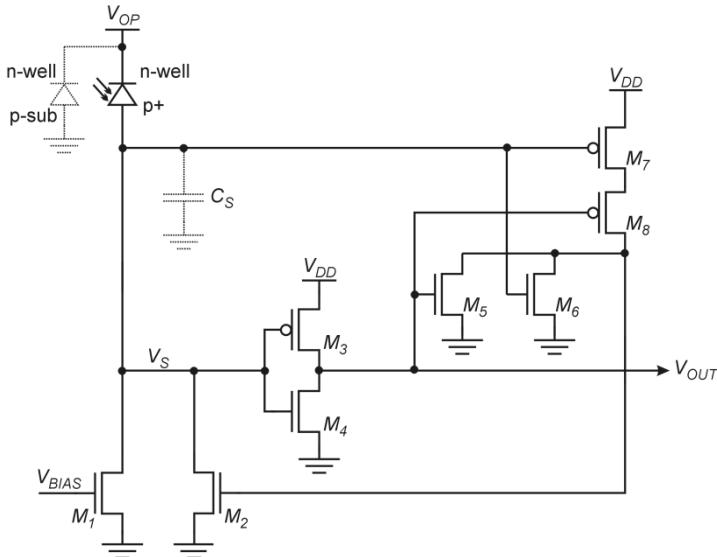


Figure 2.15. Transistor-level implementation of dual-threshold passive quenching and active recharge circuit.

The circuit of Figure 2.15 was designed in a high-voltage $0.35\mu\text{m}$ CMOS technology [153], as described in Chapter 4. Figure 2.16 shows transient simulation for V_S and V_{OUT} as well as for $I_{DS,1}$ and $I_{DS,2}$. C_S was estimated to be 35fF and R_D of the SPAD was assumed to be $2\text{k}\Omega$, based on the model of Figure 2.10. In this circuit, V_{OP} was chosen so as to set $V_{E,MAX}$ to be 3.3V , which is the nominal V_{DD} in this process for low-voltage transistors. V_{TH1} and V_{TH2} were set to 0.6V and 2.3V , respectively, by properly sizing transistors M_3 - M_8 . In the picture, when a photon detection is simulated at 10ns , V_S is raised extremely fast towards $V_{E,MAX}$, thus triggering a sharp high-to-low transition on V_{OUT} . The total quench duration is lower than 350ps . During this time, M_1 turns on and starts sinking a constant current $I_{DS,1}$ of I_0 , which was chosen to be $1\mu\text{A}$, according to Figure 2.16 (b). Note that, similarly to the single-NMOS circuit, the current peak of $I_{DS,1}$ around 10ns is only due to the drain parasitic capacitance of M_1 . As can be seen in the same picture, $I_{DS,2}$ is zero around 10ns , which shows that M_2 is completely off during quenching. This behavior is very important in order to reduce the effects of SPAD non-idealities such as afterpulsing. When V_S reaches V_{TH2} around 50ns , transistor M_2 switches on and provides a current pulse, peaking slightly below $20\mu\text{A}$, thus actively recharging the SPAD. As a result, V_S is reduced quickly below V_{TH1} , which in turn triggers V_{OUT} back to its inactive state.

The rise time of V_{OUT} is used to provide an additional time delay for M_2 to keep decreasing V_S towards ground. If a residual voltage of a few tens of millivolts remains on the V_S node at moment M_2 is turned off, M_1 continues the SPAD recharge passively before its cut off when V_S reaches ground. Finally, since $I_{DS,2}$ is unconditionally lower than I_Q , a photon detection at a time of approximately 55ns in Figure 2.16 would not lead to any unpredictable behavior. On the contrary, in such situation, the SPAD would bring V_S to $V_{E,MAX}$, thus blocking M_2 and restarting another hold-off cycle.

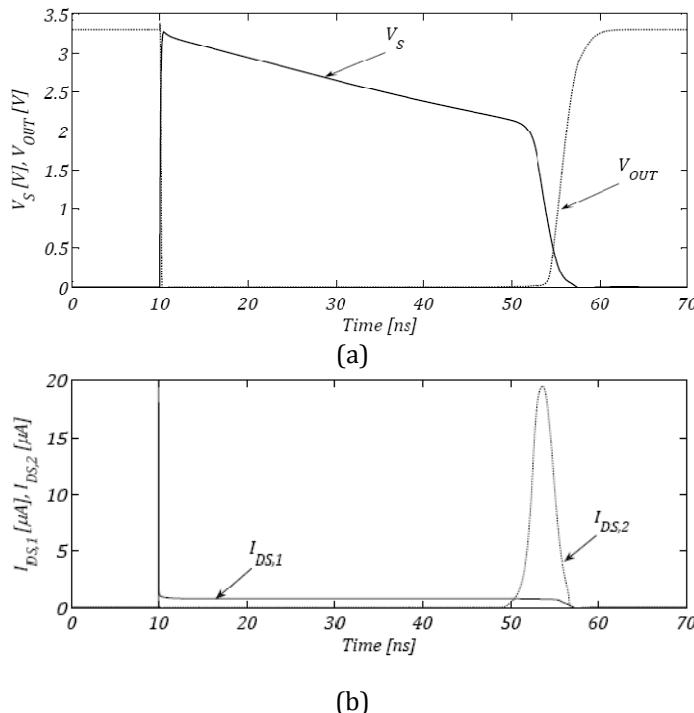


Figure 2.16. Transient simulation results for the dual-threshold quenching and recharge circuit of Figure 2.15. (a) V_S and V_{OUT} and (b) $I_{DS,1}$ and $I_{DS,2}$ as a function of time.

The circuit of Figure 2.13, whose possible implementation is shown in Figure 2.15, has a number of advantages and compares favorably to existing circuits in the literature in the context of large arrays of SPADs. Firstly, the circuit does not consume any static power, which would be a high penalty when designing large arrays of pixels. Despite its low transistor count, the circuit offers passive

quenching, followed by a hold-off time, and a reliable active recharge. A hold-off time is a key function for SPADs implemented in submicron and deep-submicron CMOS technologies to reduce afterpulsing effects. In high-quality processes, in which the density of deep-level traps is adequate, the proposed circuit also allows for a shortening of the hold-off time down to 10ns or lower by appropriately choosing I_0 and, if necessary, V_{TH2} . It is also possible to modulate T_{OFF} during operation by adjusting I_0 actively. Finally, the dual-threshold circuit does not require any passive device at pixel level, which tends to occupy an unacceptable area, and utilizes a low number of small transistors. These characteristics make this front-end circuit very suitable for low-pitch single-photon pixels in large array implementations.

2.4.3 EVENT-DRIVEN READOUT CIRCUIT

The digital nature of the signals generated in Geiger mode, which leads to significant advantages, also hold at least one drawback to be solved at circuit level. Indeed, since the parasitic capacitance of a SPAD is fully discharged upon photon detection, it is necessary to add extra circuitry to store photo-generated information. For instance, it might be important to add a counter so as to store the information on how many photons the pixel has detected in a given time interval. On a conventional image sensor, this information is stored on the photodiode capacitance (as an integrator). Since the number of transistors required implementing at least one counter is by far larger than the number of transistors required at the pixel level for a CMOS APS image sensor, the design of small SPADs pixels involves very important challenges.

A basic idea to address this challenge is to conceive a circuit that allows the designer to place all the circuits required for storage and timing evaluation outside the pixel array. Nonetheless, since the arrival time of photons is of extreme importance in time-correlated applications, the circuit also needs to preserve photon arrival time with high precision. Moreover, such a circuit should require as few transistors as possible at the pixel level to maximize fill factor.

The simplest conventional time-preserving readout circuit could be an M-to-1 multiplexer covering a range of pixels. For instance, a multiplexer covering all the pixels within a given row or column. An M-to-1 multiplexer allows one to route the output of a selected pixel directly to the global output for all M pixels in the range. Using this approach, it is possible to preserve photon timing, but it prevents simultaneous detection in all pixels of a given range (e.g. all the pixels in a row or in a column). As a result, every pixel in the range should be

sequentially selected during the integration time, which in turn leads to an image acquisition time a factor M slower than it would be on a conventional image sensor. This limitation is not acceptable in most applications, especially when the image sensor is intended to operate under low light conditions. Yet, when light illumination on a row or column is low, an alternative approach is possible.

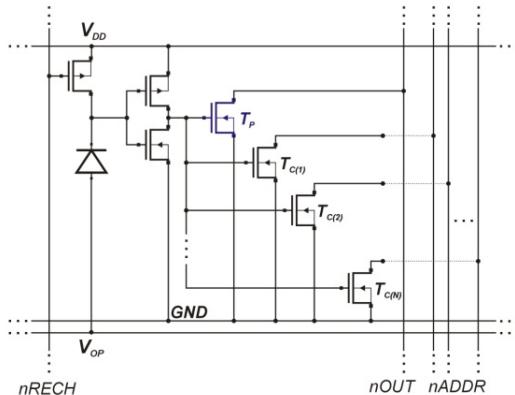


Figure 2.17. Pixel circuitry based on the event-driven readout principle

Under low light conditions, an interesting non-conventional approach to address the readout challenge was investigated in this thesis. The corresponding circuit, referred to as an event-driven readout, was first reported in [154]. In this approach a row, a column or any range of pixels is transformed into a digital bus, which is accessed by a pixel in an asynchronous fashion. In a given implementation, a bus consisting of a high-speed output line $nOUT$ and a set of N address lines $nADDR$ is employed. Signal $nOUT$ is used to propagate the photon-generated digital pulse from the SPAD, where photon detection occurred, to the bottom of the bus, where time-of-arrival can be evaluated, for instance by means of a time-to-digital converter. Time-uncorrelated photon counting or other operation modes may also be performed outside the pixel array, thus enabling resource sharing and/or better utilization of silicon area. The high-speed output line is optimized so as to limit its timing jitter contribution, thus guaranteeing high timing resolutions. Assuming that the event-driven circuit is implemented along one column and that it handles N_{ROW} pixels, the address bus encodes the row at which the detection pulse originated. A simple version of an event-driven pixel is shown in Figure 2.17. The inverter, triggered by photon detections drives the NMOS transistor M_P , which is used to pull down $nOUT$. This line, at high logic state when inactive, propagates a low

logic pulse, where the high-to-low transition reproduces the arrival time of photons. The address corresponding to the appropriate line is sent along address bus $nADDR$, using transistor array $M_{C(1)}-M_{C(N)}$ as shown in the figure. The timing pulse and address propagation assumes that the bus is inactive, i.e. it is not being utilized by another pixel on the same pixel column.

The address bus may be implemented in a number of ways. For instance, the bus may be implemented in a parallel fashion. The required number of address lines N per column is computed as

$$N = \lceil \log_2 N_{ROW} \rceil, \quad (2.14)$$

where N_{ROW} is the number of rows in the array, i.e. the number of pixels in a column. This readout method is scalable and the logarithmic dependence of N from N_{ROW} suggests that it is indicated for large arrays, as far as silicon area utilization is concerned. However, the scalability of the approach for large arrays has the effect of reducing the maximum photon flux that can be detected before saturating the readout system. A simplified illustration of the overall readout circuit in a column is shown in Figure 2.18. The figure shows the pull-up resistors at the extremities of all the signal lines in order to ensure a high logic level while not pulled down by the corresponding transistor. These resistors can be actually implemented using active devices. At the bottom of the readout channel, the timing pulse associated with a given pixel is used to latch the $nADDR$ which is transferred into $ADDR$. Once the address $ADDR$ is secured, it is used to address a memory location in which the photon event or timing measurement is stored. At the same time, $nRECH$ signals availability of the channel upon readout cycle completion. Thus, the recharge of the SPAD pixel that fired can now be completed. The adjustable delay of Figure 2.18 is utilized to avoid setup time violations. A collision resolution mechanism may be implemented based on the event-driven approach. Note that the frontend circuits proposed in Sections 2.4.1 and 2.4.2 may be effectively combined with the event-driven readout.

The proposed event-driven readout principle is particularly indicated for applications where the arrival of photons is sparse in time. This is generally the case in applications based on TCSPC and more specifically in setups involving fluorescence lifetime imaging, fluorescence correlation spectroscopy, etc. In those applications, typical counting rates per pixel range from a few kilohertz up to a few tens of kilohertz, while the excitation repetition rates reach few tens

of megahertz. As a factor 1000 typically exists between the excitation rate and the detection rate per pixel, the column-level event-driven approach could be utilized for instance to handle an array of 128x128 pixels. In TOF ranging applications however, a large-N event-driven readout circuit is less attractive due to the high counting rates induced by background light conditions. Low-N versions of the readout circuit on the contrary are still attractive.

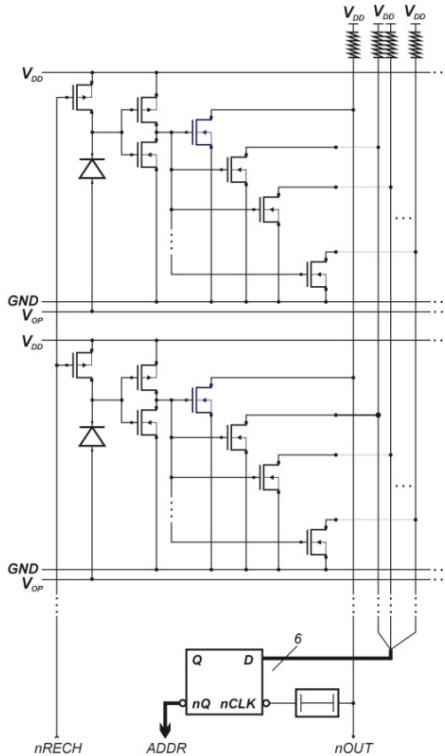


Figure 2.18. Simplified illustration of the complete column-level event-driven readout circuit.

Two single-photon sensors have been designed, fabricated and tested based on the proposed event-driven approach, as described in Sections 2.5 and 2.6. In those CMOS-integrated sensors, the event-driven readout enabled a non-sequential and quasi-full parallel frame acquisition under moderate light conditions.

2.5 SPAD INTEGRATION IN 0.8-MICRON CMOS

The first SPAD device evaluated in this thesis was based on a $0.8\mu\text{m}$ CMOS high-voltage [155]. The device chosen was the SPAD structure first reported by Rochas *et al.* in [19, 145], whose cross-section is shown in Figure 2.5 (b). This structure has been extensively characterized by Rochas in [20]. Nonetheless, since the author has not reported performance measurements of this SPAD in array formats, an open question regarding the statistical distribution of performance remained. In particular, the distribution of DCR and other parameters among all the pixels of a relatively large array was not known when this thesis started. Even more important, in an array of SPADs, it is desirable to distribute V_{OP} , which is a relatively high voltage bias, globally to all the pixels. This requirement stems from the fact that at the pixel array area, only low-voltage transistors should be used in order to reduce pixel pitch and increase fill factor. In this context, if a single SPAD in the array suffers from premature breakdown, the whole array would be compromised as it would be difficult to raise V_{OP} at the required voltage to make the remaining SPADs operate in Geiger mode. A statistical analysis of this issue in large arrays and over a number of samples was also missing in the literature at the beginning of this research.

In this thesis, in order to conduct performance evaluation in large arrays, the first approach was to design a relatively large array of a known device, hence the SPAD of Figure 2.5 (b). The device consists of a p+/deep n-well SPAD based on p-well diffusion guard-rings. Towards device performance and time-of-flight application evaluation, an array of 32x32 pixels was first reported in $0.8\mu\text{m}$ CMOS technology in [23]. Performance characterization was later reported in a journal article [156]. Another array of 64x48 SPADs was also designed in the same technology and reported in [157]. Since the latter design shared the same SPAD structure, it has also been considered in the experimental evaluation of device performance, which is described in this Section.

The sensor, comprising 32x32 pixels, was implemented based on a very simple digital readout circuitry owing to the digital nature of SPADs [156]. The functional diagram of the sensor is shown in Figure 2.19 (a). The sensor uses two power supplies: V_{DD} of 5V and V_{OP} of -25.5V. The readout circuitry consisted of a 32-channel decoder for row selection and a 32-to-1 multiplexer for column selection. The pixel consists of a circular SPAD and a 5-transistor circuit, shown in Figure 2.19 (b). The SPAD operates in passive quenching. The p+ anode is biased at a high negative voltage V_{OP} equal to -25.5V. This voltage is common to all the pixels in the array. The deep n-well cathode is connected to the power supply V_{DD} of 5V through a PMOS transistor. The excess bias voltage $V_{E,MAX}$ is 5V.

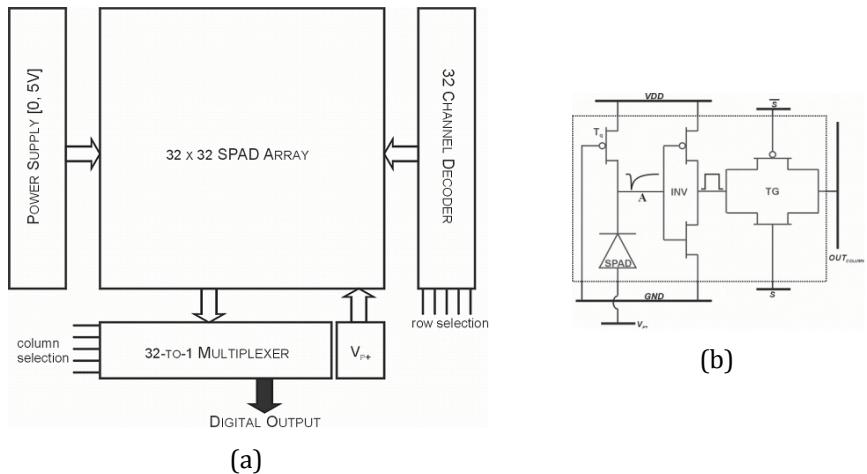


Figure 2.19. (a) Integrated circuit architecture of the single-photon sensor comprising 32x32 pixels and (b) 5T pixel circuit.

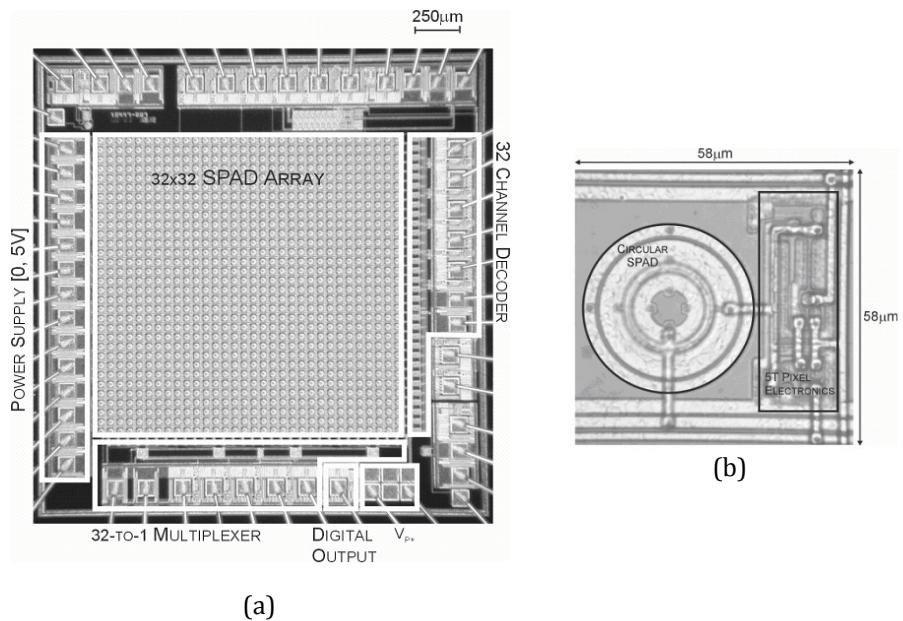


Figure 2.20. Photomicrograph of the (a) 32x32-pixel photon sensor and (b) of a zoom on pixel area.

The dead time is typically less than 40ns for the pixel. At node A, a Geiger voltage pulse of amplitude V_E reflects the detection of a single photon. The inverter stage converts this voltage swing pulse into a digital pulse. The transistor aspect ratios of the inverter are designed to set the input threshold voltage at 3.5V. The transmission gate feeds the detection signal to the column output line when the pixel is selected by the row selection decoder.

A photomicrograph of the chip is shown in Figure 2.20 (a). The chip size is 7mm². Figure 2.20 (b) shows the layout of a pixel consisting of the photodiode, the quenching circuitry and the column access. The 5T pixel occupies a square area of 58μmx58μm. The active area of the SPAD is 38μm².

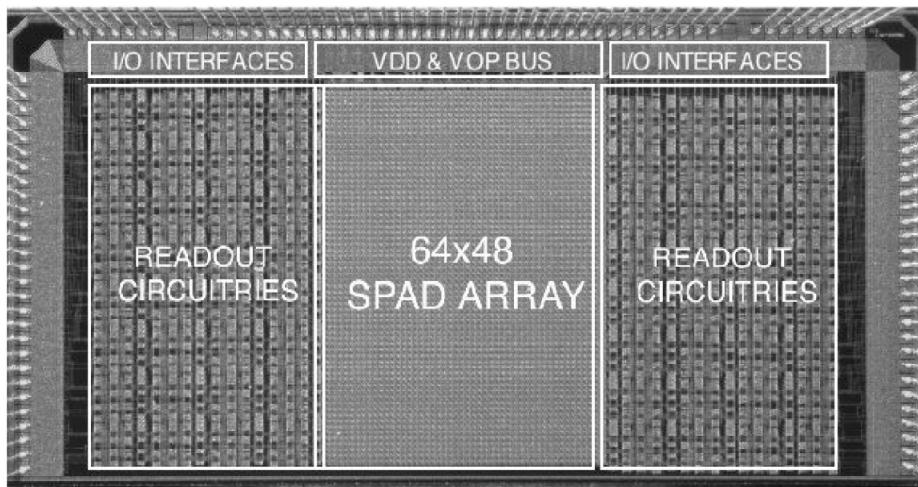


Figure 2.21. Photomicrograph of the sensor comprising 64x48 and based on an event-driven readout circuit.



Figure 2.22. Photomicrograph of the pixel in the 64x48 array.

The second design, comprising 64x48 pixels [157] was based on the event-driven readout circuit, described in Section 2.4.3. Figure 2.21 shows the photomicrograph of the sensor. The total chip area was 7.7x3.6 mm² whereas the pixel, shown in Figure 2.22, occupied an area of 45x45 μm². The pixel was based on a SPAD with an active area of 38μm² and on 10T circuit, according to Figure 2.17. In this implementation, N_{ROW} was 48, thus leading to 6 address lines, in agreement with Equation (2.14). The integrated circuit incorporated a chip-level readout circuit capable of reading out the data generated in all the 64 column-based event-driven readout, which allowed for a non-sequential and quasi-fully parallel frame acquisition. However, the column and therefore pixel dead time was excessively low, 600ns, thus restricting the sensor utilization only under very weak light signals. These limitations mainly originate from a low off-chip readout speed, partially due to the old 0.8μm CMOS technology used. Nonetheless, since the chip was fully functional and accurate under low light exposure, this sensor has also been utilized for SPAD characterization in this technology.

2.5.1 PHOTON DETECTION PROBABILITY

Photon detection probability was measured based on a measurement setup that has been assembled for this purpose. A 150W Xenon lamp was focused on the entrance slit of a monochromator, model Oriel MS 257. The output slit of the monochromator was used to illuminate the input port of an integrating sphere, thus providing a monochromatic light beam. The integrating sphere had three output ports, two of which were protected with baffles to prevent direct exposure from the input beam. As a result, in the baffle-protected ports, a monochromatic beam, whose intensity was uniformly distributed over the port aperture, was accurately obtained. At the first port, a calibrated S1226-8BQ photodiode from Hamamatsu was used as reference. Its photogenerated current was measured by means of a high precision electrometer, model 6517A from Keithley, and used to determine the photon flux at the output ports. The device under test (DUT) was placed on the second output port. In order to measure the photon counting rate in DUT, a PM6681 counter/frequency-meter from Fluke was utilized.

Based on the proposed setup, PDP measurements as a function of incident photon wavelength were obtained. Figure 2.23 shows the measured PDP as a function of the photon wavelength for a typical pixel of the 32x32 pixel-sensor with a nominal V_E of 5V at room temperature. It is higher than 20% between 430nm and 570nm with a peak at 26% at 550nm. At 700nm the PDP is still 8% without any post process treatment of the silicon chip surface.

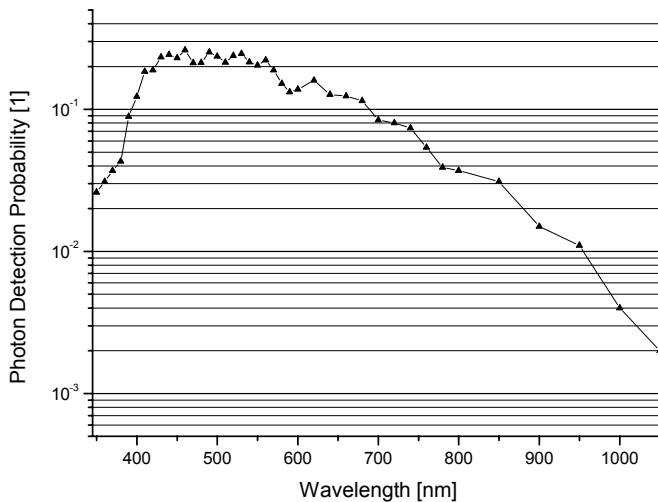


Figure 2.23. Photon detection probability as a function of photon wavelength for the p+/p-well/deep nwell SPAD in a HV 0.8 μ CMOS technology.

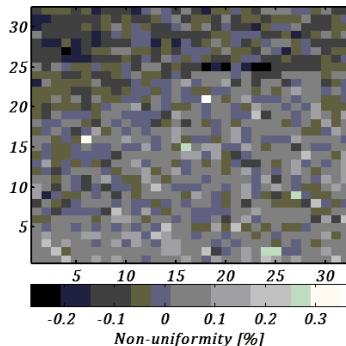


Figure 2.24. PDP non-uniformity across the sensor.

In Figure 2.24, a picture of PDP non-uniformity across the 32x32-pixel array is shown. The sensor was illuminated with a uniform light of intensity 2.5×10^{-5} W/cm² to validate the PDP uniformity. Prior to the illumination measurement, DCR had been measured for all pixels in the array and every pixel's mean value has been subtracted from the acquired image. This calibration procedure was performed to assess sensitivity uniformity independently from pixel-to-pixel

DCR variation. The SPAD array exhibited very good¹ pixel uniformity in terms of PDP, fill factor and dead time [156].

Very similar PDP performance has been measured for the sensor based on 64x48 pixels [157].

2.5.2 DARK COUNT RATE

Figure 2.25 plots DCR as a function of pixel location for the 32x32-pixel sensor. DCR was measured under nominal V_E at room temperature and at $T=0^\circ\text{C}$. At room temperature, the small active area of the SPAD and the outstanding cleanliness of the CMOS process lead to a mean value of the DCR of 350Hz over the whole sensor array. For $T=0^\circ\text{C}$, the mean value of the DCR dropped below 75Hz.

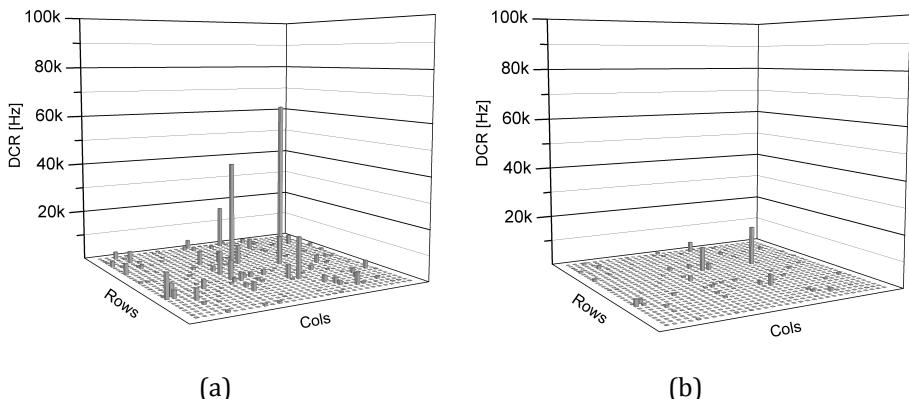


Figure 2.25. DCR as a function of pixel location of the 32x32-pixel sensor under nominal V_E : (a) at room temperature and (b) at $T=0^\circ\text{C}$.

DCR was also measured on the sensor comprising 64x48 pixels the result is shown in Figure 2.26. The mean value of DCR on the latter sensor was 376Hz at room temperature, which was relatively similar to the results obtained from the former one. In both cases, most of pixels exhibit outstanding DCR performance. Nonetheless, a small fraction of pixels show higher DCR, reaching 70 kHz in some cases. Sensitivity measurements showed that those pixels were operating similarly to low-DCR pixels in terms of PDP and actual active area.

¹ Within noise and systematic error limits.

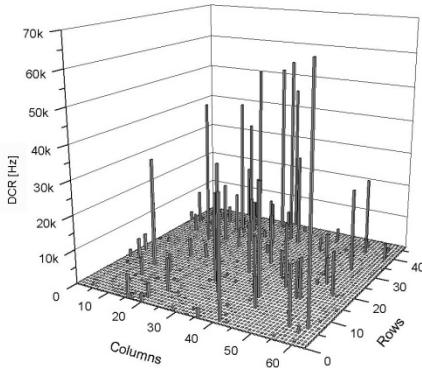


Figure 2.26. DCR as a function of pixel location of the 64x48-SPAD sensor under nominal V_E and at room temperature.

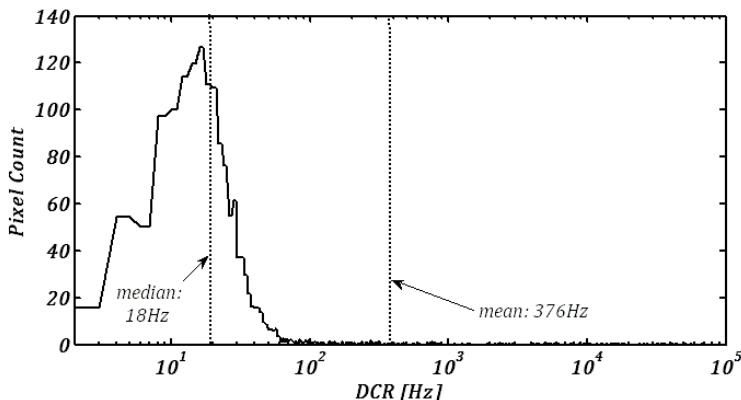


Figure 2.27. DCR distribution over 64x48 pixels under nominal V_E and at room temperature.

A distribution plot of DCR is shown in Figure 2.27 to provide some insights on the statistics of DCR over a large number of pixels. As can be seen in the picture, most pixels exhibit a DCR much lower than the mean value. Actually, the median value of DCR was only 18Hz. This result leads to an outstanding median DCR density of $0.5\text{Hz}/\mu\text{m}^2$. In many applications, the median DCR is a more suitable noise parameter than the mean value. This is particularly true in applications where high-DCR pixels may be deactivated and replaced by a linear combination of neighboring pixels. Moreover, since the mean DCR stems from a

low number of very high-DCR pixels, it may suffer from high chip-to-chip variability. Median DCR on the other hand offers a more appropriate measure for comparison purposes.

The origin of high-DCR SPADs within arrays dominated by low-DCR pixels may be explained by local defects. These imperfections are typically due to lattice defects or impurity contaminations that lead to G-R centers with energy level very close to middle of silicon bandgap. Spatial correlation measurements showed that pixels with high DCR are not correlated with surrounding pixels. This result shows that local defects impact a very small area compared to the pixel area, which was, in this case, $45 \times 45 \mu\text{m}^2$. Moreover, since no correlation was found in neighboring pixels, the solution in which a pixel output is replaced by a combination of neighbor pixels becomes effective. In applications where such pixel replacement is not acceptable, it is actually possible to rely on high-DCR pixels as far as sensitivity is important. In those pixels, illumination measurements showed that, besides their high noise level, PDP and active area uniformity were not impaired by the same defects causing high DCR.

2.5.3 TIMING RESOLUTION

Timing resolution was measured based on the TCSPC technique. A fast laser source with pulse width of 80ps and repetition rate of 40MHz emitting a beam with a wavelength of 637nm was used to illuminate the SPAD. The time interval between the laser output trigger and the leading edge of the SPAD signal was measured via a high performance oscilloscope operating as a TDC. The oscilloscope, a LeCroy 8600A, featured 20GS/s and 3ps of uncertainty. A histogram was built as the time interval measurements were repeated over very large number of times. In order to prevent the typical pile-up effect, optical neutral density filters were used to reduce the SPAD firing rate to a few tens of kHz. The resulting jitter measurement histogram is shown in Figure 2.28. The overall timing jitter was 115ps FWHM.

This measurement comprises jitter contributions due to the illumination source, to the SPAD, and to the readout circuit. A correction for the illumination pulse width yields an estimate of the circuit timing resolution (IRF) of 82ps FWHM. Note that, during this measurement, although a single pixel was investigated, the remaining pixels in the array of 32x32 SPADs were active. Since these pixels are likely to generate noise in the readout circuit by electrical coupling and digital noise in power lines, the measurement was in accordance with the scalability of SPAD to large arrays. Moreover, this additional contribution was

believed to justify the mismatch with respect to the 50-ps jitter result Rochas *et al.* had obtained with a similar device [20].

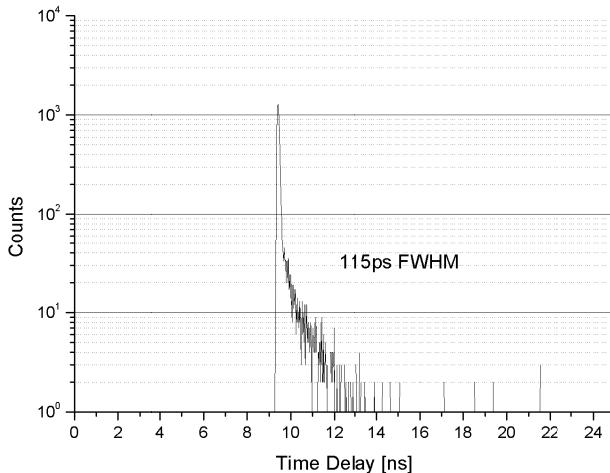


Figure 2.28. SPAD jitter measured on the 32x32 image sensor.

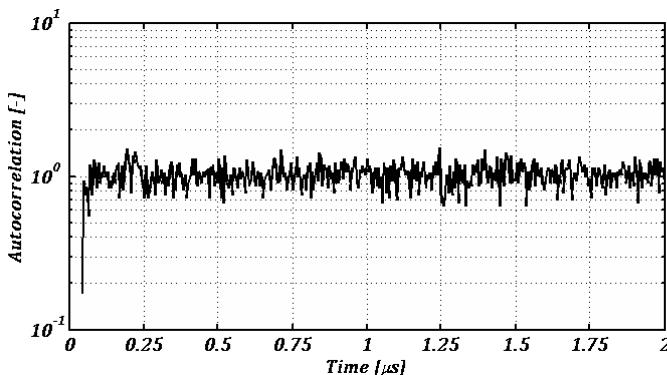


Figure 2.29. Autocorrelation measurement of the SPAD output signal for afterpulsing evaluation.

2.5.4 AFTERPULSE PROBABILITY

Afterpulsing effects in the 0.8 μm SPAD were evaluated using the array of 32x32 pixels. The measurement setup was based on an autocorrelation computation of the SPAD output signal. The output signal waveform was recorded under a

constant photon (and dark) count rate, and over a long period of time. A high-performance oscilloscope, LeCroy 8600A, was used to sample and store the waveform with a temporal resolution of 4ns. Since the SPAD dead time was 32ns, it was possible to sample 8 voltage values per photon detection, enough for afterpulsing evaluation. Furthermore, since the SPAD output delivers digital pulses with 30ns of duration, the autocorrelation was not applied directly to the voltage waveform. Instead, in order to reduce errors, a processing algorithm that calculates the autocorrelation of arrival times of Geiger pulses was conceived and utilized. In Figure 2.29, a plot of the normalized autocorrelation is shown.

The autocorrelation exhibited a measurement gap at the origin due to the dead time of 32ns. Indeed, the SPAD is unable to generate additional digital pulses within a window of approximately 30ns from each photon detection. After this period of time, a peak in the autocorrelation waveform would be expected, if the SPAD exhibited afterpulsing effects, followed by an exponential decay. As can be seen in Figure 2.29, the autocorrelation waveform is perfectly constant after the dead time, which shows that this device is completely free from afterpulsing effects.

2.5.5 CROSSTALK PROBABILITY

One of the most important concerns in view of building arrays of SPADs is crosstalk between pixels. As described in Section 2.3, previous implementations of SPADs in hybrid approaches were highly challenged by crosstalk. In this technology, crosstalk probability was evaluated based on the chip of 32x32 pixels. In that sensor, the pixel pitch was 58 μ m, lower than any previous reported implementation.

As described in Section 2.2.6, crosstalk effects may be decomposed in two components: electrical crosstalk and optical crosstalk. While the mechanisms of electrical crosstalk in SPADs are similar to those affecting conventional photodiodes, optical crosstalk in SPADs is rather unique. During an avalanche, some photons may be emitted from a SPAD, thus leading to an electroluminescence phenomenon. These photons may be detected by neighboring pixels in an array of SPADs and cause optical crosstalk.

Crosstalk effects were evaluated based on two different approaches. In the first approach, the availability of high DCR pixels in the SPAD array was used to estimate optical crosstalk performance. Since photons are expected to be emitted during an avalanche, crosstalk effects should arise even from dark counts. Consequently, by comparing the DCR between two neighbor pixels, one

may estimate if optical crosstalk occurs. As shown in Figure 2.25 (a), on a given pixel, the DCR was as high as 60kHz. If the neighbor pixels were suffering from optical crosstalk, in principle, they should exhibit non-negligible DCR. In fact, neighbor pixels exhibited dark count rates as low as 50Hz. As a result, if optical crosstalk occurs, it is probably below 0.1%.

In the second approach, overall crosstalk was measured by illuminating a single pixel in the center of the array using a highly focused laser beam through the optics of a microscope. A custom mechanical holder was fabricated to fix an optical fiber on the output port of the microscope. Typically, the output port of the microscope is designed to hold an image sensor. It is possible consequently to adjust the microscope lenses to enforce image formation between the output port and the sample plane. As a result, the light injected from the core of the optical fiber was imaged onto a single point at the sample plane, where the device under test was placed. Furthermore, the light beam spot was visible through the binocular port of the microscope, thus, it was possible to adjust the beam to excite only the active area of a single pixel. Figure 2.30 shows the resulting plot of the crosstalk measurements whereby the excited pixel's neighbors showed a response less than 0.01% of the center pixel.

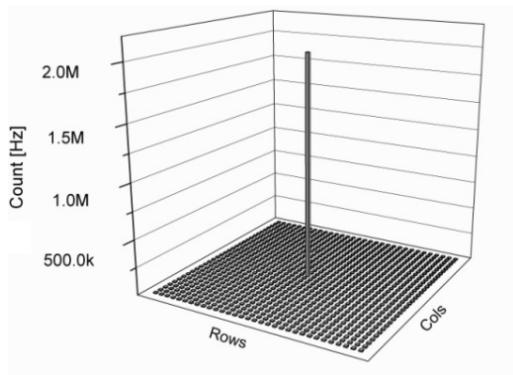


Figure 2.30. Crosstalk measurement result.

As expected, optical crosstalk is negligible due to the small capacitance in SPADs fully integrated with front-end electronics. Electrical crosstalk was eliminated by design. Minority carriers diffusing in the substrate cannot effectively reach the multiplication region of a pixel since they would be collected by the deep n-tub/p-sub junction [156].

2.6 SPAD INTEGRATION IN 0.35-MICRON CMOS

In this Section, the successful technology migration of SPADs from the 0.8 μm towards deep-submicron technology is described. To the best of our knowledge, we were the first to demonstrate deep-submicron SPAD devices. In view of the advantages provided by emerging high-voltage CMOS technologies, the candidate technology was a 0.35 μm HV CMOS process [153]. Indeed, as described in Section 2.3, these CMOS technologies require particular fabrication procedures that are in line with basic requirements for SPAD fabrication. Furthermore, the 0.35 μm technology provides very important advantages over 0.8 μm . Besides additional implant layers, typical in HV processes, it features 4 metal layers for interconnection, two polysilicon layers, and a number of different transistors ranging from 3.3V to 50V [153, 158]. Transistor cutoff frequencies are considerably higher than in 0.8 μm . Moreover, the process offers additional miniaturization features introduced in CMOS since then, e.g. via/contact stacking. As a result, this technology represents a very good compromise between, on the one hand, miniaturization and speed capabilities required for circuits and, on other hand, low-doping concentrations and fabrication cleanliness required for SPAD performance.

The basic device described in this Section has been originally published in [154]. Thanks to this contribution, the paper was elected among DATE's 30 most influential articles of the last 10 years (1997-2007). Since the publication of [154], significant improvements in PDP were achieved and a complete characterization was published in [159]. Moreover, different device sizes have been designed. In this Section, characterization data is reported for a circular device with a diameter of 7 micrometers. Characterization data reported here was based on the larger sensor array designed in this technology [160, 161]. This sensor is described in detail in Chapter 3, where a photomicrograph of the pixel shows the fabricated SPAD.

The SPAD structure, whose cross-section is depicted in Figure 2.31, consists of a circular dual junction structure: p+ anode/deep n-well/p-substrate. The p+ anode/deep n-well junction forms the avalanche multiplication region where the so-called Geiger breakdown occurs. The deep n-well/p-substrate junction allows the p+ anode to be biased independently from the p-substrate. It additionally prevents electrical crosstalk due to minority carriers diffusing in the substrate, and improves the timing jitter of the SPAD, similarly to the SPAD in 0.8 μm CMOS.

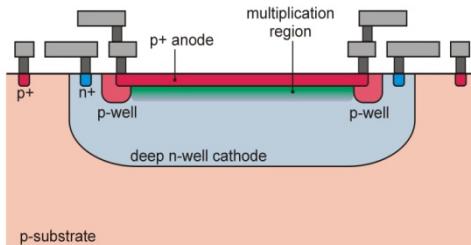


Figure 2.31. Cross-section of the SPAD structure in a $0.35\mu\text{m}$ CMOS.

A p-well guard-ring surrounds the p+ anode to prevent premature breakdown. Breakdown voltage measurements have been done on the SPAD structure. In this technology, without the guard-ring, the breakdown voltage of a typical p+/deep n-well junction is about 11V. The use of the guard-ring allowed the breakdown voltage to be increased to 17.7V. Conversely, the breakdown voltage of the deep n-well/p-substrate junction is typically 70V, thus allowing the deep n-well cathode to be comfortably biased with respect to substrate.

2.6.1 PHOTON DETECTION PROBABILITY

In order to evaluate PDP performance, the same optical setup described in Section 2.5.1 was utilized. The nominal power supply of the $0.35\mu\text{m}$ is 3.3V. Thus, when utilizing only low-voltage transistors in pixel circuits, V_E of SPADs is typically limited to 3.3V. In some cases, depending on the circuit topology, it is possible to increase V_E of a junction built-in voltage, i.e. approximately 0.7V. As a result, PDP was characterized for V_E of 3.3V and 4V. While lower values of V_E are irrelevant in most applications, higher values of V_E require the utilization of high-voltage transistors at pixel level. Moreover, as described in the next Section, DCR increases sharply with V_E .

The measured PDP as a function of photon wavelength is depicted in Figure 2.31, with V_E as a parameter. At the nominal V_E , the maximum value of PDP was 35.4% at 460nm. It is enhanced to 40% when V_E is increased to 4V. These results were unexpectedly better than PDP measured in the devices fabricated in the $0.8\mu\text{m}$ CMOS. Enhanced sensitivity in the UV/blue range may be explained by the lack of passivation on the $0.35\mu\text{m}$ device. Indeed, a relatively opaque polyimide layer forming the passivation layer in the $0.35\mu\text{m}$ process had to be prevented by design. As a result, the $0.35\mu\text{m}$ device eventually obtained a thinner and better optimized dielectric stack. UV/blue range sensitivity was also justified by a shallower p+/deep n-well junction, when compared to the

$0.8\mu\text{m}$ process. The response to red/IR range was relatively similar to the old device. In fact, the depth of the deep n-well/p-substrate in this process is very similar to the equivalent depth in the $0.8\mu\text{m}$ CMOS.

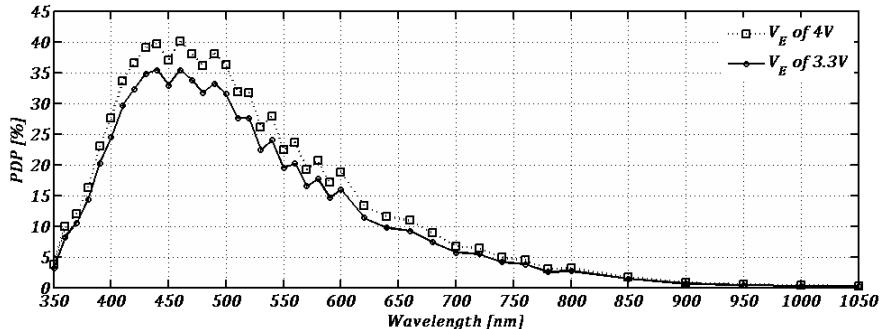


Figure 2.32. PDP as a function of wavelength, with excess bias voltage as a parameter.

2.6.2 DARK COUNT RATE

In view of the size of the array investigated here, the measurement of DCR distribution over a very high number of Geiger-mode photodiodes became possible, and it is therefore reported for the first time. The DCR distribution over 16384 pixels, at several temperatures and for V_E of 3.3V, is shown in Figure 2.33. At 27°C , the median value of DCR was 694Hz whereas the mean value was higher, at 2.4kHz, mostly due to a small number of highly noisy pixels. The measured median DCR also stays fairly constant among devices from the same fabrication batch. The median DCR measured at room temperature over a few samples of the design reported in this work spreads only 96Hz, from the best device to the noisiest one considered.

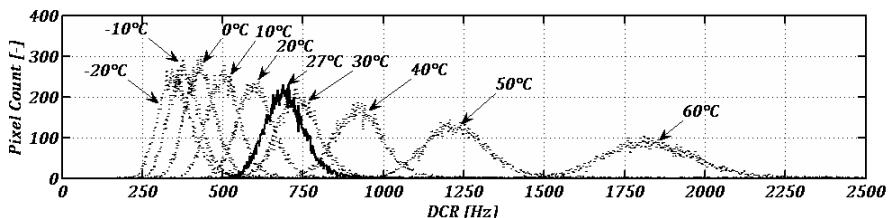


Figure 2.33. DCR distribution over the array for V_E of 3.3V and with temperature as a parameter.

As can be seen in Figure 2.34, the median DCR is a strong function of temperature for temperatures higher than 0°C. Below this temperature, only limited performance gain is obtained when reducing temperature due to the tunneling contribution to DCR. Also visible in Figure 2.34, the statistical spreading of DCR over pixels increases monotonically with temperature from -20°C up to 60°C.

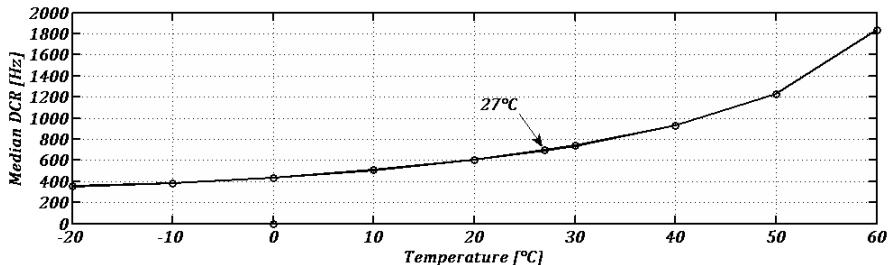


Figure 2.34. Median DCR as a function of temperature.

In Figure 2.35, the DCR of a typical pixel is plotted as a function of V_E , using the temperature as a parameter. Since the graph is a semi-log plot, it can be seen that DCR showed the typical exponential dependence from V_E .

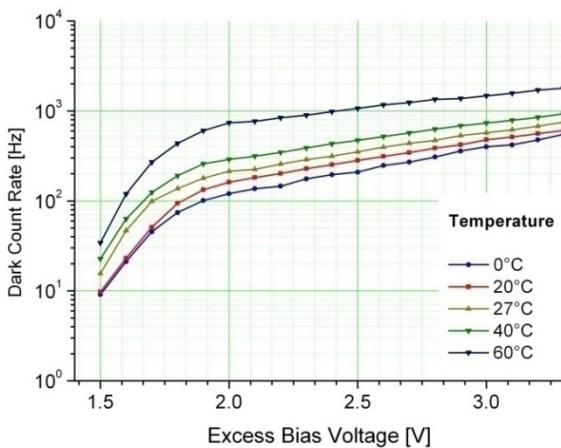


Figure 2.35. DCR of a typical pixel as a function of V_E .

