

Large-Format Time-Gated SPAD Cameras for Real-Time Phasor-Based FLIM

Présentée le 18 mai 2021

Faculté des sciences et techniques de l'ingénieur
Laboratoire d'architecture quantique
Programme doctoral en microsystemes et microélectronique

pour l'obtention du grade de Docteur ès Sciences

par

Arin Can ÜLKÜ

Acceptée sur proposition du jury

Prof. M. A. Ionescu, président du jury
Prof. E. Charbon, Dr C. Bruschini, directeurs de thèse
Prof. E. Gratton, rapporteur
Dr X. Michalet, rapporteur
Prof. N. Quack, rapporteur

Acknowledgments

First of all, I would like to thank my thesis advisor, Prof. Edoardo Charbon, for giving me the opportunity to do a PhD in his laboratory at EPFL, and for his continuous support in my entire thesis work. Under his leadership, I had the chance to constantly pursue challenging research goals, while having access to abundant resources and being surrounded by knowledgeable and supportive colleagues. I would like to equally thank my thesis co-advisor, Dr. Claudio Bruschini, for his supervision throughout my PhD and his invaluable guidance on all of my publications, including this thesis. In both experimental methods and academic writing, Claudio, like Edoardo, always expected high standards, which ultimately led to good outcomes.

Additionally, I would like to express my gratitude to Dr. Xavier Michalet from the University of California at Los Angeles (UCLA). In our collaborative project, I had the chance to work closely with him. Both his theoretical and experimental contributions were an essential part of the work on FLIM that is presented in this thesis. I would also like to thank him for being available for help whenever I had any questions in his field.

In AQUA lab, I had the chance to work closely with several colleagues, to whom I would like to extend my special thanks. Firstly, I would like to thank Andrei Ardelean. I worked in the same research project and shared an office with Andrei for more than three years. He made significant contributions to all aspects of my work. In addition to co-authoring a paper with me, he generously offered me help in multiple fields including FPGA design, system development and data analysis. I would also like to thank Dr. Ivan Michel Antolovic. Our work on microlenses reached its current state under Michel's coordination. He also helped me in most of my optical experiments, in both experimental setup design and data analysis. I would like to thank Dr. Augusto Carimatto. While working on his project for a brief period of time, I gained extensive experience in CMOS IC design, which has proved to be very useful during my own tape-outs. I am also grateful to Augusto for his friendship and personal support. I would like to thank Dr. Samuel Burri and Dr. Chockalingam Veerappan. The image sensors that I designed during my PhD contain significant elements from their previous work. I would like to additionally thank Samuel for his help during the system design phase of my research. I would like to thank Paul Mos, Dr. Michael Wayne, Baris Can Efe and Ermanno Bernasconi for their work on the camera systems based on the image sensors that I designed. Thanks to their efforts, the specifications of the cameras are continuously improving, which is opening the way to many new potential applications and users.

I would also like to thank all current and former members in AQUA lab at EPFL and TU Delft, including Francesco Gramuglia, Simone Frasca, Ekin Kizilkan, Utku Karaca, Dr. Kazuhiro Morimoto, Andrada Muntean, Dr. Myung-Jae Lee, Dr. Emanuele Ripicini, Dr. Chao Zhang, Pouyan Keshavarzian, Ming-Lo Wu, Jiuxuan Zhao, Jad Benserhir, Dr. Preethi Padmanabhan, Dr. Augusto Ximenes, Dr. Scott Lindner, Dr. Andrea Ruffino and Dr. Feng Liu. I would like to thank the AQUA lab secretary Brigitte Khan for her assistance in administrative matters.

During my PhD, I had the chance to work in collaboration with several academic groups inside and outside EPFL. At UCLA, we collaborated with the research group of Prof. Shimon Weiss. I would like to thank Prof. Weiss for this collaboration and for the opportunity to visit his laboratory twice to conduct FLIM experiments. In his group, I would like to thank Dr. Rinat Ankri and Dr. Yung Kuo for their work in our publications and their help during our visits. At the University of Wisconsin-Madison, our collaboration with Prof. Mohit Gupta's

group led to an important publication in the field of computational photography. I am thankful to Prof. Gupta and Sizhuo Ma for expanding the applications of our cameras to new fields. At EPFL, I would also like to thank Dr. Nicolas Deschames and Dr. Arne Seitz for their contributions in the microscopy experiments.

Finally, I would like to thank my parents for their unconditional love and support.

Abstract

Fluorescence lifetime imaging microscopy (FLIM) is an imaging modality often used to monitor biochemical properties of a cell or a tissue. In addition to conventional fluorescence microscopy features, such as selective labeling and non-invasiveness, FLIM enables high background rejection and insensitivity to fluorophore concentration, tissue thickness and photobleaching. In this context, the phasor method has recently gained popularity in FLIM owing to its simple arithmetic operations and graphical representation. However, its widefield implementation has been mostly limited to systems that are expensive, fragile, or of limited flexibility.

This thesis aims to overcome these challenges using SPAD technology by exploiting its single-photon sensitivity and CMOS integration capability. To achieve this goal, two large-format time-resolved SPAD imagers were designed. Time gating was selected over time-stamping due to simpler pixel implementation and higher photon-count rates, along with high spatial resolution. Of the two imagers, the first, SwissSPAD2, employs a 512×512 SPAD array, which generates photon-counting binary frames at up to 97,700 frames-per-second. SwissSPAD2 features a single time gate, which can reach a minimum width of 10.8 ns separated by steps of 17.9 ps. The SPAD in each pixel achieves one of the best photon detection probabilities (up to ~50% at 520 nm) and dark count rates (median of 7.5 cps) among current standard CMOS SPADs. The pixel native fill factor is 10.5%; microlenses were deposited on the pixel array achieving a concentration factor of up to 4.2 at the optimal photon angle of incidence. This resulted in an effective fill factor of 44.1%. The second imager, SwissSPAD3, comprises 500×500 SPAD pixels. Each pixel exploits two contiguous parallel time gates, allowing 100% duty cycle. It also has a minimum gate width of 1 ns, which, to the best of our knowledge, is the lowest gate width achieved by a large-format SPAD imager to date. The microlensed version of SwissSPAD3 is also available; its characterization is expected to yield similar concentration factors as for SwissSPAD2.

Using SwissSPAD2, widefield phasor-FLIM was demonstrated with a time-gated large-format SPAD imager for the first time. It was shown that the camera is able to distinguish two single-exponential lifetimes with 1.4 ns difference using as low as 16 gate positions, corresponding to an equivalent acquisition frame rate of 12.1 fps. Moreover, FLIM-FRET analysis capabilities were validated by a double-exponential mixture analysis. In this experiment, the relation between the phasor ratios of five mixtures of two single-exponential fluorescence dyes and their respective volume fractions were consistent with expectations. The photon economy was found to be mainly determined by the overall temporal resolution of the photon time of arrival, and is largely independent of the detected photon count. This finding indicates that the gate width does not pose a fundamental limitation to the minimum detectable lifetime, allowing the estimation and separation of lifetimes that are much shorter than the gate width.

This thesis demonstrates that time-gated SPAD imagers are competitive detectors for widefield FLIM. Future steps, such as integrated data processing, pixel pitch miniaturization, increased number of shorter gate channels, and microlens optimization are expected to further exploit the potential of SPAD technology.

Keywords: single-photon avalanche diode, SPAD, fluorescence lifetime imaging microscopy, FLIM, phasor analysis, time gating, time-resolved, image sensor, widefield, CMOS.

Sunto

L'imaging del tempo di vita in fluorescenza (FLIM) è una tecnica di microscopia molecolare spesso usata per monitorare le proprietà biochimiche di una cellula o di un tessuto. Oltre alle caratteristiche della semplice microscopia a fluorescenza i vantaggi del FLIM includono un'elevata reiezione di fondo e l'insensibilità alla concentrazione di fluorofori, allo spessore del tessuto e al photobleaching. In questo contesto, il metodo phasor ha recentemente raggiunto una certa notorietà nella comunità di utenti FLIM grazie alle sue semplici operazioni aritmetiche e alla rappresentazione grafica. Tuttavia, la sua implementazione ad ampio campo è stata per lo più limitata a sistemi dove i flussi di fotoni sono limitati o sensori con sensibilità a singolo fotone limitata.

Questa tesi mira a superare queste limitazioni sfruttando la sensibilità degli SPAD e l'integrazione CMOS. Per raggiungere questo obiettivo, sono stati progettati due imager di grande formato con alta risoluzione temporale. Un otturatore elettronico o gate temporale è stato scelto piuttosto che un convenzionale time stamping a causa dell'implementazione più semplice dei pixel e dei tassi di conteggio dei fotoni più alti, insieme all'alta risoluzione spaziale. Il primo imager, SwissSPAD2, impiega una matrice SPAD di 512×512 pixel, che genera fotogrammi binari fino a 97.700 fotogrammi al secondo. SwissSPAD2 dispone di un singolo gate temporale, che può raggiungere una durata minima di 10,8 ns ed è posizionabile a passi di 17,9 ps. Lo SPAD in ogni pixel raggiunge una delle migliori combinazioni di probabilità di rivelazione di fotoni (fino al ~50%) e rumore in oscurità (mediana di 7,5 cps) tra gli SPAD CMOS standard. Il fattore di riempimento nativo dei pixel è del 10,5%; sono state depositate microlenti sull'array di pixel ottenendo un fattore di concentrazione fino a 4,2 (all'angolo di incidenza ottimale dei fotoni). Questo ha portato ad un fattore di riempimento effettivo del 44,1%. Il secondo imager, SwissSPAD3, comprende 500×500 pixel SPAD. Ogni pixel sfrutta due gate temporali paralleli contigui, permettendo un duty cycle del 100%. Ha anche una durata minima di gate di 1 ns, che, a nostra conoscenza, è la più bassa mai raggiunta finora da un imager SPAD di grande formato. È disponibile anche la versione con microlenti di SwissSPAD3; la sua caratterizzazione dovrebbe dare fattori di concentrazione pari a quelli di SwissSPAD2.

Usando SwissSPAD2, è stato dimostrato il phasor-FLIM ad ampio campo con un imager SPAD di grande formato provvisto di gate temporale. La camera è in grado di distinguere due tempi di vita mono-esponenziali con una differenza di 1,4 ns usando solo 16 posizioni di gate, corrispondenti ad un'acquisizione di 12,1 fps. Inoltre, la capacità d'analisi FLIM-FRET è stata convalidata da un'analisi della miscela doppio-esponenziale. In questo esperimento, la relazione tra i rapporti di phasor di cinque miscele di due coloranti a fluorescenza mono-esponenziale e le loro rispettive frazioni di volume erano coerenti con i modelli teorici. Si è scoperto che l'economia dei fotoni è determinata principalmente dalla risoluzione temporale complessiva del tempo di arrivo dei fotoni ed è ampiamente indipendente dal conteggio dei fotoni rilevati. Questa scoperta indica che la larghezza del gate non pone una limitazione fondamentale al tempo di vita minimo rilevabile, permettendo la stima e la separazione di tempi di vita molto più brevi della larghezza del gate.

Questa tesi dimostra che gli imager SPAD provvisti di gate temporali sono ottimi candidati per il FLIM ad ampio campo. In futuro si prevede di integrare l'elaborazione dei dati, la miniaturizzazione dei pixel, l'aumento del numero di canali di gate più corti e l'ottimizzazione delle microlenti.

Parole chiave: diodo fotorivelatore a singolo fotone, SPAD, l'imaging del tempo di vita in fluorescenza, FLIM, analisi phasor, gate temporale, a risoluzione temporale, sensore d'immagine / imager, ampio campo, CMOS.

Contents

Acknowledgments	i
Abstract.....	iii
Sunto.....	v
Contents	vii
List of figures	xi
List of tables	xxiii
List of acronyms	xxv
Chapter 1 Introduction.....	1
1.1 Time-resolved imaging and applications	1
1.2 FLIM overview	2
1.2.1 Förster resonance energy transfer (FRET).....	4
1.2.2 Confocal vs. widefield FLIM	5
1.3 Phasor approach in FLIM	10
1.3.1 Time-correlated single-photon counting (TCSPC)	15
1.3.2 Analog frequency domain FLIM.....	17
1.3.3 Digital frequency domain FLIM (DFD-FLIM) / time gating.....	21
1.4 Sensor architectures for widefield FLIM.....	23
1.4.1 Gated intensified CCD (ICCD).....	24
1.4.2 Position-sensitive anode detectors	26
1.4.3 Gated CMOS image sensor (CIS).....	27
1.4.4 Single-photon avalanche diode	29
1.5 Contributions	31
1.6 Organization of the thesis.....	32
Chapter 2 The theory and practice of time-gated FLIM.....	33
2.1 Overview	33
2.2 Conventional implementations of time gates	34
2.2.1 Time gating in rapid lifetime determination method.....	34
2.2.2 Time gating for the least squares method.....	37

2.3	Our gating approach for phasor-FLIM	38
2.4	Gate parameters	40
2.4.1	Gate width	41
2.4.2	Transition time (rise and fall time)	41
2.4.3	Skew.....	42
2.4.4	Gate step.....	44
2.4.5	Number of gate positions	44
2.4.6	Number of gate channels	49
2.5	Performance limitations of time-gated phasor-based FLIM using SPADs	49
2.5.1	Photon sensitivity	50
2.5.2	Photon economy.....	51
2.5.3	Lifetime range.....	53
2.5.4	Dynamic range	54
2.5.5	Accuracy.....	57
Chapter 3	SwissSPAD2/3: Large-format time-gated SPAD imagers.....	59
3.1	Motivation.....	59
3.2	SPAD structure	59
3.3	SwissSPAD2/3: Design evolution	61
3.3.1	Pixel architecture	61
3.3.2	Time gate implementation	68
3.3.3	Readout and system architecture	71
3.4	Sensor characterization	76
3.4.1	Photon detection probability (PDP).....	76
3.4.2	Dark count rate (DCR).....	77
3.4.3	Crosstalk	80
3.4.4	Frame rate	81
3.4.5	Dynamic range	84
3.4.6	Gate characteristics	87
3.5	Microlens design and characterization	94
3.5.1	Motivation and theory.....	94
3.5.2	Design of the microlenses in SwissSPAD2	97
3.5.3	Measurement results.....	99

3.6	State-of-the-art comparison	103
3.7	Future work.....	106
Chapter 4	Application results	109
4.1	Sensor performance trade-offs in the target applications	109
4.2	Phasor-based widefield FLIM.....	112
4.2.1	Methodology	112
4.2.2	Phasor analysis of fluorescent solutions	120
4.2.3	Influence of frame rate on phase lifetime precision	123
4.2.4	Influence of gate width on phase lifetime determination	125
4.2.5	Dye mixture analysis.....	127
4.2.6	Phase lifetime map for complex samples.....	129
4.2.7	State-of-the-art comparison.....	130
4.2.8	Discussion	135
4.3	Phasor analysis in highly scattering medium	136
4.4	Computational photography for high-speed, low-light-level scenes	137
Chapter 5	Conclusions.....	141
5.1	Summary	141
5.2	Future work.....	143
Appendix A:	SwissSPAD2 crosstalk characteristics	145
References.....		147
Chip gallery.....		167
List of publications		169
Curriculum vitae		171

List of figures

Figure 1:1 Jablonski diagram showing the excitation and de-excitation processes of a fluorescent molecule [32, 33]. Adapted from [32].	3
Figure 1:2 Schematic of (a) confocal and (b) widefield microscopy. Adapted from [38]. The confocal microscopy setup illustrated in (a) is a sample-scanning one, in which the sample is scanned along the x,y and z axes. The red dot on the object focal plane is the point in the sample that is captured by the detector.	5
Figure 1:3 Point spread function (PSF) of a widefield (left) and confocal (right) microscope [51].	8
Figure 1:4 (a) Conceptual illustration of phasor analysis [81]. Z_{avg} , the phasor of a fluorescence decay, is on the universal semicircle if the decay is single-exponential. As lifetime increases, a single-exponential decay phasor approaches the origin along the semicircle. (b) The phasor plot of a mixture with two single-exponential components [81]. The phasor ratio can be calculated from d_1 and d_2 , the distances between the phasor of its mixture and the phasors of its components.	12
Figure 1:5 Microscopic images and phasor representation of CHO-K1 cells transfected with paxillin EGFP in a collagen matrix [77]. The phasor plot in (A) contains the phasors of two different image with the cell and the collagen matrix (B-D), and only collagen matrix (E-G). The three fluorescence contributions are EGFP (region 1 in the phasor plot), collagen (region 2) and background autofluorescence (region 3). The pixels whose phasors are in region 1 and region 2 are highlighted in (C, F) and (D, G), respectively. From the absence of phasors of region 1 in (F) and the clear correlation between the highlighted regions in (D, G) and the structural differences in the image, it can be concluded that the phasor representation can be used to distinguish different parts of an image with distinct decay characteristics.	13
Figure 1:6 (a, b) Conceptual illustration of TCSPC. (a) The sample is excited multiple times by a laser pulse, and the time of arrival of each detected emitted photon is recorded. (b) Subsequently, a histogram of the time of arrival of all detected photons is generated. The shape of the histogram represents the properties of the measured fluorescence decay function. (c, d) Data acquisition schemes of (c) confocal and (d) widefield TCSPC for a FOV captured by 8×8 pixels. Each gray square which represents a pixel in the scene is illuminated in a raster-scanning configuration in (c), an in parallel in (d). The illumination, which in reality covers the entire pixel in (c) and FOV in (d), is represented by circular red dots that only cover a part of the corresponding area, for illustrative purposes...	16
Figure 1:7 Principles of analog frequency-domain FLIM [96].	18

Figure 1:8 Operating principles of digital frequency domain FLIM (DFD-FLIM) with heterodyning [89].	21
Figure 1:9 Conceptual illustration of time-gated FLIM [117].	23
Figure 1:10 (a) Diagram of an ICCD camera structure [8] (b) Gate profiles of an ICCD camera for various gate widths [133].	24
Figure 1:11 (a) Diagram of the H33D detector, a position-sensitive detector with a cross delay-line (XDL) anode, designed for widefield FLIM [7]. (b) The experimental setup used in [7] to perform widefield FLIM. For each detected photon, the time of arrival with respect to the laser pulse is detected by the TDC, and the XY position is detected by the cross delay-line (XDL) anode.	27
Figure 1:12 (a) Schematic, (b) layout and (c) timing diagram of a 4-tap CMOS lock-in pixel [9].	28
Figure 1:13 (a) Schematic of a SPAD pixel with its most basic elements [173]. The voltage spike at the floating node of the SPAD, caused by the avalanche current, is converted to a digital pulse by an inverter. (b) Conceptual image of a SPAD imager in 3D stacked process technology [15].	30
Figure 2:1 Conceptual illustration of time-gated FLIM in time domain. The intensity of a fluorescence emission decay, which is generated as a response to an excitation pulse, is measured for different gate windows. This diagram represents only one example of gating; time gating can employ various gate configurations which are discussed throughout the chapter. Adapted from [188].	33
Figure 2:2 (a) The gates used in the original rapid lifetime determination (RLD) method for single-exponential decay and an unknown base line [191]. (b) The variant of the original RLD method for no base line [73]. (c) Ratio of relative standard deviations (σ/τ) of RLD with no base line and weighted linear least squares (WLLS) method, for different combinations of relative gate width ($\Delta t/\tau$) and total decay function area ($A\tau$) [73].	35
Figure 2:3 (a) A generalized single-exponential two-gated RLD scheme with custom gate width and delay parameters [187]. (b) Precision (relative standard deviation) comparison of the optimized RLD method with standard RLD and weighted nonlinear least-squares (NLLS) methods [187].	37
Figure 2:4 (a) Diagram of the time-gated FLIM acquisition method for histogram generation to be used in the least squares method [129]. Multiple gates with a width of TG record the intensity of the fluorescence decay with a constant gate step s . In the standard acquisition method, all gates have equal integration times. For optimal SNR, the gates sampling the later parts of the decay must have relatively longer integration times than the earlier gates. (b) Plot showing the optimal integration time ratio and gate separation for a two-gated single-exponential lifetime analysis using the scheme in (a) [129]. The color bar represents the lifetime determination error of a measurement in which the two variables, integration time ratio and gate separation, correspond to the x and y coordinates of a data point, respectively. Integration time ratio is defined as the ratio between the integration duration (acquisition time) of gate 2 and gate 1. Gate	

separation defines the delay between the rising edges of two gates, in terms of the lifetime. The red curve shows the gate separation that yields the lowest error as a function of integration time ratio. The area in black corresponds to the set of integration time ratio and gate separation that yields the lowest error. 38

Figure 2:5 Conceptual diagram of overlapping gates, which are defined as gate positions whose corresponding time windows overlap with each other. Overlapping gates are generated when the delay between adjacent gate positions (gate step) is smaller than the gate width. Overlapping gates allow sampling a fluorescence decay with a short lifetime with sufficiently high photon count in each gate position for a decent SNR. 39

Figure 2:6 Position of the first and last gate in the data acquisition process. This positioning allows the detection of only one fluorescence decay, which makes the lifetime estimation method insensitive to any mismatch between the phasor frequency and laser PRF. 40

Figure 2:7 A diagram illustrating the gate width. 41

Figure 2:8 A diagram illustrating the rise and fall times of the gate window profile. 42

Figure 2:9 Conceptual illustration of skew in an integrated circuit. The propagation delay of the signal across the wires results in a delay between the signal edges (t_e) at different points of the signal distribution network..... 43

Figure 2:10 The description of gate step in (a) non-overlapping and (b) overlapping gates. 44

Figure 2:11 A diagram illustrating the required conditions for proportionality between the total detected photon count (C_{det_total}) and number of gates in the acquisition (G), in time-gated phasor-FLIM. Measurement window (D), which is defined as $D = G \times GS$, is equal to the laser period (T_{lsr}). Three different gate configurations are represented. (a) Overlapping gates where GW is not an integer multiple of GS . (b) Overlapping gates where GW is an integer multiple of GS . Note that contiguous gates are a subset of this category, in which $GW = GS$. (c) Separated gates. As analytically shown in Equations (2.11)-(2.14), C_{det_total} is proportional to G independently of the measured function only if $GW \bmod GS = 0$. In this theoretical analysis of an ideal square gate, pile-up effect and temporal changes in the fluorescence response are neglected. 46

Figure 2:12 (a-b) The Fourier transform of a single-exponential decay function sampled with (a) 8 and (b) 32 contiguous ideal square gates [85]. The curve in dotted line shows the continuous Fourier transform of the function. (c) Modified phasor plot where the SEPL is recalculated from the modified version of the single-exponential Fourier transform for contiguous gates with widths shorter than the lifetime [85]. The number of gate positions, which is denoted by K in these figures, is denoted by G in the thesis..... 48

Figure 2:13 Illustration of various terms used in the calculation of C_{det}/C_{imp} in Equation (2.22), for a system with two non-overlapping gate channels, where $T_{lsr} \gg \tau$... 51

Figure 2:14 A diagram illustrating the lifetime range of a FLIM method. The blue curve is the F value distribution of a particular FLIM method for a range of lifetimes. The lifetime range, τ_{range} , is defined as the difference of the maximum and minimum lifetimes which can be estimated by the method with an F value no higher than F_{max} , denoted by τ_{max}

and τ_{min} , respectively. F_{max} is a user-defined value based on minimum required photon economy of a particular application, which can also be derived by the target precision and photon count (N) values, from the relation expressed in Equation (2.31). 54

Figure 2:15 The theoretical relation between the detected and measured photon counts in a SPAD imager with an active clock-driven recharge and 8-bit output image format. The red curve represents an ideal shot-noise-limited system with no pile-up or saturation mechanism, where the measured and detected photon counts are equal. The black curve represents a SPAD imager with the aforementioned characteristics, in which the pile-up effect introduces an upper bound to the maximum detectable photon count. The analytical expression for the black curve was retrieved from [203], and expressed in Equation (2.33). 55

Figure 3:1 Cross-section of the p-i-n SPAD structure used in SwissSPAD2 and SwissSPAD3 [212]. 60

Figure 3:2 (a) Pixel schematic of SwissSPAD2. The pixel consists of 11 NMOS transistors, 7 with thick-oxide and 4 with thin-oxide gate. The pixel stores a binary photon count in its memory capacitor. The in-pixel gate defines the time window, with respect to a 20 MHz external trigger signal, in which the pixel is sensitive to photons. (b-c) Pixel micrograph of SwissSPAD2 with (b) round SPAD and (c) square SPAD with rounded corners. Scale bar: 5 μm 62

Figure 3:3 Timing diagram of the operation of SwissSPAD2. (a) Global shutter: upper panel: gate window and readout signal sequence for two consecutive binary frame time windows. Lower panel: gate shifting mechanism during the transition between two b-bit frames. (b) Rolling shutter: gate window and readout signal sequence for two consecutive binary frame time windows. The gate shifting mechanism described in (a) also applies here. 65

Figure 3:4 (a) Pixel schematic of SwissSPAD3. (b) Pixel micrograph of SwissSPAD3, with a round SPAD. Scale bar: 5 μm 66

Figure 3:5 Timing diagram showing the SwissSPAD3 pixel output for different photon detection scenarios. The photons in the diagram represent the photons that are detected by the SPAD. *Out_1* is activated if at least one photon is detected inside the frame. *Out_2* is activated if the first photon in the frame is outside the gate signal pulse window, and not activated if it is inside. Both *Out_1* and *Out_2* are insensitive to subsequent photons detected in a frame. Note that the 01 state is not possible and represents thus a fault if it occurs. 67

Figure 3:6 Timing diagram of the operation of SwissSPAD3 in rolling shutter mode. 67

Figure 3:7 Conceptual diagram illustrating the relationship between the gate window and the gate signals in (a) SwissSPAD2 and (b) SwissSPAD3. 69

Figure 3:8 Three alternative gate signal distribution schemes in a pixel array: (a) No signal tree, (b) column-level signal tree, (c) pixel level H-tree. 70

Figure 3:9 Sensor architecture block diagram of (a) SwissSPAD2 and (b) SwissSPAD3. The photomicrographs of (c) SwissSPAD2 and (d) SwissSPAD3. 72

- Figure 3:10 System architecture block diagram of SwissSPAD2 for (a) 8-bit continuous mode imaging and (b) 1-bit burst mode imaging. 74
- Figure 3:11 Relations between bit depth and compression ratio, frame rate and camera data rate of SwissSPAD2. Frame rate and camera data rate are calculated for the maximum binary frame rate of 97.7 kfps. It can be observed that for continuous operation at the fastest binary frame rate, a minimum of 5-bit data compression is required. 75
- Figure 3:12 Photon detection probability (PDP) characterization of the p-i-n SPAD used in SwissSPAD2 and SwissSPAD3, for wavelengths between 400 and 860 nm [202]. . 76
- Figure 3:13 Dark count rate (DCR) characterization of SwissSPAD2. (a) 2D DCR map. This map shows that the hot pixels are randomly scattered across the field of view. (b) Population distribution of DCR at various excess bias voltages. The disparity between mean and median DCR is due to the high counts of the hot pixels. (c) The population distribution of DCR for five image sensor samples. Spatial resolution: 472×256... 78
- Figure 3:14 Dark count rate (DCR) characterization of SwissSPAD3. Spatial resolution: 499×498. (a) 2D DCR map. This map shows that the hot pixels are randomly scattered across the field of view. $V_{ex} = 6$ V. (b) Population distribution of DCR at various excess bias voltages. The difference between mean and median DCR is due to the high counts of the hot pixels..... 79
- Figure 3:15 Crosstalk percentage values of a SwissSPAD2 sample (Sample 1 of 5) [81]. 81
- Figure 3:16 Grayscale images of a rotating fan captured in global shutter mode at (a) 4.6 kfps for 4 bits (b) 273.5 fps for 8 bits. The binary frame rate is 69.7 kfps. Spatial resolution: 472×255. 81
- Figure 3:17 Binary images of a rotating fan captured at 96.8 kfps in rolling shutter mode. Spatial resolution: 512×256. 82
- Figure 3:18 A sequence of 1,000,000 binary frames capturing a digital stopwatch at 985 kfps in rolling shutter. (a) Original binary frames. (b) 8-bit images formed by the average of 255 adjacent frames including the index displayed in each image. Spatial resolution: 512×24. 83
- Figure 3:19 Dynamic range characterization of SwissSPAD2 for 8-bit images. Number of pixels: 472×256. The images are captured under 13 different light levels ($LL0 - LL12$). Exposure time: $LL0 - 10$: 1.63 ms, $LL11$: 3.26 ms, $LL12$: 6.52 ms. (a) Standard deviation and (b) signal to noise ratio (SNR) of the *measured photon count* m_i as a function of the mean measured photon count. (c) Standard deviation and (d) SNR of the *estimated detected photon count* n_i as a function of the mean measured photon count. (e) Standard deviation and (f) SNR of the *estimated detected photon count* as a function of the estimated mean detected photon count. There is a consistent match between the measured SNR values and the theoretical SNR values based on the model that includes both shot noise and pile-up. (g) Measured photon count spatial distribution under low light level. The match between the measured and Poisson fit values demonstrates the single-photon counting ability of the imager. (h) Measured photon count distribution under higher light level. A good agreement with the Poisson fit values can still be

observed, despite the mildly visible effects of pile-up, as expected. The distributions in (g) and (h) were generated after the removal of hot pixels which consist of approximately 2% of the entire array.86

