Near-Infrared Optical Tomography with Single-Photon Avalanche Diode Image Sensors

THÈSE N^{O} 6481 (2015)

PRÉSENTÉE LE 10 FÉVRIER 2015

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR

GROUPE SCI STI EC

PROGRAMME DOCTORAL EN PHOTONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Juan MATA PAVIA

acceptée sur proposition du jury:

Prof. O. Martin, président du jury
Prof. E. Charbon, Prof. M. Wolf, directeurs de thèse
Dr D. Boiko, rapporteur
Prof. M. Frenz, rapporteur
Prof. P. Seitz, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

To my family…

Abstract

In the past decades different methods based on magnetic resonance imaging (MRI) and positron emission tomography (PET) have been developed to image a variety of hemodynamic parameters in tissue. However, simultaneous high resolution imaging of oxy- and deoxyhemoglobin is not yet possible in clinical practices with millimeter accuracy. Near-infrared Spectroscopy (NIRS) is a well established technique that has been widely utilized in research and clinical practice to monitor the concentration of different chromophores in blood. It is capable of measuring a variety of parameters such as oxygenation levels in tissue, hematocrit, and cytochrome c oxidase levels among others. By increasing the amount and complexity of the light sources and detectors integrated in NIRS systems, in the last years researchers have been able to acquire tomographic images of such chromophores. This new imaging modality is known by different names: near-infrared optical tomography (NIROT), optical tomography (OT), diffuse optical tomography (DOT), or diffuse optical imaging (DOI). One of the main factors that limited the quality of the tomographic images obtained with NIROT has been the low number of sources and detectors employed in NIROT systems.

The integration of the first single-photon avalanche diode (SPAD) together with time-to-digital converters (TDCs) in CMOS enabled a whole new range of possibilities in the field of single-photon detection. In this thesis the application of a SPAD-TDC image sensor in NIROT is presented for the first time. The main objective was to develop a new system that could perform acquisitions nearly in real time and that was capable of delivering tomographic images in a short period of time for medical evaluation. A new optical setup was conceived based on this detector to take advantage of the large amount of information delivered by the SPADs. By employing line-shaped illumination sources instead of point like sources, the target is more homogeneously illuminated and consequently a reduced number of sources is necessary. It was experimentally demonstrated that a resolution of 5 mm is possible with this

new NIROT system.

New algorithms that reduced the ill-posed nature of the image reconstruction problem were developed thanks to the wide-field time-resolved measurements delivered by the SPAD image sensors. The large datasets obtained with our NIROT system and its time-resolved capabilities enabled the development of fast methods capable of reconstructing an image with millimeter resolution in a few seconds with a laptop computer.

We also analyze the effect of microlenses on the light sensitivity of the image sensor, showing that it is possible to increase it by a factor of 10 under certain conditions.

A new image sensor with 4×400 pixels implemented in a 3D CMOS technology for NIROT applications is also presented. To our knowledge, this is the first backside illuminated SPAD array that incorporates in-pixel TDCs. A novel TDC architecture was introduced that reduces the energy consumed per conversion, consequently allowing a high number of TDCs working in parallel.

The study of how NIROT can benefit from SPAD image sensors is presented in this thesis. Despite their current limitations, they enable the implementation of systems with thousands of detectors capable of millimeter resolution. Further developments in SPAD array architectures and TDCs will continue improving the performance of time-resolved NIROT systems.

Key words: Near-infrared Spectroscopy (NIRS), Near-infrared Optical Tomography (NIROT), Optical Tomography (OT), Diffuse Optical Tomography (DOT), Single-photon Detector, Single-photon Avalanche Diode (SPAD),Time-resolved Imaging, Time-to-digital Converter (TDC), Inverse Problems, Microlenses, Time-correlated Single-Photon Counting (TCSPC).

Zusammenfassung

Obwohl große Fortschritte in der medizinischen Bildgebung seit den letzten Jahren zu verzeichnen sind, gibt es bisher noch keine etablierte bildgebende Methode um die Gewebeoxigenierung mit einer Auflösung im Millimeterbereich zu messen. Unterschiedliche Methoden, wie z.B. die Magnetresonanztomographie (MRT) und Positronen-Emissions-Tomographie (PET) wurden zwar in den letzten Jahren entwickelt um unterschiedliche hämdynamische Parameter zu bestimmen, jedoch ist eine hoch aufgelöste simultane bildgebende Messung von Oxy- und Deoxihämglobin für klinische Zwecke noch nicht verfügbar. Die Nahinfrarot-Spektroskopie (NIRS) ist indes eine gut etablierte Technik die schon im Bereich der Forschung und klinischen Praxis eingesetzt wird da mit ihr die Konzentration von verschiedenen Chromophoren bestimmt werden kann, sowie andere Parameter wie z.B. unter anderem die Gewebesauerstoffsättigung, der Hämatokrit-Wert und die Konzentration der Cytochrome-c-Oxidase.

In den letzten Jahren konnte durch verschiedene Forschungsgruppen gezeigt werden, dass es möglich ist tomografische Bilder solcher Chromophoren-Konzentrationen zu ermitteln wenn die Anzahl und Komplexität der Lichtquellen und Detektoren in integrierten NIRS-Systemen erhört wird. Diese neue bildgebende Technik ist unter verschiedenen Bezeichnungen bekannt: „near-infrared optical tomography“ (NIROT), „optical tomography“ (OT), „diffuse optical tomography“ (DOT) oder „diffuse optical imaging“ (DOI). Einer der Hauptgründe warum die Qualität der Bildgebung mit NIROT bisher limitiert war ist die Tatsache, dass nur wenig Lichtquellen und Detektoren verwendet wurden.

Die Zusammenführung von Einzelphotonen-Avalanche-Photodioden-Detektoren (SPAD) mit Zeit-Digital-Wandlern (ZDW) mit der CMOS-Technik ermöglicht eine Vielzahl an neuen Möglichkeiten der Einzelphotonen-Detektion. In dieser Dissertation wird die Realisierung und Anwendung eines SPAD-ZDW Bildsensors erstmals vorgestellt. Das Hauptziel dieser Arbeit war es ein neues System zu entwickeln, dass die Bilderfassung in nahezu Echtzeit

ermöglicht und gleichzeitig die Bildrekontstruktion in einer relativ kurzen Zeit durchführt um medizinische Anwendungen zu ermöglichen. Ein neuer optischer Aufbau wurde basierend auf diesem Detektor entwickelt welcher den Vorteil der großen Datenmenge, die durch die SPADS geliefert wird, voll ausnützt. Durch die Verwendung einer linienförmigen Beleuchtung, anstatt einer punktförmigen, wird das Zielobjekt homogener beleuchtet, wodurch letztlich weniger Einzel-Lichtquellen notwendig werden. Es konnte experimentell gezeigt werden, dass eine Auflösung bis 5 mm mit diesem neuen NIROT-System möglich ist.

Neue Algorithmen, die die Problematik des schlecht gestellten Rekonstruktionsproblems mindern, konnten entwickelt werden da die grossflächige zeitaufgelöste Bilderfassung mittels des SPAD-Systems hierfür neue Möglichkeiten bietet. Die große Anzahl an gemessenen Informationen und die gute Zeitauflösung des NIROT-Systems ermöglichten zudem die Entwicklung von zeitsparenden Methoden um eine Bildrekonstruktion im Millimeterbereich vorzunehmen, sodass nun ein Bild innerhalb weniger Sekunden mit einem Laptop berechnet werden kann.

Die Auswirkung der Anbringung von Mikrolinsen auf die Lichtsensitivität des Bildsensors wurde untersucht mit dem Ergebnis, dass unter bestimmten Bedingungen eine Verstärkung um den Faktor 10 dadurch erzielt werden kann.

In der Dissertation wird des Weiteren ein neuer NIROT-Bildsensor mit einer Auflösung von 4×400 Pixeln vorgestellt. Der Sensor ist implementiert in einer 3D CMOS-Technologie. Unserem Wissen nach ist es der erste SPAD-Sensor mit Rückseitenbeleuchtung welcher Pixelbasierte ZDWs verwendet. Zudem wird eine neue ZDW-Architektur vorgestellt die es ermöglicht weniger Energie für die Konvertierung zu verwenden, was es dadurch wiederum möglich macht eine hohe Anzahl von ZDWs gleichzeitig parallel zu betreiben.

In der Dissertation wird auch darauf eingegangen wie NIROT durch die Verwendung von SPAD Bildsensoren profitieren kann. Abgesehen von den gegenwärtigen Limitierungen bietet dieser Ansatz der Konstruktion von Systemen mit tausenden von Detektoren eine Bildauflösung im Millimeterbereich. Es kann mit Sicherheit davon ausgegangen werden, dass die weiter Entwicklungen der SPAD- und ZDW-Technologie in der Zukunft die Leistung von zeitaufgelösten NIROT-Systemen weiter erhöhen wird.

Stichwörter: Nahinfrarot-Spektroskopie (NIRS), Near-infrared Optical Tomography (NIROT), Optical Tomography (OT), Diffuse Optical Tomography (DOT), Einzelphotonen-Detektor, Einzelphotonen-Avalanche-Photodioden-Detektoren (SPAD), zeitaufgelöste Bildgebung, Zeit-Digital-Wandler (ZDW), inverse Probleme, Mikrolinsen, zeitkorrelierte Einzelphotonen-Zählung.

Contents

List of Figures

List of Tables

List of Acronyms

APD	avalanche photodiode
ART	algebraic reconstruction technique
ASIC	application specific integrated circuit
BSI	back side illuminated
CG	conjugate gradient
CMOS	complementary metal-oxide-semiconductor
CPS	counts per second
CT	computed tomography
CW	continuous wave
DAC	digital-to-analog converter
DCR	dark count rate
DNL	differential nonlinearity
DNWELL	deep n-well
DOI	diffuse optical imaging
DOT	diffuse optical tomography
FD	frequency domain
FEM	finite element methods
FWHM	full width at half maximum

GPU	graphical processor unit
Hb	hemoglobin
HHb	deoxygenated hemoglobin
INL	integral nonlinearity
IRF	impulse response function
LSB	least-significant bit
LSQR	least square QR
MC	Monte Carlo
MCP	microchannel plate
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
NIR	near-infrared region
NIRI	near-infrared imaging
NIROT	near-infrared optical tomography
NIRS	near-infrared spectroscopy
NLDD	N low doped drain
O_2Hb	oxygenated hemoglobin
OT	optical tomography
PET	positron emission tomography
PDE	photon detection efficiency
PDP	photon detection probability
PMT	photomultiplier tube
RTE	radiative transfer equation
SiPM	silicon photomultiplier
SNR	signal to noise ration
SPAD	single-photon avalanche diode

SRAM	static random access memory
SP-LSQR	subspace preconditioned least square QR
STI	shallow trench isolation
SVD	singular value decomposition
TCSPC	time-correlated single-photon counting
TD	time domain
TDC	time-to-digital converter
TSV	through-silicon via
TSVD	truncated singular value decomposition
WTA	winner-take-all

1 Introduction

In order to describe how NIROT is capable of imaging oxygenation states of different body parts, it is necessary to understand how light interacts with tissue. In this chapter the mechanisms that determine the light transport in biological tissue are presented. The different techniques that so far have been utilized to measure their optical properties have been constrained by the type of light sources and detectors employed. An overview on the available light detectors used in NIROT systems and how SPADs integrated in CMOS technology change the current paradigm is also introduced.

1.1 Light transport in tissue

Light transport in biological media is mainly governed by two mechanisms, scattering and absorption [Tuchin, 2007]. Light scattering in tissue is produced by changes of the refractive index at the boundary between structures present in the media. Such structures are cell membranes, cell nucleus, mitochondria, and other intracellular organisms. The scattering caused by these structures can be classified in two types, Rayleigh and Mie scattering. Rayleigh scattering is produced by structures that are much smaller than the light wavelength, and it causes isotropic scattering. It is strongly wavelength dependent varying with λ^{-4}, where λ is the wavelength of the light. In human tissue, Rayleigh scattering is produced by structures smaller than 200nm such as viruses, collagen fibrils and other intracellular organelles [Tuchin,

2007, Jacques, 2013, Björn, 2007]. Structures of the same size or larger than the wavelength are the cause of Mie scattering. This is type of scattering is characterized by its anisotropy, it scatters stronger in the forward than in the reverse direction, and by having a weak wavelength dependence. It is mainly caused by the refraction index mismatch between different interfaces such as extracellular membrane, cytosol, cell nucleus, and mitochondria, to name a few [Tuchin, 2007, Jacques, 2013, Björn, 2007]. When light travels through most biological tissue, it undergoes a combination of Mie and Rayleigh scattering. However for wavelengths above 600 nm, the Mie scattering dominates. This means that light entering biological medium undergoes several scattering events before completely losing its directionality as is it shown in Fig. 1.3. The scattering coefficient μ_s signifies the probability that a photon is scattered per unit length, and determines the propagation speed of light in the medium. Another parameter that also characterizes scattering media is the so called reduced scattering coefficient μ_s', which determines the probability per unit length that a photon completely loses its directionality and its direction becomes random.

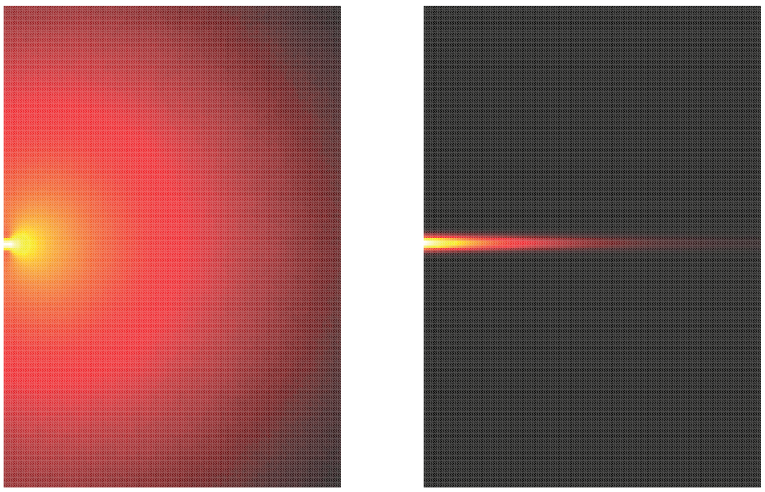

Figure 1.1 – (left) A collimated beam of light hits a highly scattering medium that scatters the photons mostly in the forward direction. After a certain distance the beam completely loses its directionality. (right) Simulation of a collimated light beam entering an absorbing non-scattering medium. Both pictures have been obtained with Monte Carlo simulations based on radiative transfer theory.

When light travels through a purely absorbing, non-scattering medium, its intensity decays at an exponential rate with the traveled distance (Fig. 1.3). In a medium composed by different

chromophores, the total absorption coefficient μ_a, defined as the probability per unit length of a photon being absorbed, is determined by the concentration and absorption coefficients of each of them. In particular, the most abundant chromophore in the human body is water. Depending on tissue and individual, it can account for 60% to 90% of the tissue's volume. Water strongly absorbs light below 200 nm and above 900 nm, however, between these two wavelengths its absorption is very low compared to other chromophores present in most tissues. Lipids have a similar optical window to water, being fairly transparent for wavelengths below 900 nm. However, the concentration of lipids depends highly on the type of tissue: most of the adipose tissue is subcutaneous or it is located between internal organs. In the optical region where water and lipids absorb less, hemoglobin (Hb) in its oxygenated (O_2Hb) and deoxygenated (HHb) forms are the main absorbers in tissue. Hb is carried by the red blood cells which constitute about 45% of whole blood volume. Hb is mainly responsible for the oxygen delivery from the lungs to the rest of the body. For wavelengths above 650 nm Hb does not absorb much light. In comparison its absorption coefficient is two orders of magnitude lower than for wavelengths below 600 nm. In the range between 650 nm and 900 nm, a subregion of the near-infrared range (NIR), the light absorption in tissue is minimal (see Fig. 1.2), this is the so called biological optical window. Within this range of wavelengths light penetrates several centimeters deep in the tissue before being absorbed, allowing the study of the optical properties of deep tissue. Although there are other chromophores that absorb light in the NIR such as melanin, cytochrome c oxidase or other forms of hemoglobin, their contribution to the total absorption is by far smaller compared to O_2Hb and HHb, and therefore they are not usually taken in consideration in the study of light absorption in tissue.

The above mentioned chromophores have a unique spectrum (see Fig. 1.2), therefore it is possible to calculate their concentrations in tissue from their absorption coefficients at different wavelengths. Since tissue is a highly scattering and absorbing medium, measuring its optical properties is a challenging task that requires disentangling the effects of absorption and scattering.

In most biological soft tissue, the scattering is much larger than the absorption, and the way photons travel through it can be modeled as a diffusive process. Fig. 1.3 shows an example of absorption and scattering coefficients measured on breast tissue [Pifferi et al., 2004]. For

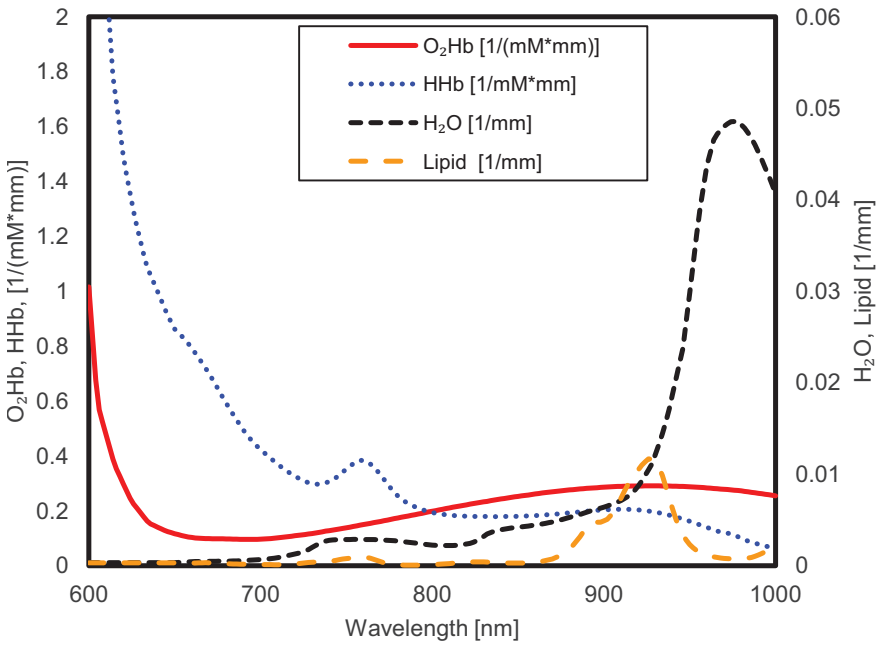

Figure 1.2 – Absorption spectra for O_2Hb, HHb, water and lipid, the main absorbers in tissue in the near-infrared region [Wray et al., 1988].

wavelengths between 650 nm and 950 nm the scattering is more than two orders of magnitude larger than the absorption. Similar results have been obtained for different types of tissue [Sandell and Zhu, 2011].

1.2 Near-infrared spectroscopy

Near-infrared spectroscopy (NIRS) is relatively novel optical method capable of monitoring different chromophore concentrations in biological tissue [Wolf et al., 2007] using near-infrared light. Although the absorbing and scattering properties of biological tissue have been studied for several years [Cutler, 1929, Sugioka, 1996, Scholkmann et al., 2014], it was not until the late 70's that this technique was applied to monitor hemodynamics in the brain [Jobsis, 1977]. Since then NIRS has been utilized to measure the concentration of chromophores in a variety of tissues such as muscle [Vardi and Nini, 2008, Wolf et al., 2003], brain [Wolf et al., 2008] or breast [Grosenick et al., 2003, Leff et al., 2008], to name a few.

The first NIRS instruments were based on measurements of continuous wave (CW) light

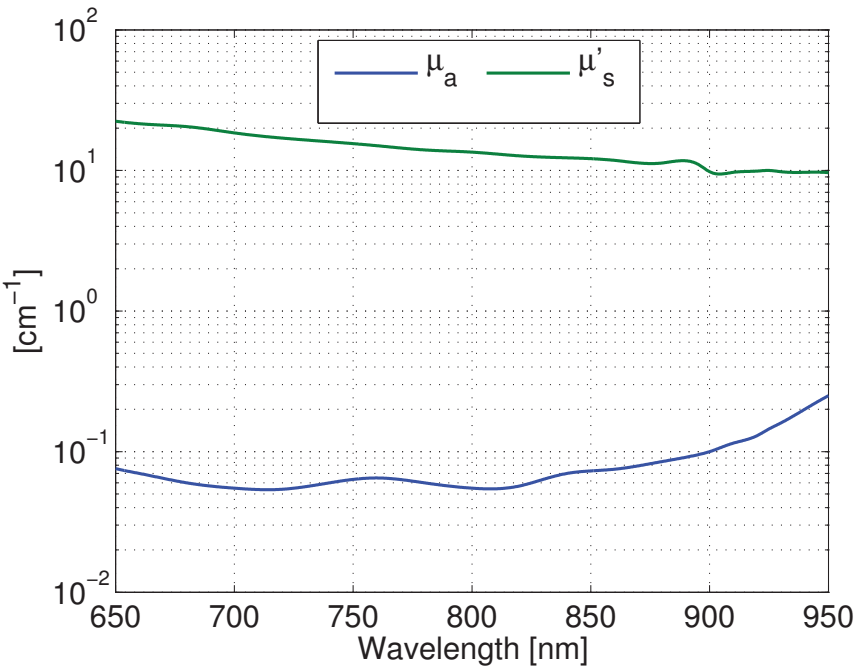

Figure 1.3 – Example of the in-vivo reduced scattering and absorption spectra of normal breast tissue versus wavelength in a single patient [Pifferi et al., 2004].

at two different wavelengths [Ferrari and Quaresima, 2012] using a single optical channel, e.g. the combination of one light source and one detector. With these systems it was only possible to obtain relative changes of chromophores as measuring the scattering coefficient was not possible. In order to determine the scattering coefficient, it is necessary to measure the propagation speed of light in the tissue under study. With the advent of new instrumentation capable of determining time-of-flight of photons [Delpy et al., 1988, Patterson et al., 1989], information about the tissue scattering coefficient became available, enabling the calculation of absolute values of chromophore concentrations. In these systems, pulses of light of a few picoseconds illuminate the tissue. The backscattered light is then detected with fast single-photon detectors which are connected to time-correlated single-photon counting (TCSPC) hardware [Becker, 2005]. These first time-domain (TD) NIRS instruments were large and required long acquisition times, i.e. they were not appropriate for in-vivo experiments. Thus they were mostly utilized to create reference tables of scattering coefficients of different tissues, whereas CW systems were employed for in-vivo experiments. Frequency domain (FD) systems in which intensity light sources are employed, were introduced as an alternative to TD measurements to measure the scattering coefficient [Gratton et al., 1993]. These devices had a

much higher operation speed compared to TD systems and were easier to scale, making them suitable for in-vivo operation. Although alternative methods for measuring the scattering coefficient based on spatially modulated light [Abookasis et al., 2009, Cuccia et al., 2009] and short source-detector distances [Bevilacqua and Depeursinge, 1999] have also been developed, TD and FD systems are still the preferred choice for the determination of absolute concentration values of chromophores in turbid media.

One of the main limitations in NIRS as compared to other spectroscopic techniques such as magnetic resonance spectroscopy (MRS), is its limited penetration depth. In reflection mode systems, e.g. systems in which the sources and detectors are positioned on the same plane of the tissue under study (see Fig. 1.4), the sensitivity depth is limited to 2 cm to 3 cm. Transmission mode NIRS setups, e.g. the sources and detectors are placed at opposite sides of the tissue (see Fig. 1.4), offer a better depth sensitivity that is approximately 4 cm to 6 cm. However, transmission mode measurements are restricted to tissues that are thinner than 6 cm, otherwise the measured signal is too weak to perform a measurement in an acceptable period of time. Reflection mode systems offer a wider range of applications, because they are less restricted by the geometry of the tissue that is being measured. The depth sensitivity in reflection mode systems is proportional to the source-detector distance, therefore by increasing it in reflection mode NIRS systems, it is possible to obtain a higher depth sensitivity. Nevertheless, with larger source-detector distances the signal weakens and in most of the cases it is not possible to increase the power of the light source to compensate for it since it may damage the tissue.

Another property of NIRS that can be seen in Fig. 1.4 is that the sensitivity is much higher in the regions next to the sources and detectors. In order to obtain some degree of depth discrimination and to be more sensitive to deep tissue, multi-channels systems with different source-detector distances were introduced [Scholkmann et al., 2014]. This reduces the influence of superficial tissues.

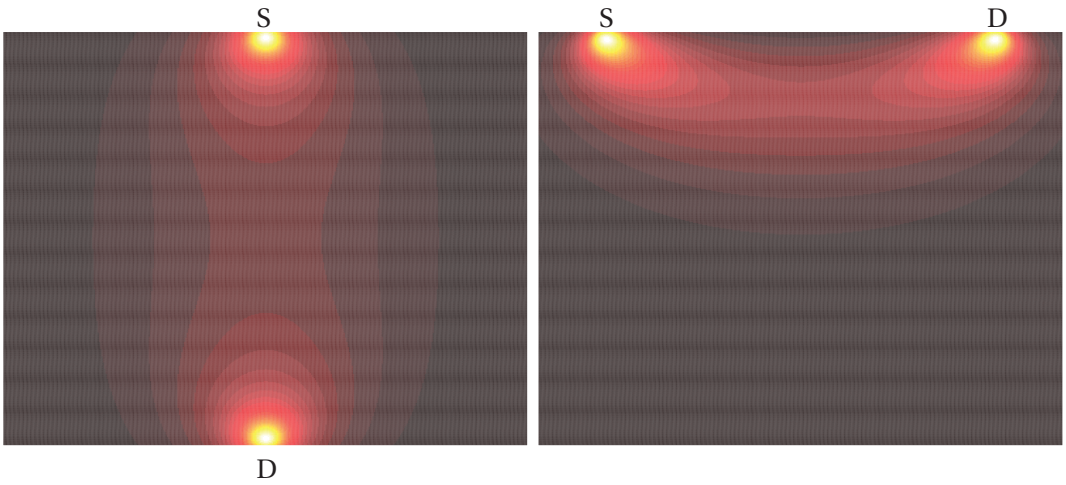

Figure 1.4 – (left) Example of a sensitivity map for a transmission mode NIRS channel, e.g. source-detector pair. A light source positioned at S illuminates a laterally infinite slab, a detector located at D measures the transmitted light. (right) Sensitivity map for a reflection mode channel. Probability density of photons passing through a specific area when traveling from S to D. Brighter areas represent a higher probability.

1.3 Near-infrared optical tomography (NIROT)

Multi-channel systems allowed not only to improve the depth sensitivity in NIRS systems, but also enabled the acquisition of tomographic images of chromophore concentrations. Methods similar to the ones already available for other tomographic techniques [Kak and Slaney, 1988] were developed. The main difference between NIROT and other established tomographic methods such as X-ray computed tomography (CT) or magnetic resonance imaging (MRI), is that for NIROT the absorbing and scattering nature of tissue makes the image reconstruction task a mathematically ill-posed problem. Moreover the technical difficulties for implementing systems with dense grids of sources and detectors made the image reconstruction problem not only ill-posed, but also underdetermined. Algebraic reconstruction techniques were consequently developed to perform image reconstructions from a limited set of measurements. These new approaches were based on a two step procedure: first a light transport model is defined for the target under study, this is the so called forward problem. This model is then inverted to obtain the optical properties of the tissue from a set of measurements. This step is known as the inverse problem [Arridge and Schotland, 2009]. Most image reconstructions algorithms can be classified in three types: perturbation models based on regular geometries [Graber et al., 1993, O'Leary, 1996], finite element methods (FEM) [Arridge et al., 1993, De-

hghani et al., 2009a], and Monte Carlo techniques [Hayakawa et al., 2001, Fang and Boas, 2009]. The first two methods are generally based on the diffusion approximation of the radiative transport equation (RTE) [Arridge, 1999], whereas the Monte Carlo methods are mostly based on the RTE.

1.3.1 Perturbation methods on regular geometries

Linearized solutions of the diffusion equation for regular geometries have been widely used to model heterogeneous highly scattering tissue due to their relatively simple implementation and computation efficiency [Graber et al., 1993, O'Leary, 1996]. Their limited applicability to targets with regular geometries such as semi-infinite mediums or spheres, made it less attractive. It has however been proven that is possible to expand the signals measured on irregular geometries to an infinite plane geometry [Ripoll and Ntziachristos, 2006]. Methods based on dense source and detector matrices have demonstrated how it is possible to implement fast reconstruction algorithms paving the path for future real time systems [Konecky et al., 2008]. The application of structured illumination as opposed to point light sources have shown how acquisition times can also be dramatically reduced [Cuccia et al., 2009, Lukic et al., 2009]. So far systems based on structured illumination working in reflection mode were only able to image objects a few millimeters deep in the tissue, limiting their range of application. Moreover, only CW systems integrating CCD or CMOS cameras have been employed in wide-field NIROT systems. Tailored fast reconstruction algorithms for time-resolved wide-field NIROT are still missing in the state-of-the-art, and will be object of study in this thesis.

1.3.2 Finite element methods

Numerical methods based on the local solution of the diffusion equation in a meshed space such as FEM, have become one of the most popular methods to define the forward model. This is due to their inherent flexibility to model irregular objects and the ability to include prior information from other tomographic sources such as MRI or CT [Arridge et al., 1993, Dehghani et al., 2009a]. Nonetheless these methods are computationally expensive requiring supercomputers for high numbers of sources and detectors such as wide-field systems based on integrated detectors. This is their main drawback since the only possibility for obtaining

tomographic images at higher resolution is with large datasets.

1.3.3 Monte Carlo methods

Monte Carlo methods have also been employed [Hayakawa et al., 2001] in NIROT, which with the advent of graphical processor unit (GPU) based computation [Fang and Boas, 2009], became very attractive as they allowed the parallel simulation of many photons using a desktop PC. A Monte Carlo based algorithm calculates the sensitivity matrix of the tissue, that is the formulation of the forward problem. The inversion step is generally performed with algebraic reconstruction techniques. These methods are particularly interesting in cases in which the light transport in tissue can no longer be modeled with the diffusion approximation [Migacheva et al., 2010] such as for very short source-detector distances. In most NIROT applications were this distance can be made long enough such that the diffusion approximation is still valid, Monte Carlo approaches are not attractive as they require very long computation times.

1.4 Photodetectors in NIRS and NIROT applications

NIROT and NIRS systems require detectors with different characteristics that are optimized for the detection of the corresponding light features. In CW systems it is important to be able to detect signals with a wide dynamic range, because in most of the cases light coming from sources placed at different source-detector distances is measured. Another property of CW devices is the integration of a high number of sources and detectors to obtain a higher spatial resolution [Habermehl et al., 2012]. As the information per channel is less than that obtained in FD or TD systems a higher number of sources and detectors are integrated in order to obtain more data for the reconstructions. Therefore photodiodes are the most widespread detectors in CW setups. They are robust, can be easily integrated in high numbers, and offer a wide dynamic range [Strangman et al., 2002]. They do not have internal signal amplification giving a one-to-one correspondence between the detected photons and the free electrons generated in the semiconductor junction [Liu, 2005].

FD systems employ modulated light sources working from a few tens to hundreds of megahertz.

They require fast, highly sensitive detectors capable of detecting the modulation frequency. Although there are fast photodiodes capable of working at the required speeds for FD devices, in order to achieve this fast operation they need to be extremely miniaturized, considerably reducing their light detection capabilities. Thus photodetectors with internal amplification such as photomultiplier tubes (PMTs), microchannel plates (MCPs), or avalanche photodiodes (APDs) are the preferred choices in FD systems, because they offer high sensitivity, and fast operation speeds. All these photodetectors are generally large, they require stabilized temperatures and a bias voltage of few hundred volts to obtain a stable constant gain. Additionally, some of them need to be constantly cooled for low noise operation, and consequently cannot be compactly integrated in high numbers.

Most TD setups are based on the TCSPC principle. They employ the same type of detectors as in FD systems operating in the single-photon detection regime. APDs working in this regime are also known as Geiger-mode APDs or single-photon avalanche diodes (SPADs) [Cova et al., 1983]. They can be integrated in arrays [Renker, 2007] enabling new types of devices such as silicon photomultipliers (SiPMs) that can be applied in NIRS and NIROT [Zimmermann et al., 2013].

Recently, researchers have demonstrated the benefits of wide-field NIROT with a variety of illumination systems [Konecky et al., 2008, Cuccia et al., 2009]. New high sensitive CCD and CMOS cameras enabled the implementation of these non-contact setups which allowed the development of fast reconstruction algorithms in NIROT.

The integration of SPAD detectors in CMOS technology [Rochas et al., 2003a] allowed on-chip integration of time discriminators, opening a new range of possibilities in the field of wide-field TD measurements. Since then, SPAD arrays have been utilized in a variety of applications where TD measurements were required [Charbon and Donati, 2010]: fluorescence lifetime imaging microscopy (FLIM) [Li et al., 2010], positron emission tomography (PET) [Meijlink et al., 2011, Mandai and Charbon, 2013b], or time-resolved Raman spectroscopy [Maruyama et al., 2013].

In particular, the integration of SPADs with time-to-digital converters (TDCs) in the same chip made it possible to perform simultaneous TCSPC measurements with a very large number of

detectors [Niclass et al., 2008]. As previously mentioned, TD measurements enable simultaneous resolution of scattering and absorbing objects, as well as higher depth sensitivity. Thus, SPAD image sensors enable wide-field TD measurements with a monolithic image sensor, bringing new possibilities to NIROT. In this thesis it will be demonstrated how SPAD image sensors change the paradigm of NIROT systems bringing together wide-field and time-resolved measurements.

1.5 Thesis motivation and aims

Oxygen is one of the most important substances in the human body. An average 80 kg human male consumes approximately 18 l of oxygen in an hour [Kappagoda et al., 1979]. Oxygen consumption is related to almost every metabolic process, however, there is no clinical method to image it in a fast and reliable way. Hb being the main oxygen transporter in blood, measuring its concentration in its oxygenated and deoxygenated forms determines the amount of oxygen that is locally consumed. So far it has only been possible to simultaneously measure HHb and O_2Hb with optical methods, however, the resolution and accuracy obtained with NIROT were not enough to apply it in clinical practice.

There are many clinical fields that would benefit from in-vivo oxygenation measurements in tissue. Several studies have shown that tumor oxygenation is directly related to the tumor's resistance to therapy (radiotherapy and chemotherapy) and to the prognosis of the patient's survival rate [Vaupel et al., 2006, Vaupel and Mayer, 2014, Mayer and Vaupel, 2013, Movsas et al., 2002, Höckel et al., 1996]. The survival rate of patients with hypoxic tumors is approximately half that of the other patients. Oxygenation is probably the most important but not yet routinely measured parameter in cancer treatment.