Although it is perfectly acceptable in most applications, the overall DCR performance of the 0.35μm CMOS SPAD was relatively inferior to the 0.8μm

equivalent structure. The median value of DCR density, at room temperature, has increased from $0.5\text{Hz}/\mu\text{m}^2$ in the previous device to $18\text{Hz}/\mu\text{m}^2$. The higher dark count rates may be explained by a number of factors. In the $0.35\mu\text{m}$ CMOS technology, the doping concentrations of both the p+ and deep n-well layers are significantly higher than in $0.8\mu\text{m}$ CMOS [153, 155]. As described in Sections 2.2 and 2.3, when both anode and cathode implants are very high, tunneling-induced carrier generation is enhanced in the high field region of the SPAD. Furthermore, since the $0.35\mu\text{m}$ technology involves a significant number of additional back and front-end fabrication steps, the risk of contamination during the IC processing is higher than in $0.8\mu\text{m}$. As a result, the density of G-R centers and traps is expected to be higher in this technology, thus favoring thermal generation as well as trap-assisted tunneling generation. Process cleanliness is a difficult issue to deal with from a design standpoint.

2.6.3 TIMING RESOLUTION

Similar to the timing resolution characterization of $0.8\mu\text{m}$ SPADs, described in Section 2.5.3, timing resolution of the new $0.35\mu\text{m}$ SPAD was measured based on the TCSPC technique, using a similar setup. An extremely fast laser source featuring femtosecond pulse width and repetition rate of 80MHz was used to illuminate the SPAD. The time interval between the laser output trigger and the leading edge of the SPAD signal was measured via a high performance oscilloscope operating as a TDC. The oscilloscope, a LeCroy 8600A, featured 20GS/s and 3ps of uncertainty. A histogram was built as the time interval measurements were repeated over very large number of times. In order to prevent the typical pile-up effect, optical neutral density filters were used to reduce the SPAD counting rate to a few tens of kHz.

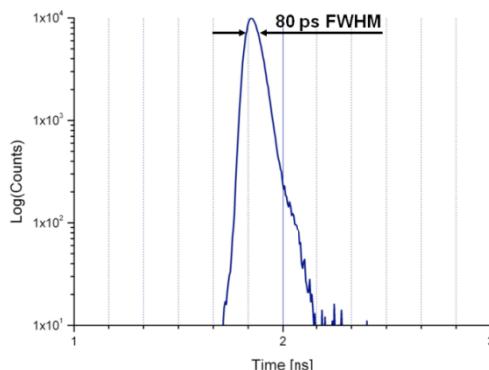


Figure 2.36. Measured SPAD jitter for the proposed $0.35\mu\text{m}$ CMOS SPAD.

The timing resolution measurement is shown in the semi-log histogram of Figure 2.36. The FWHM of the overall histogram peak was 80ps. Since a femtosecond laser source was utilized, its effect in timing resolution was negligible. As a result, the timing resolution of the SPAD is the measured IRF.

2.6.4 AFTERPULSE PROBABILITY

Afterpulsing effects in the $0.35\mu\text{m}$ SPAD were evaluated using an integrated passive quenching circuit providing a dead time of 40ns. The measurement setup was based on an autocorrelation of the SPAD output signal, as described in Section 2.5.4. Since the SPAD output signal generates digital pulses with duration of approximately 35 ns, the autocorrelation was not applied directly to the voltage waveform. Instead, a special acquisition and processing setup that calculates the autocorrelation of arrival times of Geiger pulses was designed and developed. From the autocorrelation measurement, it is possible to quantitatively compute afterpulse probability, as described in [55]. Figure 2.37 shows a plot of resulting afterpulse probability as a function of dead time. In the picture, a plot of the measured autocorrelation is also shown. Unlike the autocorrelation measurement obtained with the $0.8\mu\text{m}$ SPAD, this device clearly showed a peak and exponential decay in the autocorrelation waveform, after the dead time.

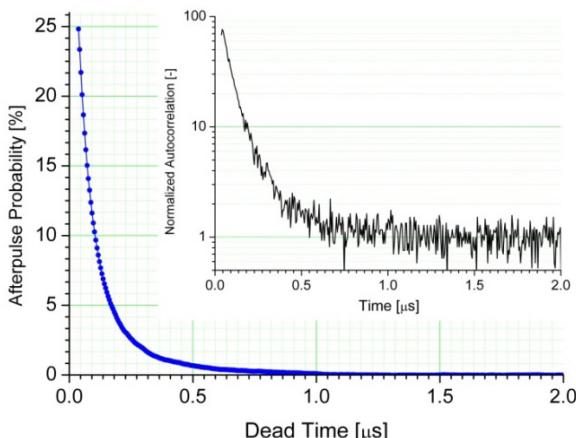


Figure 2.37. Afterpulse probability as a function of dead time for the proposed $0.35\mu\text{m}$ CMOS SPAD.

Afterpulsing effects are very pronounced in this implementation. At the shortest dead time, it was as high as 25%. As described in Section 2.2.4, when a Geiger discharge occurs, a very large number of carriers traverse the junction and some of them may be captured by trapping centers. The release of these carriers produces afterpulses. As a result, the density of deep-level traps in the multiplication region is of high importance for afterpulsing. In the 0.35 μm CMOS sensor under test, afterpulsing effects were thus probably due to inferior technology cleanliness, when compared to the 0.8 μm CMOS.

2.6.5 CROSSTALK PROBABILITY

In this technology, crosstalk probability was evaluated based on an alternative setup. The output signals of two adjacent SPADs in arrays were analyzed by means of a cross-correlation measurement. Similarly to the setup used in afterpulsing measurements, the proposed setup computed the cross-correlation of Geiger pulses detection times, over a very high number of detections. Thanks to the availability of different SPAD arrays, crosstalk probability was evaluated for two pixel pitches. Measured cross-correlation of adjacent pixels at pitches of 25 μm and 35 μm are shown in Figure 2.38 (a) and (b), respectively.

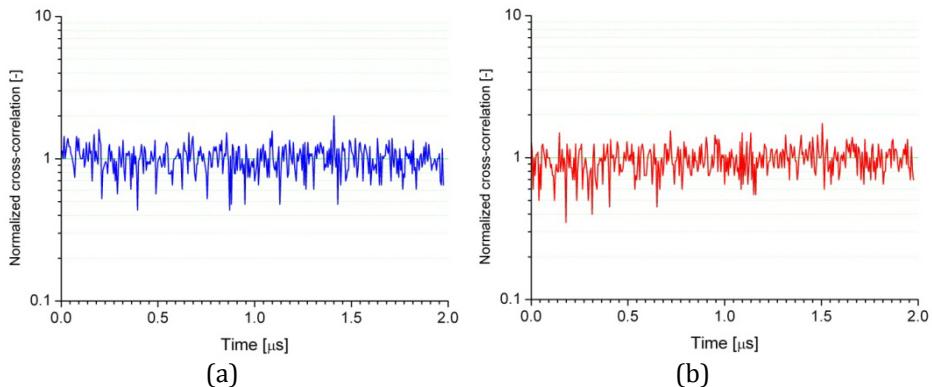


Figure 2.38. Cross-correlation between the output signals of two adjacent SPADs in arrays with: (a) pitch of 25 μm , and (b) pitch of 35 μm .

As can be seen in Figure 2.38, no correlation was detected at the origin of cross-correlation measurements. It was then concluded that at these pixel pitches, crosstalk probability was negligible.

2.7 SPAD INTEGRATION IN 130-nm CMOS

In view of a further migration of SPADs towards a technology platform that provides sufficient level of miniaturization, the design of a new device structure based on a state-of-the-art CMOS imaging process [162] has been investigated. The results of this investigation, which were published in [163, 164], are described in this Section. The SPAD presented here is amenable to the design of large arrays and, in principle, it enables the choice of any front-end and readout architectures proposed in this thesis. Due to the reduced breakdown voltages, the structure is interesting in the context of applications where power dissipation is a critical factor. Advanced CMOS technology provides a level of miniaturization that is important to design smaller front-end circuits. Thus fill factor can be improved and/or new functionality can be added in SPAD arrays.

Figure 2.39 (a) shows the cross-section of the proposed SPAD. It consists of a p+ anode within an n-well cathode where p+ and n-well are the implantations of source/drain and bulk, respectively, of standard 1.2V PMOS transistors. This configuration allows for a full isolation of the p+ anode from the p-substrate. In addition, the configuration also enables coupling relatively high bias voltages necessary in SPADs to low-voltage CMOS logic. The planar multiplication region was enabled by means of a p-well guard ring, where p-well is the bulk of isolated 1.2V NMOS transistors. A useful feature of this technology is the availability of a buried n-type isolation layer that allows for a full isolation of p-well within n-well from p-substrate. This layer was used to prevent a punch-through of the p-well guard-ring to p-substrate. The combination of n-well and buried n-isolation layer was the lowest doping concentration feasible in this technology for the cathode.

The photomicrograph of the proposed device is shown in Figure 2.39 (b). The structures visible in the figure include the octagonal anode, guard ring, and metal interconnect. The additional function of the metal is that of preventing the guard ring to be exposed to light for characterization purposes. The anode measures 10 μm in the picture.

A major improvement in this design is the physical separation of the STI interface from the SPAD multiplication region, thus having a beneficial impact on DCR. In standard deep-submicron CMOS, it is not possible to prevent STI by means of a drawn layer. As a general rule, STI is etched everywhere so that all the p+ and n+ implantations are surrounded by STI to improve isolation. It is possible however to draw a polysilicon gate of a standard transistor that represents a stop mask for n+ and p+ implantations. STI can therefore be

effectively separated from the surroundings of the anode by drawing a superposition of polysilicon, thin-gate-oxide, p+, and diffusion layers around the p+ anode. In order to prevent the formation of a high-electric field within the thin-gate-oxide layer, the polysilicon gate is kept at the same potential as the p+ and p-well layers, ensured by ohmic contacts.

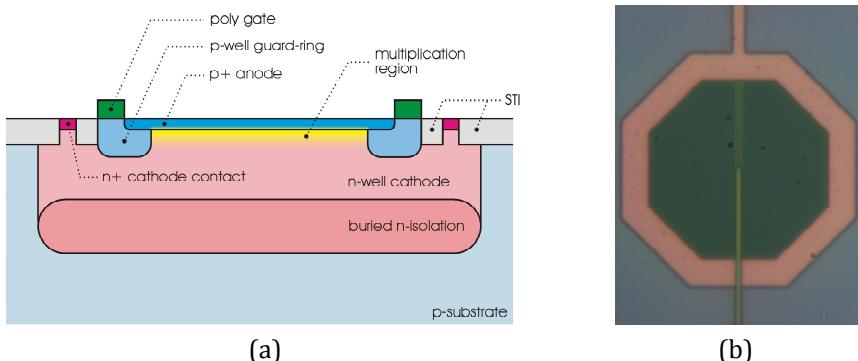


Figure 2.39. (a) Cross-section of the SPAD in a 130nm CMOS: p+ anode within n-well cathode, based on the typical p-well guard-ring (not to scale). (b) Photomicrograph of the actual SPAD.

Since the polysilicon gate prevents the p+ to be implanted, the result of the fabrication process is a p-well extension of p+ completely free of STI, whose extension can be adjusted as desired. Around the p-well guard-ring, there is still a STI ring. This STI interface, in particular at the depletion region between the p-well guard-ring and n-well cathode, may induce a large density of generation centers. Nonetheless, the p-well guard-ring lowers the electric field around the SPAD sufficiently to prevent impact ionization but it is enough to collect most of the carriers generated at the STI/p-well interface. As a result, this structure allows a small parasitic current to flow from cathode to anode without triggering avalanche events, thus reducing DCR.

The diode was tested in a number of ways. First, the I-V characteristic was measured statically using a standard semiconductor analyzer. Figure 2.40 shows the I-V characteristics of the diode in reverse bias. The picture shows that the reverse current close to breakdown voltage approaches 600pA. This relatively large current would suggest that DCR tends to be high. For instance, if we suppose that all the carriers were collected by the multiplication region, the device would not properly operate in Geiger mode as its DCR would be of the order of 3-4 GHz. In this Section, it will be shown that the structure properly

operates in Geiger mode and exhibits acceptable levels of DCR. As previously described, most of the reverse current is expected to be generated at the periphery of the SPAD, at the STI/p-well/n-well interface, where impact ionization is prevented by the p-well guard-ring.

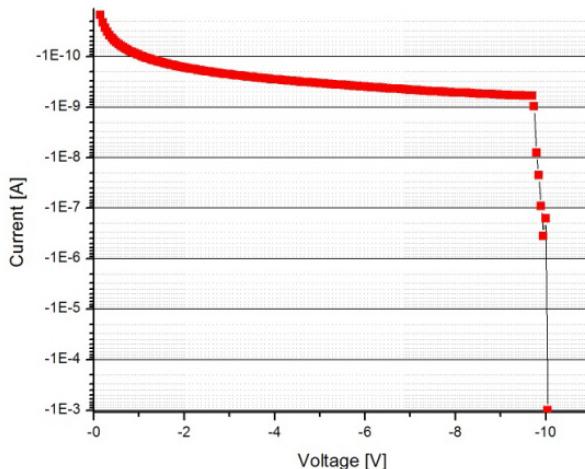


Figure 2.40. I-V Characteristic of proposed structure at room temperature.

The diode was operated in Geiger mode using both passive and active quenching circuitries. The schematic setup of the passive quenching configuration is shown in Figure 2.41 (a). The $20\text{k}\Omega$ quenching resistance R_Q , placed at the anode of the p-n junction, causes an increase of its potential in case of avalanche. If the reverse bias voltage across the junction decreases towards breakdown voltage, the avalanche current is reduced to a level of the order of tens of microamperes and eventually stops. Avalanche quenching is followed by an exponential recharge to allow the voltage across the junction to return to its initial value of V_{OP} . This voltage satisfies the following condition

$$V_{OP} = |V_{BD}| - V_E, \quad (2.15)$$

where V_{BD} , and V_E are the breakdown and excess bias voltage, respectively. The plot of Figure 2.41 shows the recharge phase of the probed voltage as a function of time for different values of V_{OP} . The simple exponential behavior is due to the RC recharge. R accounts for the resistive path to ground and C for the overall capacitance at the probing node. Due to the fact that this device does not have integrated quenching circuitry, the term C is dominated by the parasitic

capacitance of external components. It has been estimated to be 10pF, a factor 70 to 100 larger than the expected SPAD junction capacitance. The dead-time under this condition is estimated to be 450ns. As described in Section 2.2.4, afterpulsing probability depends independently on dead time, due to trap lifetimes, and on the parasitic capacitance as it increases the number of carriers traversing the multiplication region, thus filling up traps. Thus, a characterization of afterpulsing probability under this condition is meaningless, since it gives no insight on the true potential of the device and of its internal capacitance. We assume that afterpulsing characterization under the present condition would be incorrect and thus irrelevant.

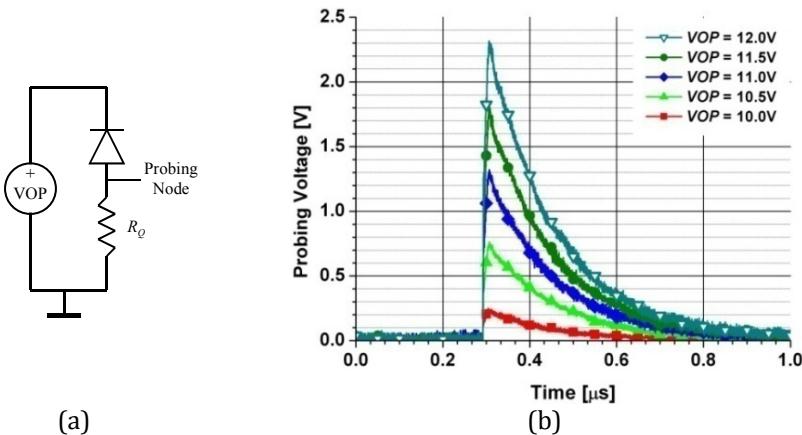


Figure 2.41. (a) Passive quenching setup. A probe senses the voltage across quenching resistor R_Q . (b) Plot of the time-dependent output of the probe voltage as a function of time for several values of V_{OP} .

In order to precisely investigate DCR and PDP in this work independently of dead time and afterpulsing effects, an alternative setup involving the use of an external gated active recharge circuit combined with TCSPC was used. This technique is often used in the characterization of III-V SPADs, which exhibit significantly higher DCR and afterpulsing effects [165]. In most active quenching and recharge setups, an active circuit replaces the quenching resistance, thus allowing one to reduce the recharge time to a few tens of nanoseconds. The proposed experimental setup is based on a commercially available gated active recharge circuit [166] and is described as follows.

V_{OP} is maintained below V_{BD} at the beginning of each event measurement cycle. V_{OP} is then quickly increased to its nominal level, according to Equation (2.15),

so as to recharge the SPAD. The time interval between the SPAD recharge signal and the moment a first Geiger event occurs is measured using a high precision time-to-digital converter. V_{OP} is subsequently kept below V_{BD} during a hold-off time of the order of 500 μ s. This hold-off time is chosen large to prevent any afterpulses. As this measurement cycle is repeated a large number of times, a histogram is built conforming to the TCSPC measurement technique. The resulting histogram shows an exponential decay, similarly to a typical fluorescence lifetime measurement [57]. The inverse of the mean value of the histogram provides the desired counting rate. Any timing offset between full SPAD recharge and Geiger pulse leading edge is removed prior to computing the counting rate. The active recharge circuit conveniently performs fast active recharge and also provides a trigger signal used to compute time-interval measurements as described above. As detector dead time and afterpulsing do not impair the measurement even at high counting rates, this technique is used to measure DCR as well as PDP.

In order to correctly characterize the measurements presented hereafter, the breakdown voltage was firstly measured for the structure as a function of temperature. Hence, V_{OP} was set for a given temperature to reflect the correct excess bias voltage according to Equation (2.15). Figure 2.42 shows a plot of the breakdown voltage as a function of temperature.

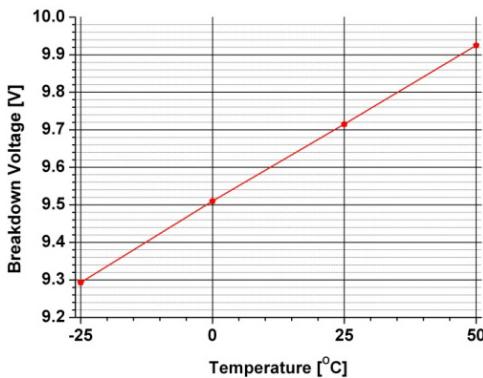


Figure 2.42. V_{BD} as a function of temperature.

2.7.1 PHOTON DETECTION PROBABILITY

PDP measurements were performed for two excess bias voltages, within the entire spectrum of interest (350–1000nm). These measurements were based on

the same optical setup described in Section 2.5.1. Furthermore, although DCR is presented in next Section, full DCR measurements were performed previously to the PDP characterization. As a result, the mean value of DCR contribution was suppressed from each counting rate used in the measurement of PDP. Figure 2.43 shows a plot of the PDP at room temperature.

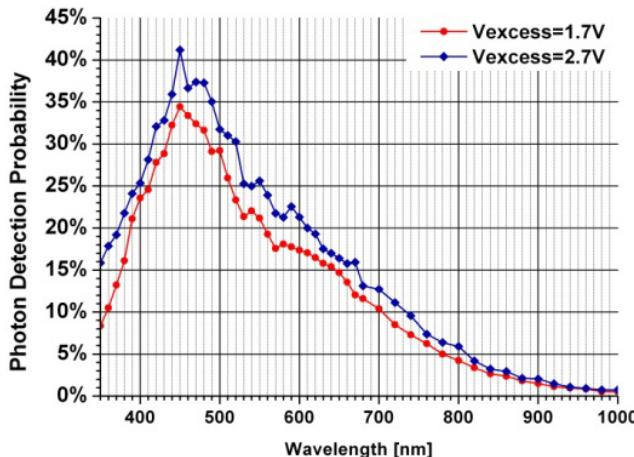


Figure 2.43. PDP as a function of photon wavelength, with V_E as a parameter.

PDP outperformed all expectations as its measurements showed values in the range of previously reported SPADs in $0.8\mu\text{m}$ and $0.35\mu\text{m}$ CMOS technologies, whose multiplication regions are wider and deeper. However, the result of a shallower multiplication region was a shift of the maximum of detection probability from 550nm in $0.8\mu\text{m}$ devices to 450nm in this technology. Relatively good PDP performance partially resulted from the use of enhanced dielectrics for optical detection, available in this imaging CMOS technology [162]. In Figure 2.39 (b), it is possible to notice a darker region in the middle of the SPAD, where optimized dielectrics was used, if compared to the remaining area of the picture, where only partial optimization was used. This darker region suggests that the light reflection coefficient at the center of the SPAD was noticeably lower than it would have been if we utilized a non-optimized optical stack. Note that, since the DCR had been measured previously to PDP, it is not responsible for any artificially increased PDP, which would be thought to falsify PDP measurements.

2.7.2 DARK COUNT RATE

In $0.8\mu\text{m}$ CMOS SPADs, DCR can be as low as a few tens of Hertz and it is a strong function of temperature and of excess bias voltage, as described in Section 2.5.2. In this 130nm CMOS technology however, DCR was very high, at least based on the proposed structure. Figure 2.44 (a) shows a plot of DCR as a function of temperature for two excess bias voltages, measured using the TCSPC method described above.

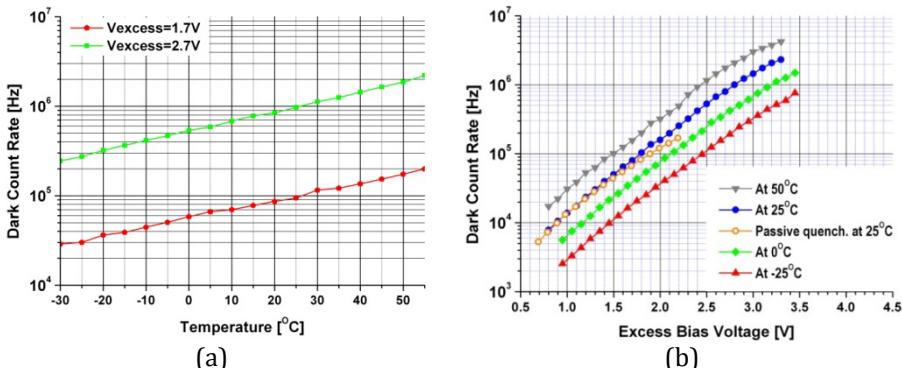


Figure 2.44. (a) DCR dependence on temperature. (b) DCR as a function of excess bias voltage at four temperatures. Temperatures are set with an uncertainty of $\pm 2^\circ\text{C}$.

Besides its higher absolute values, if compared to previous devices proposed here, DCR also exhibits weaker dependence on temperature. This suggests that DCR has a non negligible tunneling contribution. In fact, this behavior was expected in view of the relatively higher doping levels of both p+ and n-well layers available in this CMOS technology. Figure 2.44 (b) shows a plot of DCR as a function of V_E for four different temperatures based on the TCSPC measurement. It also shows a curve of DCR measured using the passive quenching setup of Figure 2.41 (a), for a temperature of 25°C . Since the measurements based on the passive quenching setup strongly saturate due to dead time, the errors in those measurements compared to the TCSPC method are significant for any DCR higher than a few tens of kHz.

As can be seen in Figure 2.44 (b), DCR reaches prohibitive levels as V_E exceeds 2V. However, depending on the amount of parasitic light in a given application, higher levels of DCR may be tolerated. For instance, as described in Chapter 3, noise contributions in a TOF 3D image sensor is in general dominated by the

parasitic background light when it operates outdoor. In such cases, in order to improve PDP and increase overall signal-to-noise ratio, higher values of V_E could be acceptable.

2.7.3 TIMING RESOLUTION

Timing resolution was also characterized based on the TCSPC technique. The measurement setup was the one utilized in jitter characterization of previous devices, described in Section 2.5.3. A fast laser source with pulse width of 80ps and repetition rate of 40MHz emitting a beam with a wavelength of 637nm was used to illuminate the SPAD. The time interval between the laser output trigger and the leading edge of the SPAD signal, operated with the active recharge circuit, was measured via a high performance oscilloscope operating as a TDC. A histogram was built as the time interval measurements were repeated a large number of times. In order to prevent the typical pile-up effect, optical neutral density filters were used to reduce the SPAD firing rate to a few tens of kHz. The resulting jitter is reported in the normalized histogram shown in Figure 2.45.

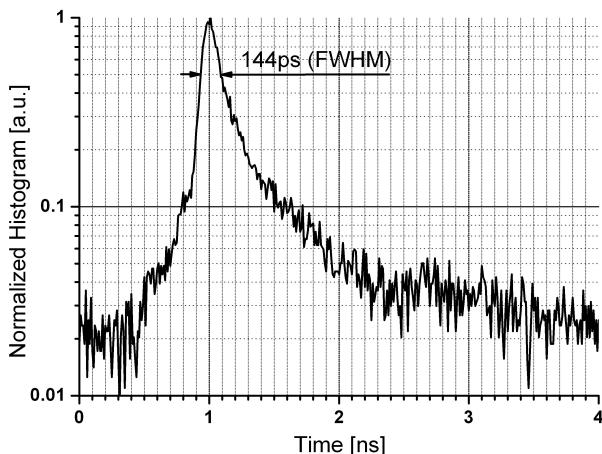


Figure 2.45. Timing resolution measurement.

The FWHM value of the resulting pulse was 144ps. Since this result comprises the illumination pulse width of 80ps, a correction is necessary to obtain the sole circuit IRF. After the correction, the circuit timing resolution was 119ps FWHM. Nonetheless, this timing uncertainty also includes the uncertainty of the hybrid active recharge setup. It is believed that lower timing uncertainties can be expected when the SPAD is monolithically integrated with its front-end circuit.

2.8 CONCLUSION

In this Chapter, the operating principle of single-photon avalanche diodes was described. Following the description of the most important performance parameters, a study of the relevant state-of-the-art for SPADs was proposed, from a device engineering standpoint. In this thesis, the investigation of large SPAD arrays in CMOS technology was carried out, for the first time. Previously in the art, only small and not scalable arrays of pixels were implemented, thus leaving a number of challenges unaddressed. In particular, the statistical study of the performance of large arrays of SPADs implemented on a single CMOS integrated circuit was missing in the art. Moreover, existing front-end circuits in the art were not suitable for large pixel arrays.

In this Chapter, the evaluation of a large array comprising 1024 pixels in $0.8\mu\text{m}$ CMOS was described. In that design, the quenching resistor was replaced by a single MOS transistor, thus achieving passive quenching with optimized pixel area. The array characterization also demonstrated that, although some pixels exhibit very high DCR, all the pixels behaved similarly in terms of photon detection probability. Outstanding pixel PDP uniformity was reported. Furthermore, since at least one terminal of the SPAD needs to be biased with a relatively high voltage (V_{OP}), -25.5V in that design, it is typically necessary to connect at least one terminal of all the SPADs in an array to the same voltage. In view of the high voltage required for V_{OP} , this bias distribution should not involve in-pixel low-voltage transistors. In this context, if one or more devices exhibited field defects such as premature breakdown, they would prevent the remaining pixels in the array to be properly biased. The evaluation, in this thesis, of a number of large arrays samples over many designs demonstrated that the guard-ring structure is reliable against this effect.

An analytical study of the basic passive quenching circuit, based on a single MOS transistor, was introduced in this Chapter. In particular, the proposed operating mode, i.e. as a current source, leads to superior performance when compared to the typical passive quenching using a resistor or the triode region of a MOS transistor. The proposed single-MOS front-end circuit exhibits good performance in applications requiring very small pixels and/or high pixel fill factor. In order to improve the typical trade-off between dead time and afterpulsing, a hold-off time after SPAD quenching is important. However, as described in Section 2.4, existing circuits that implement hold-off time are excessively large to be implemented pixel-wise on an image sensor. A new SPAD front-end circuit, called dual-threshold passive quenching and active recharge circuit, was introduced to provide an adjustable hold-off time. This circuit only

requires 8 small transistors at the pixel level. Finally in the context of SPAD ancillary circuits, another challenge that is related to SPADs is the requirement of an external integrator, typically a counter, to store light information. This challenge arises from the fact that the SPAD capacitance is fully discharged on each photon detection. For applications requiring very small pixels with high fill factor, a special circuit, called event-driven readout circuit, was introduced. This circuit allows the designer to place ancillary circuits used to store light information outside the pixel array, thus improving pixel fill factor.

For the design of relatively complex circuits to fully exploit the benefits of SPADs, a technology migration towards state-of-the-art CMOS technology was compulsory. The $0.8\mu\text{m}$ CMOS technology is very effective in SPAD performance and is currently utilized in applications requiring a single detector. However, it has a number of limitations in terms of miniaturization and transistor speed, which are of high importance on a fully-integrated 3D image sensor. In the framework of this thesis, we have introduced the first successful implementations of SPADs in $0.35\mu\text{m}$ and 130nm CMOS technologies. In particular, the $0.35\mu\text{m}$ CMOS technology is a very good compromise between photodetector performance and circuit optimization. It has been therefore utilized in the fully integrated 3D image sensors described in Chapters 3 and 4, where array sizes as large as 16384 pixels are reported.

Device characterization of the SPADs introduced in this thesis as well as array performance of the former $0.8\mu\text{m}$ SPAD were reported in this Chapter. For convenience, a summary and comparison of device performance is shown in Table 2.1. As can be seen, SPADs designed in the $0.8\mu\text{m}$ CMOS technology feature better performance over more advanced CMOS technologies, except for PDP and power consumption. PDP was lower in the $0.8\mu\text{m}$ technology mostly due to non-optimized passivation layers. On the devices implemented in $0.35\mu\text{m}$, the passivation layer was prevented by design. Since the 130nm CMOS technology was an optimized process for imaging, special dielectrics also lead to good PDP results.

As expected, the breakdown voltage as well as the excess bias voltage monotonically decreases over technology generations. Given that the SPAD capacitance is discharged upon photon detections by approximately the excess bias voltage, power consumption is highly optimized in more advanced CMOS technologies. Note that, in the case of SPADs, power consumption depends on the square of V_E and linearly on parasitic capacitance.

Table 2.1. Performance summary for the devices investigated in this thesis.

Performance Parameter	$0.8\mu\text{m}$ CMOS SPAD	$0.35\mu\text{m}$ CMOS SPAD	$0.13\mu\text{m}$ CMOS SPAD
Breakdown voltage at room temperature (27°C)	25.5V	17.7V	9.7V
Typical excess bias voltage	5V	3.3V	1.7V
Excess bias to breakdown voltage ratio	19.6%	18.6%	17.5%
Typical active area	$38\mu\text{m}^2$	$38\mu\text{m}^2$	$78\mu\text{m}^2$
Maximum photon detection probability at typical V_E	26% at 550nm	35.4% at 460nm	34.5% at 450nm
Median (mean) dark count rate at room temperature (27°C)	18Hz (376Hz)	694Hz (2.4kHz)	N/A ^{a)} (30kHz)
Dark count rate density (median)	$0.5\text{Hz}/\mu\text{m}^2$	$18\text{Hz}/\mu\text{m}^2$	$385\text{Hz}/\mu\text{m}^2$ ^{a)}
Typical dead time	32ns	40ns	N/A ^{b)}
Afterpulse probability	Negligible	25%	N/A ^{b)}
Crosstalk probability	Negligible	Negligible	N/A ^{b)}
Timing resolution (FWHM)	82ps	80ps	119ps
Largest imager resolution	64x48 [157]	128x128 [160]	N/A ^{a)}
Smallest pixel pitch	45μm	25μm	N/A ^{a)}

a) Only a single device was implemented at the time of this dissertation.

b) Since the device was not monolithically integrated with its front-end circuit, full device characterization was not performed.

Among the SPAD implementations, DCR increased considerably from the $0.8\mu\text{m}$ CMOS towards 130nm. This noise figure might restrict the utilization of the 130nm SPAD in some applications. However, as described in the next Chapters, 3D image sensors need to reliably perform under relatively high background light. As a result, although a DCR as high as 30kHz is not desirable, it could be suitable in applications of interest in the framework of this thesis.

Afterpulsing probability also evolved unfavorably with technology scaling. While afterpulsing effects were absent in the $0.8\mu\text{m}$ SPAD, in $0.35\mu\text{m}$, they became an important issue to be addressed. Unfortunately, at the time of writing this dissertation, afterpulsing probability characterization was not available for the 130nm CMOS device described in this Chapter. Indeed, the device was not fully integrated with its front-end circuits. Thus, a characterization of afterpulsing probability under this condition is meaningless, as it gives no insight on the true potential of the device and of its internal capacitance when the SPAD is monolithically integrated with its quenching and recharge circuit.

A regrettable feature of the devices investigated in this thesis is the limited detection efficiency in the red and near infrared spectral range. A number of applications require good photon detection probability in those spectral ranges, including 3D image sensors based on TOF. The devices described in this Chapter featured a PDP of approximately 3% at 850nm due to the shallow p+/n-well active junction. As discussed and experimentally demonstrated in the next Chapters, the PDP achieved by the proposed devices is enough to improve the state-of-the-art in depth image sensing, even at 850nm. However, a margin for improvement clearly exists. In order to improve PDP performance, Xiao, Popovic, *et al.* recently introduced a new device structure that improves PDP, especially in the red spectral region [167]. The device was implemented using the same 0.35 μ m CMOS technology utilized in this thesis, with a diameter of 5 μ m. Nonetheless, instead of using the shallow p+ layer as anode, the authors were able to successfully implement a SPAD based on a p-tub/deep n-tub junction. Since the doping concentration of the p-tub layer is much lower than that of the p+, the structure also leads to an outstanding DCR performance in 0.35 μ m CMOS, lower than 50Hz at an excess bias voltage of 5V to the 50V of V_{BD} . Future investigations of fully integrated array implementations based on this device are expected.

Finally, the evaluation of fully integrated SPADs in 130nm CMOS with their front-end circuits is underway [168]. The level of miniaturization offered by the 130nm CMOS technology is expected to highly improve image resolution in SPAD-based sensors. However, afterpulsing performance is still an open question. Furthermore, it would be highly desirable to improve DCR performance to levels achievable in 0.35 μ m CMOS.

3. OPTICAL TIME-OF-FLIGHT 3D IMAGING BASED ON TCSPC

3.1 INTRODUCTION

In this chapter, we study the performance of solid-state TOF 3D image sensors based on time-correlated single-photon counting (TCSPC). TCSPC is a technique used in the measurement of time properties of low light level signals with picosecond resolution [56, 57]. This technique has been used for many decades in research, especially in particle physics and bioimaging. In optical TOF rangefinders, it was successfully demonstrated in [169], where a scanner based approach was adopted. The rangefinder was based on a highly optimized optical setup using two SPADs, as reference and measuring detectors, respectively. A pulsed laser source was used to generate a collimated spot on the target and a pan and tilt platform enabled the system to scan the target surface, thereby generating depth images. Outstanding timing resolution of SPADs and the 20-ps laser source allowed the authors to achieve distance precisions better than 30 μ m in a range of 1-25 meters, using the TCSPC technique. The laser average power was 0.25mW and its highest repetition rate was 25MHz. The laboratory-style optical setup and optical reference path also contributed in achieving high measurement stability and repeatability over long period of times. This example shows the potential of TCSPC in optical TOF rangefinders. However, it also shows some limitations that are of importance in this thesis. As far as acquisition speed is concerned, the main limitation of this approach is the mechanical scanner. It is highly desirable to acquire a high number of target points simultaneously in linear or 2D array implementations. This is a very important motivation for introducing large arrays of SPADs. Mechanical moving parts in scanners also lead to practical limitations and they require costly setups that are cumbersome and difficult to miniaturize.

This Chapter is organized as follows. In Section 3.2, the principle of solid-state optical TOF 3D image sensors based on TCSPC technique is described. Operating assumptions in TCSPC are also discussed in that Section. In Sections 3.3 and 3.4, a theoretical model for the evaluation of ranging performance is introduced. The model aims at validating design parameters in the context of reliable TOF

evaluations, thus helping in the design of device, circuit, and system components. Since the proposed evaluation methodology is based on photon counting rates, a model that relates environmental, device, circuit, and system parameters to actual photon counting rates is described in Section 3.4. In Section 3.5, the design of the first fully-integrated TCSPC image sensor in CMOS is described. The experimental evaluation of this sensor in a short range scenario, based on extremely low illumination power budget is described in Section 3.6. In Section 3.7, the experimental evaluation of a CMOS SPAD imager as a TCSPC rangefinder in a distance range of up to 70 meters is reported. In this distance range, a number of design, operation and safety concerns with respect to the illumination system exist. These issues are discussed and addressed in the same Section. Finally, Section 3.8 concludes this Chapter.

3.2 TCSPC PRINCIPLE AND ASSUMPTIONS

In Figure 3.1, a schematic setup of a scannerless 3D image sensor based on TCSPC is shown. A pulsed light source, typically a laser, at a repetition rate f_0 is used to illuminate the scene in front of the image sensor with a given field-of-view (FOV). A lens or diffuser is used in front of the light source to achieve the required FOV. A single-photon image sensor is placed on the focal plane of an imaging lens so as to match the FOV, yet providing enough depth-of-field to cover all the range of interest. Since image formation is obtained on the detector plane, the light reflected by two different points on the target's surface are imaged onto independent points on the image sensor. As a result, independent and ideally simultaneous TOF distance measurements may be performed for all the points of target surface without the need for any mechanical moving part.

For every light pulse sent, the light source also provides an electrical reference signaling the moment the pulse was emitted. This trigger signal may be used as a START signal for all the pixels in the image sensor. SPADs at pixel level generate independent STOP signals upon photon detections, which are used in the TOF measurement. For each pixel, a high-speed timer is used to measure time interval between START and STOP and the resulting TOF is recorded in an individual histogram. The timer typically consists of a time-to-digital converter (TDC). Note that a time-to-amplitude (TAC) followed by an analog-to-digital converter (ADC) may be considered as well. These circuits may be implemented on a CMOS integrated circuit, ideally monolithically with the SPAD array. They are commonly referred to as TDCs in this Chapter.

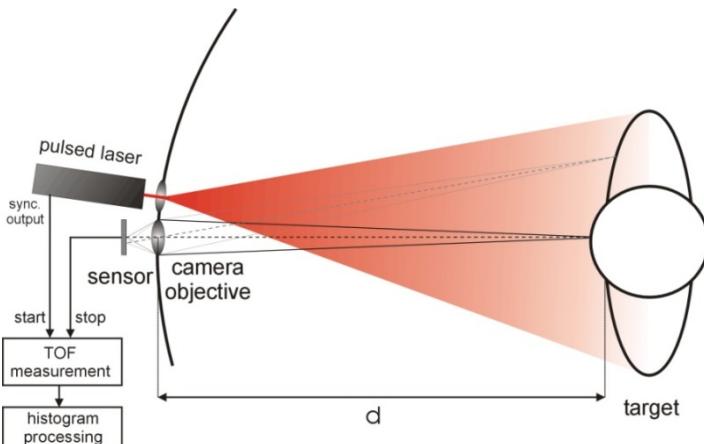


Figure 3.1. Principle of solid-state optical TOF 3D image sensor based on TCSPC.

Instead of measuring TOF based on a single back-reflected photon, in TCSPC, the detection principle typically involves a large number of detections. Figure 3.2 illustrates the basic TCSPC principle and how a histogram is built for each pixel.

TCSPC takes advantage of the fact that for low-intensity and high-repetition rate signals, the probability of detecting one photon in one signal period is far less than one. The detector signal is a train of randomly distributed pulses corresponding to the detection of individual photons. In fact, there are many signal periods without photons. In a number of periods, there is one photon detection only. Periods with more than one photon are very rare, though not impossible. When a photon is detected, the arrival time of the photon with respect to the beginning of the period (START) is measured. These events are collected in a memory by incrementing the word value in a memory location with an address proportional to the detection time. After many period cycles, the distribution of the detection times, i.e. the waveform of the incident optical pulse, is stored in the memory. The histogram contents accurately reproduce the illumination pulse shape, convolved with the instrument response function (IRF) of the measuring circuit comprising the SPAD and the TDC. Even though it is nowadays possible to design TDC and histogram memory with resolutions of 1-5ps [170], the effective TCSPC resolution is limited by the IRF. Mathematically, the IRF is given by the convolution between the single-photon detector timing jitter and the timing jitter of the TDC. Note that TCSPC does not use any time gating. Assuming an ideal photodetector, all the incident photons

are recorded and they contribute consequently to the measurement. Moreover, given that period cycles with more than one signal photon are rare, the detection efficiency is still very close to ideal even assuming a photodetector with dead-time, such as SPADs.

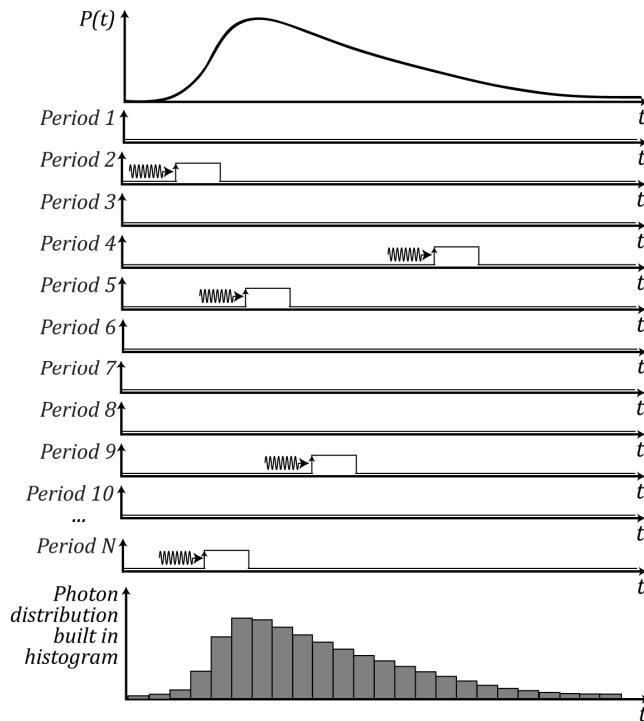


Figure 3.2. TCSPC measurement waveform.

Since the START signal resets the TOF measurement at the beginning of each illumination cycle, the longest measured *TOF* is the illumination period T_0 , given by

$$T_0 = \frac{1}{f_0}. \quad (3.1)$$

TOF is then computed from the histogram data by determining the position of the reproduced peak within the histogram. Finally, the distance z of a target point, corresponding to a given pixel, is computed as

$$z = \frac{c \cdot TOF}{2}, \quad (3.2)$$

where c is the speed of light.

A number of algorithms exist to accurately find the position of the signal peak in the histogram. An updated review on advanced and highly accurate algorithms are proposed in [171]. Advanced algorithms are practical in setups involving a low number of parallel single-photon detectors due to high requirement in computing power and complexity. Most of the algorithms described in [171] are difficult to implement in hardware. In this thesis, we felt that it was important to rely on simple algorithms that may be fully implemented in hardware and massively replicated on a single integrated circuit. For instance, a fast and hardware-inexpensive algorithm may be based on the timing position of the peak's centroid, as discussed in Section 3.3.

The main advantage of TCSPC over other techniques is its robustness against parasitic background light. Parasitic background light (BG) is due to other sources of light that may simultaneously illuminate the target in uncontrolled environments. Examples of BG are daylight solar irradiance in outdoor applications and artificial light illumination in indoor applications. In practical setups, bandpass optical filters are used to block a large fraction of BG when its spectral contents are out of the spectral band of the signal light source. For instance, near infrared laser diodes emitting at 850nm-wavelength may be used as signal light source. In this case, a bandpass filter centered at 850nm with a bandwidth of 10-30nm could be utilized to block BG outside this spectral range. Due to practical limitations, it is very difficult to further reduce the filter's bandwidth so as to completely block BG, thus allowing 1% to 5% of BG power to reach the sensor. Transmission bandwidth of the optical filter needs to be large enough to accommodate wavelength shifts of the light source due to temperature variations. In outdoor applications, solar irradiance may reach 1000 Watt per square meter in a clear sky condition at noon. The ranging sensor should be therefore robust enough to accurately determine TOF under a signal to background ratio (SBR) much lower than unity, even when assuming bandpass optical filtering. In this context, TCSPC offers an important advantage that may be illustrated by a measurement example. In Figure 3.3, a histogram measured for a single pixel using an actual TCSPC setup is shown. The measurement setup was based on a pulsed laser source with a repetition rate of 40MHz and pulse duration of 80ps FWHM. A $0.8\mu\text{m}$ CMOS SPAD sensor [156] was used to detected back-reflected photons under a strong artificial BG illumination. TOF measurements were performed by using an external TDC with 100ps resolution. The combination of SPAD and TDC circuitries showed

an IRF of 100ps FWHM. When turning off the BG light source, a signal count rate (SG) of 100kHz was measured. The BG source was then turned on together with the laser source and adjusted to provide a total count rate of 10.1MHz, thus leading to a SBR of -40dB (1/100).

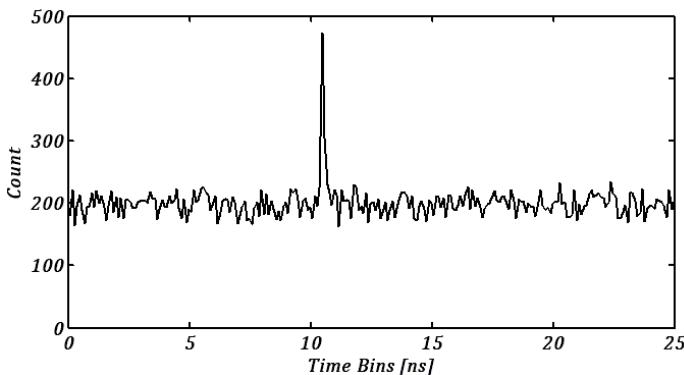


Figure 3.3. Example of measured histogram using a TCSPC setup. The measurement conditions were: signal count rate 100x lower than background count rate (SBR = -40dB), light pulse duration of 80ps (FWHM) repeated at 40MHz, IRF and histogram resolution of 100ps. The histogram contains 50k measurements.

As can be seen in the plot of Figure 3.3, since photons detected from BG arrive randomly with respect to the laser source pulses, their arrival time measurements appear uniformly distributed in the histogram. All the photons detected from the SG source, on the other hand, lead to TOF measurements around a constant value, of approximately 11ns. As a result, they form a signal peak in the histogram that may be effectively discriminated from BG. Note that, since the illumination repetition rate was 40MHz, the histogram range was limited to 25ns. Even though BG appears uniformly distributed, it exhibits a random component that is due to photon shot noise in BG. SG also suffers from shot noise but its effect is less noticeable in the example. Successful TOF detections occur when SG may be extracted from BG under the effect of those noise sources. In the example, the SG peak is doubtless detectable given the large margin between its maximum value and the mean BG level. Furthermore, TOF may be computed by using only a very small number of time bins around the peak position, e.g. by determining the timing position of the peak's centroid. As a result, the noise contribution of BG that appears in all the remaining time bins does not impact the value of TOF. In this case, it may be concluded then

that TCSPC leads to a nearly full suppression of BG noise contribution in the measurement of TOF, when the signal peak is reliably discriminated from BG.

By visual inspection of Figure 3.3, it is easy to conclude that the area under the background level, at approximately 200 counts, is proportional to the BG photon counting rate whereas the area between the peak and the BG level is proportional to the SG photon counting rate. Note that the integral of the histogram over all the time bins leads to the total number of photon detections, which were 50'000 in the example. It is easy to predict that it is possible to further increase the BG contribution in the example and still benefit from an acceptable margin between SG and BG. Details on how much BG may be increased are discussed in the model of Sections 3.3 and 3.4.

TCSPC does not only lead to advantages. The benefits briefly exemplified in last paragraphs are obtainable at the price of high circuit complexity. On a TCSPC-based TOF image sensor, each pixel must be connected to a TDC and it must count on an individual memory circuit to store a full histogram. TDCs may be shared among pixels and they are likely to be highly miniaturized in deep-submicron CMOS technologies. Memory circuits on the other hand may represent a challenge for a high-resolution imager. For instance, the histogram of Figure 3.3 requires approximately 2kbits of memory, which should be multiplied by the number of pixels in the image sensor. Another important consideration is the location of those circuits. If they are designed at the pixel level, it is very difficult to achieve high active area fill factor even assuming microlenses. On the other hand, if they are placed outside the pixel array, timing preserving, area inexpensive, and very fast readout circuitries must be designed. When it comes to design a linear or small array of pixels, those difficulties may be effectively addressed with no penalty in active area fill factor.

3.3 MODELING OF PERFORMANCE BASED ON TCSPC

References [172, 173] present two mathematical models for performance evaluation that can be applied to rangefinders based on SPADs. In both models, it was assumed that SPADs operate in “gated” mode. Gating is referred to the action of opening a pre-defined time window during which the SPAD may detect a photon. For instance, if the rangefinder can estimate by any other mean that the target is approximately at a given distance, it can open a detection window around the estimated time to perform an accurate measurement. As a result, the background light, that could potentially make the SPAD trigger at any time before the time of interest, is preempted by the gating action. This technique is

particularly useful for single-shot measurements to reduce the probability that the SPAD is in dead time during the time of interest due to a background photon. Single-shot measurements are particular attractive when the range of detection is very large, reducing the time available for averaging.