Figure 3:20 Gate characteristics of SwissSPAD2. (a-b) The conceptual illustration of the experiment which was conducted to generate the gate profile. Gp : gate position, $Tlsr$: laser period, Gs : gate step. For visualization purposes, $Tlsr = 5Gs$ was selected, as opposed to $Tlsr = 2,800Gs$, which represents the actual experimental settings. (c) The gate window profile. The transition times and the gate width are annotated in the figure. The gate width is user-programmable, and the minimum gate width in the internal laser trigger mode is 10.8 ns. The response of every 4th pixel is plotted. (d-e) Spatial distribution of the (d) rising edge and (e) falling edge position of the gate window. The insets show the 2D color map of the gate edge positions. Experimental parameters: Laser controller trigger mode: internal, laser PRF: 20 MHz, array size: 472×256, shutter mode: global shutter, bit depth: 10, gate step: 17.86 ps, number of gate positions: 4,000, integration window: 71.4 ns.89

Figure 3:21 Gate characteristics of two gate channels of SwissSPAD3 for the minimum achieved gate width. (a-c) and (d-f) represent the characteristics of the first (G1) and the second gate channel (G2), respectively. (a, d) Gate window profile. The transition times and the gate width are annotated in the figure. Percentage of plotted pixels: G1: 4%, G2: 21%. (b-c, e-f) Spatial distribution of the (b, e) rising edge and (c, f) falling edge position of the gates. The insets show the 2D color map of the gate edge positions. Experimental parameters: Laser wavelength: 637 nm, laser controller trigger mode: internal, laser PRF: 40 MHz, array size: 499×249, binary frame rate: 49.8 kfps, shutter mode: rolling shutter, bit depth: 8, gate step: 17.9 ps, number of gate positions: 1,600, integration window: 28.6 ns, Vex: 6 V.90

Figure 3:22 Gate characteristics of two gate channels of SwissSPAD3 with comparable gate widths. (a-c) and (d-f) represent the characteristics of the first (G1) and the second gate channel (G2), respectively. (a, d) Gate window profile. The transition times and the gate width are annotated in the figure. Percentage of plotted pixels: G1: 11%, G2: 10%. (b-c, e-f) Spatial distribution of the (b, e) rising edge and (c, f) falling edge position of the gates. The insets show the 2D color map of the gate edge positions. Experimental parameters: Laser wavelength: 637 nm, laser controller trigger mode: internal, laser PRF: 40 MHz, array size: 499×249, binary frame rate: 49.8 kfps, shutter mode: rolling shutter, bit depth: 8, gate step: 17.9 ps, number of gate positions: 1,600, integration window: 28.6 ns, Vex: 6 V.91

Figure 3:23 Photon count distribution characteristics of the fully open gate positions in SwissSPAD3. (a-b) Gate window shape for (a) gate 1 (G1) and (b) gate 2 (G2). Plotted pixel percentage: G1: 10%, G2: 14%. (c-d) Theoretical and experimental SNR difference of the detected photon counts of the fully open gate positions of (c) G1 and (d) G2. The difference between the theoretical SNR which is limited by the shot noise and pile-up, and the measured SNR, displayed in (c-d), was calculated for each pixel separately. (e-f) Spatial distribution of the difference between the theoretical and experimental SNR, for (e) G1 and (f) G2. As the SNR of G1 and G2 is only 0.12 dB and 0.058 dB lower than the theoretical

values, respectively, it can be concluded that the photon counts of the fully open gate positions are independent of the gate positions themselves, and limited by shot noise and pile-up. This proves that in the flat parts of the gate window, there is negligible deviation from the characteristics of a rectangular gate. Experimental parameters: Pixel format: (a-b): 499×249, (c-f) 499×498, laser PRF: 40 MHz, laser wavelength: 510 nm, number of fully open gate positions: $G1$: 480, $G2$: 580, V_{ex} : 6 V. 93

Figure 3:24 Illustration of the angles used during the discussion of microlenses. p is the photon trajectory, θ is the angle of incidence, and φ is the azimuth. 95

Figure 3:25 A typical optical stack for CMOS/SPAD cameras that shows the media in the trajectory of the photons until their detection..... 96

Figure 3:26 (a) Diagram of a telecentric optical system, often used in microscopes. High α is desirable for high SNR and resolving power, whereas the ability to reduce θ allows the maximization of the concentration factor of microlenses. The two downsides of this approach, zero angle of view and the large focal length requirement for tube lenses, do not introduce any negative effect to the system due to the lack of need for the sense of depth in microscopy and the relatively higher availability of space for the tube lens compared to a photography setup. (b) Diagram of a conventional single-lens optical system typically used in photography. High θ is desirable to improve photon sensitivity by collecting more light from the object, and minimize depth of field. In addition, concerns for compactness introduce an upper limit for f , further restricting the lower bound of θ . These constraints limit the maximum achievable concentration factor for microlenses. 98

Figure 3:27 Microlens design parameters. Adapted from [233]..... 99

Figure 3:28 Photomicrograph of a group of pixels in SwissSPAD2 with microlenses. Scale bar: 200 μm 100

Figure 3:29 Fluorescence intensity image of a *convallaria majalis* sample using a SwissSPAD2 (a) without and (b) with microlenses [81]. Experimental parameters: V_{ex} : 6.5 V, array size: 453×210, bit depth: 10, integration time: 3.21 ms, $\lambda_{emission}$: 607 nm, pile-up correction: on. Hot pixels with 1% highest dark count rate in the array were corrected using an interpolation method based on setting their intensity values to the mean of the four nearest-neighbor pixels. Different color brightness scales were used in the two images to achieve maximum contrast. 100

Figure 3:30 Concentration factor characterization of multiple SwissSPAD2 samples with microlenses at various excess bias voltages. SPAD shape: (a) round, (b) square with rounded corners. Measurements were done at optimal angle of incidence for each sample; the tuning of the angle was performed during the calibration stage prior to the measurements. 102

Figure 3:31 Concentration factor spatial distribution of two SwissSPAD2 samples with the highest measured average CF for their respective SPAD types. SPAD shape: (a) round (sample: EOR), (b) square with rounded corners (sample: D6S). Measurements were done at optimal angle of incidence for each sample; the tuning of the angle was performed

during the calibration stage prior to the measurements. $V_{\text{ex}} = 6.5 \text{ V}$, Binning: 16×16	103
Figure 3:32 Photon detection probability (PDP) and dark count rate (DCR) comparison between the p-i-n SPAD used in SwissSPAD2/3 and state-of-the-art standard CMOS-based FSI SPADs [117]. DCR represents the median value. The PDP and 0° C DCR results are based on nanoSPAD [218], a chip that employs a variant of the p-i-n SPAD design used in SwissSPAD2/3 with $107.5 \mu\text{m}^2$ active area as opposed to the $28.3 \mu\text{m}^2$ active area in the case of SwissSPAD2/3.	104
Figure 4:1 Readout sequence of the pixels in one half of SwissSPAD2 with 512×256 pixels. From this diagram, it can be observed that while the total frame readout time is directly proportional to the number of rows that are selected, it can only be reduced by up to a factor of four by reading out fewer columns.	110
Figure 4:2 (a) Trade-off between frame rate and duty cycle as a function of the binary exposure time, T_{exp} , in global shutter mode. The calculations are based on Equations (4.1) and (4.3). $T_{\text{read}} = 10.24 \mu\text{s}$, $T_{\text{lsr}} = 50 \text{ ns}$, $GW = 50 \text{ ns}$. (b) Trade-off between F-value and duty cycle as a function of the gate width, GW , in global shutter mode. The calculations are based on Equations (2.27) and (4.3). $T_{\text{exp}} = 10 \mu\text{s}$, $T_{\text{read}} = 10.24 \mu\text{s}$, $T_{\text{lsr}} = 50 \text{ ns}$, $\tau = 5 \text{ ns}$	111
Figure 4:3 (a) Gate window characteristics of SwissSPAD2. The response of every 4 th pixel in the 472×256 array is plotted. While the width of the plotted gate is 13.1 ns , the minimum achievable gate width is 10.8 ns . (b) Conceptual illustration of scanning the fluorescent decay with the time gate. The areas in red and blue indicate the gate window and the fluorescence decay, respectively. The scanning is performed with a uniform gate step, across the 50 ns laser period.	113
Figure 4:4 Conceptual diagram of the widefield FLIM setup, used in the experiments reported in this chapter.	114
Figure 4:5 (a) Illustration of the uncorrelated background correction method employed in the FLIM experiments reported throughout this chapter. The red curve shows the gated fluorescence response of the reference sample, and the gate positions located between the two vertical dashed lines are selected to calculate the mean of the uncorrelated background signal, on a pixel-per-pixel basis. It is desirable to include as many gate positions as possible in this selection, in order to calculate the mean background signal with minimum shot noise. (b) Illustration of a fluorescence emission response where two adjacent periods have a significant overlap. In the scenarios where the calibration sample displays these characteristics, our background correction method cannot be applied to the calibration sample.	116
Figure 4:6 Phasors of two single-exponential uniform fluorescent dye solutions (a) before and (b) after uncorrelated background correction. More information on the particular data set and the experiment is available in Subsection 4.2.2 and Figure 4:9.	117
Figure 4:7 Conceptual illustration of multi-exponential decay analysis. The phasor of the mixture (colored green) is located on the line segment between the phasors of its components (colored red). The volume fraction of the components can be derived from	

the phasor ratio using Equation (4.14), which is calculated from the ratio of the distances of the mixture phasor to its components (Equation (4.13)). 120

Figure 4:8 Gate response of the detector to four different fluorescent samples with various lifetimes and emission intensities, which are used in the experiments throughout this chapter. The samples are: (a) ATTO 550, (b) Cy3B, (c) Rhodamine 6G (R6G) and (d) Quantum Dot (QD) 585. The blue and red signals indicate the gate photon counts before and after pile-up correction, respectively. The signals represent the convolution of the real emission profile and the IRF. The plots represent the response of a single pixel with the coordinates of (193, 190). Experimental parameters: laser PRF: 20 MHz, gate width: 13.1 ns, bit depth: 10, background correction: off. 121

Figure 4:9 Phasors of Rhodamine 6G (R6G) ($\tau = 4.08$ ns) and Cy3B ($\tau = 2.8$ ns) solutions using data sets with (a) 16, (b) 140 and (c) 2,800 gate positions. The ATTO 550 ($\tau = 3.6$ ns) sample was used for calibration. Each phasor point represents a 4×4 ROI. All ATTO 550 phasor locations are identical, since IRF deconvolution was performed separately for each ROI. A decrease in the number of gates leads to a proportional decrease in the recorded number of photons, which increases the dispersion of the phasors of the ROIs due to shot noise. The separability of R6G and Cy3B ($\Delta\tau_{real} = 1.28$ ns, $\Delta\tau_{meas} = 1.4$ ns) is achieved down to 16 gates, corresponding to an acquisition frame rate of 12.1 fps. Experimental parameters: laser and phasor frequency: 20 MHz, gate width: 13.1 ns, array size: 472×256 , binning: 4×4 , bit depth: 8 (R6G & Cy3B), 10 (ATTO 550), pile-up correction: on, background correction, on, percentage of removed pixels: 0% (R6G, Cy3B), 0.5% (ATTO 550). 122

Figure 4:10 Influence of acquisition frame rate on (a) lifetime accuracy, (b) lifetime standard deviation, (c) number of recorded photons, and (d) F-value, based on FLIM experiments on ATTO 550 and R6G solutions ($\# Gates = G$). (a) The dashed lines show the expected lifetimes of the dyes obtained from the literature. Error bars represent the standard deviation of the lifetimes of all ROIs. (b-c) The dashed lines are the fitted lines to the data points. They indicate that $\sigma\tau$ and N are proportional to $G - 1/2$ and G , respectively. (d) The dashed lines show Monte Carlo simulation results. A possible cause of the discrepancy between the Monte Carlo and measurement results of R6G is the SNR drop due to pile-up, which is excluded from the Monte Carlo simulation. Experimental parameters: laser and phasor frequency: 20 MHz, gate width: 13.1 ns, array size: 472×256 , binning: 4×4 , bit depth: 8, pile-up correction: on, background correction: on. ... 124

Figure 4:11 Influence of gate width on the phase lifetime determination performance, using ATTO 550 and Rhodamine 6G (R6G) dye solutions. (a) Accuracy as a function of gate width. The dashed lines show the expected lifetimes of the dyes obtained from the literature. Error bars show the standard deviation over all ROIs. (b-c) Standard deviation as a function of gate width, for (b) 140 and (c) 16 gate positions. (d) F-value, an indicator of photon economy, as a function of gate width. In (b-d), the plain and dashed lines represent the analytical shot-noise model and Monte Carlo simulation results, respectively. In (d), filled and open dots represent $G = 16$ and $G = 140$, respectively. These figures show that while the accuracy remains relatively constant at the available gate width range, the photon economy improves as the gate width decreases.

Experimental parameters: laser and phasor frequency: 20 MHz, gate width: 13.1 ns, array size: 472×256, binning: 4×4, bit depth: 8, pile-up correction: on, background correction: on. 126

Figure 4:12 Mixture analysis results of Cy3B and Rhodamine 6G (R6G) dyes. (a) Phasor plot of R6G, Cy3B, and five different mixtures of the two dyes with different volume fractions. Two separately acquired Cy3B data sets were used for calibration and actual measurement, in order to include the shot noise introduced by all samples. Calibration was performed based on the Cy3B lifetime of $\tau = 2.5$ ns, which was separately measured in a reference TCSPC setup. (b) Measured Cy3B phasor ratios of the mixtures as a function of their volume fractions. The dashed line represents the relation between the phasor ratio and volume fraction, shown in Equation (4.14) [90], using the value for the parameter $\mu\chi$ obtained by fitting. The data points and error bars represent the mean and standard deviation of the calculated phasor ratios of the ROIs, respectively. Experimental parameters: laser and phasor frequency: 20 MHz, number of gate positions: 234, gate width: 22.8 ns, array size: 248×160, binning: 8×8, bit depth: 10, pile-up correction: 10, background correction: on, percentage of removed pixels: 1%. 128

Figure 4:13 Lifetime analysis of a dried quantum dot sample. (a) Fluorescence intensity image. Scale bar: 25 μm . (b) Phase lifetime image. (c) An image that combines both intensity and phase lifetime information. Colors represent the lifetime according to the color spectrum in (b), and the brightness represents the intensity. (d) Phasor distribution of the sample. In the color-coding method in (b-c), two values representing high and low lifetimes were selected from (d) as $\tau_H = 16.7$ ns and $\tau_L = 13.9$ ns which are marked on the phasor plot with green and red dots, respectively. The mapping of the pixels to the color spectrum in (b) was done based on their phasor ratios with respect to the two reference points. Colors represent the phasor density of the areas in the phasor plot. It can be observed that the lifetime distribution is not dependent on fluorescence intensity, as expected. 130

Figure 4:14 Fluorescence lifetime imaging of a cell labeled with a fluorescent protein (FP), behind tissue-like phantom with a thickness up to 1.5 mm [211]. (a) Illustration of the sample structure. A549 cells transfected with mCyrFP1, and unlabeled A549 control cells are placed in two parallel capillaries, which align with left and right halves of the field of view (FOV), respectively. (b, d) Fluorescence intensity images of the two capillaries with (b) no phantom and (d) 1.5 mm phantom. (c, e) Phasor plots of the sample with (c) no phantom and (e) 1.5 mm phantom. (f) Phase lifetimes of transfected cells, control cells and phantom autofluorescence at various phantom thickness values. 137

Figure 4:15 Conceptual illustration of the quanta burst photography method [217]. 138

Figure 4:16 Motion correction performance comparison of conventional and quanta burst photography in a scene under high, medium and low light levels [217]. The scene was captured with a 2,000-frame binary sequence using SwissSPAD2. The image sequence used by conventional burst photography was generated by adding read and quantization noise to the binary image sequence after pile-up correction, in order to emulate the raw images captured by a conventional CMOS camera. The image of conventional single

photography was generated by adding read noise and quantization error to the sum of the raw binary images. The trade-off between blur correction and SNR in conventional photography, which is particularly visible in the low-light scene, is absent in quanta burst imaging. This is enabled by the ability of SwissSPAD2 to reach high frame rates without additional noise..... 139

List of tables

Table 1:1 Evolution of preferred time-resolved imaging method among large-format SPAD imagers.....	23
Table 3:1 Performance comparison of SPAD image sensors with high spatial resolution [117] [81].....	105
Table 4:1 Phase lifetime and standard deviation (in ns) obtained from Figure 4:9. The measured phase lifetimes are slightly smaller than the literature values (Cy3B: 2.8 ns, R6G: 4.08 ns) and the standard deviation scales with $G - 12$	123
Table 4:2 Performance comparison of state-of-the-art FLIM systems with SwissSPAD2.	133

List of acronyms

ADC	analog-to-digital converter
APD	avalanche photodiode
AU	Airy unit
BRAM	block random-access memory
BSI	back side illumination
CAD	computer-aided design
CCD	charge-coupled device
CF	concentration factor
CIS	CMOS image sensor
CMM	center of mass method
CMOS	complementary metal-oxide-semiconductor
CMS	correlated multiple sampling
CNN	convolutional neural network
CPU	central processing unit
CW	continuous-wave
DCR	dark count rate
DDR	double data rate
DFD-FLIM	digital frequency domain FLIM
DFT	discrete Fourier transform
DOF	depth of field
DOM	draining-only modulation
DR	dynamic range
DSM	deep submicron
EMCCD	electron-multiplying charge-coupled device
EWC	electron well capacity
FCS	fluorescence correlation spectroscopy
FD-FLIM	frequency domain FLIM
FF	fill factor
FIFO	first-in first-out
FLIM	fluorescence lifetime imaging microscopy
FOV	field of view
FP	fluorescent protein
FPGA	field-programmable gate array
FRET	Förster resonance energy transfer
FSI	front side illumination
FWHM	full width at half maximum
G-APD	Geiger avalanche photodiode
GCR	global count rate
GPU	graphics processing unit
ICCD	intensified charge-coupled device
IRF	instrument response function
ISM	image scanning microscopy
LCR	local count rate
LED	light-emitting diode
LEFM	lateral electric field charge modulation

LiDAR	light detection and ranging
LSM	least squares method
LUT	lookup table
MCG	multi-collection-gate
MCP	microchannel plate
MLE	maximum likelihood estimation
MOS	metal-oxide-semiconductor
NA	numerical aperture
NIR	near-infrared
NIROT	near-infrared optical tomography
NLLS	nonlinear least-squares
NLOS	non-line-of-sight
NMOS	n-type metal-oxide-semiconductor
PCB	printed circuit board
PDE	photon detection efficiency
PDP	photon detection probability
PET	positron emission tomography
PMOS	p-type metal-oxide-semiconductor
PMT	photomultiplier tube
PPD	pinned photodiode
PRF	pulse repetition frequency
PSF	point spread function
QD	quantum dot
QE	quantum efficiency
QIS	quanta image sensor
RA	resistive anode
RAM	random-access memory
RF	radio-frequency
RLD	rapid lifetime determination
ROI	region of interest
SEPL	single-exponential phasor locus/loci
SNR	signal-to-noise ratio
SNSPD	superconducting nanowire single-photon detector
SPAD	single-photon avalanche diode
SPSD	single-photon synchronous detection
SWIR	short-wave infrared
TAC	time-to-amplitude converter
TCSPC	time-correlated single-photon counting
TDC	time-to-digital converter
ToF	time-of-flight
USB	universal serial bus
WLLS	weighted linear least-squares
XDL	cross delay-line

Chapter 1 Introduction

1.1 Time-resolved imaging and applications

Time-resolved imaging is a technique used in a wide range of scientific and commercial applications. By detecting the time of arrival of photons, or the number of photons in a given time window, this method allows the extraction of various parameters about the scene or target object [1]. By performing these photon operations on multiple pixels, one can build a map of the times of arrival or time-gated counts on photons impinging the image plane [2]. In its early days, time-resolved imaging was performed using complex setups that employed either multiple sensors [3], structured illumination patterns [4], or frequency modulation of the light or the detector [5]. Aside from the complexity of the systems, these methods were also suffering from deficiencies in sensitivity, accuracy and computational power. For these reasons, time-resolved imaging remained confined to scientific applications with limited commercial use. The development of pulsed lasers with sub-ns pulse widths, the emergence of scalable single-photon detectors, and the improvements of speed of operation through integration resulted in an increase in the performance and a decrease in the complexity of time-resolved systems.

Until recently, the only available hardware for time-resolved imaging were single-channel detectors, such as photomultiplier tubes (PMT) [6]. While achieving excellent temporal resolution, down to tens of picoseconds, these detectors had to be scanned across a scene to construct an image. For applications targeting high-spatial-resolution imaging, scanning was not the optimal solution due to its global photon count rate limitation and its demand for higher illumination levels to achieve a given detected photon count rate per pixel. To overcome these problems, the development of time-resolved image sensors that do not require scanning became a necessity. The first generation of these sensors was designed by either adding position sensitivity to a single-channel detector [7] or by modulating the sensitivity of conventional image sensors with image intensifiers [8]. The former approach still did not address the speed problem, whereas the latter was not suitable for applications which required both single-photon sensitivity and high dynamic range. In the time-resolved image sensors available today, both of these issues have been solved: single-photon sensitivity has been achieved in several time-resolved image sensor technologies and photon timing capability is available at the single pixel level. In light of this evolution, time-resolved image sensors are increasingly replacing single-channel detectors in various applications [9, 10].

Time-of-flight (ToF) imaging is one of the major techniques for which time-resolved imaging is essential. In this technique, the travel time of a photon from a source to the scene and back to the detector is used to extract the distance of the scene from the detector. Light detection and ranging (LiDAR), a method that is based on ToF, is currently very popular in a variety of applications including automotive [11], robotics [12] and space technology [13]. A different technique that depends on ToF is non-line-of-sight (NLOS) imaging. In this method, the depth map of a scene outside the line of sight of the sensor is reconstructed from the timing information and location of the photons arriving from the scene to the detector after being reflected by intermediate

objects. NLOS imaging could have a disruptive impact for applications such as robotic vision, medical imaging and automotive safety [14].

The timing of photons can provide valuable information in life sciences as well. In fluorescence lifetime imaging microscopy (FLIM), the statistical distribution of the timing response of fluorescent molecules can allow the extraction of a wealth of information, including oxygen concentration, ion concentration and pH of the environment where a fluorescent molecule is [15, 16, 17, 18, 19, 20]. It can also be useful for the detection of cancerous cells through changes in the fluorescence timing response of certain markers. FLIM-FRET (Förster resonance energy transfer) imaging is a branch of FLIM that allows for the monitoring of interactions between molecules (e.g. proteins) at distances far below the diffraction limit [21]. FRET can also be employed in the detection of conformational changes of certain molecules in a non-destructive way [21]. This technique is a valuable tool for different areas of cancer research, such as anti-cancer drug delivery monitoring [21]. In positron emission tomography (PET), the position of a radioactive tracer inside the human body can be monitored by detecting the time of arrival of two gamma photons that are simultaneously emitted in opposite directions by the annihilation of an electron and a positron. This detection uses luminescence and scintillators (hence PMTs or silicon photomultiplier). Based on this principle, PET provides a comprehensive 3D map of metabolic activity, which is particularly useful in mapping the spread of cancer [22].

In some optical modalities, time-resolved imaging can be used to improve their information content. Near-infrared optical tomography (NIROT) aims to obtain vital information about a biological tissue such as oxygenation by measuring the absorption and scattering characteristics of light at multiple wavelengths [23]. In NIROT, time-resolved imaging offers two advantages. Firstly, it facilitates the separation of absorption and scattering effects, since they influence the time-of-flight distribution in different ways (intensity attenuation and broadening of timing distribution, respectively). Secondly, it allows the analysis of deep tissue by selecting only the late photons in the ToF distribution [24]. Another application that can use temporal information for enhanced capabilities is Raman spectroscopy. This method aims to measure chemical properties of molecules by observing the intensity distribution of Raman scattering in a spectrum of wavelengths [25]. In Raman spectroscopy, time-resolved imaging can be used to suppress unwanted signal, such as fluorescence [26], resulting in an increase of the signal-to-noise ratio (SNR).

Time-resolved imaging is also a required feature for correlation measurements. The correlation function of the inter-arrival times of a sequence of photons can be constructed by exploiting high frame rates. Correlation is used in multiple applications. In fluorescence correlation spectroscopy (FCS), for example, the fluctuations in fluorescence intensity over time are observed to obtain characteristic information, such as molecular concentration and diffusion coefficients [15, 27]. Similarly, cross-correlation measurements can be used for quantum imaging applications, like ghost imaging [28] and quantum image distillation [29].

1.2 FLIM overview

Fluorescence microscopy is an imaging technique that is extensively used in life sciences. Its advantages include the ability to provide high contrast and to selectively label the features of interest. Among the different fluorescence imaging modalities, FLIM recently gained popularity by enabling the analysis of chemical and biochemical properties of proteins with insensitivity to variations in the fluorescence intensity [30, 17, 31].

The main goal of FLIM is to measure the fluorescence lifetime, or lifetimes, of molecules. While each fluorophore has a characteristic lifetime, such lifetime varies with the biochemical environment around the

fluorophore. Fluorescence lifetime is a measure of the average delay between excitation and emission of a fluorophore, which in general is distributed exponentially. The causes of this variation can be understood from the Jablonski diagram, where the excitation and de-excitation processes in a fluorophore are schematically illustrated [32, 33] (Figure 1:1). When a fluorophore absorbs a photon, it is promoted from the ground state (S_0) to the first singlet state (S_1). After the absorption, the molecule retains this extra energy for a certain time before undergoing de-excitation and thus returning to its ground state. For each molecule, there are several alternative mechanisms for de-excitation, each of which occur at a specific rate. Firstly, the energy drops within a state from its initial vibrational level after the excitation to the lowest vibrational level. This non-radiative process, called vibrational relaxation, occurs in picoseconds. After remaining in the lowest vibrational level of the S_1 state for a random duration, the molecule returns to the ground state through photon emission (fluorescence) or non-radiative relaxation. The magnitude of the energy loss during this inter-state relaxation depends on the vibrational level of the molecule at the end of this process. In the case of fluorescence, this variation results in a spectrum of emission wavelengths. Finally, if the energy of the molecule is still at a higher vibrational level, it drops to the ground state through another vibrational relaxation process. Due to the non-radiative energy loss during vibrational relaxation, the wavelength of the emitted photon during fluorescence is mostly higher than the wavelength of the absorbed photon. This effect, named Stokes' shift, is exploited to filter out the excitation light before it reaches the detector.