Brain injury in premature infants is a very common medical complication with tremendous consequences as these children will have motor and cognitive disabilities in their adult life [Volpe, 2009]. The development of NIRS and NIROT systems capable of detecting the oxygenation to prevent brain lesions has made significant progress in the past [Wolf and Greisen, 2009, Wolf et al., 2012], however the low precision and spatial resolution of these devices (maloxygenation can be a localized phenomenon) has made them so far impractical in clinical

practices.

The main aim of the present Ph.D. dissertation was to implement a new NIROT system with millimeter resolution capable of imaging absorbing objects a few centimeters deep in the tissue. High resolution image reconstructions have so far only been possible to be generated with a very high number of sources and detectors, either with CCD or CMOS integrated sensors or by scanning the tissue with a single detector. Although the later approach allows using any type of detector, the required measurement time makes it unfeasible for in-vivo applications. In the other hand time-resolved measurements provide better depth accuracy and the possibility of simultaneously imaging absorption and scattering coefficients. The combination of wide-field NIROT with integrated sensors and time-resolved measurements will bring the best of both worlds together.

Wide-field time-resolved NIROT with a monolithic SPAD image sensor offers the possibility of performing TD measurements over a large area without the necessity of raster scanning the detection points. The combination of non-contact wide-field images with time-resolved measurements is still an unexplored paradigm that can solve many of the current problems of NIROT. The development of a new optical system for NIROT based on a SPAD image sensor and new reconstruction algorithms that would take advantage of the large datasets obtained from the future system were the main objective of the first part of the thesis.

In order to exploit the benefits that SPAD image sensors can bring to NIROT, the aim of the second part of this thesis was to investigate the design of a new application specific integrated circuit (ASIC) for NIROT. New low power building blocks that allowed the massive integration of TDCs were studied as well as architectures and methods to increase the detector's light sensitivity.

1.6 Contributions and organization of the thesis

In chapter 2 the mathematical background that is the foundation for wide-field fast image reconstruction algorithms is introduced. The difficulties for obtaining high resolution images with NIROT and possible solutions are presented together.

Chapter 3 introduces a new fast reconstruction algorithm based on wide field TD NIROT. This algorithm takes advantage of line-shaped illumination to reduce the ill-posedness of the image reconstruction problem. The advantages of TD over CW measurements regarding resolution and depth accuracy are also studied.

The operation principles of SPADs and its current state of the art are reported in chapter 4. Here the SPAD image sensor that is integrated in the experimental setup presented in the next sections is also introduced. The limitations of the image sensor and a new sliding scale method to improve the TDCs' linearity are also described. In chapter 5 a microlens array fabricated on top of the SPAD image sensor to improve the sensor's light sensitivity is presented.

Chapter 6 presents a novel optical setup for NIROT based on a SPAD image sensor. Experimental results obtained from intralipid phantoms are reported showing the benefits of using these new types of sensors in NIROT. To our knowledge this is the first NIROT system based on a monolithic SPAD imager capable of performing measurements in the TD.

A new SPAD image sensor is presented in chapter 7. A new resource sharing scheme was implemented to make a more efficient use of the resources combined with a new TDC architecture that reduced the power consumption during a conversion. The sensor was built in a new 3D wafer stacking technology that allowed the implementation of back side illuminated SPADs optimized for NIR wavelengths.

The thesis ends with a conclusion and an outlook in the future of wide-field time-resolved NIROT and SPAD image sensors for TCSPC applications.

2 Near-infrared optical tomography: theoretical background

The aim of this chapter is to describe the theory of light propagation in tissue and how this can be applied to obtain tomographic images. The concepts introduced in it will be the basis of the algorithms presented in the following chapters.

2.1 The diffusion equation

Light propagating in a highly scattering and low absorbing medium, such as near-infrared light traveling through tissue, can be modeled with the RTE [Arridge, 1999]. This is a rather complex model which can be simplified with the P_N approximation in which the angular dependency of the light intensity is expanded in spherical harmonics. If the expansion is truncated at N=1, that is the light intensity has no angular dependency, and the source term is purely isotropic, then the photon energy density as a function of time and position in tissue $u(\mathbf{r}, t)$ is given by the following equations:

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} + c\mu_a(\mathbf{r})u(\mathbf{r}, t) + \nabla \cdot \mathbf{J}(\mathbf{r}, t) = S_0(\mathbf{r}, t), \tag{2.1}$$

$$\nabla u(\mathbf{r}, t) + \frac{3\partial \mathbf{J}(\mathbf{r}, t)}{c^2 \partial t} + \frac{\mathbf{J}(\mathbf{r}, t)}{D(\mathbf{r})} = 0. \tag{2.2}$$

Here c is the speed of light in the medium, $\mathbf{J}(\mathbf{r}, t)$ is the photon current, $S_0(\mathbf{r}, t)$ is the source term expressing the number of photons per unit volume per unit time positioned at $\mathbf{r}$, $\mu_a(\mathbf{r}$ is

the medium's absorption coefficient, and D is the diffusion coefficient defined as

$$D(\mathbf{r}) = \frac{c}{3(\mu_s'(\mathbf{r}) + \mu_a(\mathbf{r}))}. \tag{2.3}$$

The reduced scattering coefficient $\mu_s'(\mathbf{r})$ is given by

$$\mu_s'(\mathbf{r}) = (1-g)\mu_s(\mathbf{r}) \tag{2.4}$$

where g is the anisotropy factor, which indicates the amount of light that is scattered in the forward direction. It is defined as the average cosine of the single event scattering angle. In most biological tissue light is scattered in the forward direction, resulting in $g \approx 0.9$. Eq. 2.1 and Eq. 2.2 are only valid if the distances between sources and detectors are larger than the average distance the photons travel between isotropic scattering events $1/\mu_s'$.

Given a harmonic source term with time dependency $e^{i\omega t}$ where ω is the modulation frequency, Eq. 2.1 and Eq. 2.2 can be rewritten as

$$(i\omega + c\mu_a(\mathbf{r}))\,u(\mathbf{r},\omega) + \nabla\cdot\mathbf{J}(\mathbf{r},\omega) = S_0(\mathbf{r},\omega), \tag{2.5}$$

$$\nabla u(\mathbf{r},\omega) + \Big(\frac{3i\omega}{c^2} + \frac{1}{D(\mathbf{r})}\Big)\mathbf{J}(\mathbf{r},t) = 0 \tag{2.6}$$

respectively. Assuming $(3i\omega/c^2 << 1/D)$, which is true for most human tissue for modulation frequencies up to 3 GHz to 5 GHz, 2.6 becomes

$$\mathbf{J}(\mathbf{r},t) = -D(\mathbf{r})\nabla u(\mathbf{r},\omega), \tag{2.7}$$

this equation is known as the Fick's law. Substituting it in Eq. 2.5 we obtain the diffusion equation

$$(i\omega + c\mu_a(\mathbf{r}))\,u(\mathbf{r},\omega) - \nabla\cdot(D(\mathbf{r})\nabla u(\mathbf{r},\omega)) = S_0(\mathbf{r},\omega). \tag{2.8}$$

2.2 Homogeneous solutions in media with regular geometries

In a medium with $\mu_a(\mathbf{r}) = \overline{\mu_a}$ and $D(\mathbf{r}) = \overline{D}$, Eq. 2.8 can be written as a Helmholtz equation

$$(\nabla^2 + k_0(\omega)^2)\, u(\mathbf{r},\omega) = -\frac{S_0(\mathbf{r},\omega)}{\overline{D}}. \tag{2.9}$$

where

$$k_0(\omega)^2 = \frac{-c\overline{\mu_a} - i\omega}{\overline{D}}. \tag{2.10}$$

Eq. 2.9 can be analytically solved for regular geometries given a boundary condition. An infinite space would represent the simplest case, in it the only boundary condition would be that the field must be zero at infinity as we move away from the sources. For a point source centered at $\mathbf{r_0}$, $S_0(\mathbf{r},\omega) = \delta(\mathbf{r}-\mathbf{r_0})A(\omega)$, the solution to 2.9 is [Mandelis, 2001]:

$$u(\mathbf{r},\omega) = \frac{e^{ik_0(\omega)|\mathbf{r}-\mathbf{r_0}|}}{4\pi D|\mathbf{r}-\mathbf{r_0}|}, \tag{2.11}$$

which corresponds to the Green's function of Eq. 2.9. For any other space in which a transition to a non-scattering medium occurs, a boundary condition needs to be specified. There are several approaches to describe how the diffuse wave behaves at the boundary with a medium in which the diffusion equation is not valid anymore. The two most common boundary conditions are the partial-current and the extrapolated boundary conditions [Haskell et al., 1994]. The extrapolated boundary condition is a simple widespread model primarily used to describe the boundary condition of planar geometries such as semi-infinite planes or infinite slabs. According to this, the energy density at an extrapolated boundary located at $z = l$ outside the turbid media will be zero, $u(\boldsymbol{\rho}, z = l, \omega) = 0$. The light source illuminating the surface of the medium needs to be positioned at $z = -1/\mu_s'$, so that it can be considered anisotropic. It is possible then to apply the method of images to define a set of sources that satisfy the boundary condition. In the semi-infinite case this would require the positioning of a negative source at $z = 2l + 1/\mu_s'$ with the same magnitude as the real illumination term outside the medium. The value of the extrapolated distance l depends on the refractive index mismatch between the turbid medium and the other medium [Haskell et al., 1994]. This results in the following

expression of $u(\mathbf{r},\omega)$ for a semi-infinite medium:

$$u_0^{semi}(\mathbf{r},\omega) = u_0^{inf}(\boldsymbol{\rho}, z+1/\mu_s',\omega) - u_0^{inf}(\boldsymbol{\rho}, z-2l-1/\mu_s',\omega). \tag{2.12}$$

For a laterally infinite slab a similar solution can be derived using an infinite number of sources [Patterson et al., 1989]. Using cylindrical and spherical coordinate systems closed solutions for other regular geometries such as infinite cylinders or spheres have been calculated [Arridge et al., 1992].

2.3 Heterogeneous solution in media with regular geometries

One of the most common methods to calculate $u(\mathbf{r},\omega)$ in a heterogeneous medium is the perturbation theory. Given that the bulk optical properties of the medium can be measured, it is possible to calculate their spatial changes by measuring the changes in $u(\mathbf{r},\omega)$ produced by these perturbations. The two most frequent methods are the Born and Rytov approximations. The derivation of the formulas has been already reported in several publications [Kak and Slaney, 1988, O'Leary, 1996, Haensse, 2005, Konecky, 2008], therefore we briefly summarize them.

We consider a heterogeneous media with the following optical properties

$$\mu_a(\mathbf{r}) = \overline{\mu_a} + \delta\mu_a(\mathbf{r}), \tag{2.13}$$

$$D(\mathbf{r}) = \overline{D} + \delta D(\mathbf{r}). \tag{2.14}$$

In the first Born approximation we express the generated field in this medium as

$$u(\mathbf{r},\omega) = u_0(\mathbf{r},\omega) + u_s(\mathbf{r},\omega), \tag{2.15}$$

where $u_0(\mathbf{r},\omega)$ is the field in the homogeneous medium and $u_s(\mathbf{r},\omega)$ is the perturbed field due to the heterogeneities. In the first Born approximation the assumption that $u_0(\mathbf{r},\omega) >> u_s(\mathbf{r},\omega)$ is considered. Applying this assumption together with Eq. 2.13,2.14,2.15, the solution to Eq.

2.8 can be expressed in terms of Eq. 2.9 Green's function

$$u_s(\mathbf{r},\mathbf{r_s},\omega) = -\int d^3\mathbf{r}' \left[c\delta\mu_a(\mathbf{r}')u_0(\mathbf{r}',\mathbf{r_s},\omega)g(\mathbf{r},\mathbf{r}',\omega) - \delta D(\mathbf{r}')\nabla u_0(\mathbf{r}',\mathbf{r_s},\omega)\cdot\nabla g(\mathbf{r},\mathbf{r}',\omega)\right]. \quad (2.16)$$

The Green's function $g(\mathbf{r},\mathbf{r_s},\omega)$ will depend on the geometry of the medium, and can be calculated as the solution of Eq. 2.9 to a $\delta(\mathbf{r})$ stimuli. If point like illumination sources are employed then $g(\mathbf{r},\mathbf{r}',\omega)$ and $u_0(\mathbf{r}',\mathbf{r_s},\omega)$ yield the same mathematical expression.

In the Rytov approximation $u(\mathbf{r},\omega)$ is expressed as

$$u(\mathbf{r},\omega) = u_0(\mathbf{r},\omega)e^{\psi(\mathbf{r},\omega)} \quad (2.17)$$

Assuming that the spatial changes in $\psi(\mathbf{r},\omega)$ are small compared to the changes in the optical properties, the perturbed field can be expressed as

$$\psi(\mathbf{r},\mathbf{r_s},\omega) = \frac{-1}{u_0(\mathbf{r},\mathbf{r_s},\omega)} \int d^3\mathbf{r}' \left[c\delta\mu_a(\mathbf{r}')u_0(\mathbf{r}',\mathbf{r_s},\omega)g(\mathbf{r},\mathbf{r}',\omega) - \delta D(\mathbf{r}')\nabla u_0(\mathbf{r}',\mathbf{r_s},\omega)\cdot\nabla g(\mathbf{r},\mathbf{r}',\omega)\right]. \quad (2.18)$$

Defining $\phi(\mathbf{r},\mathbf{r_s},\omega) = \psi(\mathbf{r},\mathbf{r_s},\omega)u_0(\mathbf{r},\mathbf{r_s},\omega)$, Eq. 2.18 becomes

$$\phi(\mathbf{r},\mathbf{r_s},\omega) = -\int d^3\mathbf{r}' \left[c\delta\mu_a(\mathbf{r}')u_0(\mathbf{r}',\mathbf{r_s},\omega)g(\mathbf{r},\mathbf{r}',\omega) - \delta D(\mathbf{r}')\nabla u_0(\mathbf{r}',\mathbf{r_s},\omega)\cdot\nabla g(\mathbf{r},\mathbf{r}',\omega)\right], \quad (2.19)$$

which present the same kernel as the Born solution in Eq. 2.16.

In Eq. 2.16 and Eq. 2.18, the left-hand side term is measured at the boundary of the medium whereas the right-hand side term is calculated analytically. In order to calculate the measured term u_s and ψ in Eq. 2.16 and Eq. 2.18 respectively, the term u_0 is necessary. Although in both cases the homogeneous term u_0 could in principle be analytically calculated, better results are obtained when this is measured on a reference homogeneous medium. This is especially true for the Rytov case since the different experimental factors that are normally characterized by constant coupling coefficients, such as laser intensity or detector sensitivity, will be automatically canceled by dividing both terms.

2.4 The inverse problem: calculating the optical properties

Once the so-called forward problem is defined with Eq. 2.16 or Eq. 2.19, it is possible to discretize their kernel to generate an equation system that can be expressed in matrix form

$$\Phi_{M\times 1} = \left(A^{\delta\mu_a}_{M\times N}\, A^{\delta D}_{M\times N}\right)\begin{pmatrix}\delta\mu_{a\,N\times 1}\\ \\ \delta D_{N\times 1}\end{pmatrix}, \tag{2.20}$$

where N is the number of source-detector pairs multiplied by the number of modulation frequencies, M the number of voxels in which the integrated volume is discretized, and Φ is the right-hand side term in Eq. 2.16 and Eq. 2.19. The matrix elements in 2.20 are

$$\delta\mu_{a\,j} = \delta\mu_a(\mathbf{r}_j) \tag{2.21}$$

$$\delta D_j = \delta D(\mathbf{r}_j) \tag{2.22}$$

$$A^{\delta\mu_a}_{i,j} = c\,u_0(\mathbf{r}_j, \mathbf{r}_{\mathbf{s}i}, \omega_i)\, g(\mathbf{r}_i, \mathbf{r}_j, \omega_i)\, dV \tag{2.23}$$

$$A^{\delta D}_{i,j} = \nabla u_0(\mathbf{r}_j, \mathbf{r}_{\mathbf{s}i}, \omega_i) \cdot \nabla g(\mathbf{r}_i, \mathbf{r}_j, \omega_i)\, dV \tag{2.24}$$

$$\Phi_i = u(\mathbf{r}_i, \mathbf{r}_{\mathbf{s}i}, \omega_i) - u_0(\mathbf{r}_i, \mathbf{r}_{\mathbf{s}i}, \omega_i) \qquad \text{(Born approximation)} \tag{2.25}$$

$$\Phi_i = \ln\left(\frac{u(\mathbf{r}_i, \mathbf{r}_{\mathbf{s}i}, \omega_i)}{u_0(\mathbf{r}_i, \mathbf{r}_{\mathbf{s}i}, \omega_i)}\right) u_0(\mathbf{r}_i, \mathbf{r}_{\mathbf{s}i}, \omega_i) \qquad \text{(Rytov approximation)} \tag{2.26}$$

where dV is the volume of the voxel. Inverting the matrix in Eq. 2.20, it is possible to calculate the optical properties of the tissue under study from a set of measurements. Due to the absorbing and scattering nature of the tissue, the problem is ill-posed, i.e. there is not a unique solution. Due to the exponential nature of the Green's functions obtained in NIROT, the inversion problem resembles a Laplace inversion in the real line, which is the paradigm of an ill-posed problem. A matrix problem is considered ill-posed when its condition number ,i.e. the ratio between its highest and lowest singular values, is infinite [Demmel, 1987]. Moreover the number of voxels is usually in the order of tens of thousands to a few million, making the matrix inversion a critical step that needs regularization in order to converge to a meaningful solution. In many NIROT systems implemented with discrete components, the number of measurements is substantially lower than the number of voxels, making the equation's system

more ill-posed and also underdetermined.

A variety of inversion methods have been applied to the matrix inversion problem in NIROT such as algebraic reconstruction technique (ART), conjugate gradient (CG), truncated singular value decomposition (TSVD), or least square QR (LSQR) to name a few [Arridge, 1999, Arridge and Schotland, 2009, Konecky, 2008]. Most of these methods are iterative processes that seek to minimize a function that evaluates the solution's residual error. In order to improve the convergence speed and to reduce the solution space, different regularization schemes are applied in most of these techniques. Among them the Tikhonov regularization is one of the most widely used. In a system $Ax = b$, where A is the matrix that has to be inverted in order to calculate the unknown x and b is the measured term, the Tikhonov regularization consist of minimizing the error defined by

$$\epsilon = \|A\mathrm{x} - \mathrm{b}\|_2^2 - \lambda \|L\mathrm{x}\|_2^2, \tag{2.27}$$

where ϵ is the residual error, L is the regularization term, and λ is the regularization parameter that controls the weight of the regularized norm on the residual error. L is defined to give preference to a set of solutions with determined characteristics, such as solutions with smaller norms or with a certain degree of smoothness. In most inversion algorithms when applied to NIROT image reconstructions, regularization operators that enforce smooth and small solutions are chosen. However, one has to be careful when choosing λ and L as they can considerably distort the reconstruction if their contribution to ϵ is too high, consequently reducing the resolution of the images and even resulting in a non-valid solution. A high λ value would lead to an error minimization problem in which the contribution of the $\|A\mathrm{x} - \mathrm{b}\|_2^2$ term would be minimal, consequently the inversion algorithm would look for a solution that mostly tries to minimize $\|L\mathrm{x}\|_2^2$. Although different methods for calculating the optimal λ parameter have been described, such as the L-curve method [Calvetti et al., 2004], their computation complexity is higher than solving the inverse problem itself, and they do not guarantee that the calculated solution is the closest to the real solution. In problems where the structure of the matrix A, which depends on the source-detector configuration and the optical properties of the medium, and b, the measurement term, do not strongly vary, it is reasonable to employ the same λ parameter. Therefore in NIROT problems, where the source-detector positions

are usually fixed and the optical properties of tissue vary within a known range it is possible to always use a fixed λ term. In order to find the optimum regularization parameter, it is possible to define a problem in which an object with certain optical properties is positioned in a medium with different optical properties. The forward problem and the inverse solution can then be calculated for a set of λ values. By evaluating the obtained solutions in terms of object position, form and reconstructed values, it is possible to find the optimum λ.

The subspace preconditioned LSQR (SP-LSQR) algorithm [Jacobsen et al., 2003] is a multilevel method in which the solution space is split into two subspaces. In a first step the solution is searched in a predefined subspace of solutions with a small dimension. Once the optimal solution in this subspace has been found then the algorithm expands the search to the remaining subspace. We conducted a comparison of different numerical inversion methods applied to NIROT. A testbench was defined in which a 1 cm diameter sphere with a 4:1 absorption contrast with the background, was placed in a semi-infinite medium with several sources and detectors placed equidistant on the boundary surface. The field at the boundary of the medium was calculated with 2.16, and the inverse problem was defined as in 2.20. The inversion step was solved with five different methods: CG, ART, LSQR, SP-LSQR with a regularization operator that minimizes the norm of the solution with an identity matrix, and SP-LSQR with a regularization parameter that gives preference to smoother solutions. Both SP-LSQR methods were implemented with a subspace defined with a basis of cosines. The normalized error obtained with each method after every iteration is depicted in Fig 2.1. CG and LSQR yield very similar results as they are algebraically equivalent. It is remarkable also how the SP-LSQR algorithm is capable of finding solutions with a smaller error from the first iteration, this is due to the reduced subspace of solutions that is employed during the first part of the algorithm. The SP-LSQR with a smoothness operator (SP-LSQR-L) performs better than the one with the identity matrix (SP-LSQR-N). The reason for this is that in NIROT it is very difficult to reconstruct objects with high frequency components since the medium itself acts as a low pass filter. Therefore even in cases where the optimal solution would not be smooth, it is more beneficial for the inversion step to define some smoothness restrictions to make the algorithm avoid solutions with high frequency features.

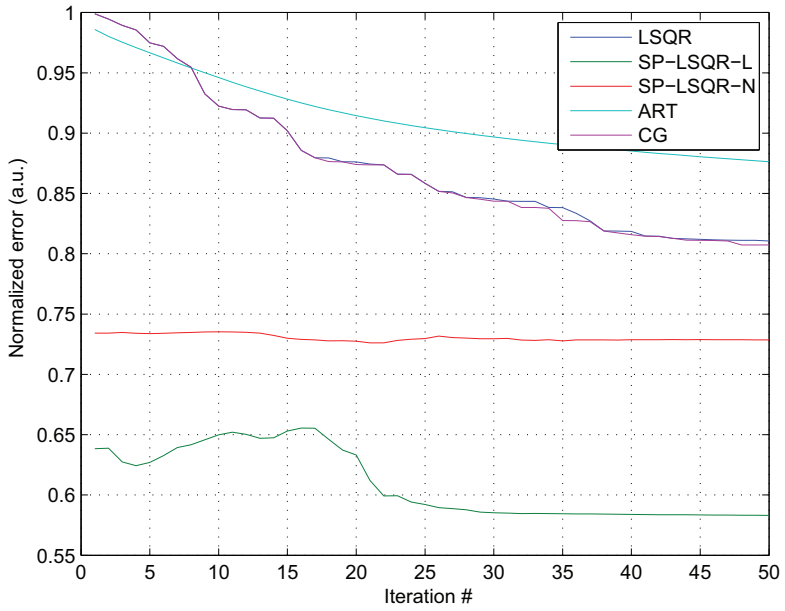

Figure 2.1 – Normalized error evolution of a NIROT image reconstruction problem for different matrix inversion procedures.

2.5 Conclusion

In this chapter it has been described the theory and the mechanism necessary to reconstruct an image in NIROT. It has been shown that this is an ill-posed problem that in many cases is also underdetermined. In addition the numerical calculations necessary to solve the inverse problem are computationally intensive, requiring large amounts of RAM memory in systems with many sources and detectors. The development of optimized algorithms that reduce the ill-posed nature of the inverse problem and that are less computationally demanding is therefore a critical aspect that needs to be addressed in NIROT.

3 Fast reconstruction method for time-resolved wide-field NIROT using line-shaped illumination

In this chapter a novel approach is presented to reduce the ill-posedness of the image reconstruction problem. This method takes advantage of line-shaped light sources to reconstruct images with less calculations compared to traditional NIROT image reconstruction algorithms.

Advances in CMOS and CCD imaging processes in recent years resulted in sensors with high photon detection efficiency suitable for low light applications such as optical tomography or fluorescence molecular tomography [Konecky et al., 2008][Ripoll, 2010], enabling for the first time wide-field tomographic systems based on integrated sensors. Furthermore, recent advances in single-photon avalanche diodes (SPADs) allowed the design of very compact sensors with the capability to determine the time-of-flight of photons [Niclass et al., 2008], enabling wide-field time-resolved NIROT as we have demonstrated [Mata Pavia et al., 2011, Mata Pavia et al., 2014b].

The large amount of information delivered by these new systems enabled the development of new fast reconstructions algorithms that drastically reduced the ill-posed nature of the inverse problem [Lukic et al., 2009][Markel and Schotland, 2004]. Image reconstructions in transmission mode setups making use of these new types of algorithms yielded high resolution image reconstructions at a few centimeters depth [Konecky et al., 2008]. However, image reconstructions from reflection mode setups so far yielded high resolutions only up to a few millimeters in depth [Konecky et al., 2009]. The main reason is that in most of the

fast reconstruction algorithms in reflection mode the illumination sources are positioned inside the field of view of the camera. This poses two problems: i) the high intensity of the illumination creates internal reflections in the camera lens, which are orders of magnitude higher than the signals of the objects of interest located a few centimeters deep in the medium and ii) since the distance between sources and detectors is short, i.e. most detected photons only traveled through the most superficial layers of the medium, little information is provided on its structure at a few centimeters depth [Dehghani et al., 2009b]. For SPAD sensors this poses an additional problem because the signals obtained with short source detector separations are much faster than the impulse response function of most SPAD detectors.

We present a new fast reconstruction algorithm for reflection and transmission setups based on illumination sources outside the field of view of the camera. This increases the contrast for objects located a few centimeter deep in the medium and avoids the problem of internal reflections in the camera lens.

3.1 Theory

We consider the propagation of diffuse light in a homogeneous absorbing and highly scattering medium with absorption coefficient $\overline{\mu}_a$ and reduced scattering coefficient $\overline{\mu}_s'$, where $\overline{\mu}_s' >> \overline{\mu}_a$. The photon density in the medium produced by a source $S_0(\mathbf{r}, \mathbf{r_s}, \omega)$, where ω is the intensity modulation frequency and $\mathbf{r_s}$ the position of the source, is described by the diffusion equation defined in Eq. (2.8).

We then consider an absorbing heterogeneity $\delta\mu_a(\mathbf{r})$ in the medium. The change in the diffuse wave provoked by this is calculated by means of the first Born approximation (Eq. (2.16)) for any laterally infinite medium, $-\infty < x < \infty, -\infty < y < \infty, z_1 < z < z_2$, as

$$u_s(\mathbf{r}, \mathbf{r_s}, \omega) = -\int d^3\mathbf{r}' c\delta\mu_a(\mathbf{r}') u_0(\boldsymbol{\rho}' - \boldsymbol{\rho_s}, z', z_s, \omega) g(\boldsymbol{\rho} - \boldsymbol{\rho}', z, z', \omega) \tag{3.1}$$

where $\mathbf{r} = (x, y, z)$, and $\boldsymbol{\rho} = (x, y)$. We define the position of the source term relative to the

measuring position such that $\boldsymbol{\rho_s} = \boldsymbol{\rho} + \boldsymbol{\Delta\rho_s}$. If we apply the Fourier transform defined by

$$\widetilde{u}(\mathbf{q}, z, \omega) = \int d^2\boldsymbol{\rho}\, e^{-i\mathbf{q}\cdot\boldsymbol{\rho}}\, u(\boldsymbol{\rho}, z, \omega), \tag{3.2}$$

where $\mathbf{q} = (q_x, q_y)$, Eq. (3.1) becomes

$$\widetilde{u_s}(\mathbf{q}, z, z_s, \boldsymbol{\Delta\rho_s}, \omega) = \\ -\int d^3\mathbf{r}' d^2\boldsymbol{\rho} e^{-i\mathbf{q}\cdot\boldsymbol{\rho}}\, c\delta\mu_a(\mathbf{r}')\, u_0(\boldsymbol{\rho}' - \boldsymbol{\rho} - \boldsymbol{\Delta\rho_s}, z, z_s, \omega)\, g(\boldsymbol{\rho} - \boldsymbol{\rho}', z, z_s, \omega). \tag{3.3}$$

Applying the change of variables $\boldsymbol{\rho}'' = \boldsymbol{\rho} - \boldsymbol{\rho}'$, Eq. (3.3) is re-written as

$$\widetilde{u_s}(\mathbf{q}, z, z_s, \boldsymbol{\Delta\rho_s}, \omega) = -\int d^3\mathbf{r}' e^{-i\mathbf{q}\cdot\boldsymbol{\rho}'}\, c\delta\mu_a(\mathbf{r}')\, K(\mathbf{q}, \boldsymbol{\Delta\rho_s}, z, z', \omega), \tag{3.4}$$

where

$$K(\mathbf{q}, \boldsymbol{\Delta\rho_s}, z, z', \omega) = \int d^2\boldsymbol{\rho}'' e^{-i\mathbf{q}\cdot\boldsymbol{\rho}''}\, u_0(-\boldsymbol{\rho}'' - \boldsymbol{\Delta\rho_s}, z, z', \omega)\, g(\boldsymbol{\rho}'', z, z', \omega). \tag{3.5}$$

Eq. (3.4) may be expressed in terms of the two dimensional Fourier transform of $\delta\mu_a(\mathbf{r}')$

$$\widetilde{u_s}(\mathbf{q}, z, z_s, \boldsymbol{\Delta\rho_s}, \omega) = -\int dz\, c\widetilde{\delta\mu_a}(\mathbf{q})\, K(\mathbf{q}, \boldsymbol{\Delta\rho_s}, z, z', \omega). \tag{3.6}$$

Based on Eq. (3.4), it is possible to formulate a one-dimensional equation system for each $\mathbf{q}$ by measuring at different modulation frequencies ω or source-detector relative positions $\boldsymbol{\Delta\rho_s}$. If point-like light sources are applied, for each detection point the source needs to be repositioned so that $\boldsymbol{\rho_s} = \boldsymbol{\rho} + \boldsymbol{\Delta\rho_s}$. This means that the number of measurements will at least equal to the number of detectors. The measurement time will therefore be similar to previous scan-based algorithms [Konecky et al., 2008]. However by replacing point-like light sources with sources generating a homogeneous field that is constant in one of the transversal directions, x or y, it is only necessary to scan in the other direction, thereby considerably reducing the measurement time. If we illuminate our medium with a line-shaped light source parallel to the x-axis, the source term is $S_0(\mathbf{r}, \mathbf{r_s}, \omega) = A(\omega)\delta(y - y_s, z)$ where y_s is the position of the line on the y dimension and $A(\omega)$ is the amplitude at the specified modulation frequency

ω, the homogeneous field in an infinite medium will become

$$u_0(\mathbf{r}-\mathbf{r_s},\omega) = \frac{i}{4D} H_0^{(1)}(k_0(\omega)|\mathbf{R}-\mathbf{R_s}|), \tag{3.7}$$

where $\mathbf{R_s} = (y_s, z_s)$ is the position of the line-shaped illumination source, and $H_0^{(1)}$ is the Hankel function of the first order. Incorporating Eq. (3.7) into Eq. (3.5) yields

$$K(\mathbf{q}, \Delta y_s, z, z') = \int dy'' e^{-iq_y y''} u_0(-y'' - \Delta y_s, z, z', \omega) \int dx'' e^{-iq_x x''} g(\boldsymbol{\rho}'', z, z', \omega). \tag{3.8}$$

The one-dimensional equation system defined by Eq. (3.6) and (3.8) may be independently solved for each measurement with a variety of inverse methods [Arridge and Schotland, 2009]. It is also possible to calculate an inverse matrix M^{-1} that will directly map $\widetilde{u_s}$ to $\delta\widetilde{\mu_a}$ such that $\delta\widetilde{\mu_a} = M^{-1}\widetilde{u_s}$ for a medium with specific $\overline{\mu}_s'$ and $\overline{\mu}_a$ by applying analytical methods described in [Markel and Schotland, 2004]. In conclusion, we have presented a fast method based on line-shaped illumination, that reduces the ill-posedness of the inverse problems and is computationally efficient.

3.2 Simulations

Numerical simulations were performed to test the performance of the reconstruction algorithm based on Eq. (3.8). Expressions for u_0 and g were calculated for a semi-infinite medium using extrapolated boundary conditions [Haskell et al., 1994]. A semi-infinite medium with optical properties $\overline{\mu}_s' = 7\ \mathrm{cm}^{-1}$ and $\overline{\mu}_a = 0.05\ \mathrm{cm}^{-1}$ served as background. The detection points were located on the surface of the medium over an area of 6.1 cm×6.1 cm forming a grid of 32×32 points with a separation of 0.2 cm. Pulsed light sources with 50 ps full width at half maximum (FWHM) and 80 MHz repetition rate generated lines of light at four different relative positions Δy_s=[-2 cm, -1 cm, 1 cm, 2 cm] for each detector row. The detectors acquired the diffuse time-resolved signal with 97 ps resolution. In order to apply the algorithm presented here, the simulated time-resolved source signal was decomposed in harmonic frequencies with a fast Fourier transformation. Eq. (3.6) and (3.8) were used to calculate u_s at the detector plane. This resulted in measurements at 65 different positive discrete frequencies that ranged

from 0 GHz to 5.12 GHz with 80 MHz increment. Four absorbing spheres with 3 mm radius and $\delta\mu_a = 0.5\ \mathrm{cm}^{-1}$ were positioned at two different depths of the phantom. At z=-2 cm two spheres were positioned at $\boldsymbol{\rho}$=(0.7 cm, 0.7 cm) and at $\boldsymbol{\rho}$=(-0.7 cm, -0.7 cm), and the other two were placed at a depth z=-3 cm at $\boldsymbol{\rho}$=(0.7 cm, -0.7 cm) and at $\boldsymbol{\rho}$=(-0.7 cm, 0.7 cm). The inverse problem was solved with the subspace preconditioned LSQR [Jacobsen et al., 2003] due to its regularization capabilities and its faster performance compared to the traditional LSQR algorithm. A linear regularization operator that penalized results with higher norm and gave preference to smooth solutions was employed.