In [172], a probabilistic model was studied to assess the probability of correct detection and the probability of false alarms. The model is suitable when the number of averaged illumination pulses is limited to 1-10. However, TCSPC relies on a relatively large number of measurements for histogram processing, which leads to excellent noise rejection. The model described in [172] is therefore not used in this thesis. Reference [173] is also based on gated operation of single-photon detectors. The special case considered by the authors is characterized by means of an artificial photon detection probability exponentially decreasing with respect to *TOF*. This undesirable behavior only occurs when gating is assumed. In this thesis, due to the short distance ranges of interest, a continuous free-running mode is more appropriate. In free-running mode, the single-photon detector may be triggered at any time. Upon photon detection, the usual dead time applies. Assuming a dead time lower than the illumination repetition period, the TCSPC principle does not exhibit the artificial behavior experienced in [173]. As a result, the model proposed in [173] is not used in this Section for general performance evaluation, except for the estimation of distortion.

As briefly mentioned, a very important assumption in TCSPC is that, on average, much less than a photon should be detected per illumination cycle. When this condition is not met, the signal experiences distortion. This is commonly known as “pile-up” effect in TCSPC [56, 57]. Many authors reported correction algorithms for pile-up distortions, e.g. in [174]. However, these algorithms may require excessive processing power on a 3D image sensor operating in real time. A simpler solution to deal with pile-up is to prevent it by reducing the detector count rate. For instance, the detector count rate may be reduced by means of an attenuator or filter. This practical solution is well known by a number of researchers who build TCSPC optical setups. In 3D imaging, it was considered in [172, 173]. Pile-up distortion may be analytically evaluated by assuming for instance a Gaussian shape for the illumination optical pulse [173], i.e.

$$P_s(t) = \frac{Q_p}{\sqrt{2\pi}\sigma_\tau} \exp\left\{-\frac{(t - TOF)^2}{2\sigma_\tau^2}\right\}, \quad (3.3)$$

where Q_p is the optical energy per illumination pulse falling on the detector, σ_τ is the Gaussian half-pulse width, and $P(t)$ is the optical power within one illumination cycle. Assuming Poisson statistics in the detection process, the distorted optical power $\tilde{P}_S(t)$ becomes

$$\tilde{P}_S(t) \cong P_S(t) \exp \left\{ -\frac{M}{2} \left[1 + \operatorname{erf} \left(\frac{t - TOF}{\sqrt{2}\sigma_\tau} \right) \right] \right\}, \quad (3.4)$$

where erf is the *Gauss error function* and M is the mean number of signal photon detections per illumination cycle, i.e.

$$M = \frac{\lambda_{SG} \bar{P}_S}{hc} \eta_{SPAD}. \quad (3.5)$$

In (3.5), λ_{SG} is the illumination wavelength, η_{SPAD} is the SPAD photon detection probability, \bar{P}_S is mean value of the signal optical power falling on the SPAD active area, and h is Plank's constant.

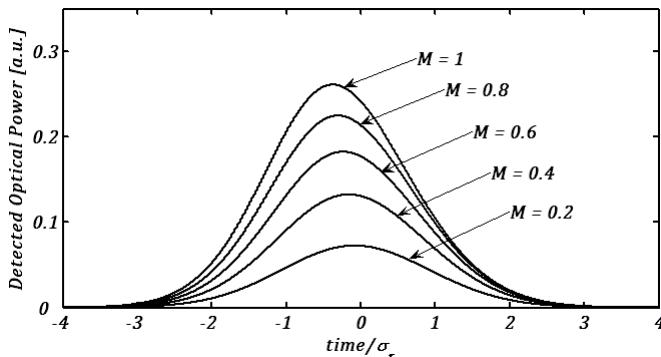


Figure 3.4. Simulated detected optical power as a function of time, normalized to σ_τ using M as a parameter and assuming $TOF = 0$.

Equation (3.4) suggests that the measured optical pulse using TCSPC suffers from a non-symmetrical distortion. In Figure 3.4, the shape of optical pulses as they would appear in the resulting histogram is plotted for a few values of M lower than unity. In the graph, the horizontal axis was normalized to σ_τ and a TOF around origin was assumed to see the timing effects of TCSPC sampling. As can be seen, sampled optical pulses exhibit shape distortion that affects their timing properties.

This effect is basically due to the dead time of single-photon detectors and it is noticeable when, on average, approximately one or more signal photons are detected per cycle. The dead time of single-photon detectors, such as SPADs, is typically longer than the optical pulse width. In these cases, in a given illumination cycle, when a photon from the leading portion of the optical pulse is detected, the detector is no longer able to detect a photon from the trailing section of the same pulse. The probability of detecting photons in the trailing section of optical pulses depends therefore on the probability that no photon has been detected from the leading section. When the optical power is increased, more photons that belong to the leading portion are detected and fewer photons are detected from the trailing part, thus leading to pulses that appear asymmetric with respect to time. When determining TOF , the effects of pile-up distortion produce systematic timing errors. Figure 3.5 shows the errors in determining TOF , measured as the mean value of distorted optical pulses, as a function of M . TOF errors (ΔTOF) are also normalized to σ_τ . An upper bound of 0.1 may be defined for M as ΔTOF may be considered to be negligible under this condition, i.e. $|\Delta TOF| < 0.028\sigma_\tau$ for any $M \leq 0.1$. For higher values of M , it might be appropriate to correct TOF by post-processing.

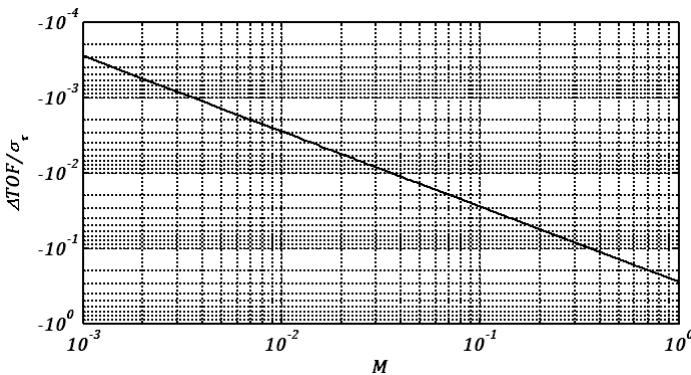


Figure 3.5. Theoretical mean error on TOF , normalized to σ_τ , as a function of M .

A TCSPC-based system can be optimized by increasing the illumination repetition rate f_0 to the maximum allowed. Assuming the average optical power \bar{P}_S of the illumination is kept constant in (3.5) when f_0 is increased, then the detector count rate remains constant and M therefore decreases. However, f_0 also defines the rangefinder distance range R_D as

$$R_D \leq \frac{c}{2f_0}. \quad (3.6)$$

An upper bound for f_0 is consequently defined as a function of R_D .

Moreover, since M depends strongly on the target distance and reflectivity, it is difficult to choose a fixed attenuation factor that covers the full range of a 3D image sensor. Practical implementations of this method will make designers opt for a tunable attenuator. If the single-photon detector is implemented as a SPAD, as in this thesis, one may use the reverse bias of the device to vary its PDP as a function of the actual counting rate in a closed-loop circuit. In this Section, distortion effects in TCSPC due to pile-up will be addressed by the introduction of such an attenuation factor.

Assuming a Gaussian illumination pulse according to (3.3), an average photon detection rate B_R due to BG and DCR, an average photon detection rate from active signal S_R , then we may express SBR as

$$SBR = \frac{S_R}{B_R}. \quad (3.7)$$

Hence, the total average photon detection rate T_R of a SPAD is therefore $S_R + N_R$.

3.3.1 ANALYTICAL MODEL

The design of a TCSPC system may be highly facilitated when an analytical model that relates sensor performance to design parameters is utilized. Examples of design parameters are illumination power, illumination pulse duration, SBR , and TDC resolution, to name a few. In this Section, a simple and fairly accurate analytical model is presented to make the decision on design parameters easier. In addition, robustness of the distance measurement is also taken as a design parameter.

When a TCSPC measurement is carried out based on the setup of Figure 3.1, histogram data are generated for every pixel in the image sensor. A typical example of histogram is shown in Figure 3.3. Accurate measurements of TOF are achieved when the signal peak is easily separated from the BG level. Depending on individual SG and BG conditions of each pixel, the BG level, peak height, and peak position might vary considerably. Nonetheless, the basic shape of histograms related to different pixels in array keeps unchanged. It is therefore possible to relate the shape of a generic TOF histogram, under some

assumptions, to design parameters. When examining a typical histogram $h^*(\tau)$, one may distinguish two components as

$$h^*(\tau) = s^*(\tau) + b^*(\tau), \quad (3.8)$$

where $s^*(\tau)$ and $b^*(\tau)$ are random variables describing the signal peak and the uniformly distributed BG component, respectively. τ represents bin locations in the histogram and it is therefore discrete, i.e.

$$\tau \in \left[0, \Delta t, 2\Delta t, 3\Delta t, \dots, \left(\left\lceil \frac{T_0}{\Delta t} \right\rceil - 1 \right) \Delta t \right], \quad (3.9)$$

where Δt is the histogram resolution. In many cases, Δt is chosen to match the TDC resolution. In order to simplify the notations hereafter, $h^*(\tau)$, $s^*(\tau)$ and $b^*(\tau)$ are described by their mean values $h(\tau)$, $n(\tau)$, and $s(\tau)$, respectively.

The IRF response of the TCSPC detection front-end may be assumed to be Gaussian, i.e.

$$IRF(\tau) = \frac{1}{\sqrt{2\pi}\sigma_{IRF}} \exp \left\{ -\frac{1}{2} \left(\frac{\tau}{\sigma_{IRF}} \right)^2 \right\}, \quad (3.10)$$

with

$$\sigma_{IRF} = \sqrt{\sigma_{SPAD}^2 + \sigma_{TDC}^2}. \quad (3.11)$$

Although this assumption may seem to be inaccurate at first glance for those who are familiar with SPAD timing resolution (see Chapter 2), it is in fact a good approximation in TOF measurements under low *SBR* conditions. The typical diffusion tail present in SPAD timing responses is very often submerged under the background level of the TCSPC histogram. Note that σ_{SPAD} is taken as a standard deviation of the SPAD timing resolution, i.e., it is approximately 0.425 times its FWHM. Concerning the σ_{TDC} component, it comes from the timing jitter in the TDC and in any intermediary circuitry. σ_{TDC} is typically a normal random variable, which further justifies Equation (3.11).

Assuming that T is the total histogram acquisition (or integration) time, it follows that $b(\tau)$, the contribution from BG and DN in the TCSPC measurement, is a constant given by

$$b(\tau) = \bar{b} = \frac{B_R T \Delta t}{T_0}. \quad (3.12)$$

When Δt is smaller than σ_τ and σ_{IRF} , $s(\tau)$ may be accurately expressed as

$$s'(\tau) \cong \frac{S_R T \Delta t}{\sqrt{2\pi}\sigma_S} \exp\left\{-\frac{(\tau - TOF)^2}{2\sigma_S^2}\right\}, \quad (3.13)$$

where $\sigma_S = \sqrt{\sigma_\tau^2 + \sigma_{IRF}^2}$ is the resulting signal half-pulse width. This expression originates from the convolution between the density of photon detection rate and IRF, sampled at intervals of Δt . In the general case, when Δt is larger than σ_S , another convolution with a time window function (or rect function) of width Δt is necessary, i.e.

$$s(\tau) = \frac{S_R T}{2} \left[\operatorname{erf}\left(\frac{\tau - TOF + \frac{\Delta t}{2}}{\sqrt{2}\sigma_S}\right) - \operatorname{erf}\left(\frac{\tau - TOF - \frac{\Delta t}{2}}{\sqrt{2}\sigma_S}\right) \right]. \quad (3.14)$$

Figure 3.6 shows the mean value of histogram $h(\tau)$ and a snapshot of $h^*(\tau)$ for a given set of parameters. $h^*(\tau)$ is shown to illustrate the effect of its randomness due to $s^*(\tau)$ and $b^*(\tau)$.

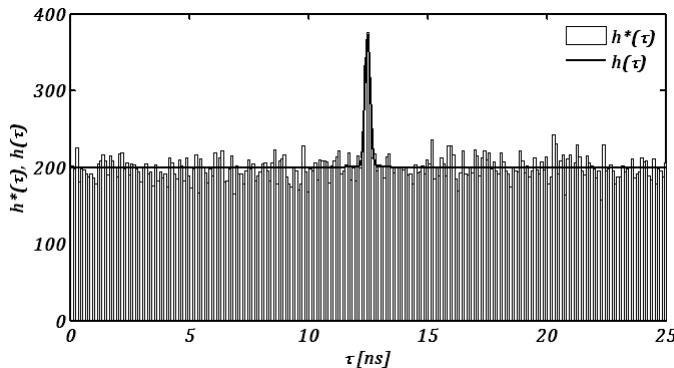


Figure 3.6. Example of $h^*(\tau)$ and $h(\tau)$ for a given set of parameters: $T_0 = 25\text{ns}$, $TOF = T_0/2$, $\sigma_\tau = 100\text{ps}$, $\sigma_{IRF} = 50\text{ps}$, $\Delta t = 100\text{ps}$, $S_R = 10\text{kHz}$, $S_B = 1\text{MHz}$, and $T = 50\text{ms}$.

Among different algorithms, in this model, an algorithm that uses the timing position of the signal's peak as a reference to perform noise rejection is

proposed. This approach has a number of advantages. It is very simple and may require a very small silicon area when implemented on an integrated circuit. As a result, it is suitable to be implemented pixel-wise on an image sensor. Moreover, the values taken by $h(\tau)$ will vary over a very large range due to measurement conditions such as SBR , lens aperture, target's reflectivity, etc. It would be therefore difficult, for instance, to use a fixed threshold value for $h(\tau)$, i.e. $h_{TH} \in (\bar{b}, s(\tau)]$, that would separate the signal peak from the background. Conversely, the timing position of the signal's peak in the histogram, for a fixed target, does not vary as a function of environment conditions. As a result, although the proposed algorithm is very simple, it exhibits an important level of robustness against measurement conditions. Let the timing position of the signal's peak be TOF' and its half width be $N_{HW}\Delta t$, where N_{HW} is an integer, then one could accurately determine TOF very simply as following

$$TOF' = j\Delta t \quad (3.15)$$

such that

$$h(j\Delta t) = \max\{h(k\Delta t)\}, \forall k \in [0, 1, 2, \dots, \left(\left\lceil \frac{T_0}{\Delta t} \right\rceil - 1\right)], \quad (3.16)$$

then we use the time interval

$$\tau \in [TOF' - N_{HW}\Delta t, TOF' + N_{HW}\Delta t] \quad (3.17)$$

to determine a mean value of τ . The mean value of τ may be estimated as the abscissa of the centroid of $h(\tau)$, i.e.

$$TOF = \bar{\tau} = \frac{\sum_{i=\frac{TOF'}{\Delta t}-N_{HW}}^{\frac{TOF'}{\Delta t}+N_{HW}} h(i\Delta t) \cdot i\Delta t}{\sum_{i=\frac{TOF'}{\Delta t}-N_{HW}}^{\frac{TOF'}{\Delta t}+N_{HW}} h(i\Delta t)}, i \in \mathbb{Z}. \quad (3.18)$$

Note that $N_{HW}\Delta t$ is known *a priori* and it does not depend, to a first approximation, on measurement conditions. Indeed, $N_{HW}\Delta t$ is the half pulse width of the resulting convolution expressed by $s(\tau)$ in Equation (3.14). N_{HW} is therefore chosen at design time to accommodate the worst-case spreading of the signal pulse. In addition, TOF' can be effortlessly computed during histogram generation using a simple hardware register when fully digital

implementation is adopted, thus reducing the algorithm processing only to interval given by (3.17).

Based on the proposed algorithm, we can determine the conditions under which one could, with a desired certainty, accurately find TOF . The computation of TOF' is a crucial step to retrieve the actual TOF value. As one may expect, the probability of an incorrect estimation of TOF increases when SBR is exceedingly low. Let us assume that one requires a reliable measurement of TOF with a given confidence level stated by a probability P . For instance, let us assume the user desires to rely on 99.7% of the TOF measurements, i.e. $P=0.997$. Then, we can predict that the signal peak is successfully discriminated from noise, by successfully obtaining TOF' , when

$$h(TOF') - \alpha_h \sigma_h > b(TOF') + \alpha_b \sigma_b, \quad (3.19)$$

where σ_h and σ_b are the standard deviations of the quantities $h^*(TOF)$ and $b^*(TOF)$ respectively. α_h and α_b are parameters used to set the confidence level of the measurement given by

$$\alpha_h = \sqrt{2} \cdot \text{erf}^{-1}(P), \text{ and} \quad (3.20)$$

$$\alpha_b = \sqrt{2} \cdot \text{erf}^{-1} \left(1 - \frac{1-P}{\frac{T_0}{\Delta t} - 1} \right) \cong \sqrt{2} \cdot \text{erf}^{-1} \left(1 - \frac{\Delta t}{T_0} (1-P) \right). \quad (3.21)$$

where erf^{-1} is the *inverse error function*. Although $h^*(TOF)$ and $b^*(TOF)$ are Poisson-distributed, in Equations (3.20)-(3.21) they are approximated by normal distributions with standard deviations σ_h and σ_b as

$$\sigma_h = \sqrt{h(TOF')}, \text{ and} \quad (3.22)$$

$$\sigma_b = \sqrt{b}. \quad (3.23)$$

Note that a Poisson random variable with large mean value may be well approximated by a normal random variable with the same mean and variance. In the context of this model, we are interested in analyzing system performance under strong background illumination so as to size, for instance, the illumination power budget. This implies that the mean value of $b(\tau)$ is typically larger than a few tens, which results in a good approximation. Once the system

is designed, an operation under lower background illumination makes the condition of (3.19) anyway easier to be met. Approximation errors due to small $b(\tau)$ mean value, under this condition, cannot lead to false detections.

Based on Equations (3.7), (3.8), (3.12), (3.14), (3.20)-(3.23), and on relation (3.19), one may determine the minimum signal count rate S_R necessary to perform a correct detection with confidence level given by P via the expression

$$S_{R,MIN} = \frac{2}{T} \left[\frac{\operatorname{erf}^{-1}(P) \sqrt{\operatorname{erf}\left(\frac{\Delta t}{\sqrt{8} \cdot \sigma_S}\right)} + \frac{\Delta t}{SBR \cdot T_0} + \operatorname{erf}^{-1}\left(1 - \frac{\Delta t}{T_0}(1-P)\right) \sqrt{\frac{\Delta t}{SBR \cdot T_0}}}{\operatorname{erf}\left(\frac{\Delta t}{\sqrt{8} \cdot \sigma_S}\right)} \right]^2. \quad (3.24)$$

When Δt is assumed to be smaller than σ_S , one may determine $S_{R,MIN}$ based on Equation (3.13), which results in a slightly simpler expression [175],

$$S_{R,MIN}' = \frac{4\pi \cdot \sigma_S^2}{T \cdot \Delta t} \left[\operatorname{erf}^{-1}(P) \sqrt{\frac{1}{\sqrt{2\pi}\sigma_S} + \frac{1}{SBR \cdot T_0}} + \frac{\operatorname{erf}^{-1}\left(1 - \frac{\Delta t}{T_0}(1-P)\right)}{\sqrt{SBR \cdot T_0}} \right]^2, \quad (3.25)$$

but inaccurate for $\Delta t \geq \sigma_S$.

$S_{R,MIN}$ may be used to determine the minimum illumination power necessary to ensure a reliable measurement of TOF' . Note that once TOF' is measured, the actual TOF is computed based on Equations (3.15)-(3.18). If the condition established by $S_{R,MIN}$ is not met, there is no guarantee that a reliable measurement of TOF is achieved. On the contrary, when $S_{R,MIN}$ is not satisfied, there is a significant probability that any BG bin location may be wrongly mistaken for the true TOF' location, thus leading to unpredictable TOF results. One may argue that it is possible that different algorithms, e.g. in [171], would allow TOF to be confidently determined even when the condition on $S_{R,MIN}$ is not entirely satisfied. Nonetheless, this model was appropriately chosen to rely on an algorithm that is fast, accurate and inexpensive.

Figure 3.7 shows a first example of $S_{R,MIN}$ from (3.24) as compared to $S'_{R,MIN}$ from (3.25), for a given set of parameters. The minimum signal count rate required was plotted as a function of Δt , normalized to σ_S . As can be firstly observed, $S'_{R,MIN}$ monotonically decreases for any $\Delta t/\sigma_S$. $S_{R,MIN}$ follows the same monotonic behavior for $\Delta t/\sigma_S$ lower than unity but eventually reaches a minimum value at $\Delta t/\sigma_S$ of approximately 3. Since $S'_{R,MIN}$ diverges highly from $S_{R,MIN}$ for $\Delta t/\sigma_S$ larger than unity, it is not helpful for a complete analytical study. The existence of a

minimum value of $S_{R,MIN}$ also implies that a trade-off exist with respect to Δt . This trade-off may be optimally exploited so as to reduce the illumination power. When Δt is very small, the mean value of each histogram bin $h(\tau)$ for a given set of parameters is also small. Since every bin value $h(\tau)$ is described by a Poisson random variable, a lower mean value leads to comparatively high fluctuations. When Δt is increased, $h(\tau)$ initially increases proportionally to Δt and its fluctuation reduces if compared to its mean value. Eventually, Δt becomes larger than the full width of the signal peak, thus reducing the contrast of $s(\tau)$ with respect to $b(\tau)$. Another important remark on Figure 3.7 is that reliable signal peak detection occurs at extremely low signal count rates, down to a few kilohertz. Note that since a SBR of -40dB was assumed, the corresponding background count rate is 100 times higher.

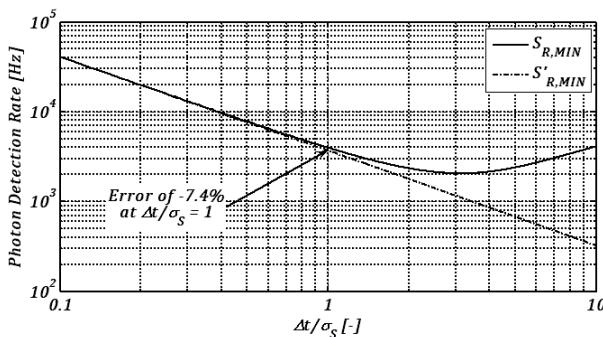


Figure 3.7. Comparison of $S_{R,MIN}$ and $S'_{R,MIN}$ as a function of Δt for a given set of parameters: $T_0 = 25\text{ns}$, $\sigma_s = 100\text{ps}$, $SBR = 1/100$ (-40dB), $P = 0.997$, and $T = 50\text{ms}$. Δt was normalized to σ_s so as to show how $S'_{R,MIN}$ diverges from $S_{R,MIN}$ for $\Delta t \geq \sigma_s$.

In Figure 3.8, $S_{R,MIN}$ is plotted again as a function of Δt using SBR as a parameter. The plot covers four values for SBR : 0dB, -20dB, -40dB, and -60dB. In the plot, the point indicating the minimum value reached by each $S_{R,MIN}$ curve is indicated. As can be seen, the optimum Δt varies with SBR . When SBR is extremely low, e.g. -60dB, a Δt of approximately $3\sigma_s$ is more appropriate. It is reasonable to decide on Δt assuming the worst case SBR given that it is much easier to meet the $S_{R,MIN}$ constraint for higher SBR . In the case of SBR of -60dB, a signal photon count rate of approximately 16.6 kHz is necessary. This figure is an interesting result, given that only 1 out of 1000 photons detected originates from the illumination system. Assuming that an optical bandpass filter placed in front of the sensor provides an additional 40dB of background rejection, it is

possible to afford 100'000 incident background photons on the target surface for each incident signal photon. In other words, assuming that a target surface is illuminated with 1000 W/m^2 due to solar irradiation, it is only necessary to spend 1 mW/m^2 of signal power. Actual S_R is still somewhat abstract up to this point as it depends on a number of sensor and environment parameters such as incident background irradiance, bandpass filter specifications, and lens f/#. In Section 3.4, a model that relates S_R to remaining (non TCSPC) parameters is presented.

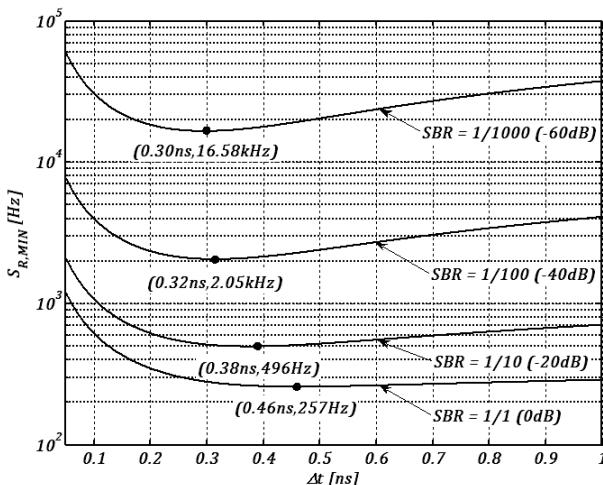


Figure 3.8. $S_{R,MIN}$ as a function of Δt using SBR as a parameter. The remaining parameters are: $T_0 = 25\text{ns}$, $\sigma_s = 100\text{ps}$, $P = 0.997$, and $T = 50\text{ms}$.

Under the condition of $S_{R,MIN}$, Equation (3.24), one may determine the standard deviation of TOF , i.e. σ_{TOF} . In particular, it is advantageous to use the benefits of averaging to improve precision. TOF is defined here as the centroid position of the light pulse within the interval given by (3.17). When performing each TOF measurement using (3.18), the fluctuations of $b^*(\tau)$ and $s^*(\tau)$ lead to an uncertainty on TOF . Let us assume that there exist two contributions to such uncertainty. Let us also assume that these contributions are two statistically independent random variables with zero mean. Then, the variance of TOF may be modeled as a linear combination of $\sigma_{TOF,B}^2$ and $\sigma_{TOF,S}^2$, i.e.

$$\sigma_{TOF}^2 = \frac{\sigma_{TOF,B}^2 \left[\sum_{i=\frac{TOF'}{\Delta t}-N_{HW}}^{\frac{TOF'}{\Delta t}+N_{HW}} b(i\Delta t) \right]^2 + \sigma_{TOF,S}^2 \left[\sum_{i=\frac{TOF'}{\Delta t}-N_{HW}}^{\frac{TOF'}{\Delta t}+N_{HW}} s(i\Delta t) \right]^2}{\left[\sum_{i=\frac{TOF'}{\Delta t}-N_{HW}}^{\frac{TOF'}{\Delta t}+N_{HW}} h(i\Delta t) \right]^2}, \quad (3.26)$$

the variances due to the fluctuations of $b^*(\tau)$ and $s^*(\tau)$, respectively.

Considering that every point of $b^*(\tau)$ within interval (3.17) is also a statistically independent random variable whose mean value is \bar{b} and whose standard deviation is σ_b , then, by applying the formula of error propagation on Equation (3.18) without the $s(\tau)$ component, $\sigma_{TOF,N}$ is determined as

$$\sigma_{TOF,B}^2 = \frac{(2N_{HW} + 1)^2 - 1}{2N_{HW} + 1} \cdot \frac{\Delta t^2}{12\bar{b}} = \frac{(2N_{HW} + 1)^2 - 1}{2N_{HW} + 1} \cdot \frac{SBR \cdot \Delta t \cdot T_0}{12 \cdot T \cdot S_R}. \quad (3.27)$$

The term $\sigma_{TOF,S}$ is determined by inspection, assuming a reduction of the single-shot error σ_S by square root of the number of signal measurements stored in $s(\tau)$,

$$\sigma_{TOF,S}^2 = \frac{\sigma_S^2}{\sum_{i=\frac{TOF'}{\Delta t}-N_{HW}}^{\frac{TOF'}{\Delta t}+N_{HW}} s(i\Delta t)} = \frac{\sigma_S^2}{T \cdot S_R \cdot \operatorname{erf}\left\{\frac{(2N_{HW} + 1)\Delta t}{\sqrt{8} \cdot \sigma_S}\right\}}. \quad (3.28)$$

Finally, σ_{TOF} is determined by combining (3.27) and (3.28) into (3.26) as

$$\sigma_{TOF} = \frac{1}{\sqrt{T \cdot S_R}} \sqrt{\frac{\frac{[(2N_{HW} + 1)^2 - 1](2N_{HW} + 1)\Delta t^3}{12 \cdot SBR \cdot T_0} + \operatorname{erf}\left\{\frac{(2N_{HW} + 1)\Delta t}{\sqrt{8} \cdot \sigma_S}\right\}\sigma_S^2}{\frac{(2N_{HW} + 1)\Delta t}{SBR \cdot T_0} + \operatorname{erf}\left\{\frac{(2N_{HW} + 1)\Delta t}{\sqrt{8} \cdot \sigma_S}\right\}}}. \quad (3.29)$$

Note that N_{HW} may be any positive integer number. Moreover, σ_{TOF} does account for any quantization error due to Δt .

In order to evaluate σ_{TOF} as a function of the distance, we introduce an attenuation factor in the model so as to reduce the pile-up effect, i.e.

$$S_R + B_R \leq f_0. \quad (3.30)$$

The attenuation is applied simultaneously on the signal and on background. As a result, the pile-up effect is greatly attenuated giving that M is limited to

$$M \leq \frac{1}{1 + \frac{1}{SBR}}. \quad (3.31)$$

Since the attenuation is applied on both signal and background, according to (3.30), it additionally leads to a maximum feasible signal counting rate $S_{R,MAX}$ evaluated as

$$S_{R,MAX} = \frac{f_0}{1 + \frac{1}{SBR}}. \quad (3.32)$$

Equation (3.32) comes from that fact that, when the total count rate is attenuated to be lower or equal to f_0 , S_R also experiences a maximum limit. For instance, when B_R is zero, SBR is infinity and $S_{R,MAX}$ is f_0 . When B_R is present, it consumes a share of the f_0 bandwidth, thus restricting the maximum S_R . Since in many situations SBR is very low, it is interesting to compare $S_{R,MAX}$ to $S_{R,MIN}$ so as to determine a theoretical limit for SBR , under the assumption of relation (3.30).

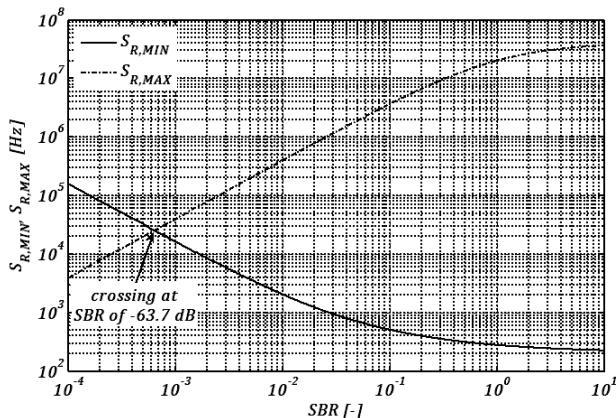


Figure 3.9. $S_{R,MAX}$ and $S_{R,MIN}$ as a function of SBR for a given set of parameters: $T_0 = 25\text{ns}$ ($f_0 = 40\text{ MHz}$), $\Delta t = 300\text{ps}$, $\sigma_s = 100\text{ps}$, $P = 0.997$, and $T = 50\text{ms}$.

Figure 3.9 shows such a comparison as a function of SBR , assuming the same parameters used previously. In the plot, a crossing point of $S_{R,MAX}$ and $S_{R,MIN}$

signals the minimum theoretical SBR for the set of parameters considered. In the example, the corresponding SBR was approximately 6.5×10^{-4} (-63.7dB).

Finally, as one may expect, $S_{R,MIN}$ is monotonic with respect to SBR . For very low SBR , $S_{R,MIN}$ is inversely proportional to SBR , i.e.

$$\lim_{SBR \rightarrow 0} (S_{R,MIN}) = \frac{2\Delta t}{SBR \cdot T_0 \cdot T} \left[\frac{\operatorname{erf}^{-1}(P) + \operatorname{erf}^{-1}\left(1 - \frac{\Delta t}{T_0}(1-P)\right)}{\operatorname{erf}\left(\frac{\Delta t}{\sqrt{8} \cdot \sigma_s}\right)} \right]^2. \quad (3.33)$$

Since the crossing between $S_{R,MAX}$ to $S_{R,MIN}$ typically occurs at very low SBR , one may combine Equations (3.32) and (3.33) to estimate approximately the minimum feasible value of SBR that still leads to a successful measurement of TOF :

$$SBR_{FEASIBLE} > \sqrt{\frac{2\Delta t}{T}} \left[\frac{\operatorname{erf}^{-1}(P) + \operatorname{erf}^{-1}\left(1 - \frac{\Delta t}{T_0}(1-P)\right)}{\operatorname{erf}\left(\frac{\Delta t}{\sqrt{8} \cdot \sigma_s}\right)} \right]. \quad (3.34)$$

The limit value of $SBR_{FEASIBLE}$ is plotted in Figure 3.10 as a function of Δt for the same parameters used in Figure 3.9. The bottom of the curve is reached at Δt of 300ps, thus in agreement with crossing point of Figure 3.9.

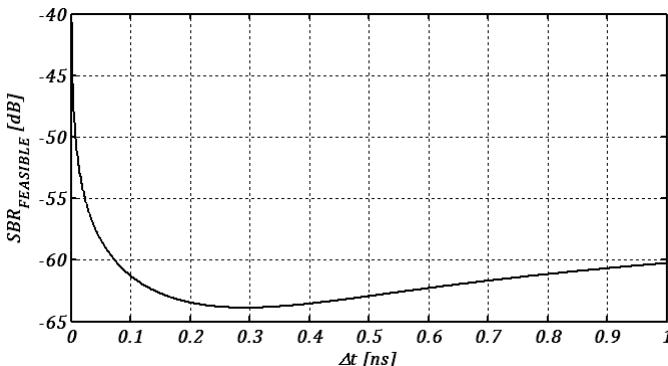


Figure 3.10. Limit value of $SBR_{FEASIBLE}$ as a function of Δt for: $\sigma_s = 100\text{ps}$, $P = 0.997$, and $T = 50\text{ms}$.

3.4 MODELING OF ENVIRONMENT CONDITIONS AND OPTICAL COMPONENTS

In Section 3.3, ranging performance was modeled as a function of photon counting rates. Signal and background photon counting rates are related to a number of system and environment parameters. In this Section, we discuss such relation. These parameters are addressed in an analytical model that may be used to assess ranging performance. Once more, an analytical model is very useful to give an insight on how the detection system behaves when one or more parameters are changed and optimized. It is also useful in a what-if or worst-case analysis of a prospective system.

3.4.1 BACKGROUND LIGHT

In the setup of Figure 3.1, a target placed in front of the TOF ranging sensor is exposed to the active illumination power and to parasitic background light. A fraction of the total incident optical power is reflected on the target's surface and then it is imaged on the single-photon sensor at focal plane. It is advantageous to determine the photon flux due to background light as a function of a simple parameter such as total background illuminance, measured in kilolux. Illuminance data is readily available in the literature for many environmental conditions and may be easily utilized. In particular, worst-case scenarios, e.g. maximum solar illuminance at noon in blue-sky conditions, are very useful in testing and sizing illumination power budget as well as optical components. Illuminance data are also easily measurable in the field by means of lux-meters. Illuminance being a photometric measure, it needs to be converted to a radiometric quantity before useful photon flux can be modeled. Spectral contents of sunlight irradiance may be modeled with Planck's law of blackbody radiation by setting the emitting temperature to 6000K [176]. Equation (3.35) gives a blackbody radiation expression that is normalized so as to lead to unity integral over the spectrum. K is Boltzmann's constant whereas T is the blackbody temperature. This equation is only used to represent the spectral distribution of sunlight radiation.

$$\widehat{I_{BG}}(\lambda) = \frac{1}{\lambda^5 \left(e^{\frac{hc}{\lambda KT}} - 1 \right)} \left(\int_0^{\infty} \frac{1}{\lambda^5 \left(e^{\frac{hc}{\lambda KT}} - 1 \right)} d\lambda \right)^{-1}. \quad (3.35)$$

Sunlight irradiance I_{BG} in Watt per square meter as a function of wavelength is obtained by multiplying $\widehat{I_{BG}}$ by the assumed background illuminance L_{BG} , measured in kilolux, divided by the radiation luminous efficacy K_{RBB} of blackbody at T of 6000K, i.e.

$$I_{BG}(\lambda) = \widehat{I_{BG}}(\lambda) \frac{L_{BG}}{K_{RBB}} \cong 5.13 \times 10^{-23} \frac{L_{BG}}{\lambda^5 \left(e^{\frac{2.398 \times 10^{-6}}{\lambda}} - 1 \right)}. \quad (3.36)$$

In equation (3.36), K_{RBB} was assumed to be 99.3 lumen per Watt, according to [176]. Note that K_{RBB} is not conversion efficiency; it is simply a measure of the effectiveness of an optical beam in stimulating the perception of light in the human eye. I_{BG} is plotted in Figure 3.11 as a function of wavelength for two values of illuminances: 100 klux and 20 klux.

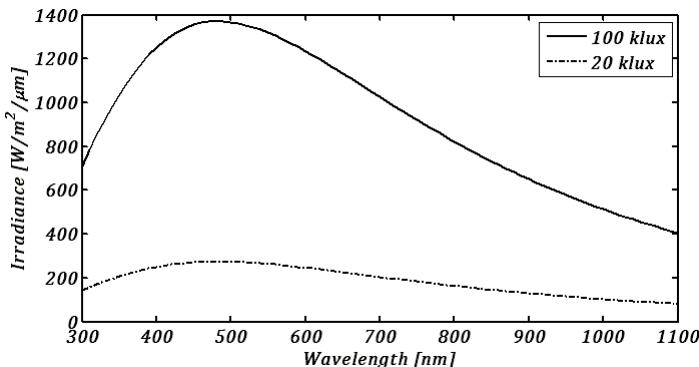


Figure 3.11. Background irradiance I_{BG} as a function of wavelength for two values of illuminance: 100 klux and 20 klux. I_{BG} of Equation (3.36) was multiplied by 10^{-6} to express irradiance per unit of spectrum (μm).

It should be noted that equation (3.36) gives a relatively accurate distribution of extraterrestrial solar spectrum. Sunlight actually experiences absorption in the atmosphere, notably due to water vapor and CO₂, which results in absolute errors with respect to what is predicted in Equation (3.36). One should bear in mind that this expression is an approximation and it is used in this thesis thanks to its simplicity. Large absolute errors may occur in regions of the spectrum where optical absorption is pronounced. A central wavelength of interest in this thesis is 850nm. At this wavelength, errors due to absorption in atmosphere are negligible [176] and Equation (3.36) may be safely considered for performance evaluation.

3.4.2 ILLUMINATION

Modeling illumination is achieved by using design parameters that characterize the illumination system. The main parameter is the illumination mean power P_{SG} . It is desirable to relate the signal irradiance on the target surface at a given distance to P_{SG} . In this context, an assumption should be made on the distribution of the signal power over the target area at a given distance. When a simple lens or diffuser is used to generate an illumination field-of-view, the resulting optical irradiance typically follows a Gaussian distribution with respect to the off-axis distance. Such a non-uniform distribution of irradiance favors the signal at pixels that are close to the optical axis and weakens the signal of peripheral pixels. The Gaussian-shaped beam also results in a circular illumination pattern that does not match square or rectangular forms of pixel arrays at the sensor. Optimized optical diffusers that create a uniform irradiance at the target surface are available and would lead to an excellent solution. These diffusers also generate non-circular illumination patterns that lead to more efficient target-to-sensor matching. Since different solutions exist, in this model, it is assumed that all the illumination power is uniformly distributed over a circular area on the object plane. This assumption leads to more conservative results for pixels close to optical axis and possibly optimistic results on peripheral pixels.

Let us define an emission field-of-view angle α_{FV} measured at the diameter of the circular illumination pattern formed by the source. Considering these parameters, it is possible to determine the signal irradiance I_{SG} on the target surface as

$$I_{SG} = \frac{P_{SG}}{\pi \tan^2\left(\frac{\alpha_{FV}}{2}\right) z^2}, \quad (3.37)$$

where z is the target's distance with respect to the illumination system. As a result of the field-of-view, the signal irradiance decreases with the square of the target's distance. Note that if a square illumination pattern is assumed, the factor $1/\pi$ of Equation (3.37) may be simply replaced by $1/4$.

I_{SG} is not valid at very short distances. This is typically not an issue since a minimum target distance should be assumed. Indeed, the imaging lens is typically focused to infinity to increase depth-of-field.

Finally, the illumination system emits light centered in spectrum on a characteristic wavelength λ_{SG} and with a spectrum broadening that should be as narrow as possible.

3.4.3 OPTICAL BAND-PASS FILTER

A very important optical component in the rangefinder setup is the band-pass filter centered at the emission wavelength of the illumination system (λ_{SG}). An inexpensive type of optical band-pass filter that may be utilized in a wide range of applications is a narrow-band interference filter (NBIF). These filters may be customized in shape and in spectral response relatively precisely. Central band-pass wavelength mismatches are as low as a few nanometers.

The main parameters of NBIFs are a) central pass-band wavelength (λ_F), b) pass-band FWHM in nanometers (F_{BW}), c) transmission at the pass-band (T_F), and d) transmission in stop-band. Although NBIFs with pass-band FWHM as low as 1 nanometer are commercially available, one should bear in mind that the effective λ_F varies according to the incident angle of optical rays. Depending on the position of the NBIF with respect to the imaging lens, the pass-band FWHM needs to be large enough to prevent peripheral optical rays to be blocked. Moreover, λ_F needs to accurately match the illumination central emission wavelength λ_{SG} . Since λ_{SG} of solid-state optical sources, which are cost and power effective, may vary significantly over temperature, extra margin on the pass-band FWHM should be allocated. Typically, more appropriate values of pass-band FWHM are in a range of 10 to 30 nanometers, especially when laser diodes are employed. Alternatively, when light emitting diodes (LEDs) are utilized, this range may be increased up to 60 nanometers or more due to the broad emission spectrum inherent to those devices.

Assuming that the NBIF exhibits a spectral transmission given by $F_{PB}(\lambda)$, it is possible to determine the pass-band filtered background irradiance $I_{BG,F}$ in Watt per square meter as

$$I_{BG,F} = \int_0^{\lambda_{OFF}} F_{PB}(\lambda) I_{BG}(\lambda) d\lambda, \quad (3.38)$$

where λ_{OFF} is the cut-off wavelength in Silicon. Note that, since the transmission in stop-band of typical NBIF is very low, Equation (3.38) may be simplified to

$$I_{BG,F} \cong T_F \int_{\lambda_F - \frac{F_{BW}}{2}}^{\lambda_F + \frac{F_{BW}}{2}} I_{BG}(\lambda) d\lambda \cong T_F F_{BW} I_{BG}(\lambda_F). \quad (3.39)$$

A further approximation was applied on the right-hand side of Equation (3.39). Assuming F_{BW} is very narrow, $I_{BG}(\lambda)$ may be considered constant in that range. Note that Equation (3.38) may also be conveniently used when full spectral characterization is provided by the filter supplier.

3.4.4 IMAGING LENS

Another important component in the rangefinder optical system is the imaging lens. Despite the complexity an image lens design may involve, its light gathering performance may be simply modeled via its f-number ($f_\#$) i.e. the ratio of its aperture diameter to its focal length. Assuming an irradiance I_{OBJ} at object plan, the imaging lens gathers a fraction of I_{OBJ} and redirect it to the focal plane as

$$I_{IMG} = \frac{T_L}{[2f_\#(1+m)]^2} I_{OBJ} \cong \frac{T_L}{(2f_\#)^2} I_{OBJ}, \quad (3.40)$$

where I_{IMG} is the obtained irradiance at focal plane, T_L is the lens transmission, and m is the lens magnification. As suggested in the right-hand side of Equation (3.40), m is typically negligible due to the large difference between the target distance and the focal length. T_L of the lens is typically 90% in medium to good quality lenses.

3.4.5 TARGET REFLECTIVITY

Although the target does not belong itself to the rangefinder optical setup, its optical properties are of high importance for the overall performance. In particular, the reflectivity of objects at the central emission wavelength of the illumination system should be as high as possible. Moreover, objects with diffuse reflection are preferred over specular reflection. Specular reflections typically lead to sensor saturation and other unwanted effects.

In the framework of this thesis, an important class of possible targets includes humans. For instance, in applications that involve the detection of humans, it is very important to predict ranging performance as a function of reflectivity of usual clothes. Although the method may seem purely empirical, we have

characterized a number of different types of clothes with respect to their reflectivity spectral response. In Figure 3.12, a plot of object reflectivity is shown for a few different usual clothes as a function of incident optical wavelength. Since the near infrared spectral region is of interest in rangefinder applications not to disturb users, the characterization was performed from 500 up to 1500 nanometers.

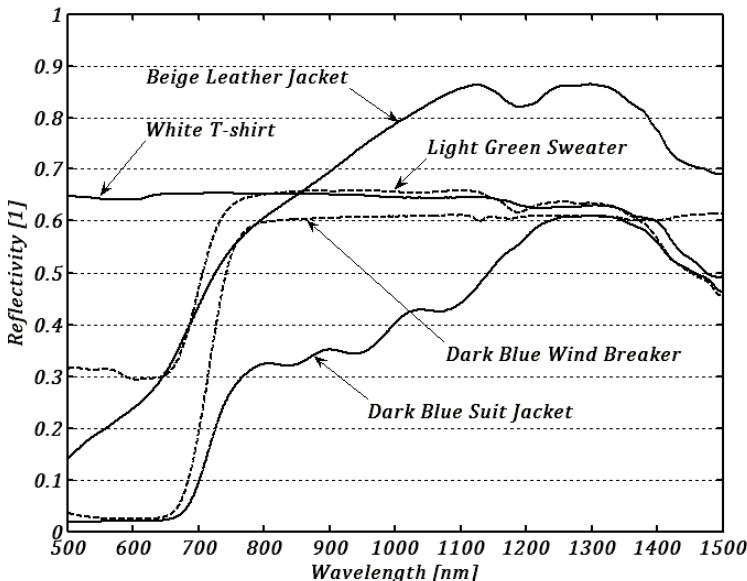


Figure 3.12. Reflectivity as a function of wavelength for a few pieces of usual clothes.

As can be seen in Figure 3.12, object reflectivity in near infrared is often unrelated to their visible color. In particular, dark objects, such as a dark blue suit jacket, which exhibit almost negligible reflectivity in the visible range, reveal reasonable values above 800 nanometers. Most of the tested pieces of clothes showed reflectivity above 50% at 850nm, an interesting central emission wavelength due to solid-state optical source availability. Performance evaluations are typically reported for an object reflectivity R_0 of 50%.

According to the measurements, it would be attractive to use a central wavelength at approximately 1300nm. However, this value is above the cut-off wavelength of silicon and it is therefore beyond the scope of this thesis.

3.4.6 SINGLE-PHOTON IMAGE SENSOR

Signal and background counting rates are determined based on geometric and detection parameters of the image sensor. Since the irradiance falling on the image sensor, through the NBIF, is considered to be almost monochromatic, detector sensitivity may be simply described by the SPAD photon detection probability at λ_{SG} , namely PDP_{SG} .

In order to model the gathering area of a SPAD, we use two parameters: active area fill factor ϕ_{FF} and pixel pitch δ_{PX} . In a number of implementations, it is highly desirable to implement microlenses at the pixel level so as to concentrate light on the SPAD active area. In these cases, ϕ_{FF} may be increased to reflect the effective light gathering fill factor.

3.4.7 SIGNAL AND BACKGROUND COUNT RATES

Based on the parameters described above, it is possible to relate signal and background detection rates, S_R and B_R respectively, as a function of distance.