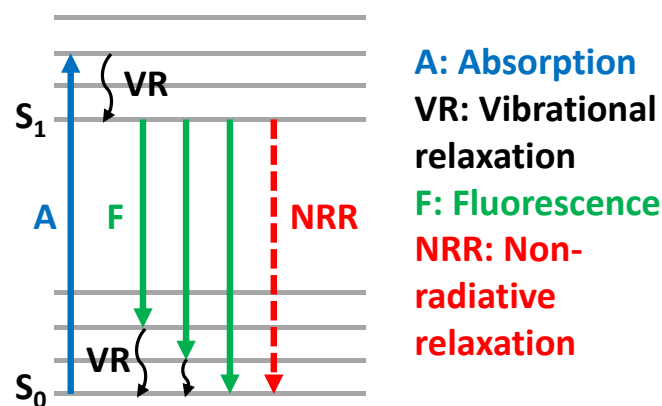


Figure 1:1 Jablonski diagram showing the excitation and de-excitation processes of a fluorescent molecule [32, 33]. Adapted from [32].

The statistical distribution of the delay t between the absorption and the emission results in a single-exponential decay model for the recorded emission intensity, given by

$$I(t) = I_0 e^{-t/\tau}, \quad (1.1)$$

where I_0 is the emission intensity at $t = 0$ and τ is the time constant of the decay. The time constant τ of this equation defines the fluorescence lifetime. The lifetime is determined by the rates of all possible relaxation processes, including all radiative and non-radiative mechanisms. The relation between lifetime and de-excitation rates k_i of multiple processes is expressed as

$$\tau = \left(\sum_{i=1}^N k_i \right)^{-1}, \quad (1.2)$$

where k_i is the decay rate of the relaxation process i , and N is the number of possible relaxation processes. The rates of these processes may be influenced by various chemical and biochemical factors such as oxygen level, pH and ion concentration in the environment, which are key indicators of the cellular functionality. As a

consequence, any variation in these indicators changes the lifetime of fluorophores located in the environment. As τ is independent of I_0 in Equation (1.1), the lifetime is independent of the emission intensity. Additionally, any uncorrelated background signal is time-independent; thus, it appears as a constant baseline superimposed to the decay, which does not affect lifetime. This factor is an important feature of FLIM that facilitates uncorrelated background suppression.

Most biologically relevant fluorophores exhibit multi-exponential decays, in which the emission response is expressed by the sum of multiple exponential decay functions [34]:

$$I(t) = \sum_{i=1}^n I_0(i) e^{-t/\tau_i}. \quad (1.3)$$

A common biological mechanism causing the multiexponential decay responses is the reduction of the lifetime by collisional quenching [35]. Multiexponential behavior can also be observed in a measurement if separate fluorophores with different lifetimes are detected by the same pixel in the image due to their proximity. Unlike in a single-exponential decay, the relative intensity coefficients of the decay components, $I_0(i)$, affect the overall timing characteristics of multiexponential decays. These coefficients can be influenced by various parameters. The first parameter is quantum yield, which indicates the percentage of absorbed photons that are followed by emission. The second parameter is the concentration ratio of the different fluorescent components, which can be different fluorescent molecules or the parts of the same molecule. Irrespective of the physical phenomena behind this behavior, the multiexponential nature of the sample increases the complexity of lifetime analysis.

1.2.1 Förster resonance energy transfer (FRET)

A popular FLIM application is FRET analysis. FRET is a non-radiative energy transfer between two fluorophores that are separated by nanometer-level distance. When two fluorophores, also called “donor” and “acceptor”, with different lifetimes undergo FRET, this adds a new de-excitation path to the donor, the fluorophore which sends the energy, thereby reducing its fluorescence lifetime.

The parameter that determines the change in lifetime due to FRET is the FRET efficiency (E), which indicates the quantum yield of this energy transfer process. The relation between FRET efficiency and donor-acceptor distance (r) is given by

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}, \quad (1.4)$$

where R_0 is the Förster radius, defined as the distance at which the FRET efficiency is 50% [36]. This equation shows that the probability of FRET is inversely proportional to the sixth power of the distance. As a consequence, the lifetime is extremely sensitive to changes in the distance between the fluorophores. Aside from distance, lifetime is also sensitive to the orientation of the fluorophores, due to the dependence of the Förster distance on a parameter called dipole orientation factor appearing in the expression of the Förster radius [37]. The dependence on distance and orientation allows the detection of changes in the geometry of the protein binding sites via FLIM-FRET, which would otherwise not be possible using conventional imaging techniques as the sizes of these proteins are below the diffraction limit.

1.2.2 Confocal vs. widefield FLIM

FLIM can be performed in one of two alternative microscopy geometries: confocal (scanning) and widefield microscopy. The concepts of these two techniques are illustrated in Figure 1:2 [38]. In confocal microscopy (Figure 1:2(a)), the excitation light is concentrated on a single spot in the sample and the emission light generated by the excited spot is collected by a single-channel detector. An image of an area in the sample is acquired by scanning the position of the excitation beam across the sample and collecting the emission signal of each point by the detector in a sequential fashion. To filter the out-of-focus emission light, a pinhole is located on the focal point of the tube lens in front of the detector. The focusing of the excitation light on a single point and the presence of the pinhole result in low depth of field (DOF), which is defined as the distance between the nearest and farthest depths at which the image of an object on a given image plane is “in focus”.

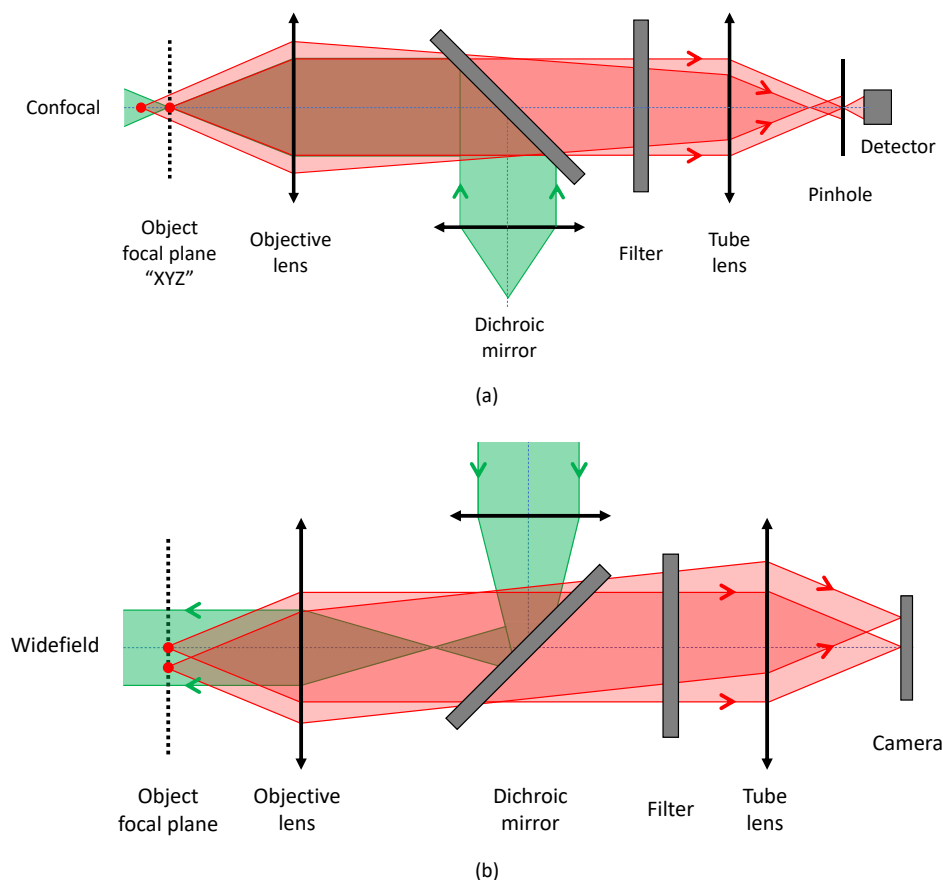


Figure 1:2 Schematic of (a) confocal and (b) widefield microscopy. Adapted from [38]. The confocal microscopy setup illustrated in (a) is a sample-scanning one, in which the sample is scanned along the x,y and z axes. The red dot on the object focal plane is the point in the sample that is captured by the detector.

The application of confocal microscopy in FLIM was presented for the first time in the 1980s by making minor modifications to a standard confocal microscope setup [39]. A standard confocal FLIM setup uses a photomultiplier tube (PMT) or a single-photon avalanche diode (SPAD) as the detector, although hybrid detectors [40] and superconducting nanowire single-photon detectors (SNSPDs) [41] are also gaining popularity for improved photon count rate and timing resolution. For timing of the detected photons, they employ either a time-to-amplitude-converter (TAC) or a time-to-digital converter (TDC). The timing resolution of PMTs and SNSPDs can reach around 30 ps [35] and 3 ps [41], respectively. While the timing performance of TDCs depends on various design constraints, they also can achieve down to less than 1 ps resolution [42, 43].

Confocal microscopy offers several advantages for FLIM. The need for only one single-channel detector as opposed to an imager allows the use of detectors with very high timing resolution; overall system jitter below 100 ps is widely available in these setups [44]. Also, its low DOF enables the optical sectioning of the sample and suppresses out-of-focus background noise. On the other hand, confocal microscopy introduces a trade-off between speed (frame rate), excitation intensity and spatial resolution in the final image. Firstly, the currently available scanning mechanisms have fundamental limits in either the speed or the scanning area [18]. Secondly, the acquisition of sufficient statistics from a pixel becomes more difficult as the dwell time decreases. This difficulty arises from the maximum allowed excitation intensity without significant photobleaching [7, 45] and the maximum photon count rate of the detectors and timing electronics [20]. When the dead time of the system is exceeded, for instance in the cases where multiple photons are emitted per pulse, the underestimation of the photon counts at large delays lowers the accuracy of the measured lifetime, an effect also called “pile-up” [46]. Despite this trade-off, in general, confocal microscopy is the technique that offers the best SNR and accuracy for experiments without strict frame rate or spatial resolution requirements.

In some applications of FLIM, however, the primary goal is to analyze a large field of view, and fast data acquisition is more important than achieving the highest accuracy and precision. For instance, when FLIM is used for blood oxygenation detection [47], the primary requirement is to distinguish two lifetimes with sufficient contrast at video-rate (25-30 fps). Additionally, in some *in vivo* diagnostic applications, such as endoscopic FLIM, it is highly preferred to use a detector that does not require scanning due to physical limitations [48]. Confocal FLIM does not offer the optimal solution for these applications. Finally, particularly in imaging with high spatial resolution, scanning requires high excitation power due to the low ratio between the dwell time of a pixel and the frame acquisition duration. This results in an elevated risk of photobleaching for some biological samples [45]. To address the shortcomings of confocal FLIM, widefield FLIM was introduced as an alternative choice. Widefield microscopy, illustrated in Figure 1:2(b), illuminates the entire sample area simultaneously, allowing the detection of the emitted signal in the entire field of view in parallel. In this approach, the excitation beam is collimated by the objective lens and transmitted to the sample with normal incidence uniformly. This collimation is achieved by focusing the excitation beam on the back focal plane of the objective lens. The emitted light from the sample, which is located on the focal plane of the objective, is collimated by the objective lens and refocused to the image sensor which is located on the focal plane of the tube lens.

Widefield FLIM is generally more compatible with high-speed data acquisition at high spatial resolution than confocal FLIM. When implemented in combination with an image sensor, widefield FLIM enables one to achieve a global photon count rate that scales with the number of pixels, to a certain degree. The bottleneck is the data transfer rate of the sensor. The global photon count rate, GCR of a widefield microscope is expressed as a function of local (pixel) photon count rate LCR and number of pixels N as

$$GCR = LCR \times N, \quad (1.5)$$

whereas the GCR of a confocal microscope is given by

$$GCR = LCR. \quad (1.6)$$

According to Equations (1.5) and (1.6), for a pair of widefield and confocal systems whose LCR are limited by the dead time of the photodiode, the GCR of the widefield system is significantly higher at high spatial resolution. Conversely, when the GCR of the two systems are equal due to a common throughput limitation of the readout network, the widefield system can reach the maximum GCR at a substantially lower LCR . This

capability brings two advantages. Firstly, it reduces the required excitation light power, resulting in lower risk of photobleaching [45]. Secondly, it makes the widefield system less susceptible to afterpulsing, since lower LCR corresponds to larger photon inter-arrival times.

The most important characteristic disadvantage of widefield microscopy is the relatively inferior DOF. In microscopy, lower DOF is often preferred as it allows the suppression of the signal coming from outside the target object plane, thereby allowing z-sectioning with a finer resolution. The primary factor that determines the DOF of a microscope is its axial (z-axis) resolution. Therefore, a brief discussion of lateral and axial resolution of a microscope is necessary to understand the difference of DOF in confocal and widefield microscopes.

The lateral and axial resolution of a microscope are mainly determined by the properties of its point spread function (PSF), which is defined as the response of the microscope to a point light source. The two main factors that determine PSF in a microscope are diffraction and aberration. Diffraction is defined as the bending of light waves when they travel through a narrow aperture, and aberration is defined as the spread of a focused light source by a lens due to its properties, such as curvature and refractive index. The microscopes with diffraction-limited objectives produce a PSF in the form of an Airy pattern.

Although various criteria are used to define resolution, one of the most common definitions is Rayleigh's criterion, which considers two points as "resolved" if their minimum distance is equal to the distance between the center and the first destructive interference band in the Airy pattern [49]. For widefield microscopes, the Rayleigh criterion for lateral and axial resolution is given by [50]

$$R_{xy} = 0.61 \frac{\lambda}{NA}, \quad (1.7)$$

and

$$R_z = 2 \frac{\lambda n}{NA^2}, \quad (1.8)$$

respectively, where λ is the light wavelength, n is the refractive index of the medium and NA is the numerical aperture. NA is determined by the maximum angle of the light cone that the objective lens can accept from a point object, and is expressed as

$$NA = n \sin \theta, \quad (1.9)$$

where n is the refractive index of the medium where the lens is placed ($n = 1$ for air, $n = 1.3$ for water and $n = 1.51$ for immersion oil), and θ is the maximum half-angle of the light cone.

In confocal microscopes, the width of the PSF is smaller than a widefield microscope, both laterally and axially. The reason for this is that unlike widefield microscopy where the sample is illuminated with uniformly distributed light, the illumination pattern of the sample is already in the form of a PSF in a confocal microscope. Therefore, after the emission light passes through the objective, the resulting PSF at the image plane is equivalent to the product of the PSF of the objective with itself at each x and y point, expressed as

$$PSF_{confocal}(x,y) = PSF_{obj}(x,y) \times PSF_{obj}(x,y), \quad (1.10)$$

which results in a narrower shape compared to widefield microscopy, as shown in Figure 1:3 [51].

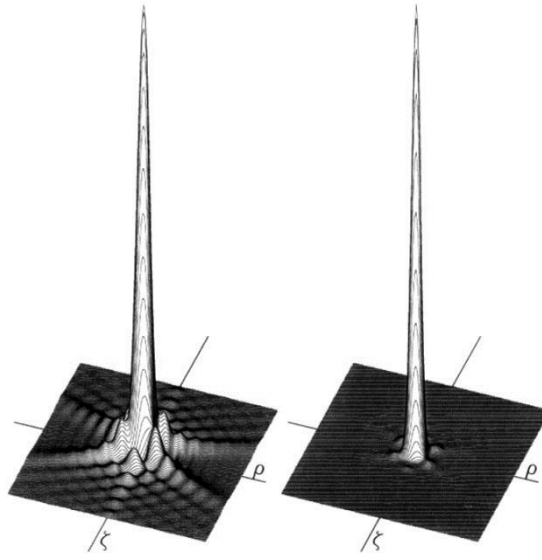


Figure 1:3 Point spread function (PSF) of a widefield (left) and confocal (right) microscope [51].

Considering the narrower PSF, the Rayleigh criterion for confocal microscopes for lateral and axial resolution can be rewritten as

$$R_{xy} = 0.44 \frac{\lambda}{NA}, \quad (1.11)$$

and

$$R_z = 1.5 \frac{\lambda n}{NA^2}, \quad (1.12)$$

respectively [52], which correspond to 28% improvement in lateral resolution and 25% improvement in axial resolution compared to widefield microscopy.

In addition to having wider DOF, a second potential disadvantage of widefield FLIM is an overall sensitivity reduction due to the limited fill factor (ratio of photosensitive to total pixel area) for some of the detector technologies, including SPAD arrays. As in widefield microscopy if the image of the entire scene is projected on a detector with multiple pixels, any photo-insensitive part of the pixel area of the detector leads to missed photons, thereby reducing the fill factor unless compensated by employing microlenses for example. On the other hand, in confocal microscopy, the entire emission signal can be focused on the photon-sensitive part of the single-pixel detector, enabling 100% fill factor. Furthermore, in confocal microscopy, a wider spectrum of sensitivity can be reached by replacing silicon detectors with other materials. Due to their incompatibility with the monolithic integration of circuit electronics, these materials are not suitable for widefield microscopy unless 3D integrated circuits are used.

The third potential disadvantage of widefield FLIM is related to timing resolution, which is defined as the timing precision of a single photon (not to be confused with the timing precision of the resulting lifetime measurement). In confocal FLIM, the time-stamping of a photon is often performed by a TDC, which can be implemented as a discrete component without significant area, power and bit depth constraints. On the other hand, time-resolved imaging in widefield systems is performed either by gating, which characteristically has lower

timing resolution than TDCs, or by TDCs with stricter design constraints. If the TDCs are designed with comparable constraints to their confocal counterparts, it can lead to a sacrifice in maximum GCR due to the high throughput demand, or a sacrifice in fill factor. The overall timing resolution of state-of-the-art widefield and confocal FLIM systems with their respective trade-offs are discussed in depth with examples in Chapter 4, Subsection 4.2.7.

Over the years, however, the characteristic limitations of both approaches have been challenged and the performance gap between the two methods is *de facto* getting narrower. For instance, various optical sectioning techniques were developed to overcome the depth of field inferiority of widefield microscopy. The methods that were successfully demonstrated include structured illumination [53], spinning disk [54] and light sheet microscopy [55]. Two-photon excitation, a depth of field enhancement technique which is conventionally used in confocal microscopy, was also applied to widefield FLIM in [56]. The timing resolution disadvantage of widefield multi-channel detection systems has also been mitigated in recent works. This development is mostly enabled by large-format SPAD sensors with in-pixel TDCs rivaling the performance of the discrete TDC blocks of confocal systems [10, 57]. On the other hand, confocal systems are improving as well. The global photon count rates of confocal TCSPC systems have increased owing to the improvements in dead time. With optimizations in TDC design and the development of hybrid photodetectors, the overall dead time of TCSPC systems was reduced to the nanosecond level [58]. This achievement also enabled the detection of multiple photons per excitation pulse without large measurement accuracy errors due to pile-up. In addition, the maximum global count rate was extended by increasing the number of detection channels.

An alternative microscopy approach consists in increasing the number of parallel illumination points in confocal microscopy. This method, called multifocal FLIM, scans a large field of view faster than the conventional confocal approach without sacrificing its characteristic performance advantages. A multifocal FLIM system generates a 2D array of focused laser beams which is scanned across the sample with a galvanometer to generate high spatial resolution. This approach combines the advantages of confocal microscopy in depth of field and widefield microscopy in speed. While displaying promising performance improvements in the version with 8×8 beamlet pattern [59], the scaling of this method has its own challenges. In order to avoid sacrificing speed and increasing the photobleaching rate through high excitation power, the ratio between the number of pixels and the number of laser beams must be kept as low as possible. However, a high number of laser spots increases the spot non-uniformity, complicating the calibration and correction processes [60]. In addition, the simultaneous detection of all laser spots requires a large-format camera, which leads to sacrifices in timing resolution, like in widefield FLIM. For these reasons, the spatial resolution of this technique remains limited to this date. Furthermore, this method suffers from additional practical implementation issues such as alignment difficulties and crosstalk between spots [15].

A second alternative microscopy approach is called image scanning microscopy (ISM), which was first proposed as a way to achieve super-resolution in confocal microscopy [61], and recently was applied to FLIM [62]. In standard confocal microscopy, when the pinhole diameter is reduced below 1 Airy unit (AU), the lateral resolution can exceed the diffraction limit of the particular wavelength, i.e. super-resolution can be achieved [63]. However, this comes at a cost of SNR loss, as the pinhole rejects a part of the desired signal by only partially transmitting the Airy disk. In ISM, instead of being filtered out by a pinhole, the Airy disk is projected on a group of pixels in an image sensor, which is placed on the focal plane corresponding to the target object z -axis. For each scanned point in the object, the responses of different detector pixels are assigned to different pixels in the reconstructed image, in such a way that the images of multiple point objects partially overlap

[61]. As a result, ISM combines the SNR of a 1 AU pinhole diameter and the lateral resolution of a <1 AU pinhole diameter, thereby eliminating the aforementioned trade-offs in standard confocal microscopy across the x , y and z axes [64].

Despite the lateral resolution advantages compared to standard confocal microscopy, the ISM method still suffers from the characteristic downsides of single-point scanning. Like standard confocal microscopy, the global count rate of an ISM system is defined by Equation (1.6); it is not influenced by the number of pixels in the frame (i.e. the number of scanned points in the sample). Therefore, it becomes more challenging to compete with global photon count rates of widefield systems as the spatial resolution increases, even considering a potential improvement in the local photon count rate resulting from a decrease in the overall dead time due to the use of multiple pixels in the ISM detector.

1.3 Phasor approach in FLIM

In its early years, FLIM has generally suffered from low data processing speed. While the data acquisition, i.e. the detection of sufficient photons to estimate lifetime with acceptable precision, could reach video-rate speeds, the speed of data processing, i.e. the estimation of the lifetime from the photon time-of-arrival distribution, could often not keep up. To make this technique valuable for applications demanding real-time refresh rate, such as dynamic imaging of fast moving cells [18], dynamic changes in ion concentrations [65, 66] and interactions between proteins [67], speed improvement was a key requirement. For this reason, the development of fast FLIM systems has become an active field of research [18]. Concerning data acquisition, limitations in photon count rate, sensitivity and timing resolution determined the minimum duration needed to acquire sufficient statistics for reliable lifetime determination. Another key aspect was related to the lifetime calculation procedure starting from the acquired raw data. In early FLIM demonstrations, the least squares method (LSM) was predominantly used to estimate lifetime [68]. This fitting-based method compares the acquired decay profile with theoretical exponential models, and finds the best fitting model by iteratively tuning the decay parameters. Alternatively, the maximum likelihood estimation method (MLE), which uses a different algorithm to perform fitting, was proposed as an alternative to LSM [69]. Although MLE has outperformed LSM under certain experimental settings, it does not offer a significant improvement in computational complexity. In general, these methods are computationally expensive and unsuitable for high refresh rates, particularly in multi-exponential decay; furthermore, they require *a priori* information about the decay models of the samples, which is also undesirable.

Solutions to overcome this limitation can be classified in two categories. The first category focuses on improving the computational power of the data processing hardware. For instance, in [70], high speed FLIM was demonstrated by using GPU acceleration. On the other hand, the solutions in the second category aim to develop alternative lifetime estimation algorithms with lower computational complexity. The novel lifetime estimation approaches that have become popular include Laguerre expansion techniques [71], rapid lifetime determination (RLD) [72, 73] and center of mass method (CMM) [74, 75]. However, these methods achieve higher speed at the expense of penalties in other performance parameters, such as reduced photon economy, which is a measure of precision of a FLIM system for a given number of detected photons, for RLD and requirement of short gate widths in CMM. In recent years, deep learning-based FLIM has emerged as another viable option for fast FLIM [76]. This technique overcomes the low data processing speed by replacing the fitting process with a novel method based on a convolutional neural network (CNN). The main relative disadvantage

of this approach is the dependence of the lifetime estimation accuracy on the design of the training process of the neural network.

The phasor method has emerged as an alternative for fast lifetime analysis, which performs the computations using basic arithmetic operations. It also differentiates from some of the existing lifetime analysis methods by operating without *a priori* information about the exponential model of the decay, which is also referred as being model agnostic [77]. In addition, it offers a graphical representation of the lifetime distribution in a sample. This representation is particularly useful for applications where observing the general characteristics of the sample is more important than finding the lifetime itself, such as in cancer detection [78]. Today, the phasor method is offered in several commercial FLIM systems for data analysis [79, 80].

A phasor is a vector in the complex plane with a magnitude m and phase φ . These two parameters represent the relation between the characteristics of a sinusoidal excitation and emission signal in frequency-domain FLIM. In the complex plane, the real and imaginary parts of a phasor are represented by g and s , respectively, as shown in Figure 1:4(a) [81]. In a typical time-correlated single-photon counting (TCSPC) system, the phasor of a decay is calculated by

$$z = g + is = me^{i\varphi} = \frac{\sum_{k=1}^G I_k e^{i2\pi f t_k}}{\sum_{k=1}^G I_k}, \quad (1.13)$$

where G is the number of bins, I_k is the intensity of the bin k , f is the phasor frequency, and t_k is the time delay representing the particular bin, also called the nanotime of the bin. The coordinates of the average decay phasor, g_{avg} and s_{avg} , are equal to the average of the coordinates of each bin, g_k and s_k , weighted by the photon count of the corresponding bin, I_k , as shown in

$$g_{avg} = \frac{\sum_{k=1}^G I_k g_k}{\sum_{k=1}^G I_k} = \frac{\sum_{k=1}^G I_k \cos(2\pi f t_k)}{\sum_{k=1}^G I_k} \quad (1.14)$$

and

$$s_{avg} = \frac{\sum_{k=1}^G I_k s_k}{\sum_{k=1}^G I_k} = \frac{\sum_{k=1}^G I_k \sin(2\pi f t_k)}{\sum_{k=1}^G I_k}. \quad (1.15)$$

For single exponential decays, g_{avg} and s_{avg} are located on a semicircle in the first quadrant on the complex plane, called the universal semicircle. On the universal semicircle, the phasors of longer lifetimes are closer to the origin than the shorter lifetimes. The phase lifetime of a single-exponential decay is described by [7, 77]

$$\tau = \frac{1}{2\pi f} \left(\frac{s_{avg}}{g_{avg}} \right). \quad (1.16)$$

Unlike conventional fitting algorithms, this calculation only consists of simple arithmetic operations, thereby making it suitable for fast data analysis. The phasors of multi-exponential decays, on the other hand, are located inside the universal semicircle, between the phasors of its single-exponential components. In the case of a double-exponential lifetime, the phasor is on the line segment between its two components, each of which is on the semicircle. The relative concentrations of the single-exponential phasor components, also called the phasor ratio (r), can be calculated from the ratio of the lengths of the line segments between the measured phasor and its two components, given by

$$r_1 = \frac{d_2}{d_1 + d_2}, \quad (1.17)$$

where r_1 is the phasor ratio of component 1, and d_1 and d_2 are the lengths of the line segments between the mixture phasor and the components 1 and 2, respectively. The relation between the phasors of a mixture and its components is illustrated in Figure 1:4(b) [81]. Multicomponent analysis can also be performed with multiple harmonics [82].