Fig. 3.1 shows the results of the reconstruction. All four objects are clearly detected at their correct positions. As expected, the reconstructed shape and values of $\delta\mu_a$ strongly depend on the depth of the inclusion. For z=-3 cm the reconstructed objects are much bigger, present a smoother shape, and have a lower absorption coefficient. This is due to the ill-posed nature of the problem, i.e. it is the more difficult to reconstruct objects the further they are from the surface. In the slices presented in Fig. 3.1 it is apparent that the reconstructed objects do not exhibit the same resolution in the x and y dimensions. The reconstructed shapes in the slices are wider in the x than in the y dimension, this effect is more pronounced in the slice at z=-3 cm. This is caused by the line-shaped illumination, which creates a photon detection probability distribution that is slightly wider than that obtained with point-like illumination sources, resulting in reconstructions with wider FWHM and therefore lower resolution. This only affects the resolution in the x dimension, because in the y dimension the source is effectively a point, therefore providing a slightly higher resolution. This effect may be resolved by scanning the surface of the medium additionally with line sources parallel to the y-axis. Although increasing the spatial resolution, this would double the measurement time. Later we present a comparison between the resolution in the x and y dimensions when illuminating with line-shaped sources.

Although the main purpose of this algorithm is to enable time-resolved wide-field applications, it is also valid for CW measurements. In this case, the amount of information for each source position is much lower. Therefore, many more sources at different positions are necessary to obtain similar results. In order to assess the performance of the presented algorithm in CW and time-resolved measurements, a testbench was designed to determine the resolution

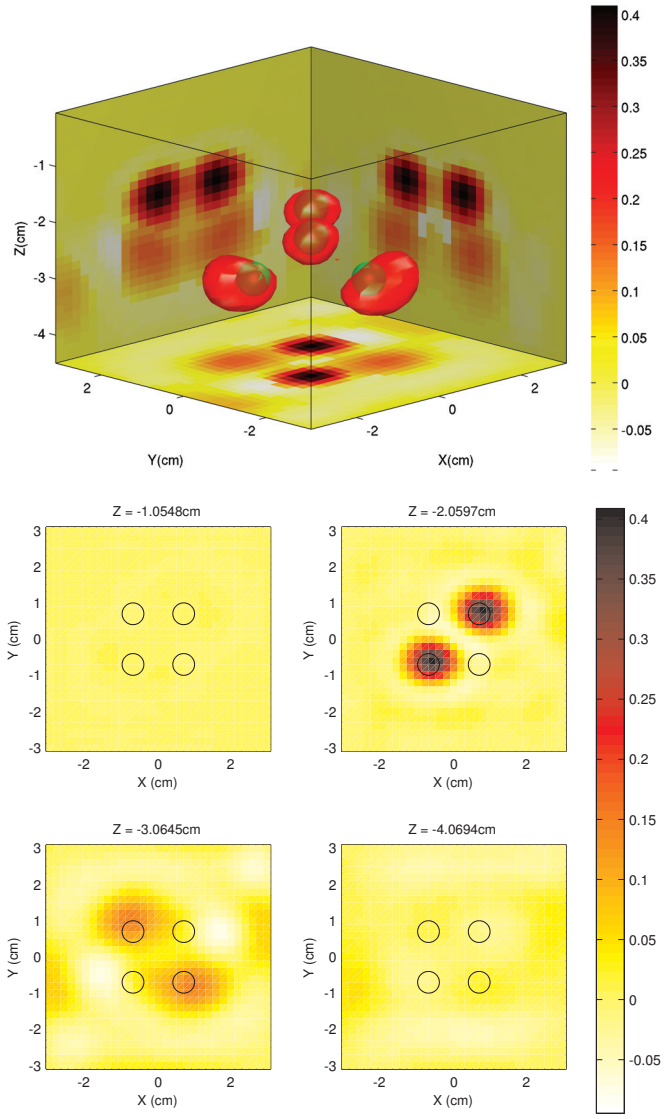

Figure 3.1 – Top: 3D reconstruction of four spheres with identical optical properties and an absorption contrast of 11:1 compared to the background medium. The 3D renderings of the original spheres (green) are visible with the rendering of the reconstructions at their respective FWHM (red). In the x=0, y=0, and z=0 planes projections of the maximum values of $\delta\mu_a$ in each dimension are depicted. Bottom: Slices of the reconstructions at depths z=[-1 cm, -2 cm, -3 cm, -4 cm].

and position accuracy. The background optical properties and the detector grid remained the same as before. For the time-resolved simulations the same four relative source positions Δy_s=[-2 cm, -1 cm, -1 cm, 2 cm] were utilized as before, whereas for CW case a total of 72 sources per detector row were positioned within the intervals [-4 cm, -0.5 cm] and [0.5 cm, 4

cm] with a separation of 0.1 cm. Reconstructions of a single sphere with contrast ratio 4:1 and 0.34 cm diameter placed at different depths in the middle of the field of view of the detector grid were performed to assess the resolution and the location accuracy in the z dimension. The results are displayed in Fig. 3.2 and Fig. 3.3.

Fig. 3.2 shows that for both CW and time-resolved cases the resolution in the x dimension is lower than in the y dimension due to the line-shaped illumination as previously explained. The resolution for time-resolved measurements is higher compared to CW as the sphere is deeper in the medium. Fig. 3.3 shows how the depth of the inclusion is only accurately detected up to a certain depth in CW reconstructions, whereas in time-resolved cases, the reconstructed depth is very accurate at all depths. These results show the benefits of time-resolved over CW measurements under similar conditions in wide-field NIROT. Moreover time-resolved measurements allow the simultaneous reconstruction of changes in absorption and scattering [Arridge and Schotland, 2009], a similar kernel as in Eq. 3.6 can be derived for changes in scattering.

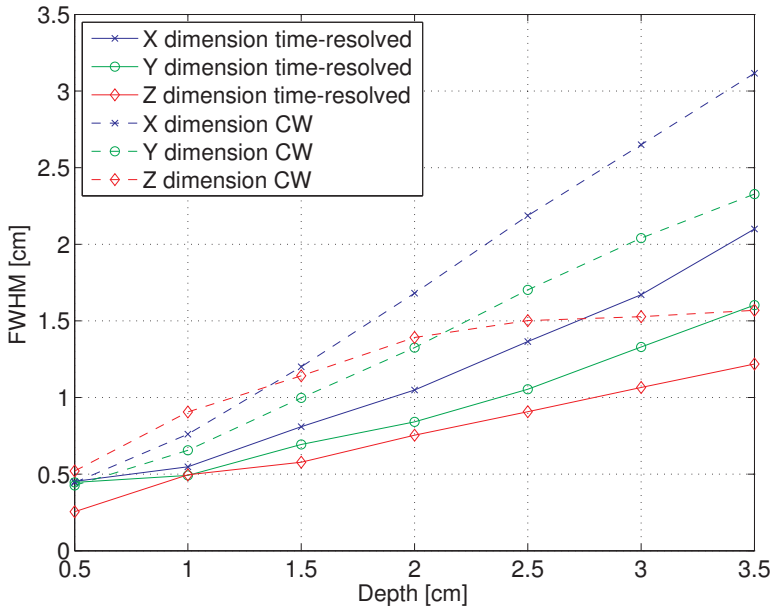

Figure 3.2 – FWHM of the image reconstruction of one sphere placed at different depths for time-resolved and CW data. Reconstructions with time-resolved data yield better resolutions since they provide more information. The resolution in the x dimension is in both cases higher than in the y dimension due to the line-shaped illumination.

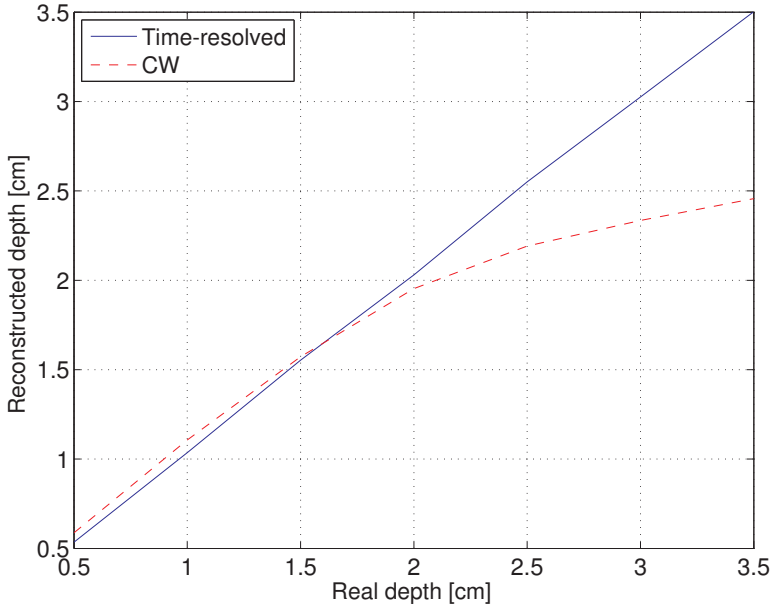

Figure 3.3 – Depth of the maximum $\delta\mu_a$ value of one sphere located at different depths for time-resolved and CW data. In CW the position of the sphere is only well determined up to a certain depth, while with time-resolved data the depth of the sphere is correctly resolved for all depths, demonstrating the benefits of time-resolved measurements for better depth accuracy in NIROT.

One critical aspect in NIROT is the time necessary to perform a measurement, therefore it is important to develop algorithms that minimize the number of acquisitions. For the time-resolved measurements in the presented algorithm 32×4 acquisitions, 4 relative positions for 32 detector rows, are necessary. However, if we account for the total number of sources' absolute positions instead of the relative positions, and we assume that there are several detector rows located at different distances away from the light source working in parallel, then the total number of acquisitions is considerably reduced. This would represent the case of a system in which the distance between the line source and the field of view of the camera is fixed and the whole system is moved to scan the surface of the object under study. In this case the total number of acquisitions required for the time-resolved and CW cases are 74 and 196 respectively. This is lower than what is required for algorithms based on structured light [Lukic et al., 2009]. Another important aspect is the time needed to reconstruct an image, especially in applications that require constant monitoring. Each reconstruction presented in this work required a computation time <15 s on a quad-core laptop computer.

3.3 Conclusion

We have presented a new fast reconstruction algorithm for time-resolved and CW NIROT setups that works in reflection mode, in which the illumination sources are always outside the field of view of the image sensor, solving the problems associated with small source detector distances and the internal light reflections in the camera lens. This allows the detection of absorption heterogeneities located a few centimeters deep in the tissue. Although only numerical simulations in reflection mode were presented, the algorithm can also be adapted for transmission mode systems as the above mentioned equations work for any laterally infinite medium.

4 CMOS Single-photon avalanche diode image sensors

The implementation of the first SPAD in CMOS technology [Rochas et al., 2003a] opened a new range of possibilities in the field of single-photon detection, allowing the on-chip integration of processing circuitry that so far had to be implemented on a separate chip. This was particularly interesting in applications requiring time-resolved measurements. In these applications it is critical to detect the time of arrival of the photons with the highest possible precision. This may best be achieved in architectures where the detection and time-of-arrival coding is realized in the same integrated circuit. Since they have become available in CMOS technology [Rochas et al., 2003b], SPAD arrays integrating time discrimination circuitry on chip have been utilized in a variety of applications requiring time-resolved measurements [Charbon and Donati, 2010]: near-infrared imaging (NIRI) [Mata Pavia et al., 2011, Mata Pavia et al., 2012], fluorescence lifetime imaging microscopy (FLIM) [Li et al., 2010], positron emission tomography (PET) [Meijlink et al., 2011, Mandai and Charbon, 2013b], and time-resolved Raman spectroscopy [Maruyama et al., 2013], to name a few.

4.1 SPAD operation principle and characteristics

When a photon is absorbed in the depletion region of a p-n junction it generates two free carriers, an electron and a hole. These free carriers either recombine between them, or they drift to the edges of the junction due to the electric field in the depletion region, generating

a current if they are connected to an electric circuit. The amount of carriers that recombine depend on the time they stay in the depletion region, which in turn is given by the speed at which they drift. To decrease the number of carriers that recombine and thus increase the generated current, an external voltage can be applied to the p-n junction to increase the electric field across the depletion region and thus the speed at which the carriers drift. When the electric field is high enough, a carrier can accumulate sufficient kinetic energy to generate a new electron-hole pair by impact ionization. These newly generated carriers will also be accelerated by the electric field creating new free electron-hole pairs by impact ionization. This process is known as an avalanche multiplication, and if it is not externally quenched, the generated current will be so high that it would eventually destroy the junction. The minimum voltage at which avalanches occur is known as the breakdown voltage, and the area in which the avalanches start is known as multiplication region.

SPADs or Geiger-mode APDs are a p-n junction reverse biased beyond breakdown. The avalanche building process is very fast, generating a high current within picoseconds upon the absorption of the photon. This effect has been exploited to detect photon arrival times with very high time resolutions [Cova et al., 1981].

A CMOS SPAD is usually formed by a flat p-n junction parallel to the photon detection plane surrounded by a guard ring, which spatially stops the avalanche process and prevents premature edge breakdown. The planar geometry of the diodes causes the fact that the electrical field at the edges of the p-n junction is higher than in the center, where the multiplication region is. Thus the breakdown voltage is lower at the edges, producing premature edge breakdown. If this is the case, most avalanche multiplication events are not optically but thermally generated at the edges of the SPAD. This results in a high noise with consequent reduction of sensitivity. The guard ring placed around the SPAD is usually lightly doped; its purpose is to reduce the electrical field at the edges, effectively preventing premature edge breakdown. This is one of the most critical aspects in the design of SPAD structures that probably has the highest impact on their performance. More details about the design of SPADs in CMOS technology can be found in [Fishburn, 2012, Niclass, 2008, Rochas, 2003].

4.1.1 Quenching and recharge

As previously mentioned, when an avalanche is produced in the multiplication region the generated current can destroy the device if it is not properly quenched. This is achieved by a circuit which lowers the voltage across the p-n junction below the breakdown voltage directly after an avalanche event, thus reducing the electric field at the p-n junction and effectively stopping the avalanche. After this, the SPAD bias voltage needs to return to the initial value, above the breakdown voltage. This is accomplished by the recharge circuit. The time elapsed during the quench and recharge phases is known as the dead time. During this period of time the SPAD is less sensitive to incoming photons and the probability that an avalanche is triggered is very low.

The simplest quenching and recharge circuits are implemented with a resistor or a MOS transistor connected in series with one of the terminals of the SPAD. When an avalanche is triggered, the current flowing through the quenching device increases the voltage across it, thus effectively reducing the SPAD's bias voltage and consequently quenching the avalanche. Once the avalanche has been stopped, the same device recharges the SPAD until it reaches the voltage prior to the avalanche event. This process is presented in Fig. 4.1 where the quenching and recharge circuit is implemented with a PMOS transistor.

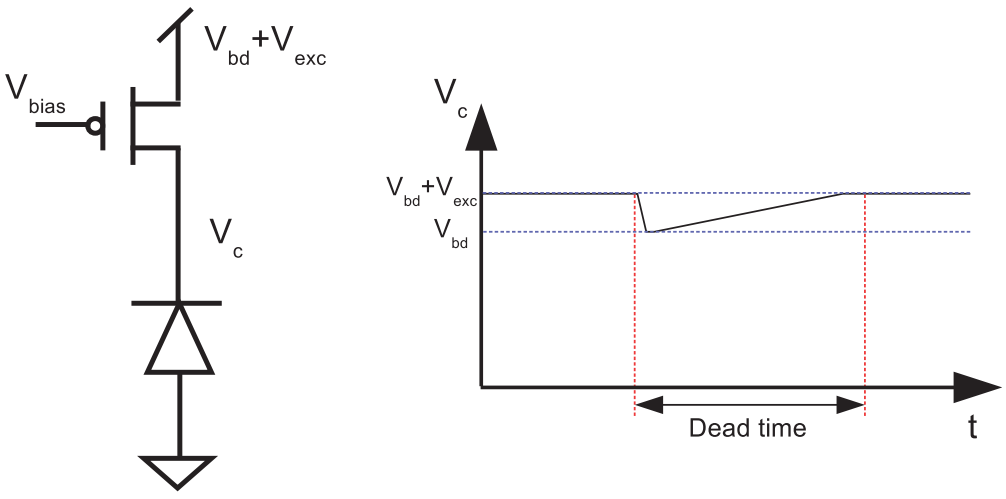

Figure 4.1 – SPAD quenching and recharge circuit implemented with a PMOS transistor. V_{bd} is the SPAD's breakdown voltage and V_{exc} the excess bias voltage.

4.1.2 Photon detection probability

The photon detection probability (PDP) in SPADs is defined as the probability of a photon that arrives at the detector's active area to generate an avalanche that can be detected by the discriminating circuit. The PDP is wavelength dependent and it is strongly related to the optical properties of the material in which the detector is implemented and the depth of the multiplication region. Photons absorbed too close to the surface generate electron-hole pairs that are likely to recombine before they can reach the multiplication region and start an avalanche. The same happens with photons that are absorbed too deep in the device. The depth at which photons are absorbed in a given material is defined as the mean penetration depth. Fig. 4.2 shows the mean penetration depth at which photons of different wavelengths are absorbed in silicon [Green, 2008]. In most SPADs implemented in CMOS planar processes the depletion region is located at depths between 100 nm and 1 μm, therefore the wavelength at which the PDP is highest between 400 nm and 600 nm. The PDP is related to the ionization rate in the depletion region, which in turn depends on the electric field across the p-n junction. Therefore it is possible to adjust the PDP by modifying the excess bias voltage. Although the PDP increases with the excess bias voltage, so does the dark count rate (DCR). Therefore, for each application, the PDP-DRC ratio can be adjusted to the optimum by the excess bias voltage. [Charbon and Fishburn, 2011] and [Rochas, 2003] present detailed derivations of the PDP as a function of the ionization.

4.1.3 Dark count rate

Avalanches triggered by carriers not generated by the absorption of a photon are known as dark counts. They occur also in complete darkness. The DCR is the amount of dark counts per unit time that are generated in a SPAD and is expressed in counts per second (CPS). Dark counts in SPADs can be classified in two types: thermally generated and tunneling assisted. Spontaneous transitions of electrons from the valence band to the conduction band due to thermal effects may occur in the depletion region of the SPADs. However, the large band-gap in most semiconductor materials makes these transitions very rare under normal conditions. Traps are defects in the semiconductor's crystalline lattice that can store carriers and release them after a period of time. Traps generate intermediate energy levels between the valence

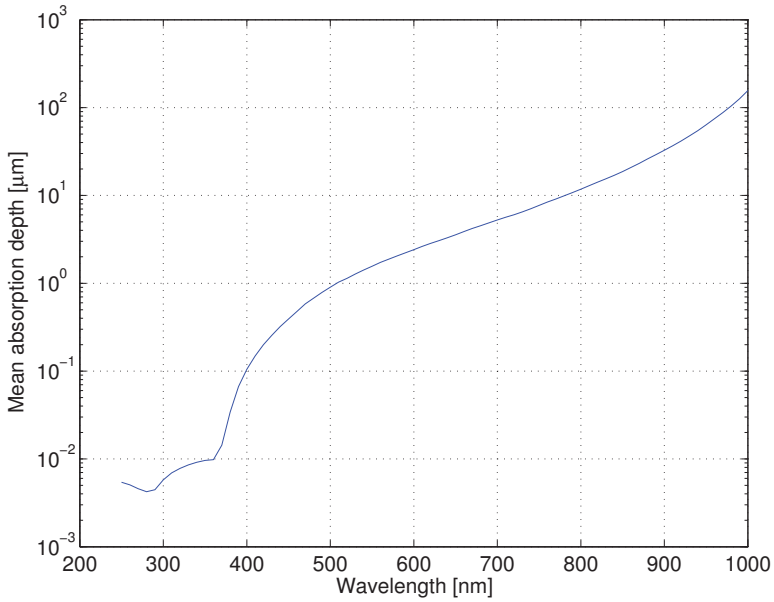

Figure 4.2 – Absorption depth in silicon for different light wavelengths. The absorption depth corresponds to the inverse of the absorption coefficient.

and conduction bands, increasing the amount of thermally generated carriers that reach the conduction band [Shockley and Read, 1952]. Thermally generated dark counts are highly temperature dependent, i.e. they increase exponentially with temperature. The amount of dark counts generated by quantum tunneling is highly dependent on the electric field in the depletion region, thus doping concentrations and excess bias voltage play a major role. As for thermally generated dark counts, the presence of traps in the lattice substantially magnifies this effect.

4.1.4 Afterpulsing

Carriers that are generated during the avalanche process may be trapped in the depletion or multiplication regions of the SPAD. After a period of time these charges are released provoking a secondary avalanche that is known as afterpulse. In order to reduce the amount of afterpulsing, the recharge time can be set so that when the carrier is released from the trap, the SPAD is not yet active.

4.1.5 Timing jitter

The statistical variations in the time elapsed between the absorption of the photon and the start of the avalanche process, determines the jitter of the electrical signal that is generated. The timing uncertainty of the avalanche triggering upon a photon reception depends on the absorption depth. Photons absorbed in the depletion region generate carriers that drift towards the multiplication region where the avalanche process is started, the depth at which the photon is absorbed therefore determines the jitter. Similarly the lateral position of the photon absorption also influences the avalanche build up time [Spinelli and Lacaita, 1997], having an influence on the jitter. The timing variations due to the described effects follow a Gaussian distribution that dominates the jitter of the generated electrical signal.

Photons absorbed beyond the depletion region generate carriers that recombine most of the times, however, some of them diffuse to the multiplication region generating an avalanche. The deeper a photon is absorbed, the longer it takes to trigger an avalanche and the lower the probability is that the carrier reaches the multiplication region. Since diffusion of carriers is a slow process, the timing jitter induced by this effect is orders of magnitude higher than the Gaussian jitter. This results in a long exponential tail in the time distribution of the generated signal that appears after the initial Gaussian peak.

4.1.6 CMOS SPAD's state of the art

In [Veerappan and Charbon, 2014] a comparison of state-of-the-art CMOS SPADs has been published including data from recent publications [Gersbach et al., 2008][Niclass et al., 2007][Richardson et al., 2009a][Bronzi et al., 2012][Webster et al., 2012b][Webster et al., 2012a][Mandai et al., 2012][Wu et al., 2013][Leitner et al., 2013][Niclass et al., 2008]. The main figures of merit when comparing the SPADs' performance are PDP and DCR. Ideally the PDP should be the highest and the DCR the lowest possible. However, depending on the application, a lower PDP or higher DCR may be acceptable. Figures 4.3, 4.4 and 4.5 summarize the SPAD's state-of-the-art according to their PDP and DCR performance.

SPAD structures with isolated anodes and cathodes exhibit shallow multiplication regions and narrow depletion regions, therefore their detection peak is between 400 nm and 500 nm, and

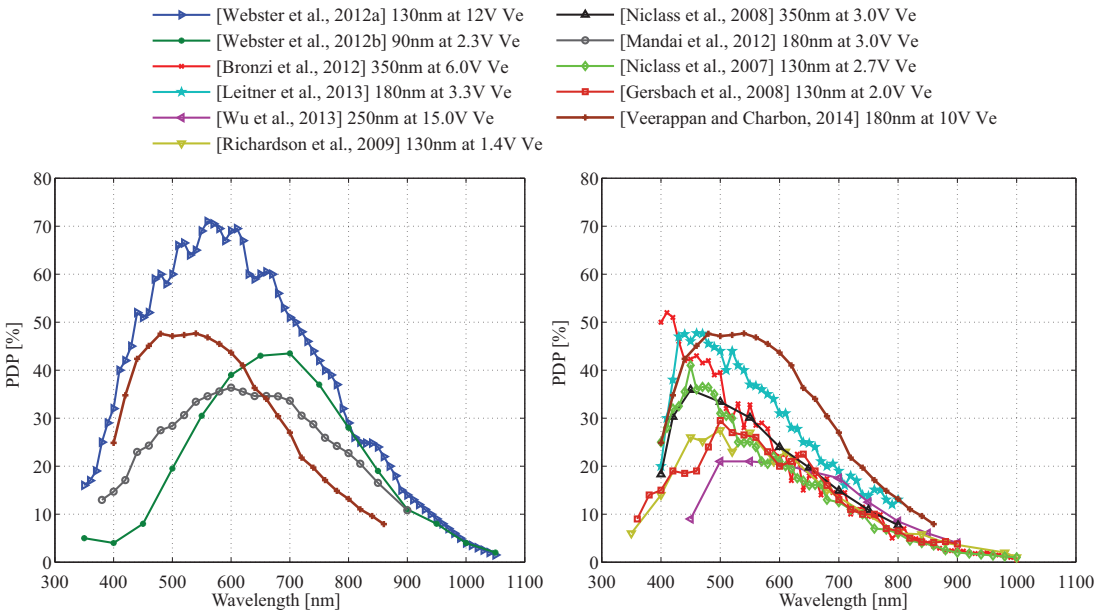

Figure 4.3 – Comparison of state-of-the-art PDP in non-isolated devices (left), and in substrate isolated SPADs (right). The CMOS technology feature size and excess bias voltage (V_e) are specified for each SPAD. *Picture credit: C. Veerappan ©2014 IEEE*

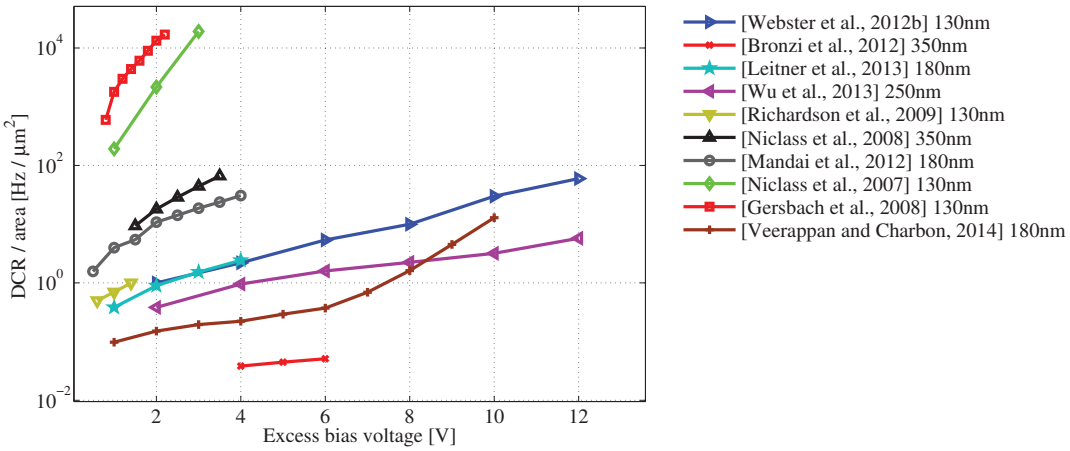

Figure 4.4 – Comparison of state-of-the-art SPAD's DCR. *Picture credit: C. Veerappan ©2014 IEEE*

their PDP is spectrally rather narrow compared to devices that share their anode (Fig. 4.3). Structures with isolated anode provide an accurate timing response and are easy to integrate in standard CMOS circuits, because they produce a low voltage output signal.

SPADs that share one of their terminals, usually the anode as most CMOS processes utilize

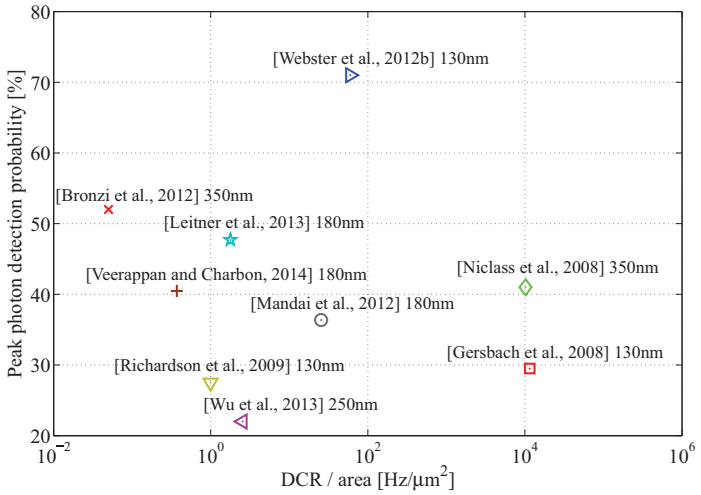

Figure 4.5 – PDP versus DCR. Comparison of the state-of-the-art for different structures implemented in CMOS processes with diverse feature sizes. *Picture credit: C. Veerappan ©2014 IEEE*

a P-substrate, present a wider spectral response, and a a peak performance almost in the near-infrared region, between 600 nm and 700 nm. However their time response is normally not as accurate as the one achieved with isolated devices. In addition their interface with standard CMOS devices poses a problem, because they have a high voltage output.

The noise in SPADs is influenced by the doping concentrations and the integrity of the crystalline structure in the multiplication and depletion regions, and in the guard ring. In deep submicron processes high doping concentrations are employed leading to a high number of trap-assisted tunneling dark counts. The first SPADs implemented in 130 nm presented a large DCR due to this effect [Niclass et al., 2007, Gersbach et al., 2008]. Structures employing lightly doped materials have led to devices with improved levels of DCR (Fig. 4.4). In conclusion, the PDP and DCR vary substantially between different types of SPADs and need to be optimized for a specific application and CMOS process.

4.2 The LASP image sensor

The LASP image sensor was, to our knowledge, the first SPAD array that integrated on-chip TDCs [Niclass et al., 2008]. It consists of a SPAD array of 128×128 SPAD pixels connected to a bank of 32 TDCs. The fill-factor of the pixel is 6%. The TDCs have a time resolution, 1 least-

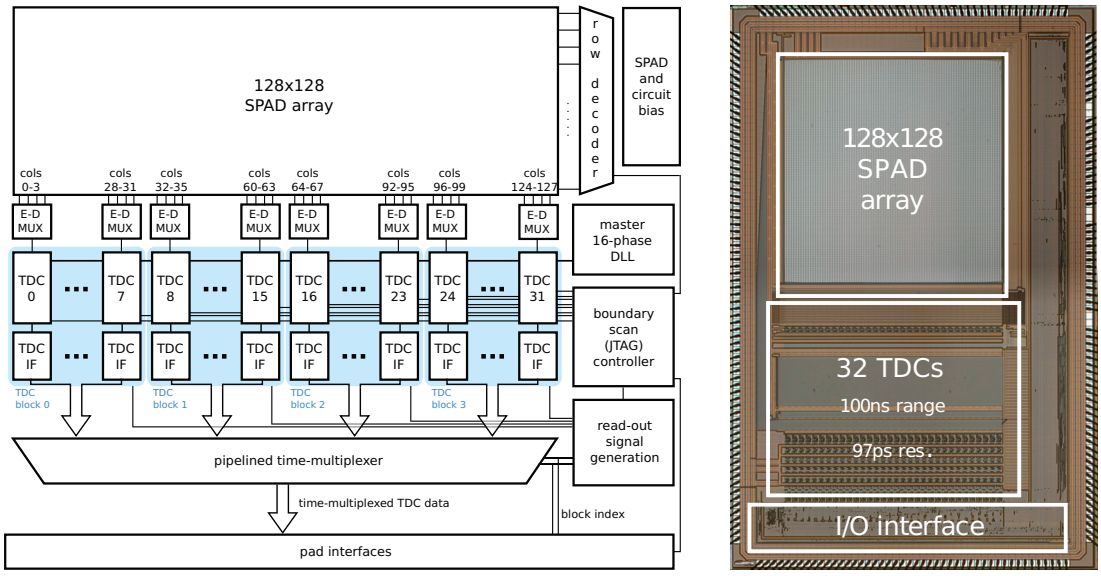

Figure 4.6 – (left) Block diagram of the LASP image sensor.(right) Microphotograph of the sensor fabricated in a standard CMOS HV 0.35 μm process. The dimensions of the active area are 3.2×3.2 mm^2. *Picture credit: C. Niclass ©2008 IEEE*

significant bit (LSB), of 97 ps and a range of 100 ns. When a row of pixels is selected, a direct connection between those pixels and the TDC bank is established via a 4-to-1 event-driven multiplexer. Consequently the sensor is capable of providing time resolved measurements for each of its pixels independently, making it suitable for TCSPC measurements. Since only one row can be selected at any time, the readout of all rows requires a relatively long acquisition time. This is the most critical aspect in terms of operation speed of the sensor. Fig. 4.6 shows a schematic of the sensor's architecture and a microphotograph of the chip.

The TDC integrated in this SPAD image sensor is based on a three-level interpolation in order to achieve a relatively large range of operation of 100 ns, while reaching a high temporal resolution of 97 ps in a compact manner. Fig. 4.7 shows a basic diagram of the interpolation principle implemented in the TDC and a simplified schematic of the TDC. This TDC has two operation modes: measurement and calibration mode. The global CLK signal serves as a time invariant reference for each TDC. From the CLK signal, 16 global signals PHI[0:15] with uniformly distributed phases are generated by means of a global DLL. In each TDC, the PHI[i] signal triggers a 2-bit counter. In measurement mode the CALRQ signal is deasserted, and the input multiplexer connects the STOP signal with TRG. When a photon arrives, the TRG signal is asserted which in turn triggers a register that stores the state of the counter. The period of the

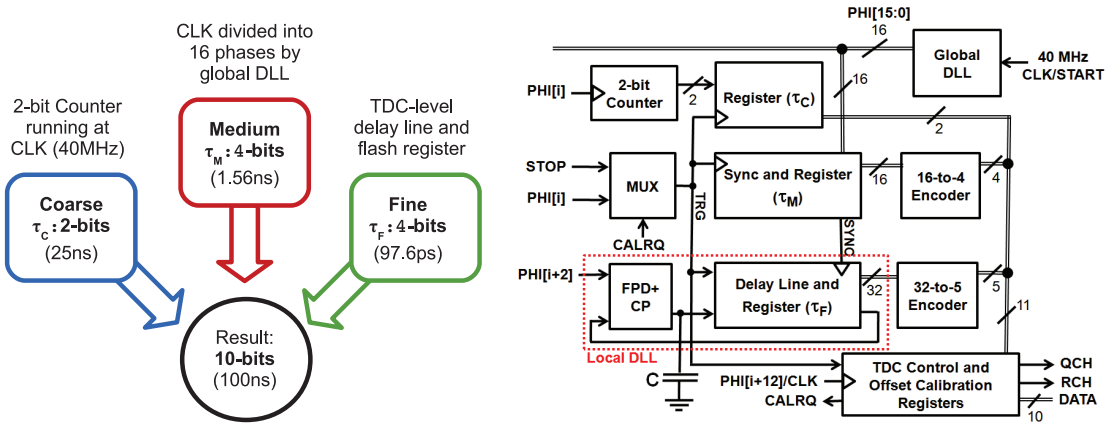

Figure 4.7 – (left)TDC interpolator principle. Coarse and medium resolution are generated globally and fine resolution is implemented on each TDC. (b) Simplified TDC block diagram . *Picture credit: C. Niclass ©2008 IEEE*

PHI[i] signal determines the resolution of the coarse resolution τ_C. Simultaneously the state of the 16 signals is locally stored in a register in each TDC, which after decoding generate the 4-bits of the medium resolution stage. The resolution of this block τ_M is thus the delay between two adjacent phases. In the fine resolution block, the TRG signal is transmitted through a 32-tap delay line whose state is sampled on the next transition of the SYNC signal. Only 16 of these delay elements are employed to generate the 4 LSBs, the rest are used to synchronize the SYNC and TRG signal in the fine resolution block. The SYNC signal is generated in the medium resolution block and it is triggered on the next PHI[0:15] transition after a photon reception. The resolution τ_F of this block is determined by the propagation delay of a single element in the delay line, which in turn is 97 ps.