Assuming that NBIF band-pass is appropriately chosen to be wide enough to accommodate the entire spectrum contents of the illumination signal, the signal irradiance at focal plane $I_{SG,FP}$ is determined as

$$I_{SG,FP}(z) = \frac{R_O T_F T_L P_{SG}}{\pi(2f_\#)^2 \tan^2\left(\frac{\alpha_{FV}}{2}\right)} \frac{1}{z^2}. \quad (3.41)$$

Based on $I_{SG,FP}$, S_R is immediately determined using the image sensor parameters as

$$S_R(z) = \frac{\lambda_{SG}}{hc} PDP_{SG} \phi_{FF} (\delta_{PX})^2 I_{SG,FP}(z). \quad (3.42)$$

The combination of (3.41) and (3.42) leads to

$$S_R(z) = \frac{\lambda_{SG}}{hc\pi} R_O T_F T_L PDP_{SG} \phi_{FF} \left(\frac{\delta_{PX}}{2f_\# \tan\left(\frac{\alpha_{FV}}{2}\right)} \right)^2 \frac{P_{SG}}{z^2}. \quad (3.43)$$

Similarly, background irradiance at focal plane $I_{BG,FP}$ is determined as

$$I_{BG,FP} = \frac{R_O T_L}{(2f_\#)^2} \int_0^{\lambda_{OFF}} F_{PB}(\lambda) I_{BG}(\lambda) d\lambda, \quad (3.44)$$

which may be approximated as

$$I_{BG,FP} \cong \frac{R_O T_L F_{BW}}{(2f_\#)^2} \frac{5.13 \times 10^{-23} \cdot L_{BG}}{\lambda_{SG}^5 \left(e^{\frac{2.398 \times 10^{-6}}{\lambda_{SG}}} - 1 \right)}, \quad (3.45)$$

where λ_F matches the illumination wavelength λ_{SG} . B_R is then determined by using the image sensor parameters, similarly to S_R .

$$B_R = \frac{\lambda_{SG}}{hc} PDP_{SG} \phi_{FF} (\delta_{PX})^2 I_{BG,FP}. \quad (3.46)$$

Note that after the NBIF filter, the background irradiance may be considered to be monochromatic at λ_{SG} and as result, PDP_{SG} is applicable. Moreover, to a first approximation B_R does not depend on target distance. By combining (3.45) and (3.46), one may determine B_R as a function of all the parameters considered, i.e.

$$B_R = \frac{1}{hc} R_O T_L F_{BW} PDP_{SG} \phi_{FF} \left(\frac{\delta_{PX}}{2f_\#} \right)^2 \frac{5.13 \times 10^{-23} \cdot L_{BG}}{\lambda_{SG}^4 \left(e^{\frac{2.398 \times 10^{-6}}{\lambda_{SG}}} - 1 \right)}. \quad (3.47)$$

Note that in Equation (3.47), L_{BG} is defined in kilolux.

Finally, it is interesting to relate optical and environment parameters to the SBR as defined in Equation (3.7), i.e.

$$SBR(z) = \frac{\lambda_{SG}^5 \left(e^{\frac{2.398 \times 10^{-6}}{\lambda_{SG}}} - 1 \right)}{1.61 \times 10^{-22} \cdot F_{BW} \tan^2 \left(\frac{\alpha_{FV}}{2} \right) L_{BG} z^2} \frac{P_{SG}}{L_{BG}} \frac{1}{z^2}. \quad (3.48)$$

As a result, SBR decreases quadratically with target distance.

In order to give some insight on typical photon detection rates based on a CMOS SPAD image sensor, S_R , B_R , and SBR are plotted in Figure 3.13. Typical design parameters were utilized assuming that the non-ambiguity distance was 3.75 meters, corresponding to a modulation or repetition frequency of 40MHz. In

this distance range, a signal power budget P_{SG} of 500mW was considered while the total background illuminance L_{BG} was 100klux. This value is typically considered to be the strongest sunlight illuminance under optimal sky conditions. Since we assumed in the calculation of B_R that the background light illuminates the target from the same angle as the illumination system, the effect of B_R is even more severe than in practice. Note that even using conservative design parameters and based on a very severe background figure, photon detection rates and SBR stay in detectable ranges.

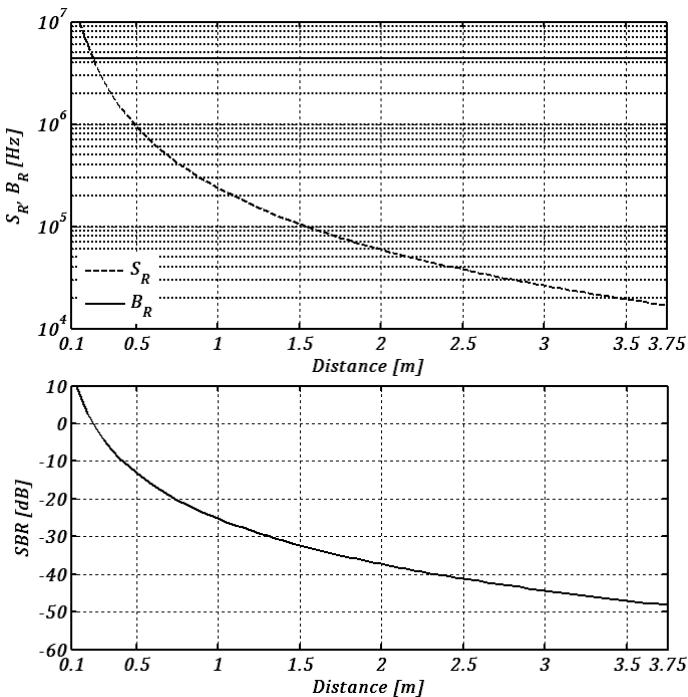


Figure 3.13. S_R , B_R (top) and SBR (bottom) as a function of distance for a given set of parameters: $R_0 = 0.5$, $T_F = 0.7$, $F_{BW} = 30\text{nm}$, $T_L = 0.9$, $f_\# = 1.4$, $\alpha_{FV} = 40^\circ$, $\lambda_{SG} = 850\text{nm}$, $PDP_{SG} = 3\%$, $\phi_{FF} = 6.15\%$, $\delta_{PX} = 25\mu\text{m}$, $L_{BG} = 100\text{klux}$, and $P_{SG} = 500\text{mW}$.

3.5 FULLY-INTEGRATED TCSPC 3D IMAGE SENSOR DESIGN

In this Section, the design and characterization of the first fully-integrated TOF 3D image sensor based on TCSPC is described. This image sensor, implemented in a $0.35\mu\text{m}$ CMOS technology, was first reported in [160]. An extended article has been accepted for publication in the J. of Solid-State Circuits [177]. For the first time, an array of SPADs was monolithically implemented on an integrated circuit comprising a bank of time-to-digital converters, thus delivering TOF measurements digitally off-chip with high-throughput.

3.5.1 IMAGER ARCHITECTURE

In Figure 3.14, a block diagram of the image sensor is shown. The sensor consists of an array of 128×128 single-photon pixels based on the $0.35\mu\text{m}$ SPAD design, discussed in Chapter 2. A bank of 32 TDCs was integrated on the same IC. TDCs are utilized to compute time-interval measurements between a global start signal and photon arrival times in individual pixels. A row selection decoder is used to activate one row of 128 pixels that have access to the bank of TDCs. In this design, the 32 TDCs are shared among 128 pixels in a given row. The sharing scheme is based on a 4:1 event-driven readout [157] that allows the 128 pixels in a row to operate simultaneously. Since every TDC is shared among four pixels, time-to-digital conversion time was highly optimized at design phase so as to maximize throughput.

The calibration of TDCs is implemented on-chip based on a master delay-locked loop (DLL) that locks against an external reference frequency, generated by a crystal oscillator. As a result, TDC resolution and linearity are maintained over process, voltage, and temperature (PVT) variations.

At the bottom of the TDC array, a high-speed digital readout circuit handles the data generated by all the TDCs. It consists of a pipelined 4:1 time-multiplexer that operates at a frequency at least four times faster than the data generated in every TDC. A readout controller generates all the signals utilized internally and implements a readout protocol interface. Most of digital building blocks in the TDC and readout circuitry provide duplicated (shadow) registers that may be read and written via an on-chip JTAG controller. The JTAG controller offered advantageous testing and characterization options.

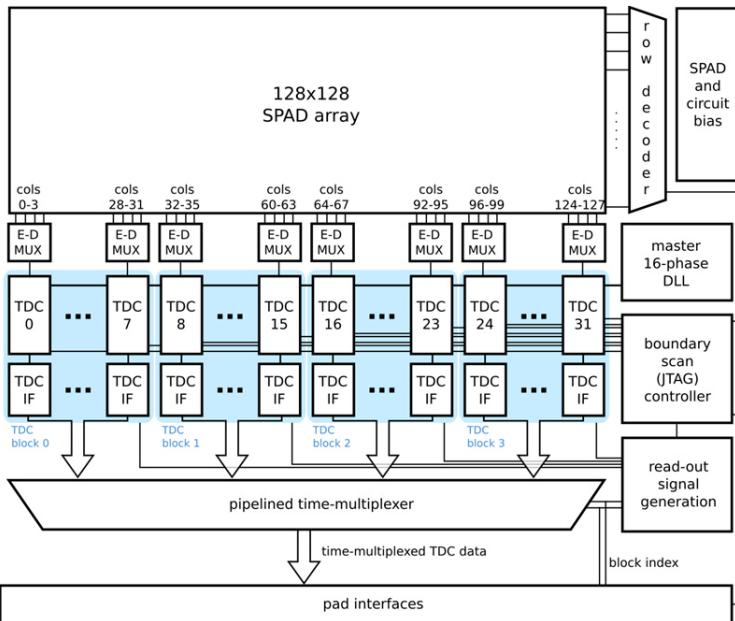


Figure 3.14. Block diagram of the image sensor. The sensor consists of a 128x128 pixel array, a bank of 32 TDCs, and a fast parallel readout circuitry. A row decoding logic selects 128 pixels that are activated for detection. The pixels are organized in groups of four that access the same TDC based on an event-driven sharing scheme.

3.5.2 PIXEL CIRCUIT

Single-photon detection capability is achieved in the pixel via a p+/deep n-well SPAD, using p-well as guard-ring, according to Chapter 2. At room temperature, the breakdown voltage V_{BD} of the SPAD in this design is 17.7V. At its cathode, a bias voltage of 21V is applied in order to operate nominally with an excess bias voltage V_E of 3.3V. Miniaturization constraints led to a solution based on a NMOS-only circuit for the pixel. The SPAD front-end and readout circuit consists of 7 NMOS transistors. Figure 3.15 shows a simplified schematic of the pixel circuit, which implements passive quenching and active recharge. A row selection transistor (M_1) decouples the SPAD anode from pixels in the rows that are not selected. At the selected row, the SPAD is charged as a result of its anode being connected to ground via quenching/recharge transistor M_2 . Passive quenching and active recharge is controlled by means of a switch that connects

the gate-to-source voltage of M_2 , $V_{GS,2}$, to either V_{QCH} or V_{RCH} . Prior to any photon detection, $V_{GS,2}$ is connected to V_{QCH} , a low voltage, thus providing a very high impedance from node V_x to ground. This impedance is typically higher than one megaohm. When a photon triggers the SPAD, the avalanche current induces a voltage across M_1 which pulls V_x towards V_E , according to the waveform shown at the bottom of Figure 3.15, thus quenching the SPAD. This voltage transition, which reflects the arrival time of a photon, is accurately transferred to the column line via the output transistor M_7 . At the bottom of the column, the corresponding TDC starts performing a time-to-digital conversion. During the conversion time, since M_2 still provides a high impedance path to ground, the SPAD is maintained in dead-time state while V_x decreases very slowly towards ground. Note that the waveform shown in Figure 3.15 is not to scale. At the moment the TDC is approaching the end of a conversion, it triggers the switch position so as to apply a recharge voltage V_{RCH} on $V_{GS,2}$, thereby performing a well defined active recharge. From this point, the pixel is able to detect photons again.

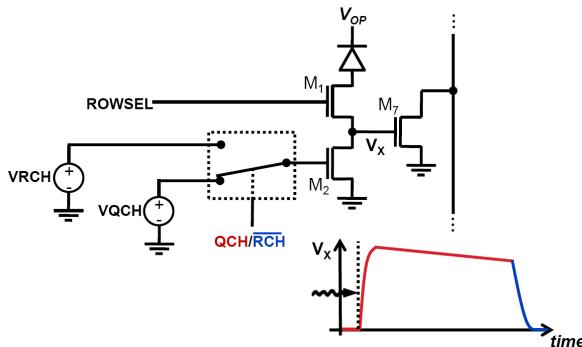


Figure 3.15. Principle of pixel circuit providing passive quenching and active recharge (top). Example of signal waveform on photon detection (bottom).

In Figure 3.16, the actual implementation of the pixel circuit is shown. Transistors M_3 , M_4 , and M_6 implement the switches to set $V_{GS,2}$. M_5 is used as a capacitor to reduce the effect of switching noise on V_{QCH} caused by charge injection from the gate of M_2 . Signals QCH and RCH are generated by the TDC controller. At the bottom of each column, a signal shaping circuit ensures that QCH and RCH are non-overlapping, so as to prevent current to flow from V_{RCH} to V_{QCH} . The column line potential is kept high by pull-up transistors at the bottom of the column when it is inactive. Bias voltages V_{OP} , V_{QCH} , and V_{RCH} are global for

the pixels in the SPAD array. In particular, V_{QCH} and V_{RCH} are distributed by means of a number of analog buffers to insure stable reference voltages.

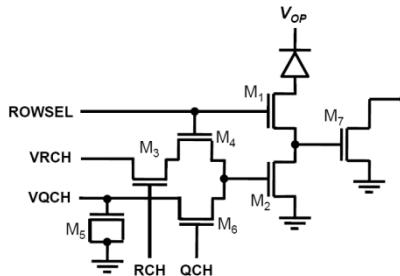


Figure 3.16. Schematics of actual pixel implementation.

3.5.3 TIME-TO-DIGITAL CONVERTER

A very critical component involved in the design of picosecond-resolution time-resolved imager is the TDC. In this design, the initial requirement specifications were: a) 10-bit resolution TDC with adjustable least significant bit (LSB) value ranging from 70 to 200 picoseconds, b) an automatic calibration against PVT variations, c) a high-throughput with a maximum conversion time of 100ns. In addition, d) acceptable reproducibility and yield were suited, and e) a scalable design was desirable in view of future deep-submicron integrations.

Taking into account all these requirements, a careful solution search leads to a TDC based on a three-level interpolator. Based on this principle, the TDC combines coarse (τ_C), medium (τ_M), and fine (τ_F) resolutions to resolve a measurement range of 10 bits with a LSB value equal to τ_F . The TDC interpolation principle is schematically shown in Figure 3.17. A global clock signal CLK modulated at a frequency of 40MHz is used as a PVT-invariant reference signal. The period of CLK is used to set τ_C , it is then split into 16 uniformly shifted phases PHI[15:0] by means of a global DLL. The delay between two adjacent phases, i.e. between PHI[i] and PHI[i+1], is used as τ_M . Typically, τ_C and τ_M are 25ns and 1.56ns, respectively. At the TDC level, a very short delay of 97.6ps is synthesized out of PHI[15:0] and used as τ_F . Finally, an appropriate interpolation based on a linear combination of τ_C , τ_M , and τ_F enables the measurement of a time interval within a typical range of 100ns. This solution has a number of advantages. Multi-level interpolation allows for the determination of relatively long time intervals with high resolution without the need of long delay lines, while distributing relatively low frequency signals only.

Long delay lines usually occupy relatively large silicon area. Moreover, when a signal propagates through a long delay line, it accumulates timing jitter, thus causing a broadening of timing uncertainty. The proposed solution is advantageous especially with large arrays of TDCs. Since τ_c and τ_M are synthesized globally, they do not use additional circuit area at TDC level. As a result, relatively small and low power TDCs may be obtained.

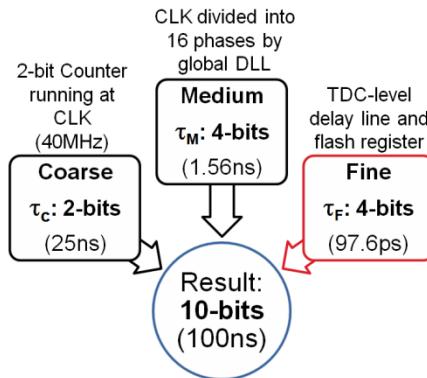


Figure 3.17. TDC interpolator principle. Coarse and medium resolution are generated globally and fine resolution is implemented on each TDC. This solution leads to efficient area utilization while achieving large measurement range.

In Figure 3.18 a simplified block diagram of the TDC is shown. An example of the waveform of key signals involved in the time-to-digital conversion is illustrated in Figure 3.19. The main TDC structure is similar to the design reported in [27]. Nonetheless, further improvements have been implemented to reduce the silicon area and to perform flash conversion, thus increasing throughput. Each TDC has an independent controller that is used as a time interpolator, to generate internal signals, and to control its operating mode. Each controller also manages the interface with the global readout circuit and with the column circuitry. The TDC supports two main operating modes: (i) a measurement mode and (ii) a calibration mode. In measurement mode, the STOP signal originated by the first of four SPADs that detects a photon is mapped to signal TRG in the column-level TDC via a multiplexer, i.e. signal CALRQ is deasserted. TRG precisely and securely registers the state of the 2-bit coarse counter as well as the state of 16 phases, thereby generating two and four bits of conversion data, respectively.

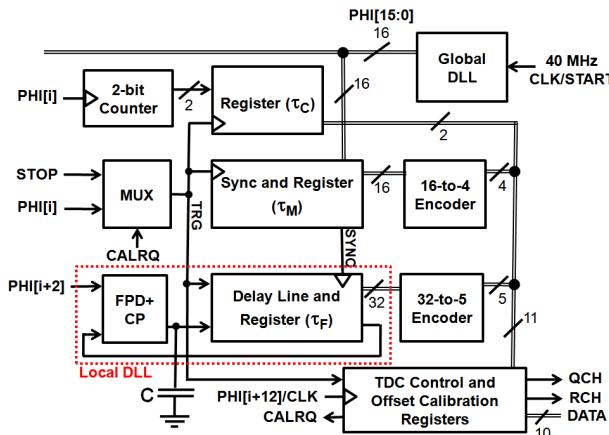


Figure 3.18. Simplified TDC block diagram.

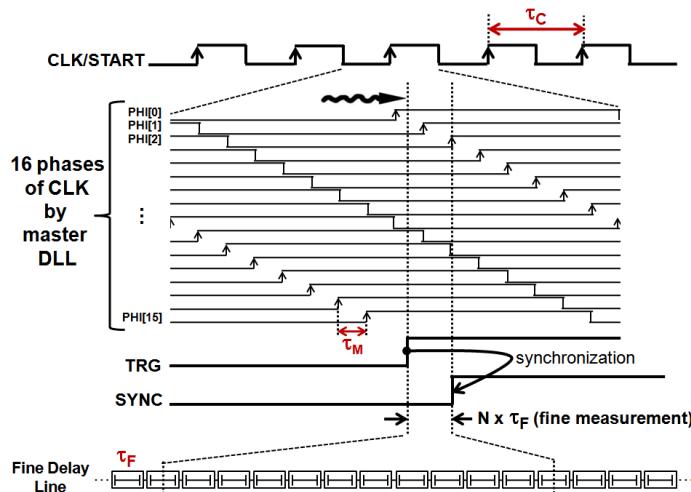


Figure 3.19. Waveform of key signals involved in a TDC.

In order to determine the four remaining least-significant bits, i.e. fine resolution, the next transition of **PHI[15:0]** following the assertion of **TRG** is of interest. Besides registering coarse and medium resolutions, signal **TRG** is also fed into a 32-tap fine delay line, whose propagation state is sampled on the next transition of **PHI[15:0]**. This is achieved by signal **SYNC**, which is generated in the medium resolution register, shown in Figure 3.20. Metastability issues make

it difficult to rely on the exact first transition of PHI[15:0] to assert SYNC. Indeed, when the setup time of a register is violated, its clock to output propagation time suffers from large dispersion, thus potentially impacting the performance of the TDC. In order to avoid this issue, the circuit of Figure 3.20 skips the first transition and securely asserts SYNC on the second transition of PHI[15:0], thus resulting in a timing accurate synchronization signal.

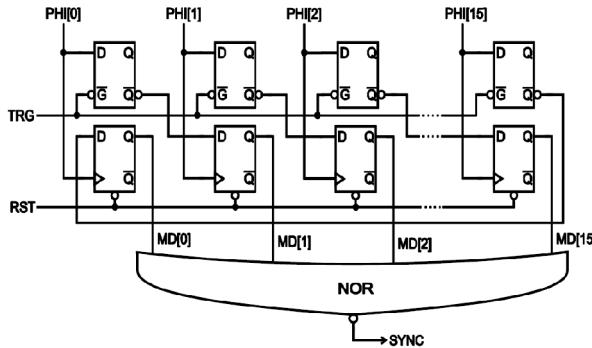


Figure 3.20. Medium resolution register and synchronization circuit.

The time interval between TRG and SYNC is measured by determining how far TRG propagates in the fine delay line before SYNC is asserted. The fine delay line and associated fast register are shown in Figure 3.21. The fine delay line consists of two chains of optimized inverters. An input STRT (TRG) signal is converted to a fully-differential signal, whose transitions are aligned via an edge aligner circuit. Each component of the differential signal propagates through a different chain of inverters. In order to achieve the specified requirement in terms of propagation delay for inverters, used as τ_f , two types of inverters have been utilized: an inverter optimized for low-to-high transitions and an inverter optimized for high-to-low transitions. Each type of inverter is placed in the delay line in such a way that its effective transition occurs during a time interval measurement, according to Figure 3.21 (a). Although the two inverters are different, their effective propagation times need to match. Note that any timing mismatch between them is canceled out during propagation of two consecutive delay cells. As a result, a potential timing mismatch does not cause a severe degradation in the TDC linearity. In order to sample the state of the propagation of STRT in the delay line, a differential D flip-flop is used as register, triggered by a STOP (SYNC) signal. The differential flip-flop, shown in Figure 3.21 (b), is based on a sense amplifier topology and it is symmetric with respect to its data input, similarly to [178].

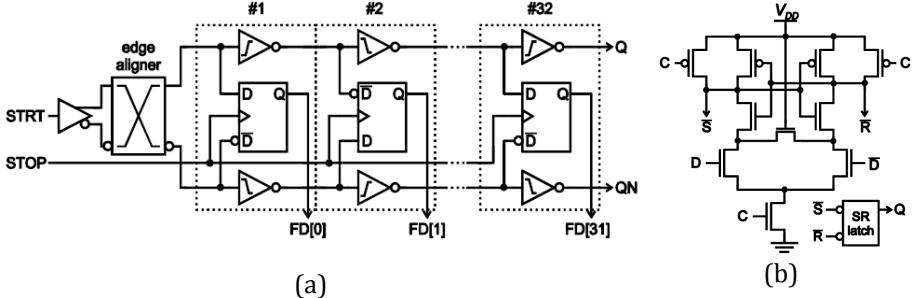


Figure 3.21. (a) Fine delay line and register. (b) Transistor-level schematics of the differential D flip-flop used as register.

Only 16 delay cells (4 bits) of the fine delay line are used for the final result. The remaining 16 delay cells are added to accommodate timing shifts due to PVT variations. In the TDC controller, an interpolator utilizes the three measurements, i.e. coarse, medium, and fine, to generate a 10-bit time-to-digital code.

Referring back to Figure 3.18, when the TDC is in calibration mode, it utilizes the full 32-tap fine delay line in combination with a phase/frequency detector (PFD) and a charge pump (CP) so to form a local DLL. In this mode, CALRQ is asserted by the TDC controller which, in turn, connects PHI[i] to TRG via the input multiplexer. The DLL locks to two non-successive phases (PHI[i] and PHI[i+2]) with a total duration of $2\tau_M$, thus generating a fine resolution τ_F of 97.66ps. The analog control voltage of the DLL is stored on a local capacitor and used in measurement mode. Note that since the calibration is performed individually in each TDC, matching requirements between TDCs may be relaxed.

Since every TDC locks to two non-successive global phases PHI[15:0], the main DLL is of high importance. Fortunately, it can be optimized with little or no constraint in silicon area as the imager only has a single instance of it. Moreover, in order to guarantee that every phase line, say PHI[i], is loaded with the same capacitance, the 32 TDCs are distributed evenly among the 16 phases. For instance, reference signal PHI[i] of the 1st and 17th TDCs is PHI[0], whereas the 2nd and 18th TDCs utilizes PHI[1] as reference, the 3rd and 19th TDCs utilizes PHI[2], and so on. Although this scheme requires that the number of TDCs is a multiple of the number of phases, it ensures no load mismatches and consequently leads to high timing accuracy between reference phases.

3.5.4 READOUT CIRCUIT

The output of all TDCs is transferred off-chip via a fast global readout circuit consisting of 32 TDC interface blocks and a pipelined time-multiplexer readout chain. Figure 3.22 shows a simplified block diagram of the readout circuit.

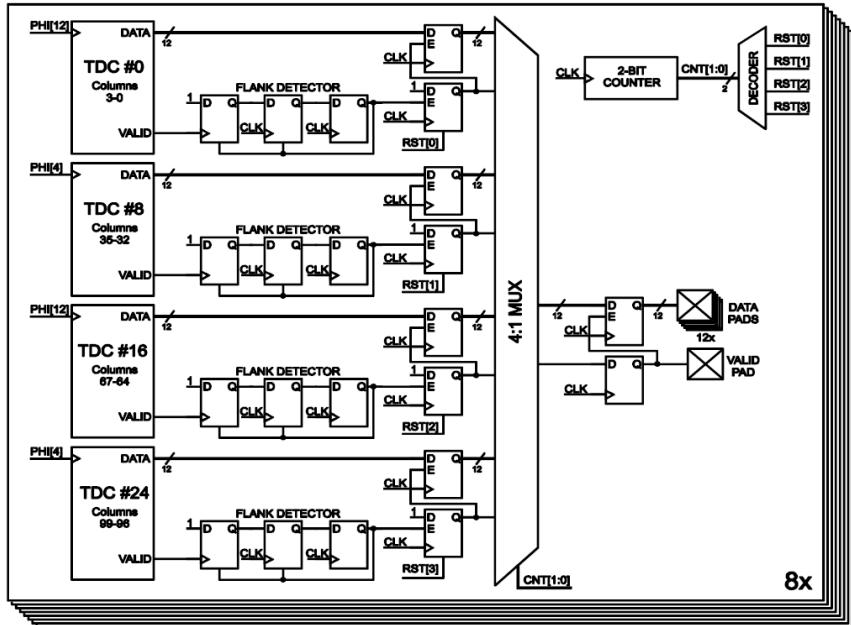


Figure 3.22. Simplified block diagram of the readout circuit. The basic circuit is instantiated 8 times in the imager so as to read out the data of all the 32 TDCs in parallel.

In the figure, the basic readout circuit, covering four TDCs, is shown. This circuit is instantiated 8 times on the image sensor, so as to read out the data from all 32 TDCs in parallel. Note that TDC data consist of 12 bits. 2 bits are used to indicate in which column the photon was detected. Data readout in a given cycle is requested by a TDC via a control signal, called VALID, indicating that valid data are available. TDCs and readout circuit operate in two different clock domains. As a result, every basic readout circuit has a TDC interface circuit to handle synchronization issues, as can be seen in Figure 3.22. The interface block securely samples the 12-bit data and VALID bit within the readout clock domain. A pipelined time-multiplexer chain transfers the TDC data along with

the VALID bit onto a set of digital pads, consisting of one output bus. In order to maximize data rate, the readout circuit operates at a frequency of at least four times faster than the TDC sample rate. As a result, every TDC may operate at maximum throughput. Moreover, as shown in Figure 3.22, data pads only toggle state when valid data are available in a readout cycle, thus minimizing power consumption.

As mentioned, the readout circuit provides several configuration/testing modes to read and modify most TDCs and readout circuit registers via the integrated JTAG controller and standard 4-bit JTAG bus.

3.5.5 IMAGER CHARACTERIZATION

This design was fabricated in a high-voltage 0.35 μm CMOS technology. The implemented IC was tested and characterized with respect to a number of performance parameters. Basic testing demonstrated full functionality of the design.

SPAD Array

Figure 3.23 shows a photomicrograph of the IC along with a detail of the pixel. The pixel pitch was 25 μm , while the active diameter of the circular SPAD was 7 μm . Thus the fill factor is 6.2%.

The SPAD utilized in this design is described in Chapter 2, where full characterization measurements are reported. In this design, thanks to the size of the array, the measurement of DCR distribution over a very high number of Geiger-mode photodiodes became possible, and it is therefore reported for the first time in Chapter 2. At 27°C, the median value of DCR was 694Hz whereas the mean value was higher, at 2.4kHz, mostly due to a small number of highly noisy pixels. In many applications, it is advantageous to electronically ignore these pixels and replace them by local averages. As a result, the median value of DCR is a better performance parameter than the mean value. The measured median DCR also stays fairly constant among devices from the same fabrication batch. The median DCR measured at room temperature over a few samples of the design reported in this work spreads of only 96Hz, from the best device to the noisiest one.

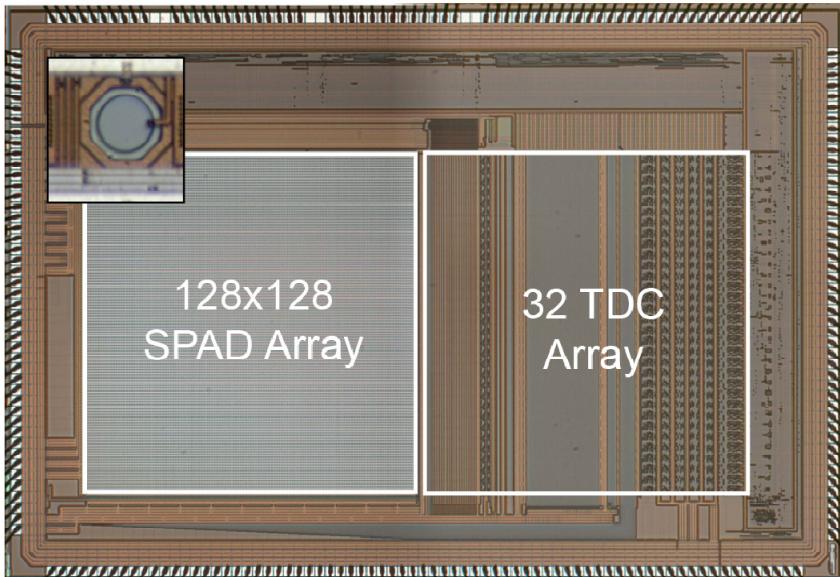


Figure 3.23. Photomicrograph of the TCSPC image sensor with a pixel detail in the inset. The integrated circuit, fabricated in an $0.35\mu\text{m}$ CMOS technology, has a surface of $8\times 5\text{ mm}^2$. The pixel pitch is $25\mu\text{m}$, which leads to an active area fill factor of 6.2%.

Time-to-Digital Converter Array

The bank of TDCs and associated circuitry were tested over a number of days of operation. TDC characterization was performed based on differential non-linearity (DNL). The integral of DNL over all the possible resulting codes of a TDC, i.e. integral non-linearity (INL), was also computed from DNL measurements. DNL was measured taking advantage of the statistical properties of uncorrelated photons, which follow a Poisson distribution. The stop signal of the TDC under test was connected to a single-photon pixel which, under uncorrelated background light, generated a train of photon-detection pulses with random arrival times. The pixel count rate was adjusted so as to provide a mean arrival time much larger than the TDC measurement range, which was 100ns in typical conditions. Since the TDC stop assertions were random in time with respect to the start signals, the probability of generating a particular TDC code within the measurement range was uniformly distributed. As a result, when a histogram on the generated TDC codes is built using this measurement setup, a uniform histogram is expected for an ideal TDC.

However, due to non-idealities within the TDC, some TDC codes account for a longer duration than the ideal TDC resolution and, as a result, those codes have higher probability of capturing events. These non-uniformities are recorded in the resulting histogram. Similarly, TDC codes with duration shorter than the ideal resolution have lower probability of being triggered by photons and exhibit lower bin values within the histogram. DNL can be consequently extracted from the histogram by comparing every bin value to the ideal expected value. DNL and INL measurements were carried out for all the 32 TDCs. Figure 3.24 shows the plot of DNL and INL for a TDC, which was chosen as the worst-case INL among all the 32 TDCs. Due to electrical crosstalk between some TDC signals and the clock, the INL response was relatively higher than expected. Its maximum value was 1.89 LSB. Note that since the full TDC range covers 4 clock cycles, DNL and INL exhibited a periodic behavior.

In order to complete TDC testing and characterization, the image sensor was illuminated with a pulsed laser source with a repetition rate of 40 MHz and pulse duration of 80ps. The trigger out signal of the laser source was used a start signal for the TDCs and as a result, time intervals from laser source to the image sensor could be measured. Figure 3.25 shows a histogram of the photon arrival times as recorded using an on-chip TDC. In the picture, a single pulse is shown although the actual histogram comprises four pulses due to TDC range, which is four times longer than the laser period. The FWHM of the recorded pulse was approximately 250ps, thus showing some additional jitter components arising from the single-photon pixel and TDC circuitry. As can be seen in the picture inset, where a semi-log plot is shown, the Gaussian shaped component in the jitter measurement indicates a typical circuit jitter contribution. The exponential tail, on the other hand, is caused by long carrier diffusion times within this SPAD device. Circuit jitter was expected to be an issue due to signal crosstalk between miniaturized pixels and TDCs. Indeed, when all the TDCs are in operation, such as when the measurement of Figure 3.25 was performed, digital switching noise is coupled to analog and power signals due to space constraints, which induce timing jitter on some internal signals of the TDCs.

In Table 3.1, a summary of design and performance parameters for the fully-integrated TCSPC image sensor is shown.

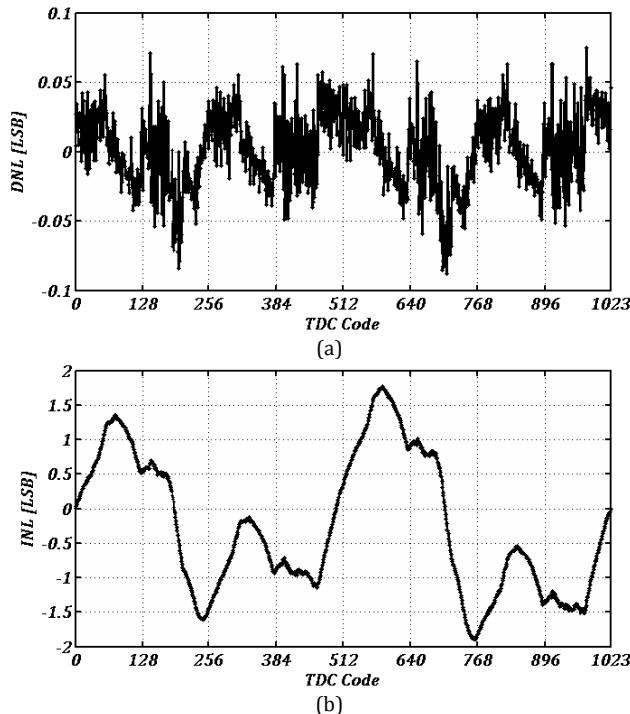


Figure 3.24. Measurements of (a) differential non-linearity (DNL) and (b) integral non-linearity (INL) for the worst case TDC at room temperature.

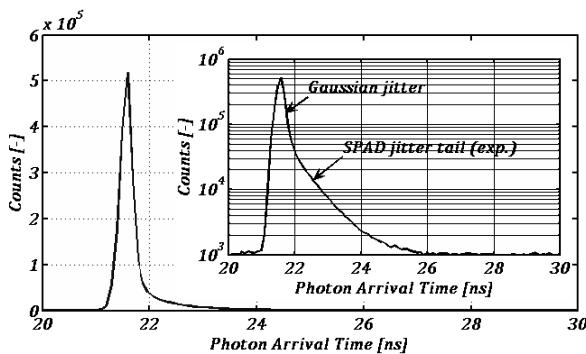


Figure 3.25. Time jitter measurement of the SPAD detector and overall circuitry using the integrated TDCs. In the inset, a logarithmic plot is shown.

Table 3.1. Summary of design and performance parameters for the fully-integrated TCSPC imager.

<i>Parameter</i>	<i>Symbol</i>	<i>Min.</i>	<i>Typ.</i>	<i>Max.</i>	<i>Unit</i>
Pixel	Image resolution		128x128		
	Pixel pitch		25		μm
	SPAD active area		38		μm ²
	Photon detection probability @ $V_E=3.3V$	PDP_{SPAD}	3	35 @ 460nm	%
	Sensitivity spectrum	λ	350	900	nm
	Median DCR at 27°C of temperature	DCR	598	646	Hz
	Dead time (TDC limited)		100		ns
TDC	Tuning of measurement range (10bits)		71.68	100	204.8
	Resolution (LSB)	τ_F	70	97.66	200 ps
	Measurement rate			10	MS/s
	Differential non-linearity	DNL		0.08	LSB
	Integral non-linearity	INL		1.89	LSB
IC	Clock frequency (tunable range)	f_0	19.5	40	55.8 MHz
	Total IO bandwidth			7.68	Gbps
	JTAG bandwidth			8	Mbps
	Static power dissipation			33	mW
	Dynamic power dissipation		150		mW

3.6 EXPERIMENTAL RESULTS IN SHORT RANGE

In this Section, experimental characterization of the TCSPC 3D imager is reported in the context of short range. In particular, the evaluation aimed at demonstrating the potential of parallel TCSPC acquisition and processing for range-finding based on extreme low-power illumination.

Ranging performance was first quantitatively evaluated by studying the distance error versus a reference panel, assumed as ground truth. A solid-state pulsed laser source emitting an average of 1mW of optical power at 635nm was used to illuminate the reference panel. The laser repetition rate was 40MHz and the duration of light pulses was 80ps (FWHM). In order to accurately determine TOF, the trigger out signal of the laser source was used as a reference signal for the imager to lock on. A diffuser was placed in front of the laser source so as to generate a light beam with a field-of-view of 5°. The sensor was equipped with an imaging lens (f/1.4) and with a narrow-band interference filter centered at 635nm to block most of background light. The passband spectral width of the optical filter was 11nm, enough to accommodate the spectral spread of the laser

source around its central emission wavelength. The evaluation of distance precision of our sensor was carried out under a constant 150lux indoor background.

Experimental measurements were also used to validate the analytical models introduced in Sections 3.3 and 3.4. The model of Section 3.4 aims at relating optical components and environmental conditions to actual photon count rates, i.e. S_R and B_R . Once count rates are known for a set of design parameters, it is possible predict performance and reliability based on the TCSPC model described in Section 3.3.

The parameters corresponding to the optical setup, reference plane (object), TCSPC image sensor, and illumination were used as input of the analytical model. Since the illumination repetition rate f_0 was 40MHz, the unambiguous distance range was 3.75 meters. However, taking into account the minimum required working distance, the first measurement distance was 40cm. As a result, the reference panel was carefully aligned in front of the sensor and its distance was varied from 40cm up to 360cm in steps of 20cm, using an alternative measuring method. Systematic errors in the alignment and displacement of the reference panel, using the alternative method, were estimated to be within +/- 3mm.

At each distance, the total photon count rate of the pixel under investigation was recorded. In Figure 3.26, the total count rate T_R , i.e. $S_R + B_R$, was plotted as a function of actual distance. In the same plot, the counting rate that was predicted by the theoretical model is also shown. However, since the experiments were carried out in indoor conditions, the blackbody radiation model, described in Section 3.4.1, exhibited a strong mismatch to the model. Although a constant illuminance of 150lux was maintained on the reference surface over the distance range, an artificial illuminance of 300lux had to be assumed in model of B_R . Based on this value, the model exhibited acceptable matching, as shown in Figure 3.26.

In order to evaluate the potential of TCSPC for a real time rangefinder, an integration time T of 50ms was considered. As can be seen in Figure 3.25, the overall jitter of the sensor was approximately 250ps FWHM. This timing jitter includes the timing resolution of the SPAD, jitter contributions in TDCs, and the illumination pulse. As a result, the parameter characterizing half of the signal pulse width σ_S was assumed to be 112ps (0.425 of FWHM). Remaining parameters were: ΔT of 97ps (TDC), $N_{HW} = 3$, and $P = 97.7\%$. Note that T_0 is the inverse of the illumination repetition rate.

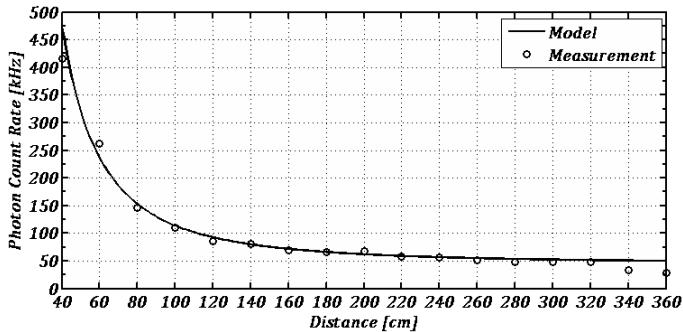


Figure 3.26. Measured and modeled total photon count rate T_R as a function of distance.

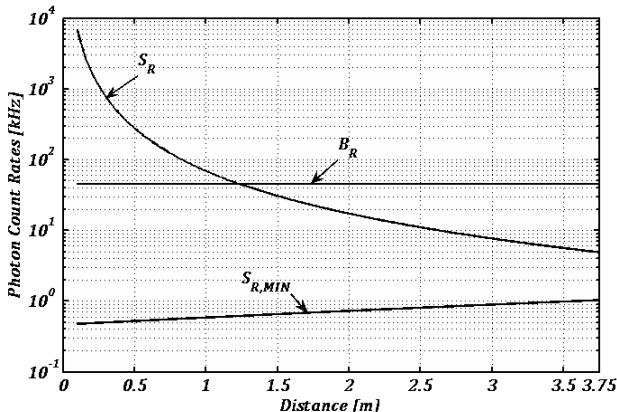


Figure 3.27. S_R , B_R , and $S_{R,MIN}$ as a function of distance, according to theoretical model. S_R fulfills the condition of $S_{R,MIN}$ over the full range. The model therefore predicts that accurate distance measurements are expected within this range.

Figure 3.27 shows a plot of modeled S_R , B_R and $S_{R,MIN}$ as a function of distance. In the plot, S_R exhibits a typical quadratic decay due to the field-of-view of 5° in the illumination system. B_R is constant over distance due to the controlled indoor illuminance. Interestingly, despite the low-power illumination of 1mW within a field of view of 5° and short integration time, the $S_{R,MIN}$ required by the model was comfortably lower than S_R over the full range. This behavior predicts that, under these conditions, the acquisition and processing of TCSPC histograms

mostly lead to accurate TOF computations. The margin between S_R and $S_{R,MIN}$ additionally suggests that presumably 100% of distance computations are accurate. Furthermore, note that throughout the distance range, S_R stays considerably lower than f_0 , i.e. $M \ll 1$. As a result, no pile-up distortion, as described in Section 3.2, should be experienced in this measurement setup.

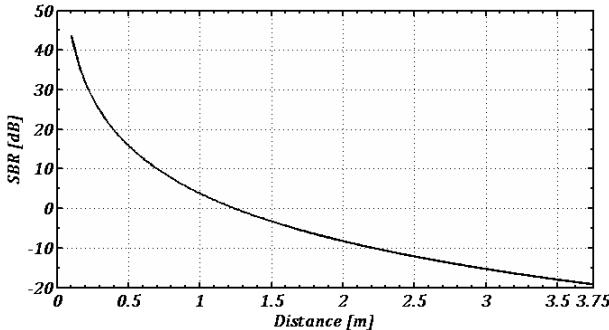


Figure 3.28. SBR as a function of distance, according to the theoretical model.

In Figure 3.28, a plot of the SBR as a function of distance is shown for convenience. The worst-case SBR is approximately -20dB, at the end of the measurement range. In Section 3.3, theoretical limits of TCSPC with respect to SBR were as low as -60dB. Despite the 150lux of background light, SBR was not challenging for the TCSPC setup. The reason for this was the use of the very narrow optical filter.

Furthermore, based on the analytical model, it is possible to estimate the distance resolution σ_{TOF} as a function of distance, shown in Figure 3.29. Sub-millimeter distance precision was predicted in most of the distance range, due to favorable SBR . However, as mentioned in Section 3.3, σ_{TOF} does not include quantization errors due to finite histogram resolution ΔT .

At each evaluation distance, a set of 10000 distance computations was recorded for the pixel under investigation. In Figure 3.30, a sample histogram for each distance step is plotted, based on the measurement conditions described previously. As can be seen, the bandpass optical filter was effective in eliminating uncorrelated photons. Remaining background photons which fall on the laser emission spectral band and possible dark counts are distributed evenly over the full histogram and therefore have a negligible impact on the

signal peak, which was accurately built around each distance, as predicted by the analytical model.

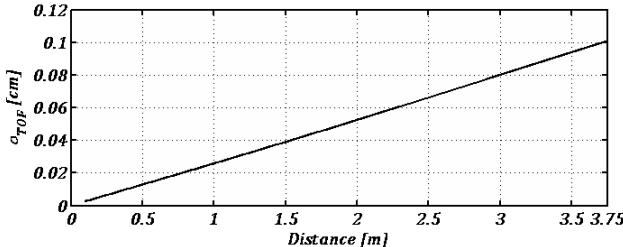


Figure 3.29. Estimated σ_{TOF} as a function of distance. Note that the model does not include quantization errors in σ_{TOF} .

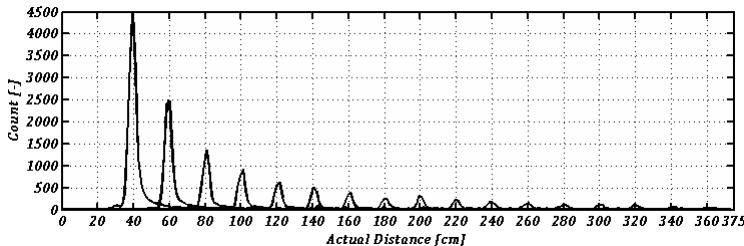


Figure 3.30. Histogram sample curves obtained when moving the target from 40cm up to 360cm with steps of 20cm. Measurements using a 40MHz pulsed laser source emitting 1mW of optical power within a field-of-view of 5°. Each histogram was built based on an integration time of 50ms.

In Figure 3.30, a quadratic decrease in signal peak as a function of distance is clearly visible due to the light source field-of-view. Figure 3.31 summarizes the statistical evaluation of range-finding performance based on the proposed setup. In Figure 3.31 (a), measured distances as a function of ground truth distance is plotted. A linear fit is also plotted in the same figure. In Figure 3.31 (b), the mean non-linearity errors (μ_{Error}) with respect to actual distances are plotted. Finally, Figure 3.31 (c), depicts the 1σ repeatability error (σ_{Error}) as a function of distance.

Despite the quadratic decrease in signal peak as a function of distance, accurate TOF evaluations were achieved throughout the range. Note that only a few tens of photons were enough to accurately determine distance, especially towards the end of the range.

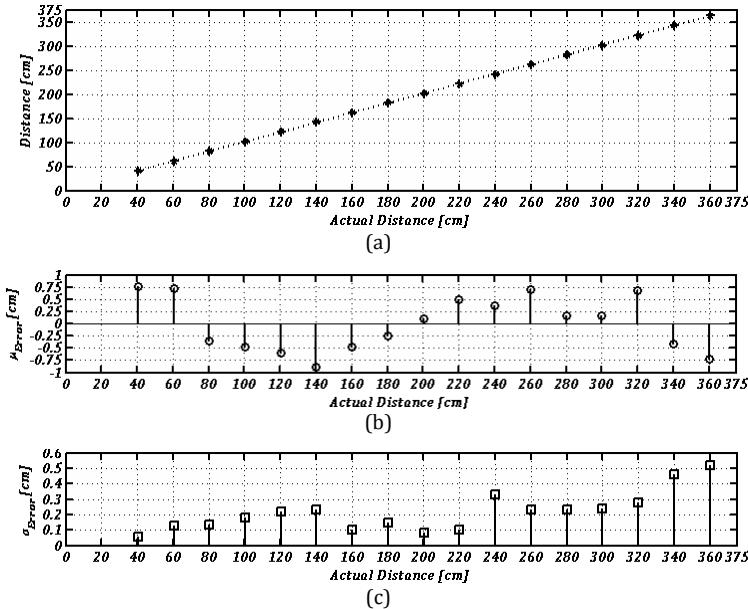


Figure 3.31. Rangefinding performance using a 40MHz pulsed laser source emitting 1mW of optical power within a field-of-view of 5°. Every distance measurement was based on an acquisition time of 50ms: (a) measured versus actual distance, (b) mean error with respect to ground truth, and (c) 1 σ repeatability errors as a function of distance

The maximum μ_{Error} was within +/- 9mm over the full range, whereas the maximum time-varying uncertainty σ_{Error} , at each distance, was 5.2mm. μ_{Error} is mostly due to TDC non-linearity, while a fraction of it may be explained by the systematic uncertainty of +/- 3mm in aligning the reference panel. At short distances, μ_{Error} was also believed to be limited by signal crosstalk within the lens cavity due to sensor-to-lens and lens-to-sensor multi-reflections. These effects become negligible at longer distances as the number of incident photons reduces strongly. Conveniently, since the signal peak spreads over at least 7 bins in the histogram, the effects of TDC non-linearity and some quantization errors were attenuated, thus leading to an outstanding μ_{Error} lower than 9mm. σ_{Error} performance was excellent over most of the range, taking into account the limited 1mW of optical power and 50ms of acquisition time. Towards the end of the distance range, the number of photons detected per histogram was only a few tens and as a result, photon shot-noise resulted in increased single-shot distance uncertainty. However, measured σ_{Error} was considerably higher than

the estimation of σ_{TOF} , based on the analytical model. In order to investigate this large mismatch, the distributions of measured distances over the 10000 measurements were carefully analyzed. In Figure 3.32 (a) and (b), measured distances, when the reference was placed at 40cm and 60cm from the imager, were plotted for the successive 10000 TOF computations.