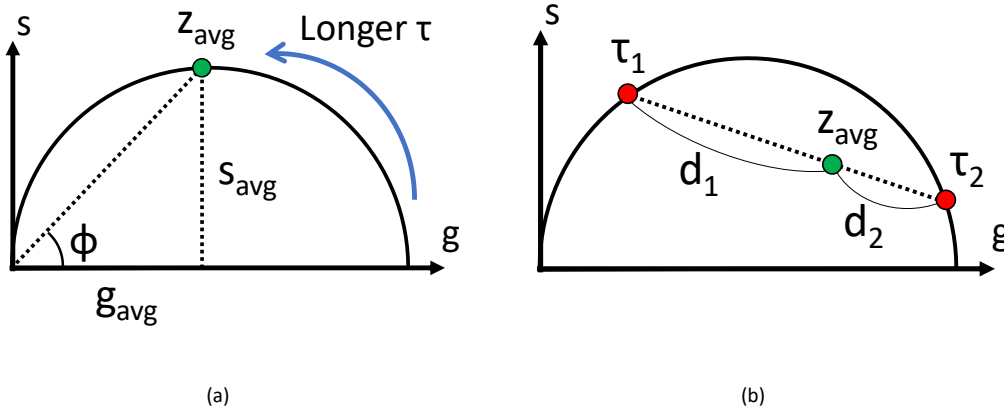


Figure 1:4 (a) Conceptual illustration of phasor analysis [81]. Z_{avg} , the phasor of a fluorescence decay, is on the universal semicircle if the decay is single-exponential. As lifetime increases, a single-exponential decay phasor approaches the origin along the semicircle. (b) The phasor plot of a mixture with two single-exponential components [81]. The phasor ratio can be calculated from d_1 and d_2 , the distances between the phasor of its mixture and the phasors of its components.

From Equation (1.13), it can be seen that the phasor calculation for a fluorescence emission function acquired in the time domain (with pulsed excitation) is equivalent to the normalized discrete Fourier transform (DFT) of the sampled function, where each bin represents a sampling point. The sampling rate f_s and measurement window D in this DFT process are equal to $f_s = 1/(t_{k+1} - t_k)$ and $D = G \times (t_{k+1} - t_k)$, respectively. The resolution f_{res} of the phasor frequency and the maximum permitted bandwidth B of the emission function are equal to $f_{res} = 1/D$ and $B = 0.5f_s$, corresponding to the fundamental frequency of the underlying Fourier series and to the Nyquist frequency, respectively.

The phasor plot is constructed by plotting the points (g_{avg}, s_{avg}) for all pixels or ROIs across the array. Furthermore, if required by the application, a color-coded 2D phase lifetime map of the sample is also generated to complement the phasor plot. The phase lifetime is calculated using Equation (1.16). Using the pixel coordinates associated with each computed phasor, pixels in the image corresponding to specific phasor values can be color-coded and displayed in a “phasor” map. This feature allows the identification of the areas in the image, which share similar biochemical characteristics. Conversely, one can plot the phasor distribution of only a specific ROI. This feature of the phasor method is demonstrated by Figure 1:5 [77]. In this figure, the phasors of two different images were combined in a phasor plot in (A). The first image (B-D) contains a CHO-K1 cell expressing paxillin-EGFP in a collagen matrix, whereas the second image (E-G) contains only a collagen matrix. The phasors plotted in (A) have three fluorescence contributions: EGFP (region 1), collagen fluorescence (region 2), and background autofluorescence (region 3). The phasors in region 1 and region 2 are highlighted in images (C, F) and (D, G), respectively. As region 1 represents EGFP which is absent in the second image, no pixels are highlighted in (F). It can also be observed that the highlighted areas in (C) and (D) closely match with the cell and collagen matrix, respectively. These results clearly demonstrate the possibility to

highlight different biological structures in the image with distinct fluorescence decay characteristics from the phasor plot.

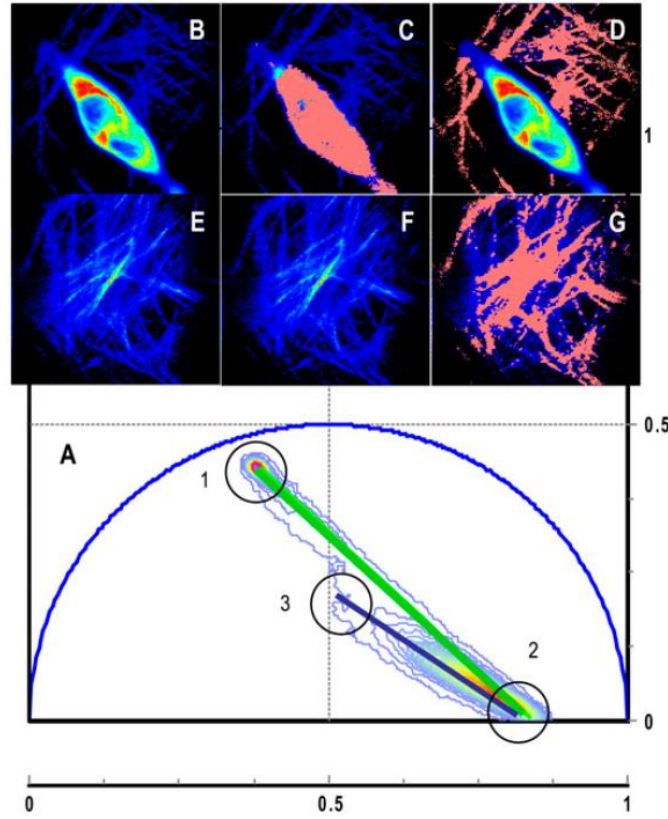


Figure 1:5 Microscopic images and phasor representation of CHO-K1 cells transfected with paxillin EGFP in a collagen matrix [77]. The phasor plot in (A) contains the phasors of two different image with the cell and the collagen matrix (B-D), and only collagen matrix (E-G). The three fluorescence contributions are EGFP (region 1 in the phasor plot), collagen (region 2) and background autofluorescence (region 3). The pixels whose phasors are in region 1 and region 2 are highlighted in (C, F) and (D, G), respectively. From the absence of phasors of region 1 in (F) and the clear correlation between the highlighted regions in (D, G) and the structural differences in the image, it can be concluded that the phasor representation can be used to distinguish different parts of an image with distinct decay characteristics.

An important step in the phasor calculation is IRF deconvolution. This step is required to compensate for the distortions in the fluorescence response function during its sampling by the detector, due to the shape of the gate window and its delay from the laser pulse. In the time domain, the measured emission function is equal to the convolution between the real emission function and the IRF, which is expressed by

$$I_{meas}(t) = I_{real}(t) * I_{IRF}(t). \quad (1.18)$$

This conversion can be translated, when working in the phasor plane, as

$$Z_{meas} = Z_{real}Z_{IRF}, \quad (1.19)$$

where z represents a phasor, which can be expressed as

$$z = me^{i\varphi}, \quad (1.20)$$

where m and φ are the modulation and the phase of the phasor, respectively. Combining Equations (1.19) and (1.20), the relation between the real and measured m and φ is given by [7, 77]

$$m_{meas}(f) = m_{real}(f) \times m_{IRF}(f), \quad (1.21)$$

and

$$\varphi_{meas}(f) = \varphi_{real}(f) + \varphi_{IRF}(f). \quad (1.22)$$

To find the phasor coordinates of the IRF, a calibration measurement is performed on a sample known to be characterized by a single-exponential decay with a known lifetime, denoted as τ_{cal} . In this calibration process, $m_{real}(f)$ and $\varphi_{real}(f)$ define the expected phasor location of the calibration sample, whereas $m_{meas}(f)$ and $\varphi_{meas}(f)$ define its measured location. $m_{real}(f)$ and $\varphi_{real}(f)$ are calculated from τ_{cal} from the equations below [90]:

$$m_{real}(f) = \frac{1}{\sqrt{1 + (2\pi f \tau_{cal})^2}} \quad (1.23)$$

and

$$\varphi_{real}(f) = \tan^{-1}(2\pi f \tau_{cal}). \quad (1.24)$$

Using $m_{real}(f)$ and $\varphi_{real}(f)$ of the calibration sample calculated from Equations (1.23) and (1.24), and the measured values $m_{meas}(f)$ and $\varphi_{meas}(f)$, we can calculate $m_{IRF}(f)$ and $\varphi_{IRF}(f)$ from Equations (1.21) and (1.22). The calculation of $m_{IRF}(f)$ and $\varphi_{IRF}(f)$ requires division and subtraction operations, respectively. It must be noted that since the IRF phasor parameters are functions of f , the calibration process must be repeated for each target phasor frequency. Due to the considerable variation of the IRF between the detector pixels, m_{IRF} and φ_{IRF} are characterized separately for each pixel to maximize phase lifetime determination accuracy.

Finally, IRF deconvolution is performed during the phasor analysis stage of each sample of interest by using the $m_{IRF}(f)$ and $\varphi_{IRF}(f)$ values which are stored after the calibration process. This deconvolution procedure is described as follows. Firstly, since the measured phasors are expressed in g and s coordinates, they are converted to m_{meas} and φ_{meas} using the equations below [7]:

$$m = \sqrt{g^2 + s^2} \quad (1.25)$$

and

$$\varphi = \tan^{-1}\left(\frac{s}{g}\right) \quad (1.26)$$

Then, using these m and φ values as m_{meas} and φ_{meas} in Equations (1.21) and (1.22), together with the previously stored $m_{IRF}(f)$ and $\varphi_{IRF}(f)$ values, m_{real} and φ_{real} can be calculated for the sample of interest.

Since the advent of phasor-FLIM in 2008 [77, 83, 84], recent efforts have been focused on adapting various fast FLIM techniques to the phasor approach. Two of these techniques are undersampling and truncation. Undersampling refers to the practice of reducing f_s below the Nyquist rate, i.e. twice the bandwidth of the sampled signal. In phasor-FLIM, this corresponds to increasing the delay between adjacent bins/gates (also called “gate step” in the rest of the thesis). The main motivation for undersampling in FLIM is to decrease the data acquisition time of one FLIM image, with a consequent increase of refresh rate. The distortions in the Fourier transform due to undersampling, which are caused by aliasing, result in the representation of the

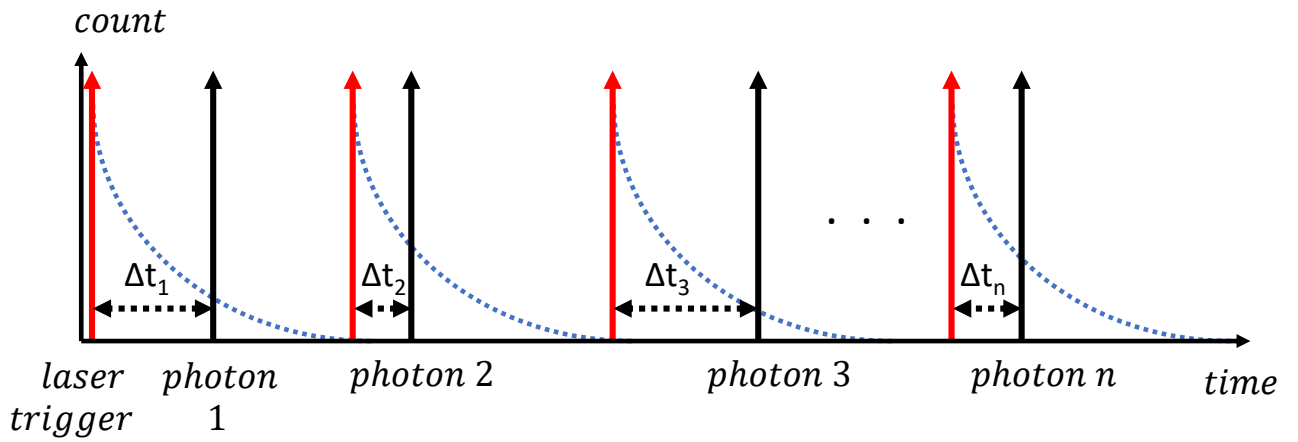
phasors in modified single-exponential phasor loci (SEPL) [86]. In [85], the phasor approach was adapted to undersampled discrete functions via the modeling of the SEPL. An alternative correction method for the lifetime estimation of discrete functions is based on rectangular or trapezoidal interpolation, which is discussed in [91]. The second technique, known as truncation, is the limitation of the measurement window to the extent that a significant part of the function area is not sampled. The main motivation for this technique is to increase the photon count rate by minimizing the variation of the intensity between the captured gate positions (note that the brightest recorded gate determines the maximum count rate of the detector). The effects of truncation and optimization techniques of the phasor analysis to truncated decays have been discussed in [86]. A more detailed discussion of undersampling and truncation with examples from the corresponding works is also available in Chapter 2.

The phasor method, by definition, entails the representation of the recorded fluorescence emission signal in the frequency domain, and the expression of its phase and normalized modulation at a user-selected frequency (also called phasor frequency). This technique can however be applied to a fluorescence emission function that is generated after excitation with either a pulsed or sinusoidally modulated light source. In the literature, the use of pulsed excitation and the sampling of the response with equally spaced time gates is generally considered as a “time-domain” approach, regardless of the subsequent lifetime analysis method [87]. Another popular opinion is to apply the term “frequency-domain FLIM” to all lifetime estimation methods based on finding the lifetime from φ or m , independent of the use of pulsed or sinusoidally modulated excitation [88, 89]. Regardless of the classification method that is adopted, the phasor method can be universally applied to both time-domain and frequency-domain data acquisition methods [77].

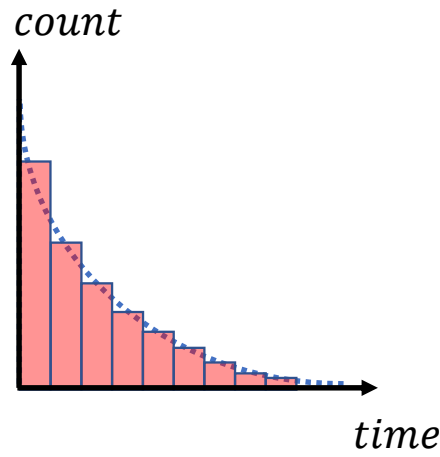
1.3.1 Time-correlated single-photon counting (TCSPC)

Time-correlated single-photon counting (TCSPC) is considered the gold standard of FLIM, and it is the primary method used for reference in the performance evaluation of novel methods. The main reason for this is that it yields the highest SNR among the available FLIM methods [92].

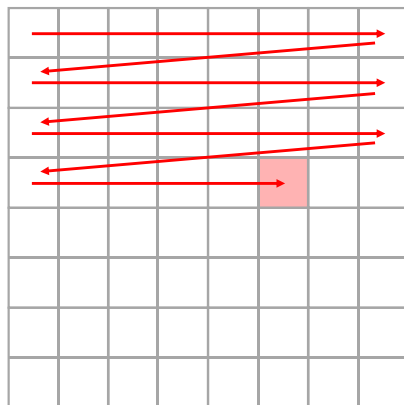
The description of the standard TCSPC operation is as follows, as illustrated in Figure 1:6(a) and (b). The fluorescent sample is illuminated by a short light pulse, typically with a width of tens of picoseconds for nanosecond-range lifetimes. For every photon emitted from the sample, the detector records the time delay (Δt) between the excitation pulse and the emitted photon (Figure 1:6(a)). A histogram showing the statistical distribution of all detected delay values is constructed after the data acquisition (Figure 1:6(b)). Subsequently, this histogram can be the input data for lifetime analysis, which can be performed using one of the available lifetime estimation methods listed in the previous section. Phasor-based TCSPC systems are also commercially available [93].



(a)

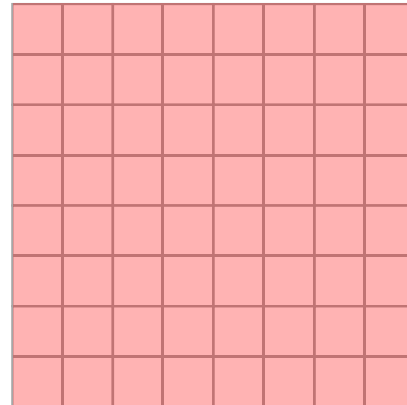


(b)



$$T_{acq} = N \times t_{acq}$$

(c)



$$T_{acq} = t_{acq}$$

(d)

Figure 1:6 (a, b) Conceptual illustration of TCSPC. (a) The sample is excited multiple times by a laser pulse, and the time of arrival of each detected emitted photon is recorded. (b) Subsequently, a histogram of the time of arrival of all detected photons is generated. The shape of the histogram represents the properties of the measured fluorescence decay function. (c, d) Data acquisition schemes of (c) confocal and (d) widefield TCSPC for a FOV captured by 8×8 pixels. Each gray square which represents a pixel in the scene is illuminated in a raster-scanning configuration in (c), and in parallel in (d). The illumination, which in reality covers the entire pixel in (c) and FOV in (d), is represented by circular red dots that only cover a part of the corresponding area, for illustrative purposes.

The illumination schemes of the FOV for confocal and widefield TCSPC are illustrated in Figure 1:6(c) and (d), respectively. In confocal TCSPC, the pixel areas in a 2D FOV are raster-scanned, and in widefield TCSPC, the entire FOV is illuminated in parallel. The total acquisition time of a frame, T_{acq} , is the product of the pixel acquisition time t_{acq} and the number of pixels N in confocal TCSPC, whereas it is equal to t_{acq} in widefield FLIM. The characteristics of confocal and widefield TCSPC that are discussed in this paragraph are however not exclusive to TCSPC; therefore, they can also be applied to all time-resolved imaging techniques.

The dead time and timing resolution of both the photodiode and the timing electronics are essential parameters for TCSPC performance, as well as the characteristics of the laser pulse. The demanding hardware requirements limited the early generation TCSPC solutions to single-channel systems, due to the general trend of performance degradation of these components in multi-channel implementations. Therefore, the application of TCSPC to FLIM only became possible with the introduction of confocal microscopes [39]. In state-of-the-art scanning-based TCSPC FLIM systems, the timing resolution of each detected photon can be as low as a few picoseconds [18]. However, due to the trade-off between spatial resolution and speed in confocal microscopy, which was discussed in Section 1.2, video-rate TCSPC FLIM for samples susceptible to photobleaching has remained challenging.

The restriction of the TCSPC method to confocal microscopy has recently been challenged by the widefield TCSPC concept. The different widefield TCSPC methods, which are reviewed in [92], can be summarized in two main categories. The architectures in the first category use microchannel plates (MCP) to achieve single-photon sensitivity by multiplying the photoelectrons before the readout stage. There are two main options for the readout stage itself: a standard CMOS or CCD camera, or a position-sensitive anode. The former option is unable to analyze nanosecond-range lifetimes due to its MHz frame rates, and the latter suffers from global photon count rate limitations due to being single-channel. The second category consists of natively single-photon counting detectors such as SPADs and superconducting detectors. The SPAD detectors are the most suitable candidates for mitigating the global count rate limitation as they support the integration of timing electronics. While the first generation of SPAD arrays with integrated TDCs only had limited number of channels or very low fill factor [10] due to the required area for a high-performance TDC, technology scaling in CMOS processes recently culminated in large-format SPAD imagers with in-pixel [57] and shared [94, 95] TDCs and reasonably high native fill factor.

1.3.2 Analog frequency domain FLIM

As an alternative approach to TCSPC, which is a time-domain method, fluorescence lifetime can also be estimated in frequency domain. In this approach, instead of a pulsed light source, the sample is excited with continuous wave light that is modulated with a sinusoidal signal, as shown in Figure 1:7 [96]. The response of the fluorophore causes a change in the phase and the modulation of the signal. The ratio between the emission and excitation modulation values, the demodulation factor, is also called “modulation” in the literature [35]. Unless stated otherwise, in this thesis, the term “modulation (m)” refers to the demodulation factor, and the term “phase (φ)” refers to the phase difference of the emission and excitation signals. These two parameters are characteristic properties of the fluorescent response of a particular sample; the lifetime of a single-exponential decay can be derived from these two parameters.

An essential hardware feature required for this approach is the sinusoidal modulation of the light source. The modulation of the CW light amplitude has been used since the introduction of early phase fluorimeters, achieving the detection of down to sub-ns lifetimes [97, 98]. In the early implementations of these devices,

the Debye-Sears modulators were used for the modulation of the excitation light. The main drawback of these devices was that they could only operate at a limited range of modulation frequencies, which curtailed the use of the devices for samples with complex decay characteristics [35]. While a significant improvement in frequency and light modulation has been achieved in the following decades, the modulation hardware used in modern times still employs radio-frequency (RF) components [99]. Due to the integration challenges of RF blocks into modern integrated systems and the high output power requirements of the power amplifier to control the light source [99], the development of more compact solutions for these systems remains difficult.

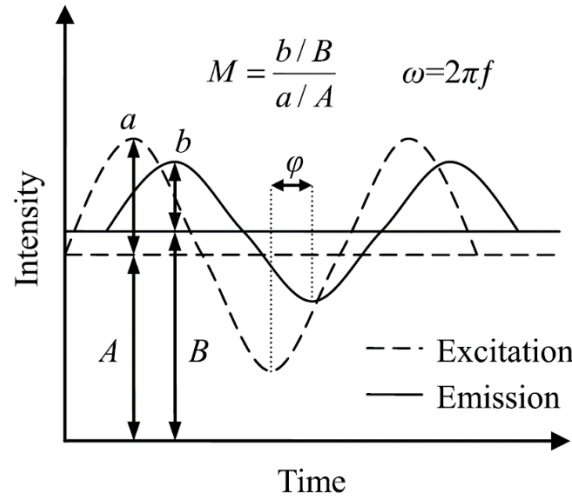


Figure 1:7 Principles of analog frequency-domain FLIM [96].

A typical analog FD-FLIM setup employs two identical PMTs as detectors. The incident light is split into the reference PMT and the sample PMT. The phase and the modulation of the sample PMT are measured sequentially for scattering and fluorescent solutions. The role of the scattering solution is to characterize the instrument response for calibration during the fluorophore measurement. Using these measurement parameters, the phase and the modulation of the fluorescent response are calculated from the equations below [100]:

$$\varphi = (\varphi_R - \varphi_F) - (\varphi_R - \varphi_S) \quad (1.27)$$

and

$$m = \left(\frac{ac}{dc}\right)_F / \left(\frac{ac}{dc}\right)_S, \quad (1.28)$$

where φ_F , φ_S and φ_R are the phase of the fluorescent sample, scattered sample and reference PMT, respectively. $\left(\frac{ac}{dc}\right)_F$ and $\left(\frac{ac}{dc}\right)_S$ are the ac/dc ratios of the fluorescent and scattered sample. Subsequently, the phase lifetime τ_φ and modulation lifetime τ_m can be calculated from [35]

$$\tau_\varphi = \omega^{-1} \tan(\varphi), \quad (1.29)$$

and

$$\tau_m = \omega^{-1} \sqrt{m^{-2} - 1}, \quad (1.30)$$

where $\omega = 2\pi f$ is the modulation angular frequency. τ_φ and τ_m are equal when the sample exhibits a single-exponential lifetime, and different in the case of a multi-exponential lifetime. For the identification of the lifetime components in a multi-exponential sample, φ and m of multiple values of ω are required [101]. Due to the sinusoidal modulation of the excitation light at a single frequency, this requires multiple data sets with different excitation frequency values in FD-FLIM, thereby potentially increasing the total data acquisition time.

Based on the maximum sensitivity of phase $\left(\frac{\partial\varphi}{\partial\tau}\right)$ and modulation $\left(\frac{\partial m}{\partial\tau}\right)$ to the change in lifetime, the optimal modulation frequency for phase and modulation lifetime is given by

$$f_{opt_ \varphi} = \frac{0.7}{2\pi\tau}, \quad (1.31)$$

and

$$f_{opt_ m} = \frac{1.4}{2\pi\tau}, \quad (1.32)$$

respectively, in the case of finite SNR [102, 88]. If phase lifetime is selected due to its insensitivity to uncorrelated background noise, $f_{opt_ \varphi}$ for 1-5 ns lifetime range can be estimated from Equation (1.31) as 22-111 MHz. However, the processing of signals at this frequency range introduces challenging requirements on the bandwidths of electronic components to be used for data processing [103, 104]. This problem is conventionally overcome by exciting the fluorescent sample at its optimal modulation frequency range, and subsequently down-converting the detector signal to lower frequency values.

This conversion is based on frequency mixing, in which the product of two signals with frequencies of f_1 and f_2 generates two distinct frequency components at $f_1 - f_2$ and $f_1 + f_2$. This operation is typically implemented by modulating the gain of the detector, effectively using it as a frequency mixer. Alternatively, it can be performed by a discrete RF mixer block at the output of the detector [105]. After the multiplication, the down-conversion is finalized by filtering the output with a low-pass filter, thereby removing the high-frequency ($f_1 + f_2$) component. This filter can be implemented by a discrete RF block, or the detector itself, for slow detectors like CCDs.

There are two main types of frequency down-conversion in analog FD-FLIM: homodyning and heterodyning. In homodyning, the excitation (f_{exc}) and detector (f_{det}) modulation frequencies are equal, resulting in a DC ($f = 0$ Hz) output, whose intensity depends on φ [103]. In heterodyning, also called the cross-correlation method, f_{exc} and f_{det} have slightly different values, generating a down-converted signal in the kHz or Hz range [100]. A relative advantage of heterodyning is the ability to identify multi-exponential samples due to the availability of m . Furthermore, the reported unreliability of high-frequency mixers for DC output signals is another factor that favors heterodyning [105]. Another source of error in homodyning is the fact that noise at DC cannot be distinguished from the signal. This can be a significant problem especially in the presence of $1/f$ noise.

In the homodyning method, the typical way of measuring φ and m is the sampling of the emission response with a gain-modulated detector by shifting the phase of the detector modulation signal relative to the excitation modulation signal. In accordance with the rules of DFT, the multiple detector phase shift values must be equidistant across one full period of the excitation modulation signal. The intensity function of the phase shift images is in the form of a cosine function, given by [35]

$$I(\varphi_D) = k \times C \times (1 + 0.5 \times m_D \times m \times \cos(\varphi - \varphi_D)), \quad (1.33)$$

where k is a constant, C is the fluorophore concentration, m_D is the detector modulation, m is the emission modulation, φ_D is the phase of the detector, and φ is the phase of the emission signal. φ and m are calculated by fitting $I(\varphi_D)$ to a cosine function, and subsequently τ_φ or τ_m is calculated from Equation (1.29) and (1.30), respectively. In the heterodyning method, as $I(\varphi_D)$ is time-variant, only a single value of φ_D is sufficient for the calculation of φ and m . Conventionally, φ can be directly measured using a two-channel frequency counter, from the delay between the rising edges of the detector and reference excitation cross-correlation (mixer output) signals [98, 100]. Alternatively, φ can be measured by synchronizing the frame rate of the camera with the detector cross-correlation signal. To achieve that without aliasing, the frame rate of the detector f_D must meet the Nyquist rate requirement of $f_D \geq 2 \times f_{cc}$, where f_{cc} is the cross-correlation signal frequency. In [106], for values of f_{exc} up to 700 MHz, the cross-correlation signal at $f_{cc} = 7$ Hz was sampled using a CCD with a frame rate of 28 Hz. After the readout, the intensity function is digitized by an analog-to-digital converter (ADC) and subsequently digitally processed. Using the intensity values of the 4 sampling points, I_{0-3} , the phase is calculated by [106]

$$\varphi = \tan^{-1} \left(\frac{I_1 - I_3}{I_0 - I_2} \right). \quad (1.34)$$

More advanced system architectures for analog FD-FLIM were developed in the following years. In [107], most of the analog electronics in the system was replaced by digital electronics, and a significant portion of the data processing was shifted to the computer. This effort led to substantial reduction in computation error and added more configurability. With the evolution of the system hardware in FD-FLIM, the efforts to perform widefield measurements have also accelerated. In [108], the PMT was replaced by a CCD camera. Due to the slow frequency response of CCD pixels, the high frequency modulation of the detector was enabled by placing a microchannel plate in front of the CCD.