The TDCs need regular calibration of the fine delay lines to compensate for PVT changes. When the TDC is in calibration mode, the CALRQ signal is asserted connecting the PHI[i] to TRG via the input multiplexer. A local DLL is formed with the fine delay line in combination with a phase frequency detector (PFD) and a charge pump (CP). The DLL is then run until it locks PHI[i] and PHI[i+2]. The resulting analog voltage is then stored in a local capacitor which is later used to bias the delay line in measurement mode.

4.2.1 New sliding scale method for improved TDC linearity

The TDC implemented in the LASP sensor presented high levels of differential nonlinearity (DNL) and integral nonlinearity (INL). When a signal with random starting time points with respect the reference clock is applied to the TDC, all the codes should have the same probability of occurrence. However, due to mismatch of the delay elements and desynchronization of the different blocks in the TDC, some codes will have a higher probability of occurrence than others. In the fine resolution block, mismatch in the delay value of the different taps causes specific codes to occur more often than others. Since two types of inverters are employed in the fine delay line, these need to be perfectly calibrated by biasing them to obtain exactly the same delay in both types. When acquiring signals in TCSPC mode, a mismatch between the delay of these two types of inverters leads to a ripple in the measured signal that has a period of 97 ps (1 LSB). The medium resolution block generates the SYNC signal that latches the values in the delay line of the fine resolution block. Therefore it is critical that these two blocks are perfectly synchronized. This is achieved by adjusting the delay between the SYNC and TRG signals at the fine resolution block with two dedicated delay lines. This calibration scheme only functions properly under the assumption that in the generation of the SYNC signal, the 16 signals that form PHI[0:15] have a phase shift equally distributed in one clock period. Any error in the phase difference cannot be compensated. Therefore it generates nonlinearities in the TDC response. Variations in the phase of PHI[0:15] result in jitter in the SYNC signal which translates in desynchronization between the medium and fine resolution blocks. When the jitter in the SYNC signal produces a positive delay, e.g. it arrives later than it should, the code that indicates that the TRG signal was transmitted through all the delay taps of the fine resolution block occurs more often than the others. In case of a negative delay, e.g. the SYNC signal arrives earlier than it should, a similar imbalance will occur. This desynchronization effect results in peaks and valleys in the TDC probability response to a time uncorrelated signal that repeats periodically every 16 LSBs, which corresponds to the period of the fine resolution block. Finally, in the coarse resolution block, the state of a two bit counter is latched upon the reception of a photon. Between this block and the medium resolution block, there is no synchronization mechanism. Any mismatch in the delays of the signals that latch the results in both blocks or in the latching mechanism themselves, will produce codes that are

shifted multiples of 1-bit from the medium resolution block. This type of error is more difficult to detect and to compensate for than the non-linearities produced in the fine and medium resolution blocks, because they are not visible in the width of the bins of the TDC probability response. If a time uncorrelated input signal is used to trigger the TDC, this nonlinear effect will produce a response similar to the one obtained when TDC bins are widened or shrunk. This behavior can nonetheless be identified by employing an input signal with a fixed delay with respect the TDC reference clock. By linearly increasing the delay of the input signal, the converted code should increase accordingly. If an abrupt change in the generated code occurs, the nonlinearity will be identified.

Fig. 4.8 shows the response of a TDC implemented in the LASP chip when it is triggered by a time uncorrelated signal. The signal exhibits a periodic behavior, since it expands over 4 clock cycles. Thus the peaks and valleys are repeated in each of the four clock periods. In this figure the nonlinearities produced by the fine and medium blocks desynchronization are clearly visible because they repeat every 16 LSBs, i.e. 1-bit of the medium interpolation block. The height of these peaks and valleys depends on the phase misalignment of the corresponding PHI[0:15] signal. In the same figure the number of counts have been classified according to their last 4-bits. In this classification the nonlinear codes are identifiable, because they present

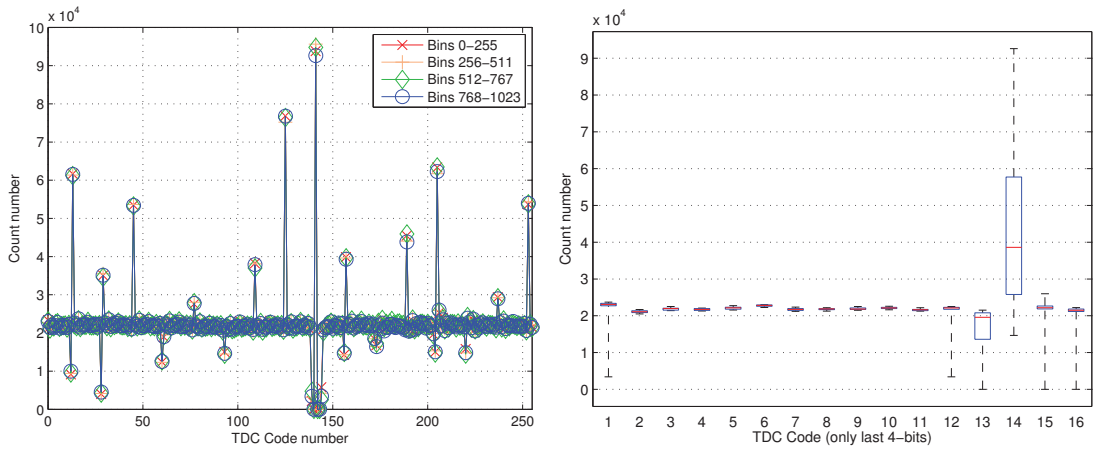

Figure 4.8 – (left) TDC response to a time uncorrelated input signal. As the TDC range covers 4 clock cycles, the response exhibits a periodic behavior. The four cycles have been superimposed to better show this effect. (right) Statistical classification of the TDC conversion codes according to their last 4-bits. The central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points.

high count number variations. It is remarkable here how linear the response is in the codes that do not suffer from the effects of desynchronization, this indicates that the linearity of the delay lines is excellent.

A sliding scale method was applied similar to the one published in [Sumner, 2001] in the results presented in [Niclass et al., 2008] to improve the linearity of the TDC. A 64 tap programmable delay line with 1 LSB delay step was employed to apply a deterministic delay on the clock signal of the TDC. After the TDC conversion, the corresponding code of the applied delay is digitally subtracted from the result. Therefore events arriving at the same time point make use of different parts of the TDC transfer curve, effectively averaging the error of the nonlinear bins over many different bins and consequently improving the DNL. It has to be noted that the averaged nonlinearities and errors in the variable time delay introduce a considerable amount of jitter in the TDC, decreasing its accuracy. To minimize this, delay errors should be much less than one LSB [Cottini et al., 1963, Knoll, 2010]. This mechanism substantially improves the DNL as shown in Fig. 4.9, though the INL does not change significantly. In order to improve INL, a variable delay that covers the whole TDC range should be employed [Sumner, 2001]. However it is impossible to obtain such a delay line that spans over such a long time range with a high accuracy. Moreover, as shown later in this section, the sliding scale mechanism does not reduce the distortion introduced by the desynchronization between the coarse and medium resolution blocks.

In the present TDC the nonlinear behavior is localized in a few bins of the TDC, which repeat periodically. In order to reduce the influence of these bins, it is possible to discard the conversions that result in one of the codes that present high levels of nonlinearity. By doing so, the TDC linearity improves dramatically, but only a selected region of the whole range is covered. The sliding scale mechanism can be applied, to recover the whole range. Fig. 4.9 shows the result of applying the presented bin gating mechanism with 16 sliding scale steps, which is the period of the nonlinear behavior. Since with this method the number of sliding scale steps required is lower than in the traditional one, the amount of jitter introduced in the system is considerably reduced.

Earlier in this chapter the nonlinearity produced by the desynchronization of the coarse and

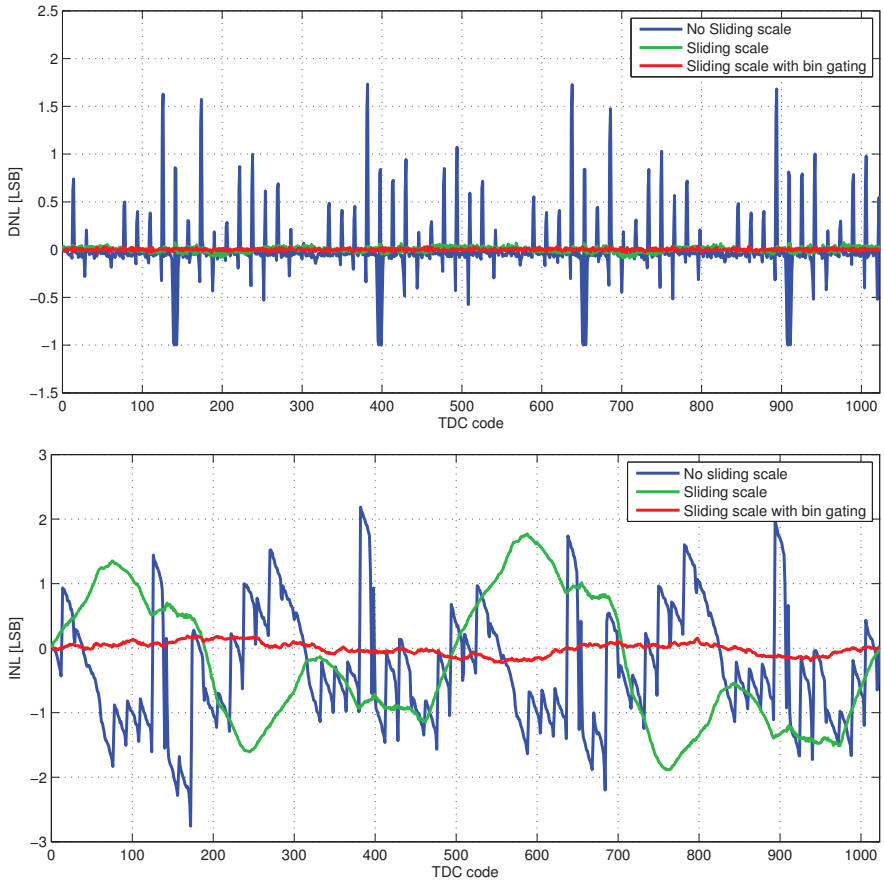

Figure 4.9 – (top) DNL measurements of the TDC, both sliding scale methods substantially reduce the DNL. (bottom) The INL in the other hand does not improve much with the traditional sliding scale mechanism. However, when the nonlinear TDC codes are excluded and the proposed bin gating method is applied, the INL dramatically improves.

medium resolution blocks were described. Since the input of the TDC in the LASP chip is directly connected to the SPAD pixels, it is not possible to generate the required input signal to quantify this type of non-linearity. It is nevertheless possible to qualitatively measure the influence of this nonlinear effect in a real measurement. Fig. 4.10 shows the acquisitions performed by a TDC when the SPADs were illuminated with a picosecond laser (SuperK Extreme, NKT Photonics) working at 650 nm. When the pulse arrival time is far from a CLK transition, small peaks appear on the measured response since some bins in the TDC probability response have a higher occurrence. When applying the sliding scale mechanism, these peaks effectively disappear, because they are averaged over 64 bins. When the pulse arrives during a transition of the CLK signal, substantially larger peaks will also appear. The latter are a consequence of

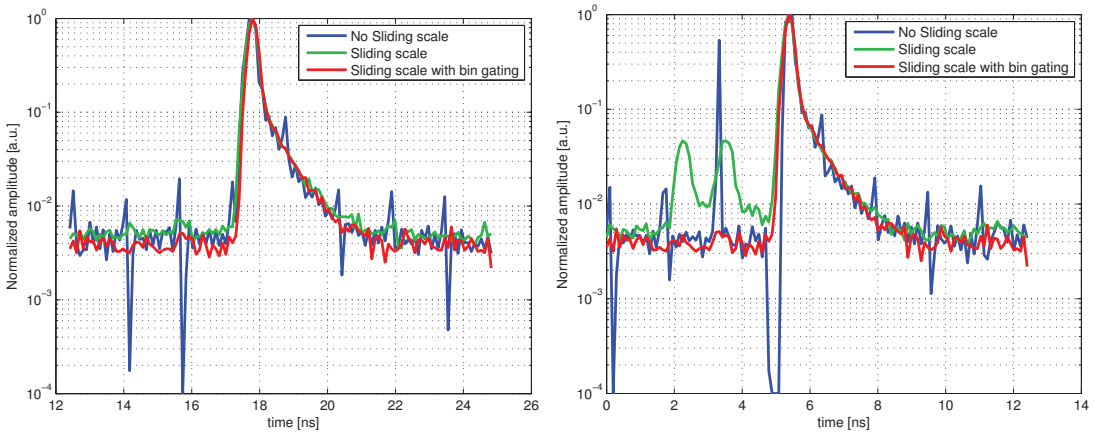

Figure 4.10 – TDC measurements of a picosecond pulse of light detected with a SPAD employing different sliding scale methods: (left) when the pulse arrives far from a transition of the coarse interpolation block and (right) when the pulse is centered in a transition of the coarse resolution block.

the nonlinear behavior produced by the desynchronization between the coarse and medium resolution blocks. Their amplitude depends on the position of the measured optical pulse. Therefore after applying the sliding scale mechanism these peaks are smoothed, but they do not completely disappear. But when the bin gating method is applied, these peaks are effectively removed.

In conclusion, a new sliding scale method has been introduced that can be applied to TDCs to improve their linearity. The method requires a low number of delay steps to be applied in the input signal, consequently reducing the amount of jitter that is introduced in the system compared to the original sliding scale method. Moreover with a very small set of delay steps it is capable of improving the DNL and INL of the TDC.

5 Microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery[1]

Since they have become available in CMOS technology [Rochas et al., 2003a], SPADs have been integrated in large arrays, enabling the implementation of application specific in-pixel standard CMOS electronics. The large amount of in-pixel circuitry integrated in some of these arrays has resulted in pixels with very low fill factor (i.e. sensitive area per total sensor area), thus reducing the PDE. A lower PDE results in a decrease of the signal to noise ratio (SNR) and therefore in longer acquisition times. Due to the complexity of the in-pixel circuits implemented in SPAD sensors, often the fill factor shrinks to 1% [Veerappan et al., 2011]. Conventional lenses focus the light in a continuous manner, that means they can focus light in an area to create an image or a single spot. However, in non-continuous detectors, such as imagers with many low fill factor pixels, most of the light hits the inactive area between pixels. An effective way to reclaim some of this lost light is the use of micro-optical devices with sizes comparable to the pixel's size. Microconcentrators, such as plano convex microlenses, have been extensively applied to CCD and CMOS sensors [Deguchi et al., 1992] to address this problem. Other geometries implemented in solar panel technologies have been proposed to recover even more light [Donati et al., 2007], however so far it is very difficult to fabricate them. The main difference between SPAD and CMOS/CCD imagers, is that the former have much lower fill factor and therefore the implementation of proper microconcentrators is critical.

[1] The material of this chapter is derived from: **[Mata Pavia et al., 2014a] Mata Pavia, J. , Wolf, M., and Charbon, E. (2014). "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery, *Opt. Express* 22(4):4202-4213.**

Though recovery factors of ×25 have been reported in detectors where the initial fill factor was particularly low [Donati et al., 2011], arrays of microlenses with high recovery factor and high uniformity have not been reported for practical sizes.

The aim of this chapter is to present the study of an array of 128x128 plano convex microlenses fabricated on the LASP SPAD array, which have 5% fill factor and 25 μm pitch [Niclass et al., 2008]. A single microlens is placed in front of each pixel, focusing the light in the active area, thus increasing the PDE. An accurate model for the microlens was conceived, verified by extensive ray-tracing simulations, and validated by measurements. The model was in excellent agreement with the measurements under a wide variety of real-life conditions (f-number, illumination levels, etc.), and will be a useful tool in the design of future arrays of microlenses.

5.1 Microlens manufacturing

The microlens arrays were fabricated by CSEM Muttenz (Switzerland). A quartz mold was utilized to imprint the microlenses in a sol-gel polymer on top of the SPAD array. The polymer used for the fabrication of the microlenses was ORMOCER®. In the manufacturing process the polymer is deposited on top of the imager, the mold is then brought into contact with the polymer and pressure is applied. The polymer is then UV cured and thermally stabilized in an oven. The detector was produced by standard CMOS 0.35 μm high voltage technology, with a pixel pitch of 25 μm and a photosensitive area of 6 μm in diameter, resulting in a fill factor of approximately 5%. A graphical representation of nine pixels is shown in Fig. 5.1. The microlenses have a radius of curvature of 21.77 μm and a conic factor k=1.6, further geometrical details can be seen in Fig. 5.1. The microlenses focus the incoming light into the photosensitive area of the pixel. Microlenses of the same mold with different heights were produced in order to assess their performance. Scanning electron microscope (SEM) pictures of sample chips were taken (Fig. 5.2) to assess the quality of the fabrication process. Profilometries of all samples before bonding were performed to verify the height of the fabricated microlenses. A Tencor® profilometer is used to measure the height of the microlenses on six points at the edges of the replicated array. Figure 5.2 shows a chip with microlenses bonded in a package.

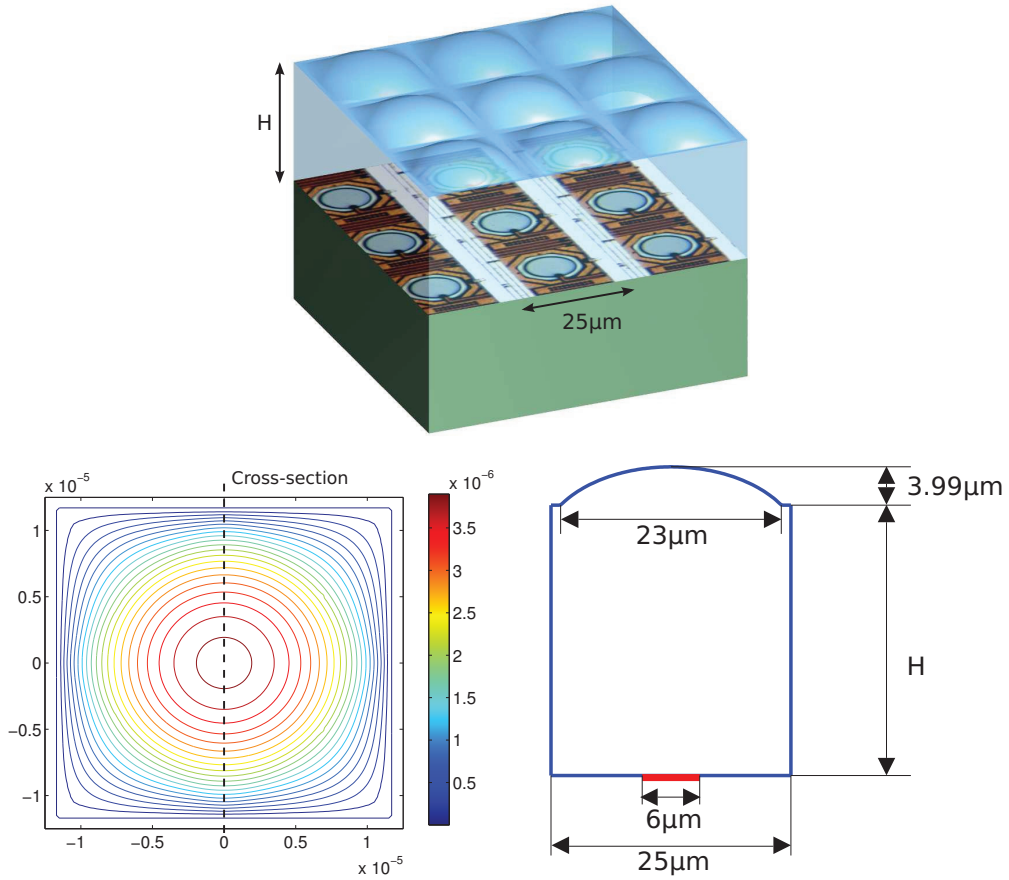

Figure 5.1 – (top) Detail of array of microlenses mounted on top of nine SPAD pixels. The height of the microlenses H can be adjusted for each sample. (bottom left) Top view of a single microlens with isolines that depict its shape. In this picture it can be seen that the base of the microlens is square, in other words the radius of curvature of the microlenses varies with the azimuthal angle. (bottom right) Cross-section across the dotted line of the microlens with geometrical details. The active area of the pixel is depicted in red.

5.2 Microlens concentration factor simulator

A ray tracer was implemented in Matlab to study the different microlens parameters and to compare the experimental results with our model. This solution was adopted in order to be able to simulate microlenses with arbitrary surfaces. As SPAD imagers are mostly targeted to imaging applications, it was decided to implement a simulator in such a way to include the optical system in which the imager would be ultimately integrated. For this reason, the simulator includes four main optical elements: an object plane, a thin lens that mimics the application's optical system, the microlenses, and the image plane. Thin lenses with different

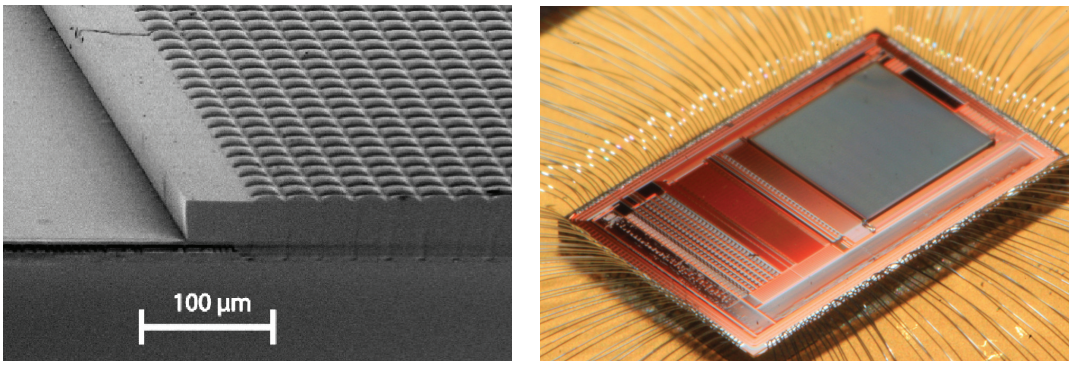

Figure 5.2 – (left) Scanning electron microscope image of the microlens array fabricated on the CMOS SPAD array. (right) SPAD array chip with microlenses bonded in a PGA package (right).

f-number and focal length can be set in order to simulate any given optical setup. It is also possible to generate collimated light in the simulator that hits the detector with different angles of incidence. The simulator was verified by comparing its results with those from a commercial ray tracer (WinLens, Qioptiq, Goettingen, Germany). An optical setup consisting of a thin lens and a circular microlens was simulated with both tools. The results showed good agreement indicating that our simulator is accurate.

The ray's origin points are equidistant in the object plane. From each of these points a specified amount of rays are then generated uniformly covering a predefined solid angle. When one of these rays interacts with a surface (thin lens, microlens, metal wire, etc.) the new trajectory and intensity of the refracted/reflected ray is calculated. The intensity of the detected light is calculated adding the energy of all the rays that hit the photosensitive area of the pixel (Fig. 5.3). An array of nine microlenses and nine detectors is simulated to account for optical crosstalk with neighboring microlenses. In order to evaluate the microlens performance, we use the concept of concentration factor CF that has been already proposed [Donati et al., 2007, Donati et al., 2011]:

$$CF = E_0/E_i, \tag{5.1}$$

where E_i is the irradiance (power per unit area) at the input of the microconcentrator, in this particular case the microlens' convex surface, and E_o is the output irradiance of the microconcentrator, that would be the irradiance at the photosensitive area of the pixel. In the current configuration E_i is equivalent to the irradiance that the same photodetector without

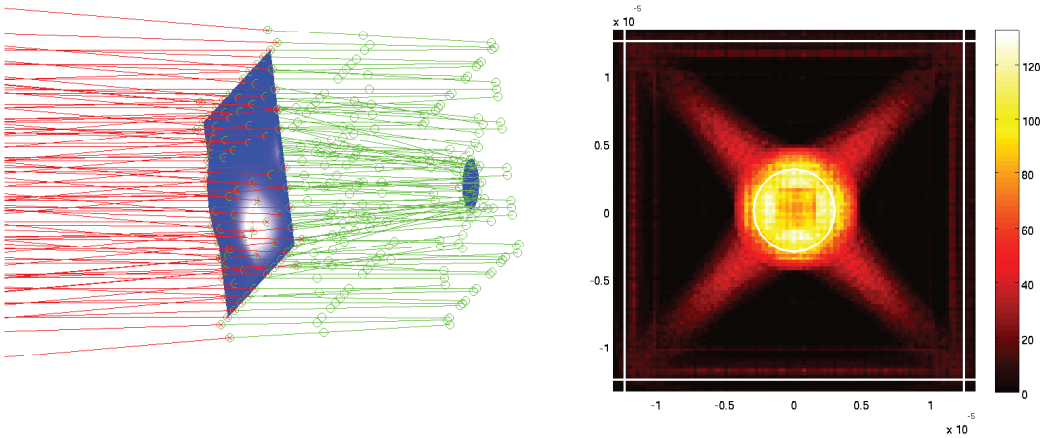

Figure 5.3 – (left) Ray tracing of light going through a microlens and then reaching the image plane, the circle on the right denotes the photosensitive area of the pixel. Although nine microlenses and nine detectors are simulated, only one microlens and detector are depicted in this figure for clarity purposes. (right) Light intensity profile in the image plane. The outer white straight lines defines the limits of the pixel and the inner white circle the pixel's photosensitive area.

microlenses would receive. For this reason, the microlens concentration factor is obtained by dividing the detected light intensity when microlenses are present by the detected light intensity when the detector has no microlenses.

Since the refractive index of the material used to fabricate the microlenses is quite uniform in the visual spectrum, less than 2.5% variation between 400 nm and 700 nm, and for ease of comparison with previous work [Donati et al., 2011] which was carried out with white light illumination, the results presented in this work assume a constant refractive index in the microlenses for all wavelengths.

In order to obtain accurate estimates, the discretization error needs to be kept very low, which means that a high number of rays needs to be generated. The simulator has been equipped with Monte Carlo capabilities enabling reliable preliminary results with a reduced number of rays. In this case the distribution of the rays origin points in the object plane is randomized with a uniform distribution, and at each of these points the solid angle is also randomly sampled with a uniform distribution.

5.3 Experimental results

The concentration factor of microlenses has been tested on the above mentioned 128x128 SPAD array. The initial microlenses were designed for a microscopy application in which the numerical aperture (NA) in the image plane was close to 0.005. The microlens designed back focal-length (BFL) for this particular application was 40 μm. The same microlens mold was applied to produce microlenses of different heights. Several samples were manufactured at a variety of heights. We report in this paper two samples at 30 μm and 70 μm height. The height of microlenses is defined as the distance from the base of the microlenses to the photosensitive area in the microchip. Figure 5.4 shows a SEM cross-section of a diced chip where the height of the microlenses can be measured.

All the simulation and experimental results reported in this section where obtained by means of a double telecentric objective lens (CVI Melles Griot 59 LGC 916 base lens with 59 LGJ 423 attachment lens) with a magnification factor $\beta = 0.07$. This objective lens is composed of two main elements, a singlet front element (double convex singlet, 200 mm focal length, 110 mm

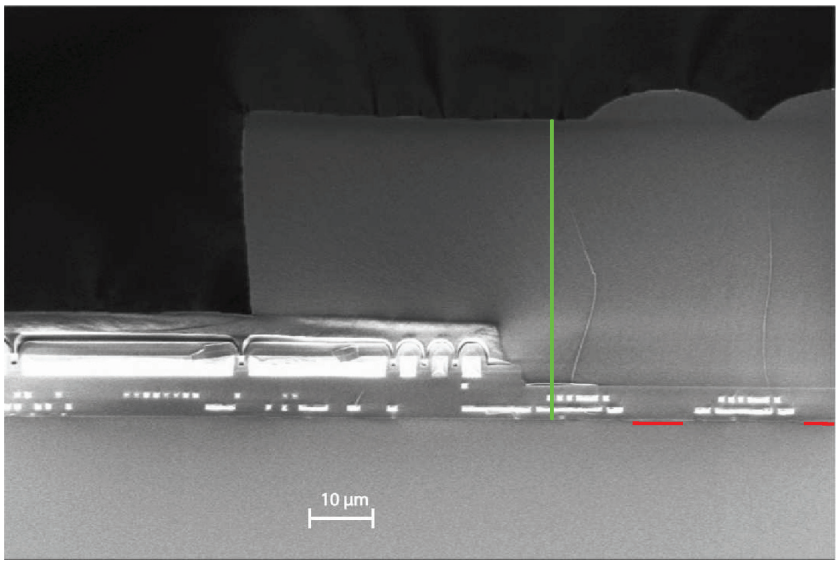

Figure 5.4 – SEM picture of the cross-section of one of the chips with microlenses. In the array's photosensitive area, marked in red, the upper layers of the microchip have been etched away to increase the performance of the sensor. Metal 1 through 4 are also visible in the cross-section, covered by 2 μm of SiO_2 passivation and a polyimide layer. The height of the microlenses is marked in green. For this particular sample the height of the microlenses was 51 μm.

diameter) and a rear objective lens (15 mm focal length, f/1.4-f/22).

Simulations showed (Fig. 5.5) that the performance at different f-numbers can be optimized by adjusting the height of the microlenses. As expected for high f-numbers, the optimal performance is achieved when the microlenses' height is similar to the calculated BFL of the microlenses. However for low f-numbers a small height of the microlenses leads to a higher concentration factor. The latter is due to the fact that the concentration factor obtained from skew rays at large apertures is higher than the one obtained from paraxial rays. Figure 5.5 right shows how the maximum concentration factor for each f-number shifts towards smaller heights as the f-number decreases.

The correct height of the microlenses is a key parameter that needs to be controlled in the fabrication process because it has a tremendous impact on the performance of the sensor. Figure 5.6 shows substantial concentration factor variations for small height changes. The microlenses' total height depends not only on the height of the sol-gel polymer, but also on the different metal and SiO_2 layers that are on top of the photosensitive area of the sensor. This means that for an accurate production of the microlenses, the actual height of the different layers in the sensor needs to be measured beforehand. Typically the CMOS process specifications state that variations of up to ± 2 μm are possible.

The heights of the different layers were measured in the SEM cross-section (Fig. 5.4). Since all

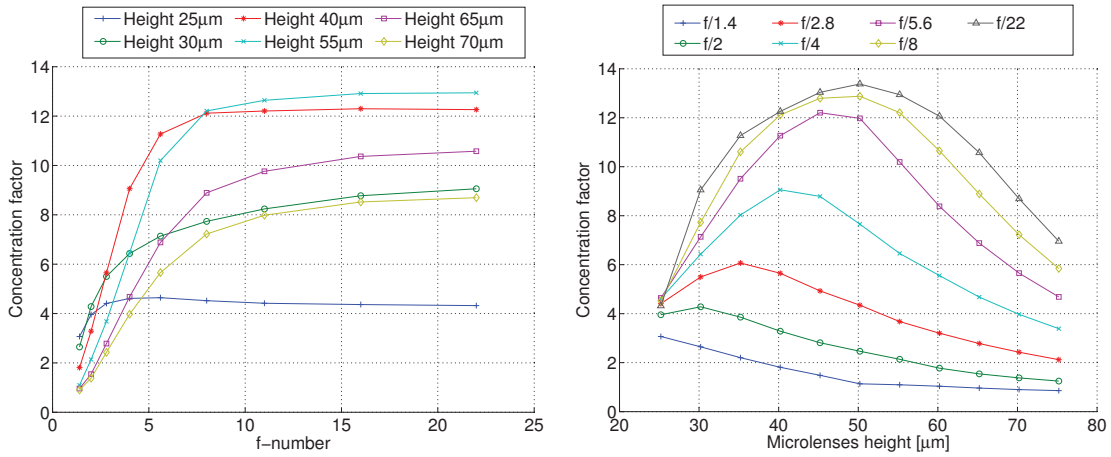

Figure 5.5 – (left) Simulated concentration factors for different heights of the microlenses. (right) Depending on the application's f-number, the optimal height of the microlenses varies.

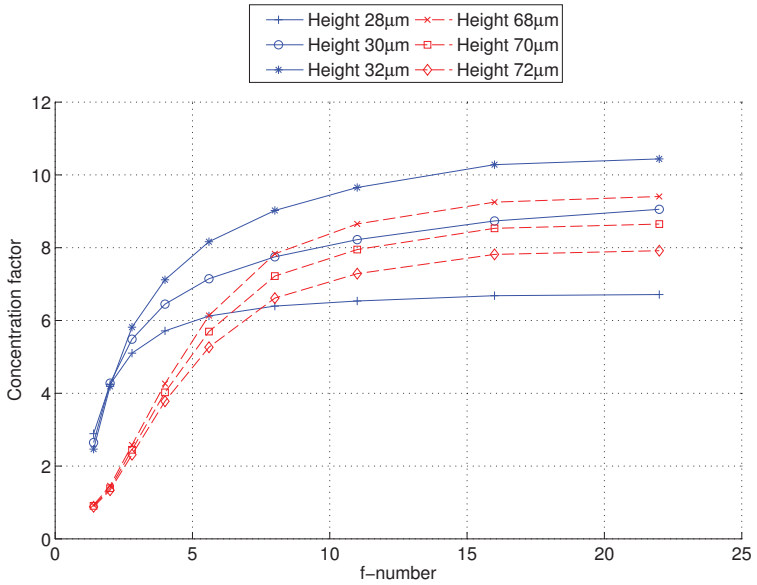

Figure 5.6 – Concentration factor variations for small height changes.

chips originate from the same wafer, we assume that the heights of the different layers were the same for all of them. The total height of the layers related to the CMOS process differed from the specified values by 1 μm. A profilometry was performed on each sample to measure the height of the imprinted microlenses with high accuracy. This value together with the height obtained from the SEM measurement corresponds to the total microlens height of each sample.