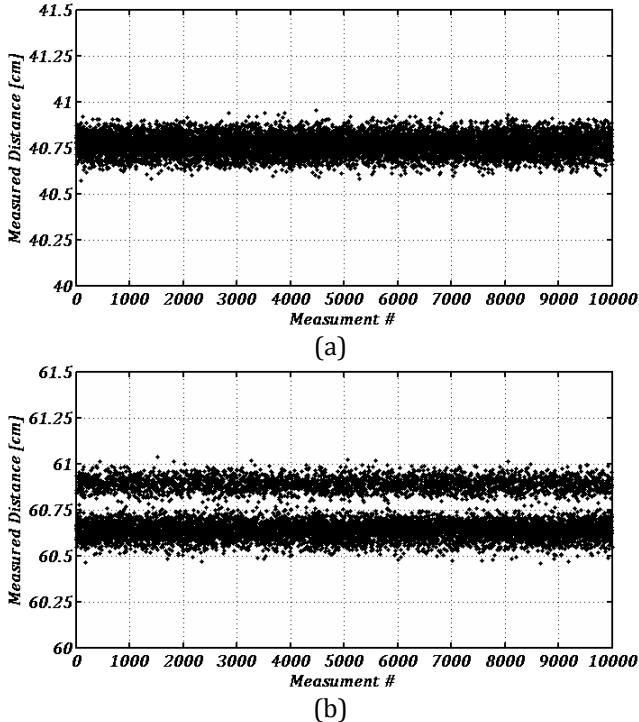


Figure 3.32. Plot of measured distance for 10000 successive TOF measurements, using the fully-integrated TCSPC imager: (a) with ground truth at 40cm and (b) with ground truth at 60cm. Note the larger than expected spreading in (b) due to quantization errors in the histogram.

While the plot of Figure 3.32 (a) exhibits a single spreading component, the measurements of Figure 3.32 (b) reveals a clear quantization error, due to finite histogram resolution ΔT . This error may be explained as follows: depending on the actual distance of the reference panel, the sampled signal peak may be located exactly between two bins in the histogram. As a result, the histogram interval around the maximum value, used to compute the peak centroid position, suffers from a timing error of one full bin width, i.e. ΔT . Note that, in

this setup, the ΔT duration leads to a distance error of 1.5cm. The computation of TOF thus suffers from quantization errors, which were not included in the model of Section 3.3. However, as expected, the computation of the peak's centroid position leads to a net improvement in quantization errors. Indeed, as can be seen in Figure 3.32 (b), the distance separation between the two clouds of points was approximately 0.3cm.

In order to further investigate the error distributions of each evaluated distance, normalized histograms based on the set of 10000 measurements were computed. These distributions are plotted in Figure 3.33 and Figure 3.34, for the whole evaluation range. In these Figures, the distributions were shown around each ground truth distance, which appears in the center of each plot. Furthermore, mean errors were not removed from the data set. A careful analysis of these distributions leads to interesting conclusions.

First, the occurrence of quantization errors among the distance is very frequent. In those cases, large random errors were obtained, as shown in Figure 3.31 (c). Otherwise, when the error distribution exhibited a single component, outstanding distance resolutions of the order of 1 millimeter were obtained.

Second, the topology adopted in the design of the integrated TDCs also played an important contribution in distance errors. As described in Section 3.5.3, the TDC is based on a multi-level interpolator. Moreover, it was strongly optimized in terms of miniaturization and measurement throughput. Non-idealities were therefore present during measurement, mainly caused by the inherent TDC topology, combined with signal crosstalk of neighbor TDCs. These non-idealities, which were characterized by means of DNL and INL in the previous Section, lead to errors that depend on the TOF to be measured. Distance errors, depicted in Figure 3.33 and Figure 3.34, are the result of a chain of processes, from photon detection to averaging computation. In particular, they involve linear and non-linear processing steps such as TDC conversion and histogram processing. Even linear computations, such as time-to-digital conversion, exhibited non-linear errors that are convolved with the remaining process. As a result, very complex mechanisms are involved in non-linearity and random distance errors.

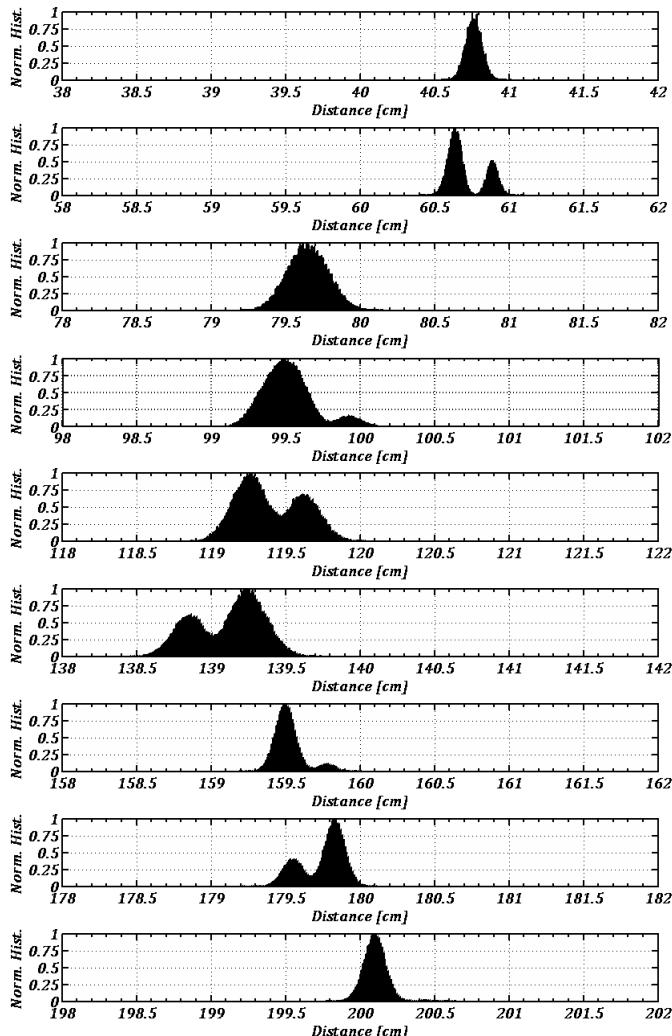


Figure 3.33. Normalized distribution of measured versus actual distances, for 10000 successive measurements: each plot, from top to bottom, corresponds to a measurement step from 40cm up to 200cm, in steps of 20cm. Note that mean errors were intentionally left in the normalized histogram. Moreover, ground truth distances coincide with the center of each plot.

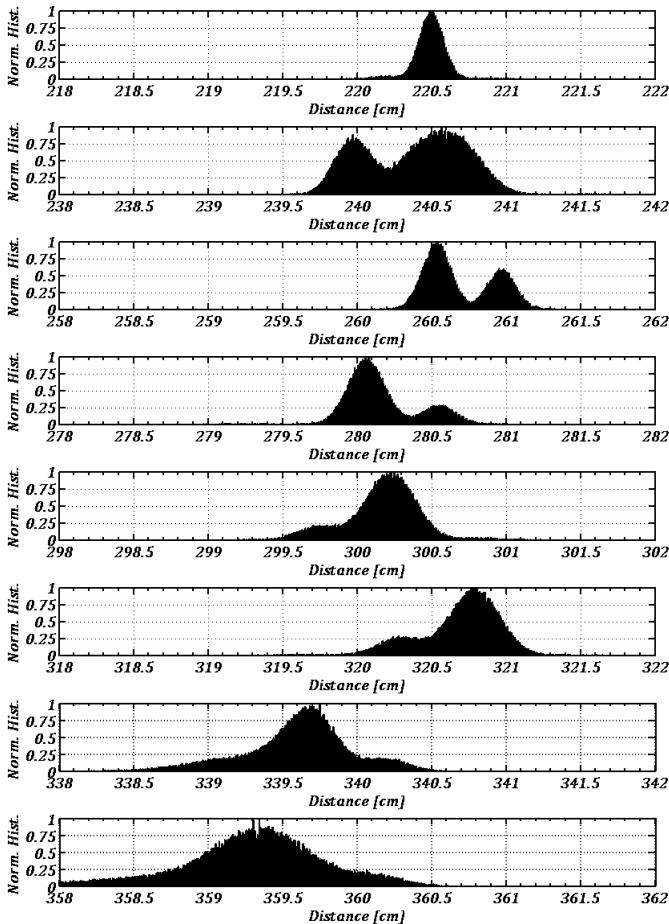


Figure 3.34. Normalized distribution of measured versus actual distances, for 10000 successive measurements: remaining measurements, from 220cm up to 360cm, in steps of 20cm.

It is therefore very difficult to analytically and accurately describe all these mechanisms. For example, despite a considerably higher SBR at 80cm than at 200cm, random errors at 80cm were larger than the errors at 2 meters. This unexpected result may be explained by the TDC interpolator. The measurements corresponding to 80cm have possibly occurred at a TDC range location that is close to the interface between fine and medium resolutions, see Section 3.5.3. On the other hand, when the distance measurements are all,

unintentionally, occurring in the middle of the fine delay line, high distance precision is obtained. Towards the end of the distance range, at 340cm and 360cm, larger dispersions are clearly visible. In those cases, photon-shot noise is likely to dominate over circuit jitter components. Indeed, only a few hundred signal photons are integrated in the histogram at those distances, as anticipated by S_R in the analytical model.

To conclude this Section, the fully-integrated rangefinder sensor was qualitatively evaluated by taking a depth snapshot of a human-size mannequin face. The model was placed at 1 meter from the sensor. In order for the light source to illuminate the model completely, a diffuser inducing a field-of-view of 30° was installed. Figure 3.35 shows the depth map of the model as well as the picture of the model, captured with a standard digital camera.

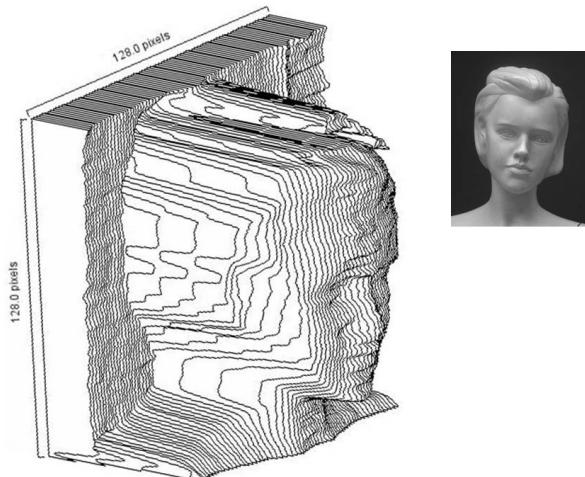


Figure 3.35. Experimental 3D image with model picture in inset. Measurement based on a target distance of 1 meter and on an illumination field-of-view of 30°. Remaining parameters kept unchanged.

In this depth map evaluation, some errors are introduced due to occlusion. Since the imaging lens and the illumination source were not concentric, a fraction of the mannequin head was not illuminated, thus generating inaccurate TOF evaluations. Note that the illumination was placed below the imaging lens.

3.7 EXPERIMENTAL RESULTS IN MEDIUM RANGE

In order to demonstrate the potential of CMOS SPADs and TCSPC for reliable depth sensing in applications involving longer distances, experimental distance evaluation in a range of several tens of meters is described in this Section. Furthermore, in order to verify the applicability of this technology in outdoor conditions, experimental measurements were carried out in daylight.

3.7.1 ILLUMINATION SYSTEM

In addition to the challenges involved in sensing technology for TCSPC, another important issue in medium ranges is the design and operation of pulsed optical sources. While low-power picosecond laser sources are commercially available, when the illumination system requires more powerful devices, additional concerns are involved. In particular, TCSPC requires very short optical pulses, which leads to extremely low optical duty cycles. When an optical mean power in the range of several Watts is required to satisfactorily deal with solar background irradiation, the optical source has to feature extremely high peak power of the order of several hundred Watts. In this context, the first question that arises is eye-safety compliance. In the next paragraphs, this issue is investigated and addressed. Furthermore, very high peak power poses extremely stringent requirements in the optical source, and associated electronics, in terms of rise and fall times. Although the design, prototyping, and characterization of high-frequency pulsed optical sources are beyond the scope of this thesis, a custom illumination system was designed, mostly due to the unavailability of suitable sources. The characteristics of this illumination prototype are described in end of this Sub-section.

In this Section, distance ranging is evaluated using hypothetical requirements for a specific application. These requirements are: a FOV of 20° , a distance range R_D of 80 meters, a frame rate of 10 fps, and reliability under 100klux of solar background illuminance. As described in Section 3.3, in range-finding TCSPC imagers, it is advantageous to operate with light pulses as short as possible. Extremely low duty cycles of the order of 1% or lower are therefore necessary. Taking into account very low duty cycles, in order to achieve the required mean power in outdoor condition, optical pulses with very high peak power are required. To mitigate this constraint, it is possible to increase the illumination repetition f_0 rate to its maximum value, for a given distance range R_D . The relationship between f_0 and R_D is given in Relation (3.6). The maximum f_0 is therefore 2MHz for 80 meters of range. Considering the hypothetical requirements and current detection parameters of CMOS SPADs, a careful

analysis of illumination power budget, based on the analytical model of Sections 3.3 and 3.4, leads to a mean optical power of 20W. These values, shown in Table 3.2, are therefore considered for eye-safety evaluation. Note that the hypothetical requirements are extremely stringent. Specifically, a FOV of 20° leads to an illuminated area of more than 600 square meters at 80 meters. Furthermore, different FOV aspect ratios may be utilized, under the same optical power budget, as long as the worst-case illuminated area stays constant.

Table 3.2. Application and optical source parameters assumed in the eye-safety evaluation.

Parameter	Symbol	Value	Unit
Maximum background illuminance	L_{BG}	100	klux
Maximum distance range	R_D	80	m
Image integration time	T	100	ms
Illumination central wavelength	λ_{SG}	850	nm
Illumination repetition rate	f_0	2	MHz
Illumination duty cycle		1	%
Illumination field-of-view	α_{FV}	20	°
Illumination mean optical power	\bar{P}_S	20	W
Approximate illumination peak power		2000	W

Eye-safety Evaluation

The IEC60825-1 standard [179] was assumed to guarantee strict user eye-safety in this theoretical evaluation. In particular, the illumination system should comply with the most restrictive classification of the standard, namely Class 1. According to the standard, there are a number of means to address eye-safety constraints under certain conditions. Here, we exploit the fact that an extended light source can safely provide more illumination power than a point light source. An extended light source is safer because it is imaged to the user's retina over a larger area, thus reducing the maximum local exposure. This condition is particularly true for wavelengths of interest in this thesis, i.e. red and near infrared. A point light source can be transformed into an extended light source by means of a diffuser. As a result, the calculation of the minimum area a diffuser needs to have, to meet the Class 1, is of interest. For the reasons described in Section 3.4.2, it is assumed that the diffuser is perfect. That is, that the light intensity after the diffuser is constant within the FOV.

An executable report has been created to size the minimum diameter that a circular diffuser has to have to comply with the Class 1 conditions, based on a symbolic mathematics tool. The result of this report, available in Appendix A, was independently validated by a specialized institute [180]. Among the three

conditions enforced by the standard, the average power limit was the most critical condition. This result states that, independently of the optical peak power, the required diffuser size is constant assuming a fixed mean power. The critical diffuser diameter, for the parameters of Table 3.2, was 2.8cm. This critical diffuser size appears to be acceptable in most applications. Furthermore, larger diameter sizes would allow some safety margin with respect to the standard.

Illumination System Characterization

An illumination prototype, based on an array of 16 laser diodes (LDs) and associated electronics, was designed and implemented. The maximum rating of power dissipation in component packages restricted the maximum achievable optical power considerably. In particular, the LD component had a constraint in terms of maximum repetition rate. In order to achieve the assumed frame rate of 2MHz, the LD array was operated in such way that only two lasers were active at a given period. As a result, while each individual LD was pulsed at a frame rate of 250kHz, the overall required repetition rate of 2MHz was achieved.

The illumination prototype was carefully characterized in the laboratory. In Table 3.3, a list of relevant illumination parameters is shown. Note that the parameters in the table include device-to-device timing and power variations.

Table 3.3. Illumination prototype parameters

Parameter	Symbol	Value	Unit
Illumination repetition rate	f_0	2	MHz
Illumination field-of-view	α_{rv}	5	°
Optical pulse rise time		3.7	ns
Optical pulse fall time		3.1	ns
Optical pulse duration (FWHM)		6.1	ns
Pulse half-width duration, approximated by a Gaussian shape pulse (to be used in the theoretical model)	σ_s	2.7	ns
Illumination duty cycle		1.22	%
Illumination mean optical power	\bar{P}_s	194	mW
Approximate illumination peak power		16	W

As can be seen in Table 3.3, while the achieved duty cycle was appropriate, the measured mean optical power was considerably restrictive. Furthermore, due to this limitation, a diffuser with 5° of FOV was utilized. Finally, Figure 3.36 shows the normalized optical power spectral density (PSD) of the illumination signal, measured with a spectrometer. The mean central emission of the prototype was at 857nm, whereas the spectrum FWHM was 6nm.

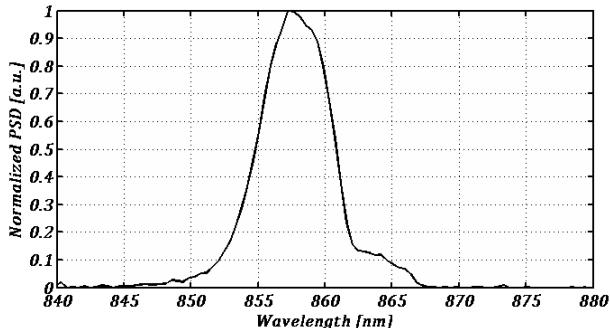


Figure 3.36. Normalized optical PSD of the illumination prototype, measured with a spectrometer.

3.7.2 RANGE-FINDING EVALUATION

Range-finding evaluation was carried out based on a setup involving a 32x32 SPAD imager implemented in 0.8 μ m CMOS [156], a commercially available TDC module, and the illumination prototype described in the previous paragraph. Distance performance was evaluated in outdoor conditions, but under moderate solar background light, due to the optical power limitation of the source. Background illuminance was approximately 4klux, instead of 100klux, which would be required in worst-case conditions. This experimental evaluation aimed at validating the theoretical model under severe conditions of *SBR*. System performance may be therefore evaluated under different conditions, by using the proposed model. Furthermore, since the illumination FOV was 5°, a direct scaling of required illumination power may be assumed, taking into account these results, to achieve similar performance in wider FOVs.

Measurements were carried out at the EPFL campus, on the roof of a building with an area of approximately 100x15 squared meters. A flat panel of approximately one square meter was covered with a pre-characterized textile, featuring a reflectivity of 65% at 850nm. The panel, used as reference, was moved from 10 to 70 meters, in steps of 10 meters. Evaluation at 80 meters was preempted by a technical limitation of the external TDC. At each distance, the reference was aligned and its distance with respect to the sensor was measured using an alternative method. Systematic errors in the measurement of the reference distance, using the alternative method, were estimated to be within +/− 15cm. Although the 32x32 SPAD array was utilized in intensity mode to take snapshots of the reference for alignment purposes, a single pixel in the center of the array was used in TCSPC mode.

In order to estimate range-finding performance based on the measurement conditions, the theoretical model was used. In Figure 3.37, estimated photon count rates, i.e. S_R , B_R , and $S_{R,MIN}$, are plotted as a function of distance. In the same figure, measured photon count rates were also plotted for comparison. Estimated count rates were based on the parameters of Table 3.4.

Table 3.4. System and environmental parameters used in the model for medium range evaluation.

Parameter	Symbol	Value	Unit
Background illuminance	L_{BG}	4	klux
Field-of-view	α_{FV}	5	°
Distance range	R_D	80	m
Illumination repetition rate	f_0	2	MHz
Illumination period	T_0	500	ns
Illumination average power	\bar{P}_S	50	mW
Illumination half pulse width	σ_S	2.7	ns
Histogram resolution	ΔT	2	ns
NBIF central wavelength	λ_F	860	nm
NBIF bandwidth	F_{BW}	12	nm
NBIF transmission in the pass-band	T_F	70	%
Imaging lens transmission	T_L	80	%
Imaging lens f-number	$f_\#$	f/1.4	m ⁻¹
Target object reflectivity	R_O	65	%
Integration time	T	100	ms
Pixel pitch	δ_{PX}	58	μm
Pixel fill factor	ϕ_{FF}	1.12	%
Photon detection probability at λ_L	PDP_{SG}	4.6	%
Measurement confidence probability for TOF evaluation	P	97.7	%
Half pulse width of signal pulse for centroid computation	N_{HW}	3	bins

Model parameters were mostly in agreement with actual values. However, the assumed mean illumination power \bar{P}_S exhibited a considerable mismatch with respect to the actual value measured in laboratory. In the model, it was assumed 50mW, instead of the actual 154mW. As depicted in Figure 3.37, based on these values, modeled and measured count rates matched accurately. After a careful post-measurement analysis, we concluded that mismatch was caused by the narrow-band interference filter. In particular, it somewhat blocked a significant fraction of the illumination signal. The main reason for this behavior was the placement of the filter between the imaging lens and the sensor array. In the lens cavity, optical rays exhibit very large incident angles with respect to the optical axis, due the small lens f-number. The spectral response of narrow-band interference filters shift as a function of the incident angle. The filter bandwidth was 12nm FWHM, centered at 860nm. The illumination spectrum width was 6nm FWHM, at 857nm. As a result, any shift in the spectral passband of the filter

leads to large signal losses. This issue leads to an evaluation of range-finding performance under extremely low SBR and S_R conditions. Note that when SBR is very low, higher S_R is required to reliably achieve TOF evaluations.

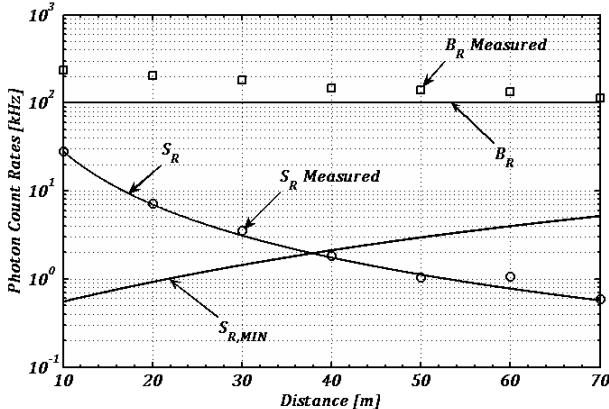


Figure 3.37. S_R , B_R , and $S_{R,MIN}$ as a function of distance, according to theoretical model. Measured S_R and B_R are also plotted as a function of distance for comparison.

In addition, modeled and measured B_R exhibited a moderate mismatch. Since a high number of TOF evaluations were performed, to allow a statistical analysis, the daylight condition varied slightly from the beginning to the end of range. Indeed, illuminance measurements performed on the target surface at each distance, confirmed this drift. Nonetheless, since S_R exhibited a much stronger dependence on distance, the slight drift of B_R had a negligible impact on the model predictions.

As can be seen in Figure 3.38, the curves of S_R and $S_{R,MIN}$ intersect at approximately 40 meters of distance. This intersection is a sign that TOF evaluations beyond 40 meters are no longer reliable, based on the assumed algorithm. In Figure 3.38, modeled SBR is plotted as a function of distance. It suggests for instance that at 40 meters, B_R is approximately 56 higher than S_R (-35dB).

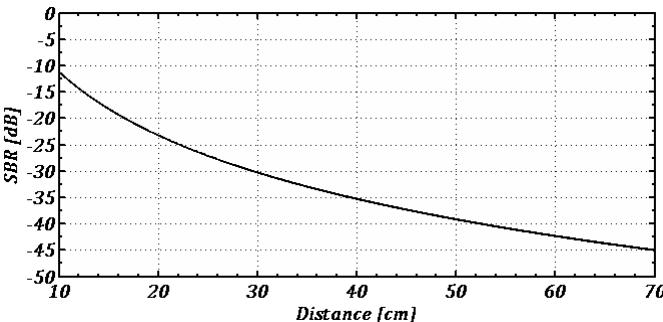


Figure 3.38. SBR as a function of distance, according to the theoretical model.

At each evaluation distance, a set of 1000 distance computations was recorded for the pixel under investigation. In Figure 3.39, a sample histogram for each distance step is plotted, based on the measurement conditions described previously. As can be seen, unlike in the short range conditions, the bandpass optical filter was not able to block all uncorrelated photons. However, as expected, background photons appeared uniformly distributed in the histograms. As can be seen in the plots, from 10 meters to 70 meters, the signal peak experiences a quadratic decay as a function of distance, due to the diffuser field-of-view of 5°. At 70 meters, the signal pulse is hardly identified, due to background noise. Note that at 70 meters, the signal optical power of 194mW (effectively 50mW) was used to illuminate an area of 29 square meters.

Note that the external TDC exhibited non-idealities at the end of the range, due to START signal dead time, as shown in Figure 3.39. This behavior preempted the investigation of rangefinder performance at 80 meters.

In Figure 3.40 (a), mean measured distance is plotted as a function of distance. Figure 3.40 (b) and (c) show mean error μ_{Error} and 1σ repeatability error σ_{Error} are plotted as a function of distance, respectively. As estimated by the analytical model, a large deviation occurred at long distances, mostly due to the limited signal count rate. For distances below 40 meters, outstanding performance was achieved. In this range, mean errors were practically within the systematic errors related to the alignment of the reference plane. Furthermore, repeatability errors were below 20cm in the same range.

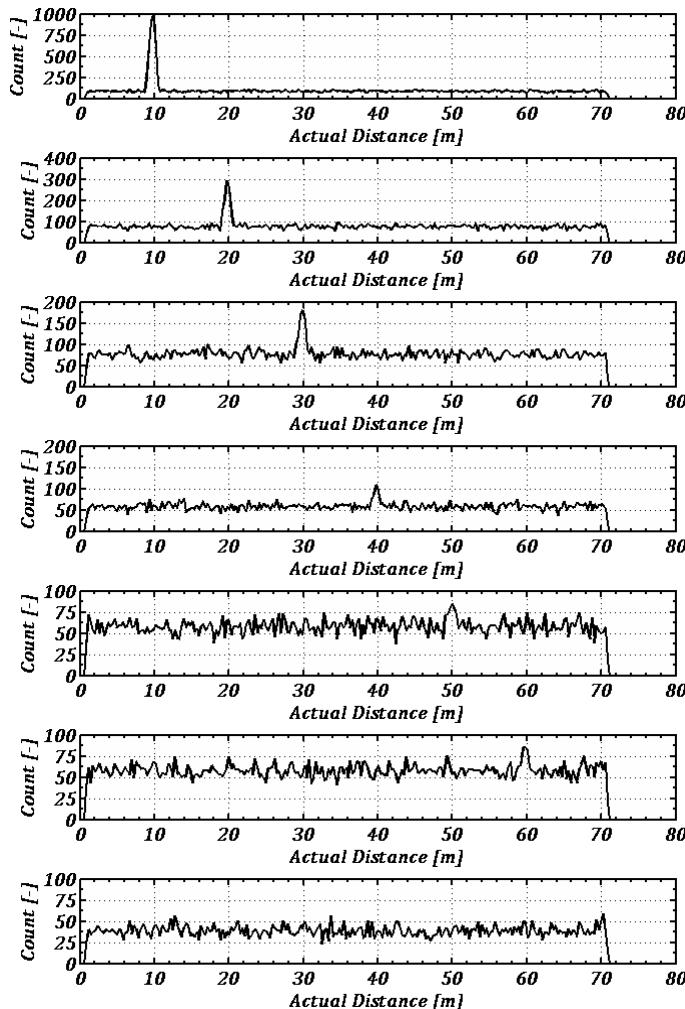


Figure 3.39. Histogram sample curves obtained when moving the target from 10m up to 70m with steps of 10m. Measurements using a 2MHz pulsed laser source emitting 194mW of optical power within a field-of-view of 5°, under 4klux of background illuminance. Each histogram was built based on an integration time of 100ms.

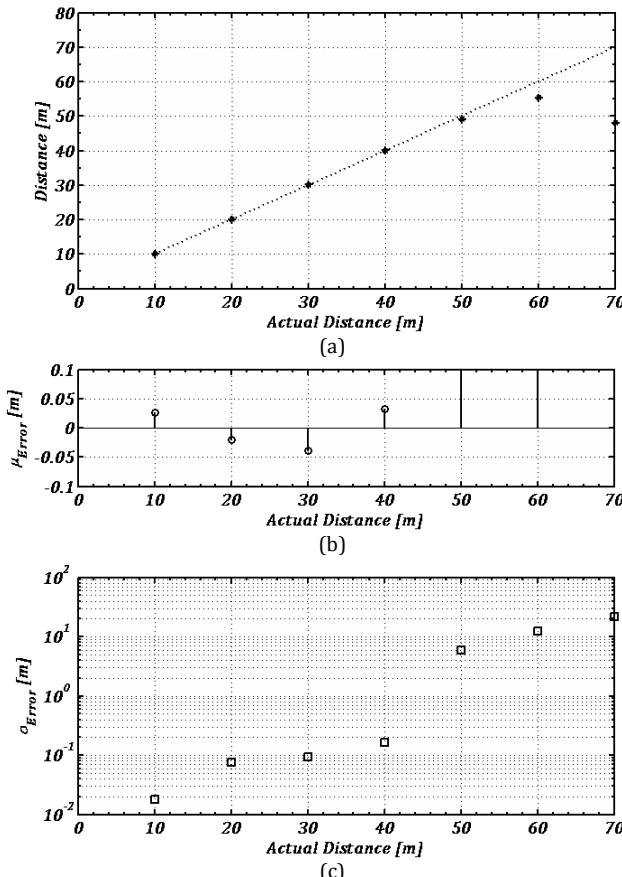


Figure 3.40. Rangefinding performance using a 2MHz pulsed laser source emitting 194mW of optical power within a field-of-view of 5°, under 4klux of background illuminance. Every distance measurement was based on an acquisition time of 100ms: (a) measured versus actual distance, (b) mean error with respect to ground truth, and (c) 1σ repeatability errors as a function of distance

In order to investigate the error distributions corresponding to 50m, 60m, and 70m in detail, recorded measured distances are plotted in Figure 3.41 (a), (b), and (c), respectively. As can be seen in Figure 3.41 (a), corresponding to 50 meters, although most of the TOF evaluations are accurate, a small number of measurements exhibited random behavior.

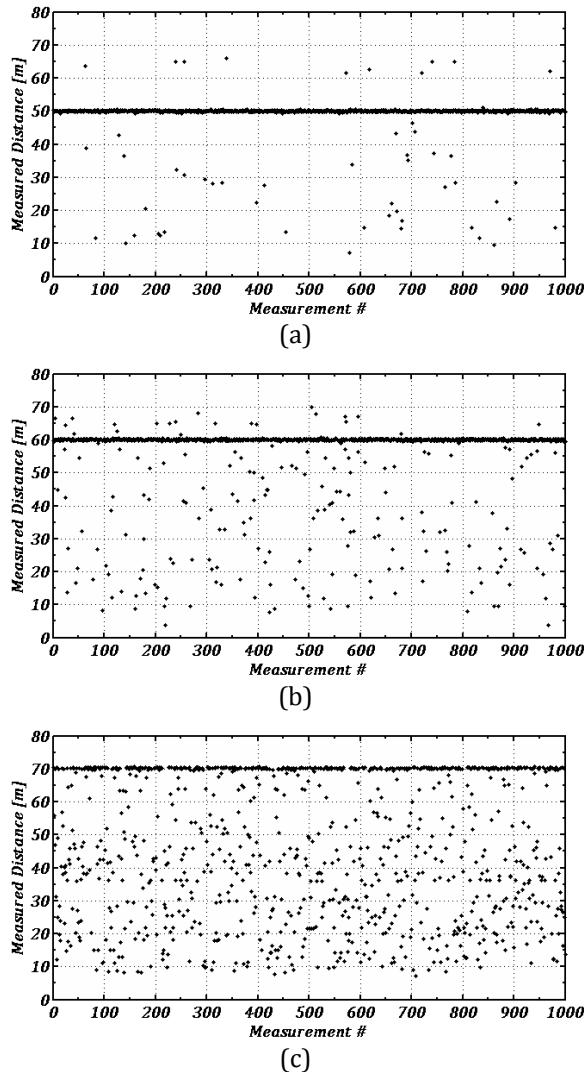


Figure 3.41. Plot of measured distance for 1000 successive TOF measurements, using the proposed TCSPC setup: with ground truth at (a) 50m, (b) 60m, and (c) at 70m.

This effect is a result of the proposed algorithm. In the assumed algorithm, the position of the signal pulse is first identified in the histogram by finding the bin

position corresponding to the maximum value. In accurate TOF evaluations, the maximum value corresponds to the signal pulse. However, when the signal pulse amplitude approaches the background noise level, there is a non-negligible probability that a noise peak is taken for the signal pulse. In that case, a random TOF is obtained, as shown in Figure 3.41 (a). Since the pulse amplitude decreases in the histogram with the distance, this effect occurs at higher frequencies in long distances, as shown in Figure 3.41 (b) and (c). Nonetheless, an important number of accurate TOF evaluations are performed at 60 and 70 meters. In Figure 3.40 (a), since the mean distances over 1000 measurements were plotted, the random TOF computations impaired considerably the perceptible linearity of the sensor. However, if those random computations are removed by an additional processing, the sensor linearity would be fairly improved at distances longer than 40 meters. For example, the use of a spatial (inter-pixel) or temporal (inter-frame) median filter would lead to higher ranging performance. These filters are typically utilized in image processing, to remove “salt and pepper” noise.

In order to quantitatively evaluate the probability of successful TOF detection, i.e. measurement reliability, the number of distance computations that occurred within a range of $\pm 50\text{cm}$ from the ground truth was computed for each distance. The ratio of this number to the total number of measurements leads to the probability of obtaining a reliable measurement, shown in Figure 3.42 as a function of the distance.

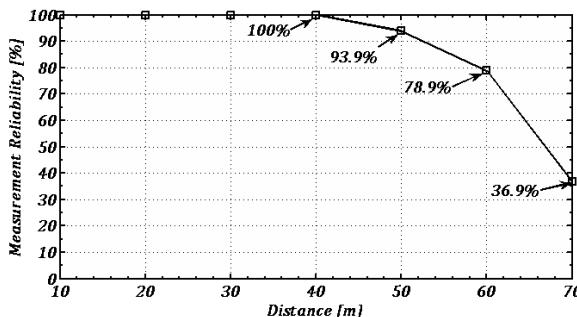


Figure 3.42. Plot of measurement reliability as a function of distance. Successful detections were assumed to be within an absolute error of $\pm 50\text{cm}$ with respect to ground truth.

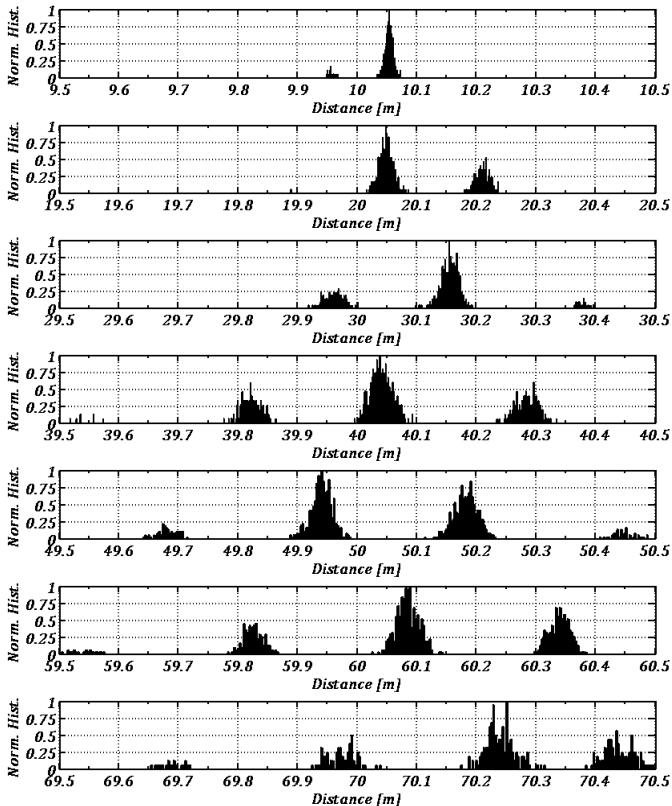


Figure 3.43. Normalized distribution of measured versus actual distances, for 1000 successive measurements: each plot, from top to bottom, corresponds to a measurement step from 10m up to 70m, in steps of 10m. Note that mean errors were intentionally left in the normalized histogram. Moreover, ground truth distances coincide with the center of each plot.

As shown in Figure 3.42, up to 40 meters, 100% of the measurements lead to successful detection. At 60 meters, almost 80% of the detections were accurate. Taking into account the optical source power, illumination field of view, and the background illuminance of 4klux, these results are among the best ever published.

To observe the distribution of TOF errors around actual distances, measured distance distributions are plotted in Figure 3.43. Similarly to the short range investigation, TOF measurements exhibited clear quantization errors. This

behavior was not surprising due to a histogram resolution ΔT of 2ns. This value was chosen in accordance with to theoretical model, so as to decrease the constraint on $S_{R,MIN}$, thus leading to higher successful measurements even at higher distances. Low histogram resolution is also beneficial in decreasing the required memory size for histogram storage and processing. This issue is a potential problem in medium to long distance ranges. Furthermore, the effects in histogram processing, described in the short range scenario, are also relevant to the results in this section. Towards the end of the measured range, additional errors were caused due to the limited reference height. Indeed, at 70 meters, the illumination spot diameter was approximately 6 meters. Since the reference height was approximately 1.5 meters, multiple reflections between ground and reference were expected to impair distance measurements.

3.8 CONCLUSION

In this Chapter, the principle and implementation of solid-state optical TOF 3D image sensors based on TCSPC have been described. In TCSPC, signal acquisition and processing differs considerably from remaining methods in solid-state depth sensing. In particular, some operating conditions should be respected to prevent signal distortion. However, when these conditions are fulfilled, TCSPC leads to superior performance that is unsurpassed by other methods, under the same illumination power budget. Conveniently, when the sensor is operated in environments that enforce very low signal-to-background ratio, these conditions are automatically fulfilled. Furthermore, extremely high background light rejection is achieved in TCSPC thanks to the confinement of signal photons in very small time durations, when compared to the illumination period. While these photons lead to sharp pulses in a TCSPC histogram, background photons are uniformly distributed over the full histogram length. Consequently, the signal pulse is typically discriminated from the background. In addition, it is possible to compute the timing position of the signal pulse, i.e. time-of-flight, virtually free from background fluctuations. To the best of our knowledge, this possibility is unique of TCSPC.

In Sections 3.3 and 3.4, an analytical model for the evaluation of ranging performance has been introduced. The model aims at validating design parameters in the context of reliable TOF evaluations, thus helping in the design of device, circuit, and system components. Since the model is analytical, it helps designers to understand typical compromises in TCSPC. It also enables theoretical, yet realistic, feasibility investigations of the TCSPC approach in different environmental conditions. Since the proposed evaluation methodology

is based on photon counting rates, a model that relates environmental, device, circuit, and system parameters to actual photon counting rates is described in Section 3.4. Note that photon count rate is a typical and convenient measure in the field of single-photon detection.

In Section 3.5, the first fully integrated system for single-photon time-of-flight evaluation has been introduced in CMOS. The IC comprises an array of 128x128 single-photon avalanche diodes with active recharge, a bank of 32 independent time-to-digital converters, and a 7.68Gbps readout system. The 32 converters are shared among 128 detectors in a row using an event-driven readout scheme. Furthermore, a continuously operating calibration scheme ensures an individual resolution of 97ps within 100ns of range and over a wide range of temperatures. The bank of converters can perform a total of 320 million single-photon time-of-flight evaluations per second. Other applications that may directly benefit from this design, beyond optical rangefinding, include fast fluorescence lifetime imaging, particle imaging velocimetry, fluorescence correlation spectroscopy, etc.

Experimental range-finding investigation has been carried out in two evaluation scenarios: (a) in a short-range, and (b) in a medium range, under 4 klux of background light.

Given that the fully-integrated TCSPC imager was specifically designed to operate at 40MHz, it has been used in the short-range scenario. This design has allowed depth map reconstructions of 3D scenes with millimeter precisions, in extremely low signal exposure. Quantitative ranging performance has been evaluated using a 1mW laser source and a field-of-view of 5°. The reference panel was simultaneously illuminated by 150lux of constant background light. Accurate distance measurements were repeatedly achieved based on a short integration time of 50ms even when signal photon count rates as low as a few thousand photons per second were available. The maximum non-linearity in distance measurement was 9 millimeters over the full measurement range. Time-varying uncertainty at the farthest distance was 5.2 millimeters. At that distance, the repeatability error distribution showed a shot-noise limited behavior. It may be concluded that, based on the proposed SPAD array design, time-varying uncertainty has limited improvement potential under the same illumination power budget. Note however that reliable TOF evaluations were repeatedly achieved based on a signal exposure as low as a few hundred photons. This signal exposure is typically challenged by noise levels in most conventional photodetectors in CCD and CMOS technologies. To the best of our knowledge, this performance was the best ever achieved by a solid-state 3D imager sensor in the art. Similarly, non-linearity errors were superior to values

reported in the art for 3D image sensors. Nonetheless, these errors may be further reduced with lower TDC non-linearity and/or with parameterized non-linearity compensation.

In the medium range scenario, an experimental range-finding evaluation was carried out using a setup involving the 32x32 SPAD imager implemented in 0.8 μ m CMOS and introduced in Chapter 2. A commercially available TDC module and the custom designed illumination system completed the setup. Distance performance was evaluated in outdoor conditions, but under moderate solar background light, due to the optical power limitation of the source. While the background illuminance was approximately 4klux, the illumination mean power was 194mW and its field-of-view was 5°. Environmental conditions as well as system component parameters were introduced in the theoretical model to estimate performance. While mismatches occurred in the modeling of optical components, when actual photon count rates were projected, the model accurately anticipated system performance. Furthermore, the mismatches in the modeling of optical components were explained by the choice of unsuitable narrow-band optical filter bandwidth. Statistical analysis of the experimental results showed that, despite the severe conditions in terms of signal-to-background ratio, 100% of time-of-flight evaluations were accurate within a range of 40 meters. In this range, non-linearity errors were within systematic errors in the alignment of the reference panel. 1 σ repeatability errors were below 20cm in the same range. Detailed statistical analysis of measured distances achieved at 50 meters, and beyond, demonstrated that a high percentage of time-of-flight evaluations were accurate. The probability of successful detection was computed based on a set of 1000 measurements. At 50 meters, 93.9% of distance evaluations were accurate. This number decreased to 78.9% and 36.9% at 60 and 70 meters, respectively. Taking into account actual signal-to-background ratios, measured on the reference surface, these results are outstanding. To the best of our knowledge, no existing solid-state 3D image sensors previously reported in the art achieved similar results.

While the ranging performance reported in this Chapter are excellent, TCSPC has some practical limitations. Its potential for some applications should be therefore slightly tempered. As described in Section 3.2 and illustrated in Section 3.5, the design of a fully integrated TCSPC microsystem involves relatively complex circuits. As a result, the implementation of large two-dimensional arrays of pixels that operate simultaneously is currently challenged by the current state of CMOS technology. On a high-resolution pixel array, the required silicon area required to implement the TDC and histogram memory would certainly lead to extremely low fill factors. In this context, the sharing of

circuit infrastructure among some pixels, such as in the design proposed here, by means for example of event-driven readout schemes, is highly encouraged. Alternatively, the applications that involve low FOV aspect ratios, yet requiring high ranging performance, are excellent candidates. A perfect example is the automotive field for the detection of vulnerable road users (see Chapter 1). In this application, low vertical field-of-view is acceptable, while wide horizontal field-of-view is typically necessary. Note that low vertical FOV is further justified if one takes into account possible multi-reflections from road to target. In this context, it is possible to design image sensors featuring a low number of rows, e.g. 8 or 16, and high horizontal resolution, e.g. 256 or 512 columns. Considering this aspect ratio, it is possible to move the required TCSPC pixel circuitry to outside the light sensing area, thus leading to high pixel fill factor. Such implementation would highly benefit from the results achieved in this thesis.

4. OPTICAL TIME-OF-FLIGHT 3D IMAGING BASED ON SPSD

4.1 INTRODUCTION

In this Chapter, an innovative approach is studied for the acquisition of TOF images based on a demodulation technique called single-photon synchronous detection (SPSD). This technique differs considerably from TCSPC and solves some of its practical limitations. In particular, it allows for TOF measurements based on single-photon detectors without the need for time-to-digital converters nor histogram storage and processing. As a result, large arrays of SPAD may be implemented, yet providing full parallelism in image acquisition, processing, and readout.

The SPSD technique is introduced in Section 4.2. Performance modeling and evaluation of TOF rangefinders based on SPSD are described in Section 4.3. In Section 4.4, the first fully-integrated 3D image sensor based on SPSD is introduced. A prototype of a real-time 3D camera based on SPSD is described in Section 4.5. Experimental validation and performance evaluation is reported in 4.6. Finally, this Chapter is concluded in Section 4.7.

4.2 SPSD PRINCIPLE AND ASSUMPTIONS

The SPSD technique, introduced in [181], involves the demodulation of the phase shift an optical signal experiences when travelling from source to target and back to the sensor. Figure 4.1 shows the basic setup involved in a solid-state 3D imager based on SPSD. It consists of a periodically modulated light source, typically a sine wave, used to illuminate the target, imaging optics, a band-pass filter used to block background light, and a single-photon image sensor. This sensor comprises an array of single-photon detector pixels with pixel-level demodulation circuitry. Depending upon the waveform used, the optical signal spectrum may contain more than a single harmonic. However, similarly to photogate-based lock-in pixels described in Chapter 1, the SPSD rangefinder determines the phase only of the first harmonic. Back-reflected photons are

imaged onto the array of pixels at the focal plane. In each pixel, a SPSD circuit performs the demodulation of the phase φ , amplitude A , and offset B of the incident optical signal. Once the phase is determined, the distance z to each point of the scene, is computed as

$$z = \frac{c}{2} \frac{\varphi}{2\pi f_0}, \quad (4.1)$$

where f_0 is the fundamental frequency.

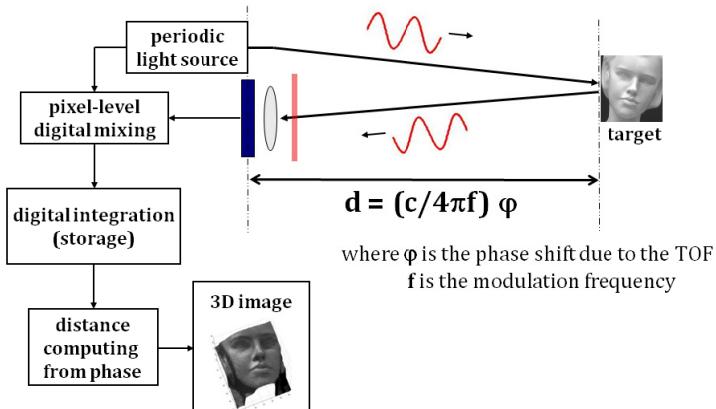


Figure 4.1. Solid-state 3D imaging setup based on the SPSD technique.

The general in-pixel SPSD demodulation circuit is shown in Figure 4.2. The SPSD demodulation principle involves a synchronous detection and sampling of the incident photon flux. A single-photon detector and its front-end circuit are used to detect individual photons. At the output of the front-end circuit, discrete digital pulses resulting from photon detections are accurately generated. The single-photon detector output is connected to the input of a demultiplexer whose selection input S is digitally modulated in synchronization with a reference signal, which is shared by the illumination source. The demultiplexer has N_C outputs, one being connected to an independent edge-sensitive counter, thus forming a bank of N_C counters.

Signal demodulation is achieved by dividing each period of the reference signal in N_C uniform time intervals, namely $\Delta T_0, \Delta T_1, \dots, \Delta T_{N_C-1}$, and modulating S such that

$$S = \begin{cases} 0 & \text{when } t \in \Delta T_0 \\ 1 & \text{when } t \in \Delta T_1 \\ 2 & \text{when } t \in \Delta T_2 \\ \vdots \\ N_C - 1 & \text{when } t \in \Delta T_{N_C-1} \end{cases}. \quad (4.2)$$

As a result, photons detected during each time interval, e.g. ΔT_1 , are counted in the corresponding counter, i.e. C_1 . Since the SPSD demodulation is carried out over a high number of illumination cycles, the distribution of counter contents converges to the incident optical signal waveform over one period.