The main important advantage of analog FD-FLIM compared to TCSPC FLIM is that it is suitable for the rapid calculation of single-exponential lifetimes. On the other hand, the sinusoidal modulation of the detector that is used in most systems for frequency down-conversion automatically reduces the duty cycle of the system (defined as the fraction of the time during which the sensor is photon sensitive) to 50%, which causes at least half of the emitted photons to be missed by the detector, without including the photons lost due to the detector QE. In addition, the implementation of the sinusoidal modulation still requires analog electronic blocks, despite the digitalization efforts of the overall system. Finally, the analysis of multi-exponential decays requires data acquisition at multiple frequencies, which increases the acquisition time.

In some new generation analog FD-FLIM systems, the sinusoidal modulation of the detector is replaced by square modulation. This switch was a crucial step to reduce the dependence of the detector hardware on analog electronics. There are nowadays several research-grade and commercial cameras that are able to perform phasor analysis using analog FD-FLIM as the acquisition method. In [109], using a lock-in CCD camera with two complementary gates [110], lifetime imaging on of single-exponential decays was performed using only a single frame. In other works, a similar acquisition scheme using the same camera was combined with phasor analysis to observe various phenomena such as the effects on the lifetime accuracy of translational motion during the measurements [111] and the actions of antimicrobial peptides on living bacterial cells [112]. Furthermore, a commercial camera based on lock-in CMOS image sensor technology is also available for phasor imaging in the frequency domain [113]. In the standard operation mode, the camera is able to

modulate a CW light source at up to 40 MHz frequency, and the two square-shaped gates are used to detect the emission signal. Alternatively, the camera can be triggered externally with pulsed excitation, as demonstrated in [114].

1.3.3 Digital frequency domain FLIM (DFD-FLIM) / time gating

The concept of digital frequency domain FLIM (DFD-FLIM) emerged from the need to address the two aforementioned architectural drawbacks of analog FD-FLIM: limited duty cycle and the dependence on RF electronic components [89]. To achieve this, two main principles were used. The first principle is the acquisition of timing information using square-shaped sampling windows in the readout electronics, instead of modulating the detector photon sensitivity. Although this difference does not change the timing behavior, it allows the implementation of the photon timing mechanism using digital logic circuits without any need for RF components. Furthermore, the continuous sensitivity of the detector permits the generation of multiple sampling periods covering the entire modulation period, which allows 100% duty cycle. The second principle is the achievement of phase delays that are significantly smaller than the length of the sampling windows. In the original demonstration of DFD-FLIM, a technique called digital heterodyning was used to implement these small phase differences [89]. This technique, illustrated in Figure 1:8 [89], is based on the modulation of the excitation light and the detector with slightly different frequencies, causing the sampling windows to shift with respect to the laser pulse in each excitation period. The magnitude of the shift is determined by the ratio between the frequency difference and the frequencies themselves. The same effect can also be achieved using alternative methods, such as the precise shifting of the phase of the sampling windows against the laser pulse after a certain number of frames.

Recently, with the widespread availability of high-bandwidth electronics, the implementation of digital FD-FLIM without any need for homodyning or heterodyning has also become possible. In [104], the emission response of a single-point scanning system was detected by fixed-gain avalanche photodiodes (APD) at three different spectral bands in parallel, sampled at a rate of 250 MS/s, and digitized with 14-bit amplitude resolution using an FPGA. Despite being a highly promising approach, this method is incompatible with widefield FLIM due to large raw data size (~ 0.4 GB/s/channel), and not single-photon sensitive.

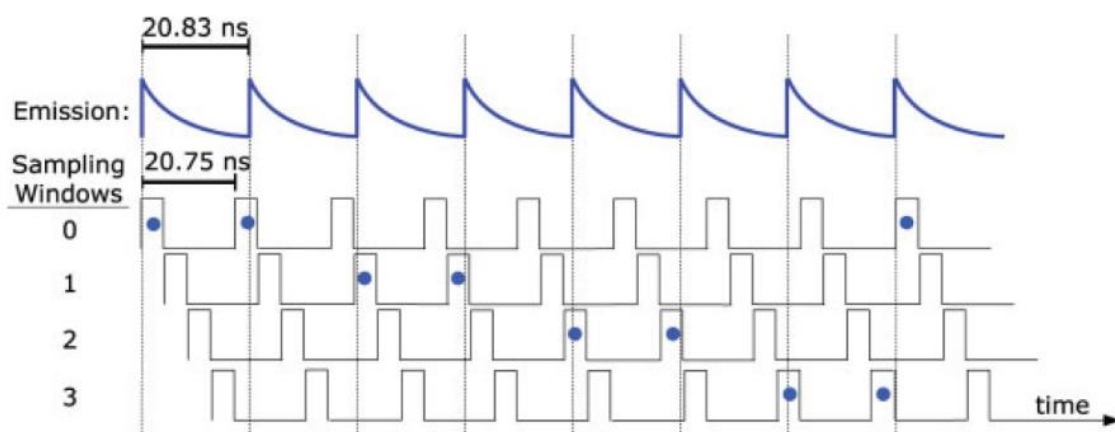


Figure 1:8 Operating principles of digital frequency domain FLIM (DFD-FLIM) with heterodyning [89].

Time-gated FLIM is another common technique, particularly among solid-state time-resolved image sensors. The first work to demonstrate time-domain FLIM, which dates back to 1960s, used the time gating approach

[115]. In this work, which targeted lifetimes in the range of nanoseconds, the gates were implemented by modulating the gain of a PMT with a square pulse, which turned it on and off rapidly. An alternative gating method which did not require any detector sensitivity modulation was based on sampling the time of arrival of an emission pulse, and subsequently registering it to a channel of an oscilloscope controlled by the sampling unit, based on its time of arrival and amplitude [116]. This method was superior to the previous one in terms of both duty cycle and gate precision. In these initial setups, an important performance challenge was caused by the noise in the flashlight pulses [35]. This fundamental problem was later eliminated with the development of picosecond and femtosecond pulsed lasers. These two methods form the basis of the currently used time-gated FLIM architectures, despite substantial changes in the hardware components and capabilities over time. In most new-generation systems, the sampling-based gating method is chosen, in which the oscilloscope is replaced by integrated readout electronics. This architectural choice also enables the generation of multiple gate channels, which increases the speed and photon efficiency.

Time gating can be considered as a variant of DFD-FLIM, when used for the phasor method. The conceptual illustration of the time-gated data acquisition for phasor-FLIM is shown in Figure 1:9 [117]. In this method, according to the principles of DFT, a gate with a fixed width is shifted with uniform steps across the fluorescence emission signal for a measurement window that is equal to the laser period. While Figure 1:9 represents pulsed excitation, the excitation can be either pulsed or sinusoidally modulated. In time-resolved CMOS and CCD sensors, digital time gates are usually implemented as taps or buckets, which are multiple electron wells in a pixel. By external voltage modulation, each detected photoelectron is directed towards one of these taps. In SPAD imagers, the gate is typically implemented as a transistor that controls the current between the photodiode and the memory capacitor. As the pixel format of SPAD imagers became larger, the SPAD imagers that employ time-gating as the primary time-resolved imaging method increasingly became more common, as shown in Table 1:1. The reasons for this were the implementation difficulties of in-pixel TDCs in large-format imagers without a sacrifice in fill factor, and the substantial global photon count rate penalty of the shared TDC architectures. Although in-pixel TDC architectures are being demonstrated in increasingly larger pixels formats [57], the relative simplicity of the gating architectures and their support for higher spatial resolution still remain valid.

Despite its aforementioned advantages, the photon timing precision (temporal resolution) in time gating is inferior to the TCSPC method, for several reasons. Firstly, the minimum width of a time gate is larger than the bin width of the state-of-the-art TCSPC methods. Secondly, narrower gate windows require larger number of distinct gate profiles with simultaneous data acquisition capabilities (also referred to as “gate channels” in this thesis) to achieve 100% duty cycle. In practice, the implementation of more than four gate channels in a pixel is challenging due to its negative impact on the fill factor and high throughput demand from the readout circuits. On the other hand, this gap in photon economy is largely offset by a substantial gain in maximum photon count rate compared to scanning-based systems.

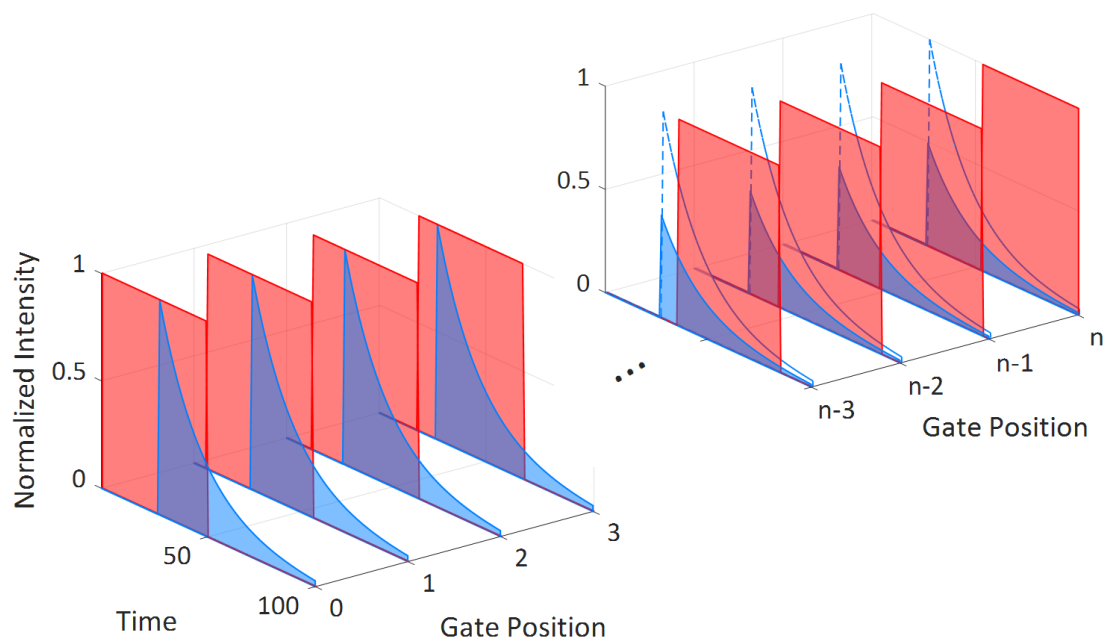


Figure 1:9 Conceptual illustration of time-gated FLIM [117].

Table 1:1 Evolution of preferred time-resolved imaging method among large-format SPAD imagers.

Reference	Pixel Format	Year of Publication	Primary Time-Resolved Imaging Method
[10]	160×128	2011	TCSPC
[57]	192×128	2019	TCSPC
[118]	256×256	2015	TCSPC
[119]	256×256	2018	Gating
[120]	512×128	2014	Gating
[121]	320×240	2016	Gating
[117]	512×512	2019	Gating
[122]	1024×1000	2019	Gating
[123]	1200×900	2020	Gating

The FLIM method proposed in this thesis, which is based on a widefield detector equipped with time-gated pixels, can be included in this category.

1.4 Sensor architectures for widefield FLIM

In widefield FLIM, the sensor is one of the most important components, since its specifications contribute significantly to the overall system performance in multiple aspects. Advances in sensor technology in the last few decades have enabled significant improvements in widefield FLIM performance as far as frame rate, spatial and temporal resolution, photon sensitivity and dynamic range are concerned, expanding the practical applications of widefield FLIM and making it a strong competitor to confocal FLIM [124]. With the current trend of integrating other system blocks into the sensor architecture [10, 57], the detector is increasingly becoming the most critical component in a widefield FLIM setup.

1.4.1 Gated intensified CCD (ICCD)

Charge-coupled device (CCD) was one of the mainstream technologies for conventional photography, and the most popular choice in this field for decades [125]. Time-resolved imaging was achieved with a single CCD camera after the introduction of lock-in pixels [5], which is discussed in Subsection 1.4.3. However, these cameras do not achieve single-photon sensitivity due to the readout noise introduced by the amplification stage, which renders them incompatible with low-light-level imaging, an essential requirement for fluorescence microscopy. This problem was solved by electron multiplying CCD (EMCCD) technology, in which the photoelectrons are multiplied in shift registers before the amplification stage, thus reducing the relative magnitude of the readout noise compared to the electron count of a single photon [126]. While EMCCDs became a primary technology choice for fluorescence imaging, they cannot be used for FLIM due to the lack of time-resolved imaging capability using a single camera unit. CCD cameras based on multi-collection-gate (MCG) technology, which employ pixels with multiple contiguous gates [127], can potentially be used for time-resolved imaging by synchronizing their gates with a pulsed or modulated light source. However, this implementation has not been demonstrated to date.

The intensified CCD (ICCD) was originally invented for high-speed imaging at ultra-low-light level [128], and later also used for time-resolved imaging [129, 47, 130]. In addition to its gating mechanism, it also offers sensitivity to low photon counts, making this technology attractive for scientific applications. The structure of an ICCD, as displayed in Figure 1:10(a) [8], consists of a standard CCD sensor and an image intensifier tube that is placed before the sensor [8]. In the image intensifier, the incident photon is first converted to an electron by a photocathode. A stack of microchannel plates (MCP) behind the photocathode multiplies this electron. The multiplied electrons are then converted back to photons by a phosphor screen and directed to the CCD active area using fiber or lens coupling. The intensifier adds two extra features to the camera. Firstly, it enables low-light-level sensitivity. The multiplication of photoelectrons before the readout stage minimizes the SNR drop caused by readout noise [131]. Secondly, it adds time-resolved imaging capability. Indeed, unlike the CCD pixel, the image intensifier can be turned off and on rapidly by modulating the voltage of the photocathode, and/or the MCP [132]. By synchronizing this voltage with a pulsed or sinusoidal illumination source, time-gated imaging can be achieved [133, 134]. The width of the gate windows implemented in this configuration can reach 200 ps, as shown in Figure 1:10(b) [133].

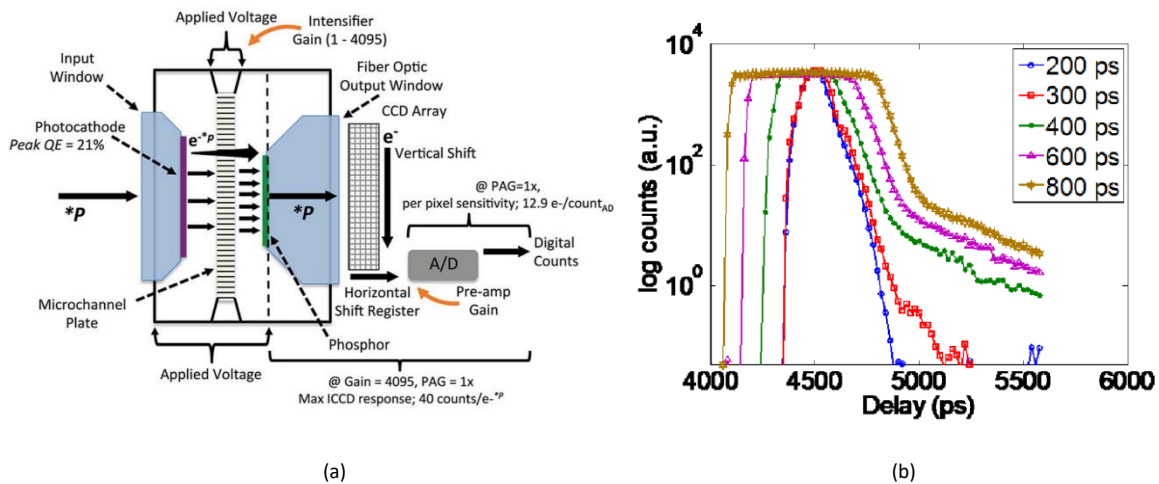


Figure 1:10 (a) Diagram of an ICCD camera structure [8] (b) Gate profiles of an ICCD camera for various gate widths [133].

Despite being a popular detector choice for widefield FLIM, the ICCD has several structural disadvantages. Firstly, its duty cycle is below unity by design. Since gating is achieved by turning off the image intensifier in a certain time window, the photons arriving outside this time window are missed. Unlike alternative technologies, where multiple gates can be generated by readout electronics while keeping the detector permanently on, the expansion to multiple gates is not possible in ICCDs. This limitation introduces a trade-off between the duty cycle and photon economy.

Image intensifiers, while being of key importance for single-photon sensitivity, introduce several performance limitations to the ICCD. Firstly, they cause a drop in the overall quantum efficiency (QE) compared to a CCD camera without intensifier. In general, the QE of the photocathodes does not exceed 40-50% [135]; this value is even lower for the variants with a wider spectral range [136, 137]. Furthermore, the fill factor of MCPs is also less than 100% due to the area occupied by the spaces between the microchannel holes [92]. The second problem is the high operation voltage of the MCP. In order to accelerate the electrons through the microchannels to achieve the desired gain level, the required voltage is in the order of kilovolts. Thirdly, the lifetime of an MCP is limited, which reduces the overall robustness of the system [92].

Since ICCDs employ a CCD sensor, the architectural limitations of CCDs apply to ICCDs as well. A major disadvantage of CCDs that eventually led to them being overtaken by CMOS imagers in photography is cost. Though CCDs can be very fast, up to 1 Gfps in burst mode [127], CCD technologies are not compatible with standard CMOS nodes. Thus, implementing fast ADCs on chip is generally not an option and so the readout is analog, usually through a single channel. As the noise of the charge amplifier stage increases with bandwidth, CCD cameras are relatively slow devices in continuous mode, especially, with large pixel counts.

An additional problem is caused by the early-stage electron multiplication scheme. While the multiplication of electrons before the CCD stage enables single-photon sensitivity, it also results in the saturation of the electron wells in the CCD pixels. This introduces a trade-off between single-photon sensitivity and dynamic range, which can be controlled by tuning the gain of the intensifier [138]. In state-of-the-art commercial ICCD systems, the dynamic range can reach 85 dB at ICCD gain below 10, compared to only around 55 dB at a gain of 1,000 [139].

Finally, the gating mechanism employed in image intensifiers offers only partial closing of the gate in ICCDs, thereby leading to the detection of photons when the gate window is closed. This undesired effect, also-called “bleed-through” effect, is quantified by the parameter called “on/off ratio”, defined by [132]

$$On/Off\ Ratio = \frac{Real\ Signal\ (cts)}{Leakage\ (cts/ms) \times Gate\ Width\ (ms)}. \quad (1.35)$$

For popular ICCD camera models, on/off ratios in the order of 10^7 can be achieved [140]. However, it was also reported that the on/off ratio is wavelength dependent, and can decrease to the level of 10^4 at ~ 200 nm [132]. To overcome this limitation, a different method was developed which proposes gating the MCP along with the photocathode, which results in an on/off ratio up to $> 10^8$ at ~ 190 nm [132]. However, this method comes with the challenges of generating a very sharp square pulse in the 500-1000 V range, such as damping of the signal to avoid ringing. The integrity of the pulse signal is more crucial in the gating of MCPs than the gating of a transistor in solid-state devices, as the photon sensitivity varies with fluctuations in the gate voltage in the former, but not in the latter. In addition, the requirement to keep the pulse width of the MCP larger than the pulse width of the photocathode introduces a fundamental limit to the improvement of the on/off

ratio, as the leakage will still occur in the time window where the MCP voltage is high and photocathode voltage is low.

1.4.2 Position-sensitive anode detectors

A different approach to widefield FLIM is the use of a TCSPC system with an additional 2D localization feature. This approach aims to combine the advantages of single-channel detectors in timing resolution with the wide-field imaging capability. One of the primary motivations for the development of these detectors is the performance drawbacks of the ICCDs existing at the time in terms of gate width and low-light-level performance [141].

The idea of adding sub-nanosecond timing feature to a 2D single-photon detector first emerged in late 1980s [142]. The detector developed in this work had two main components. A combination of MCPs and resistive anodes (RA) was used as a 2D single-photon detector [143], and a standard TCSPC system with an avalanche photodiode (APD) and a TAC was used for photon timing. A further improved version of this detector with the same architecture was capable of detecting 1,024 decays in parallel with 80 ps timing resolution [141], which was a timing performance comparable to the state-of-the-art single-channel MCP detectors at the time [144]. The working principles of the 2D detectors are similar to the image intensifier, with one major difference: the multiplied electrons are detected by the anode, which captures their time of arrival and 2D position, instead of being converted back to photons via a phosphor screen.

While position sensing can be achieved with several alternative anode architectures [92], the most widely used architecture is the cross delay-line (XDL) anode. In Figure 1:11(a), the diagram of the H33D detector, a position-sensitive single-photon detector with an XDL anode, is displayed [7]. This anode structure is formed by two layers of meandering wires that scan the entire area under the photocathode and the MCPs. These two wires provide the X and Y coordinates of the arriving electron. The coordinates are calculated from the propagation delay difference between the two ends of each wire, which is measured by TDCs at the edges. The advantage of the XDL anodes is the high spatial resolution with only a single readout channel for each axis. However, this advantage comes at the expense of significant dead time. Since only one photon must be read until the propagation across the entire wire is completed, the dead time of the anode increases with the length of the wires, whereby longer wires are necessary to achieve high spatial resolution. A typical experimental setup of widefield FLIM for these detectors is shown in Figure 1:11(b), where it is indicated that for each detected photon, time-of-arrival information and XY position are recorded by a TDC and the XDL anode structure, respectively [7].

Anode-based detectors have comparable performance to ICCDs, since they use a similar mechanism for the photon multiplication stage. However, several differences arise from the use of separate timing and position detection mechanisms. Firstly, the use of single-channel TDCs for photon timing allows anode-based detectors to reach picosecond timing resolutions, compared to the hundreds of picosecond for ICCDs. Secondly, the duty cycle of these detectors is 100%, owing to the lack of detector sensitivity modulation. Finally, due to the absence of a camera sensor, anode-based detectors can be implemented with larger active areas, the main limitation usually being determined by the maximum achievable MCP diameter.

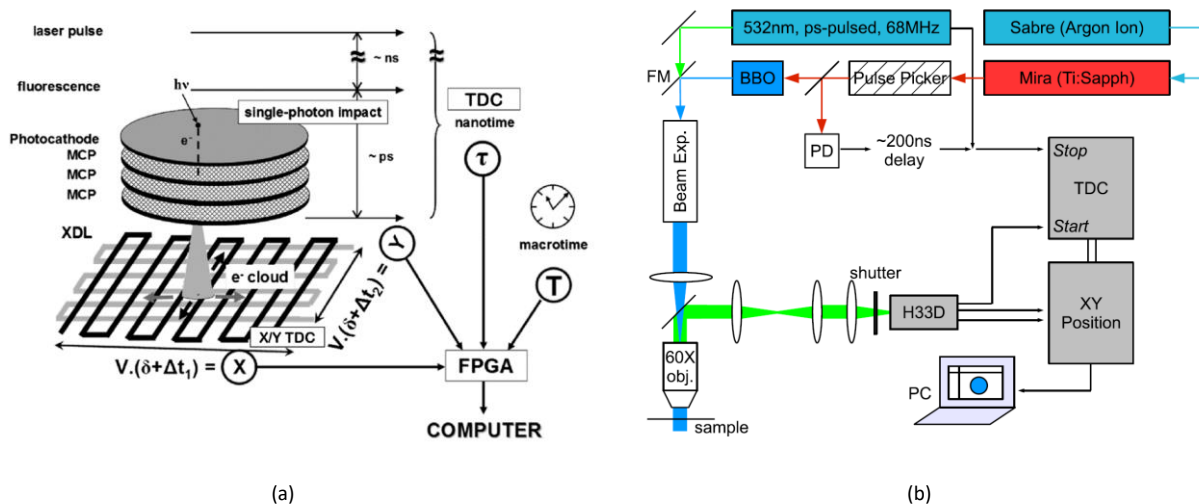


Figure 1:11 (a) Diagram of the H33D detector, a position-sensitive detector with a cross delay-line (XDL) anode, designed for widefield FLIM [7]. (b) The experimental setup used in [7] to perform widefield FLIM. For each detected photon, the time of arrival with respect to the laser pulse is detected by the TDC, and the XY position is detected by the cross delay-line (XDL) anode.

The most important disadvantage of these detectors is their global and local photon count rate limitation. Since they employ only a single-channel timing mechanism, the combined dead time of the anode and the timing electronics determines the maximum photon count rate in the entire field of view. As a consequence, these detectors are not suitable for high-speed FLIM. Furthermore, the disadvantages of photocathode and MCPs, which were explained in the Subsection 1.4.1, also apply to these detectors. These problems include low QE of the photocathode, and size limitation and high voltage operation requirement of the MCPs.

1.4.3 Gated CMOS image sensor (CIS)

CMOS image sensors (CIS), which were initially considered as more affordable but noisier alternatives to CCD cameras, have emerged as a valid alternative for scientific imaging in the last decade owing to the substantial progress in their noise performance thanks to the introduction of the pinned photodiode (PPD) [145, 146]. This trend has recently culminated in the development of single-photon sensitive CIS cameras with sub-electron readout noise [147, 148]. To complement the photon-counting ability with time-resolved imaging feature, CIS cameras with different variants of in-pixel gating were introduced. The efforts to add ToF capability to conventional image sensors started with the invention of lock-in pixels. This concept, first implemented in CCD technology [149] and later adapted to CMOS [150, 151], divides the exposure time of a frame into multiple equally-sized sample windows. The initial motivation behind the lock-in pixel was to perform 3D ranging using active illumination with sinusoidal modulation. Owing to its capability of storing multiple frames in the same pixel, it was also considered as a suitable hardware technology for fast motion detection.

In these pixels, the gates (also called taps) are implemented by generating multiple electron wells and directing the photoelectrons to the selected well by modulating the electric field with an external voltage. Since the invention of the first lock-in pixel with two taps, several features were added to these architectures to improve the SNR to the level required by demanding applications such as FLIM. The draining-only modulation (DOM) technique substantially improved low-light level sensitivity by minimizing the pixel-to-pixel sensitivity variation [152]. Lateral electric field charge modulation (LEFM) technique enabled first lock-in pixels with sub-ns time resolution [153]. Finally, with the extension of the number of taps from two to four, the lifetime analysis of double-exponential fluorophores from a single frame became possible [9].