The concentration factor is empirically computed as the ratio of the average counts in a sensor with and in a sensor without microlenses, using a given primary lens and a well controlled illumination of the scene. To compensate for variations between different chips, the average counts of each sensor are compared using no primary lens and the same diffused light, where microlenses have no effect and cause no concentration. The ratio between the two measured counts is used to correct for any technological differences. Different scenes will result in different average counts, however, the ratio between different sensors should stay the same, as long as no saturation is observed and the majority of pixel counts are above dark count rate (DCR). Experimental measurements showed a good correspondence between samples with and without microlenses. The intensity in the sample with microlenses was approximately 5% lower, which was probably caused by the fact that the diffuse light still had some directivity.

With the above mentioned inter-sample correction factor, measurements were performed on

the two samples with 30 μm and 70 μm height microlenses, and on another sample without microlenses which served as reference. In the experiments a constant white light source illuminated two targets and then measurements were performed at different f-numbers. The f-number is set by adjusting the iris diaphragm of the objective lens. The reflected light by a white target was measured during the higher f-number measurements in order to have enough signal above the noise level. In the low f-number measurements a black target reflected the light coming from the illumination system so the signal would not saturate the sensor. The concentration factors were experimentally measured on the two samples and then compared with the simulations. Figure 5.7 shows the median concentration factor of the pixels in each sensor obtained experimentally and the simulated values for each f-number. This comparison showed good agreement.

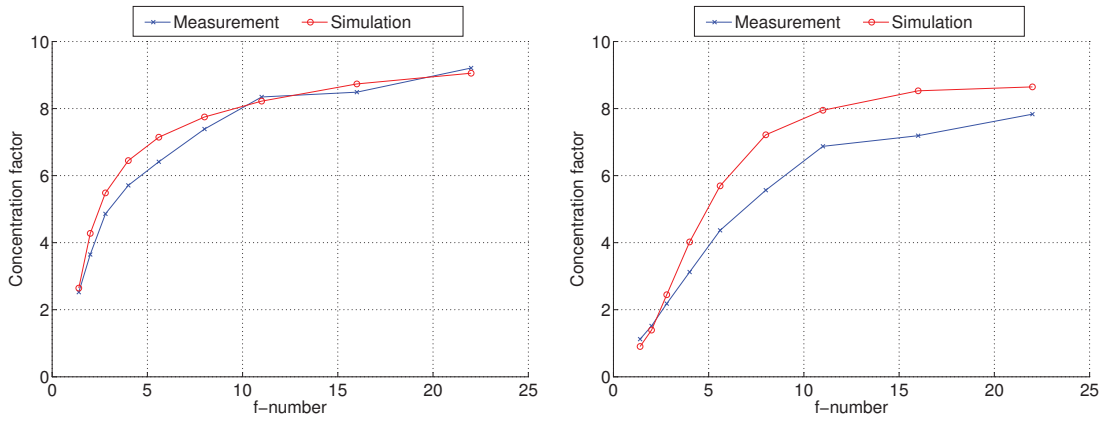

Figure 5.7 – Concentration factor measurements and simulations of a sensor with 30 μm (left) and 70 μm (right) height microlenses. The sample with 70 μm height microlenses presented a higher missmatch between simulations and measurements. This was probably due to some deviation in the fabrication parameters.

Figure 5.8 shows images obtained with two SPAD sensors, one with 30 μm height microlenses and another one without them. All the images were taken using the same USAF resolution target under the same lighting conditions. The acquisition time for each picture was 200 μs, two acquisitions were performed for each sensor, one with the illuminated target and another with the sensor in complete darkness to measure the DCR for each pixel. The DCR value was then subtracted from the picture of the USAF resolution target to enhance the contrast and minimize the effects of the DCR on the concentration factor calculation. Hot pixels, i.e. pixels with unusual high DCR, were discarded and the value at its location was calculated

with a median filter. These pixels account for approximately 1% of the total amount of pixels, therefore this filtering process has no visible effect on the quality of the final image. In the pictures the intensity scale of the images was adapted for each f-number to enhance the visibility of the images. The scale of the concentration factor images has been kept constant to facilitate the comparison between them. Here the improvement due to microlenses is clearly visible, especially at high f-numbers. Moreover, since the main source of noise in SPAD imagers is the DCR, and this is not altered by the microlenses, an increase in intensity due to the microlenses means an increase in SNR. This effect is, however, not visible in these examples, as they had been compensated for DCR. Nevertheless, it is not always possible in every application to make this compensation and therefore microlenses are generally beneficial to increase signal-to-noise ratio and intensity.

The uniformity of the concentration factor is also observable in the images of Fig. 5.8. The concentration factor for each image and its calculated standard deviation are shown in Table 5.1. The standard deviation is relatively high, compared to previous studies [Donati et al., 2011]; it is caused mainly by the telecentric error, i.e. the main beam for each image point is not parallel to the optical axis of the imaging system, and it increases with the f-number. For low f-numbers the size of the focal spot is substantially larger than the pixel's active area, therefore a small focal spot decentering will have almost no effect on the irradiance within the pixel active area. As the f-number increases, the size of the focal spot decreases. When the focal spot reaches a size similar to the pixel active area the centering of the focal spot becomes more critical. In this case a small displacement of the focal spot has a tremendous impact as the irradiance in some parts of the pixel active area drops to almost zero. The microlenses at the center of the image circle, i.e. the area at the image plane that is uniformly exposed by the optics without significant shadings on the margin, receive light rays at an angle that approaches 90 degrees, while periphery microlenses are hit at an angle, as a consequence the focal spot in periphery pixels will be decentered and therefore they will experience reduced concentration factors at high f-numbers. The same effect occurs in case of misalignment between pixels and microlenses. At low f-numbers the focal spot centering is not that critical and therefore the concentration factor will not vary with respect to perfectly aligned microlenses, whereas at high f-numbers a small misalignment between microlenses and detectors can have

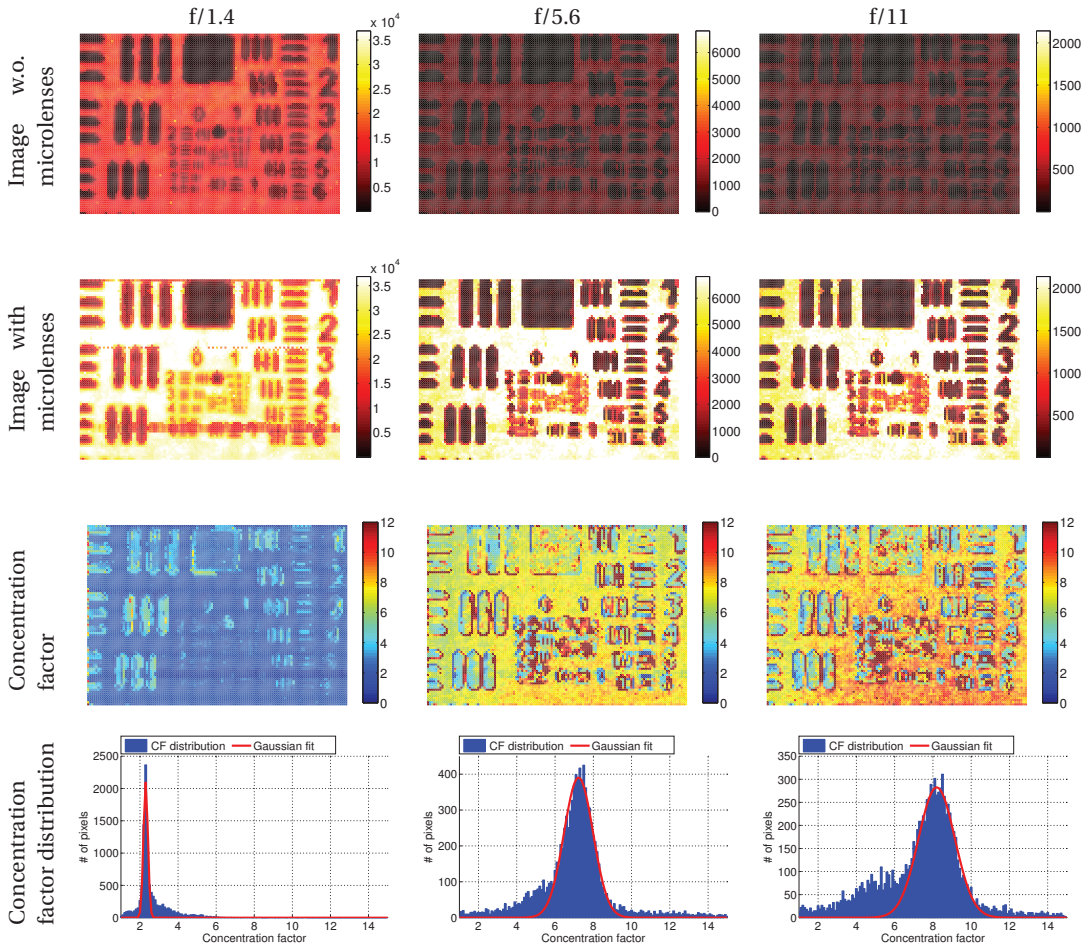

Figure 5.8 – Image crops obtained from SPAD sensors with 30 μm height microlenses and without microlenses at different f-numbers. The concentration factor obtained with the microlenses increases with the f-number, and so does its standard deviation.

a large impact in the concentration factor (Fig. 5.9). This explains the mismatch between the simulations and measurements at high f-numbers as shown in Fig. 5.7 for the 70 μm sample.

	f/1.4	f/2.8	f/5.6	f/11	f/22
Concentration factor	2.32	5.34	7.36	8.08	9.34
Standard deviation	0.18	0.30	0.76	0.95	1.70

Table 5.1 – Concentration factor and its standard deviation for different f-numbers.

The most accurate way to determine the true value of the height of microlenses is to apply a destructive method such as dicing a sample for SEM measurements. Alternatively, we propose

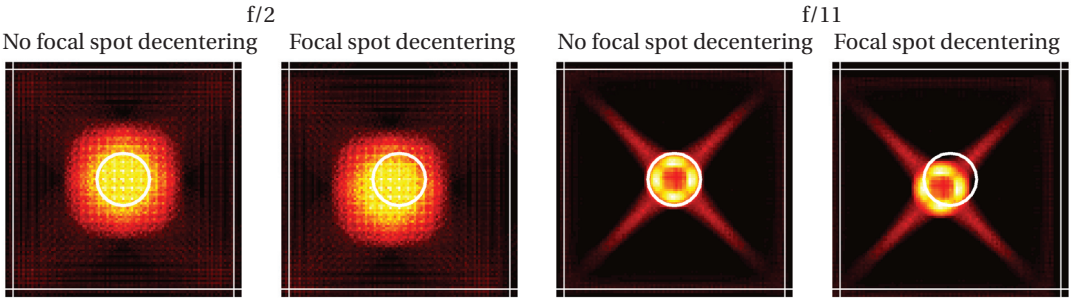

Figure 5.9 – Light intensity profile in the image plane for f/2 and f/11 when the focal spot is perfectly centered and in case of focal spot decentering due to microlenes misalignment or telecentric error. For high f-numbers a small focal spot decentering will dramatically reduce the irradiance in certain parts of the pixel active area (enclosed in a white circle). For f/2 the concentration factor dropped from 4.3 to 4.0 when the misalignment was introduced, whereas for f/11 the concentration factor changed from 8.1 to 6.8. In these simulations microlenses with 30 μm were used and the focal spot was decentered 1 μm in the X and Y directions.

a novel non-destructive method based on projecting angular incident collimated light on the sensor, which reveals the exact height of the microlenses. If the microlenses and the detectors are well aligned, the measured intensity will be maximum for light incident perpendicularly to the chip. As the polar angle of incidence increases, the focus spot moves away from the photosensitive area and therefore the measured intensity decreases. At a certain polar angle the focus spot will enter the next neighboring pixel. This angle depends on the design of the microlenses and the height of the imprinted microlenses. Therefore by measuring at which polar angle the peak in intensity appears, it is possible to calculate the height of the microlenses. Figure 5.10 shows the angle dependency simulations for microlenses of different heights and the experimental measurement performed on one of the sensors. The secondary peak position versus the height of the microlenses is reported in Fig. 5.11.

There is good agreement between simulations and measurements since the microlens array and the sensor were well aligned and no significant gradient in the concentration factor, and therefore in the height of the microlenses, was observed. Measurements at 0° and 90° azimuthal angles were performed to verify that the alignment was correct. In case of a misalignment between the sensor and the microlenses, a complete scanning in polar and azimuthal angles would enable the measurement of the misalignment and/or tilt of the microlenses and to still determine their height. In case of misalignment between the microlenses and the

sensor, the intensity peak would not be at 0° azimuthal and polar angles, but at a different one. By measuring at which polar and azimuthal angles the peak occurs, it is possible to measure the misalignment. If these angles are not constant for all the pixels, that means that there is not only a misalignment but that the microlens array is tilted.

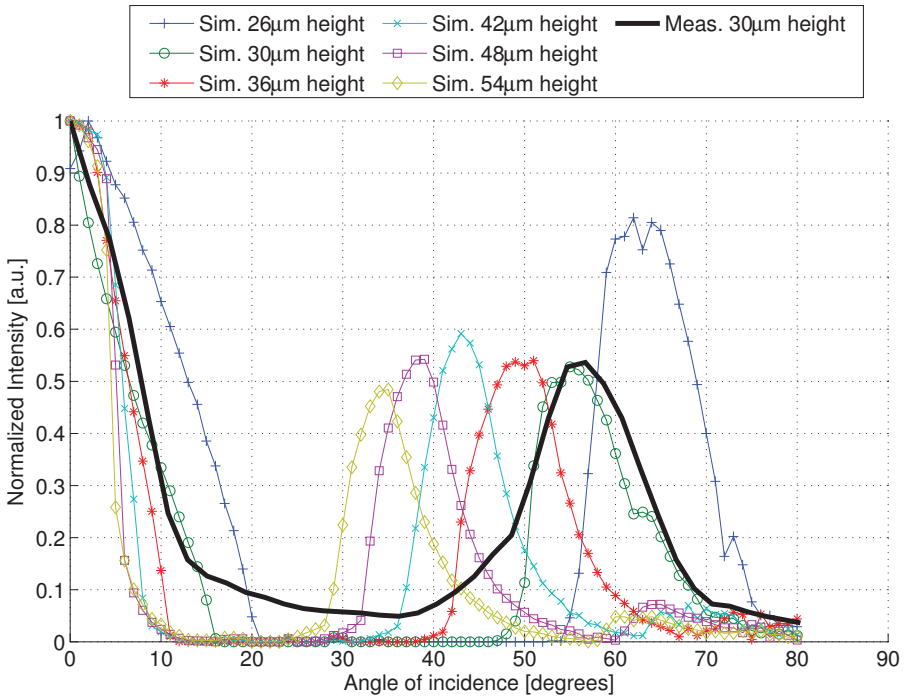

Figure 5.10 – Light intensity dependence versus angle of incidence for different simulated microlenses heights and measurements obtained from a sensor with 30 μm height microlenses.

5.4 Conclusion

We have shown the relation between concentration factor of imprinted microlenses and f-number. We demonstrated that the optimal height of the microlenses depends strongly on the f-number of the whole optical system.

We have developed a reliable simulation software, whose projected values agree well with the experimentally measured values. We also introduced a new contact-less non-destructive method to determine the height of the microlenses, which can be applied in mass production. These results show that microlenses can compensate substantially for very low fill factors such as for SPAD imagers and enable the development of new microlens designs tailored for each

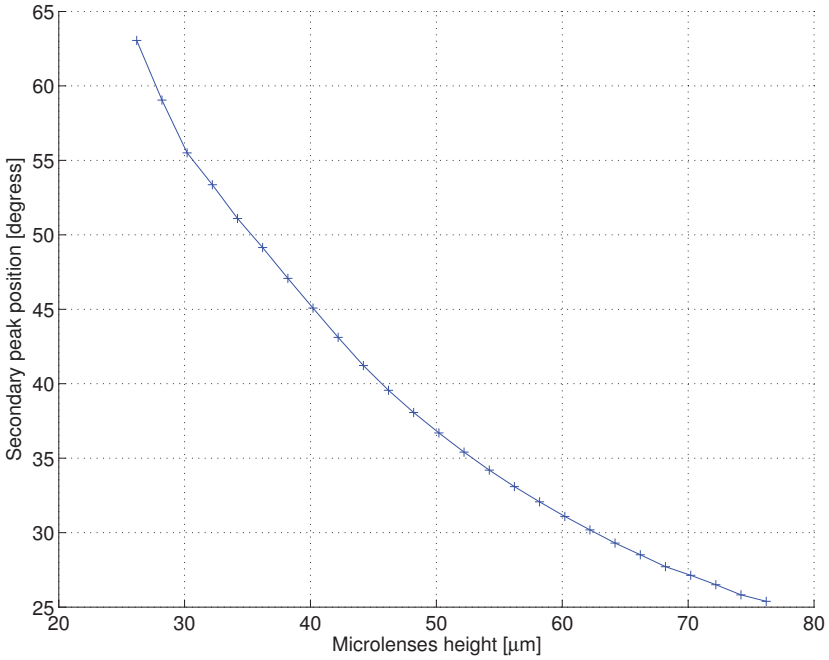

Figure 5.11 – Secondary peak position variation with respect to the microlenses height.

specific application.

6 Single-photon avalanche diode imagers applied to near-infrared optical tomography[1]

In an attempt to recover most of the backscattered light from tissue and to reduce interferences from ambient light, NIROT devices have traditionally employed fiber based sources and detectors. Such optical probes were directly attached to the surface of the tissue. On the one hand this leads to a high number of photons and hence a high SNR, on the other hand it limits the scalability of NIROT devices, making it impossible to construct systems with more than a few dozens of sources and detectors. It is desirable to increase the number of sources and detectors to obtain more information in order to reduce the ill-posed nature of the reconstruction problem as described in chapter 3.

The aim of this chapter is to present a NIROT setup based on the LASP image sensor which integrates 16,384 detectors, each capable of providing independent timing information with picosecond resolution. We also systematically analyze the benefits and limitations of current SPAD image sensors when operating in TCSPC regime with an emphasis in NIROT applications.

[1] The material of this chapter is derived from: **[Mata Pavia et al., 2014b] Mata Pavia, J., Wolf, M., Charbon, E. (2014). Single-Photon Avalanche Diode Imagers Applied to Near-Infrared Imaging, *Selected Topics in Quantum Electronics, IEEE Journal of*, 20(6).**

6.1 Optical setup

NIROT systems can be classified into two types depending on the position of its sources and detectors. In transmission mode, the subject under study is illuminated from one side and the backscattered light is measured at its opposite side. This approach restricts the range of objects that can be studied, because for objects thicker than 6 cm almost no light will reach the detectors. For this reason we decided to implement a NIROT setup in reflection mode where sources and detectors are placed on the same side of the object. This reflection approach enables the study of a wider range of tissues, and is therefore clinically more relevant. However, it limits the depth sensitivity to ~3 cm. The light backscattered close to the sources has a very fast response that cannot accurately be measured by TCSPC in our SPAD image sensor. Therefore a sufficient distance between sources and detectors is necessary. To measure backscattered light at very short source-detector distances requires dedicated discrete SPADs [Pifferi et al., 2008, Dalla Mora et al., 2012].

As described previously in chapter 3, fast reconstruction algorithms published so far [Konecky et al., 2008, Lukic et al., 2009] require placing the light sources inside the field of view of the camera when working in reflection mode, which is impossible as explained before for our SPAD sensor. Furthermore, the fast algorithm presented in chapter 3 requires a relatively high number of source positions to perform a reconstruction. Since it is only possible to measure with a row of pixels at a time with the LASP image sensor, the total measurement time would take several hours with this algorithm. Therefore, a new illumination system was conceived which makes a compromise between the complexity of the reconstruction algorithm and the required number of source positions, as explained in the next section. The solution was to illuminate the object with two fixed parallel lines outside the field of view of the camera.

Fig. 6.1 shows the schematic of the NIROT setup and a picture of its implementation. The light source is a 785 nm, 10 mW laser with FWHM of ~140ps at 3mW operation (Becker&Hickl BHLP-700). The repetition rate of the laser was 80MHz, and it was set to deliver an average output power of 3 mW in order to keep the laser's time point spread function much faster than the measured signals. The lines are generated by a collimator and a line diffuser; a telecentric objective lens with a magnification factor $\beta = 0.07$, projects the backscattered light

at the surface of an intralipid phantom to the SPAD image sensor. For each line source an acquisition is performed. Fig. 6.3 shows the typical response obtained in time resolved NIROT measurements. The pulse shape is captured by TCSPC as a reconstructed time histogram and its shape is determined by the absorption and scattering properties of the medium.

Intralipid phantoms with similar absorption and scattering properties as human tissue were produced for the experiments. These phantoms generally consist of a tank filled with a liquid composed of distilled water, intralipid emulsion, and Indian ink. These components can be mixed in different proportions to adjust the absorption and scattering coefficients. In this liquid we immersed the targets of the reconstructions. They were small objects made of silicon, carbon powder, and TiO_2. Although in principle it is possible to resolve scattering and absorption of these targets, in this work we only focused on absorbing targets, because they are clinically more relevant. The theoretical framework described in the next section only takes into account changes in the absorption coefficient, but could be further developed for absorption and scattering changes. For this reason the targets and the intralipid phantom have the same scattering coefficient and only the absorption coefficient is varied in the targets in our experiments.

6.2 Reconstruction algorithm

6.2.1 Theory

The LASP's operating restrictions made the algorithm described in chapter 3 impractical for our image sensor. Therefore a new algorithm that balanced the computation power with the number of measurements required to obtain a reconstructed image was developed. The starting point of the algorithm is Eq. 2.16 for absorption changes in the presence of line-like illumination. Applying the Fourier transformation defined by

$$\widehat{u}(\mathbf{R}, q_x, \omega) = \int dx\, e^{-iq_x x} u(\mathbf{R}, x, \omega), \tag{6.1}$$

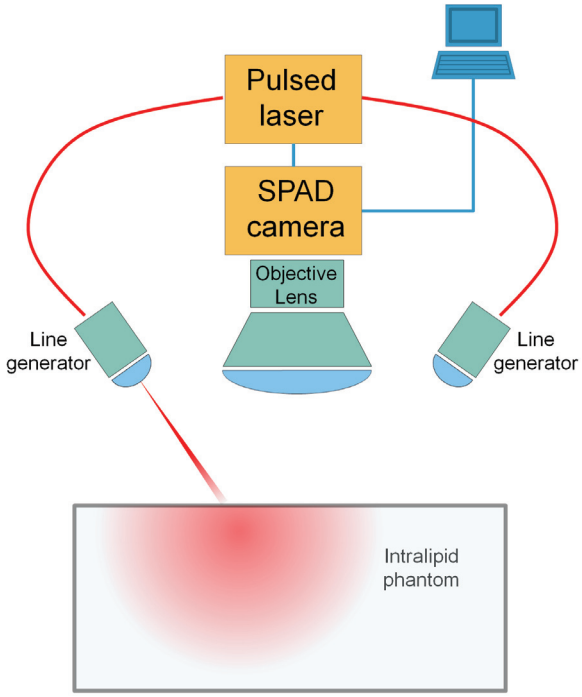

Figure 6.1 – Schematic of the optical setup. A picosecond laser generates pulses of light that are projected as lines on the surface of the object under study. The light propagates through the object (diffusion process) and light emerging from the surface is captured by the objective lens and projected onto the SPAD sensor. Data is digitized in the camera and transferred to a computer.

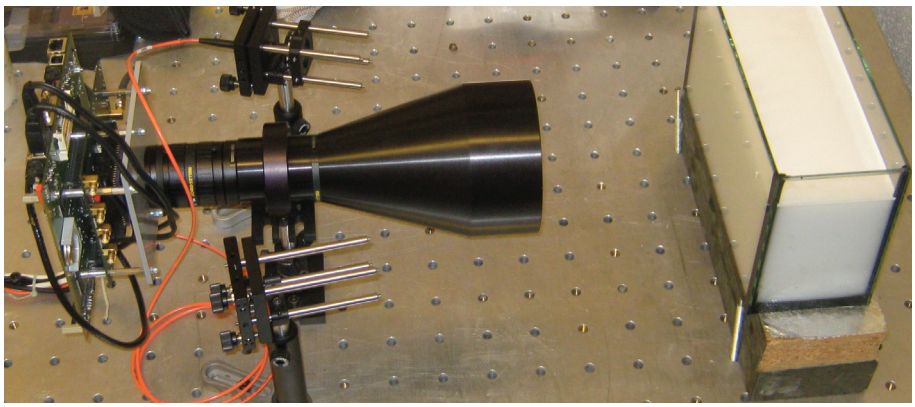

Figure 6.2 – Picture of the NIROT setup in the lab with the tank containing the intralipid phantom.

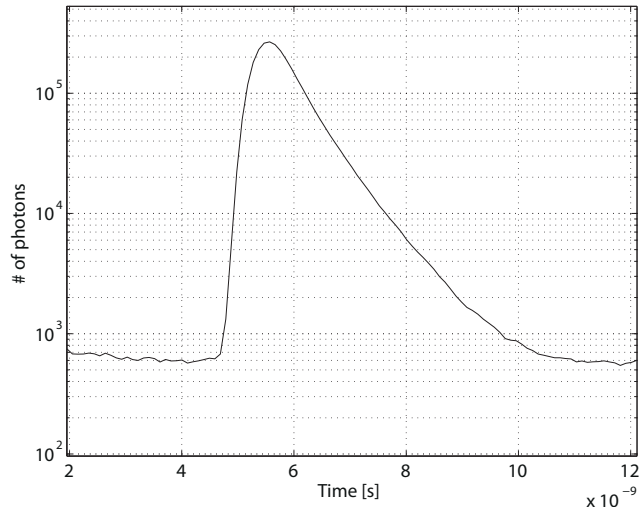

Figure 6.3 – Response measured from an intralipid phantom with $\mu_a = 0.07$ cm^{-1} and $\mu_s = 0.07$ cm^{-1} at a source-detector separation of 2 cm.

where $\mathbf{R} = (y, z)$, Eq. 2.16 results in

$$\widehat{u}_s(\mathbf{R}, q_x, \mathbf{r_s}, \omega) = -\int d^3\mathbf{r}' c\delta\mu_a(\mathbf{r}') u_0(\mathbf{r}' - \mathbf{r_s}, \omega) \int dx e^{-iq_x x} g(\mathbf{r} - \mathbf{r}', \omega). \tag{6.2}$$

Applying the change of variable $x'' = x - x'$, Eq. (6.2) results in

$$\widehat{u}_s(\mathbf{R}, q_x, \mathbf{r_s}, \omega) = -\int d^3\mathbf{r}' c\delta\mu_a(\mathbf{r}') u_0(\mathbf{r}' - \mathbf{r_s}, \omega) e^{-iq_x x'} \int dx'' e^{-iq_x x''} g(\mathbf{R} - \mathbf{R}', x'', \omega). \tag{6.3}$$

The homogeneous field $u_0(\mathbf{r}' - \mathbf{r_s}, \omega)$ produced by an infinite line of light that is parallel to the x-axis, will be x independent as in Eq. 3.7. Thus Eq. 6.3 can be rewritten as

$$\widehat{u}_s(\mathbf{R}, q_x, \mathbf{r_s}, \omega) = -\int d^2\mathbf{R}' c\widehat{\delta\mu_a}(\mathbf{R}', q_x) u_0(\mathbf{R}' - \mathbf{R_s}, \omega) \widehat{g}(\mathbf{R} - \mathbf{R}', q_x, \omega). \tag{6.4}$$

In an infinite medium, $u_0(\mathbf{R}' - \mathbf{R_s}, \omega)$ will be defined as in Eq. 3.7, whereas $\widehat{g}(\mathbf{R} - \mathbf{R}', q_x, \omega)$ can be numerically calculated from Eq. 2.11. An analytical expression for $\widehat{g}(\mathbf{R} - \mathbf{R}', q_x, \omega)$ can also be derived from the solution of the homogeneous Helmholtz equation in two dimensions. The Green's function expressed in Eq. 2.11 is the solution to the homogeneous Helmholtz equation

$$(\nabla^2 + k_0(\omega)^2) g(\mathbf{r}, \omega) = 0. \tag{6.5}$$

Applying the transformation defined by Eq. 6.1 into Eq. 6.5 results in

$$(\nabla_{yz}^2 + k_0(\omega)^2 - q_x^2) D\widehat{g}(\mathbf{R}, q_x, \omega) = 0, \tag{6.6}$$

which is a two dimensional Helmholtz equation whose solution in an infinite medium [Mandelis, 2001] can be written as

$$\widehat{g}(\mathbf{R}, q_x, \omega) = \frac{i}{4D} H_0^{(1)}(\gamma(\omega, q_x)|\mathbf{R}|), \tag{6.7}$$

where $H_0^{(1)}$ is the Hankel function of the first order and

$$\gamma(\omega, q_x)^2 = k_0(\omega)^2 - q_x^2. \tag{6.8}$$

Eq. 6.4 can be independently inverted for each q_x component. After $\widehat{\delta\mu_a}(\mathbf{R}, q_x)$ is calculated for a range of equidistant q_x, it is possible to apply the Fourier inversion operator to calculate $\delta\mu_a(\mathbf{r})$. The advantage of Eq. 6.4 with respect to Eq. 2.16 lies in the reduced computational complexity of the inversion problem. Since in Eq. 6.4 the integral equation is only two dimensional, although many more equations need to be inverted, the computational power required to reconstruct $\widehat{\delta\mu_a}(\mathbf{R}, q_x)$ is substantially lower than that for the three dimensional equation in Eq. 2.16. As an example, we compare the computational complexity necessary to calculate $\delta\mu_a(\mathbf{r})$ when singular value decomposition (SVD) is used to invert the integral equations. In an infinite medium with $N \times N \times N$ voxels, measurements are acquired at $N \times P$ points at Q different modulation frequencies. The computational complexity of a reconstruction algorithm based on Eq. 2.16 is $O((N \times P \times Q)^2 \times N \times N \times N)$, whereas for Eq. 6.4 is $O(((P \times Q)^2 + log(N)) \times N \times N \times N)$. Not only is the complexity considerably reduced, but the algorithm based on Eq. 6.4 requires solving many independent inverse problems and may therefore easily be parallelized, whereas a reconstruction algorithm based on Eq. 2.16 would only have a single inverse problem. The resulting matrices are dense and therefore parallelization in the inversion step itself is very difficult and can only be implemented up to a certain degree.

The Green's function (Eq. 6.7) and homogeneous field (Eq. 3.7) have been formulated for an infinite medium. Their equivalent for a semi-infinite medium is straightforward to calculate

with Eq. 2.12, and for an infinite slab they can be calculated applying the method of images described in [Haskell et al., 1994]. It has been recently suggested that boundary removal techniques [Ripoll and Ntziachristos, 2006] could be applied in fast reconstruction algorithms so they can work with media with arbitrary geometries [Arridge and Schotland, 2009]. As previously mentioned, although in this work the target of the reconstructions were only absorption changes in the medium, it is possible to expand the algorithm to simultaneously reconstruct changes in scattering and absorption [Markel and Schotland, 2004].

6.2.2 Implementation

The implemented reconstruction algorithm was based on Eq. 6.4, for a semi-infinite medium using extrapolated boundary conditions [Haskell et al., 1994]. The main requirement for the algorithm was to work with a reduced number of rows of pixels and source positions to perform measurements in a short period of time. For each source position and pixel row a new acquisition needs to be performed. In the implementation of the algorithm presented in this contribution, only three rows of pixels and two source lines placed at opposite sides of the detector were necessary. With more sources and detectors it would be possible to increase the volume that can be reconstructed and the accuracy of the reconstructions would improve as well. However, since the ultimate goal is to perform measurements on human subjects the total acquisition time is a critical factor that needs to be kept as short as possible.

The data obtained from the detector is time-resolved, however, the presented algorithm only works with FD data. Therefore a fast Fourier transform is required to convert the measured signals in frequency domain data. Since the time resolution of the TDC is 97 ps, the maximum modulation frequency is ~5 GHz. Although lock-in image sensors capable of directly measuring the amplitude and phase of a frequency modulated light signal exist and could be applied to NIROT setups, due to their current limitations we decided that SPADs are, at the time this research was conducted, better suited for our purposes. Lock-in image sensors provide a faster operation as they do not require a photon starving mode as for SPADs working in TCSPC mode. They also offer better SNR when one or only few modulation frequencies are employed. However the maximum modulation frequency they can currently achieve is far from the 5 GHz that we obtain with the SPAD image sensor integrated in our NIROT system.

Moreover, the current implementation of the reconstruction algorithm requires 32 modulation frequencies, and such a high number of frequencies has so far not been achieved with lock-in image sensors.

Due to the ill-posed nature of the problem, the specific algorithm inverting Eq. 6.4 has a large impact on the quality of the reconstructions. Therefore, inversion algorithms require some type of regularization. In our implementation, the subspace preconditioned LSQR algorithm presented in [Jacobsen et al., 2003] was chosen to solve the inversion problem due to its iterative nature and its Tikhonov regularization scheme.

6.3 Results

6.3.1 Simulations

We performed simulations to evaluate the performance of the reconstruction algorithm in the absence of noise. In order to assess the ideal resolution of our setup together with the algorithm, a testbench that mimicked the experimental setup was implemented. Fig. 6.4 shows a schematic of the setup where the sources, detectors, and objects with increased absorption coefficient are visible. In the testbench a semi-infinite medium with μ_a=0.1 cm^{-1} and μ_s'=10 cm^{-1} with three differently shaped objects with optical properties μ_a= 0.3 cm^{-1} and μ_s'=10 cm^{-1} was modeled. This resulted in an absorption contrast ratio of ~3, which is a moderate contrast in NIROT, e.g. the optical contrast between blood vessels and its surrounding tissue is usually two orders of magnitude higher.