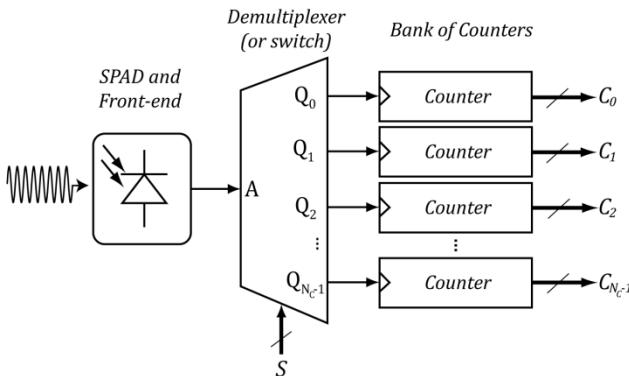


Figure 4.2. General in-pixel SPSD demodulation circuit. A single-photon detector and its front-end circuit are used to detect individual photons. The output of the single-photon detector is connected to a N_C -to-1 digital demultiplexer, which, in turn, connects the single-photon signal synchronously with S to one of the edge-sensitive counters.

An illustration of the SPSD principle is shown in Figure 4.3, wherein N_C was assumed to be 4, and the illumination was based on a sinusoidal wave, i.e.

$$P_S(t) = A\cos(2\pi f_0 t - \varphi) + B. \quad (4.3)$$

In the picture, $P_S(t)$ represents the incident optical power that falls onto the single-photon detector of a given pixel. In the picture, 20 signal periods are illustrated. Note that the sinusoidal waveform of P_S is an ideal representation. In fact, since the incident optical power that enters the active area of the single-photon detector is extremely low, P_S is a collection of discrete photonic events that follows a sinusoidal distribution in time. In most cases, many periods of P_S do not contain any photon and periods with more than a single photon are rare.

One may characterize the envelop (or mean value) of P_S by an amplitude A , an offset B , and a phase φ , as shown in Figure 4.3. In the picture, the first photon detection occurs in the second period. Since state of S at that moment was 1, the C_1 counter was incremented. Similarly, on the second photon detection, the state of S was 0, thus resulting in a C_0 increment. Note that as the number of photon detection increases, the distribution of the counter values tends towards the mean value of P_S over one period, shown in the right-hand side of the picture.

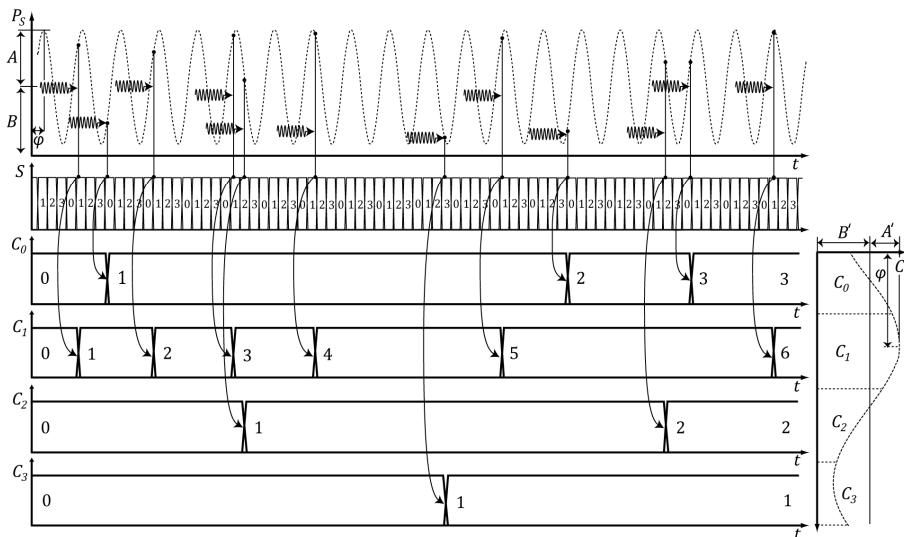


Figure 4.3. Waveform of SPSD demodulation signals for a pixel based on four counters. P_s is the incident optical power, which is proportional to the incident photon flux, S is the demultiplexer selection signal, and C_0 , C_1 , C_2 , and C_3 are the counter values. Note that, over time, the distribution of the counter values converges to the incident power waveform within one period.

Thanks to the digital nature of single-photon detectors, phase demodulation may be achieved in fully digital implementations. Note that the demultiplexer may be replaced by a simple switch. However, taking into account the nature of the signals involved, a fully digital implementation leads to the best performance/silicon area utilization ratio. Furthermore, in the illustration of the SPSD principle, no assumption was made with respect to the depth of the counters. For the moment, ideal counters are assumed for the calculation of A , B , and φ . Non-idealities are then introduced later.

Although the SPSD implementation principle differs considerably from the CCD/CMOS lock-in pixels based on photogates, once the counter values $\{C_k\}$ are acquired, subsequent calculations of A , B , and φ in SPSD are equivalent to the lock-in approach. Unlike CCD/CMOS lock-in pixels however, in SPSD, the acquisition of $\{C_k\}$ values does not involve any analog processing or analog-to-digital conversion, thus leading to a virtually noise free detection and demodulation approach. Furthermore, as described in [181], the SPSD in-pixel circuit is amenable to implementing multiple access techniques to enable coexisting rangefinding systems in close proximity based, for example, on FDMA, CDMA, etc.

Referring back to the general case of N_c sample values and any periodic illumination waveform, since the SPSD demodulation may be seen as a synchronous sampling of the incident optical signal, one may utilize the theory of discrete Fourier transform (DFT) to relate the quantities of interest to the sampled values [35]. In particular, it is assumed for the moment that the number of samples N_c is high enough to satisfy the Nyquist-Shannon sampling theorem for a given illumination waveform. The sampled sequence F_n is

$$F_n = \frac{1}{N_c} \sum_{k=0}^{N_c-1} C_k e^{-i2\pi \frac{n}{N_c} k}. \quad (4.4)$$

Since we want to determine the phase of the first harmonic only, we can determine A' , B' , and φ in the case of $n = 1$, i.e.

$$A' = \frac{2}{N_c} \sqrt{\left(\sum_{k=0}^{N_c-1} C_k \cos\left(2\pi \frac{k}{N_c}\right) \right)^2 + \left(\sum_{k=0}^{N_c-1} C_k \sin\left(2\pi \frac{k}{N_c}\right) \right)^2}, \quad (4.5)$$

$$B' = \frac{1}{N_c} \sum_{k=0}^{N_c-1} C_k, \quad (4.6)$$

$$\varphi = -\arctan \left\{ \frac{\sum_{k=0}^{N_c-1} C_k \sin\left(2\pi \frac{k}{N_c}\right)}{\sum_{k=0}^{N_c-1} C_k \cos\left(2\pi \frac{k}{N_c}\right)} \right\}, \quad (4.7)$$

where A' and B' are respectively the amplitude and offset of the sampled waveform, shown on the right-hand side of Figure 4.3. Since the SPSD sampling

intervals ΔT_i are not infinitesimally small, the amplitude to offset ratio between the incident optical signal and the sampled waveform is not constant. Assuming that the sampling intervals ΔT_i are of constant duration ΔT , then the coefficients $\{C_k\}$ are expressed by a convolution of the incident optical power P_S with a sampling window (*rect* function) over a high number of illumination periods, i.e.

$$C_k = \frac{\eta_{SG} \lambda_{SG} T}{hc} \left\{ P_S(t) * \text{rect} \left(\frac{t - k \frac{T_0}{N_C}}{\Delta t} \right) \right\}, \quad (4.8)$$

where T is the integration time, λ_{SG} is the illumination signal wavelength, and η_{SG} is the PDP at λ_{SG} . In Equation (4.8), h and c are respectively Plank's constant and the speed of light. Similarly to CCD lock-in pixels, the SPSD demodulation process is characterized by an amplitude demodulation efficiency lower than unity, due to the limited bandwidth involved (i.e. finite ΔT). This demodulation efficiency, known as demodulation contrast, is given by

$$c_D = \text{sinc} \left(\pi \frac{\Delta T}{T_0} \right), \quad (4.9)$$

due to the convolution of a *rect* function with the input signal. Note that, unlike the photogate lock-in pixels described in Chapter 1, the theoretical demodulation contrast c_D is not limited in SPSD by pixel non-idealities. The reason for this is that SPADs feature a timing resolution of a few tens of picoseconds, thus the theoretical (sharp) definition of sampling windows applies in practice. In other words, the sampling operating is virtually ideal.

Based on these equations, it is possible to relate the incident optical amplitude and offset, for a sinusoidal illumination, as

$$A = \frac{hc}{\eta_{SG} \lambda_{SG} T} \frac{A'}{\text{sinc} \left(\pi \frac{\Delta T}{T_0} \right)}, \quad (4.10)$$

$$B = \frac{hc}{\eta_{SG} \lambda_{SG} T} \text{sinc} \left(\pi \frac{\Delta T}{T_0} \right) B'. \quad (4.11)$$

An important particular case, illustrated in Figure 4.3, is the SPSD implementation based on four samples ($N_C = 4$). In this case, Equations (4.5)-(4.7) become

$$A' = \frac{1}{2} \sqrt{(C_3 - C_1)^2 + (C_0 - C_2)^2}, \quad (4.12)$$

$$B' = \frac{1}{4} (C_0 + C_1 + C_2 + C_3), \quad (4.13)$$

$$\varphi = \arctan \left\{ \frac{C_3 - C_1}{C_0 - C_2} \right\}, \quad (4.14)$$

respectively. These Equations were originally described in [10], for CCD lock-in pixels. The four sample case is a very good compromise between high demodulation contrast and pixel complexity. Note that, for a pure sine wave illumination, it is required that at least three samples be implemented to respect the Nyquist-Shannon sampling theorem. In the case of four samples, and assuming a ΔT of $T_0/4$ as illustrated in Figure 4.3, the demodulation contrast c_D is 90%. For example, if eight samples are implemented ($N_C = 8$) with ΔT of $T_0/8$, then c_D would be 97.4%. That is, an efficiency improvement of 7% at the price of 100% of increase in circuit area.

Furthermore, although φ is the most important result for a TOF rangefinder, A and B also carry interesting pixel information. A may be conveniently used to determine whether a pixel signal, in a given acquisition frame, has a sufficiently strong amplitude, so as to be considered as a reliable measurement. Indeed, pixel signals with negligible amplitude could be simply disregarded. B may also be used to compute intensity images.

4.3 MODELING OF PERFORMANCE BASED ON SPSD

Although the SPSD implementation is very different from existing photogate lock-in image sensors, the computation of distance based on $\{C_k\}$ values is the same as in lock-in rangefinders. As a result, the analysis of distance errors in SPSD may be carried out by adapting existing models of lock-in pixels, reported for example in [33, 35]. However, thanks to the digital nature of the SPSD implementation, a number of non-idealities existing in lock-in pixels, such as readout noise and gain non-linearity, can be ignored.

In order to simulate the performance of SPSD implementations in the context of background light, the model of environment conditions and optical components, described in Section 3.4, may be used with the theory in this Chapter as well.

4.3.1 DISTANCE RESOLUTION

Range-finding precision in SPSD may be derived from Equation (4.1) to depend on design or environment parameters. The r.m.s. distance resolution σ_{Error} is expressed as

$$\sigma_{Error} = \frac{R_D}{2\pi} \sigma_\varphi, \quad (4.15)$$

where σ_φ is the 1σ phase error and R_D is the non-ambiguous distance range, i.e.

$$R_D = \frac{c}{2f_0}. \quad (4.16)$$

In [35], an expression for σ_φ was determined by propagating the errors of the sample values $\{C_k\}$ into φ , in Equation (4.7) [35], i.e.

$$\sigma_\varphi^2 = \sum_{k=0}^{N_C-1} \left| \frac{\partial \varphi}{\partial C_k} \right|^2 \sigma_{C_k}^2, \quad (4.17)$$

where $\sigma_{C_k}^2$ is the variance of C_k . Note that $\{C_k\}$ are random variables that are affected by a number of noise sources. In SPSD, since the integration and readout of $\{C_k\}$ are virtually noiseless, the main noise contributions are photon shot noise and DCR. Furthermore, given that the number of photon detections and the number of dark counts integrated over time are well approximated by Poisson variables, $\{C_k\}$ may also be assumed to be Poisson. Consequently, the variance of C_k is given by their mean value.

In the case of four samples per period, i.e. $N_C = 4$, Equation (4.15) may be rewritten as a function of A' and B' as

$$\sigma_{Error} = \frac{R_D}{\sqrt{8}\pi} \frac{\sqrt{B'}}{A'}. \quad (4.18)$$

In (4.18), the quantities A' and B' may be computed as a function of the demodulation contrast c_D , integration time T , and photon counting rates S_R and B_R , as introduced in Section 3.4. S_R and B_R are the mean photon counting rates owing to the illumination and background light, respectively. Thus,

$$\sigma_{Error} = \frac{R_D}{\sqrt{8}\pi c_D c_S} \frac{1}{\sqrt{T \cdot SBR \cdot S_R}}, \quad (4.19)$$

where c_s is the signal modulation contrast due to non-idealities in the illumination source. In (4.19), the contribution of DCR was conveniently ignored as it is typically negligible compared to the background count rate. In situations where this assumption does not hold, a general expression for the distance resolution is given by

$$\sigma_{Error} = \frac{R_D}{\sqrt{8\pi c_D c_{SSR}}} \sqrt{\frac{B_R + DCR}{T}}. \quad (4.20)$$

4.3.2 NON-LINEARITY ERRORS

Similarly to photogate lock-in pixels, SPSD may also suffer from non-linearity errors. These errors are mostly due to harmonic distortion in the illumination source, which cause aliasing errors, as described in the next paragraph. However, since neither analog processing nor conversion is required in SPAD image sensors, typical errors that cause non-linearity in analog signals do not affect the SPSD demodulation. Examples of analog non-idealities are gain non-linearity, gain non-uniformity in pixels, ADC non-linearity, full-well saturation, etc. In SPSD, under some conditions of non-saturation, the demodulation process is nearly ideal. Saturation effects in SPSD exist though and may be divided in three categories: (i) saturation due to limited integrator capacity, (ii) saturation due to SPAD dead time, and (iii) saturation due to pile-up effect. The saturation effect due to limited integrator capacity is effectively addressed by design techniques. Note that the equivalent in SPSD of photogate full-well capacity is simply counter depth. Counters are typically digital circuits and their depth (capacity) may be increased easily using standard design techniques. Moreover, means exist to prevent counters to saturate, as described in Section 4.4. The effect of dead time is related to imager dynamic range and it is described in Section 4.3.3. Finally, the effects of pile-up are described in this Section succinctly, after a review of the effects of aliasing in SPSD.

Aliasing Errors

The sampling nature of SPSD demodulation inherently presupposes incident optical signals with restricted bandwidth. This is particularly true if one takes into account the lack of low-pass filter prior to the sampling operation. Consequently, aliasing issues are dealt with during the design of the illumination system.

Assuming that a perfect sinusoidal illumination system is utilized, more than two samples per period are required in SPSD to prevent aliasing, according to

Nyquist-Shannon sampling theorem. Typically, a SPSD pixel may provide four $\{C_k\}$ values, i.e. four samples per period. As a result, no aliasing effects occur and the measurement characteristic is perfectly linear. However, in practice, it is difficult to build illumination systems that exhibit only the fundamental frequency. The main reason for this difficulty is the linearity requirement in LED or LD drivers, which in turn, call for high-power consumption. In order to build low-cost and low-power arrays of LED/LD drivers, typically non-linear circuit topologies such as push-pull amplifiers are utilized. As a result, the illumination signals usually exhibit harmonic distortions that are sampled by the SPSD principle, thus leading to non-linearity errors due to aliasing. Note that this effect also occurs in lock-in pixels.

Fortunately, thanks to the availability of four (or more) $\{C_k\}$ per pixel, these errors are partially mitigated. A quantitative analysis of aliasing effects for a number of potential illumination waveforms are presented by Lange in reference [35]. In Table 4.1, a summary of the maximum non-linearity errors for typical illumination waveform is shown.

The results in Table 4.1 are based on $N_c = 4$ and $\Delta T = T_0/2$. As can be seen, the demodulation based on four samples is affected with non-linearity errors as high as 5.4% when the 3rd harmonic is present in the illumination signal. However, as demonstrated in [35], and is noticeable in the table, even harmonics do not impair linearity at all. This is explained by the sampling window duration, ΔT , which plays an important role in reducing the effects of aliasing. The input signal is convolved with the sampling window, i.e. *rect* function. This operation implies that the optical signal is multiplied by a *sinc* function in the frequency domain, thus providing a frequency filtering action. Even harmonic frequencies are therefore eliminated because the *sinc* function is zero at those frequencies. Furthermore, since the *sinc* function vanishes for very high frequencies, it progressively attenuates odd harmonics, as can be seen in Table 4.1.

From an illumination system design standpoint, it would be convenient to utilize square wave illumination as digital drivers provide cheap and power-efficient solutions. Nonetheless, as can be seen in Table 4.1, such solution leads to non-linearity errors as high as 1.7%. A clear compromise should therefore be found in terms of harmonic distortion versus illumination power efficiency (and cost).

Table 4.1. Maximum non-linearity errors due to aliasing in the SPSD demodulation for a number of typical illumination waveforms.

Illumination Waveform	Harmonics	Maximum Non-Linearity Errors [%]
$\cos(\omega_0 t)$	1	0
$\cos^2(\omega_0 t)$	1	0
$\cos(\omega_0 t) + \cos(2\omega_0 t)$	1, 2	0
$\cos(\omega_0 t) + \cos(3\omega_0 t)$	1, 3	± 5.4
$\cos(\omega_0 t) + \cos(4\omega_0 t)$	1, 4	0
$\cos(\omega_0 t) + \cos(5\omega_0 t)$	1, 5	± 3.2
$\cos(\omega_0 t) + \cos(6\omega_0 t)$	1, 6	0
$\cos(\omega_0 t) + \cos(7\omega_0 t)$	1, 7	± 2.2
Square wave: $\sum_{m=0}^{\infty} \frac{\sin\left((2m+1)\omega_0\left(t + \frac{T_0}{4}\right)\right)}{2m+1}$	1, 3, 5, 7, ...	± 1.7

Pile-up Distortion

As described in Chapter 3, a very important assumption in TCSPC is that, on average, less than a photon should be detected per illumination cycle. When this condition is not met, the signal experiences distortion. This is commonly known as pile-up effect [56, 57]. It turns out that pile-up effects may also impair SPSD performance. In particular, as discussed in Section 3.3, the asymmetric distortions that the input signal experiences may impair measurement linearity. In Section 3.3, a parameter M was introduced to evaluate the pile-up effect. It is the mean number of photon detections per illumination period, i.e.

$$M = \frac{\frac{\lambda_{SG} \bar{P}_S}{hc} \eta_{SG}}{f_0}. \quad (4.21)$$

Pile-up distortion is noticeable when M is higher than 1. In short-range measurements, the modulation period T_0 is comparable or shorter than typical values of SPAD dead time, i.e. in the range of 20 to 40 nanoseconds. As a result, when M is close to 1, the SPAD image sensor would be at the limit of its saturation. This is as rather exceptional region to operate in and it would occur only when the target is extremely close to the image sensor or it induces specular reflection. In those cases, SBR and noise conditions are so favorable

that parameterized correction for pile-up would be effective. Note that only photons from the active illumination are taken into account in Equation (3.5).

Conversely, in long-range distances, T_0 is considerably longer than SPAD dead time. In those cases, when objects are placed close to the image sensor (or induce specular reflection), M is likely to be higher than 1. However, when T_0 is longer than the dead time, distortion effects that are visible in pulsed illuminations, such as in TCSPC, tend to be negligible in sinusoidal modulation. The main reason for this is that in a sinusoidal wave, photons are randomly detected over the full period. Although the typical dead time applies on each photon detection, when this process is averaged over a high number of cycles, the illumination waveform is eventually sampled uniformly over the period.

Independently of distance range, an effective solution to deal with pile-up is to prevent it by reducing the detector count rate, by controlling the SPAD's photon detection probability. This may be achieved electrically pixel-wise by adjusting the SPAD biasing. As a result, by insuring that the maximum count rate of a pixel is f_0 or a fraction of it, pile-up effects are eliminated. Note that pile-up effects only occur in pixels that exhibit high signal counting rate. Most of pixels exhibit signal counting rates that are much lower than f_0 . Those pixels benefit from the maximum detection efficiency.

4.3.3 DYNAMIC RANGE

The dynamic range (DR) is defined as the maximum output swing divided by the temporal noise in the dark. The output swing is limited at the low end by the product of DCR and the integration time. At the high end, the maximum count rate ever achieved by a SPAD is given by the inverse of dead time (DT). As a result, the maximum achieved output value is expressed by the ratio of the integration time to the dead time. Finally, in the dark, the temporal noise is expressed by time-varying component of DCR. Since the number of dark counts detected in the integration time follows a Poisson distribution [20], its noise component is easily determined. Thus,

$$DR = 20 \cdot \log \left[\sqrt{\frac{T}{DCR}} \left(\frac{1}{DT} - DCR \right) \right] \cong 20 \cdot \log \left(\sqrt{\frac{T}{DCR}} \cdot \frac{1}{DT} \right). \quad (4.22)$$

Note that a factor 20 was used to determine DR in decibels. This was intentionally done to allow DR of SPADs to be equitably compared to the dynamic range of CCD and CMOS image sensors.

When evaluating Equation (4.22) for typical parameters of CMOS SPAD in $0.35\mu\text{m}$, i.e. DCR of 600Hz and DT of 40ns, and for an integration time of 20ms, the dynamic range of a SPAD image sensor was estimated at 107dB. This result compares favorably with most of conventional CCD and CMOS image sensors.

4.4 FULLY-INTEGRATED SPSD 3D IMAGE SENSOR DESIGN

The design and characterization of the first fully-integrated TOF 3D image sensor based on SPSD [182] is described in this Section. For the first time, a large array of SPADs was implemented with fully parallel image acquisition and readout capability. The circuit delivers TOF measurements and intensity images at very high frame rates.

4.4.1 IMAGER ARCHITECTURE

The proposed image sensor takes advantage of a fully digital implementation, from photon detection to depth imaging. A simplified block diagram of the image sensor is shown in Figure 4.4. It consists of an array of 60×48 single-photon pixels, each one comprising its own SPSD demodulation circuit based on two 8-bit counters. The sensor also includes a bias generation circuit, a JTAG controller for testing/characterization purposes, and a fast readout circuit.

4.4.2 PIXEL CIRCUIT

Single-photon detection at pixel level with high timing resolution is achieved by means of a $0.35\mu\text{m}$ CMOS SPAD, introduced in Chapter 2, Section 2.6. Its front-end circuit involves 8 MOS transistors that perform passive quenching and active recharge, based on the dual-threshold scheme introduced in Section 2.4.2. Figure 4.5 shows the schematics of the complete pixel circuit. Active quenching is achieved by adequately choosing two different thresholds for the inverter and nor gates. At the inverter output, a digital inverted pulse reflects the detection of a photon. Its leading edge, i.e. high-to-low transition, accurately indicates the arrival time of the photon.

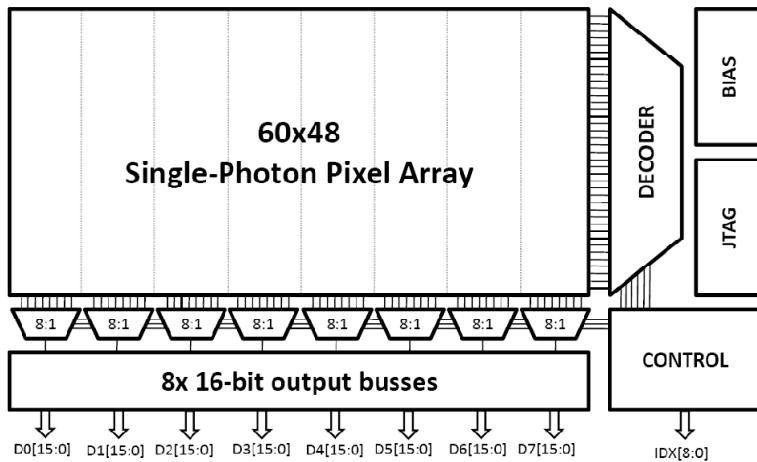


Figure 4.4. Block diagram of the image sensor. The sensor consists of a 60x48 pixel array, a JTAG controller, and a fast parallel readout circuitry. Every pixel has two 8-bit counters with overflow prevention mechanism.

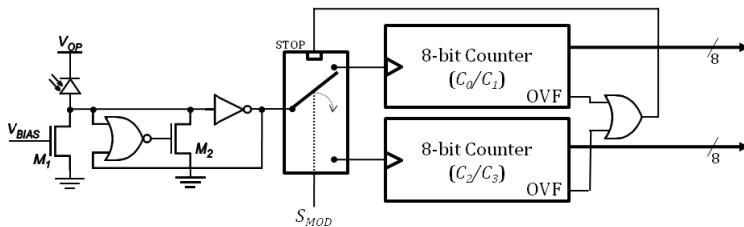


Figure 4.5. Pixel circuit comprising front-end, digital demodulation, and storage devices. Passive quenching and active recharge ensures higher dynamic range with little impact on pixel size. Fully digital implementation enables noise-free demodulation and readout.

Demodulation Circuit based on Two Counters

As shown in Figure 4.5, the demodulation circuit consists of a 2:1 multiplexer driven by a global signal S_{MOD} , synchronized with the light source, and two 8-bit counters. Each counter has a parallel tri-state output bus and a signal, namely OVF, indicating that the next increment would result in an overflow state. The OVF signals of both counters are combined, via an OR gate, to block the demultiplexer in a state in which neither counters could be incremented. Once

one counter reaches its maximum value, the pixel is blocked until the next readout and reset operation.

As described in Section 4.2, in order to unambiguously demodulate the signal phase, at least three counters are theoretically necessary. Practically, it is possible however to use only two counters and generate four samples. Figure 4.6 shows an example of illumination waveform and demodulation waveforms, as adopted in this design. The sensor operates in an interlaced detection scheme based on two acquisition phases. In the first acquisition phase, the pixel-level demultiplexer switches between two counters so as to generate two samples, C_0 and C_2 , corresponding to 0° and 180° phase with respect S_{MOD} . Once these two samples are acquired and readout, the sensor operates in the second acquisition phase, in which S_{MOD} is delayed by a quarter of period with respect to the reference signal. As a result, the same in-pixel counters are used to acquire samples C_1 and C_3 , corresponding to 90° and 270° of phase. Note that, although only two counters are used, the demodulation circuit does not miss any photon detection, unless the counter maximum value is reached.

In Equations (4.12)-(4.14), all the four samples $\{C_k\}$ are assumed to be acquired simultaneously. When the objects in the scene are not static, the acquisition of four samples based on two counters may suffer from higher motion artifacts, than it would have if the four $\{C_k\}$ were acquired simultaneously. In order to solve this problem, the acquisition of C_0/C_2 and of C_1/C_3 should be interlaced at frequency sufficiently high that moving objects appear static and thus affects all $\{C_k\}$ simultaneously. Note that conventional motion artifacts may persist depending on the actual frame rate achieved by the image sensor.

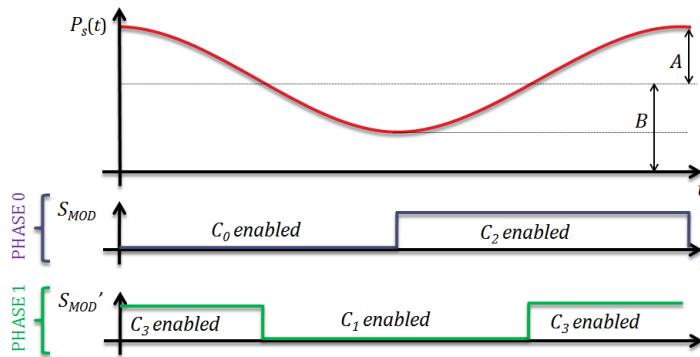


Figure 4.6. Incident optical signal and demodulation waveforms.

Based on this particular implementation, the sampling window duration ΔT was $T_0/2$, thus leading to a maximum demodulation contrast c_D of 64%, according to Equation (4.9). This implementation achieves an interesting compromise between pixel size (and fill factor) and overall demodulation contrast.

4.4.3 READOUT CIRCUIT

In this implementation, the readout circuit is based on a controller that allows the image sensor to operate autonomously, only requiring a clock signal. As shown in Figure 4.4, the pixel matrix area is divided in 8 blocks, each one consisting of 8x48 pixels and being handled by an independent readout block. The first and last readout blocks handle six active and two blocked columns each. A decoder, driven by the readout controller, selects a row. In that row, a pipelined sequence of readout and reset is achieved in the 8 blocks in parallel, thanks to the 8 digital output buses of 16 bits implemented. In each row, the readout sequence is operated as follows. The first pixels in all the 8 blocks, are read out, then, when the second group of pixels in all the blocks are read out, the first ones are simultaneously reset to zero. When the readout circuit finishes reading out the eighth pixel group in all the blocks, it spends one additional cycle to reset them, before switching to the next row. As a result, 9 cycles of clock are necessary to read out and reset a full row. Since the sensor comprises 48 rows, the full frame rolling readout requires exactly 432 cycles. Note that in each readout cycle a digital signal, $IDX[8:0]$, indicates the address of the pixels in the blocks that are currently being read out, and that each 16-bit bus outputs the contents of the two in-pixel counters simultaneously.

The readout circuit was designed to run at a clock frequency of up to 40MHz. At that frequency, a frame acquisition and readout takes 10.8 μ s. This time is short enough to be used in the interlaced acquisition of C_0/C_2 and C_1/C_3 , thus preventing motion artifacts. Moreover, since a pixel may be read out and reset in only 10.8 μ s, its 8-bit counters marginally reach their maximum values, assuming a dead time of 40ns.

4.5 3D CAMERA PROTOTYPE

4.5.1 IMAGE SENSOR

The image sensor design described in the previous Section was fabricated in a 0.35 μ m CMOS, shown in Figure 4.7. The sensor has a surface of 6.5x5.5 mm². The pixel matrix area occupies most of the IC area. Global distribution of S_{MOD} is

implemented symmetrically, from a pad in the center of the top part of the padring.

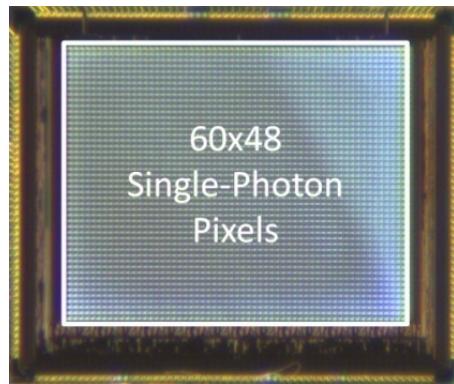


Figure 4.7. Photomicrograph of the SPSD image sensor. The circuit, fabricated in $0.35\mu\text{m}$ CMOS technology, has a surface of $6.5\times 5.5\text{mm}^2$. The pixel pitch is $85\mu\text{m}$.

The image sensor was then mounted on a custom prototype of camera, based on a FPGA for data interface and USB controller to provide a link with any PC. The camera prototype is shown in Figure 4.8 (a).

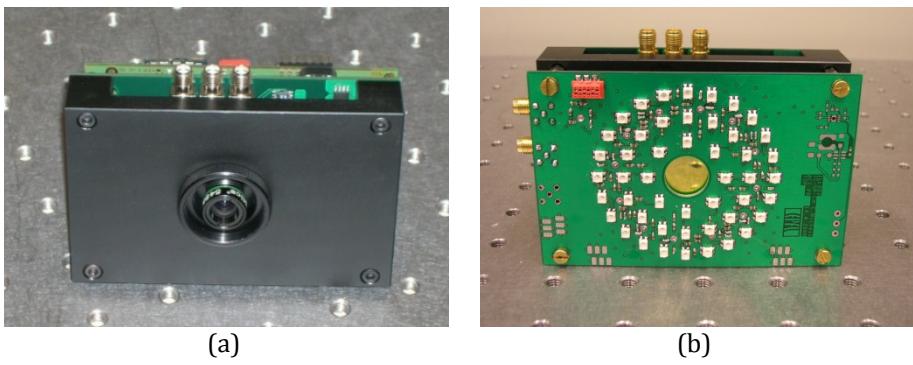


Figure 4.8. Picture of the SPSD camera prototype. (a) Camera casing without illumination system and (b) complete prototype. The 3D camera prototype measures approximately $11\times 7\times 6\text{cm}^3$, nonetheless a clear margin for miniaturization exists.

4.5.2 ILLUMINATION SYSTEM DESIGN

In order to evaluate depth sensing performance, an illumination system was designed and implemented. The illumination system was specifically designed to fit on the camera prototype. Figure 4.8 shows a picture of the full prototype. The illumination was based on array of 48 infrared LEDs emitting at 850nm. In order to prevent occlusion of the illumination beam, the array of LEDs was distributed along 3 concentric circles sharing the same center of the imaging lens. Furthermore, in order to reduce the effects of multiple paths, from different LEDs, the outer LED circle of the illumination was kept small. The illumination system was designed to emit a variable illumination power up to 800mW in average, within a field of view of 50°.

LED drivers may be operated at a maximum frequency of 40MHz. Special placement and routing techniques were adopted for LEDs and drivers so as to prevent LED-to-LED timing skews. Moreover, thermal dissipation was handled with proper layout techniques as well as with adapted driver packages.

4.5.3 CAMERA OPTICS

The custom made camera casing supports a C-mount lens or a M12-mount with adapter. First experimental results, described in the next Section, were based on an f/1.4 imaging lens mounted on a C-mount-to-M12 adapter. The imaging lens field-of-view was approximately 50°/40° (H/V) to match the illumination system field-of-view.

In order to block a fraction of background light, a narrow-band interference optical filter centered at 850nm was placed in front of the imaging lens. The FWHM of the filter was 40nm.

4.6 EXPERIMENTAL RESULTS

In this Section, the proposed image sensor and illumination system are first characterized. The SPSD-based 3D camera prototype was then utilized in the laboratory to assess range-finding performance within a range of a few meters.

Generally, 3D images were captured in indoor condition, characterized with a measured background illuminance of 150 lux.

4.6.1 SINGLE-PHOTON IMAGE SENSOR

The single-photon image sensor described in Section 4.4 was firstly tested and characterized. The SPAD front-end circuit based on the dual-threshold passive quenching and active recharge circuit, introduced in Chapter 2, was fully functional. The nominal dead time of 40ns was set to operate in the 3D camera prototype.

SPADs were characterized with respect to DCR. The distribution of DCR over the 3072 pixels is shown in Figure 4.9. Unexpectedly, the median value of DCR was approximately three times lower than in the sensor described in Chapter 3, whose characterization data was discussed in Chapter 2. The median value in the 60x48 image sensor was 245Hz, whereas the mean value was 1260Hz. Similarly to the previous sensor, higher mean value is mostly due to a small number of highly noisy pixels. Typically, in an imaging application, these pixels are deactivated without any impact to neighbor pixels.

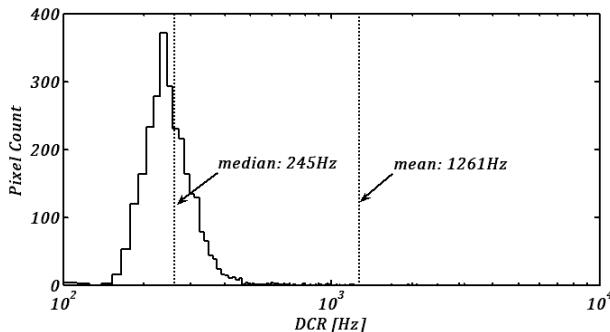


Figure 4.9. DCR distribution over the array of SPADs. The median value was 245Hz, whereas the mean values was relatively higher, at 1261Hz, due to a small number of highly noisy pixels.

Noise performance from SPAD sensors processed in different fabrication batches are likely to suffer from non-negligible spreading in terms of DCR and afterpulsing effects. Furthermore, as experienced by Rochas in [20], noise performance from CMOS SPAD sensors depends on the evolution of the CMOS foundry, which is typically difficult to anticipate from a circuit designer standpoint. For instance, the $0.8\mu\text{m}$ CMOS process utilized by Rochas *et al.* experienced a major modification, which in turn have led to extremely good DCR and afterpulsing performance [20].

However, afterpulsing performance might have been improved by means of the proposed front-end circuit. As described in Chapter 2, the dual-threshold circuit performs a hold-off time over most of the dead time. This hold-off time has a beneficial effect on afterpulsing probability.

4.6.2 ILLUMINATION CHARACTERIZATION

Experimental evaluation started with the investigation of the illumination system. Preliminary measurements showed that the system achieves the best compromise between modulation contrast (c_s) and operating frequency (f_0) when operated at 30MHz. The system may be optionally set to operate at any frequency up to 40MHz as well. However, as indicated by Equation (4.20), distance repeatability error decreases with the inverse of $c_s f_0$ product. Note that distance range R_D is inversely proportional to f_0 , according Equation (4.16). As a result, it is important to increase the $c_s f_0$ product to achieve the best precision within a given illumination power budget, provided that f_0 satisfies the range of interest. Consequently, experimental range-finding performance was evaluated based on illumination repetition rate of 30MHz.

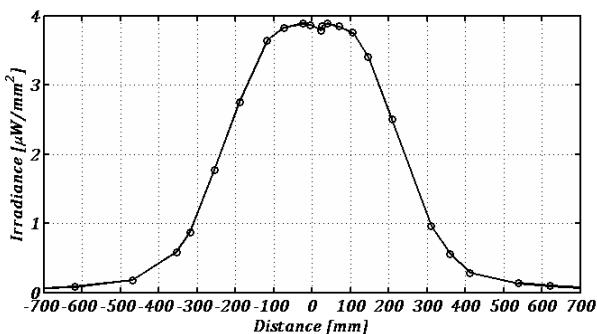


Figure 4.10. Irradiance of illumination source as a function of off-axis distance, measured at 40cm from prototype. Assuming symmetry of revolution, the integral of the irradiance leads to 800mW of optical power.

In order to quantitatively determine the total optical power emitted by the camera prototype, a reference photodiode was utilized to scan the illumination beam at precisely 40cm from the prototype. In order to determine the lateral displacement of the photodiode with respect to the center of the illumination beam, its distance with respect to a fixed reference was measured at each sampling point. Figure 4.10 shows the distribution of illumination irradiance as a function of the off-axis distance. The measurements were carried out along a

horizontal axis whose center coincides with the center of the illumination system. In view of the illumination topology, it is possible to assume symmetry of revolution so as to determine the total power. By numerically integrating the irradiance over the area and along the beam radius, a total illumination power of approximately 800mW was determined.

As described in Section 4.3.2, a very important concern in the illumination system is harmonic distortion. The SPSD demodulation technique perfectly tolerates even harmonic distortion, however, when odd harmonics exist, non-linearity errors are expected. In the design of the illumination system, an important trade-off between power efficiency and distance accuracy exists. In order to design an illumination that emits a pure optical sine wave, linear components are required. Unfortunately, linear components are expensive and typically lead to poor electrical-to-optical conversion efficiency. As a result, the illumination prototype is based on a pseudo sinusoidal modulation, featuring nonetheless some harmonic distortions. In order to experimentally evaluate the quality of the illumination waveform, a high-speed photodiode with a cut-off frequency of 500MHz was utilized to sample the illumination beam. The photodiode output was analyzed on a high performance oscilloscope, capable of sampling the photodiode output signal with a sampling rate of 10GS/s. Finally, a fast Fourier transform (FFT) was computed on the measured waveform. Figure 4.11 shows a plot of the illumination power spectral density as a function of frequency, based on the FFT computation.

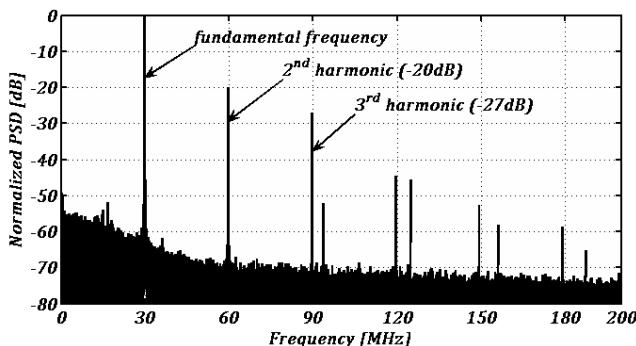


Figure 4.11. Normalized PSD of illumination via FFT computation.

As can be seen in the plot of Figure 4.11, non-negligible harmonic distortion exists. While the second harmonic is filtered out by the SPSD demodulation based on four $\{C_k\}$ samples, the third harmonic produces non-linearity errors.

The 3rd harmonic power was approximately -27dB lower than the fundamental frequency.

Finally, the illumination power actually measured by the 3D camera prototype also depends on the narrow-band interference filter placed in front of the image lens. In order to evaluate how much signal power is lost in the optical filter, a calibrated spectrometer was used to analyze the illumination optical spectrum at a fixed distance, along the optical axis. Once the illumination spectrum was measured, the 40nm optical filter was carefully placed and aligned in front of the spectrometer (optic fiber) input, thus leading to filtered illumination spectrum. A plot of the original and filtered (normalized) optical PSD as a function of wavelength is shown in Figure 4.12.

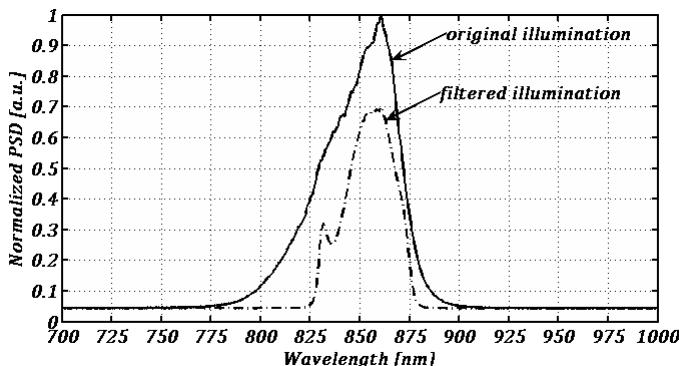


Figure 4.12. Illumination optical spectrum measured with a spectrometer.

Based on the optical PSD measurements, it can be concluded that the optical filter was effective in blocking light components out of the illumination band. However, it also had a negative impact on actual signal power. The filter transmission at 850nm was much lower than expected from its datasheet. Furthermore, the band-pass width of 40nm seemed to be restrictive as it also blocked an important fraction of the signal power. Nonetheless, experimental performance evaluation was based on this setup due to the unavailability of broader band-pass interference filters. The ratio of optical power integrated over the PSD between the filtered and non-filtered optical signals indicates the average filter transmission. It was only 56%, which shows that a large margin exists for improvement in signal gathering or in reducing illumination power consumption, keeping the performance reported in this Chapter constant.

4.6.3 RANGE-FINDING PERFORMANCE

Quantitative evaluation of ranging performance was performed by statistically measuring distance error between the prototype and a reference panel, assumed as ground truth. The reference panel was carefully aligned in front of the camera prototype and its distance was varied from 40cm up to 240cm, in steps of 20cm. In order to ensure accurate reference distances, an alternative measuring method was utilized. Systematic distance errors in the alignment and measurement of the reference plane were assumed to be within +/- 3mm. At each distance, 5000 measurements were computed and stored based on an integration time of 45ms. This integration time leads to 22 frames per second.

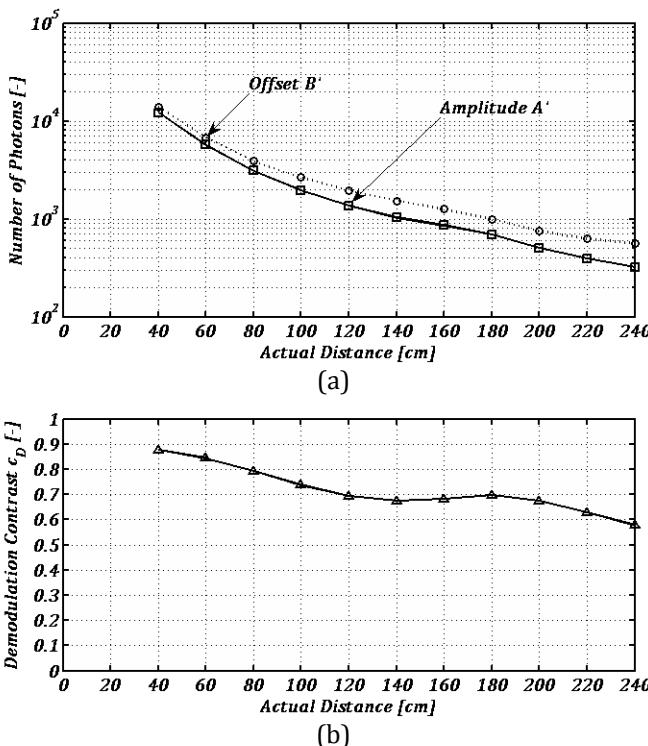


Figure 4.13. (a) Recorded mean offset B' and mean amplitude A' , as a function of distance. (b) Measured mean demodulation contrast c_D as a function of distance. Each distance measurement was statistically evaluated with 5000 measurements using an integration time of 45ms.

In Figure 4.13 (a), the measured mean offset B' and amplitude A' are plotted as function of distance. Figure 4.13 (b) shows a plot of the measured demodulation contrast. Both offset A' and B' follow a quadratic decay with the distance, especially in the beginning of the range. This implies that B' is mostly due to the active illumination, whose field-of-view of 50° leads to such quadratic decay in that range. This is particularly true in SPSD as demodulation is performed virtually free of noise, when background light and DCR are negligible. As can be seen in Figure 4.13 (b), the maximum achieved demodulation contrast is over 85%. Towards the end of the measurement range, the illumination signal is very weak and the background light component in B' makes c_D to decrease with distance. However, it is still higher than 55% throughout the measurement range. High demodulation contrast was expected owing to excellent timing resolution of SPADs. However, one may note that 85% is higher than theoretical maximum c_D of 64%. The reason for this mismatch is that the theoretical c_D was based on a perfect sinusoidal waveform, whereas the actual illumination approaches a square waveform.

In Figure 4.14 (a), measured distance versus actual distance is plotted. Every distance indicated corresponds to the mean value over 5000 measurements. In Figure 4.14 (b), the mean error (μ_{Error}) with respect to ground truth is plotted, whereas 1σ repeatability error, i.e. σ_{Error} , is plotted in Figure 4.14 (c). Obtained non-linearity errors were higher than expected. However, these results compare favorably with non-linearity performance of state-of-the-art 3D cameras based on CCD/CMOS lock-in pixels [33]. Similarly to [33], the main source of non-linearity errors is given by aliasing effects due to higher harmonics in the illumination waveform. An improved illumination source is expected to reduce these errors. The maximum mean error was lower than 12cm throughout the measurement range. It is below 8cm within a range of 2 meters. These errors may be reduced by means of parameterized correction algorithms.

σ_{Error} results were very close to performance predicted by theory, taking into account actual A' and B' values. Interestingly, the measurement conditions led to σ_{Error} that depends linearly on the distance. A linear fit was also plotted in Figure 4.14 (c), which indicates an increase of σ_{Error} at a pace of 1.51cm per meter within the measurement range. In order to see the distribution of the error at a given distance, a histogram was built for the measurements taken at 1 meter of distance. Figure 4.15 shows a plot of the obtained histogram and of a Gaussian fit. The mean distance error was subtracted from each measurement so as to build a histogram with zero mean. As can be seen, the random component of distance errors approaches a normal distribution relatively well. As a result,

σ_{Error} is expected to be improved by averaging similarly to a normal random variable.

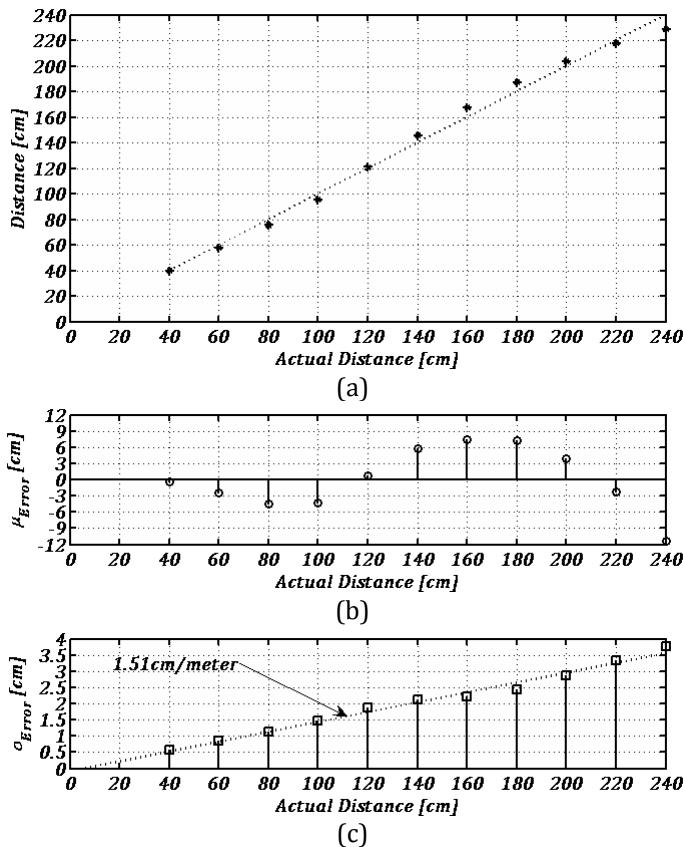


Figure 4.14. Range-finding performance of the proposed 3D camera prototype, based on $N_C = 4$: (a) Measured versus actual distance, compared to the ideal characteristic; (b) Mean non-linearity error as a function of distance; and (c) 1σ distance error as a function of distance. Each distance was statistically evaluated with 5000 measurements using an integration time of 45ms.