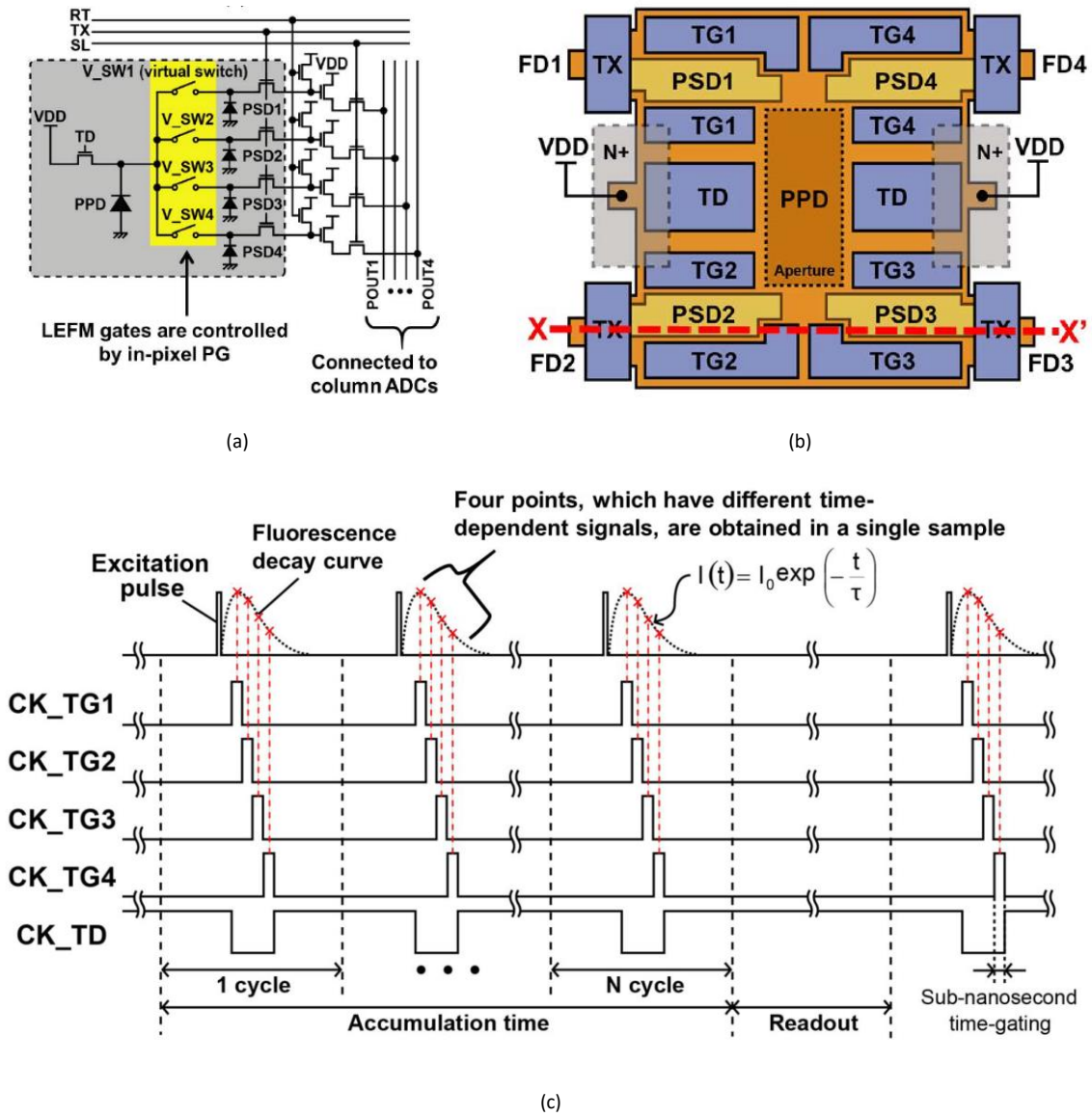


Figure 1:12 (a) Schematic, (b) layout and (c) timing diagram of a 4-tap CMOS lock-in pixel [9].

A key requirement for making CIS cameras compatible with FLIM was single-photon sensitivity, since low-light-level detection is essential for many biologically relevant fluorescent samples. While long integration times can be proposed to overcome the readout noise limitation, this approach does not represent a feasible solution for real-time FLIM, which introduces an upper limit for the acquisition time of a gate image. For this purpose, the dominant noise mechanisms and noise reduction techniques in this technology were investigated [154, 155, 156]. In new generation CIS cameras, sub-electron noise performance was reached with the help of various circuit-level techniques such as column/pixel-level amplification [157, 158] and correlated multiple sampling (CMS) [159], or modifications in the CMOS process [160, 161]. After being demonstrated in standard CIS pixels, some of these techniques were applied to lock-in pixels to achieve similar results [147].

The major performance challenges of the gated CIS technology are as follows. Firstly, the area occupied by the gating mechanism reduces the fill factor; the degradation of fill factor becomes more significant as the number of taps increases. This effect is illustrated in Figure 1:12; the PPD forms the active area of the pixel, whereas the blocks colored in blue and yellow are transistors and storage photodiodes which constitute the taps [9].

Secondly, the readout noise in CMOS sensors highly depends on analog performance, and the reduction of the readout noise to single-photon counting levels requires substantial circuit level optimization [162]. The overall degradation in analog performance with technology node shrinking prevents the implementation of these sensors in advanced CMOS processes that allow the integration of complex electronics into the pixel. Finally, the techniques employed for sub-electron readout noise may also cause various unwanted performance penalties. Depending on the implementation technique, CMS requires additional active [163] or passive circuitry [164], or causes readout speed degradation [165]. Furthermore, high-gain column amplification to increase photon sensitivity can reduce the dynamic range by limiting the maximum photon count through saturation [166].

1.4.4 Single-photon avalanche diode

A single-photon avalanche diode (SPAD) is a single-photon sensitive photodiode that offers sub-ns temporal resolution at room temperature. As a solid-state device, the SPAD rose to prominence as a promising technology to fill the gap between the photocathode-based single-photon detectors with limited quantum efficiency and high operating voltage, and established image sensor technologies with no single-photon sensitivity and poor temporal resolution.

A SPAD is an avalanche photodiode (APD) designed to operate above breakdown, typically formed by a p-n or a n-p junction. SPADs are also called Geiger APDs or G-APDs [167]. In this regime, an electron excited by a photon absorbed in the high electric field region (also called multiplication region) excites multiple additional electrons due to its high kinetic energy, before losing its energy through recombination. The large current that is generated through this fast electron multiplication can be easily detected by readout electronics. The magnitude of this current, called avalanche current, does not indicate the detected photon count anymore, in contrast to an APD operated in linear mode. Therefore, the avalanche current is typically directly converted to a digital voltage, and the relatively small fluctuations in the voltage level throughout the readout process do not affect the detected photon count. As a consequence, SPADs have virtually no readout noise.

Over decades, SPADs have transformed from a merely experimental endeavor to a mainstream time-resolved imaging technology with a widespread academic and commercial use. The first remarkable development in SPADs was the use of a planar p-n junction for its design [168, 169], which made them compatible with CMOS fabrication processes unlike their predecessors. The development of epitaxial devices demonstrated substantial performance improvements in terms of timing resolution [170]. This was followed by the design of first SPADs in standard CMOS processes, using a 0.8 μm technology node [171]. The shift from custom CMOS to standard CMOS initially led to a performance drop due to reduced flexibility in fabrication parameters. On the other hand, it allowed the quenching and readout circuitry to be integrated with the SPAD itself for the first time, paving the way for integrated SPAD pixels, as well as large area smart imaging sensors.

The first fully integrated CMOS SPAD pixel consisted of a SPAD, a passive quenching resistor and a comparator for detecting the edge of the voltage spike due to the avalanche current [172], as shown in Figure 1:13(a) [173]. The blocks that were later integrated into SPAD pixels include digital counter, active quenching and recharge, time gate and time-to-digital converter (TDC). The integration of more complex blocks into SPAD pixels became possible as a result of technology scaling in CMOS. Development of SPADs in deep submicron (DSM) technology [174] and 3D stacked processes [175] represented two major milestones further increasing pixel complexity (Figure 1:13(b) [15]). In light of these technology developments, the spatial resolution of SPAD

sensors increased over the years, and finally culminated in two distinct megapixel SPAD arrays in 2019 and 2020, respectively [122, 123].

SPADs have gained attention as a detector technology for FLIM since their early years. In confocal FLIM, SPADs were considered as a solid-state alternative to PMTs. The main advantages of SPADs were the lack of fundamental QE/PDP limitation that PMTs had due to the use of photocathodes, and lower operation voltages. The first TCSPC experiments using SPADs were conducted in the work [46]. In this work, it was observed that the timing performance of SPADs was limited by their diffusion tail, which is caused by the movement of the minority carriers across the depletion region [176], and predominantly occurred at higher visible and NIR wavelengths. Owing to the improvements in photon sensitivity and timing resolution particularly in custom technology versions, SPADs can currently compete with PMTs and hybrid photo detectors as the primary single-photon detector choice in life sciences [177, 178].

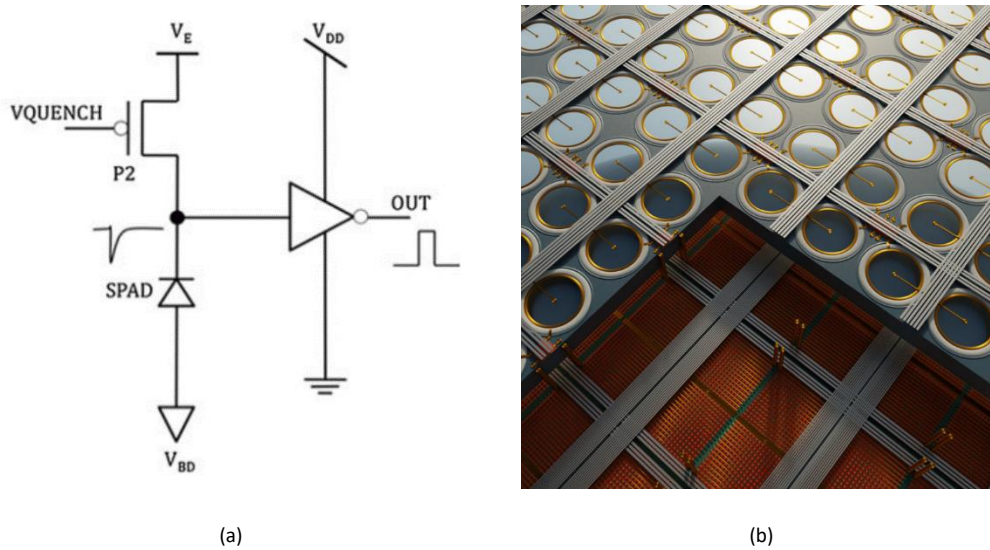


Figure 1:13 (a) Schematic of a SPAD pixel with its most basic elements [173]. The voltage spike at the floating node of the SPAD, caused by the avalanche current, is converted to a digital pulse by an inverter. (b) Conceptual image of a SPAD imager in 3D stacked process technology [15].

SPADs have increasingly become a common detector of choice in widefield FLIM, as well, owing to their strengths as time-resolved image sensors. In [179], a SPAD imager with 32×32 pixels and in-pixel TDCs was used for TCSPC FLIM. Although the spatial resolution and the 119 ps timing resolution were behind the state-of-the-art systems at the time, the work demonstrated widefield FLIM for the first time with a fully integrated time-resolved imager. For the next generation SPAD-based widefield FLIM imagers with larger spatial resolution, in-pixel TDC architectures were temporarily abandoned due to their significant fill factor penalty, and replaced with either column-shared TDCs with event driven architectures [180], or time-gating [181]. In recent years, however, in-pixel TDC approaches were again made possible for large SPAD arrays thanks to the shrinking CMOS technology nodes [57].

Despite its rapid progress in performance, the SPAD technology still faces several important challenges to date that influence its status as a detector choice. Since dark count, which is the dominant noise source in SPADs, is proportional to the exposure time, it may introduce a limitation on the minimum photon flux that can be detected, depending on the DCR of the SPAD model and the properties of the process technology. On the contrary, scenes with low illumination can be detected in conventional cameras via long exposure as their dominant noise source, readout noise, is not affected by the exposure time. Band-to-band tunneling, a major

contributor of DCR, is higher for narrow junctions with large electric field [182]. These characteristics have several implications in performance. Firstly, they cause a trade-off between DCR and PDP, both of which are critical performance parameters. Secondly, they cause an increase in DCR with technology scaling, due to the narrower junction sizes in smaller technology nodes. Another challenge of SPADs is related to jitter. Jitter degrades as the size of the multiplication region increases, which is necessary for a wide spectral response. Jitter may also be higher in SPADs with large active areas, causing a trade-off between jitter and fill factor. Afterpulsing, which is the generation of delayed counts after an initial photon detection, and recharge time introduce a limitation to the minimum photon inter-arrival time that can be detected by SPADs, which is also called the dead time. Particularly in large-format SPAD imagers, crosstalk, defined as a false photon detection event caused by another photon detection event in a neighboring pixel, can also represent another source of noise. Finally, while the PDP of state-of-the-art SPADs have exceeded photocathode-based detectors, it is still behind the QE of CMOS image sensors.

In summary, in addition to the single-photon sensitivity and excellent timing capability, the two main strengths of SPADs which place them in a unique position among scientific detectors are the absence of readout noise and monolithic integration with CMOS electronics. The former results in the ability to detect single-photons without significant dynamic range degradation, whereas the latter allows digital on-chip pre-processing, enabling extremely high local and global photon count rates.

1.5 Contributions

The contributions of the thesis are summarized below.

A large-format SPAD camera was employed for phasor-FLIM for the first time. Prior to SPAD technology, the detector choices for widefield FLIM were non-solid-state technologies such as position-sensitive anode detectors and ICCDs, and solid-state detectors with limited single-photon detection ability, like lock-in CIS cameras. With SwissSPAD2, it was shown that the gate width limitations of large-format SPAD arrays do not present an obstacle to the measurement of short lifetimes with high precision. The demonstration of this concept paves the way for the use of next generation SPAD imagers with in-pixel data processing for video-rate phasor-FLIM at high spatial resolution.

The second contribution was in the spatial resolution of SPAD arrays. At the time it was published, SwissSPAD2 achieved the highest spatial resolution (512×256 pixels in the camera module, 512×512 pixels in the image sensor chip) among SPAD imagers. Furthermore, the SPAD design employed by SwissSPAD2 reports one of the best combinations of PDP and DCR among SPADs that were fabricated in standard CMOS process technology and used in large-format imagers.

With its dual-gated architecture, SwissSPAD3 became the largest time-resolved SPAD imager with 100% duty cycle. In SPAD imagers, the use of multiple complementary gates with pixel architectures such as single-photon synchronous detection (SPSD) was generally limited to arrays with small number of pixels, due to the fill factor penalty caused by the gate electronics area [183, 184]. In SwissSPAD3, by deriving the second gate from the first gate and the intensity count, two natively contiguous gates were distributed across the pixel array, without any need for special signal matching techniques. SwissSPAD3 also achieved the shortest gate width in large-format SPAD imagers (1 ns). In this sensor, a minimum gate width of 1 ns was measured without significant degradation in the other gate parameters compared to longer gate widths in the same device. Finally, it

achieved state-of-the-art gate shape uniformity and gate skew, with a substantial improvement compared to SwissSPAD2.

1.6 Organization of the thesis

The thesis is organized as follows. Chapter 2 is dedicated to time-gated FLIM. It starts with the description of the general principles of this technique, and explains its different and most widely adopted variations. It also describes the gating approach used in the experiments described in this thesis. Furthermore, several important gate parameters and their impacts on the FLIM performance are also discussed. Chapter 3 presents SwissSPAD2 and SwissSPAD3, two widefield time-gated SPAD image sensors. In this chapter, the design choices, trade-offs and the characterization results of the sensors are presented. Chapter 4 is dedicated to the various scientific applications which were demonstrated using SwissSPAD2 and SwissSPAD3. The chapter contains the descriptions of the methodology of each experiment, followed by the measured results. Subsequently, a discussion on the achieved performance, the fundamental performance limitations, and a comparison with state-of-the-art methods are included. Finally, Chapter 5 summarizes the conclusions of the thesis and proposed future work.

Chapter 2 The theory and practice of time-gated FLIM

2.1 Overview

Time-gated systems achieve time-resolved imaging by dividing the response of the scene into multiple time windows, and measuring the photon count of each window sequentially or in parallel (Figure 2:1). Due to inherent hardware limitations on the number of parallel gates and focus on data acquisition speed and photon sensitivity, gates sample a fluorescence exponential function generally with a smaller number of data points and larger time windows compared to TCSPC. For years, research efforts have aimed to find the optimal gating parameters which yield the best overall lifetime estimation performance [185, 186, 72, 187]. Early studies in this direction mostly focused on minimizing the accuracy and precision loss compared to TCSPC while achieving significantly higher speed [73]. Furthermore, the target applications of these studies were mostly exploiting the discrimination of two single-exponential decays, or qualitatively observing temporal changes in the lifetime of a sample and thus, speed was emphasized over accuracy and precision.

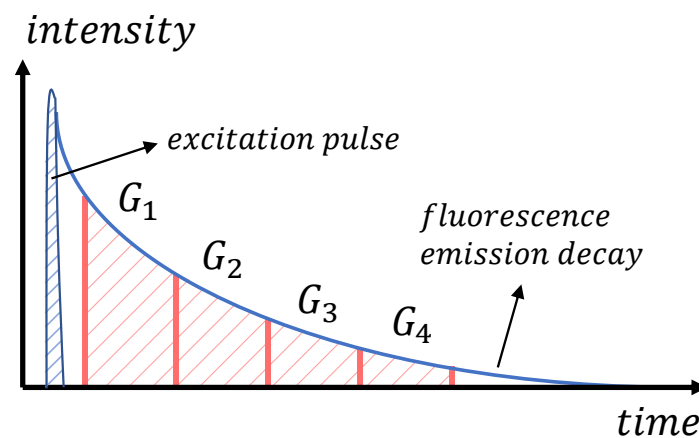


Figure 2:1 Conceptual illustration of time-gated FLIM in time domain. The intensity of a fluorescence emission decay, which is generated as a response to an excitation pulse, is measured for different gate windows. This diagram represents only one example of gating; time gating can employ various gate configurations which are discussed throughout the chapter. Adapted from [188].

Recently, the accuracy and precision gap between TCSPC and time gating has gradually narrowed for several reasons. Thanks to the advances in speed and size of electronics, and the implementation of time-resolved image sensors in 3D stacked processes, shorter gate widths and higher number of gate channels no longer have to introduce substantial penalties in duty cycle and fill factor, respectively. Indeed, even though the minimum achievable gate width is often still larger than a TCSPC bin, high sampling rates can be achieved with overlapping gating schemes, where the delay between the gates can reach values comparable to the typical TCSPC bin sizes. Potential penalties in speed due to the increase in raw data size are mitigated by the application of new data analysis techniques such as the phasor method [77]. As a consequence of all of these developments, time-gated imaging is increasingly becoming a serious competitor to TCSPC even in more

complex experimental setups featuring multi-exponential decays and with higher demands for accuracy and precision.

In time-gated systems, the shape of the gate determines the instrument response function (IRF) of the complete instrument, which is defined as its response function to direct illumination by the excitation light source in absence of any other sources of illumination. The IRF contains the essential parameters that are ultimately influencing the lifetime estimation performance, as discussed in detail in Section 2.4. All sources of temporal variation in the system contribute to the gate shape, including the pulse width of the light source, and the jitter of the photodiode(s) and the gate signals. In systems with multiple gate channels, the uniformity of the gate shape across the pixels is also an important parameter. A significant non-uniformity between the gate channels can complicate the calibration and correction processes, potentially resulting in reduced lifetime estimation precision and/or in frame rate.

2.2 Conventional implementations of time gates

2.2.1 Time gating in rapid lifetime determination method

Rapid lifetime determination (RLD) is a method that emerged to address the need of fast fluorescence decay analysis with a strong emphasis on the identification and discrimination of different lifetimes in a sample, or rapid lifetime changes. Biologically relevant phenomena which require fast FLIM include changes in ion concentration and polarization/depolarization of the cell membrane [189]. RLD is generally more suitable for single-exponential decays than multi-exponential decays [190]; up to 100 fps frame rate was achieved using single-exponential samples with relatively low complexity [189]. Furthermore, RLD requires prior knowledge that a single exponential is a good model of the sample [73].

The aim of the original RLD method is to calculate the lifetime of a single-exponential decay with an unknown base line, which could be caused by uncorrelated noise sources such as background illumination [191]. The decay function analyzed in this method is expressed as

$$I(t) = Ae^{-t/\tau} + B, \quad (2.1)$$

where A is the preexponential factor, τ is the lifetime and B is the base line. The decay window is sampled by three contiguous gates with equal width, as shown in Figure 2:2(a) [191]. The single-exponential lifetime is calculated by

$$\tau = \frac{-\Delta t}{\ln\left(\frac{D_1 - D_2}{D_0 - D_1}\right)}, \quad (2.2)$$

where Δt is the gate width and D_i is the photon count of gate index i . If it is initially known that the sample displays no or negligible base line, the lifetime estimation of a single-exponential decay requires only two gates (Figure 2:2(b) [73]), and can be rewritten as [73]

$$\tau = \frac{-\Delta t}{\ln(D_1/D_0)}. \quad (2.3)$$

The required number of gates increases further with the number of exponential decay components. In accordance with its original goal of providing high contrast at maximum available speed, it was estimated that single-exponential RLD offers 10-100 times faster speed compared to the least squares method (LSM), a

conventional data analysis method for FLIM, at the expense of only 30-40% lower accuracy [73]. Subsequently, multiple research efforts focused on the customization of the gate configuration in the RLD method for performance improvements. These configurations include separated and equal-width gates [185], separated and unequal-width gates [186], and overlapping gates [72].

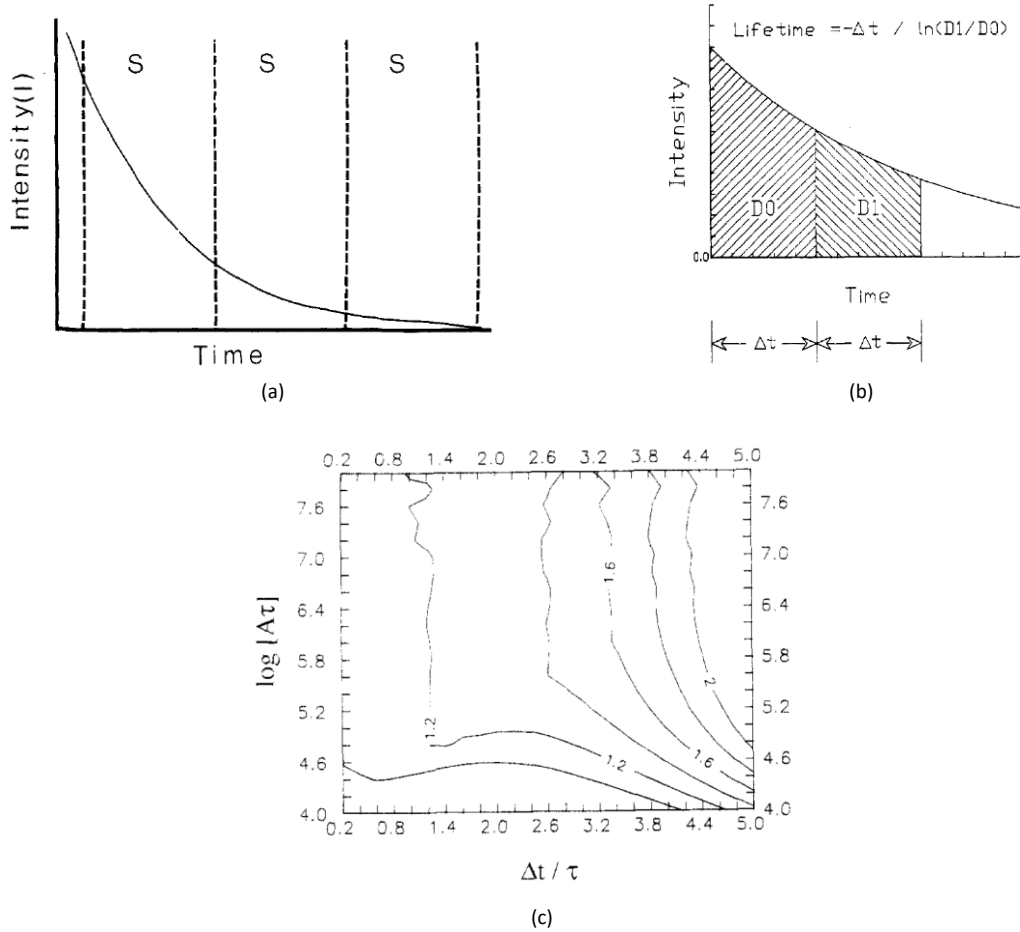


Figure 2:2 (a) The gates used in the original rapid lifetime determination (RLD) method for single-exponential decay and an unknown base line [191]. (b) The variant of the original RLD method for no base line [73]. (c) Ratio of relative standard deviations (σ/τ) of RLD with no base line and weighted linear least squares (WLLS) method, for different combinations of relative gate width ($\Delta t/\tau$) and total decay function area ($A\tau$) [73].

Quantitative performance characterization of the time-gated RLD method was conducted first in [191] by experimentally comparing the mean and the standard deviation of certain lifetimes estimated by RLD and more established FLIM data analysis methods. Subsequently, a theoretical error characterization for single-exponential decays was performed in [73]. In this work, the precision of RLD was compared with the weighted linear least-squares (WLLS) method under the same experimental conditions using both error propagation analysis and Monte Carlo simulations. Using error propagation, the standard deviation of the lifetime, σ_τ , was derived as

$$\sigma_\tau = \frac{-\Delta t}{(\ln(D_1/D_0))^2} \left(\frac{\sigma_{D_0^2}}{D_0^2} + \frac{\sigma_{D_1^2}}{D_1^2} \right)^{1/2}. \quad (2.4)$$

According to Equation (2.4), the standard deviation of the lifetime estimation is proportional to the gate width Δt , as long as D_0 and D_1 are not functions of Δt . Under these conditions, $\sigma_{D_0^2}$ and $\sigma_{D_1^2}$ also remain constant, as they are governed by Poisson statistics. Since D_1/D_0 depends on the lifetime, and D_0 and D_1

both depend on the preexponential factor (A) in the decay function shown in Equation (1.1), it can be concluded that σ_τ depends on both A and τ . The results of this analysis, with a good agreement between the error propagation and Monte Carlo methods, indicate that there is an optimum value of $\Delta t/\tau$ which yields the lowest relative standard deviation of the lifetime (σ_τ/τ). In this optimum region, the standard deviation of the lifetime estimated by the RLD method is 30-40% larger than in the WLLS method, whereas RLD can even outperform WLLS at lower photon counts (Figure 2:2(c) [73]). On the other hand, an experimental speed comparison with the WLLS method shows that the single-exponential RLD achieves up to 350-830 times faster speed. In summary, this analysis concludes that RLD offers a better trade-off between precision and speed in single-exponential analysis than WLLS.

In [72], where overlapping gates were first proposed for RLD, the error analysis in [73] was repeated to compare overlapping and contiguous gates. In the overlapping scheme, the overlap between the gates was set to 50%. In single-exponential decay measurement with two gate positions, the results showed that for the optimal value of $\Delta t/\tau$, both configurations achieved similar precision. However, the overlapping gated configuration had a decisive advantage as $\Delta t/\tau$ increases, leading to the conclusion that it offers improvements in the lifetime range under fixed measurement parameters. This comparison was also performed for double-exponential decays with known and unknown lifetimes. For unknown lifetimes, the number of gate positions was increased to 4, due to the increase in the number of variables. The analysis was performed with only Monte Carlo simulation; error propagation was omitted due to its relative complexity for double-exponential models. It was demonstrated that contiguous gates were superior for known lifetimes and overlapping gates were superior for unknown lifetimes. These findings are consistent with the higher lifetime range of the overlapping gated scheme, which offers an advantage when gating settings are set independently of the lifetimes.

A similar error analysis was also performed for a generalized two-gated single-exponential RLD model with custom gate width and delay parameters, as illustrated in Figure 2:3(a) [187]. The motivation of this effort was to reduce the dependence of the RLD method on the prior information on the lifetime by devising a gating scheme with a wider lifetime range at the peak precision value. In this work, it was found that in the optimized gating scheme, gates overlap by 75% and the second gate has a width of 4-12 $\Delta t/\tau$, where Δt is the width of the first gate. Figure 2:3(b) [187] shows the precision of the optimized RLD compared to the standard RLD and weighted nonlinear least-squares (NLLS) methods. Optimized RLD achieves higher precision than standard RLD at all gate widths, and higher precision than NLLS at relatively lower gate widths.

The initial time gating experiments with the RLD approach predominantly used a gated ICCD, which captured gate images sequentially [192]. Subsequently, various techniques with the ability to acquire multiple gates in parallel were developed. This trend eventually culminated in the acquisition of the entire decay in a single image. The first implementation of this technique (also called “single-shot” FLIM) was done in [189] by projecting the original and delayed versions of the same image on two halves of a gated ICCD. This approach, which was chosen due to the lack of a parallel gating feature in an image intensifier, has several disadvantages. It decreases the spatial resolution by dividing the camera field of view into two halves, and increases the setup complexity due to high dependence on optical components. Furthermore, while it addresses the speed problem by reducing the acquisition time by 50%, it does not increase the duty cycle. A different approach with a segmented intensifier achieves single-shot FLIM at the expense of similar trade-offs [193]. Later, with the introduction of lock-in pixels which support up to four gating channels to this date, single-shot FLIM was performed with no spatial resolution penalty [109]. This method was subsequently demonstrated with gated SPAD arrays [184].