Fig. 6.4 also shows the targets and the corresponding reconstruction. As expected the edges of the objects are not well detected, however the position and shape of the reconstructions match the targets with high accuracy. The reconstruction of the long cube-shaped object is particularly remarkable since traditionally NIROT algorithms lack accurate depth resolution. In the cross-sections of Fig. 6.5 the high accuracy of the reconstructed absorption coefficient is displayed. In the pictures it is visible that in certain areas negative absorption values are reconstructed. Although this is physically impossible, it often occurs in NIROT reconstructions when in the inversion step no restrictions are set on the sign of the solution. Reconstructions

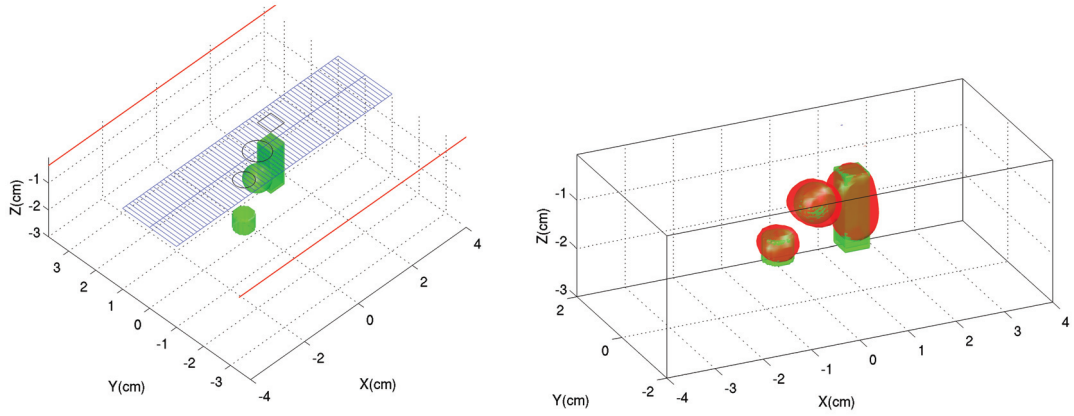

Figure 6.4 – (left) Three dimensional representation of the setup in the simulation testbench. In a semi-infinite medium three objects with higher absorption coefficients are placed at different positions. The XY plane at z=0 is the boundary of the medium. Two infinite lines of light are projected on the boundary plane at y=[-3.5, 3.5] to illuminate the medium. Three lines of 64 pixels located at y=[-1,0,1] measure the backscattered light from the medium. (right) Reconstructed three dimensional images of the objects superimposed on top of the original targets.

with 1.25 mm×1.25 mm×1.25 mm voxels and 32 temporal modulation frequencies required ~20 s of CPU time on a desktop computer.

To assess the depth sensitivity of our algorithm a testbench with a single absorbing sphere placed at different depths was simulated. Fig. 6.6 shows the superimposed results from the three different simulations. These simulations show more accurate results than previous time resolved NIROT algorithms based on the usage of discrete components [Selb et al., 2007].

6.3.2 Experimental results

Experimental measurements were performed with an intralipid phantom with optical properties μ_a=0.07 cm^{-1} and μ_s'=5 cm^{-1} which contained two cylinders of 5 mm diameter separated by 5 mm placed at z=-1 cm with optical properties μ_a=0.25 cm^{-1} and μ_s'=5 cm^{-1}. μ_a and μ_s' have been reduced compared to those employed in the simulation in order to increase the intensity of the measured light, consequently reducing the influence of noise, and to ensure that the measured signals had a much slower response than the system's impulse response function (IRF). If the system's IRF was comparable or slower than the measured signals, then it would be necessary to deconvolve the measured signal with the system's IRF before

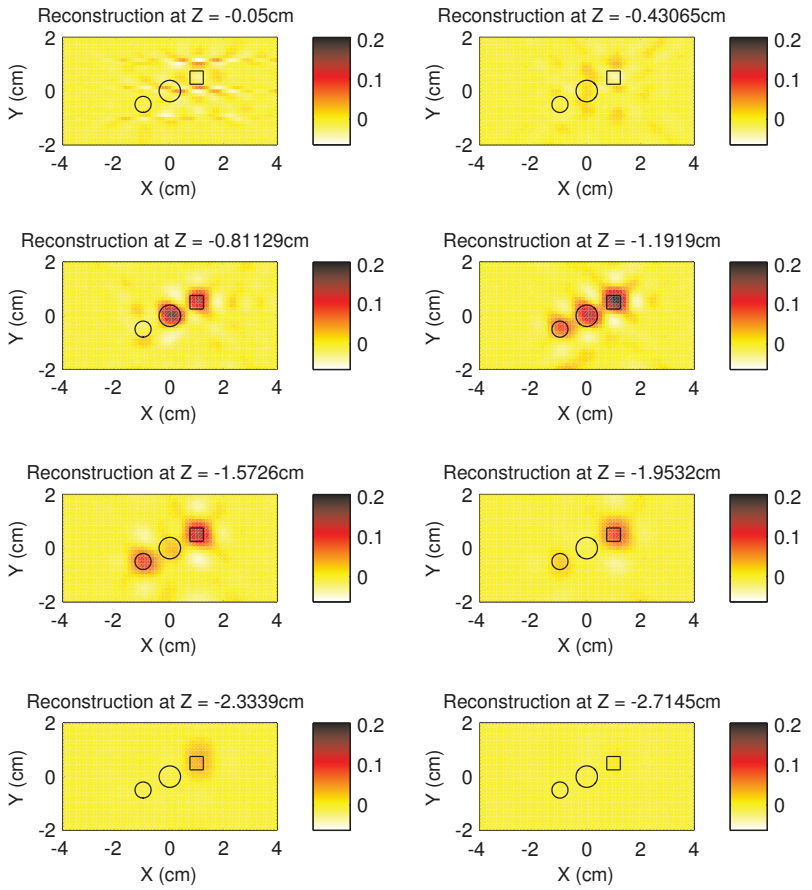

Figure 6.5 – Cross-sections of the reconstruction at different depths. The reconstructed differences in the value of the absorption coefficient to the background absorption are very close to the original ones of $\delta\mu_a$ =0.2 cm^{-1}.

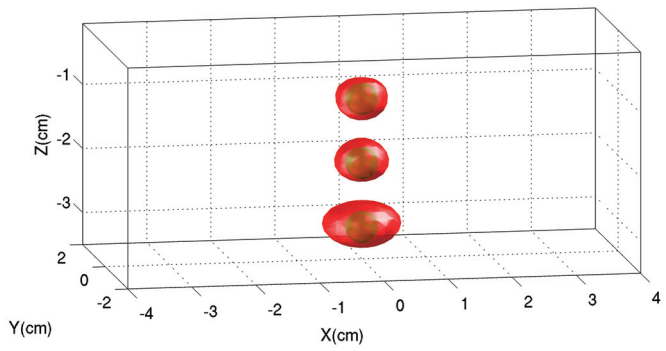

Figure 6.6 – Superimposition of the three reconstructions of a sphere placed at different depths of the phantom.

performing any reconstruction. Trials with higher absorption and scattering showed poor results in terms of SNR indicating that the setup needs to be improved for higher absorbing or scattering media. The positions of the sources and detectors were also modified to increase the SNR: the sources are in this case closer to the detectors to increase the intensity of the measured signal, and the distance between detectors has been reduced so that the difference of signal level among them is smaller. In the current setup the source lights are placed at y=[-2.65, 2.65], whereas the detectors are placed at y=[-0.6,0,0.6]. In this case the 128 pixels in each row were binned into 32 detectors to improve the SNR. The acquisition time for each row of pixels was about 6 minutes. The power of the laser and the source-detector distances were adjusted so that the maximum detection rate in the pixels was never higher than 1% of $1/T_0$, where T_0 is the TDC's conversion period, which in this particular case was 100 ns. This was done in order to avoid pile-up effects in the histograms and to ensure that the pixels were always working in single-photon detection regime. In the reconstruction a voxel of size 0.94 mm×1.44 mm×1.25 mm and 32 temporal modulation frequencies were employed. Fig. 6.7 shows the experimental setup and the reconstructed image. Both cylinders are clearly resolved indicating that a resolution of at least 5 mm is achieved with the current setup and reconstruction algorithm.

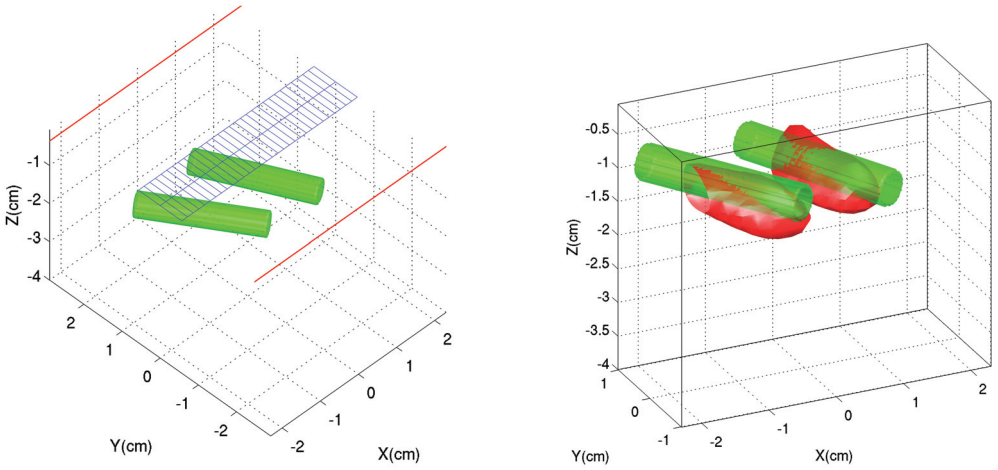

Figure 6.7 – (left) Representation of the intralipid phantom with the two immersed cylinders. The XY plane at z = 0 represents the wall of the tank on which the light is projected and where the backscattered photons are measured. (right) Reconstructed three-dimensional images of the objects superimposed on top of the original targets.

The limited field of view of the setup (only three pixel rows placed at intervals of 6 mm were employed) naturally limited the volume in which the objects were reconstructed. This is the reason why the ends of the cylinders were not visible in the reconstructed image, since they were outside the field of view of the sensor.

Fig. 6.8 shows the cross section of the reconstruction at different depths. In these cross sections both cylinders are clearly differentiated reaching a peak $\delta\mu_a$ of ~0.13 cm^{-1} , which is close to the actual $\delta\mu_a$ of the target cylinders of 0.17 cm^{-1}.

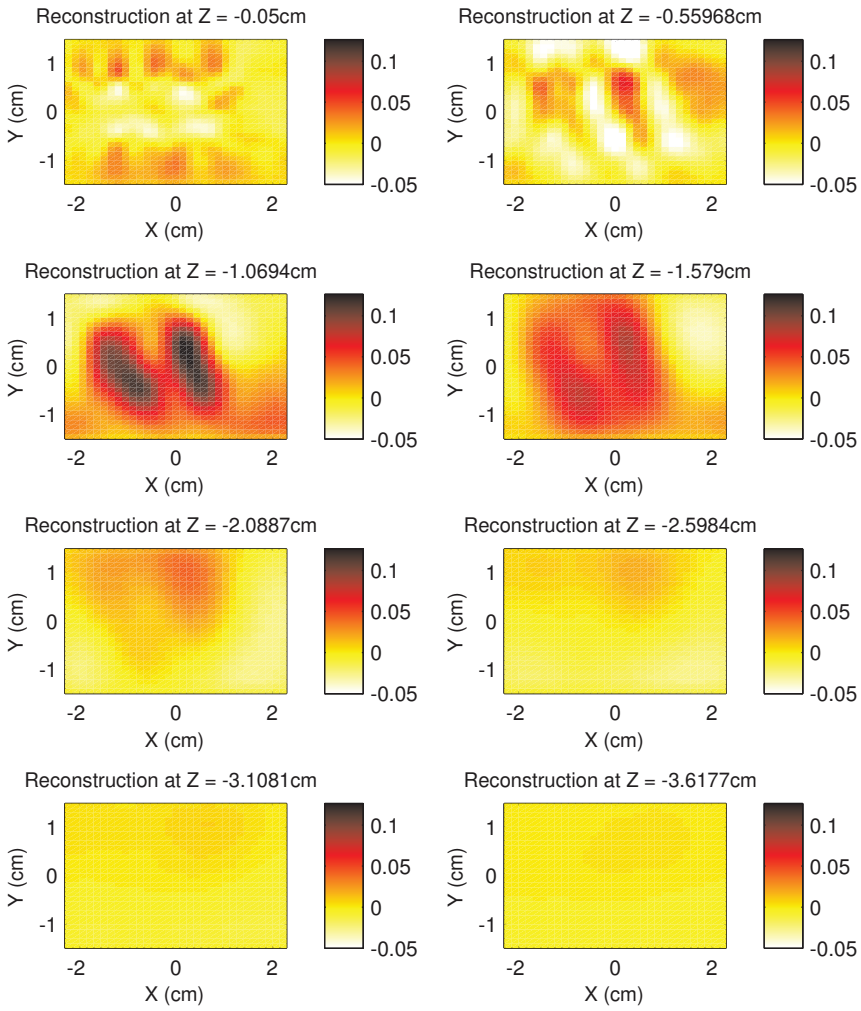

Figure 6.8 – Cross-sections of the reconstruction at different depths. The reconstructed differential absorption values $\delta\mu_a$ are lower than the targets' ones of 0.18 cm^{-1}. However, the integral of the absorption coefficient in the reconstructed objects should be similar to the integral of the absorption coefficient in the target objects.

6.4 Discussion and perspective

The possibility of performing high-resolution time resolved measurements with SPAD image sensors opened a new era in many biomedical and biological imaging fields [Charbon and Donati, 2010]. Their suitability for NIROT has been probed in this chapter, enabling for the first time NIROT devices with high spatial resolution and time resolved measurements.

There are certain limitations in actual SPAD image sensors that could have a large impact in the performance of NIROT devices if they were addressed. The relatively small field of view of the presented system was limited by the size of the sensor's active area (3.2 mm×3.2 mm). Although sensors with larger active areas have been developed [Veerappan et al., 2011], they are still small in comparison to commercial CCD and CMOS sensors. So far most SPAD pixels integrated in image sensors exhibit low fill-factors, making them inefficient in low light applications like NIROT. Although we have shown in a previous chapter how microlenses can improve the fill-factor in SPAD image sensors [Mata Pavia et al., 2014a], their performance strongly depends on the optical system where they are integrated. In particular, the image sensor that was employed in the presented experiments did not have microlenses. Nevertheless there have been recent advances in the development of SPAD pixels [Mandai and Charbon, 2013a], showing that it is possible to reach fill-factors similar to those in commercial CMOS sensors by implementing square SPADs with rounded corners.

At the moment the most critical factor for SPAD image sensors when applied to in-vivo biomedical applications is the measurement time. Most sensors architectures are based on sequential reading of the pixels similar to those used in CMOS sensors. However due to the low activity rate of the pixels in time resolved measurements it is necessary to read every pixel for long periods of time to obtain enough statistics of the photons. Event-driven approaches similar to the ones implemented in digital SiPMs [Mandai and Charbon, 2013a], in which concurrent reading of the pixels is possible will drastically reduce the measurement time with SPAD sensors making them suitable for in-vivo measurements.

6.5 Conclusion

In this chapter it has been demonstrated how NIROT can benefit from high accuracy time-resolved measurements offered by SPAD image sensors. Reconstructed images with millimeter resolution were obtained with a SPAD image sensor prototype. NIROT will benefit in the coming years from the new advances that are undergoing in this still young technology. We believe that high resolution NIROT systems with fast acquisition times can open a new dimension in clinical applications, by enabling the measurement of oxygenation state of organs, e.g. the brain, muscle, liver among others, in real time and with sub-centimeter resolution.

7 SPAD linear array implemented in 3D wafer stacking CMOS process

In chapter 6 it was presented how SPAD image sensors can be applied to NIROT, paving the way for new contactless setups and dedicated reconstruction algorithms. The image sensor that was employed in the setup presented in chapter 6, incorporated 128×128 pixels. However, as only 32 TDCs were integrated on chip, the time necessary to acquire an image, even for a limited number of columns, was too long for in-vivo wide-field NIROT applications. The ideal sensor for NIROT requires a large array of pixels where each SPAD pixel is coupled with and is adjacent to a TDC, so as to guarantee low skew and high parallelism, and a high PDP in the near-infrared range. With the availability of SPADs in deep-submicron CMOS technologies, such imagers have become possible, however with very low fill factor [Gersbach et al., 2012, Bronzi et al., 2014, Richardson et al., 2009b]. A possible solution is resource sharing; an example of this approach was achieved in SPADnet, where 720 SPADs share multiple TDCs [Braga et al., 2014, Mandai and Charbon, 2013a]. However, at over 600 μm and 800 μm respectively, the resulting pixel pitch is very large and unsuitable for NIROT.

In this chapter, we propose a novel backside-illuminated (BSI) SPAD array integrated in a 3D multi-wafer stacking CMOS process that enables superimposition of the pixel detector to the ancillary circuitry. A new resource sharing scheme based on a winner-take-all (WTA) circuit is introduced to enable the integration of a large number of pixels, along with a new TDC architecture for low power operation. The SPAD structures, which employ the bulk substrate as a shared anode, are implemented in a separate substrate from the CMOS circuitry.

7.1 CMOS 3D wafer stacking process

The image sensor was implemented in the 3D two-tier Tezzaron's FaStack® process, which is based on a CMOS 130 nm technology. In this technology two standard CMOS wafers are bonded face-to-face utilizing copper to copper bonds called supercontacts, which are hexagonal metal features. The top tier implements through-silicon-vias (TSVs), i.e. 6μm deep pillars made of tungsten. After wafer-to-wafer bonding, the substrate of the top tier is thinned to 4.2 μm, leaving the TSVs exposed; metalization is added to create the bond pads. The supercontacts have a 4 μm pitch and TSVs are approximately 1.2 μm in diameter. Fig. 7.1 shows a representation of the cross-section of this technology.

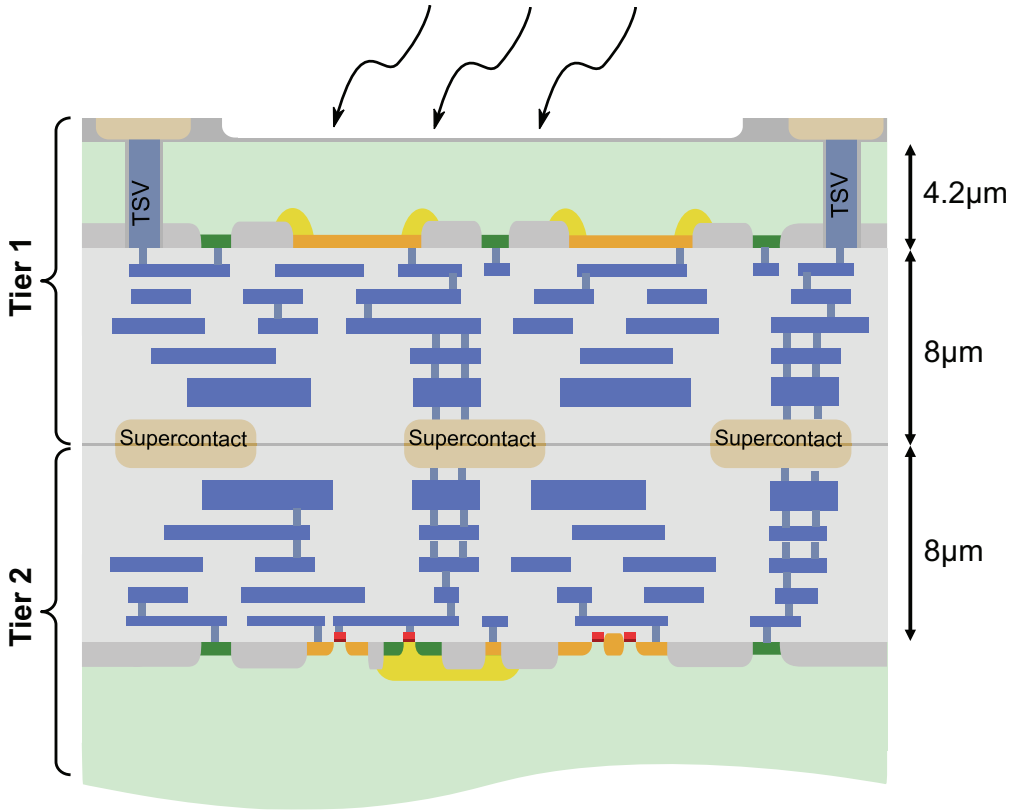

Figure 7.1 – Cross-section of the Tezzaron 3D wafer stacking process. The wafers are bonded face to face by the top metalization through supercontacts. The tier 1 substrate wafer is thinned to 4.2 μm. Through-silicon vias (TSV) allow the interconnection between the the metal layers of tier 1, and the bond pads at its back side. The SPAD pixels are placed in tier 1 with the minimum number of transistors necessary for operation to maximize the fill-factor. The rest of the circuitry is placed in tier 2.

This process allows the integration of SPADs in the top tier, together with the minimum amount

of circuitry for its operation, while the rest of the circuits based on standard CMOS are placed on the bottom tier. This architecture offers a number of benefits with respect to traditional designs with only one substrate. As most of the circuits will be located in the bottom tier, SPAD arrays with very high fill-factor can be implemented in the top tier. Moreover the fact that the pixels will be BSI will increase the fill factor since there are no metalization layers in the substrate that could block or reflect light. BSI SPADs are especially interesting in applications that make use of near-infrared light, because their multiplication region can be engineered to be a few μm deep, and a deeper multiplication region leads to a higher sensitivity of the SPADs in the NIR range, as shown in Chapter 4.

7.2 Image sensor architecture

Fig. 7.2 shows a block diagram of the image sensor. It consists of an array of 4×400 SPAD pixels implemented in tier 1 connected to an array of 2×100 processing blocks located in tier 2. Each processing block includes a WTA circuit, a TDC, and a local 2-port static random access memory (SRAM) block. The SPADs are clustered in groups of 2×4 pixels which are connected to the same processing block. Upon detection of a photon, the TDC in the corresponding processing block calculates the timestamp associated with the photon's arrival time. This timestamp is locally stored in the processing block until it is read out. All memory positions in the processing blocks are accessible through the read out multiplexer.

In Fig. 7.3 the schematic of a processing block is shown. Each block integrates a full TDC along with a redistribution mechanism based on a winner-take-all (WTA) architecture with a collision-free event-driven backend. This circuitry is used to connect the outputs of 8 SPADs to one TDC while preserving the identity of the triggering SPAD. This sensor was built to target TCSPC applications in which the SPADs work in photon-starved mode. In these applications the pixels have a maximum detection rate of about 1%, therefore sharing one TDC among 8 pixels will have a very low impact on the overall performance of the image sensor. In this particular case the probability that more than one pixel detects a photon in the same clock period is 0.3%.

Upon a TDC conversion, the resulting timestamp is stored in a SRAM location. The location in

the SRAM memory block will be determined by the address coded in the WTA circuit. When two or more pixels fire simultaneously, the generated code in the WTA circuit will be a non-valid address, and thus the result will not be stored in the SRAM memory. Each pixel has a dedicated position in the SRAM memory that is flagged as unread after a new conversion is stored on it. After the contents of a SRAM location are read out, its unread flag is automatically deasserted. Each memory location in the SRAM memory can be accessed at the same time for read and write operations.

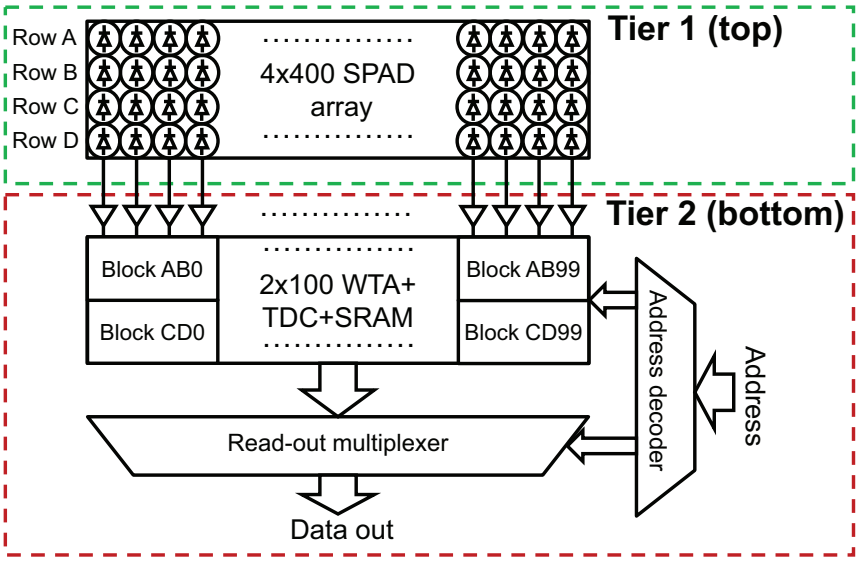

Figure 7.2 – Block diagram of the 3DAPS sensor. Every four pixels in rows A and B (and in C and D) form a cluster that is connected to a WTA+TDC+SRAM memory located in tier 2

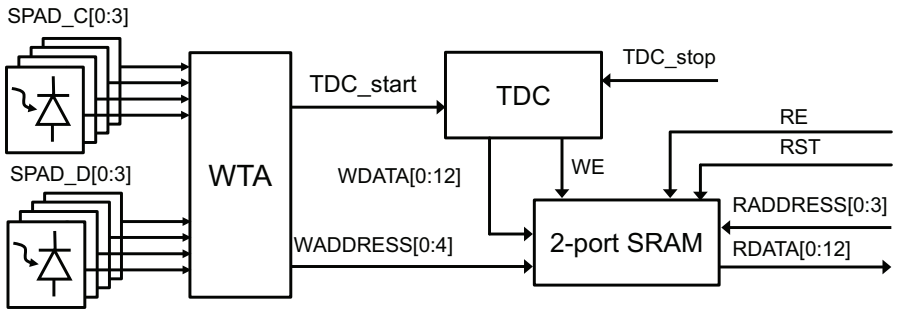

Figure 7.3 – Each cluster has 8 SPAD pixels from two different rows (A and B, or C and D) connected to the winner-take-all (WTA) circuit which triggers the conversion in the TDC and writes its result in the local SRAM memory.

7.3 Pixel design

Four different SPAD structures were implemented in the array, these are shown in Fig. 7.4. SPAD A consists of a circular n+/p-well junction, where the high-field multiplication region is generated. This is surrounded by a n-well/p-well junction, which has a higher breakdown voltage, acting as the guard ring that prevents premature edge breakdown of the SPAD. This structure has been reported in a similar technology [Mandai et al., 2012] yielding PDP values above 30% for 700 nm. SPAD B is similar to SPAD A but it includes a polysilicon ring on top of the guard ring to prevent direct contact between shallow trench isolation (STI) and multiplication region. This approach has been applied to previous SPAD structures [Niclass et al., 2007], showing better DCR results since the contact between the STI and multiplication region is a source of deep-level traps that increase DCR and afterpulsing. In SPAD C the same structure as in SPAD A is adopted, however in this case the multiplication region is formed by a NLDD (N low doped drain) layer and the p-well. The NLDD layer is more lightly doped than N+ and consequently it contains less carrier traps, thus improving DCR and afterpulsing. In SPAD D the multiplication region is formed by the junction between deep n-well (DNW) and p-substrate. The guard ring is formed by the inherent retrograde doping profile of the DNW while blocking the implantation of p-well around the multiplication region. Similar structures have been implemented in the past [Webster et al., 2012a].

The quenching circuitry is implemented in tier 1 with a PMOS transistor working in linear mode. The transistor has to be biased such that the equivalent ohmic resistance is ~100 kΩ to promptly quench the avalanche upon photon detection [Ghioni et al., 2007]. Along with the quenching mechanism, a follower is used to drive the metal connecting the two tiers. Although it would be possible to directly connect the output of the SPAD to the buffer in the bottom tier, this would increase the parasitic capacitance at the cathode and therefore the current flowing through the diode during an avalanche event. This would negatively impact on afterpulsing, timing resolution and power consumption [Niclass, 2008]. For SPADs A, B and C, both PMOS transistors are placed in the top tier isolated from the substrate with a DNW. For SPAD D the quenching and follower transistors are located in the bottom tier since the DNW's breakdown voltage would be the same as for the SPAD D.

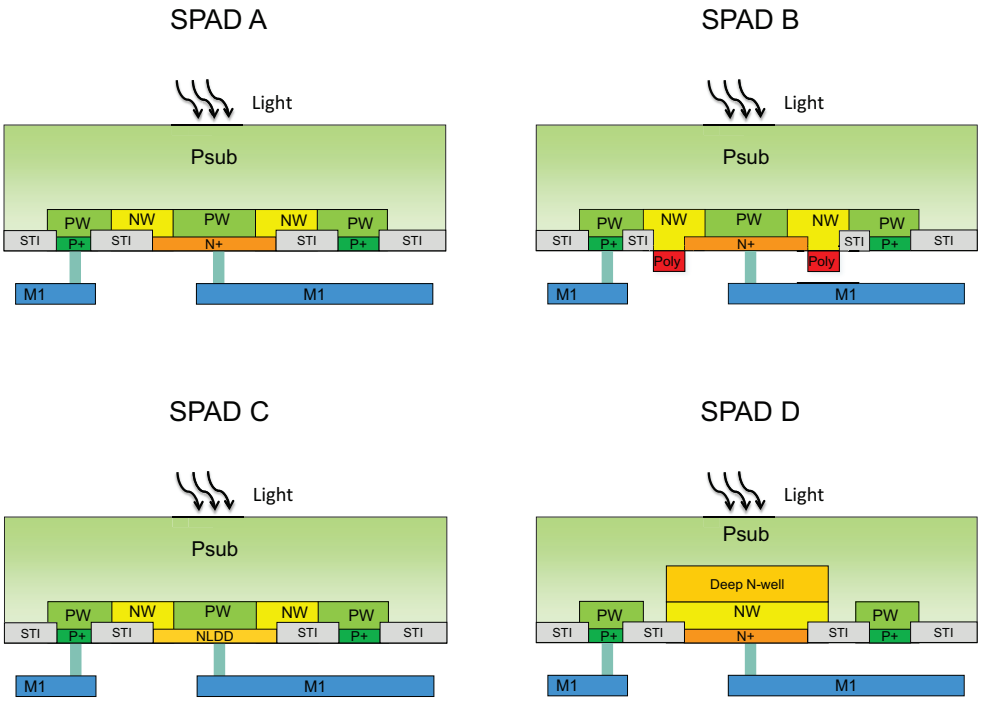

Figure 7.4 – SPAD structures in the array

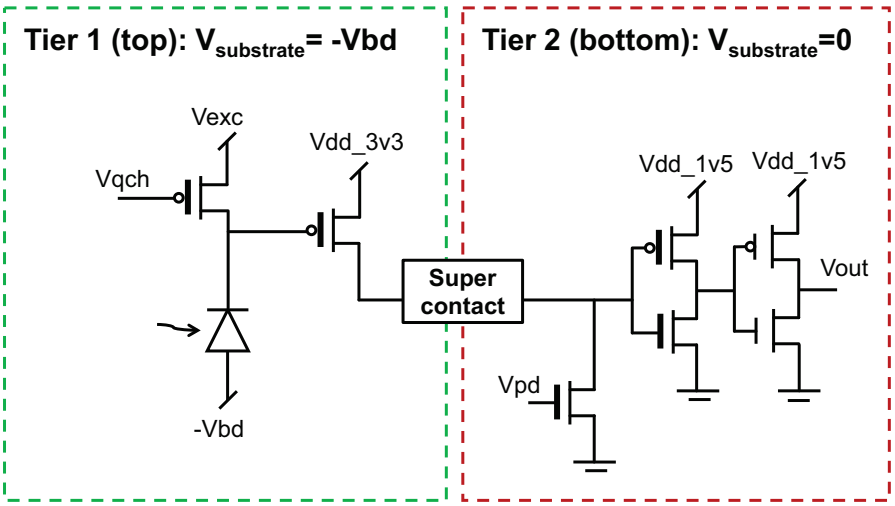

Figure 7.5 – SPAD pixel circuit including quenching transistor and buffer. The number of transistors in tier 1 has been kept low to maximize the space available for SPADs. A follower drives the buffer in tier 2 to minimize the parasitic effects of the supercontact (the contact between the tiers)

7.4 Winner-take-all circuit

A WTA circuit similar to the one introduced in [Niclass et al., 2006] was implemented on each cluster to share the utilization of the TDC among 8 pixels and to make a more efficient use of silicon area. In the WTA circuit the output of each pixel is connected to a digital bus with 5 address lines through pull-down transistors. A dedicated extra line *TDC_start* is used to propagate the digital pulses from the pixels to the TDC, where a time-to-digital conversion will be started. Each pixel generates a unique code in the bus that is latched at the end of the bus by the *TDC_start* signal. Fig. 7.6 shows a schematic of the WTA circuit.

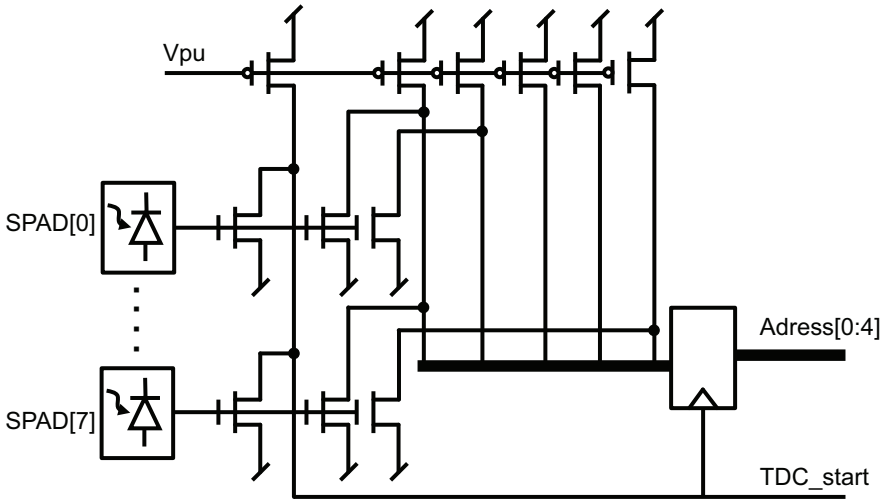

Figure 7.6 – Schematic of the winner-take-all circuit. The collisions in the bus, in the event of two or more pixels firing simultaneously, generate a non-valid code.

In the original WTA implementation presented in [Niclass et al., 2006], in the event of two pixels firing simultaneously the address provided when using a binary code produced an incorrect address. E.g. if pixels 1 and 2, with digital codes ‘001’ and ‘010‘ respectively, receive a photon at the same time, the generated code in the bus would be ’011’ which corresponds to pixel 3, thus providing invalid data for pixel 3.

In order to detect collisions between two or more pixels, we present a new address coding that identifies these collisions in the address bus. With these new collision detection coding, invalid addresses are generated in the event of two or more pixels firing simultaneously. The collision detection code is based on the fact that each code has three ones and two zeros. Table 7.1 presents the codes for each pixel in the present WTA implementation. In a pulled-up

bus when two pixels fire simultaneously, the number of zeros will be more than three and the code indicates an invalid event. In the event of an invalid code in the address bus, the corresponding TDC result will not be stored in the SRAM memory and consequently the data will not be corrupted. The probability of two or more photons being detected in the same clock period can be calculated with a Poisson cumulative probability function. In the proposed WTA implementation where 8 pixels are connected to the same bus, for a pixel detection rate of 1%, the probability that two photons are detected during the same clock period is 0.3% which accounts for 4% of the total number of events in the bus. An increase in the pixel detection rate will result in a higher rate of collisions in the bus, i.e. for a pixel detection rate of 10% the amount of collisions in the bus will account for 35% of the total number of events in the bus.