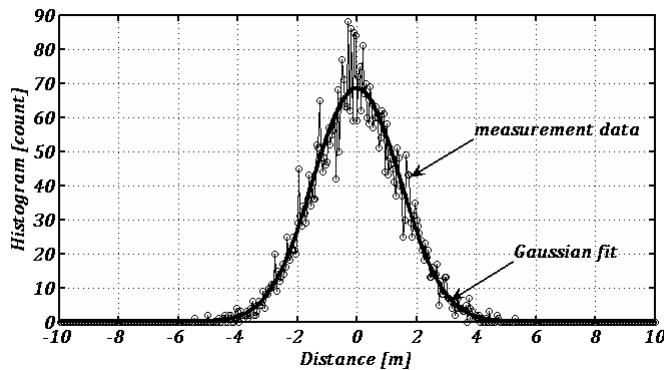


Figure 4.15. Measurement error distribution using 5000 measurements at a distance of 1 meter, each measurement based on 45ms of integration time. In the plot, a comparison with a Gaussian fit is shown. The mean distance was subtracted from each measurement to show the symmetry of σ_{Error} .

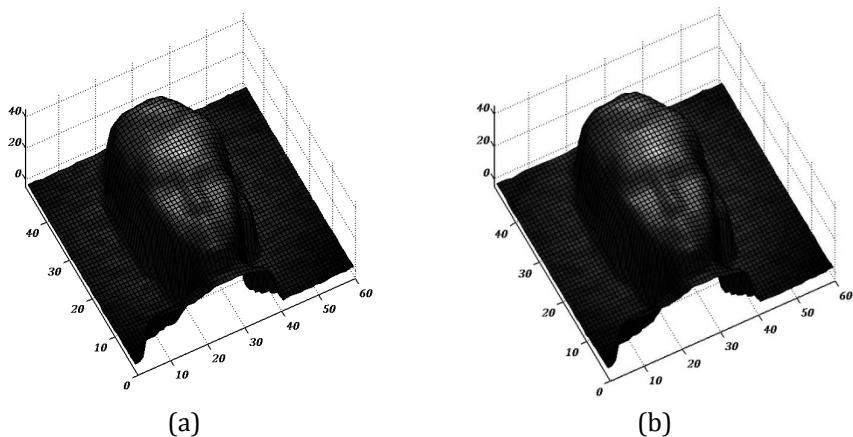


Figure 4.16. Sample 3D images of a human-sized mannequin head acquired with two integration times: (a) 500ms and (b) 1000ms. Depth axis is in centimeter.

Finally, the 3D camera prototype and SPSD technique was evaluated qualitatively by capturing 3D snapshots of several objects. Figure 4.16 (a) and

(b) shows two snapshots of a human-sized mannequin head, acquired with 500ms and 1s of integration time, respectively. Figure 4.17 (a), (b), and (c) shows sample 3D images of a table lamp acquired with 50ms, 500ms, and 1s of integration time, respectively. As can be seen, the effects of averaging, when increasing the integration time, are clearly visible from (a) to (c).

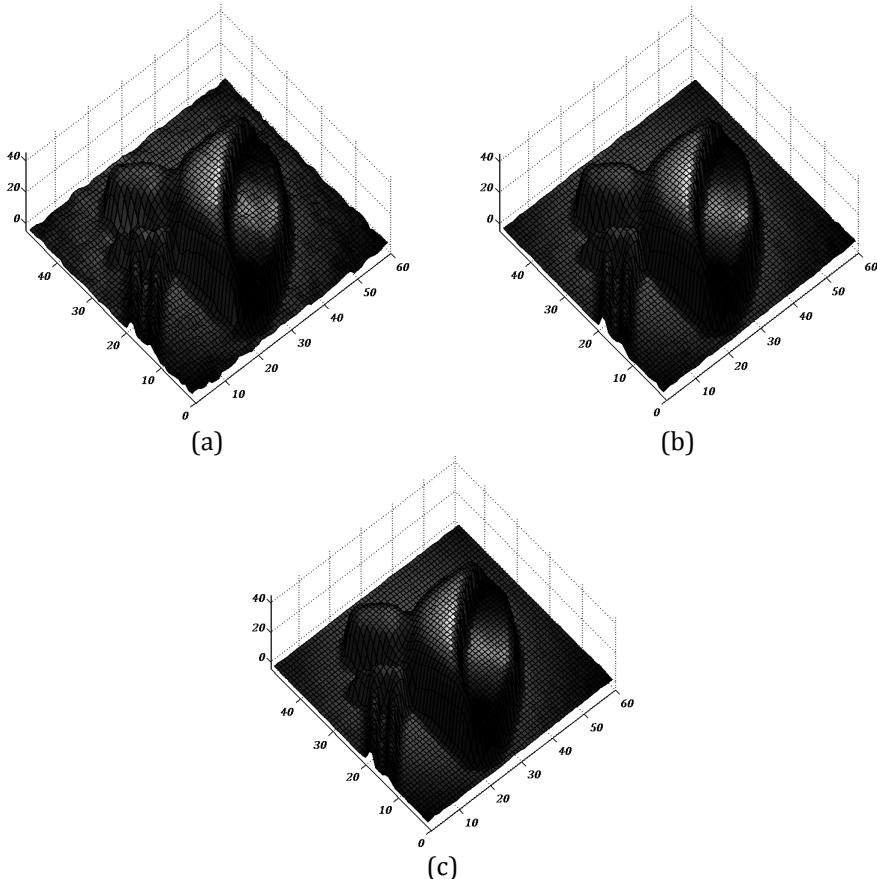


Figure 4.17. Sample 3D images of a table light in front of flat background acquired with three integration times: (a) 50ms, (b) 500ms, and (c) 1s. Depth axis is in centimeter.

In Table 4.2, a summary of the performance achieved by the 3D camera prototype is shown. These results, demonstrates the feasibility of the proposed

SPSD technique. It also shows the performance that is achievable with a fully digital depth image sensor in CMOS technology.

Table 4.2. Summary of design and performance parameters for the SPSD 3D camera prototype.

Parameter	Value	Unit
CMOS technology feature size	0.35	µm
SPAD image sensor resolution	60x48	-
SPAD active area	38	µm ²
SPAD median DCR	245	Hz
SPAD dead time	40	ns
Pixel pitch	85	µm
Imaging lens f-number	1.4	m ⁻¹
Illumination central wavelength	850	nm
Narrow band optical filter width	40	nm
Illumination frequency	30	MHz
Illumination field-of-view	50	°
Illumination average power	800	mW
Resolvable distance range	5	m
Integration time used in performance evaluation	45	ms
Maximum distance non-linearity error (μ_{Error}) up to 2.4m	11	cm
Maximum distance repeatability error (σ_{Error}) up to 2.4m	3.8	cm
Maximum feasible frame-rate a),b)	46296	fps

a) Assuming a complete SPSD demodulation per frame.

b) Currently not achieved due to prototype limitations.

4.7 CONCLUSION

In this Chapter, a novel technique called single-photon synchronous detection (SPSD) was introduced. Among a number of applications that may directly benefit from this technique, solid-state 3D image sensors based on the time-of-flight is the most obvious example. SPSD allows for the sampling and demodulation of optical signals that are continuously modulated in time, by means of single-photon detectors. Despite the discrete detection nature of SPADs, when they are utilized within a SPSD demodulation circuit, it is possible design 3D cameras that measure the phase shift taken by a sinusoidal modulated light beam when it travels to and back from the objects in the scene. Furthermore, such implementations are inherently digital sensors, from photon detection to distance computation. As a result, robust and noise-free demodulation enables depth map evaluations in a wide range of measurement conditions. Digital implementations additionally have the advantage of

favorably evolve with newer CMOS technologies. They also allow the utilization of well established digital design flows in CMOS technology.

Since SPSD is based on a continuous modulated illumination, the theory of lock-in pixels in CCD/CMOS applies, to a certain extent. In this Chapter, an adapted theory was introduced for SPSD. However, thanks to the digital nature of the SPSD implementations, a number of non-idealities existing in lock-in pixels, such as readout noise and gain non-linearity, are not relevant. Furthermore, thanks to picosecond resolution of SPADs, extremely high demodulation contrasts are achievable.

The design of the first fully-parallel implementation of single-photon image sensor in CMOS technology was described in this Chapter. The sensor was based on 60x48 SPSD pixels, thus enabling for the first time real-time 3D image acquisition based on CMOS SPADs. The sensor exhibited a median DCR of 245Hz and a pixel dead time of 40ns. The median DCR was approximately three times lower than in the sensor described in Chapter 3, whose characterization data was discussed in Chapter 2. A 3D camera prototype was designed and built based on the SPAD image sensor. An array of 48 IR LEDs was utilized to generate an illumination beam with a field of view of 50°.

Experimental results showed that the SPSD rangefinder was effective. Demodulation contrast as high as 85% was demonstrated. To the best of our knowledge, this value was superior than any demodulation contrast achieved using lock-in pixels in any technology, under the same conditions. The conditions were: an incident 850nm optical signal and modulated at frequency of 30MHz. Note that modulation contrast depends strongly on modulation frequency and on incident optical wavelength.

Distance measurement performance was characterized with a maximum non-linearity error of 11cm within a range of 2.4 meters. In the same range, the maximum repeatability error was 3.8cm. Furthermore, the repeatability error showed an approximate linear dependence on distance. It was 5 mm at 40cm and increased with a pace of approximately 1.5cm per meter of range.

We believe that non-linearity errors were caused mostly due to aliasing effects as a result of strong harmonic distortion in the illumination system. Harmonic distortion in the illumination system was verified experimentally by means of a fast photodiode and FFT analysis. Improvements in the illumination source will certainly lead to better non-linearity performance. As an alternative, since these errors showed to be systematic, a parameterized error compensation processing is likely to significantly reduce the errors.

Regrettably, at the time this thesis was being written, a non optimized optical filter with a band-pass width of 40nm had to be used in 3D camera prototype. As shown experimentally, the filter blocks 46% of the incident illumination power. Since repeatability errors are improved when the illumination power is increased, by simply replacing the filter with a more appropriate one, repeatability errors would be immediately reduced by more than 25%. Note that σ_{Error} depends on the square root of the incident optical power.

Significant improvements in distance resolution are also expected by improving light detection in pixels by means of microlenses or by improving photon-detection probability. While microlenses are readily feasible, the improvement of SPAD detection efficiency at 850nm is a more challenging achievement. As mentioned in Chapter 2, SPAD structures that utilize deeper layers in CMOS are currently being investigated [167]. These devices are likely to improve overall performance in SPAD sensors operated in red and near infrared spectral regions. Currently, the photon detection probability of the proposed SPAD in this Chapter is 3% at 850nm. As a result, a clear margin for improvement exists. By combining potential improvements of both microlenses and PDP, an enhancement of a factor of 100 in light detection efficiency is foreseen. This improvement would lead to millimeter distance resolution. Alternatively, the improvement could also lead to a very low power illumination of a few tens of milliwatts, keeping current distance resolution unchanged.

As an immediate solution, the use of visible light in the illumination system would enhance light detection efficiency in the current implementation by a factor of 10 (see Chapter 2). Distance resolution, in turn, would be improved by a factor of more than 3, thus leading to sub-centimeter 1σ errors in most of the distance range reported in this Chapter.

To conclude this Chapter, it has to be said that SPSD in 3D imaging is in its beginnings. For example, as a matter of comparison, the lock-in pixel based on photogate in CCD/CMOS has been investigated and improved for more than 12 years [10]. Continued engineering practice in technology, device, and circuits is expected to highly enhance SPSD over the years following this thesis.

5. CONCLUSION

5.1 MAIN ACHIEVEMENTS

The challenges of this investigation have been successfully addressed. For the first time, solid-state 3D image sensors based on single-photon detectors have been fully integrated in CMOS technology. The state-of-the-art in single-photon detection as well as in solid-state 3D imaging has been considerably enhanced.

We have successfully implemented SPADs in $0.35\mu\text{m}$ and in $0.13\mu\text{m}$ CMOS for the first time. With the successful technology migration of SPADs from the original $0.8\mu\text{m}$ CMOS, practical single-photon imaging in CMOS technology has been enabled. Device characterization carried out in this thesis, over a number of designs and in three CMOS technologies, has shown that SPADs may be effectively arranged in high-resolution array formats. For example, a single-photon image sensor featuring 128×128 pixels with a pitch of $25\mu\text{m}$ has been introduced. To the best of our knowledge, this is the largest array of single-photon detectors ever reported. At this pitch, we have confirmed negligible pixel crosstalk probability. While SPAD noise performance varies from device-to-device on arrays, we have not noticed any device failure among the designs evaluated, based on approximately several tens of samples. This result was made possible by robust SPAD structures and mature CMOS technologies. The distribution of the SPAD bias voltage, which is not supported by low-voltage transistors in CMOS, may be therefore globally shared among all the pixels on an image sensor. Furthermore, since SPADs are biased well above the breakdown voltage, small mismatches in devices with respect to breakdown voltage have not impacted photon-detection uniformity. The uniformity of photon detection probability in this thesis favorably compares with conventional image sensors. Similar pixel uniformity has been shown in SPAD timing resolution. Note that these investigations were unaddressed in art before.

In order to implement large arrays of SPADs, area-optimized front-end and readout circuits have been introduced. Passive quenching based on a single MOS transistor has been investigated for the first time. The proposed SPAD biasing topology, in which the SPAD anode terminal is sensed, has enabled us to reduce the effective SPAD parasitic capacitance from 68fF in [20] to 25fF , for

similar device sizes. Based on this parasitic capacitance, the proposed single-NMOS passive quenching is an important candidate for high-resolution single-photon imagers. As demonstrated in this thesis, this circuit achieves better recharging performance than resistor-based passive quenching, provided that appropriate SPAD and transistor biasing is implemented. When afterpulsing performance needs to be optimized, the proposed dual-threshold passive quenching and active recharge circuit is a promising solution. This circuit utilizes the SPAD capacitance value to generate an adjustable hold-off time, thus improving afterpulsing performance under a constant dead time figure. In order to allow the design of single-photon imagers that operate in parallel image acquisition, without the use of complex circuits at pixel level, an event-driven readout scheme has been introduced. This circuit optimizes the reuse of circuit infrastructure among a large number of pixels, thus leading to highly miniaturized designs.

In order to demonstrate the potential of time-correlated single-photon counting (TCSPC) in solid-state 3D imaging, a complete theoretical and experimental investigation was conducted. An analytical model for the evaluation of ranging performance based on TCSPC was introduced. The model aims at validating design parameters in the context of reliable time-of-flight evaluations, thus helping in the design of device, circuit, and system components. A fully integrated TCSPC system for single-photon time-of-flight evaluation has been introduced in CMOS. The IC comprises an array of 128x128 single-photon avalanche diodes with active recharge, a bank of 32 independent time-to-digital converters, and a 7.68Gbps readout system. The 32 converters are shared among 128 detectors in a row using an event-driven readout scheme. Furthermore, a continuously operating calibration scheme ensures an individual resolution of 97ps within 100ns of range and over a wide range of temperatures. The bank of converters can perform a total of 320 million single-photon time-of-flight evaluations per second. To the best of our knowledge, this is the first time a fully integrated TCSPC microsystem was monolithically integrated on a single IC. This sensor was used to characterize range-finding performance within a short-range of 3.75 meters. The optimized design has allowed us to acquire depth maps of 3D scenes with millimeter precisions, in extremely low signal exposure. Quantitative ranging performance has been evaluated using a 1mW laser source with a field-of-view of 5° as active illumination. Accurate distance measurements were repeatedly achieved based on a short integration time of 50ms under 150lux of background light. The maximum non-linearity in distance measurement was 9 millimeters over the full measurement range. Time-varying uncertainty at the farthest distance was 5.2 millimeters. As far as we know, these results are the best ever achieved by a

solid-state 3D image sensor. In the same distance range, the best devices reported in the art, and summarized in Chapter 1, typically exhibit non-linearity errors of the order of several centimeters and centimeter repeatability errors. Furthermore, these performances are achieved with illumination power budgets that are orders of magnitude higher than that reported here.

In order to evaluate the reliability of TCSPC under very low signal-to-background ratio conditions, an experimental evaluation was carried out in a range of 70 meters. Distance performance was tested in outdoor conditions, but under moderate solar background light, due to the optical power limitation of the source. While the background illuminance was approximately 4klux, the illumination mean power was 194mW and its field-of-view was 5°. Statistical analysis of the experimental results showed that, despite the severe conditions in terms of signal-to-background ratio, 100% of time-of-flight evaluations were accurate within a range of 40 meters. In this range, non-linearity errors were below systematic errors in the alignment of the reference panel, approximately ±15cm. 1 σ repeatability errors were below 20cm in the same range. Mean distance evaluations at 50, 60, and 70 meters exhibited high deviations from ground truth. However, a statistical analysis demonstrated that a high percentage of time-of-flight evaluations were accurate. When overall mean and random errors were computed at these distances, false alarms led to considerably large deviations. Nonetheless, at 50 meters, 93.9% of distance evaluations were accurate. This number decreased to 78.9% and 36.9% at 60 and 70 meters, respectively. Taking into account actual signal-to-background ratios, measured on the reference surface, these results are excellent. Assuming a similar figure for illumination power, field-of-view, and acquisition time, we were unable to find similar or better performance in the literature.

TCSPC does not only lead to advantages. The outstanding results summarized in last paragraphs are obtainable at the price of high circuit complexity. On a TCSPC-based time-of-flight image sensor, each pixel should be connected to a TDC circuit and it must count on an individual memory circuit to store a full histogram. These issues may be difficult to be addressed in a high-resolution 3D image sensor when circuit infrastructure, such as TDCs, is not shared among a number of pixels. In order to address these constraints, an innovative approach has been introduced. The method, called single-photon synchronous detection (SPSD), differs considerably from TCSPC and solves most of its practical limitations. In particular, it allows for time-of-flight evaluations based on single-photon detectors without the need for time-to-digital converters nor histogram processing. As a result, large arrays of SPAD may be implemented, yet providing full parallelism in image acquisition, processing, and readout. The approach has

been theoretically and experimentally investigated. The design of the first fully-parallel implementation of a single-photon image sensor in CMOS has been introduced. The sensor is based on 60x48 SPSD pixels, thus enabling real-time 3D image acquisition based on CMOS SPADs for the first time. A 3D camera prototype was designed and built based on the SPSD image sensor. Experimental results showed that the SPSD rangefinder was effective. Demodulation contrast as high as 85% has been demonstrated. As far as we know, this value was superior than any demodulation contrast achieved using lock-in pixels in any technology, under the same conditions. Distance measurement performance was characterized with a maximum non-linearity error of 11cm within a range of 2.4 meters. In the same range, the maximum repeatability error was 3.8cm. We believe that non-linearity errors were caused mostly due to strong harmonic distortion in the illumination system. Harmonic distortion in the illumination system was experimentally verified.

While the SPSD prototype has exhibited similar performance achieved by other state-of-the-art 3D imager approaches, the latter have benefitted from several years of optimizations. Continued engineering practice in technology, device, and circuits is expected to highly enhance SPSD over the years following this thesis.

5.2 OUTLOOK FOR IMPROVEMENTS

Depending on the actual rangefinder implementation of interest, among TCSPC and SPSD, different improvements are foreseen.

In SPSD, significant improvements in distance resolution are expected by improving light detection in pixels. For example, light detection efficiency may be improved by means of microlenses or by enhancing photon-detection probability. While microlenses are readily feasible, the improvement of SPAD detection efficiency at 850nm is a more challenging achievement. SPAD structures that utilize deeper layers in CMOS have been investigated [167]. These devices are likely to improve overall performance in SPAD sensors operated in red and near infrared spectral regions. Currently, the photon detection probabilities of the investigated SPADs are approximately 3% at 850nm. As a result, a clear margin for improvement exists. By combining potential improvements of both microlenses and PDP, an enhancement of a factor of 100 in light detection efficiency is foreseen. This improvement would lead to millimeter distance resolution in SPSD. Alternatively, the improvement could also lead to a very low power illumination of a few tens of milliwatts,

keeping current distance resolution unchanged. Although these improvements would generally lead to superior performance in TCSPC, the actual performance enhancement is tempered, when compared to SPSD.

As an immediate solution, the use of visible light in the illumination system would enhance light detection efficiency in the current implementation by a factor of 10. However, the applicability of visible light obviously depends upon application requirements. In SPSD, this enhancement would lead to a distance resolution improvement of a factor three, approximately. In TCSPC, since non-linear processing is involved, the proposed theoretical model may be utilized to simulate improvement potential, in a particular scenario.

In order to further miniaturize pixels and increase image resolution, the evaluation of fully-integrated imagers in 130nm CMOS is a promising research direction. However, afterpulsing performance is still an open question. Moreover, it would be desirable to improve DCR performance to levels achievable in 0.35 μ m CMOS. Furthermore, detection efficiency in the infrared spectral region tends to be lower in more advanced CMOS technology. The suitability of this approach depends on the applicability of visible light in a given application. Alternatively, a state-of-the-art CMOS technology may be customized to better support SPADs, thereby solving most of the remaining issues. This approach is widely adopted in CMOS image sensors and could be foreseen for SPADs. Since a typical SPAD optimization for low-noise and high infrared detection efficiency would involve deep and relatively low doping concentrations, specific implants may be added in early processing steps of the CMOS technology. As a result, it is possible to predict little impact on the implants and layers that are added subsequently in the process. However, this approach is typically preempted by economic challenges. It therefore relies on the successful use of 3D image sensors, or another SPAD application, in high volume applications.

An important outlook in short time for this thesis is the evaluation, and possibly customization, of the proposed designs in specific applications. In the context of a specific application, an important margin for the optimization of the illumination source exists. While TCSPC provides superior ranging performance, it enforces more stringent speed constraints in the illumination source. Conversely, SPSD requires illumination sources that exhibit comparatively significantly lower bandwidth. As a result, light-emitting diodes are good candidates in SPSD. However, special treatment of harmonic distortion should be envisioned for SPSD. In both cases, electro-optical conversion efficiency exhibits an important potential for improvement.

APPENDIX A:

CLASS 1 EYE SAFETY ESTIMATION

BASED ON IEC608251

Circular diffuser with 20° of field-of-view and 850nm wavelength

It is assumed in this calculation that the laser source is covered with a diffuser whose diameter should be determined in accordance with the IEC60825-1 standard, for Class 1 range. In addition, as a result of previous estimates, it is assumed that the diameter of diffuser is larger than 10mm. This will simplify the calculations since C_6 and T_2 become constants.

Every surface point of the diffuser emits light within a cone of 20° of field-of-view. The light radiance is constant at any position within the cone of light. A central emission wavelength of 850nm is assumed for the laser source.

A.1 Problem Parameters

When dealing with the standard, we are interested in computing the exposure of light within a human eye. The worse case aperture of the eye (pupil) is defined in the standard as 7mm. Consequently, the eye's opening surface SE in squared meters [m^2] is given by

$$SE = \pi \left(\frac{0.007}{2} \right)^2. \quad (\text{A.1})$$

The human eye is considered to stand in front of the source at a distance L [m]:

$$L = 0.1. \quad (\text{A.2})$$

As stated previously, the diffuser emits light within an angle θ [rad] as follows.

$$\theta = 20^\circ. \quad (\text{A.3})$$

According to the standard, the maximum angular subtense of the apparent source defined by α_{max} , corresponds to a source whose diameter is given by D_{max} in [m]:

$$D_{max} = 0.01. \quad (\text{A.4})$$

The average power of the illumination system P_{mean} [W] is given as follows:

$$P_{mean} = 20. \quad (\text{A.5})$$

The illumination is pulsed at a frequency f_0 of 2MHz, whose period T_0 [s] is given by

$$T_0 = \frac{1}{2 \cdot 10^6}. \quad (\text{A.6})$$

For repetitively pulsed illumination sources, the standard states three conditions to be fulfilled:

- i. the exposure of any single pulse should not exceed the MPE for the duration of a single pulse (T_0 in this case);
- ii. the average power for a pulse train should not exceed the MPE over the duration T_2 ;
- iii. the average exposure from all of the pulses within the pulse train should not exceed the MPE of an single pulse, multiplied by a correction factor C_5 , where C_5 is a function of the number of pulses within a train of pulses.

In addition, for condition (iii) only, the standard states that, if the actual single pulse duration is below the duration given in Table 9 (i.e. T_1), a number of successive pulses should be summed up so as to form a single pulse of duration T_1 . Under the assumptions of this study, the standard gives T_1 (Table 9) and T_2 (Notes to tables 1 to 4) as follows:

$$T_1 = 18 \cdot 10^{-6} \quad (\text{A.7})$$

$$T_2 = 100 \quad (\text{A.8})$$

Under the same assumptions, one can also determine the following corrections factors (from notes to tables 1 to 4):

$$C_4 = 10^{0.002(850-700)} \quad (\text{A.9})$$

$$C_5 = \left(\frac{T_2}{T_1} \right)^{-0.25} \quad (\text{A.10})$$

$$C_6 = 66.7. \quad (\text{A.11})$$

A.2 Light Exposure Formulation

In order to determine the amount of light that enters into the eye aperture, one needs to calculate the total surface ($SS [m^2]$) that is illuminated by the source with a diameter d at the distance L . Geometrically, SS depends on the diameter d of the source which is the parameter to be found. Thus,

$$SS(d) = \pi \left(\frac{d}{2} + L \tan \frac{\theta}{2} \right)^2. \quad (\text{A.12})$$

In addition, since the source diameter d is assumed to be larger than D_{max} , one needs to determine the fraction of the total source power (or exposure energy) that should be considered for MPE. Given an input PEX that could be either a power [W] or exposure [J] measure and the source diameter d , than the fraction of power (or exposure) of interest PS [W or J] can be determined as follows:

$$PS(PEX, d) = PEX \left(\frac{D_{max}}{d} \right)^2. \quad (\text{A.13})$$

Finally, the power or exposure PE [W or J] emitted by the source that enters the eye aperture is proportional to the fraction of source exposure and to the eye opening surface SE , and is inversely proportional to the total illuminated surface SS , as follows:

$$PE(PEX, d) = SE \frac{PS(PEX, d)}{SS(d)}. \quad (\text{A.14})$$

As can be seen, PE is formulated so as to depend on the source power (or energy) and on the source diameter d . We will change in the next sections the power (or exposure energy) according to each condition among (i), (ii), and (iii).

A.3 Condition (i): single pulse

In order to determine the exposure energy for a single pulse (ESINGLE), we take into account the average source power and the time duration of each pulse, i.e. T_0 . Then,

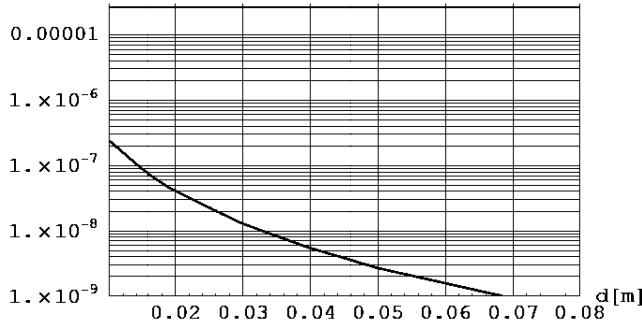
$$ESINGLE = P_{mean} T_0. \quad (\text{A.15})$$

The AEL (*EAELSINGLE*) for a single pulse can be found in Table 1 of the standard for the duration of T_0 :

$$EAELSINGLE = 2 \cdot 10^{-7} C_4 C_6. \quad (\text{A.16})$$

Now, by plotting *EAELSINGLE* and *PE* for *ESINGLE* as a function of the source diameter d , we can determine d by looking at the intersection between *EAELSINGLE* (constant) and *PE*.

EAELSINGLE [J] and *PE* (*ESINGLE*) [J]



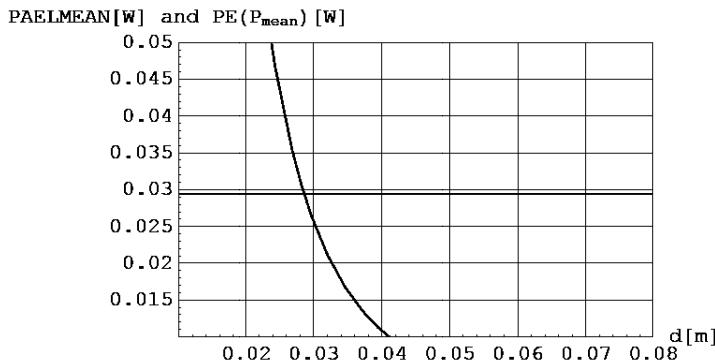
As can be seen in the picture above, *PE*(*ESINGLE*) is always lower than *EAELSINGLE* when the diameter d is larger than D_{max} . This means that for any d larger than D_{max} , condition (i) is fulfilled.

A.4 Condition (ii): average power

In this condition, we are interested in average power of the illumination source. We can directly consider the average power P_{mean} for plotting *PE* as a function of the source diameter. From Table 1 of the standard, we will consider *AEL* for the duration of T_2 . Since this exposure unit is Joule, we will convert it to power (*PAELMEAN*) so that we can compare it with P_{mean} , i.e., we divide the exposure given by Table 1 by T_2 :

$$PAELMEAN = \frac{7 \cdot 10^{-4} \cdot T_2^{0.75} C_4 C_6}{T_2}. \quad (\text{A.17})$$

Now, we can again plot $PAELMEAN$ (constant) and PE for P_{mean} as a function of d . The intersection between these curves gives us the minimum diameter d .



As can be seen in the figure above, $PAELMEAN$ and P_{mean} intersect for d of 0.028m or, equivalently, **2.8cm**.

A.5 Condition (iii): train of pulses

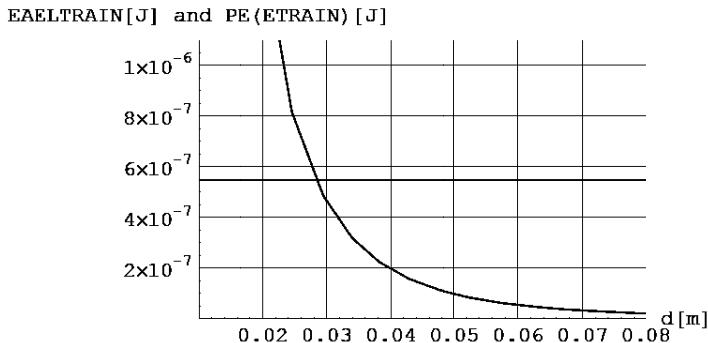
Finally, we are left with the last condition requested for class 1 compliance. The standard requires us to compare the exposure of a full train of pulses to the AEL of a single pulse multiplied by a correction factor C_5 . We can define the AEL for the pulse train as $EAELTRAIN$:

$$EAELTRAIN = 2 \cdot 10^{-7} C_4 C_5 C_6. \quad (\text{A.18})$$

In addition, we define the exposure of train of pulse as $ETRAIN$, then

$$ETRAIN = P_{mean} T_1. \quad (\text{A.19})$$

As in the previous section, we can find the diameter d that fulfills the condition (iii) by plotting $EAELTRAIN$ and PE for $ETRAIN$ as a function of d .



According to the figure, the intersection between $EAELTRAIN$ (constant) and $PE(ETRAIN)$ gives us a diameter d of **2.8cm**, similar to the condition (ii). Since this diameter value is the most restrictive among all the three conditions, it defines the diameter the diffuser needs to have to comply with class 1.

REFERENCES

1. P. Besl, *Active, optical range imaging sensors*. Machine Vision and Applications, 1988. **2**: pp. 127-152.
2. R. Jarvis, *Range Sensing for Computer Vision in Three-Dimensional Object Recognition Systems*. Three-Dimensional Object Recognition Systems, ed. A. K. Jain and P. J. Flynn. 1993, Amsterdam: Elsevier Science.
3. R. Dandliker, Y. Salvade, and E. Zimmermann, *Distance measurement by multiple-wavelength interferometry*. Journal of Optics, 1998. **29**: pp. 105-114.
4. J. Paakkari, H. Ailisto, and H. Kopola, *Flatness and dimensional control of planar objects with a novel modular moire system*. Journal of Optics, 1998. **29**(3): pp. 174-178.
5. W. Koechner, *Optical Ranging System Employing a High Power Injection Laser Diode*. IEEE Transactions on Aerospace and Electronic Systems, 1968. **4**(1): pp. 81-91.
6. J. P. Anthes, P. Garcia, J. T. Pierce, P. V. Dressendorfer, I. K. Evans, B. D. Bradley, J. T. Sackos, and M. M. LeCavalier, *Non-scanned lidar imaging and applications*. Proceedings of SPIE Vol. 1936, Oct. 1993: pp. 11-22.
7. S. A. Weiss, *Laser-Radar Imaging without Scanners*. Photonics Spectra, April 1994: pp. 22-29.
8. M. R. Muguirra, J. T. Sackos, and B. D. Bradley, *Scannerless Range Imaging with a Aquare Wave*. Proceedings SPIE, 1995. **2472**: pp. 106-113.
9. H. Kume, K. Koyama, K. Nakatsugawa, S. Suzuki, and D. Fatlowitz, *Ultrafast microchannel plate photomultipliers*. Applied Optics, 1988. **27**(6): pp. 1170-1178.
10. T. Spirig, P. Seitz, O. Vietze, and F. Heitger, *The Lock-In CCD-Two-Dimensional Synchronous Detection of Light*. IEEE Journal of Quantum Electronics, 1995. **31**(9): pp. 1705-1708.
11. R. Miyagawa and T. Kanade, *CCD-Based Range-Finding Sensor*. IEEE Transactions on Electron Devices, 1997. **41**(10): pp. 1648-1652.
12. R. Lange and P. Seitz, *Solid-State Time-of-Flight Range Camera*. IEEE Journal of Quantum Electronics, 2001. **37**(3): pp. 390-397.
13. P. Palojärvi, K. Määttä, and J. Kostamovaara, *Integrated Time-of-Flight Laser Radar*. IEEE Instrumentation and Measurement Technology Conference, 1996: pp. 1378-1381.

References

14. H. Ailisto, V. Heikkinen, R. Mitikka, R. Myllylä, J. Kostamovaara, A. Mäntyniemi, and M. Koskinen, *Scannerless imaging pulsed-laser range finding*. Journal of Optics A: Pure and Applied Optics, 2002. **4**: pp. 337-346.
15. R. Schwarte, Z. Xu, H. Heinol, J. Olk, and B. Buxbaum, *New optical four-quadrant phase detector integrated into a photogate array for small and precise 3D cameras*. Proceedings of SPIE - Three-Dimensional Image Capture, 1997. **3023**: pp. 119-128.
16. R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte, *A new electrooptical mixing and correlating sensor: Facilities and Applications of the Photonic Mixer Device (PMD)*. Proceedings of SPIE, Sensors, Sensor Systems, and Sensor Data Processing, 1997. **3100**: pp. 245-253.
17. M. Vaidyanathan, A. Joshi, S. Xue, B. Hanyaloglu, M. Thomas, M. Zandian, D. Edwall, G. Williams, J. Blackwell, W. Tennant, and G. Hughes, *High Performance Ladar Focal Plane Arrays for 3D Range Imaging*. IEEE Aerospace Conference Proceedings, 2004: pp. 1776-1781.
18. B. F. Aull, A. H. Loomis, D. J. Young, R. M. Heinrichs, B. J. Felton, P. J. Daniels, and D. J. Landers, *Geiger-mode avalanche photodiodes for three-dimensional imaging*. Lincoln Laboratory Journal, 2002. **13**(2): pp. 335-350.
19. A. Rochas, M. Gani, B. Furrer, P.-A. Besse, R. S. Popovic, G. Ribordy, and N. Gisin, *Single photon detector fabricated in a complementary metal-oxide-semiconductor high-voltage technology*. Review of Scientific Instruments, 2003. **74**(7): pp. 3263-3270.
20. A. Rochas, *Single photon avalanche diodes in CMOS technology*. 2003, PhD thesis dissertation Nr. 2814, Ecole Polytechnique Fédérale de Lausanne.: Switzerland.
21. C. Niclass, A. Rochas, P. A. Besse, and E. Charbon, *A CMOS Single Photon Avalanche Diode Array for 3D Imaging*. IEEE International Solid-State Circuits Conference, Feb. 2004: pp. 121-122.
22. C. Niclass, A. Rochas, P. A. Besse, and E. Charbon, *A CMOS 3D Camera with Millimetric Depth Resolution*. IEEE Custom Integrated Circuits Conference, Oct. 2004: pp. 705-708.
23. C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, *Towards a 3D Camera Based on Single Photon Avalanche Diodes*. IEEE Journal of Selected Topics in Quantum Electronics, 2004. **10**(4): pp. 796-802.
24. D. Stoppa, L. Pancheri, M. Scandiuzzo, A. Simoni, V. L., and G.-F. D. Betta, *A CMOS Sensor based on Single Photon Avalanche Diode for Distance Measurement Applications*. IEEE Instrumentation and Measurement Technology Conference 2005: pp. 1162-1165.

25. D. Stoppa, L. Pacheri, M. Scanduzzo, L. Gonzo, G.-F. D. Betta, and A. Simoni, *A CMOS 3-D Imager Based on Single Photon Avalanche Diode*. IEEE Transactions on Circuits and Systems, 2007. **54**(1): pp. 4-12.
26. C. Bamji and E. Charbon, *Systems for CMOS-compatible three-dimensional image sensing using quantum efficiency modulation*. US Patent 6,580,496, 2003.
27. A. Mantyniemi, T. Rahkonen, and J. Kostamovaara, *An Integrated 9-Channel Time Digitizer with 30ps Resolution*. IEEE International Solid-State Circuits Conference, Feb. 2002. **1**: pp. 266-267.
28. Canesta. *Product Main Page*. [cited May. 2008]; Available from: <http://www.canesta.com>.
29. S. B. Gokturk, H. Yalcin, and C. Bamji, *A Time-Of-Flight Depth Sensor - System Description, Issues an Solutions*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2004.
30. T. Oggier, R. Kaufmann, M. Lehmann, B. Büttgen, S. Neukom, M. Richter, M. Schweizer, P. Metzler, F. Lustenberger, and N. Blanc, *Novel Pixel Architecture with Inherent Background Suppression for 3D Time-of-Flight Imaging*. Proceedings of SPIE-IS&T Electronic Imaging, 2005. **5665**: pp. 1-8.
31. M. Lehmann, L. Cavalier, S. Lauxtermann, M. Richter, F. Lustenberger, N. Blanc, and P. Seitz, *Smart Pixels for future 3D-TOF sensors*. IEEE Workshop on CCD and Advanced Image Sensors, 2005.
32. B. Büttgen, T. Oggier, M. Lehmann, R. Kaufmann, and F. Lustenberger, *CCD/CMOS Lock-In Pixel for Range Imaging: Challenges, Limitations and State-of-the-Art*. 1st Range Imaging Research Day, 2005: pp. 21-32.
33. B. Büttgen, *Extending Time-of-Flight Optical 3D-Imaging to Extreme Operating Conditions*. 2006, PhD thesis dissertation, University of Neuchâtel: Switzerland.
34. T. Möller, H. Kraft, J. Frey, M. Albrecht, and R. Lange, *Robust 3D Measurement with PMD Sensors*. 1st Range Imaging Research Day, 2005.
35. R. Lange, *3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology*. 2000, PhD Dissertation, University of Siegen: Germany.
36. B. Büttgen, T. Oggier, R. Kaufmann, P. Seitz, and N. Blanc, *Demonstration of a Novel Drift Field Pixel Structure for the Demodulation of Modulated Light Waves with Application in Three-Dimensional Image Capture*. Proceedings of SPIE-IS&T Electronic Imaging, 2004. **5302**: pp. 9-20.
37. B. Büttgen, T. Oggier, M. Lehmann, R. Kaufmann, S. Neukom, M. Richter, M. Schweizer, D. Beyeler, R. Cook, C. Gimkiewicz, C. Urban, P. Metzler, P. Seitz, and F. Lustenberger, *High-Speed and High-Sensitive Demodulation Pixel for 3D-Imaging*. Proceedings of SPIE-IS&T Electronic Imaging, 2006. **6056**.

References

38. B. Büttgen, F. Lustenberger, and P. Seitz, *Demodulation Pixel based on Static Drift Fields*. IEEE Transactions on Electron Devices, 2006. **53**(11): pp. 2741-2747.
39. R. Jeremias, W. Brockherde, G. Doemens, B. Hosticka, L. Listl, and P. Mengel, *A CMOS Photosensor Array for 3D Imaging Using Pulsed Laser*. IEEE International Solid-State Circuits Conference, Feb. 2001: pp. 252-253.
40. P. Mengel, G. Doemens, and L. Listl, *Fast Range Imaging by CMOS Sensor Array Through Multiple Double Short Time Integration (MDSI)*. IEEE International Conference on Imaging Processing, 2001: pp. 169-172.
41. O. M. Schrey, O. Elkhalili, P. Mengel, M. Petermann, W. Brockherde, and B. J. Hosticka, *A 4x64 Pixel CMOS Image Sensor for 3D Measurement Applications*. IEEE European Solid-State Circuits Conference, 2003: pp. 333-336.
42. D. Stoppa, L. Gonzo, M. Gottardi, A. Simoni, and L. Viarani, *A Novel Fully Differential Pixel Concept for Indirect TOF 3D Measurement*. IEEE Instrumentation and Measurement Technology Conference, 2003: pp. 152-155.
43. L. Viarani, D. Stoppa, L. Gonzo, M. Gottardi, and A. Simoni, *A CMOS test chip for 3D vision applications with high dynamic range*. European Conference on Circuit Theory and Design, 2003. **3**: pp. 233-236.
44. L. Viarani, D. Stoppa, L. Gonzo, M. Gottardi, and A. Simoni, *A CMOS Smart Pixel for Active 3-D Vision Applications*. IEEE Sensors Journal, 2004. **4**(1): pp. 145-152.
45. D. Stoppa, L. Viarani, A. Simoni, L. Gonzo, M. Malfatti, and G. Pedretti, *A 16x16-Pixel Range-Finding CMOS Image Sensor*. IEEE European Solid-State Circuits Conference, 2004: pp. 419-422.
46. O. Elkhalili, O. M. Schrey, W. Ulfing, W. Brockherde, B. J. Hosticka, P. Mengel, and L. Listl, *A 64x8 Pixel 3-D CMOS Time of Flight Image Sensor for Car Safety Applications*. IEEE European Solid-State Circuits Conference, 2006: pp. 568-571.
47. G. Yahav, G. J. Iddan, and D. Mandelboum, *3D Imaging Camera for Gaming Application*. IEEE International Conference on Consumer Electronics, 2007: pp. 1-2.
48. A. Manassen and A. Sander, *Compound optical modulator for generating multiplexed optical signals for transmission in communication networks, has stack of quantum well structures having electronic that generates electric field orthogonal to layers of modulator*. PCT Patent WO2003001282-A, 2001.
49. 3DV Systems, *ZCam Product Data Sheet*. 2007.
50. T. Ushinaga, I. A. Halin, T. Sawada, S. Kawahito, M. Homma, and Y. Maeda, *A QVGA-size CMOS time-of-flight range image sensor with background light charge draining structure*. Proceedings of SPIE-IS&T Electronic Imaging, 2006. **6056**.

51. S. Kawahito, I. A. Halin, T. Ushinaga, T. Sawada, M. Homma, and Y. Maeda, *A CMOS Time-of-Flight Range Image Sensor With Gates-on-Field-Oxide Structure*. IEEE Sensors Journal, 2007. **7**(12): pp. 1578-1586.
52. M. Kawakita, K. Iizuka, H. Nakamura, I. Mizuno, T. Kurita, T. Aida, Y. Yamanouchi, H. Mitsumine, T. Fukaya, H. Kikuchi, and F. Sato, *High-definition real-time depth-mapping TV camera: HDTV Axi-Vision Camera*. Optics Express, 2004. **12**(12): pp. 2781-2786.
53. T. Oggier, B. Büttgen, F. Lustenberger, R. Becker, B. Rüegg, and A. Hodac, *SwissRanger SR3000 and First Experiences based on Miniaturized 3D-TOF Cameras*. 1st Range Imaging Research Day, 2005: pp. 97-108.
54. B. F. Aull, *Geiger-Mode Avalanche Photodiode Arrays Integrated with CMOS timing Circuits*. Meeting of the Boston Chapter of the IEEE Electroc Devices Society, 1998.
55. W. J. Kindt, *Geiger mode avalanche photodiode arrays for spatially resolved single photon counting*. 1999, PhD thesis dissertation, Technische Universiteit Delft. ISBN: 90-407-1845-8.
56. D. V. O'Connor and D. Phillips, *Time-Correlated Single-Photon Counting*. 1984, London: Academic Press.
57. W. Becker, *Advanced Time-correlated Single-Photon Counting Techniques*. 2005, Berlin: Springer.
58. R. M. Heinrichs, B. F. Aull, R. M. Marino, D. G. Fouche, A. K. McIntosh, J. J. Zaykowski, T. Stephens, M. E. O'Brien, and M. A. Albota, *Three-Dimensional Laser Radar with APD Arrays*. Proceedings of SPIE, Laser Radar Technology and Applications VI, 2001. **4377**: pp. 106-117.
59. R. M. Marino, T. Stephens, R. E. Hatch, J. L. McLaughlin, J. G. Mooney, M. E. O'Brien, G. S. Rowe, J. S. Adams, L. Skelly, R. C. Knowlton, S. E. Forman, and W. R. Davis, *A compact 3D imaging laser radar system using Geiger-mode APD arrays: system and measurements*. Proceedings of SPIE, Laser Radar Technology and Applications VIII, 2003. **5086**: pp. 1-15.
60. B. F. Aull, A. H. Loomis, D. J. Young, A. Stern, B. J. Felton, P. J. Daniels, D. J. Landers, L. Retherford, D. D. Rathman, R. M. Heinrichs, R. M. Marino, D. G. Fouche, M. A. Albota, R. E. Hatch, G. S. Rowe, D. G. Kocher, J. G. Mooney, M. E. O'Brien, B. E. Player, et al., *Three-dimensional imaging with arrays of Geiger-mode avalanche photodiodes*. Proceedings of SPIE, Semiconductor Photodetectors, 2004. **5353**: pp. 105-116.
61. B. F. Aull, J. Burns, C. Chen, B. J. Felton, H. Hanson, C. Keast, J. Knecht, A. H. Loomis, M. Renzi, A. Soares, V. Suntharalingam, K. Warner, D. Wolfson, D. Yost, and D. J. Young, *Laser radar imager based on three-dimensional integration of Geiger-mode avalanche photodiodes with two SOI timing-circuit layers*. IEEE International Solid-State Circuits Conference, 2006: pp. 304-305.

References

62. PReVENT. *Integrated Project. European Automotive Industry and European Commission.* 2007 [cited May 2008]; Available from: <http://www.prevent-ip.org>.
63. P. Mengel, L. Listl, B. König, C. Toepfer, M. Pellkofer, W. Brockherde, B. Hosticka, O. El Khalili, O. Schrey, and W. Ulfig, *Three-Dimensional CMOS Image Sensor for Pedestrian Protection and Collision Mitigation*. The International Forum on Advanced Microsystems for Automotive Applications (AMAA), 2006: pp. 23-39.
64. S. Hsu, S. Acharya, A. Rafii, and R. New, *Performance of a Time-of-Flight Range Camera for Intelligent Vehicle Safety Applications*. The International Forum on Advanced Microsystems for Automotive Applications (AMAA), 2006: pp. 205-219.
65. B. Buxton, *Surface and Tangible Computing, and the "Small" Matter of People and Design*. IEEE International Solid-State Circuits Conference, 2008.
66. J. Han, *Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflections*. ACM Symposium on User Interface Software and Technology, 2005.
67. J. Han. *Unveiling the genius of multi-touch interface design.* 2006 [cited May 2008]; Available from: <http://www.ted.com>.
68. Apple. *iPhone Multi-touch.* 2008 [cited May 2008]; Available from: <http://www.apple.com/iphone/technology>
69. Nintendo Corporation. *Nintendo Wii.* 2008 [cited May 2005]; Available from: <http://www.wii.com>.
70. Microsoft Corporation. *Microsoft Surface.* 2008 [cited May 2008]; Available from: <http://www.microsoft.com/surface>.
71. S. Hodges, S. Izadi, A. Butler, A. Rrustemi, and B. Buxton, *ThinSight: versatile multi-touch sensing for thin form-factor displays*. Proceedings of the 20th annual ACM symposium on User interface software and technology, 2007: pp. 259-268.
72. H. K. Lim, *The 2nd Wave of Digital Consumer Revolution Challenges and Opportunities*. IEEE International Solid-State Circuits Conference, 2008.
73. 3DV Systems. *Application Page.* 2008 [cited May 2008]; Available from: <http://www.3dvsystems.com>.
74. Mesa Imaging. *Product Main Page.* 2008 [cited May 2008]; Available from: <http://www.mesa-imaging.ch>.
75. PMDTec. *Product Main Page.* 2008 [cited May 2008]; Available from: <http://www.pmdtec.com>.
76. G. J. Iddan and G. Yahav, *3D imaging in the studio (and elsewhere...)*. Proceedings of SPIE - Three-Dimensional Image Capture and Applications IV, 2001. **4298**: pp. 48-55.