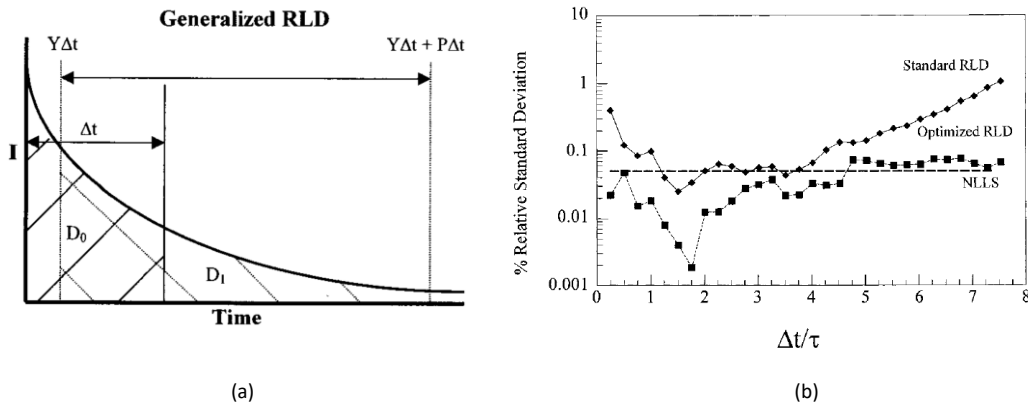


Figure 2:3 (a) A generalized single-exponential two-gated RLD scheme with custom gate width and delay parameters [187]. (b) Precision (relative standard deviation) comparison of the optimized RLD method with standard RLD and weighted nonlinear least-squares (NLLS) methods [187].

2.2.2 Time gating for the least squares method

The least squares method (LSM) is a fitting-based lifetime estimator used in time-domain FLIM since its early years. To this date, LSM is still considered among the methods that yield the best accuracy and precision [194]. This method is based on the fitting of the measured emission response to an exponential decay model after multiple iterations.

Conventionally, LSM is performed with the TCSPC method using single-channel scanning microscopy, owing to its high timing resolution. This combination, while still being considered as the “gold standard” in FLIM, has the characteristic speed limitations of both its data acquisition and data analysis methods, which have been discussed in Chapter 1, Section 1.3. Subsequently, time-gated LSM was developed as a method to implement LSM in widefield FLIM systems. This approach offers an alternative method to RLD in widefield FLIM, in which speed is sacrificed for state-of-the-art SNR performance.

The principle of this method is illustrated in Figure 2:4(a) [129]. The gates are configured to emulate TCSPC bins. Thus, the fluorescence decay is sampled using gates with equal widths. In most of its implementations, the gates are non-overlapping and separated. The gate width is often set to the minimum value permitted by the hardware specifications, in order to maximize the time-stamping precision. The gate positions, whose number is restricted due to frame rate concerns, are equally distributed across the measurement window.

This method conventionally uses ICCDs as the detectors of choice [53, 195]; it was the first established gated camera technology and it achieves the shortest gate width compared to its competitors. For this reason, multiple challenges in this method arise from the hardware limitations of ICCDs. Firstly, the number of gate channels is limited to one, which leads to a trade-off between duty cycle and photon economy. Secondly, the decay time of the phosphor screen, which is caused by the persistence of bright images for a duration of $\sim 300 \mu\text{s}$, introduces a frame rate limitation [53]. Instead of ICCDs, it is also possible to use lock-in pixels or SPAD cameras. However, the wider gates of these two technologies compared to ICCDs cause a drop in the overall photon economy.

The rules governing the required gate parameters for good SNR in this method are investigated in [196]. Even though it is theoretically possible to determine a single-exponential lifetime with only two gate positions [191], it was reported that the configuration that yields acceptable precision requires at least 5-10 gate positions. The purpose of this redundancy is to minimize random errors caused by factors such as the jitter of laser pulse and the gain factor of the detector. The minimum required measurement window was calculated

as 2.5τ for capturing $>90\%$ of the emission signal and achieving reasonably high SNR. In [195], where time-gated FLIM with an ICCD camera was performed for cancer detection, the LSM and RLD methods were compared using the same hardware. The LSM method requires 25 gate positions and achieves a refresh rate of 0.05-0.1 fps. On the other hand, the RLD method only requires 2 gate positions and reaches up to 10 fps. While the SNR of the RLD images is lower, it is generally adequate for cancer detection, which focuses on spatial lifetime contrast rather than the actual values of the lifetime. While this gating method yields acceptable results for single-exponential lifetimes, as demonstrated in the cancer detection application, its ability to resolve multi-exponential lifetimes is still inferior to conventional TCSPC-based LSM since the gate width of an ICCD camera is generally larger than the bin widths of state-of-the-art TCSPC systems [196].

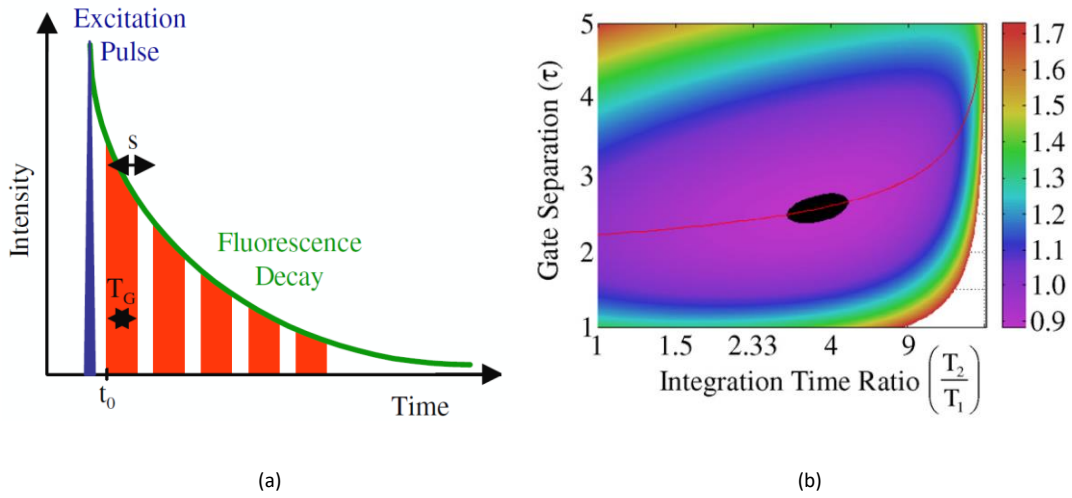


Figure 2:4 (a) Diagram of the time-gated FLIM acquisition method for histogram generation to be used in the least squares method [129]. Multiple gates with a width of T_g record the intensity of the fluorescence decay with a constant gate step s . In the standard acquisition method, all gates have equal integration times. For optimal SNR, the gates sampling the later parts of the decay must have relatively longer integration times than the earlier gates. (b) Plot showing the optimal integration time ratio and gate separation for a two-gated single-exponential lifetime analysis using the scheme in (a) [129]. The color bar represents the lifetime determination error of a measurement in which the two variables, integration time ratio and gate separation, correspond to the x and y coordinates of a data point, respectively. Integration time ratio is defined as the ratio between the integration duration (acquisition time) of gate 2 and gate 1. Gate separation defines the delay between the rising edges of two gates, in terms of the lifetime. The red curve shows the gate separation that yields the lowest error as a function of integration time ratio. The area in black corresponds to the set of integration time ratio and gate separation that yields the lowest error.

The quantitative performance analysis of this method was further investigated in [129]. While the width and the separation of the gates are kept constant throughout the decay, the work also investigates the scenario where the integration times of the gate positions are set to different values such that the photon counts of the gates have a smaller variation. Figure 2:4(b) [129] shows the distribution of the error in a two-gated configuration as a function of the integration (exposure) time ratio and the separation between the two gates. The plot shows that the highest precision is achieved at an integration ratio of around 4, and gate separation of $\sim 2.5\tau$. However, it should be noted that the optimal precision can only be reached in this method if the approximate value of the lifetime is known a priori, which is a disadvantage of this approach.

2.3 Our gating approach for phasor-FLIM

The goal of our research was to design a widefield FLIM system to perform time-domain phasor analysis. Phasor analysis, which is discussed in Chapter 1, Section 1.3, is based on the representation of a function in the frequency domain by a single point called a phasor. A phasor z , which represents only the sinusoidal component of the function at a single frequency value, is expressed in the complex plane as

$$z = g + is = me^{i\varphi}, \quad (2.5)$$

where g and s are the real and imaginary part of the phasor, and m and φ are its modulus and phase, respectively. Phasor analysis for FLIM can be applied to the fluorescence response acquired in both time domain and frequency domain, corresponding to pulsed and sinusoidally modulated excitation, respectively [77].

The standard time-domain phasor equations based on DFT (Equations (1.14) and (1.15)) only apply when the target function is measured at a constant sampling rate, and the measurement window includes most of the exponential decay area; for the measurement of truncated decays, additional conditions apply [86]. We therefore decided to scan the gate across the full laser period with equal steps, in order to be able to use the standard phasor method without additional phasor frequency restrictions. The measurement window D , defined in [74] for contiguous bins in the center of mass method as the delay between the rising edge of the first gate position and falling edge of the last gate position, was redefined in accordance with the principles of DFT as the product of the number of gate positions G and the gate step G_s , and selected as

$$D = G \times G_s = T_{lsr}, \quad (2.6)$$

where T_{lsr} is the laser period. The scanning was performed with overlapping gates, which are defined as gate positions whose corresponding time windows overlap with one another. Overlapping gates are generated when the delay between adjacent gate windows (gate step) is smaller than the gate width, which is expressed as $G_s < G_w$, as illustrated in Figure 2:5. The necessity to record sufficient photon counts at each gate position for decent SNR, which introduces a maximum gate step limitation, and the hardware-related limitations on the minimum gate width, were the two reasons for the selection of overlapping gates.

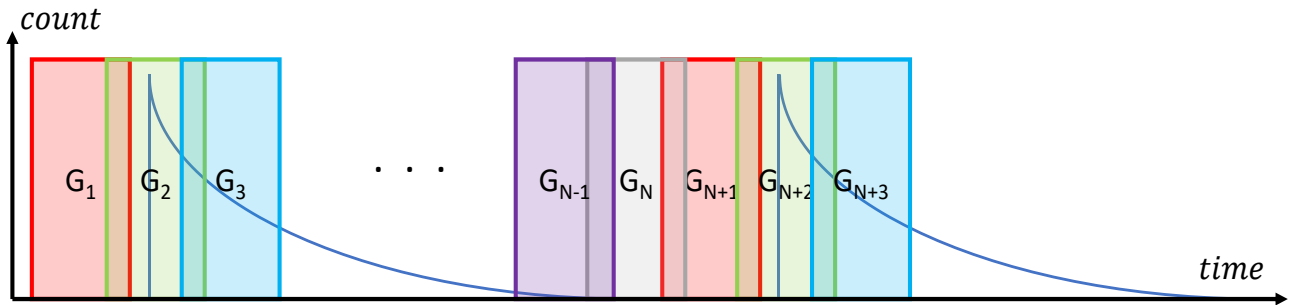


Figure 2:5 Conceptual diagram of overlapping gates, which are defined as gate positions whose corresponding time windows overlap with each other. Overlapping gates are generated when the delay between adjacent gate positions (gate step) is smaller than the gate width. Overlapping gates allow sampling a fluorescence decay with a short lifetime with sufficiently high photon count in each gate position for a decent SNR.

During phasor analysis, a nanotime was assigned to each gate, similar to the case with TCSPC bins. The nanotimes of the gates were defined as

$$t_i = G_s \times i, \quad (2.7)$$

where i is the gate position index, t_i is the nanotime of the gate i , and G_s is the gate step. The nanotime values are independent of the actual gate position; any offset added to t_i which equally applies to all values of i is indeed corrected during the calibration process via IRF deconvolution, which is discussed in Chapter 4, Section 4.2.

We used a laser with an adjustable pulse repetition frequency (PRF). The PRF of the laser was set to 20 MHz, which is the fixed value that is supported by the detector. The position of the first gate is set in such a way that its falling edge starts overlapping with the beginning of the decay only after several steps, as shown in Figure 2:6. This measure ensures that only one emission decay is used for the analysis.

The effects of the IRF on the phasors are “calibrated” by characterizing the phasor of the IRF at the frequency of interest using a reference sample with a known lifetime, and subsequently performing a simple algebraic correction of the measurements of the samples of interest, as discussed in Chapter 4, Subsection 4.2.1. This technique, which is explained in depth in Chapter 4, is only valid for the correction of effects that are independent of the lifetime; in particular, it cannot be used to correct the distortions caused by undersampling, as discussed in Section 2.4.5.

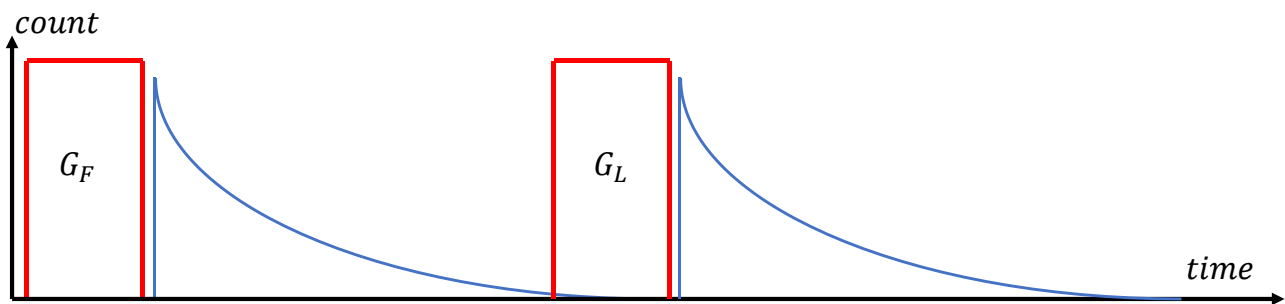


Figure 2:6 Position of the first and last gate in the data acquisition process. This positioning allows the detection of only one fluorescence decay, which makes the lifetime estimation method insensitive to any mismatch between the phasor frequency and laser PRF.

To collect sufficient photons in multiple gate positions for good precision and to minimize aliasing due to undersampling, the gate step is typically set to hundreds of picoseconds, which is significantly shorter than the gate width of at least ~ 12 ns.

While exposure times and bit depths – i.e., the number of bits used to store the photon count – of the gate positions can be adjusted, they are constant across all gate positions. Although it is demonstrated in literature that gate configurations with unequally distributed exposure times may yield superior SNR [193, 195], the dependence of the optimal configuration on the target lifetime and the number of gates complicates this optimization process; therefore, it is avoided in our approach.

Further details of this gating method are presented in Chapter 4.

2.4 Gate parameters

The characteristics of the gate influence the overall FLIM performance in multiple aspects including photon sensitivity, precision and lifetime range. These gate characteristics are represented by various parameters. The understanding of the gate parameters and their contributing factors is necessary for the assessment of the performance limitations of a gated FLIM system.

The gate window profile is defined as the normalized photon sensitivity of a pixel as a function of the delay of a laser pulse directed towards the pixel, with respect to an edge of the gate signal voltage. In other words, the gate window profile is equivalent to the pixel’s IRF for a particular gate width. By this definition, other than the specifications of the camera, the characteristics of the gate window profile are unique to the temporal profile of the laser pulse, which generally exhibits a narrow peak followed by a decaying tail, and the photodiode timing characteristics.

2.4.1 Gate width

The gate width (G_W) is defined as the full width at half maximum (FWHM) of the gate profile, as shown in Figure 2:7. The “maximum” is here replaced by the average value at the high level of the gate profile, which is equal to the sensitivity of the pixel when the gate is fully on, and is only limited by the QE/PDE of the photodiode.

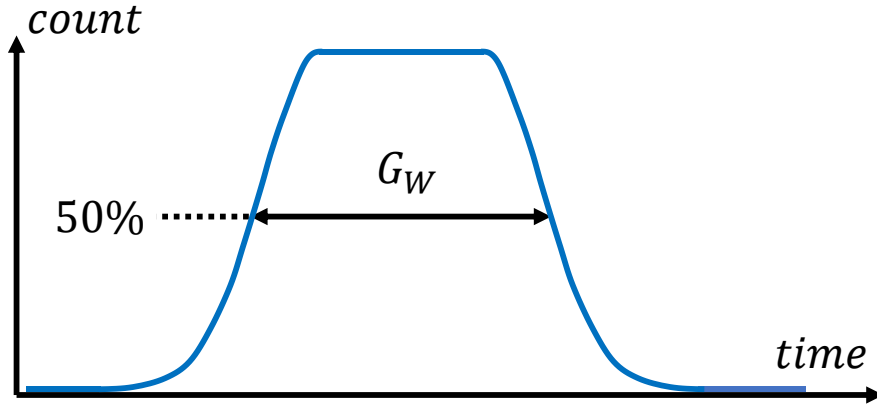


Figure 2:7 A diagram illustrating the gate width.

In time-resolved imaging, the gate width determines the photon time-stamping precision of the detector; therefore, it is analogous to the bin width of the TCSPC system. The uncertainty on the photon time-stamp measured with a square gate with width G_W is expressed as [81]

$$\sigma_W = \frac{G_W}{\sqrt{12}}. \quad (2.8)$$

This relation, which is only valid for ideal square gates with infinitesimal rise and fall times, is due to the fact that the variance of a uniform distribution between the intervals $[a, b]$ is equal to $1/12(b - a)^2$. On the other hand, for a given laser PRF, longer gates capture a larger percentage of the measurement window. Therefore, they allow the detection of more photons under the same experimental conditions.

In time-gated SPAD imagers, the minimum gate width is mainly determined by two parameters. Firstly, the power distribution characteristics of the gate signal network on the chip influence the minimum gate width. As the delay between the rising and falling edges of the gate window decreases, the possible superposition of the temporary supply voltage drops corresponding to multiple signal edges may drive the supply voltage below the critical level needed for the operation of the pixel transistors. This effect may temporarily disable the pixel operation itself, rendering it insensitive to photons. The distortion that this effect causes in the gate window shape is typically in the form of a fluctuation in the high sensitivity level of the gate, which is normally expected to be flat, as in Figure 2:7. The second limitation of the minimum gate width is introduced by the transition times. This limitation is described in the next subsection.

2.4.2 Transition time (rise and fall time)

The rise and fall times of a gate window profile are defined as the time delay between the 10% and 90% intensity levels of the rising and falling edge of the window, respectively. In Figure 2:8, the edges of the gate window are drawn as linear functions for simplicity. In reality, however, they have non-linear characteristics. In SPAD detectors, if the delay between the laser trigger edge and the avalanche current edge is considered

as a random variable, the gate window transition can be modeled as its cumulative distribution function, considering an ideal gating mechanism with no temporal uncertainty and instant switching time of the pixel sensitivity. By this definition, both the laser pulse shape and the SPAD jitter contribute to the transition time, and the characteristics of the rise and fall times are expected to be identical.

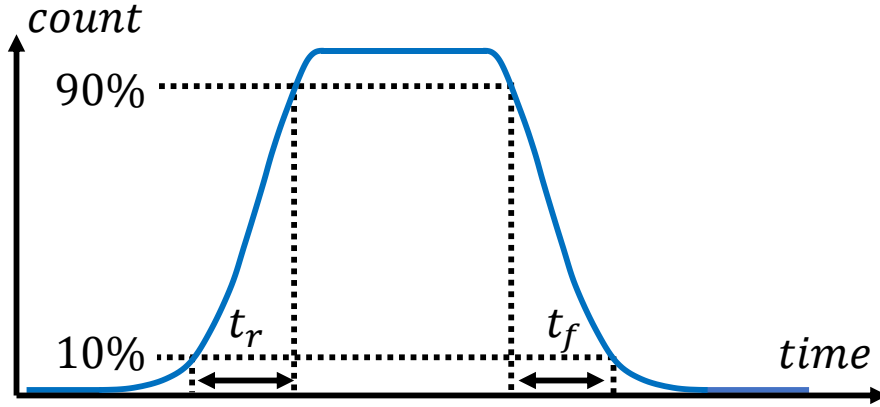


Figure 2:8 A diagram illustrating the rise and fall times of the gate window profile.

In a more realistic scenario, the gating mechanism also displays temporal uncertainty. The factors contributing to this uncertainty may include the following: the jitter between the laser trigger and the gate signal edge, the jitter between the gate signal edge and the gate transistor switching time, and the duration of the photodiode sensitivity change after the gate signal edge. As seen from the list above, the list of contributing factors related to the gating itself depends on the physical mechanism corresponding to the gate transition, as also reported in [120]. This phenomenon is the primary cause of a potential asymmetry between the rise and fall times of a gate window.

The rise and fall times of the gate introduce two effects on the timing performance of the system. Firstly, they determine the minimum width of a gate window. The shortest gate width ($G_{W_{min}}$) that can be achieved without reducing the peak sensitivity level is approximately equal to

$$G_{W_{min}} \cong \frac{1}{2}(t_r + t_f), \quad (2.9)$$

where t_r and t_f are rise time and fall time of the gate window, respectively. Secondly, the transition time affects the bandwidth of the IRF; shorter transition times indicate higher IRF bandwidth. The effects of the IRF bandwidth on the lifetime estimation performance are discussed later in the chapter.

2.4.3 Skew

Skew is a term that is often used in the distribution of clock signals in digital electronics. It is defined as the delay of one edge of a signal at various locations within the chip, as illustrated in Figure 2:9. The ideal value of the skew is zero, as the signal is intended to be transmitted synchronously to all its destinations. In time gated image sensors, the gate skew is defined as the maximum delay of a gate window edge between any two pixels in the array.

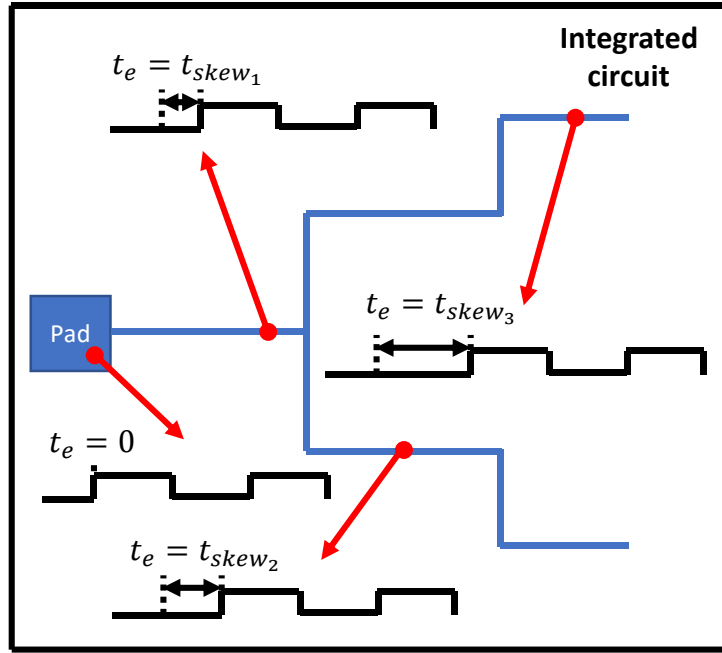


Figure 2:9 Conceptual illustration of skew in an integrated circuit. The propagation delay of the signal across the wires results in a delay between the signal edges (t_e) at different points of the signal distribution network.

The skew is primarily caused by the differences in the transmission of a signal. Any disparity in the resistance and capacitance of the transmission paths affects the propagation time of a signal along the path, thus causing a skew. The factors determining these characteristics of the paths include the physical properties of the metal wires, such as length, width and thickness; and the drivability of the active components such as inverters, buffers and logic gates. However, even if all of these parameters are designed to be equal, additional factors that arise from outside the network may also affect the skew. For instance, the lateral and vertical spacing between the signal wire and the other structures around them causes variations in the parasitic capacitance. In addition, temporary supply voltage drops due to the operation of the surrounding electronics may cause variations in the toggling speed of the buffers and inverters along the signal transmission paths.

The skew of a signal can be minimized by the use of signal distribution trees, which are designed to propagate a signal through multiple paths simultaneously. Several common signal distribution trees are discussed in Chapter 3, Subsection 3.3.2. Particularly large-format image sensors present challenges for the design of these trees, which are connected to each pixel. In monolithic image sensors, the implementation of an extensive signal tree inside the pixel array is challenging due to the large area requirement that it demands. Therefore, a widely used method to distribute the gate signal is to propagate it equally to each row or column through a signal tree, and distribute it to the pixels in a single direction using metal wires. While this method offers a good trade-off between the skew and fill factor, the skew along the direction of the distribution within the array cannot be eliminated. In time-gated imaging, the gate skew causes the different pixels to sample the decay at slightly different time delays in a single frame. This factor limits the ability for different pixels to capture spatially correlated photons. Most applications of FLIM do not include correlation measurements, and are therefore not affected by this limitation. In the imaging of fast phenomena, the skew of the beginning and the end of the exposure times of a frame becomes a critical specification, due to the demand for global shutter operation to minimize artifacts. However, the frame acquisition durations are typically several orders of magnitude greater than the skew of the gate window. In this case, the skew does not significantly limit the ability to capture fast phenomena.

The skew also introduces a deterministic temporal offset between the IRFs of different pixels, which, unlike the previous effects, does have adverse consequences for FLIM performance. For instance, as will be discussed later, the IRF deconvolution for FLIM needs to be performed at pixel level to minimize systematic errors in the lifetime estimation. This calibration stage requires the calculation and storage of IRF parameters for each pixel, as opposed to a single parameter set representing all pixels. This measure increases the computational complexity of the calibration process and introduces demanding data storage capacity during the lifetime calculations.

2.4.4 Gate step

The gate step indicates the value of the time delay between two adjacent gate positions in a measured gate response profile. The gate step is illustrated for non-overlapping separated gates and overlapping gates for in Figure 2:10(a) and (b), respectively. In discrete Fourier transform (DFT), the gate step determines the sampling frequency (f_s); f_s is equal to the inverse of the gate step. The maximum allowed value of the gate step is determined by the bandwidth of the emission signal. According to the Nyquist-Shannon sampling theorem, the requirement for the sampling frequency is expressed as

$$f_s \geq 2f_c, \quad (2.10)$$

where f_c is the cutoff frequency of a bandlimited function. If this requirement is not met, aliasing occurs, causing distortions in the locations of the phasors, as shown in Figure 2:12(c) [85].

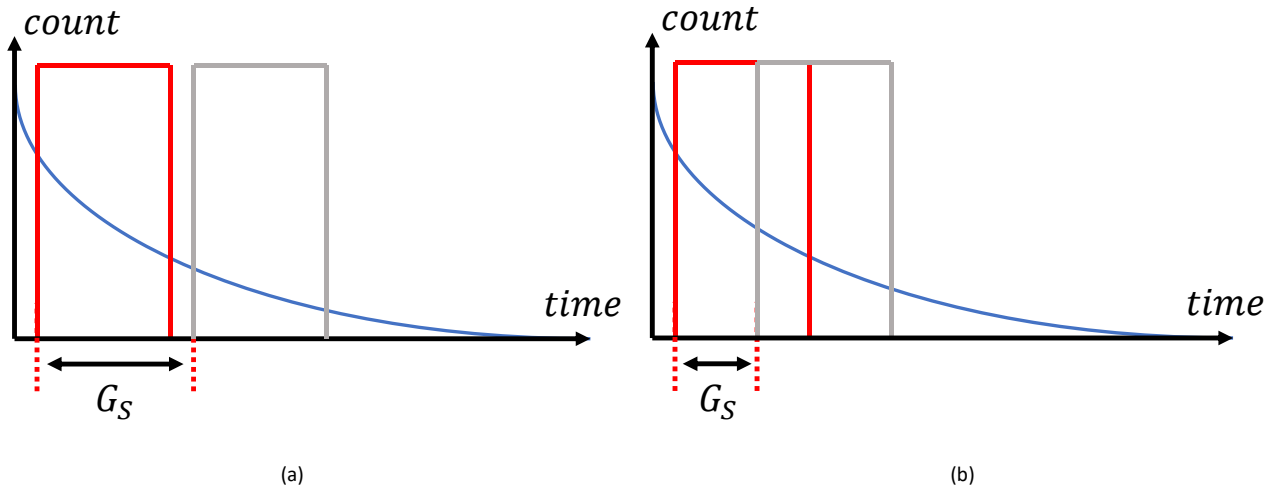


Figure 2:10 The description of gate step in (a) non-overlapping and (b) overlapping gates.