SPAD	Binary code	Collision detection code
0	000	00111
1	001	11100
2	010	11010
3	011	11001
4	100	10110
5	101	10101
6	110	10011
7	111	01110

Table 7.1 – Collision detection code table for 8 pixels.

This collision detection coding can be expanded to a bus with any number of address lines providing each address has equal numbers of ones and zeros. Consequently, the number of codes per number of address lines is less than when using standard binary coding. The maximum number of codes in an address bus with n lines is given by the number of k elements combinations of a set of n elements, where k is the number of zeros in each address code. The total number of possible codes for a bus with n lines is given by Eq. 7.1, where k has to be the integer closest to $n/2$ in order to obtain the highest number of collision detection codes.

$$\#codes = \frac{n!}{k!(n-k)!}. \tag{7.1}$$

Table 7.2 shows the maximum number of collision detection codes that can be obtained for different bus sizes. The overhead of the collision detection code is at maximum 2 address lines for the presented cases.

Bus width (in # bits)	# Collision detection codes	# Bits required for binary coding	# Bits overhead
2	2	1	1
3	3	2	1
4	6	3	1
5	10	4	1
6	20	5	1
7	35	6	1
8	70	7	1
9	126	7	2
10	252	8	2
11	462	9	2
12	924	10	2

Table 7.2 – Number of collision detection codes for different bus sizes.

7.5 Time-to-digital converter

Previous TDC implementations based on ring oscillators (RO) required operation at frequencies of a few gigahertz during the conversion period, drastically increasing the total power consumption of the chip when many of them were active [Richardson et al., 2009b]. We propose a new TDC based on a dual-speed ring oscillator that only requires gigahertz operation during a fraction of the conversion time. Upon the start of a conversion, the RO starts operating at a frequency of 240 MHz, clocking the low speed counter whose result will determine the 5 most significant bits of the measurement. At the next rising edge of the RO generated clock, after the assertion of the reference clock signal, the RO's operating frequency is boosted to

2.52 GHz. At the same time the low speed counter is deactivated and the high-speed counter is enabled. A delayed clock reference signal eventually stops the RO. The delay between the two reference clocks has to be larger than a period of the RO clock in low speed operation, otherwise the RO might not start oscillating in high-speed mode. The result of the high-speed counter together with the phase of the RO provides the other 7 bits of the measurement. This mechanism effectively measures the time difference between the start signal and the reference clock with a coarse resolution, and subsequently the time difference between the RO low speed frequency clock and the delayed clock signal with high accuracy. Therefore, in the worst case, the TDC only operates at 2.52 GHz between the rising edge of the reference clock and the rising edge of the delayed clock. Fig. 7.7 shows the block diagram of the TDC, and Fig. 7.8 the timing diagram associated with a conversion.

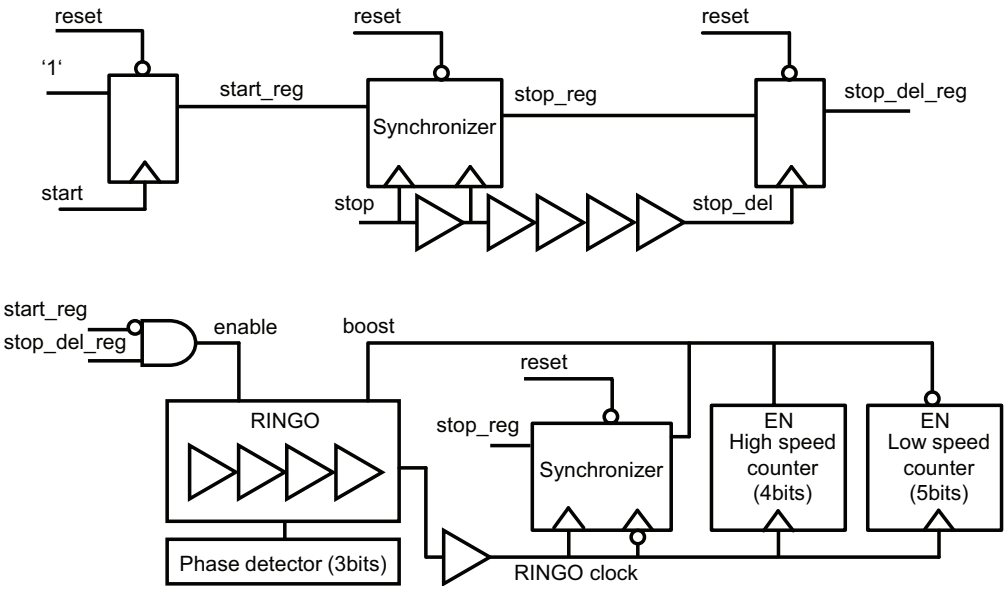

Figure 7.7 – TDC general schematic showing the main building blocks. The TDC is based on a dual-speed ring oscillator that minimizes the power consumed during a conversion.

The internal structure of the dual-speed RO is shown in Fig. 7.9. A differential inverter is employed as basic delay cell in order to equalize the delay in the transitions from '0' to '1' and from '1' to '0'. The differential stages were designed to have wide voltage swing to reduce the effects of jitter induced by power supply noise. The oscillation frequency is controlled through the tail transistors which generate the bias current, and the PMOS transistors controlled by the $\overline{boost}$ signal, which injects current in the differential pair.

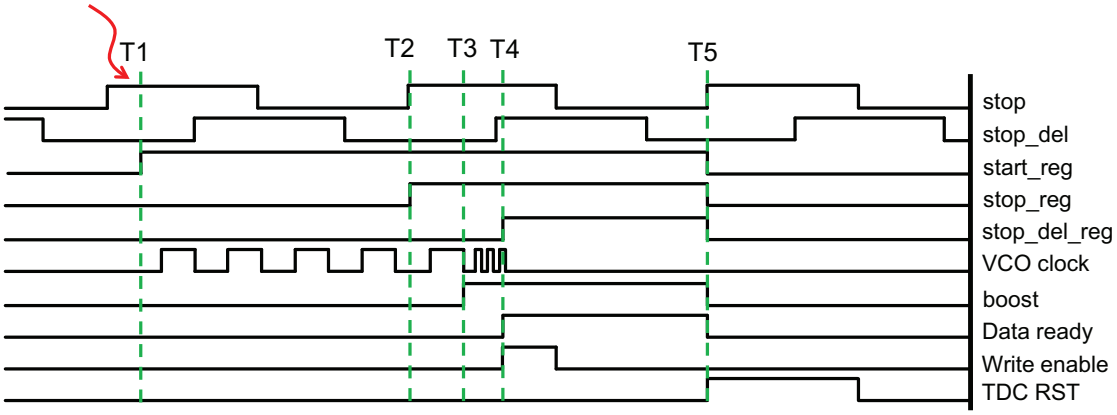

Figure 7.8 – TDC timing diagram. Upon a photon detection, the ring oscillator starts oscillating at low frequency (T1). At the next rising edge of the *stop* signal, the *stop_reg* signal is asserted, where a synchronizer is used to avoid metastability (T2). The *stop_reg* signal is re-synchronized with the ring oscillator clock to generate the *boost* signal, which activates its high-frequency operation mode (T3). The *stop_del* signal stops the ring oscillator (T4). After writing the data in the local SRAM memory, the TDC is reset and is ready to perform the next conversion (T5).

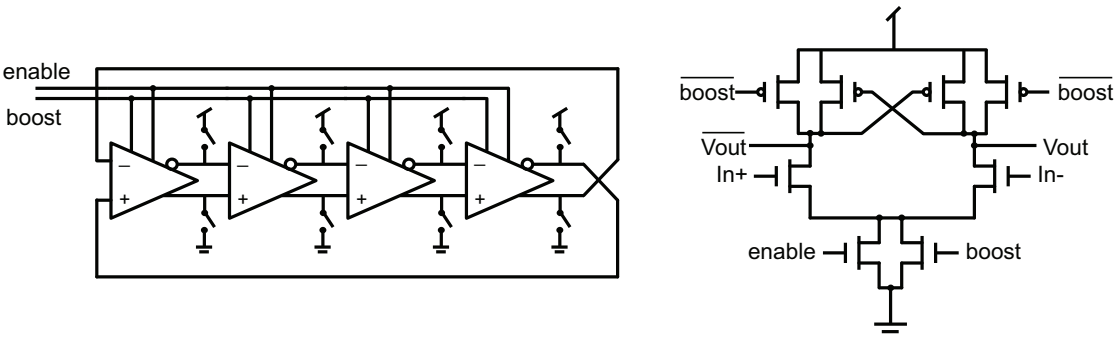

Figure 7.9 – The ring oscillator is based on a differential buffer whose propagation delay can be adjusted with the *boost* signal. Ring oscillator diagram (left) and schematic of a single delay stage (right).

In the TDC block diagram presented in Fig. 7.7 two synchronizers are employed to avoid metastability when signals travel between different clock domains. In the two cases only the transition from '0' to '1' needs to be synchronized, simplifying the design of the synchronizer. Fig. 7.10 shows the schematic of the synchronizer that is employed to control the boosting of the RO and the latching of the *stop_reg* signal.

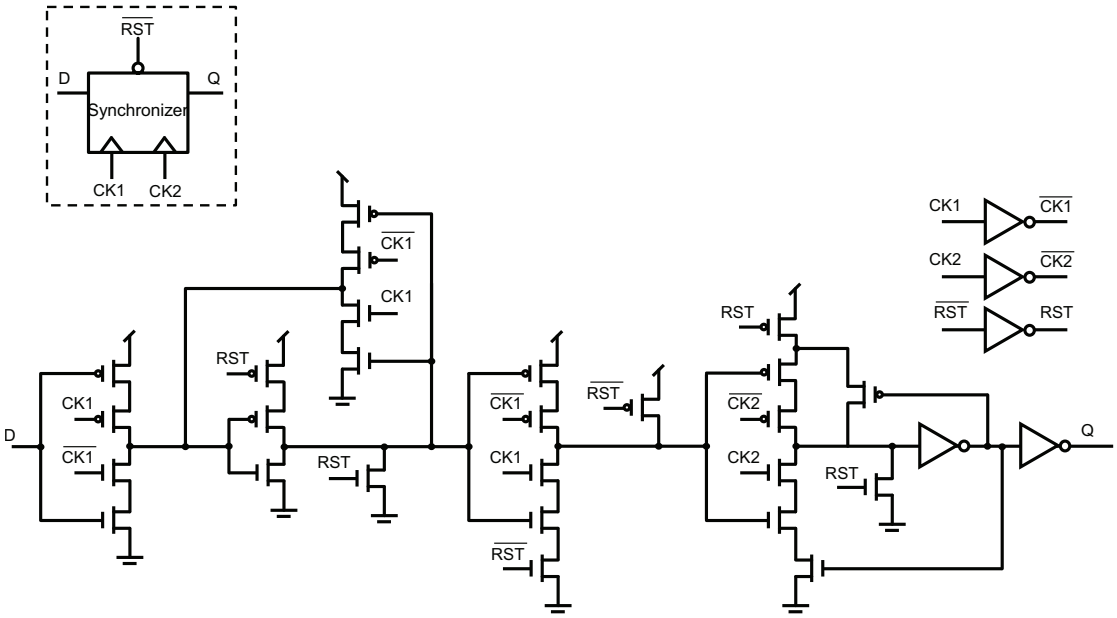

Figure 7.10 – Schematic of the synchronizer circuit implemented in the TDC. The synchronizer is made out of three buffer stages and its purpose is to deliver a '1' at its output synchronized with *CK2* when *D* becomes '1'. Once this happens, *Q* is latched to '1' and it is necessary to reset it to be operational again. Only the first stage includes a latch to minimize the area and to regenerate the integrity of the signal in the internal nodes in case of metastability.

7.6 2-port static RAM memory block

A 2-port static SRAM memory is implemented to locally store the TDC conversions. The SRAM memory consists of 8 words, one for each pixel, of 13 bits size: 12 bits to store the result of the TDC conversion, and an extra bit to indicate whether the result has been already read out or is new. Fig. 7.11 shows a block diagram of the 2-port SRAM memory. The write decoder is controlled by the WTA circuit, whereas the read decoder can be directly controlled from external pads. The two-port architecture of the SRAM memory allows simultaneous asynchronous reading and writing. This is of great benefit in cases where the reference clock is much slower than the maximum read out frequency and it eases the integration of the image sensor in test setups as clock synchronicity is not required during read out operations. Fig 7.12 shows details of the SRAM read and write buffers together with the memory cell integrated in the SRAM array.

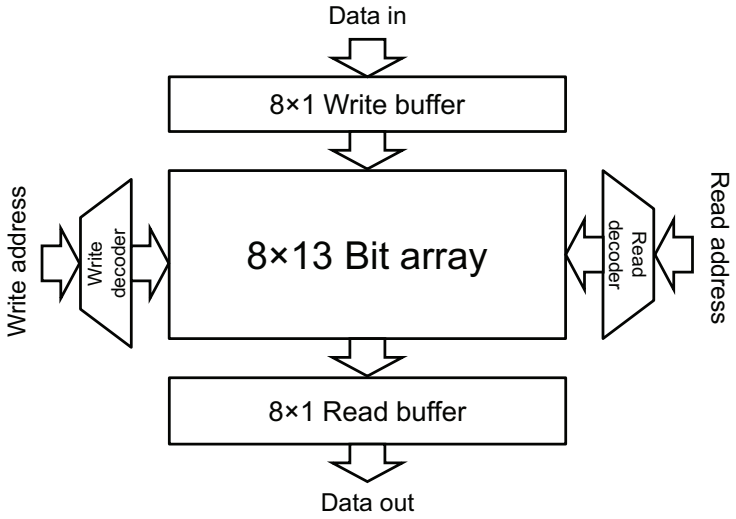

Figure 7.11 – The 2-port SRAM memory has 8 different addresses and a word length of 13 bits. Read and write operations can be performed simultaneously.

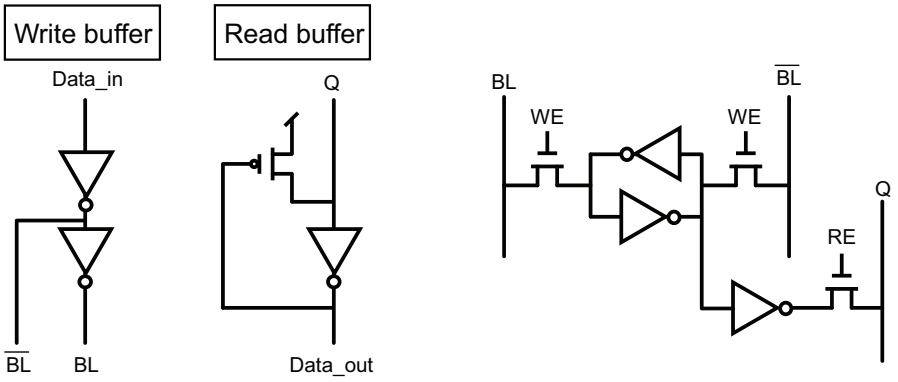

Figure 7.12 – SRAM Read and write buffers (left), and memory cell(right).

7.7 Sensor implementation

The image sensor was fabricated in the 3D two-tier Tezzaron's FaStack® process described in section 7.1, alongside with a number of different SPAD test structures in a SPAD farm. These new SPAD pixels were implemented to explore the performance of previously known and new structures in this technology that can be incorporated in future SPAD array designs. Fig. 7.13 shows a microphotograph of the chip. The SPAD test structures occupy most of the area since each pixel requires at least two pad landings for optical and electrical characterization.

Fig. 7.14 shows a layout view of a pixel cluster in the top and bottom tiers. The total area was only 64×47 μm^2, most of which is occupied by the 2-port SRAM memory. The cluster

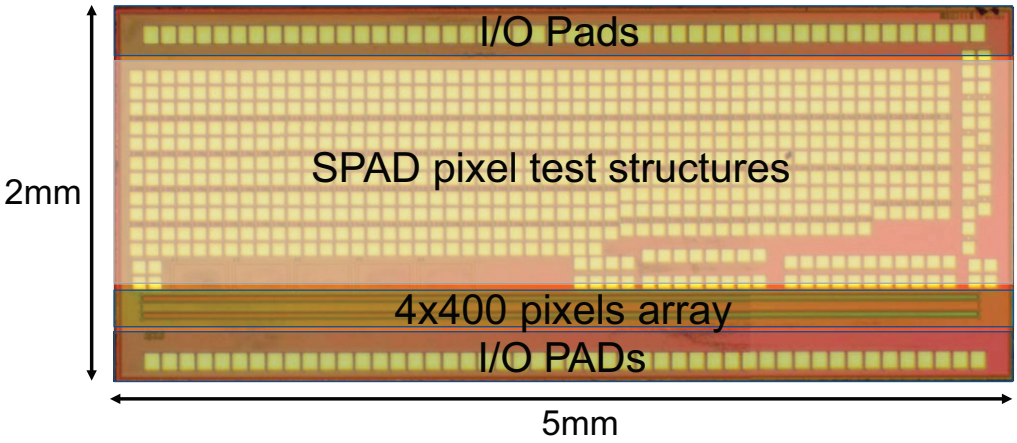

Figure 7.13 – Photomicrograph of the image sensor and the SPAD test structures. The code-name of the chip was 3DAPS.

was equipped with different design-for-test capabilities in order to be able to test each block separately. For this reason a read out multiplexer was added to have direct access to the outputs of the main building blocks, WTA, TDC and SRAM.

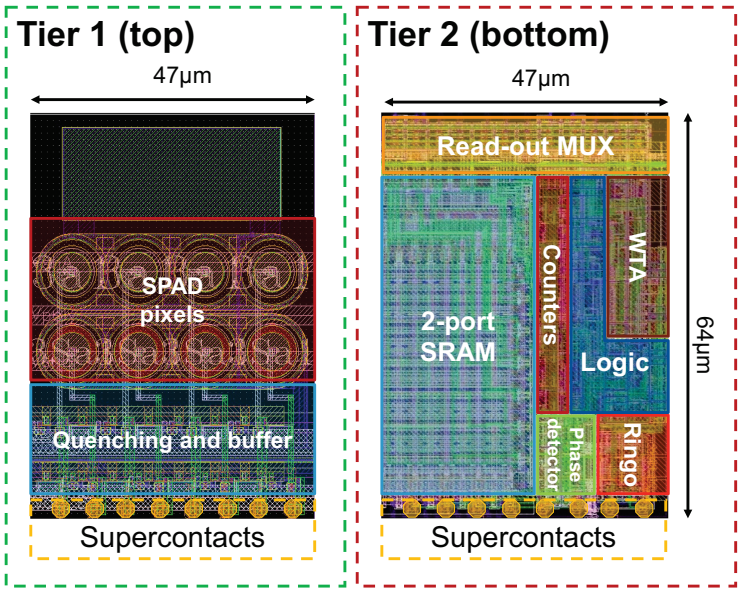

Figure 7.14 – Layout view of a pixel cluster in both tiers. The two tiers are connected through the supercontacts. The pixel pitch is 11.75 μm.

7.8 Experimental results

The image sensor requires independent power supplies and bias voltages for each of the four SPAD structures in the array. Moreover, the RO, delay lines, and the rest of the digital logic, had dedicated power domains to provide more flexibility during the test of the sensor. A total of 13 different power supply voltages and 8 bias voltages are required to operate the image sensor. In order to reduce the amount of external equipment during the test of the chip, a daughterboard was designed that incorporated a number of programmable low drop out (LDO) regulators and digital-to-analog converters (DACs). The daughterboard is connected to a Virtex® -6 FPGA ML605 evaluation kit motherboard, which programs the different voltages in the daughterboard and controls communication with the sensor. Fig. 7.15 shows a picture of the daughterboard with a 3DAPS chip connected to the Virtex-6 evaluation board. The motherboard communicates with a computer via a USB link.

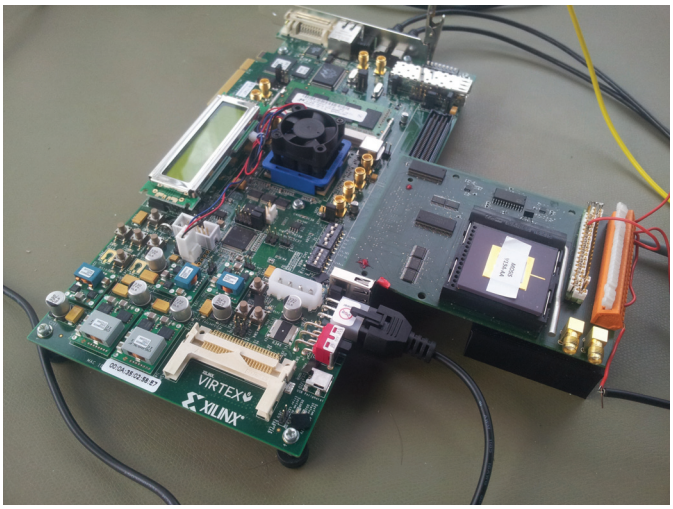

Figure 7.15 – 3DAPS sensor test setup. The packaged sensor is inserted in a zero insertion force socket in the daughterboard which is connected to a Virtex® -6 evaluation board.

The TDC was characterized individually to avoid the interference of other circuits in the results. A 40 MHz signal was generated in the FPGA and then fed to the image sensor as reference clock. The Virtex-6 FPGA integrates a programmable delay line with 18 ps taps which is employed to generate the *start* signal that triggers the conversion in the TDC. Fig. 7.16 shows the conversion results for different delay values in the *start* signal. Since the low-speed period of the RO is not a multiple of 2 of the high-speed period, and the minimum count of the high-speed counter

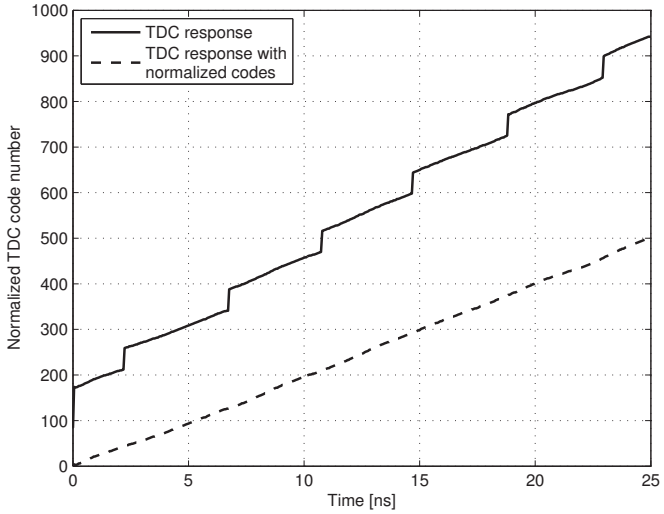

Figure 7.16 – TDC response obtained when varying the arriving time of the START signal with a programmable delay. As the low-speed period of the ring oscillator is not a multiple of 2 of the high-speed period, certain codes are unreachable. Therefore the TDC output needs to be normalized with a lookup table, which leads to a linear relationship

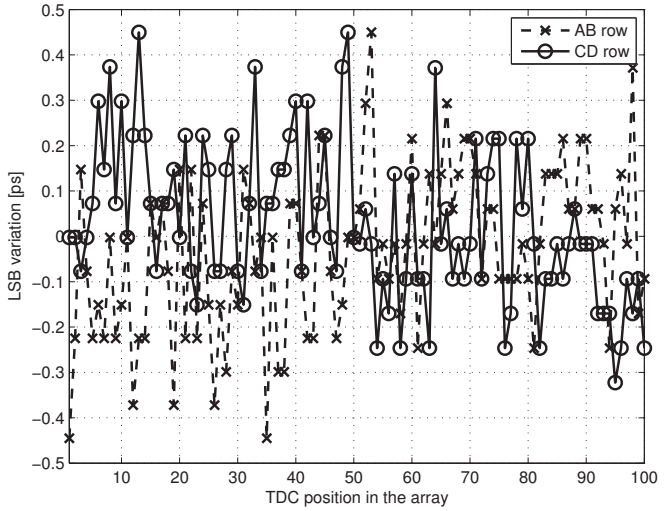

Figure 7.17 – LSB variation across the array.

is always higher than 1 (the RO has to be running during a certain period of time before it is stopped to ensure stability in the output frequency), there are certain codes that will never be reached. Thus it is necessary to use a lookup table that translates the resulting codes of the TDC into a normalized response. In Fig. 7.16 the normalized response of the TDC with a lookup table is also presented. The nominal resolution of the TDC (1 LSB) is 49.5 ps. Fig. 7.17 shows the resolution variations across the array.

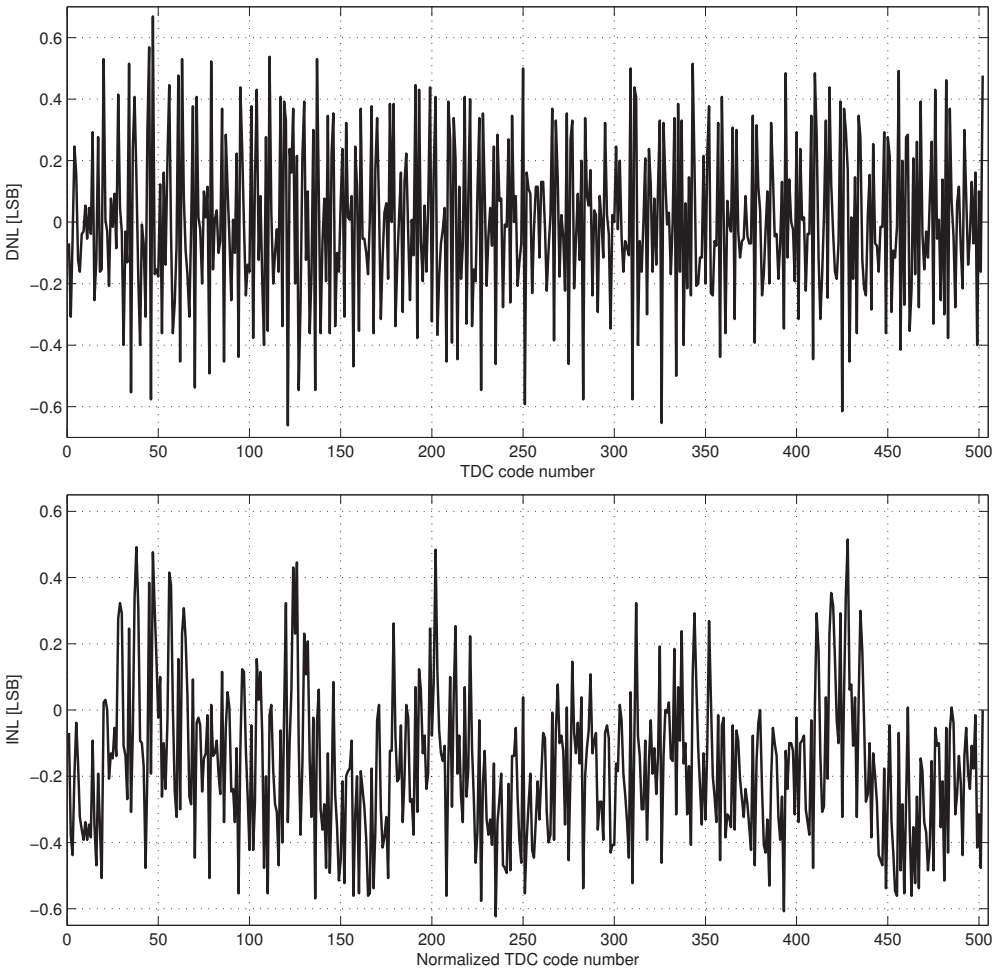

Figure 7.18 – Results of the the DNL (top) and INL (bottom) measured with a 25 ns period reference clock.

In order to measure the INL and DNL of the TDC, the SPAD pixels are employed to generate the *start* signal that triggers the conversion in the TDC. Under uncorrelated light conditions, the time of arrival of the photons detected by the SPAD should be completely random. If only one or less photons arrive during a TDC conversion period, the resulting histogram showing the distribution of times of arrival should be flat. However in reality some codes show higher occurrences than others. If enough photons are acquired, the DNL can be calculated by dividing the number of occurrences in each code by the average number of occurrences per code. The INL can be calculated by integrating the DNL. Fig. 7.18 shows the DNL and INL results, which ranged -0.62/+0.62 LSB and -0.61/+0.5 LSB respectively.

The timing characterization of the SPAD-TDC subsystem was obtained by illuminating the sensor with a supercontinuum laser (SuperK Extreme, NKT Photonics). The emitted light had a pulse width of 20 ps FWHM, while the optical power was adjusted to avoid pile-up during the time-correlated single-photon counting measurements. Fig. 7.19 shows the results obtained for SPADs A and C. The impulse response function is plotted for different NIR wavelengths showing a similar behavior for all of them. The best results in terms of signal to noise ratio were obtained with light of 750 nm wavelength, indicating the suitability of the sensor for NIROT applications. The FWHM jitter of the BSI SPADs as measured using the proposed TDC is plotted as a function of excess bias voltage.

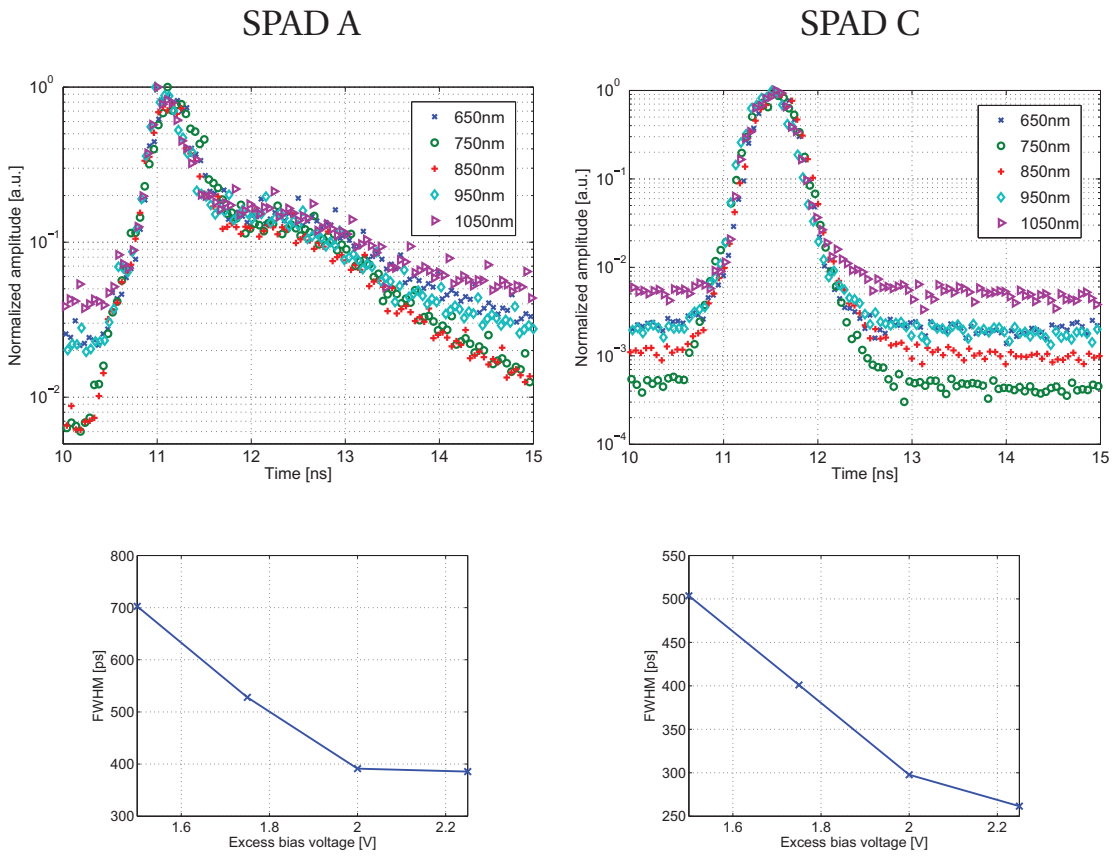

Figure 7.19 – Histograms of the response of BSI SPADs A and C at different wavelengths for an excess bias voltage of 1.5V (top). FWHM jitter as a function of the excess bias voltage for a 750 nm wavelength excitation (bottom).

7.9 Conclusions

The proposed sensor is to the best of our knowledge the first SPAD image sensor with integrated TDCs implemented in a deep-submicron 3D multi-wafer-stacking IC technology. It is also the first CMOS SPAD array with BSI pixels. The power consumption per TDC is the lowest ever reported for any SPAD-TDC, while this image sensor is the first fully dedicated to NIROT applications. Table 7.3 summarizes the performance of the chip and presents a comparison with the state of the art. The resource sharing scheme adopted in this image sensor together with the low power consumption of the presented new TDC architecture pave the way towards the implementation of larger sensors.

		This work	[Veerappan et al., 2011]	[Richardson et al., 2009b]	[Braga et al., 2014]	[Gersbach et al., 2012]
SPAD	Peak PDE [%]	11†	27.5	n.a.	45	25
	DCR [kHz]	10	0.16	n.a.	13.7	0.1
	Jitter (FWHM) [ps]	260	150	n.a.	150	185
TDC	Resolution (1 LSB) [ps]	49.5	55	52	65	119
	Range [ns]	133.3	55	53	262	100
	DNL [LSB]	-0.66/+0.66	-0.3/+0.3	-0.4/+0.4	-0.24/+0.28	-0.4/+0.4
	INL [LSB]	-0.61/0.5	-2/+2	-1.4/1.4	-3.9/+2.3	-1.2/+1.2
	Power consumption [μW @ 500 kS/s]	15	n.a.	38	94	48
Sensor	Process technology	CMOS 3D 130nm	CMOS 1P4M 0.13μm	CMOS 1P4M 0.13μm	CMOS 1P4M 0.13μm	CIS 130nm
	SPAD array size	4×400	160×128	32×32	192×480	32×32
	Chip size [mm^2]	0.77×5‡	12.3×11	n.a.	9.85×5.452	4.8×3.2
	Output data rate [Gbps]	1.04	51.2	5.12	1.6	10.24
	Clock frequency [MHz]	40	16	40	100	280

† Data obtained from a SPAD implemented in the same technology but that was not integrated in the array.
‡ Size of the image sensor with the I/O pads but without the SPAD test structures.

Table 7.3 – SPAD arrays performance comparison with the state of the art.