77. J. W. Weingarten, G. Gruener, and R. Siegwart, *A state-of-the-art 3D sensor for robot navigation*. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2004: pp. 2155-2160.
78. R. V. Bostelman, T. H. Hong, and R. Madhavan, *Towards AGV safety and navigation advancement - Obstacle detection using a TOF range camera*. International Conference on Advanced Robotics, 2005: pp. 460-467.
79. B. Jensen, J. W. Weingarten, S. Kolski, and R. Siegwart, *Laser Range Imaging Using Mobile Robots: From Pose Estimation to 3D-Models*. Proceedings of the 1st Range Imaging Research Day, ETH Zurich, 2005: pp. 129-144.
80. J. Kubacki and K. Pfeiffer, *Using Range Imaging Sensors with Color Imaging Sensors in Cognitive Robot Companions: A New and Simple Calibration Technique Based on Particle Swarm Optimization*. Proceedings of the 1st Range Imaging Research Day, ETH Zurich, 2005: pp. 43-55.
81. S. May, B. Werner, H. Surmann, and K. Pervolz, *3D time-of-flight cameras for mobile robotics*. IEEE International Conference on Intelligent Robots and Systems, 2006: pp. 790-795.
82. R. Craig, I. Gravseth, R. P. Earhart, J. Bladt, S. Barnhill, L. Ruppert, and C. Centamore, *Processing 3D Flash LADAR point-clouds in real-time for flight applications*. Proceedings of SPIE, 2007. **6555**: pp. 65550D.
83. A. K. Vadlamani and M. U. De Haag, *Aerial vehicle navigation over unknown terrain environments using flash LADAR and inertial measurements*. Proceedings of the Institute of Navigation, National Technical Meeting, 2007: pp. 1104-1118.
84. T. Bömer, *Vision Based Protective Devices (VBPD) - A Vision becomes reality*. 3rd International Conference on Industrial Automated Systems (SIAS), 2003.
85. M. Hauke and T. Bömer, *Range imaging and vision based protective devices (VBPDs) A team for safety?* Proceedings of the 1st Range Imaging Research Day, ETH Zurich, 2005: pp. 81-87.
86. European Agency for Safety and Health at Work. *Expert forecast on emerging physical risks related to occupational safety and health*. 2008 [cited May 2008]; Available from: <http://riskobservatory.osha.europa.eu>.
87. IEC International Electrotechnical Commission, *IEC 61496-4 Safety of machinery - Electro-sensitive protective equipment - Part 4: Particular requirements for equipment using vision based protective devices (VBPD)* 2007.
88. Q. Xiao and K. Dahel. *Challenges and Potential Research Areas in Biometrics, Defense R&D Canada*. 2004 [cited May 2008]; Available from: <http://www.ottawa.drdc-rddc.gc.ca>.
89. K. Bowyer, I. Kyong, P. Y. Chang, P. J. Flynn, E. Hansley, and S. Sarkar, *Multi-modal biometrics: an overview*. Second Workshop on Multi-Modal User Authentication, May 2006.

References

90. Guardia. *Products Main Page.* 2008 [cited May 2008]; Available from: <http://www.guardia.com>.
91. A. V. Agronskaia, L. Tertoolen, and H. C. Gerritsen, *Fast Fluorescence Lifetime Imaging of Calcium In Living Cells.* Journal of Biomedical Optics, 2004. **9**(6): pp. 1230-1237.
92. P. Schwille, U. Haupt, S. Maiti, and W. W. Webb, *Molecular Dynamics in Living Cells Observed by Fluorescence Correlation Spectroscopy with One- and Two-Photon Excitation.* Biophysics Journal, 1999. **77**: pp. 2251-2265.
93. A. Grinvald, D. Schmuel, D. Glaser, I. Vanzetta, E. Shtoyerman, H. Slovin, A. Sterkin, C. Wijnbergen, R. Hildesheim, and A. Arieli, *In-Vivo Optical Imaging of Cortical Architecture and Dynamics.* Modern Techniques in Neuroscience Reserach, ed. U. Windhorst and H. Johansson. 2001: Springer.
94. J. Fisher, E. F. Civillico, D. Contreras, and A. G. Yodh, *In Vivo Fluorescence Microscopy of Neuronal Activity in Three Dimensions by Use of Voltage-Sensitive Dyes.* Optics Letters, 2004. **29**(1): pp. 71-73.
95. S. Eisenberg, W. Reckers, and B. Wieneke, *Visualization and PIV Measurements of High-Speed Flows and Other Phenomena with Novel Ultra-High-Speed CCD Camera.* Proceedings of SPIE, 2003. **4948**: pp. 671-676.
96. W. Reckers, H. Schwab, C. Weiten, B. Befrui, and R. Kneer, *Investigation of Flame Propagation and Cyclic Combustion Variations in a DISI Engine using Synchronous High-Speed Visualization and Cylinder Pressure Analysis.* International Symposium fur Verbrennungdiagnostik, 2002: pp. 27-32.
97. A. Rochas, M. Gösch, A. Serov, P. A. Besse, R. S. Popovic, T. Lasser, and R. Rigler, *First Fully Integrated 2-D Array of Single-Photon Detectors in Standard CMOS Technology.* IEEE Photonics Technology Letters, 2003. **15**(7): pp. 963-965.
98. S. Cova, A. Longoni, and A. Andreoni, *Towards picosecond resolution with single-photon avalanche diodes.* Rev. Sci. Instruments, 1981. **52**(3): pp. 408-412.
99. E. R. Fossum, *Active pixel sensors: are CCDs dinosaurs?* Proceedings of SPIE Vol. 1900, Charge-Coupled Devices and Solid State Optical Sensors III, 1993: pp. 2-14.
100. R. G. W. Brown and M. Daniels, *Characterization of silicon avalanche photodiodes for photon correlation measurements. 3: Sub-Giger operation.* Applied Optics, 1989. **28**(21): pp. 4616-4621.
101. H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. MyIntyre, C. Trottier, and P. P. Webb, *Photon counting techniques with silicon avalanche photodiodes.* Applied Optics, 1993. **32**(21): pp. 3894-3900.
102. Hamamatsu Photonics. *Silicon APD product family.* 2007 [cited Dec. 2007]; Available from: <http://www.hamamatsu.com>.

103. A. I. Biber, *Avalanche Photodiode Image Sensing in Standard Silicon BiCMOS Technology*. 2000, PhD thesis dissertation Nr. 13544, Swiss Federal Institute of Technology Zurich.
104. L. Aina, A. Fathimulla, H. Hier, M. Lecates, P. Patel, and S. B. Foshee, *Non-Geiger-mode single-photon counting APDs with high detection probability and afterpulse-free performance*. Proceedings SPIE Vol. 6771, Advanced Photon Counting Techniques II, 2007.
105. A. S. Huntington, M. A. Compton, and G. M. Williams, *Linear-mode single-photon APD detectors*. Proceedings SPIE Vol. 6771, Advanced Photon Counting Techniques II, 2007.
106. PerkinElmer Optoelectronics. *SPCM-AQ Series Datasheet*. 2007 [cited Dec. 2007]; Available from: <http://opto.perkinelmer.com>.
107. IdQuantique. *id100 Series Datasheet* 2007 [cited Dec. 2007]; Available from: <http://www.idquantique.com>.
108. Micro Photon Devices. *PDM Series Datasheet*. 2007 [cited Dec. 2007]; Available from: <http://www.microphotondevices.com>.
109. R. H. Haitz, *Mechanisms contributing to the noise pulse rate of avalanche diodes*. Journal of Applied Physics, 1965. **36**(10): pp. 3123-3131.
110. S. M. Sze, *Physics of semiconductor devices*. 2nd ed. 1981: John Wiley & Sons.
111. W. Shockley and W. T. Read, *Statistics of the recombinations of holes and electrons*. Physical Review, 1952. **87**(5): pp. 835-842.
112. G. Vincent, A. Chantre, and D. Bois, *Electric field effect on the thermal emission of traps in semiconductor junctions*. Journal of Applied Physics, 1979. **50**(8): pp. 5484-5487.
113. G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, *A new recombination model for device simulation including tunneling*. IEEE Transactions on Electron Devices, 1992. **39**(2): pp. 331-338.
114. S. Cova, A. Lacaita, and G. Ripamonti, *Trapping phenomena in avalanche photodiodes on nanosecond scale*. IEEE Electron Device Letters, 1991. **12**(12): pp. 685-787.
115. A. Lacaita, M. Mastrapasqua, M. Ghioni, and S. Vanoli, *Observation of avalanche propagation by multiplication assisted diffusion in p-n junctions*. Applied Physics Letters, 1990. **57**(5): pp. 489-491.
116. A. Lacaita, S. Cova, A. Spinelli, and F. Zappa, *Photon-assisted avalanche spreading in reach-through photodiodes*. Applied Physics Letters, 1993. **62**(6): pp. 606-608.
117. G. Ripamonti and S. Cova, *Carrier diffusion effects in the time-response of a fast photodiode*. Solid-State Electronics, 1984. **28**(9): pp. 925-931.

| References

118. A. Spinelli and A. L. Lacaita, *Physics and numerical simulation of single photon avalanche diodes*. IEEE Transactions on Electron Devices, 1997. **44**(11): pp. 1931-1943.
119. R. Newman, *Visible Light from a Silicon p-n junction*. Physical Review, 1955. **100**(2): pp. 700-703.
120. R. H. Haitz, *Studies on optical coupling between silicon p-n junctions*. Solid-State Electronics, 1965. **8**: pp. 417-425.
121. R. H. Haitz, A. Goetzberger, R. M. Scarlett, and W. Shockley, *Avalanche effects in silicon p-n junctions. I. Localized Photomultiplication studies on microplasmas*. Journal of Applied Physics, 1963. **34**(6): pp. 1581-1590.
122. A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, *Avalanche effects in silicon p-n junctions. II. Structurally perfect junctions*. Journal of Applied Physics, 1963. **34**(6): pp. 1591-1600.
123. R. H. Haitz, *Model for the electrical behavior of a microplasma*. Journal of Applied Physics, 1964. **35**(5): pp. 1370-1376.
124. R. L. Batdorf, A. G. Chynoweth, G. C. Dacey, and P. W. Foy, *Uniform p-n junctions. I. Broad area breakdown*. Journal of Applied Physics, 1960. **31**(7): pp. 1153-1160.
125. A. G. Chynoweth and G. L. Pearson, *Effect of Dislocations on Breakdown in Silicon p-n Junctions*. Journal of Applied Physics, 1958. **29**(7): pp. 1103-1110.
126. S. M. Myers, M. Seibt, and W. Schroter, *Mechanisms of transition-metal gettering in silicon*. Journal of Applied Physics, 2000. **88**(7): pp. 3795-3819.
127. R. J. McIntyre, *The distribution of gains in uniformly multiplying avalanche photodiodes: theory*. IEEE Transactions on Electron Devices, 1972. **19**(6): pp. 703-713.
128. J. Conradi, *The distribution of gains in uniformly multiplying avalanche photodiodes: experimental*. IEEE Transactions on Electron Devices, 1972. **19**(6): pp. 713-718.
129. P. P. Webb, R. J. McIntyre, and J. Conradi, *Properties of avalanche photodiodes*. RCA Review, 1974. **35**: pp. 234-278.
130. S. Cova, M. Ghioni, F. Zappa, I. Rech, and A. Gulinatti, *A View on Progress of Silicon Single Photon Avalanche Diodes and Quenching Circuits*. Proceedings SPIE Vol. 6372, Advanced Photon Counting Techniques, 2006.
131. P. Antognetti, S. Cova, and A. Longoni, *A study of the operation and performances of an avalanche diode as a single photon detector*. Euratom Publications EUR 5370e, Proceedings 2nd Ispra Nuclear Electronics Symposium, 1975: pp. 453-456.
132. M. Ghioni, S. Cova, A. Lacaita, and G. Ripamonti, *New silicon epitaxial avalanche diode for single-photon timing at room temperature*. Electronics Letters, 1988. **24**(24): pp. 1476-1477.

133. A. Lacaita, M. Ghioni, and S. Cova, *Double epitaxy improves single-photon avalanche diode performance*. Electronics Letters, 1989. **25**(13): pp. 841-843.
134. 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, 1993. **14**(7): pp. 360-362.
135. S. Cova, A. Lacaita, M. Ghioni, and G. Ripamonti, *20-ps timing resolution with single-photon avalanche diodes*. Review of Scientific Instruments, 1989. **60**(6): pp. 1104-1110.
136. A. Gulinatti, P. Maccagnani, I. Rech, M. Ghioni, and S. Cova, *35 ps time resolution at room temperature with large area single photon avalanche diodes*. Electronics Letters, 2005. **41**(5): pp. 272-274.
137. E. Sciacca, A. C. Giudice, D. Sanfilippo, F. Zappa, S. Lombardo, R. Consentino, C. D. Franco, M. Ghioni, G. Fallica, G. Bonanno, S. Cova, and E. Rimini, *Silicon planar technology for single-photon optical detectors*. IEEE Transactions on Electron Devices, 2003. **50**(4): pp. 918-925.
138. M. Ghioni, A. Gulinatti, P. Maccagnani, I. Rech, and S. Cova, *Planar Silicon SPADs with 200um diameter and 35 ps photon timing resolution*. Proceedings of SPIE Vol. 6372, Advanced Photon Counting Techniques, 2006.
139. A. Restelli, I. Rech, P. Maccagnani, M. Ghioni, and S. Cova, *Monolithic silicon matrix detector with 50um photon counting pixels*. Journal of Modern Optics, 2007. **54**(2-3): pp. 213-223.
140. A. Spinelli, M. Ghioni, S. Cova, and L. M. Davis, *Avalanche detector with ultraclean response for time-resolved photon counting*. IEEE Journal of Quantum Electronics, 1998. **34**(5): pp. 817-821.
141. F. Zappa, A. Gulinatti, P. Maccagnani, S. Tisa, and S. Cova, *SPADA: single-photon avalanche diode arrays*. IEEE Photonics Technology Letters, 2005. **17**(3): pp. 657-659.
142. W. J. Kindt, N. H. Shahrjerdy, and H. W. v. Zeijl, *A silicon avalanche photodiode for single optical photon counting in the Geiger mode*. Sensors and Actuators A, 1997. **60**(1-3): pp. 98-102.
143. A. Rochas, G. Ribordy, B. Furrer, P.-A. Besse, and R. S. Popovic, *First passively-quenched single photon counting avalanche photodiode element integrated in a conventional CMOS process with 32ns dead time*. Proceedings of SPIE Vol. 4833, Applications of Photonic Technology 5, 2002: pp. 107-115.
144. A. Rochas, A. R. Pauchard, P.-A. Besse, D. Pantic, Z. Prijic, and R. S. Popovic, *Low-noise silicon avalanche photodiodes fabricated in conventional CMOS technologies*. IEEE Electron Device Letters, 2002. **49**(3): pp. 387-394.

145. A. Rochas, P.-A. Besse, and R. S. Popovic, *Actively recharged single photon counting avalanche photodiode integrated in an industrial CMOS process*. Sensors and Actuators A, 2004. **110**(1-3): pp. 124-129.
146. A. Rochas, A. Pauchard, L. Monat, A. Matteo, P. Trinkler, R. Thew, and R. Ribordy, *Ultra-compact CMOS single photon detector*. Proceedings SPIE Vol. 6372, Advanced Photon Counting Techniques, 2006.
147. R. M. Forsyth, *Technology and design of integrated circuits for up to 50V applications*. IEEE International Conference on Industrial Technology, 2003. **1**: pp. 7-13.
148. H. Finkelstein, M. J. Hsu, and S. Esener, *An ultrafast Geiger-mode single-photon avalanche diode in 0.18- μ m CMOS technology*. Proceedings of SPIE Vol. 6372, Advanced Photon Counting Techniques, 2006.
149. T. Mamamoto, *Sidewall damage in a silicon substrate caused by trench etching*. Applied Physics Letters, 1991. **25**(24): pp. 2942-2944.
150. H. I. Kwon, M. I. Kang, B.-G. Park, J. D. Lee, and S. S. Park, *The analysis of dark signals in the CMOS APS imagers from the characterization of test structures*. IEEE Transactions on Electron Devices, 2004. **51**(2): pp. 178-184.
151. S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, *Avalanche photodiodes and quenching circuits for single-photon detection*. Applied Optics, 1996. **35**(12): pp. 1956-1976.
152. F. Zappa, A. Lotito, A. C. Giudice, S. Cova, and M. Ghioni, *Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors*. IEEE Journal of Solid-State Circuits, 2003. **38**(7): pp. 1298-1301.
153. Austriamicrosystems, *0.35um 50V CMOS Process Parameters, Rev. 2.0*. 2004.
154. C. Niclass, M. Sergio, and E. Charbon, *A Single Photon Avalanche Diode Array Fabricated in Deep-Submicron CMOS Technology*. Design, Automation and Test in Europe (DATE), 2006. **1**: pp. 1-6.
155. Austriamicrosystems, *0.8um HV CMOS Process Parameters, Rev. 2.0*. 1999.
156. C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, *Design and Characterization of a CMOS 3-D Image Sensor based on Single Photon Avalanche Diodes*. IEEE Journal of Solid-State Circuits, 2005. **40**(9): pp. 1847-1854.
157. C. Niclass, M. Sergio, and E. Charbon, *A CMOS 64x48 Single Photon Avalanche Diode Array with Event-Driven Readout*. Proceedings of the 32nd European Solid-State Circuits Conference, Oct. 2006: pp. 556.
158. Austriamicrosystems, *0.35um 50V CMOS Design Rules, Rev. 2.0*. 2004.
159. C. Niclass, M. Sergio, and E. Charbon, *A single photon avalanche diode array fabricated in 0.35- μ m CMOS and based on an event-driven readout for TCSPC*

- experiments.* Proceedings of SPIE Vol. 6372, Advanced Photon Counting Techniques, 2006.
- 160. C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, *A 128x128 Single-Photon Imager with on-Chip Column-Level 10b Time-to-Digital Converter Array Capable of 97ps Resolution.* IEEE International Solid-State Circuits Conference, Feb. 2008. **1:** pp. 44-45.
 - 161. C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, *A 128x128 Single-Photon Imager with Column-Level 10b Time-to-Digital Converter Array.* IEEE Journal of Solid-State Circuits, to appear 2008.
 - 162. M. Cohen, F. Roy, D. Herault, Y. Cazaux, A. Gandolfi, J. P. Reynard, C. Cowache, E. Bruno, T. Girault, J. Vaillant, F. Barbier, Y. Sanchez, N. Hotellier, O. LeBorgne, C. Augier, A. Inard, T. Jagueneau, C. Zinck, J. Michailos, et al., *Fully Optimized Cu based process with dedicated cavity etch for 1.75 μ m and 1.45 μ m pixel pitch CMOS Image Sensors.* International Electron Devices Meeting (IEDM), 2006: pp. 1-4.
 - 163. C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, *A 130-nm CMOS single-photon avalanche diode.* Proceedings of SPIE, Optoelectronic Devices: Physics, Fabrication, and Application IV, 2007. **6766**.
 - 164. C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, *A Single Photon Avalanche Diode Implemented in 130-nm CMOS Technology.* IEEE Journal of Selected Topics in Quantum Electronics, 2007. **13**(4): pp. 863-869.
 - 165. A. Lacaita, F. Zappa, S. Cova, and P. Lovati, *Single-photon detection beyond 1 μ m: performance of commercially available InGaAs/InP detectors.* Applied Optics, 1996. **35**(16): pp. 2986.
 - 166. IdQuantique. *idQ-P Pulser Gated Active Quenching Circuit Datasheet.* 2007 [cited Dec. 2007]; Available from: <http://www.idquantique.com>.
 - 167. Z. Xiao, D. Pantic, and R. S. Popovic, *A New Single Photon Avalanche Diode in CMOS High-Voltage Technology.* Solid-State Sensors, Actuators and Microsystems Conference (Transducers), 2007: pp. 1365-1368.
 - 168. M. Gersbach, C. Niclass, E. Charbon, J. Richardson, R. Henderson, and G. Lindsay, *A Single Photon Detector Implemented in a 130nm CMOS Imaging Process.* 34th European Solid-State Device Research Conference (ESSDERC), to appear 2008.
 - 169. J. Massa, G. Buller, A. Walker, G. Smith, S. Cova, M. Umasuthan, and A. Wallace, *Optical design and evaluation of a three-dimensional imaging and ranging system based on time-correlated single-photon counting.* Applied Optics, 2002. **41**(6): pp. 1063-1070.
 - 170. S. Henzler, S. Koeppe, W. Kamp, H. Mulatz, and D. Schmitt-Landsiedel, *90nm 4.7ps-Resolution 0.7 LSB Single-Shot Precision and 19pJ-per-Shot Local Passive Interpolation Time-to-Digital Converter with On-Chip Characterization.* IEEE International Solid-State Circuits Conference, Feb. 2008: pp. 548-549.

171. G. S. Buller and A. M. Wallace, *Ranging and Three-Dimensional Imaging Using Time-Correlated Single-Photon Counting and Point-by-Point Acquisition*. IEEE Journal of Selected Topics in Quantum Electronics, 2007. **13**(4): pp. 1006-1015.
172. D. G. Fouche, *Detection and false-alarm probabilities for laser radars that use Geiger-mode detectors*. Applied Optics, 2003. **42**(27): pp. 5388-5398.
173. M. Rohner, K. Siercks, and J. Hinderling, *Single-Photon Detection for High Precision Ranging – A trade-off Study*. Proceedings of the 1st Range Imaging Research Day, ETH Zurich, 2005: pp. 33-42.
174. P. B. Coates, *The correction of photon pile-up in the measurement of radiative lifetimes*. Journal of Physics E: Scientific Instruments, 1968. **1**(2): pp. 878-879.
175. C. Niclass, M. Soga, and E. Charbon, *3D Imaging based on Single-Photon Detectors*. 8th Conf. on Optical 3D Measurement Techniques, Jul. 2007. **I**: pp. 34-41.
176. R. McCluney, *Introduction to Radiometry and Photometry*. 1994: Artech House.
177. C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, *A 128x128 Single-Photon Imager with on-Chip Column-Level 10b Time-to-Digital Converter Array*. IEEE Journal of Solid-State Circuits, 2008.
178. R. B. Staszewski, S. Vemulapalli, P. Vallur, J. Wallberg, and P. T. Balsara, *Time-to-Digital Converter for RF Frequency Synthesis in 90nm CMOS*. IEEE Radio Frequency Integrated Circuits Symposium, 2005: pp. 473-476.
179. International Electrotechnical Commission, *IEC60825-1 Safety of laser products - Part 1: Equipment classification, requirements and user's guide*. 2001.
180. SP Technical Research Institute of Sweden. *Contact page*. 2008 [cited May 2008]; Available from: <http://www.sp.se>.
181. C. Niclass, *Method and arrangement for measuring the distance to an object*. U.S. Patent Application 182,949, 2007.
182. C. Niclass, C. Favi, T. Kluter, F. Monnier, and E. Charbon, *Single-Photon Synchronous Detection*. 34th European Solid-State Circuits Conference (ESSCIRC), 2008.

LIST OF ACRONYMS

3D	Three-dimensional
ADC	Analog-to-digital converter
APD	Avalanche photodiode
AQ	Active quenching
AR	Active recharge
BG	Parasitic background light
CCD	Charge-coupled-device
CDMA	Code division multiple access
CMOS	Complementary metal oxide semiconductor
DCR	Dark count rate
DFT	Discrete Fourier transform
DN	Detector noise
DR	Dynamic range
DT	Dead-time
DUT	Device under test
FDMA	Frequency division multiple access
FFT	Fast Fourier transform
FOV	Field-of-view
FPGA	Field programmable gate array
FWHM	Full width at half maximum
G-R	Generation-recombination
H/V	Horizontal/vertical
HV	High-voltage
IC	Integrated circuit
IR	Infrared
IRF	Instrument response function
I-V	Current-to-voltage
JTAG	Joint test action group
LD	Laser diode
LED	Light emitting diode
MCT	Micro-channel plate
MOS	Metal oxide semiconductor
NBIF	Narrow-band interference filter
NMOS	N-type MOSFET transistor
OR	OR logic gate

| List of Acronyms

OVF	Overflow
PDP	Photon detection probability
PMOS	P-type MOSFET transistor
PMT	Photo-multiplier tube
PSD	Power spectral density
PQ	Passive quenching
PR	Passive recharge
QE	Quantum efficiency
SBR	Signal to background ratio
SG	Signal count rate
SPAD	Single-photon avalanche diode
SPSD	Single-photon synchronous detection
STI	Shallow-trench isolation
TAC	Time-to-amplitude
TCSPC	Time-correlated single-photon counting
TDC	Time-to-digital converter
TE	Thermoelectric
TOF	Time-of-flight
UV	Ultraviolet
W/L	Width-to-length ratio

LIST OF SYMBOLS

Symbol	Description	Unit
$b(\tau)$	Mean value of $b^*(\tau)$	[·]
$b^*(\tau)$	Background component of TCSPC histogram bin value at time τ	[·]
B_R	SPAD photon detection (counting) rate corresponding to the background component.	[Hz]
c	Speed of light in vacuum: 2.99792×10^{-8}	[m/s]
C_0	Counter value in SPSD based on 4 samples	[·]
C_1	Counter value in SPSD based on 4 samples	[·]
C_2	Counter value in SPSD based on 4 samples	[·]
C_3	Counter value in SPSD based on 4 samples	[·]
c_D	SPSD demodulation contrast	[%]
C_k	k^{th} counter (sample) value in SPSD	[·]
c_S	Modulation contrast of illumination source	[%]
C_S	Total capacitance seen by the active SPAD terminal	[fF]
erf	Gaussian error function	[·]
$f_{\#}$	Lens f-number	[·]
f_0	Repetition rate of pulse illumination or first harmonic of continuously modulated illumination.	[Hz]
F_{BW}	Narrow-band optical filter bandwidth	[nm]
$F_{PB}(\lambda)$	Narrow-band optical filter transmission as a function of wavelength	[%]
h	Planck constant: 6.62617×10^{-34}	[J s]
$h(\tau)$	Mean value of $h^*(\tau)$	[·]
$h^*(\tau)$	TCSPC histogram bin value at time τ	[·]
I_{BG}	Background light irradiance	[W/m ²]
$I_{BG,F}$	Pass-band filtered background irradiance	[W/m ²]
$I_{BG,FP}$	Background irradiance at focal plane	[W/m ²]
$I_{DS,1}$	Drain-to-source current in the quenching transistor	[A]
$I_{DS,2}$	Drain-to-source current in the recharge transistor (in dual-threshold front-end circuit)	[A]
I_{IMG}	Irradiance measured at focal (image) plane	[W/m ²]
I_o	Drain current used bias quenching transistor	[A]
I_{OBJ}	Irradiance measured on the target (object) plane	[W/m ²]
I_Q	SPAD latching current	[A]
I_{SG}	Illumination signal irradiance	[W/m ²]
$I_{SG,FP}$	Signal irradiance at focal plane	[W/m ²]
K	Boltzmann constant: 1.38066×10^{-23}	[J/K]
K_{RBB}	Radiation luminous efficacy	[lm/W]

List of Symbols

L_{BG}	Assumed background illuminance	[lux]
M	Mean number of signal photon detection per illumination cycle	[\cdot]
m	Lens magnification	[\cdot]
N_C	Number of counter (sample) values in SPSD	[\cdot]
N_{HW}	Half pulse width of signal pulse for centroid computation	[\cdot]
n_i	Intrinsic carrier concentration	[cm^{-3}]
N_P	Number of acquired illumination cycles	[\cdot]
N_{ROW}	Number of sharing rows in event-driven readout	[\cdot]
ϕ_{FF}	Pixel active area fill factor	[$\%$]
PDP_{SG}	Photon detection probability at λ_{SG}	[$\%$]
\tilde{P}_S	Distorted signal power due to pile-up effect as sampled by a TCSPC setup.	[W]
\bar{P}_S	Mean value of incident signal power falling on the SPAD active area.	[W]
P_S	Incident signal power falling on the SPAD active area	[W]
q	Elementary electron charge: 1.60218×10^{-19}	[C]
Q_P	Optical energy per illumination pulse falling on the detector in TCSPC	[J]
R_D	Space-charge and series resistance of the SPAD	[Ω]
$R_{DS,1}$	Drain-to-source resistance of quenching transistor in the triode region	[Ω]
R_o	Object/target reflectivity	[$\%$]
R_Q	Quenching resistor value	[Ω]
$s(\tau)$	Mean value of $s^*(\tau)$	[\cdot]
$s^*(\tau)$	Signal component of TCSPC histogram bin value at time τ	[\cdot]
$S'_{R,MIN}$	Approximation of $S_{R,MIN}$ for $\sigma_S \ll \Delta T$	[Hz]
S_R	SPAD photon detection (counting) rate corresponding to the signal component.	[Hz]
$S_{R,MAX}$	Maximum achievable S_R when limiting pile-up effects in TCSPC	[Hz]
$S_{R,MIN}$	Minimum required S_R in TCSPC to obtain reliable TOF evaluation	[Hz]
T	Frame integration (acquisition) time. Alternatively, absolute temperature	[s], [K]
T_0	Illumination period, the inverse of f_0 .	[s]
T_F	Narrow-band optical filter transmission in the pass-band	[$\%$]
T_L	Lens transmission	[$\%$]
T_{OFF}	Hold-off time used in front-end circuits to reduce afterpulsing effects	[s]
T_Q	Quenching duration with respect to the photon detection time	[s]
T_R	SPAD total detection (counting) rate. Alternatively, recharge time with respect to the photon detection time	[Hz], [s]
V_{BD}	Breakdown voltage	

V_{BIAS}	Quenching transistor gate-to-source bias voltage	[V]
V_{DD}	IC nominal power supply	[V]
$V_{DS,1}$	Drain-to-source voltage of quenching transistor	[V]
$V_{DS,2}$	Drain-to-source voltage of recharge transistor (in dual-threshold front-end circuit)	[V]
$V_{DS,SAT,1}$	Drain-to-source saturation voltage of quenching transistor	[V]
V_E	Excess bias voltage (dynamic)	[V]
$V_{E,MAX}$	Nominal excess bias voltage	[V]
$V_{GS,2}$	Gate-to-source voltage of the recharge transistor (in dual-threshold front-end circuit)	[V]
V_{OP}	SPAD bias voltage	[V]
V_{OUT}	Front-end output voltage	[V]
V_{RCH}	Recharge voltage used in dual-threshold front-end circuit	[V]
V_S	SPAD sensing voltage	[V]
$V_{SD,1}$	Source-to-drain voltage of (PMOS) quenching transistor	[V]
$V_{SD,SAT,1}$	Source-to-drain saturation voltage of (PMOS) quenching transistor	[V]
V_{TH}	Threshold voltage used to discriminate Geiger pulses on V_S	[V]
V_{TH1}	First threshold voltage in dual-threshold front-end circuit	[V]
V_{TH2}	Second threshold voltage in dual-threshold front-end circuit	[V]
z	Target distance with respect to the illumination system	[m]
α_{FV}	Emission field-of-view angle	[°]
δ_{PX}	Pixel pitch	[m]
ΔT	Time-to-digital resolution. Alternatively, PSD sampling window duration.	[s]
ΔTOF	Systematic TOF timing errors due to pile up effect in TCSPC	[s]
η_{SG}	SPAD photon detection probability at central emission wavelength of the illumination source	[%]
λ_F	Central wavelength of narrow band optical filter	[nm]
λ_{OFF}	Cut-off wavelength in Silicon	[nm]
λ_{SG}	Central wavelength of the illumination source	[nm]
μ_{Error}	Mean non-linearity distance error	[cm]
σ_{Error}	Repeatability distance error (r.m.s.)	[m]
σ_{IRF}	Instrument response function (r.m.s.)	[s]
σ_S	Half-pulse width in Gaussian shaped illumination pulse used in performance modeling of TCSPC measurements	[s]
σ_{SPAD}	SPAD timing resolution (r.m.s.)	[s]
σ_{TDC}	TDC timing jitter (r.m.s.)	[s]
σ_{TOF}	Predicted distance repeatability error in TCSPC	[m]
τ	Timing location of a TCSPC histogram bin	[ns]

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest acknowledgment to my advisor, Prof. Edoardo Charbon, for giving me the opportunity to work on three-dimensional image sensors in CMOS and for his technical expertise in the field. He allowed me to pursue this research work with freedom, in optimal conditions, and virtually free of financial constraints. His guidance during the last four years is highly acknowledged. I am particularly grateful to Prof. E. Charbon for introducing me to the secrets of writing scientific publications with efficiency. I also thank Prof. E. Charbon for his remarkable enthusiasm throughout these years as well as for his great availability, in spite of his demanding calendar.

I am deeply indebted to Prof. Radivoje S. Popovic, head of the Microsystems Design Group, for hosting me as invited researcher in his group during entire duration of this investigation. Prof. R. S. Popovic was at the origin of using CMOS technologies to implement single-photon detectors. The results of this research highly benefited from his vision and expertise. I also thank Prof. R. S. Popovic for kindly serving as a co-examiner of this thesis.

During these years, I had the opportunity to work with many outstanding people. I would like to express my gratitude to Dr. Pierre-André Besse for suggesting my name for this position. Dr. P.-A. Besse has contributed directly and indirectly to many aspects of this thesis. In particular, his contributions to a number of scientific articles as well as his critical inputs are highly appreciated. His analytical and honest manner of conducting research and presenting results will certainly be a solid base for my future career. I am deeply grateful to my former colleague, Dr. Alexis Rochas, for his outstanding contributions in the field of single-photon detection, one of the fundamental pillars of this thesis. He kindly taught me numerous theoretical and practical aspects involved in the design and characterization of single-photon detectors. His expertise in device research as well as the countless discussions we had together directly contributed to the achievements of this research. Dr. A. Rochas also co-examined this thesis with remarkable professionalism. I would like to thank Prof. Peter Seitz of University of Neuchâtel for accepting to co-examine this thesis. It was a great satisfaction to be examined by the co-inventor and principal researcher of the first solid-state TOF 3D camera ever built. Prof. P. Seitz always gave fruitful inputs to my thoughts whenever I had the opportunity

to meet him in scientific conferences. I also thank Prof. Giovanni De Micheli for being the president of my thesis jury, in spite of his tight schedule.

I would like to express my deep gratitude to all the members of the Microsystems Design Group (LMIS3) as well as of the Processor Architecture Lab (LAP). It was a pleasure and a continuous source of inspiration for me to be involved in two distinct research groups. I would like to particularly thank my officemate, Zhen Xiao, for his company and friendship. Zhen has been a great discussion partner in many subjects. In relation with this thesis, I highly acknowledge Zhen for the numerous discussions that we shared involving the design of mixed-signal and analog circuits, and photodiode devices. I also thank all my colleagues at LMIS3 for keeping an enjoyable and open-minded working environment. In particular, I am grateful to Marc, Giovanni, Pavel, Serge, Malika, Klaus, Ozge, Mirjana, Jens, and Bernard for sharing great moments together. I would like to specially thank Marc Lany for his outstanding inputs and his great availability to help in any circumstance. At LAP, I warmly thank all my current and past colleagues, especially Prof. P. Ienne, René, Frédéric, Claudio F., Theo, Huan, Marek, Estelle, Maximilian, Yuki, Claudio B., and Dmitri. I especially thank Claudio Favi and Theo Kluter for their wonderful help in the design of some readout circuits used in the thesis. Besides my lab members and supervisors, some people have spent precious time in design reviews and deserve explicit acknowledgements. I sincerely thank Dr. Beat Furrer for sharing his large experience in the design of circuits with me. I also thank Dr. Paul Muller for spending his time in some design reviews as well as for his friendship and encouragements. Ilona Modoux and Chantal Schneeberger, the secretaries of LMIS3 and LAP respectively, are particularly acknowledged for their enormous assistance and support in administrative matters during all these years.

I am indebted to a number of technical staff and students who directly helped in the design and test of many prototypes. I would like to thank Peter Bruehlmeier and George Vaucher of Acort for drawing the layout of a lot of PCBs designed during this thesis. It was a pleasure to work with the person who has mounted the very first modern mouse device, André Guignard. He has helped me with a number of optical and electro-mechanical prototypes. His expertise and kindness are highly acknowledged. Regarding the IT people, I would like to warmly thank Dr. Alain Vachoux for his professionalism and efficiency in the installation and maintenance of design kits and EDA tools. I also thank Yves Pilloud, Dr. Pavel Kejik, and Rodolphe Buret for the effective IT support. Special acknowledgements to Frédéric Monnier for his contributions in the design of the illumination system mounted on the SPSD prototype. Frédéric has also

helped me in the characterization of range performance of some sensors. Honoré Takeugming is kindly acknowledged for helping in some FPGA designs.

I would like to express my deep gratitude to all the industrial partners and institutions that supported this research. I would like to acknowledge the Swiss National Science Foundation for funding most of this research. I would also like to acknowledge the Sixth Framework Programme of EU for funding part of this research through the Megraframe Project. I am deeply indebted to the collaborators of the EPFL's Industrial Relations Office for their financial support and assistance. Particularly, I am grateful to Gabriel Clerc, Mehdi Jacques Aminian, and Hervé Lebret. Mehdi has been a great advisor in all IP-related subjects. Hervé is highly acknowledged for providing me with an "Innogramant", which has been utilized to fund the SPSD prototype. Although they prefer to stay anonymous, a number of industrial partners have supported this research and are highly acknowledged. I would like to explicitly express my gratitude to Mineki Soga for his enormous help and contributions all over the years.

I am indebted to the collaborators of Ingenico (Suisse) SA for providing me with the opportunity to work as a part-time engineer in a different environment. Besides the financial support, this position has been a great opportunity for me to meet highly focused engineers and managers. In many situations, that experience allowed me to have a complementary and more practical view in some subjects of my research work. I am particularly grateful to Pierre Bays for his advices and encouragements as well as to Jacques, Jean-Nicolas, Vincent, Martin, Reto, Oliver, Enrico, and Julian for our memorable collaboration over the years. I would like to sincerely thank Prof. Abbas El Gamal, head of the Information Systems Laboratory of Stanford University, for kindly hosting me in his group during the summer of 2005. This collaboration in Silicon Valley was an inestimable social and technical experience. I specially thank Sam Kavusi for the countless discussions that we shared on image sensors and for his natural kindness. I also thank Kunal Ghosh for sharing his office with me and for the pleasant time we spent together. Finally, I am grateful to all my friends that contributed directly and indirectly to this achievement. Special thanks to Marcelo and Roberta.

I would like to dedicate this thesis to my wife Grazieli for her love and encouragement during all these years, to my parents Léon and Carmen for their constant support and for allowing me to attend good schools, to my brothers Léon and Nicolas, to my sisters Angela and Angelita, to my mother-in-law Aurea, and to our godson Raphaël.

ABOUT THE AUTHOR

Cristiano Niclass was born in a historical town called Ibirama, located in the state of Santa Catarina, south of Brazil, in 1975. He followed his Swiss origins and received the M. Sc. degree in microtechnology from the Swiss Federal Institute of Technology (EPFL), Lausanne, in 2003. After his graduation, he worked as a part-time R&D engineer until May 2006, for Ingenico (Suisse) SA, in the design and prototyping of electronic payment terminals. In May 2003, he joined the Processor Architecture Laboratory of EPFL and subsequently the Quantum Architecture Group, where he worked toward the Ph.D. degree. At EPFL, he has investigated the design, implementation, and characterization of fully integrated image sensors in CMOS using single-photon avalanche diodes. He has also been involved in the design of high-speed and high-resolution data converters implemented in conventional technologies. In 2005, Cristiano Niclass has spent two months in the group of Prof. Abbas El Gamal, at the Information System Laboratory of Stanford University, where he has studied conventional image sensors in CMOS. In 2008, he obtained his Ph.D. degree from EPFL for this work on single-photon image sensors in CMOS.

Dr. Niclass has authored and co-authored nearly 30 peer-reviewed journal and conference publications. He is also the inventor and co-inventor in seven patent applications. Dr. Niclass is a member of the Institute of Electrical and Electronics Engineers (IEEE) and of the International Society for Optical Engineering (SPIE). His interests include high-speed and low-noise mixed-signal ICs with emphasis on high-performance imaging.

LIST OF PUBLICATIONS

Journal Articles as First Author

1. C. Niclass, C. Favi, T. Kluter, M. Gersbach, E. Charbon, **A 128x128 Single-Photon Imager with Column-Level 10b Time-to-Digital Converter Array**, IEEE Journal of Solid-State Circuits, December 2008.
2. C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, **A Single Photon Avalanche Diode Implemented in 130nm CMOS Technology**, IEEE Journal of Selected Topics in Quantum Electronics, July/August 2007.
3. C. Niclass, A. Rochas, P.A. Besse, R.S. Popovic, E. Charbon, **A 4 μ s Integration Time Imager Based on CMOS Single Photon Avalanche Diode Technology**, Sensors and Actuators A: Physical, 2006.
4. C. Niclass, A. Rochas, P.A. Besse, E. Charbon, **Design and Characterization of a CMOS 3D Image Sensor based on Single Photon Avalanche Diodes**, IEEE Journal of Solid-State Circuits, September 2005.
5. C. Niclass, A. Rochas, P.A. Besse, E. Charbon, **Towards a 3D Camera Based on Single Photon Avalanche Diodes**, IEEE Journal of Selected Topics in Quantum Electronics, Vol. 10, No. 4, July/August 2004.

Journal Articles as Co-author

6. M. Gersbach, D. L. Boiko, C. Niclass, C. Petersen and E. Charbon, **Fast Fluorescence Dynamics in Nonratiometric Calcium Indicators**, submitted to Applied Physics Letters, arXiv: 0807.2756, July 2008.
7. E. Grigoriev, A. Akindinov, M. Breitenmoser, S. Buono, E. Charbon, C. Niclass, I. Desforges, R. Rocca, **Silicon Photomultipliers and their Bio-medical Applications**, Nuclear Instruments & Methods in Physics Research A, Vol. 571 N. 1+2, pp. 130-133, January 2007.

Conference Articles as First Author

8. C. Niclass, C. Favi, T. Kluter, F. Monnier, E. Charbon, **Single-Photon Synchronous Detection**, 34th European Solid-State Circuits Conference (ESSCIRC), to appear 2008.
9. C. Niclass, C. Favi, T. Kluter, M. Gersbach, E. Charbon, **A 128x128 Single-Photon Imager with on-Chip Column-Level 10b Time-to-Digital Converter Array Capable of 97ps Resolution Resolution**, IEEE International Solid-State Circuits Conference (ISSCC), February 2008.

10. C. Niclass, M. Soga, and E. Charbon, **3D Imaging based on Single-Photon Detectors (INVITED)**, 8th Conf. on Optical 3D Measurement Techniques, ETH Zurich, July 2007.
11. C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, **A 130nm CMOS Single-Photon Avalanche Diode (INVITED)**, SPIE Optics East Conference, Sept. 2007.
12. C. Niclass, M. Sergio, and E. Charbon, **A single-photon avalanche diode array fabricated in 0.35um CMOS and based on an event-driven readout for TCSPC experiments**, SPIE Optics East Conference, October 2006.
13. C. Niclass, M. Sergio, and E. Charbon, **A CMOS 64x48 Single Photon Avalanche Diode Array with Event-Driven Readout**, IEEE European Solid-State Circuits Conference (ESSCIRC), September 2006.
14. C. Niclass, M. Sergio, and E. Charbon, **A Single Photon Avalanche Diode Array Fabricated in Deep-Submicron CMOS Technology**, Design, Automation and Test in Europe (DATE), March 2006
15. C. Niclass, P.A. Besse, E. Charbon, **Arrays of Single Photon Avalanche Diodes in CMOS Technology: Picosecond Timing Resolution for Range Imaging (INVITED)**, 1st Range Imaging Research Day, September 2005
16. C. Niclass, A. Rochas, P.A. Besse, R. S. Popovic, E. Charbon, **CMOS Imager Based on Single Photon Avalanche Diodes**, The 13th International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers), June 2005.
17. C. Niclass, E. Charbon, **A CMOS Single Photon Detector Array with 64x64 Resolution and Millimetric Depth Accuracy for 3D Imaging**, IEEE International Solid-State Circuits Conference (ISSCC), February 2005.
18. C. Niclass, A. Rochas, P.A. Besse, E. Charbon, **A CMOS 3D Camera with Millimetric Depth Resolution**, IEEE Custom Integrated Circuits Conference (CICC), pp. 705-708, October 2004.
19. C. Niclass, A. Rochas, P.A. Besse, E. Charbon, **A CMOS Single Photon Avalanche Diode Array for 3D Imaging**, IEEE International Solid-State Circuits Conference (ISSCC), pp. 120-121, February 2004.

Conference Articles as Co-author

20. M. Gersbach, C. Niclass, L. Carrara, M. Sergio, N. Scheidegger, H. Shea, E. Charbon, **A Study of the Effects of Gamma Radiation on CMOS Single-Photon Avalanche Diodes**, IEEE Sensors, to appear 2008.
21. M. Gersbach, C. Niclass, E. Charbon, J. Richardson, R. Henderson, L. Grant, **A Single Photon Detector Implemented in a 130nm CMOS Imaging Process**, 34th European Solid-State Device Research Conference (ESSDERC), to appear 2008.
22. Z. Xiao, C. Niclass, E. Charbon, and R. S. Popovic, **Integrated Single Photon Avalanche Detectors (INVITED)**, International Conference on Lasers, Applications, and Technologies, May 2007.

23. M. Gersbach, C. Niclass, D. Boiko, C. Petersen, and E. Charbon, **CMOS Single Photon Sensors for Bioimaging**, EOS Conference on Frontiers in Electronic Imaging, June 2007.
24. U. Lehmann, M. Sergio, S. Pietrocola, C. Niclass, E. Charbon, and M.A.M. Gijs, **A CMOS Microsystem Combining Magnetic Actuation and In-situ Optical Detection of Microparticles**, The 14th International Conference on Solid-State Sensors, Actuators and Microsystems, June 2007.
25. M. Sergio, C. Niclass, E. Charbon, **A 128×2 CMOS Single-Photon Streak Camera with Timing-Preserving Latchless Pipeline Readout**, IEEE International Solid-State Circuits Conference (ISSCC), February 2007.
26. N. Scheidegger, R. Krpoun, H. Shea, C. Niclass, and E. Charbon, **A new Concept for a Low-Cost Earth Sensor: Imaging Oxygen Airglow with Arrays of Single Photon Detectors**, 30th Annual AAS Guidance and Control Conference, February, 2007.
27. M. Gersbach, Y. Maruyama, C. Niclass, K. Sawada, and E. Charbon, **A Room Temperature CMOS Single Photon Sensor for Chemiluminescence Detection**, Conference on Miniaturized Systems for Chemistry and Life Sciences (MicroTAS), November, 2006.
28. M. Gersbach, C. Niclass, M. Sergio, D. L. Boiko, C. Petersen, E. Charbon, **Time-Correlated Fluorescence Microscopy Using a Room Temperature Solid-State Single Photon Sensor**, 9th International conference on near-field optics, nanophotonics and related techniques (NFO-9), September 2006.
29. E. Grigoriev, A. Akindinov, M. Breitenmoser, S. Buono, E. Charbon, I. Desforges, C. Niclass, and R. Rocca, **Silicon Photomultipliers and their bio-medical applications**, 1st European Conference on Molecular Imaging Technology, May 2006.

Patent Applications

1. C. Niclass, US Application Patent 7'182'949, Distance measuring arrangement for use in rangefinder system, has solid-state image sensor with array of avalanche photodiodes and circuits that process signals output by photodiodes to yield data depending on photonic wave, December 2006.
2. C. Niclass and E. Charbon, US Application Patent 6'202'129, Integrated circuit for e.g. virtual keyboard, has array of single photon avalanche diodes in which each of avalanche diodes is coupled to one read-out circuit, where avalanche diodes are coupled to one of read-out circuits, February 2006.
3. E. Charbon and C. Niclass, US Application Patent 6'202'121, Bidimensional or three-dimensional information reading transducer, has solid-state single photon detector for acquisition of image of portion of carrier, and read-out circuit for processing signal delivered by detector, February 2006.
4. C. Niclass, E. Charbon, A. Rochas, and P.A. Besse, US Application Patent 6'192'086, Integrated imager circuit, has set of single photon avalanche diodes connected to

2D readout circuits to determine intensity of light reflected by scene by counting number of photons received by diode during period of time, February 2006.

5. E. Charbon and C. Niclass, US Application Patent 6'131'480, Spectral response controlling method of light sensitive semiconductor element for image sensor, involves controlling response by applying control signal to element, December 2005.
6. P. Bays, J. Bays, C. Niclass, E. Petraglio, R. Straumann, European Application Patent 1'788'507, Electronic processing terminal e.g. payment terminal, data e.g. password, protecting method for e.g. payment transaction, involves making selection between operating mode of electronic processing terminal and non-secured operating mode, November 2006.