8 Conclusion and outlook

A new time-resolved NIROT system based on a SPAD image sensor was introduced for the first time in the present dissertation. The new NIROT instrument was based on a contactless optical setup that allowed the integration of thousands of detectors. This is an important aspect in NIROT devices since the quality of the reconstructions is highly dependent on the amount of information that is available in the reconstructions. A new illumination scheme was also implemented that made use of structured light to reduce the amount of sources and therefore the measurement time.

The large datasets provided by our new NIROT system required the use of high performance computers with a few hundred gigabytes of RAM memory. New fast reconstruction algorithms have been introduced that exploit the line-shaped structure of the light sources to reduce the ill-posed nature of the problem. Numerical and experimental results have been presented showing that it is possible to obtain sub-centimeter resolution with our system. We have also proven how time-resolved measurements improve the depth resolution in NIROT.

We have demonstrated how microlenses can improve the light sensitivity of SPAD image sensors with low fill-factors, broadening the field of application of current sensors that were not designed for low light applications such as NIROT.

A new SPAD image sensor tailored for NIROT was designed to improve the performance of our system. The main aims of this sensor were to implement SPAD pixels which were optimized

for NIR wavelengths, and to define a new architecture that allowed concurrent operation of all pixels. The sensor was fabricated in a 3D wafer stacking CMOS technology which allowed for much higher fill factors than conventional CMOS processes. To the best our knowledge, the first BSI SPAD was demonstrated in CMOS technology. A new resource sharing scheme together with a novel TDC architecture for low power operation were introduced. This more efficient allocation of resources will enable the scaling of future arrays to ever larger formats.

Further improvements in the NIROT system will broaden its field of application. The results presented in this work were obtained using media with regular geometries. The implementation of new fast reconstruction algorithms capable of working with arbitrary geometries is still a subject to much discussion in NIROT. Although algorithms employing FEM approaches seem very promising, they are computationally expensive and do not scale well with the number of sources and detectors. Simultaneous registration of the surface geometry would be a prerequisite in this case, if NIROT is not combined with other tomographic modalities such as MR or CT. This would not pose a great difficulty since the applicability of SPAD image sensors in the field of range imaging has been proven in the past.

One of the most critical aspects in the current NIROT system is the measurement time as the speed of the system will determine the type of metabolic processes that can be imaged, and its robustness against movement artifacts and environmental changes. Up to now, in time-resolve measurements this was constrained by the number of TDCs that were integrated on chip. Since the introduction of miniaturized TDCs suitable for in-pixel integration and resource sharing approaches this paradigm has shift. Now the bottleneck is in the on-chip and off-chip data transmission. Faster event-driven bus architectures need to be developed to reach the maximum physical speed in TCSPC measurements, together with the integration of a high number of low power on-chip data links that enable the transmission of the generated data. As an example we can take into consideration a system that is illuminated by a 80MHz clock. If all pixels are evenly illuminated and their photon detection rate is ~1%, each pixel will produce a timestamp at a rate of 800kHz. In a 256×256 sensor with 128 pixels connected to the same bus, each of them with 10 bit width timestamps and 7 bit address, in order to transfer all the generated data, the bus bandwidth would have to be 1.741 Gbps. For the full sensor the required I/O bandwidth is 891 Gbps for uncompressed data. Data compression and on-chip

histogramming will drastically reduce the I/O bandwidth requirements, however this is still a field that has not been explored for large number of pixels.

The results achieved in this thesis open a new paradigm in NIROT. New faster systems based on SPAD image sensors will be able to record high resolution tomographic images of chromophores in real time, enabling the study of biological processes with a new perspective.

Bibliography

[Abookasis et al., 2009] Abookasis, D., Lay, C. C., Mathews, M. S., Linskey, M. E., Frostig, R. D., and Tromberg, B. J. (2009). Imaging cortical absorption, scattering, and hemodynamic response during ischemic stroke using spatially modulated near-infrared illumination. *Journal of Biomedical Optics*, 14(2):024033–024033–9.

[Arridge et al., 1993] Arridge, S., Schweiger, M., Hiraoka, M., and Delpy, D. (1993). A finite element approach for modeling photon transport in tissue. *Medical physics*, 20(2):299–309.

[Arridge, 1999] Arridge, S. R. (1999). Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41.

[Arridge et al., 1992] Arridge, S. R., Cope, M., and Delpy, D. T. (1992). The theoretical basis for the determination of optical pathlengths in tissue: temporal and frequency analysis. *Phys. Med. Biol.*, 37:1531–1560.

[Arridge and Schotland, 2009] Arridge, S. R. and Schotland, J. C. (2009). Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12):123010.

[Becker, 2005] Becker, W. (2005). *Advanced Time-Correlated Single Photon Counting Techniques*. Chemical Physics. Springer Berlin Heidelberg.

[Bevilacqua and Depeursinge, 1999] Bevilacqua, F. and Depeursinge, C. (1999). Monte Carlo study of diffuse reflectance at source–detector separations close to one transport mean free path. *J. Opt. Soc. Am. A*, 16(12):2935–2945.

[Björn, 2007] Björn, L. (2007). *Photobiology: The Science of Life and Light*. Springer.

[Braga et al., 2014] Braga, L., Gasparini, L., Grant, L., Henderson, R., Massari, N., Perenzoni, M., Stoppa, D., and Walker, R. (2014). A fully digital 8×16 SiPM array for PET applications with per-pixel TDCs and real-time energy output. *Solid-State Circuits, IEEE Journal of*, 49(1):301–314.

[Bronzi et al., 2012] Bronzi, D., Villa, F., Bellisai, S., Markovic, B., Tisa, S., Tosi, A., Zappa, F., Weyers, S., Durini, D., Brockherde, W., and Paschen, U. (2012). Low-noise and large-area CMOS SPADs with timing response free from slow tails. In *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, pages 230–233.

[Bronzi et al., 2014] Bronzi, D., Villa, F., Tisa, S., Tosi, A., Zappa, F., Durini, D., Weyers, S., and Brockherde, W. (2014). 100 000 frames 64×32 single-photon detector array for 2-D imaging and 3-D ranging. *Selected Topics in Quantum Electronics, IEEE Journal of*, 20(6):1–10.

[Calvetti et al., 2004] Calvetti, D., Reichel, L., and Shuibi, A. (2004). L-curve and curvature bounds for tikhonov regularization. *Numerical Algorithms*, 35(2-4):301–314.

[Charbon and Donati, 2010] Charbon, E. and Donati, S. (2010). SPAD sensors come of age. *Opt. Photon. News*, 21(2):34–41.

[Charbon and Fishburn, 2011] Charbon, E. and Fishburn, M. (2011). *Monolithic Single-Photon Avalanche Diodes: SPADs*, volume 160 of *Springer Series in Optical Sciences*, pages 123–157. Springer Berlin Heidelberg.

[Cottini et al., 1963] Cottini, C., Gatti, E., and Svelto, V. (1963). A new method for analog to digital conversion. *Nuclear Instruments and Methods*, 24(0):241 – 242.

[Cova et al., 1981] Cova, S., Longoni, A., and Andreoni, A. (1981). Towards picosecond resolution with single-photon avalanche diodes. *Review of Scientific Instruments*, 52(3):408–412.

[Cova et al., 1983] Cova, S., Longoni, A., Andreoni, A., and Cubeddu, R. (1983). A semiconductor detector for measuring ultraweak fluorescence decays with 70 ps FWHM resolution. *Quantum Electronics, IEEE Journal of*, 19(4):630–634.

[Cuccia et al., 2009] Cuccia, D. J., Bevilacqua, F., Durkin, A. J., Ayers, F. R., and Tromberg, B. J. (2009). Quantitation and mapping of tissue optical properties using modulated imaging. *Journal of Biomedical Optics*, 14(2):024012–024012–13.

[Cutler, 1929] Cutler, M. (1929). Transillumination as an aid in the diagnosis of breast lesions. *Surgery, Gynecology and Obstetrics*, 48:721–729.

[Dalla Mora et al., 2012] Dalla Mora, A., Contini, D., Pifferi, A., Cubeddu, R., Tosi, A., and Zappa, F. (2012). Afterpulse-like noise limits dynamic range in time-gated applications of thin-junction silicon single-photon avalanche diode. *Applied Physics Letters*, 100(24):241111–241111–4.

[Deguchi et al., 1992] Deguchi, M., Maruyama, T., Yamasaki, F., Hamamoto, T., and Izumi, A. (1992). Microlens design using simulation program for CCD image sensor. *Consumer Electronics, IEEE Transactions on*, 38(3):583–589.

[Dehghani et al., 2009a] Dehghani, H., Eames, M. E., Yalavarthy, P. K., Davis, S. C., Srinivasan, S., Carpenter, C. M., Pogue, B. W., and Paulsen, K. D. (2009a). Near infrared optical tomography using NIRFAST: Algorithm for numerical model and image reconstruction. *Communications in Numerical Methods in Engineering*, 25(6):711–732.

[Dehghani et al., 2009b] Dehghani, H., White, B. R., Zeff, B. W., Tizzard, A., and Culver, J. P. (2009b). Depth sensitivity and image reconstruction analysis of dense imaging arrays for mapping brain function with diffuse optical tomography. *Appl. Opt.*, 48(10):D137–D143.

[Delpy et al., 1988] Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., and Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, 33(12):1433.

[Demmel, 1987] Demmel, J. W. (1987). On condition numbers and the distance to the nearest ill-posed problem. *Numerische Mathematik*, 51(3):251–289.

[Donati et al., 2007] Donati, S., Martini, G., and Norgia, M. (2007). Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits. *Opt. Express*, 15(26):18066–18075.

[Donati et al., 2011] Donati, S., Martini, G., and Randone, E. (2011). Improving photodetector performance by means of microoptics concentrators. *J. Lightwave Technol.*, 29(5):661–665.

[Fang and Boas, 2009] Fang, Q. and Boas, D. A. (2009). Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. *Opt. Express*, 17(22):20178–20190.

[Ferrari and Quaresima, 2012] Ferrari, M. and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2):921 – 935.

[Fishburn, 2012] Fishburn, M. (2012). *Fundamentals of CMOS Single-Photon Avalanche Diodes*. Ph.D. thesis, Delft University of Technology.

[Gersbach et al., 2012] Gersbach, M., Maruyama, Y., Trimananda, R., Fishburn, M., Stoppa, D., Richardson, J., Walker, R., Henderson, R., and Charbon, E. (2012). A time-resolved, low-noise single-photon image sensor fabricated in deep-submicron cmos technology. *Solid-State Circuits, IEEE Journal of*, 47(6):1394–1407.

[Gersbach et al., 2008] Gersbach, M., Niclass, C., Charbon, E., Richardson, J., Henderson, R., and Grant, L. (2008). A single photon detector implemented in a 130nm CMOS imaging process. In *Solid-State Device Research Conference, 2008. ESSDERC 2008. 38th European*, pages 270–273.

[Ghioni et al., 2007] Ghioni, M., Gulinatti, A., Rech, I., Zappa, F., and Cova, S. (2007). Progress in silicon single-photon avalanche diodes. *Selected Topics in Quantum Electronics, IEEE Journal of*, 13(4):852–862.

[Graber et al., 1993] Graber, H. L., Chang, J., Aronson, R., and Barbour, R. L. (1993). A perturbation model for imaging in dense scattering media: derivation and evaluation of imaging operators. *Medical Optical Tomography: Functional Imaging and Monitoring*, pages 121–143.

[Gratton et al., 1993] Gratton, E., Mantulin, W. W., vandeVen, M. J., Fishkin, J. B., Maris, M. B., and Chance, B. (1993). A novel approach to laser tomography. *Bioimaging*, 1(1):40–46.

[Green, 2008] Green, M. A. (2008). Self-consistent optical parameters of intrinsic silicon at 300°K including temperature coefficients. *Solar Energy Materials and Solar Cells*, 92(11):1305 – 1310.

[Grosenick et al., 2003] Grosenick, D., Moesta, K. T., Wabnitz, H., Mucke, J., Stroszczynski, C., Macdonald, R., Schlag, P. M., and Rinneberg, H. (2003). Time-domain optical mammography: Initial clinical results on detection and characterization of breast tumors. *Appl. Opt.*, 42(16):3170–3186.

[Habermehl et al., 2012] Habermehl, C., Holtze, S., Steinbrink, J., Koch, S. P., Obrig, H., Mehnert, J., and Schmitz, C. H. (2012). Somatosensory activation of two fingers can be discriminated with ultrahigh-density diffuse optical tomography. *NeuroImage*, 59(4):3201 – 3211.

[Haensse, 2005] Haensse, D. V. (2005). *Changes in Cerebral Oxygenation in Response to Various Stimuli in Newborns as Measured by Functional Near-Infrared Spectroscopy*. Ph.D. thesis, Swiss Federal Institute of Technology Zurich.

[Haskell et al., 1994] Haskell, R. C., Svaasand, L. O., Tsay, T.-T., Feng, T.-C., McAdams, M. S., and Tromberg, B. J. (1994). Boundary conditions for the diffusion equation in radiative transfer. *J. Opt. Soc. Am. A*, 11(10):2727–2741.

[Hayakawa et al., 2001] Hayakawa, C. K., Spanier, J., Bevilacqua, F., Dunn, A. K., You, J. S., Tromberg, B. J., and Venugopalan, V. (2001). Perturbation Monte Carlo methods to solve inverse photon migration problems in heterogeneous tissues. *Opt. Lett.*, 26(17):1335–1337.

[Höckel et al., 1996] Höckel, M., Schlenger, K., Aral, B., Mitze, M., Schäffer, U., and Vaupel, P. (1996). Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix. *Cancer Research*, 56(19):4509–4515.

[Jacobsen et al., 2003] Jacobsen, M., Hansen, P., and Saunders, M. (2003). Subspace preconditioned LSQR for discrete ill-posed problems. *BIT Numerical Mathematics*, 43:975–989. 10.1023/B:BITN.0000014547.88978.05.

[Jacques, 2013] Jacques, S. L. (2013). Optical properties of biological tissues: a review. *Physics in Medicine and Biology*, 58(11):R37.

[Jobsis, 1977] Jobsis, F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 198(4323):1264–1267.

[Kak and Slaney, 1988] Kak, A. C. and Slaney, M. (1988). *Principles of Computerized Tomographic Imaging*. IEEE Press.

[Kappagoda et al., 1979] Kappagoda, C. T., Linden, R. J., and Newell, J. P. (1979). A comparison of the oxygen consumption/body weight relationship obtained during submaximal exercise on a bicycle ergometer and on a treadmill. *Experimental Physiology*, 64(3):205–215.

[Knoll, 2010] Knoll, G. F. (2010). *Radiation detection and measurement; 4th ed.* Wiley, New York, NY.

[Konecky, 2008] Konecky, S. D. (2008). *Non-invasive imaging of breast cancer with diffusing near-infrared light*. Ph.D. thesis, University of Pennsylvania.

[Konecky et al., 2009] Konecky, S. D., Mazhar, A., Cuccia, D., Durkin, A. J., Schotland, J. C., and Tromberg, B. J. (2009). Quantitative optical tomography of sub-surface heterogeneities using spatially modulated structured light. *Opt. Express*, 17(17):14780–14790.

[Konecky et al., 2008] Konecky, S. D., Panasyuk, G. Y., Lee, K., Markel, V., Yodh, A. G., and Schotland, J. C. (2008). Imaging complex structures with diffuse light. *Opt. Express*, 16(7):5048–5060.

[Leff et al., 2008] Leff, D., Warren, O., Enfield, L., Gibson, A., Athanasiou, T., Patten, D., Hebden, J., Yang, G., and Darzi, A. (2008). Diffuse optical imaging of the healthy and diseased breast: A systematic review. *Breast Cancer Research and Treatment*, 108(1):9–22.

[Leitner et al., 2013] Leitner, T., Feiningstein, A., Turchetta, R., Coath, R., Chick, S., Visokolov, G., Savuskan, V., Javitt, M., Gal, L., Brouk, I., Bar-Lev, S., and Nemirovsky, Y. (2013). Measurements and simulations of low dark count rate single photon avalanche diode device in a low voltage 180-nm cmos image sensor technology. *Electron Devices, IEEE Transactions on*, 60(6):1982–1988.

[Li et al., 2010] Li, D.-U., Arlt, J., Richardson, J., Walker, R., Buts, A., Stoppa, D., Charbon, E., and Henderson, R. (2010). Real-time fluorescence lifetime imaging system with a 32×32 0.13μm cmos low dark-count single-photon avalanche diode array. *Opt. Express*, 18(10):10257–10269.

[Liu, 2005] Liu, J. (2005). *Photonic Devices*. Cambridge University Press.

[Lukic et al., 2009] Lukic, V., Markel, V. A., and Schotland, J. C. (2009). Optical tomography with structured illumination. *Opt. Lett.*, 34(7):983–985.

[Mandai and Charbon, 2013a] Mandai, S. and Charbon, E. (2013a). A 4 × 4 × 416 digital SiPM array with 192 TDCs for multiple high-resolution timestamp acquisition. *Journal of Instrumentation*, 8(05):P05024.

[Mandai and Charbon, 2013b] Mandai, S. and Charbon, E. (2013b). Timing optimization of a H-tree based digital silicon photomultiplier. *Journal of Instrumentation*, 8(09):P09016.

[Mandai et al., 2012] Mandai, S., Fishburn, M. W., Maruyama, Y., and Charbon, E. (2012). A wide spectral range single-photon avalanche diode fabricated in an advanced 180nm CMOS technology. *Opt. Express*, 20(6):5849–5857.

[Mandelis, 2001] Mandelis, A. (2001). *Diffusion-wave fields: mathematical methods and Green functions*. Springer Verlag.

[Markel and Schotland, 2004] Markel, V. A. and Schotland, J. C. (2004). Symmetries, inversion formulas, and image reconstruction for optical tomography. *Phys. Rev. E*, 70:056616.

[Maruyama et al., 2013] Maruyama, Y., Blacksberg, J., and Charbon, E. (2013). A 1024×8 700ps time-gated SPAD line sensor for laser Raman spectroscopy and LIBS in space and rover-based planetary exploration. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, pages 110–111.

[Mata Pavia et al., 2012] Mata Pavia, J., Charbon, E., and Wolf, M. (2012). 3D near-infrared imaging based on a single-photon avalanche diode array sensor: A new perspective on reconstruction algorithms. In *Biomedical Optics and 3-D Imaging*, page BW1A.5. Optical Society of America.

[Mata Pavia et al., 2011] Mata Pavia, J., Wolf, M., and Charbon, E. (2011). 3D near-infrared imaging based on a single-photon avalanche diode array sensor. In *Proc. SPIE*, volume 8088, pages 808811–808811–6.

[Mata Pavia et al., 2014a] Mata Pavia, J., Wolf, M., and Charbon, E. (2014a). Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery. *Opt. Express*, 22(4):4202–4213.

[Mata Pavia et al., 2014b] Mata Pavia, J., Wolf, M., and Charbon, E. (2014b). Single-photon avalanche diode imagers applied to near-infrared imaging. *Selected Topics in Quantum Electronics, IEEE Journal of*, 20(6).

[Mayer and Vaupel, 2013] Mayer, A. and Vaupel, P. (2013). Hypoxia, lactate accumulation, and acidosis: Siblings or accomplices driving tumor progression and resistance to therapy? In *Oxygen Transport to Tissue XXXV*, volume 789 of *Advances in Experimental Medicine and Biology*, pages 203–209.

[Meijlink et al., 2011] Meijlink, J. R., Veerappan, C., Seifert, S., Stoppa, D., Henderson, R., Charbon, E., and Schaart, D. (2011). First measurement of scintillation photon arrival statistics using a high-granularity solid-state photosensor enabling time-stamping of up to 20,480 single photons. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, pages 2254–2257.

[Migacheva et al., 2010] Migacheva, E. V., Chamot, S. R., Seydoux, O., Weber, B., Depeursinge, C., Marquet, P., and Magistretti, P. J. (2010). Determination of local optical properties of the rat barrel cortex during neural activation: Monte-Carlo approach to light propagation. In *Proc. SPIE*, volume 7715, pages 771531–771531–6.

[Movsas et al., 2002] Movsas, B., Chapman, J., Hanlon, A. L., Horwitz, E. M., Greenberg, R. E., Stobbe, C., Hanks, G. E., and Pollack, A. (2002). Hypoxic prostate/muscle pO2 ratio predicts for biochemical failure in patients with prostate cancer: preliminary findings. *Urology*, 60(4):634 – 639.

[Niclass, 2008] Niclass, C. (2008). *Single-Photon Image Sensors in CMOS: Picosecond Resolution for Three-Dimensional Imaging*. Ph.D. thesis, Swiss Federal Institute of Technology Lausanne.

[Niclass et al., 2008] Niclass, C., Favi, C., Kluter, T., Gersbach, M., and Charbon, E. (2008). A 128×128 single-photon image sensor with column-level 10-bit time-to-digital converter array. *Solid-State Circuits, IEEE Journal of*, 43(12):2977–2989.

[Niclass et al., 2007] Niclass, C., Gersbach, M., Henderson, R., Grant, L., and Charbon, E. (2007). A single photon avalanche diode implemented in 130-nm CMOS technology. *Selected Topics in Quantum Electronics, IEEE Journal of*, 13(4):863–869.

[Niclass et al., 2006] Niclass, C., Sergio, M., and Charbon, E. (2006). A CMOS 64×48 single photon avalanche diode array with event-driven readout. In *Solid-State Circuits Conference, 2006. ESSCIRC 2006. Proceedings of the 32nd European*, pages 556–559.

[O'Leary, 1996] O'Leary, M. A. (1996). *Imaging with diffuse photon density waves*. Ph.D. thesis, University of Pennsylvania.

[Patterson et al., 1989] Patterson, M. S., Chance, B., and Wilson, B. C. (1989). Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties. *Appl. Opt.*, 28(12):2331–2336.

[Pifferi et al., 2004] Pifferi, A., Swartling, J., Chikoidze, E., Torricelli, A., Taroni, P., Bassi, A., Andersson-Engels, S., and Cubeddu, R. (2004). Spectroscopic time-resolved diffuse reflectance and transmittance measurements of the female breast at different interfiber distances. *Journal of Biomedical Optics*, 9(6):1143–1151.

[Pifferi et al., 2008] Pifferi, A., Torricelli, A., Spinelli, L., Contini, D., Cubeddu, R., Martelli, F., Zaccanti, G., Tosi, A., Dalla Mora, A., Zappa, F., and Cova, S. (2008). Time-resolved diffuse reflectance using small source-detector separation and fast single-photon gating. *Phys. Rev. Lett.*, 100:138101.

[Renker, 2007] Renker, D. (2007). New trends on photodetectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 571(1–2):1 – 6. Proceedings of the 1st International Conference on Molecular Imaging Technology EuroMedIm 2006.

[Richardson et al., 2009a] Richardson, J., Grant, L., and Henderson, R. (2009a). Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology. *Photonics Technology Letters, IEEE*, 21(14):1020–1022.

[Richardson et al., 2009b] Richardson, J., Walker, R., Grant, L., Stoppa, D., Borghetti, F., Charbon, E., Gersbach, M., and Henderson, R. (2009b). A 32×32 50ps resolution 10 bit time to

digital converter array in 130nm CMOS for time correlated imaging. In *Custom Integrated Circuits Conference, 2009. CICC '09. IEEE*, pages 77–80.

[Ripoll, 2010] Ripoll, J. (2010). Hybrid fourier-real space method for diffuse optical tomography. *Opt. Lett.*, 35(5):688–690.

[Ripoll and Ntziachristos, 2006] Ripoll, J. and Ntziachristos, V. (2006). From finite to infinite volumes: Removal of boundaries in diffuse wave imaging. *Physical Review Letters*, 96:173903.

[Rochas, 2003] Rochas, A. (2003). *Single Photon Avalanche Diodes in CMOS Technology*. Ph.D. thesis, Swiss Federal Institute of Technology Lausanne.

[Rochas et al., 2003a] Rochas, A., Gani, M., Furrer, B., Besse, P. A., Popovic, R. S., Ribordy, G., and Gisin, N. (2003a). Single photon detector fabricated in a complementary metal-oxide-semiconductor high-voltage technology. *Review of Scientific Instruments*, 74(7):3263–3270.

[Rochas et al., 2003b] Rochas, A., Gosch, M., Serov, A., Besse, P. A., Popovic, R., Lasser, T., and Rigler, R. (2003b). First fully integrated 2-D array of single-photon detectors in standard CMOS technology. *Photonics Technology Letters, IEEE*, 15(7):963–965.

[Sandell and Zhu, 2011] Sandell, J. L. and Zhu, T. C. (2011). A review of in-vivo optical properties of human tissues and its impact on PDT. *Journal of Biophotonics*, 4(11-12):773–787.

[Scholkmann et al., 2014] Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., and Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, 85, Part 1(0):6 – 27. Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS).

[Selb et al., 2007] Selb, J., Dale, A. M., and Boas, D. A. (2007). Linear 3D reconstruction of time-domain diffuse optical imaging differential data: improved depth localization and lateral resolution. *Opt. Express*, 15(25):16400–16412.

[Shockley and Read, 1952] Shockley, W. and Read, W. T. (1952). Statistics of the recombinations of holes and electrons. *Phys. Rev.*, 87:835–842.

[Spinelli and Lacaita, 1997] Spinelli, A. and Lacaita, A. (1997). Physics and numerical simulation of single photon avalanche diodes. *Electron Devices, IEEE Transactions on*, 44(11):1931–1943.

[Strangman et al., 2002] Strangman, G., Boas, D. A., and Sutton, J. P. (2002). Non-invasive neuroimaging using near-infrared light. *Biological Psychiatry*, 52(7):679 – 693.

[Sugioka, 1996] Sugioka, K. (1996). From Germany to Britton. In Ince, C., Kesecioglu, J., Telci, L., and Akpir, K., editors, *Oxygen Transport to Tissue XVII*, volume 388 of *Advances in Experimental Medicine and Biology*, pages 1–11. Springer US.

[Sumner, 2001] Sumner, R. (2001). A sliding scale method to reduce the differential non linearity of a time digitizer. In *Nuclear Science Symposium Conference Record, 2001 IEEE*, volume 2, pages 803–806.

[Tuchin, 2007] Tuchin, V. (2007). *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*. Press Monographs. SPIE.

[Vardi and Nini, 2008] Vardi, M. and Nini, A. (2008). Near-infrared spectroscopy for evaluation of peripheral vascular disease. a systematic review of literature. *European Journal of Vascular and Endovascular Surgery*, 35(1):68 – 74.

[Vaupel and Mayer, 2014] Vaupel, P. and Mayer, A. (2014). Hypoxia in tumors: Pathogenesis-related classification, characterization of hypoxia subtypes, and associated biological and clinical implications. In *Oxygen Transport to Tissue XXXVI*, volume 812 of *Advances in Experimental Medicine and Biology*, pages 19–24.

[Vaupel et al., 2006] Vaupel, P., Mayer, A., and Höckel, M. (2006). Impact of hemoglobin levels on tumor oxygenation: the higher, the better? *Strahlentherapie und Onkologie*, 182(2):63–71.

[Veerappan and Charbon, 2014] Veerappan, C. and Charbon, E. (2014). A substrate isolated CMOS SPAD enabling wide spectral response and low electrical crosstalk. *Selected Topics in Quantum Electronics, IEEE Journal of*, 20(6):1–7.

[Veerappan et al., 2011] Veerappan, C., Richardson, J., Walker, R., Li, D.-U., Fishburn, M. W., Maruyama, Y., Stoppa, D., Borghetti, F., Gersbach, M., Henderson, R. K., et al. (2011). A

160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pages 312–314. IEEE.

[Volpe, 2009] Volpe, J. J. (2009). Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. *The Lancet Neurology*, 8(1):110 – 124.

[Webster et al., 2012a] Webster, E., Grant, L., and Henderson, R. (2012a). A high-performance single-photon avalanche diode in 130-nm CMOS imaging technology. *Electron Device Letters, IEEE*, 33(11):1589–1591.

[Webster et al., 2012b] Webster, E. A. G., Richardson, J. A., Grant, L. A., Renshaw, D., and Henderson, R. K. (2012b). A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm. *Electron Device Letters, IEEE*, 33(5):694–696.

[Wolf et al., 2007] Wolf, M., Ferrari, M., and Quaresima, V. (2007). Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *Journal of Biomedical Optics*, 12(6):062104–062104–14.

[Wolf and Greisen, 2009] Wolf, M. and Greisen, G. (2009). Advances in near-infrared spectroscopy to study the brain of the preterm and term neonate. *Clinics in Perinatology*, 36(4):807 – 834. Neurology of the Newborn Infant.

[Wolf et al., 2008] Wolf, M., Morren, G., Haensse, D., Karen, T., Wolf, U., Fauchère, J., and Bucher, H. (2008). Near infrared spectroscopy to study the brain: an overview. *Opto-Electronics Review*, 16(4):413–419.

[Wolf et al., 2012] Wolf, M., Naulaers, G., van Bel, F., Kleiser, S., and Griesen, G. (2012). Review: A review of near infrared spectroscopy for term and preterm newborns. *Journal of Near Infrared Spectroscopy*, 20(1):43–55.

[Wolf et al., 2003] Wolf, U., Wolf, M., Choi, J. H., Levi, M., Choudhury, D., Hull, S., Coussiratb, D., Paunescu, L., Safonova, L. P., Michalos, A., Mantulin, W. W., and Gratton, E. (2003). Localized irregularities in hemoglobin flow and oxygenation in calf muscle in patients with

peripheral vascular disease detected with near-infrared spectrophotometry. *Journal of Vascular Surgery*, 37(5):1017 – 1026.

[Wray et al., 1988] Wray, S., Cope, M., Delpy, D. T., Wyatt, J. S., and Reynolds, E. R. (1988). Characterization of the near infrared absorption spectra of cytochrome aa3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 933(1):184 – 192.

[Wu et al., 2013] Wu, J.-Y., Li, S.-C., Hsu, F.-Z., and Lin, S.-D. (2013). Two-dimensional mapping of photon counts in low-noise single-photon avalanche diodes. In *Int. Image Sens. Workshop*.

[Zimmermann et al., 2013] Zimmermann, R., Braun, F., Achtnich, T., Lambercy, O., Gassert, R., and Wolf, M. (2013). Silicon photomultipliers for improved detection of low light levels in miniature near-infrared spectroscopy instruments. *Biomed. Opt. Express*, 4(5):659–666.

List of publications

International peer-reviewed journal articles

Mata Pavia, J., Wolf, M., Charbon, E. (2014). Single-Photon Avalanche Diode Imagers Applied to Near-Infrared Imaging, *Selected Topics in Quantum Electronics, IEEE Journal of*, 20(6).

Mata Pavia, J., Wolf, M., Charbon, E. (2014). Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery, *Opt. Express* 22(4):4202-4213.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., **Mata Pavia, J.**, Wolf, U., Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology, *NeuroImage*, 85(1):6-27.

International conferences

Mata Pavia, J., and Charbon, E., Wolf, M. (2011). 3D Near-infrared Imaging based on a Single-photon Avalanche Diode Array. In *In Proceedings SPIE*, volume 8088, pages 808811-808811-6. **Invited talk**.

Mata Pavia, J., Niclass, C., Favi, C., Wolf, M., Charbon, E. (2011). 3D Near-Infrared Imaging Based on a SPAD Image Sensor, In *Intl. Image Sensor Workshop (IISW)*, page R42.

Mata Pavia, J., and Charbon, E., Wolf, M. (2012). 3D near-infrared imaging based on a single-photon avalanche diode array sensor: A new perspective on reconstruction algorithms, In *Biomedical Optics and 3-D Imaging*, page BW1A.5, Optical Society of America.

Acknowledgements

The work presented in this thesis could not have been possible without the help and support from many people. I would like to thank my advisors Martin Wolf and Edoardo Charbon for giving me the opportunity to work on this exciting project. I am very grateful to Martin Wolf for his scientific advice and his personal support during the past five years. Although I did not have many chances to meet him in person, Edoardo was always very receptive and gave me advice on any problem I encountered. I am also very thankful to him.

I would like to thank the whole BORL team for their help, both direct and indirect, in bringing this project forward. Scott Lindner, thanks for staying up all night when it was necessary to finish the work and for his invaluable help correcting texts. I wish to thank Damien de Courten for being a lovely office mate, and for teaching me that you should only go for the best in class equipment and anything else in life. Marc Hohmann was of great help characterizing the sensors, despite the short period of time he stayed with us he did tremendous work. I will never forget the nice evenings that Andreas Metz organized at bQm and their follow ups. Raphael Zimmermann, thanks for those lovely evenings of playing and singing. I would like to especially thank Iris Suter for her help in all administrative issues and nice conversations. I also wish to express my gratitude to all present and former members of the group that contributed to this work, among them: Mark Adams, Linda Ahnen, Dirk Bassler, Martin Biallas, Hans Ulrich Bucher, Simon Christen, Jean-Claude Fauchère, Cornelia Hagmann, Julia Hillebrandt, Stefan Kleiser, Brigitte Koller, Christoph Kuhn, Thomas Mühlemann, Nassim Nasseri, Daniel Ostojic, Marcin Pastewski, Felix Scholkmann, Oliver Sieber, Sonja Spichtig, Helen Stachel, Ivo Trajkovic, Flavia Wehrle, and Dominik Wyser.

I am also grateful to Markus Rudin and his group for his collaboration in the last years. I am especially thankful to Catherine Germanier, Andreas Elmer and Wuwei Ren for their help in the fluorescence experiments in mice.

I wish to express my gratitude to all people in the Circuits & Systems group at TU Delft who were really friendly during my stay in Delft. I also wish to thank Mauro Scandiuzzo for his help in the longest tape out we ever experienced and for being so close despite living so far away.

Lastly I would like to thank my family for their support and motivation. To my siblings Rosa, Andrea and Ignasi, for always believing in me and giving me their support. To my father for always encouraging us to travel and learn from other cultures, and to my mother for teaching me that what matters is to strive and not to yield. To my children Andreu, Oliver and Nadja for being patient with their dad in the long working days, and to always make me forget everything else in life when I was with them. Finally none of this work would have been possible without the unconditional help, support and motivation of my wife Divna, she has always been my guiding light, in part this has been her thesis too.

About the author

Juan Mata Pavia was born in Castelló de la Plana, Spain, in 1979. He received the M.S. degree in electrical engineering from the Royal Institute of technology (KTH), Stockholm, Sweden, and the telecommunications engineering diploma from the Polytechnic University of Valencia (UPV), Valencia, Spain, in 2004. After receiving his degree, he joined Philips Semiconductors, where he was involved in the design and test of application processors for smartphones. In 2009 he joined the Swiss Federal Institute of Technology Lausanne (EPFL) as a Ph.D. candidate at the Quantum Architecture Group (AQUA). He is currently working at the University Hospital Zurich in the development of new biomedical optic applications with single-photon image sensors.