In state-of-the-art CMOS technologies, it is possible to delay a signal by steps as small as ~ 20 ps. However, as $G = T_{lsr}/G_S$, such a step choice requires up to thousands of gate positions for phasor analysis at typical laser PRF values.

2.4.5 Number of gate positions

In time-gated FLIM, the parameter that has one of the most direct effects on the acquisition time is the number of gate positions. For a given binary frame rate and gate image bit depth, the acquisition time is directly proportional to the number of gate positions.

Since time-domain phasor analysis includes the discrete Fourier transform (DFT) of an exponential decay, it requires the sampling of the decay with equal gate steps across all gate positions, as discussed in Section 2.3.

In the cases where the emission intensity does not decay significantly across the laser period, the required data acquisition approach to minimize truncation is to scan the entire laser period by selecting the measurement window according to Equation (2.6). Under this condition, the number of gate positions is determined by the PRF and the gate step, not the gate width. This rule applies to both overlapping and non-overlapping gates, as the gate width is not a parameter that influences G in Equation (2.6).

The number of gate positions also influences the trade-off between the acquisition time and the number of detected photons. As shown in Figure 2:11, in time-gated acquisition, where gate positions with uniform steps cover the entire laser period in accordance with Equation (2.6), the total integration time $T_{int_total}(t)$, which is a function of the delay relative to the laser trigger, t , is expressed as

$$\begin{aligned} T_{int_total}(t) &= T_{int_gate} \times \left\lfloor \frac{G_W}{G_S} \right\rfloor, & \text{for } LE - t > G_W \bmod G_S \\ T_{int_total}(t) &= T_{int_gate} \times \left\lceil \frac{G_W}{G_S} \right\rceil, & \text{for } LE - t < G_W \bmod G_S, \end{aligned} \quad (2.11)$$

where T_{int_gate} is the integration time of a gate, defined as the light exposure duration during the data acquisition of the gate, G_W is the gate width, G_S is the gate step, LE is the closest leading gate edge delay to t in the gate shift direction, and “mod” is the modulo operation. By expressing G_S as $G_S = D/G$, Equation (2.11) can be rewritten as

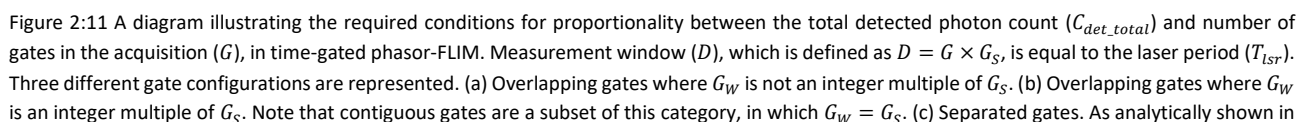
$$\begin{aligned} T_{int_total}(t) &= T_{int_gate} \times \left\lfloor \frac{G_W \times G}{D} \right\rfloor, & \text{for } LE - t > G_W \bmod G_S \\ T_{int_total}(t) &= T_{int_gate} \times \left\lceil \frac{G_W \times G}{D} \right\rceil, & \text{for } LE - t < G_W \bmod G_S. \end{aligned} \quad (2.12)$$

This relation assumes ideal square gates, and is valid for separated, contiguous and overlapping gates. The number of detected photons during the entire measurement, C_{det_total} , is then expressed as

$$C_{det_total} = \int_{t=0}^D T_{int_total}(t) \times CR(t) dt, \quad (2.13)$$

where $CR(t)$ is the photon count rate of the delay t . When $\frac{G_W \times G}{D} \in \mathbb{N}$, as the floor and ceiling functions of an integer are equal, Equation (2.13) can be rewritten as

$$C_{det_total} = T_{int_gate} \times \frac{G_W \times G}{D} \int_{t=0}^D CR(t) dt. \quad (2.14)$$



Equations (2.11)-(2.14), C_{det_total} is proportional to G independently of the measured function only if $(G_W \bmod G_S) = 0$. In this theoretical analysis of an ideal square gate, pile-up effect and temporal changes in the fluorescence response are neglected.

Equation (2.14) proves that the total detected photon count is directly proportional to the number of gate positions, when sufficient photon count is achieved to minimize statistical variations, irrespective of the characteristics of the measured function, e.g. single or multi-exponential decay, as long as $\frac{G_W \times G}{D} \in N$ for all values of G that are selected. This condition is met only when $(G_W \bmod G_S) = 0$, which encompasses all contiguous gates ($G_W = G_S$) and overlapping gates in which the gate width is an integer multiple of the gate step, and excludes separated gates. However, for overlapping gates in which $G_W \gg G_S$, C_{det_total} can be considered almost proportional to G due to the small relative difference between $\left\lfloor \frac{G_W}{G_S} \right\rfloor$ and $\left\lceil \frac{G_W}{G_S} \right\rceil$. From Equation (2.14) and the proportionality between the data acquisition time and G , the trade-off between the total data acquisition time of the decay and the number of detected photons can be confirmed. This concept is illustrated in Figure 2:11, in which the integration time uniformity of photon delays across the measurement window is shown for the aforementioned gate configurations.

The reduction of the number of gate positions can only be made possible by an increase in the gate step or a decrease in the measurement window, both of which may introduce distortions in the resulting discrete Fourier transform. Depending on the bandwidth of the system IRF, the former option may result in undersampling, thus causing a drop of lifetime precision, as discussed in Subsection 2.4.4. On the other hand, the latter option may lead to an effect called “truncation” when the measurement window excludes a significant percentage of the emission response function area [85, 86].

The effects of undersampling and truncation in time-gated phasor-FLIM were investigated in [85] and [86]. In this work, the Fourier transform of a single-exponential decay was reformulated based on discrete sampling using ideal square-shaped contiguous gates with equal width and uniform steps across the measurement window. The Fourier transform of a single-exponential decay, which is expressed in the continuous domain as

$$z(\tau, f) = \frac{1}{1 - i2\pi f\tau}, \quad (2.15)$$

was rewritten for contiguous gates with widths shorter than the lifetime as

$$z'(\tau, f) = \sum_{m=-\infty}^{\infty} \frac{1}{1 - i\left(2\pi f - 2\pi \frac{mG}{T}\right)\tau} (-1)^m, \quad (2.16)$$

where m is the gate index, f is the phasor frequency and G is the number of gate positions. In this approach, the center of a gate was selected as its nanotime. The authors concluded that the magnitude of the discrete Fourier transform deviates from the continuous Fourier transform in a certain range of frequencies as shown in Figure 2:12(a-b) [85]. In these figures, the effects of undersampling and truncation can be observed. Frequency-shifted repetitions (aliases) of the ideal continuous frequency spectrum have a periodicity of $f_s = G/D$. On the other hand, the ripples with a periodicity of $1/D$ are due to truncation; the truncated frequency spectrum matches with the continuous spectrum only at $f = 1/D$ and its harmonics. The latter effect is more visible in Figure 2:12(b) in which the distortions due to aliasing are negligible at low harmonics of $1/D$, compared to Figure 2:12(a).

Both effects lead to distortions in the phasor locations, which ultimately result in differences between the SEPL and the universal semicircle, which are undesirable for the reasons discussed in Subsection 2.4.4. These differences increase with the magnitude of the distortion caused by undersampling and truncation. A decrease in the number of gate positions is expected to result in larger modifications in the SEPL. This effect is illustrated in Figure 2:12(c) [85], where the SEPL corresponding to $G = 4, 8$ and 16 are plotted against the standard universal semicircle, which corresponds to $G = +\infty$.

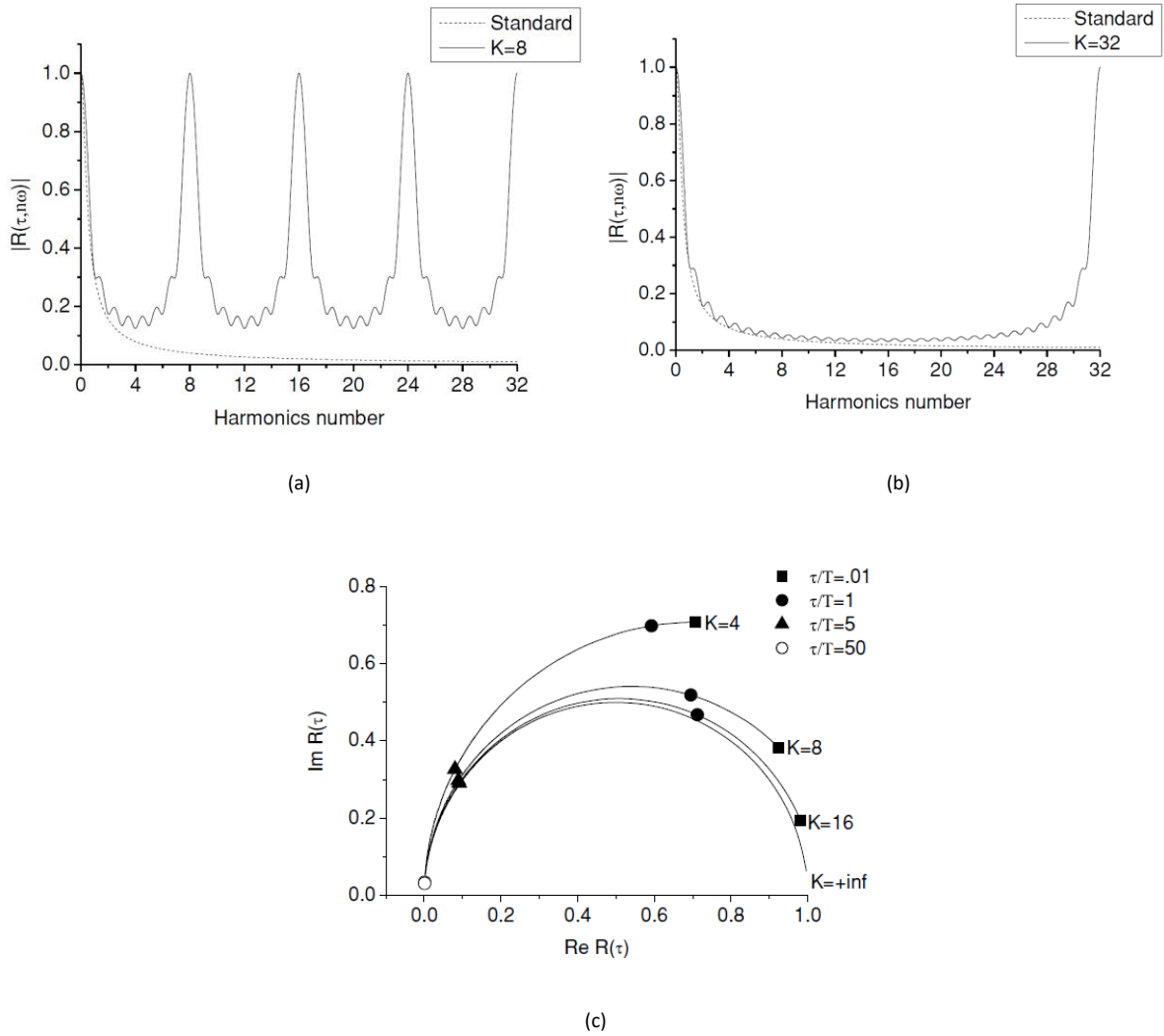


Figure 2:12 (a-b) The Fourier transform of a single-exponential decay function sampled with (a) 8 and (b) 32 contiguous ideal square gates [85]. The curve in dotted line shows the continuous Fourier transform of the function. (c) Modified phasor plot where the SEPL is recalculated from the modified version of the single-exponential Fourier transform for contiguous gates with widths shorter than the lifetime [85]. The number of gate positions, which is denoted by K in these figures, is denoted by G in the thesis.

Unlike aliasing, truncation does not distort the phasor locations at the harmonics of $1/D$. This property can be exploited to reduce the number of gate positions via truncation and still avoid phasor location distortions. In [86], the effects of decay truncation on the phasor location were theoretically investigated and the optimal phasor frequency was proposed for truncated decays. For a given laser period T_{lsr} and measurement window D , with $D < T_{lsr}$, the phasors of single-exponential decays are located on the standard universal semicircle only if the phasor frequency f is equal to $f = n/D$.

In [85], the performance of long and short gates was compared using double-exponential mixtures, as well. The separability index (S) was used as the figure of merit for this comparison. This index, which indicates the separation level of two lifetimes and was introduced in [197], is defined by:

$$S = \frac{|\tau_2 - \tau_1|}{\sqrt{|\text{var}(\tau_2) - \text{var}(\tau_1)|}}. \quad (2.17)$$

In this comparison, it was concluded that the separability of the method increases with a larger number of gates, as theoretically expected. However, it must be noted that this modified phasor method analyzes double exponential decays using the “global analysis method” [198]. This method assumes that the values of the two lifetime components of the sample are the same throughout the field of view, and only the preexponential factors of the decays may be different. This assumption makes the analysis of lifetime possible, since the data of one pixel is relevant for the neighboring pixels, and dramatically increases the statistics and accuracy [199].

2.4.6 Number of gate channels

The number of gate channels indicates the number of gates which can acquire data in parallel. There are two major advantages of having multiple gate channels. Firstly, it is possible with more than one gate channel to make the system sensitive to photons during the entire emission window. The fraction of the time during which the sensor is photon sensitive, also called the duty cycle, or temporal aperture, is a particularly important parameter for the measurements of biological systems with limited photon budget. Secondly, it enables the acquisition of the entire decay in a shorter period of time. This feature decreases the errors in the measured decays of samples with fast dynamics or high photobleaching rate. When the number of gate channels is equal to the number of gate positions, single-shot FLIM can be performed. On the other hand, there are also several limitations on the number of gate channels. Firstly, an increase in the number of channels causes fill factor degradation and an increase in the throughput demand of the readout configuration. Secondly, any gate profile non-uniformity between the channels causes systematic errors in the measurements, thus requiring more complex calibration and correction processes.

In the ICCD-based early generation time-gated FLIM systems, the number of channels was limited to one, since the gate is implemented by modulating the photodiode. In fully solid-state cameras where the gates are generated by readout electronics, including lock-in CMOS pixels and SPAD image sensors, multiple gate channels are possible. In state-of-the-art time-gated detectors, the number of gate channels ranges from 1 to 8 [18]. This number is expected to increase further with the advances in the capability of solid-state detector technologies.

2.5 Performance limitations of time-gated phasor-based FLIM using SPADs

Today, the performance of FLIM systems is influenced both by data acquisition and analysis methods, as well as by the characteristics of the detector technology. Since widefield phasor-based FLIM has only recently been demonstrated using time-gated SPAD imagers, a discussion on the performance of such systems from multiple perspectives is of key importance. The major performance parameters of a FLIM system include photon sensitivity, photon economy, lifetime range and dynamic range.

2.5.1 Photon sensitivity

Photon sensitivity is an important parameter for samples with low photon budgets, since in those cases lower detector sensitivity cannot always be compensated by higher excitation power. Photon sensitivity in SPAD imagers is represented by photon detection efficiency (PDE). PDE is the ratio between the number of detected photons and the number of incident photons on the entire pixel area. The relation between the PDE and its components, photon detection probability (PDP) and fill factor (FF), is given by

$$PDE = PDP \times FF. \quad (2.18)$$

The fill factor indicates the ratio between pixel active area, i.e. the area where the electric field is sufficient for impact ionization upon photon absorption, and the entire pixel area. The PDP is the probability that a photon incident on the active area generates an avalanche current, which in turn is expressed as

$$PDP = P_{av} \times QE, \quad (2.19)$$

where P_{av} is the avalanche probability and QE is the quantum efficiency, which indicates the probability of an impinging photon to generate a free electron (or hole). As shown in these equations, the only contributor of the PDE unaffected by photodiode physics is the fill factor. The complexity of the pixel electronics has a major influence on fill factor. In gated pixels, the number of gate transistors in the pixel is often proportional to the number of gate channels. Therefore, a trade-off exists between the number of gate channels and the PDE of a SPAD imager, for a given pixel size.

A second measure of photon sensitivity is duty cycle. The duty cycle is the fraction of the time during which the sensor is photon sensitive. In gated imaging with non-overlapping square gates whose period is equal to the laser period T_{lsr} , the relation between duty cycle (Y) and gate parameters can be expressed in rolling-shutter mode as

$$Y = \frac{G_W \times N_c}{T_{lsr}}, \quad (2.20)$$

where G_W is the gate width and N_c the number of gate channels. In large-format SPAD imagers, the smallest gate width to ensure 100% duty cycle is around 6.25 ns, provided that the number of equal-width gate channels rarely exceeds 2, the laser PRF usually does not exceed 80 MHz, and the gate period is equal to T_{lsr} . In these systems, the largest laser PRF is determined by the requirement of the laser period being several times greater than the maximum lifetime of interest, in order to avoid a significant overlap of the emission signals from two adjacent decays, for convenience.

Unlike PDE or QE, the duty cycle is not always proportional to the percentage of missed photons in the emission profile. In the approach where a decay is scanned with gates with equal widths, a period of T_{lsr} , and uniform gate steps, the ratio between detected photon count (C_{det}) and impinging photon count (C_{imp}) in a pixel, under negligible pile-up effect and when $T_{lsr} \gg \tau$, is given by

$$\frac{C_{det}}{C_{imp}} = PDE \times \frac{1}{G} \sum_{k=1}^G \frac{I_k}{I_{decay}}, \quad (2.21)$$

where k is the gate position index, G is the number of gate positions, I_k is the total intensity of all gate channels in gate position k , and I_{decay} is the total intensity of the emission profile. When the measurement window D is a multiple of T_{lsr} , Equation (2.21) can be re-written as

$$\frac{C_{det}}{C_{imp}} = PDE \times \frac{G_W \times N_c}{T_{lsr}} = PDE \times Y, \quad (2.22)$$

where Y is the duty cycle. Equations (2.21) and (2.22) show that the proportionality between the duty cycle and the percentage of missed photons is only present when D is a multiple of T_{lsr} . Note that for a gate sequence which contains multiple channels with a fixed delay between the channels, G denotes the number of positions of the entire gate sequence. The value of I_k is the sum of the intensity of all gate channels in the position k , given by

$$I_k = \sum_{j=1}^{N_c} I_{k,j}, \quad (2.23)$$

where j is the gate channel index, N_c is the number of gate channels, and $I_{k,j}$ is the intensity of the gate channel j at position k . Equations (2.22) and (2.23) apply to both overlapping and non-overlapping gates. For a system with multiple gate channels, Equation (2.22) is valid only if the total measurement window of all channels does not exceed the laser period. The terms that are used in Equations (2.22) and (2.23) are illustrated in Figure 2:13, for a system with two non-overlapping parallel gate channels.

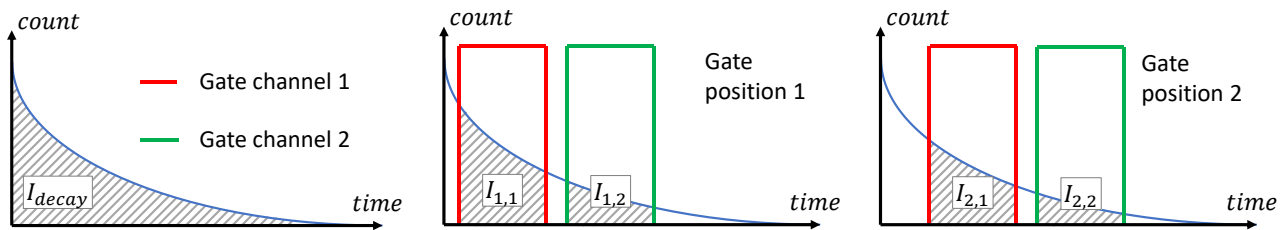


Figure 2:13 Illustration of various terms used in the calculation of C_{det}/C_{imp} in Equation (2.22), for a system with two non-overlapping gate channels, where $T_{lsr} \gg \tau$.

2.5.2 Photon economy

Photon economy is a measure of precision of a FLIM system for a given number of detected photons. Photon economy can become a critical parameter in several scenarios. Firstly, if the sample has a low photon budget due to photobleaching, the lifetime estimation must be performed with a limited photon count. Secondly, for operation modes with high frame rates, the number of photons per frame is often limited by the photon count rate of the sample, which also poses a similar challenge. In such scenarios, where the number of impinging photons is limited by the aforementioned causes, and the percentage of the detected photons is limited by the photon sensitivity of the detector, the precision of the estimation is mainly determined by photon economy.

The most broadly used parameter to indicate photon economy is the F value [44, 188, 200]. This parameter, which compares the precision of the lifetime determination of a system with the shot noise limited value, is expressed as

$$F = \frac{\sigma_\tau/\tau}{\sigma_N/N}, \quad (2.24)$$

where σ_τ is the standard deviation of the measured lifetime, τ is the true value of the target lifetime, σ_N is the standard deviation of the total number of detected photons, and N is the mean of the total number of detected photons. As shown in Equation (2.24), F value expresses the relative standard deviation of the lifetime in terms of the relative standard deviation of the number of photons, which is one of major contributors to the measurement uncertainty. In a system where the variance of the number of photons is shot-noise limited, the standard deviation of the photon counts equals the square root of the photon count itself. Therefore, Equation (2.24) can be rewritten as

$$F = \frac{\sigma_\tau \sqrt{N}}{\tau}. \quad (2.25)$$

When the standard deviation of the estimation (σ_τ) is inversely proportional to the square root of the detected photon count, the F value remains independent of the photon count. The variance of an exponential distribution whose probability density function is $\frac{1}{\tau} e^{-t/\tau}$, is equal to τ^2 . When the mean of N samples is calculated, its variance and the standard deviation are equal to τ^2/N and τ/\sqrt{N} , respectively, according to the law of large numbers. When this expression of the standard deviation is plugged in Equation (2.25), the F value is equal to 1. This proves that, if the photon emission process displays Poisson statistics, then F cannot be below 1. The F value is higher for detectors which have additional contributors to the variance of the lifetime other than the contribution of the shot noise for the exponential distribution. In a system with a non-zero gate width, the standard deviation of a single-exponential lifetime and the F value can be approximated as

$$\sigma_\tau(G_W, N) = \frac{\tau}{\sqrt{N}} \sqrt{1 + \frac{1}{12} \left(\frac{G_W}{\tau} \right)^2} \quad (2.26)$$

and

$$F_{G_W}(\tau) = \sqrt{1 + \frac{1}{12} \left(\frac{G_W}{\tau} \right)^2}. \quad (2.27)$$

This first-order estimation, derived in [81], assumes that the photon arrival times are uniformly distributed within a gate, which becomes increasingly inaccurate as G_W/τ increases. From this relation, it can be seen that the photon economy depends on the ratio between gate width and lifetime.

The F value can increase due to additional factors, such as the errors in data analysis methods [74], uncertainties in the calibration [81] and systematic errors. In experiments where the IRF is characterized by performing a reference measurement using a sample of known lifetime, statistical dispersion of the reference (i.e. calibration) and sample lifetimes are statistically independent. As a consequence, the variance of the final estimation can be approximated by the sum of the variances of the two estimations. From this relation, the standard deviation of the measured lifetime after IRF deconvolution can be written as

$$\tilde{\sigma}_{\tau_i}(G_W, N_i) = \sqrt{\sigma_{\tau_i}^2(G_W, N_i) + \sigma_{\tau_{cal}}^2(G_W, N_{cal})}, \quad (2.28)$$

where τ_i is the lifetime of the single-exponential sample of interest, τ_{cal} is the lifetime of the single-exponential calibration sample, N_i is the number of photons in the measurement of the lifetime of interest, and N_{cal} is the number of photons in the measurement of the calibration sample [81]. Plugging this expression in Equation (2.25), the F value is given by

$$F = \frac{\tilde{\sigma}_{\tau_i}(G_W, N_i)\sqrt{N}}{\tau}. \quad (2.29)$$

The most effective way to minimize the contribution of the calibration to the F value is to perform the calibration with as many photons as possible. In SPAD detectors, this can be done by choosing a calibration sample with low photobleaching rate and increasing the bit depth of the gate positions. This method maximizes N_{cal}/N_i , which causes $\tilde{\sigma}_{\tau_i}$ to be dominated by the term σ_{τ_i} .

It should be noted that the F value is a measure of precision, not accuracy. Therefore, it is affected by the statistical dispersion of the lifetime estimation, but not by systematic errors. For this reason, the F value can be less than one in some cases where the dispersion in the photon counts or the time-stamping is reduced due to quantization error. In SPAD detectors, this phenomenon occurs during the operation in the pile-up regime, where the magnitude of the error increases with the level of the pile-up. As the camera operation enters deeper into the pile-up regime, the drop in the photon count dispersion due to truncation becomes more severe. As a result, the F value gradually decreases, eventually converging to zero when the camera is fully saturated. This phenomenon is discussed in further detail in the following subsections.

2.5.3 Lifetime range

In a FLIM system, the lifetime range can be defined as the range between the maximum and minimum lifetimes that can be estimated with an acceptable accuracy and precision. Unlike the dynamic range, there is a lack of standardization on the exact definition of the lifetime range, particularly in the quantitative definition of the “acceptable” accuracy and precision, which is a key requirement for its calculation. A quantitative analysis of the lifetime range for various FLIM methods was presented in [74]. The method used to determine the lifetime range in this work is as follows. Firstly, the target precision is defined for a given recorded photon count (N) as¹

$$Precision = \frac{\tau_{meas}}{\sigma_{\tau_{meas}}}, \quad (2.30)$$

where $\sigma_{\tau_{meas}}$ is the standard deviation of the measured lifetime τ_{meas} . While a modified precision definition which includes accuracy (defined and discussed in more detail in Subsection 2.5.5) was also provided, the standard definition can also be used to determine the lifetime range, given that the errors leading to low accuracy can in principle be corrected due to their deterministic nature [74, 201]. For perfect accuracy which results in $\tau_{meas} = \tau_{real}$, the F value can then be defined in terms of precision as

$$F = \sigma_{\tau_{meas}} \frac{\sqrt{N}}{\tau_{real}} = \frac{\sqrt{N} \times \tau_{meas}}{Precision \times \tau_{real}} = \frac{\sqrt{N}}{Precision}. \quad (2.31)$$

¹ Conventionally, precision can be defined as either $Precision = \frac{\tau_{meas}}{\sigma_{\tau_{meas}}}$ or $Precision = \frac{\sigma_{\tau_{meas}}}{\tau_{meas}}$, both of which are commonly used expressions.

In [74], the maximum acceptable F value for $Precision = 40 \text{ dB} = 10^{(40/20)} = 100$ and $N = 2^{17}$ was estimated as $F_{max} \approx 4$. After the calculation of F_{max} , the lifetime range is determined from the plot of F vs. τ corresponding to the particular FLIM method/system as the set of τ values for which $F < F_{max}$, as illustrated in Figure 2:14.

As photon economy is the primary contributor of the lifetime range, the factors that influence photon economy, as discussed in Subsection 2.5.2, also influence the lifetime range. These factors include gate width, the number of detected photons in the calibration stage, and potentially the phasor frequency.

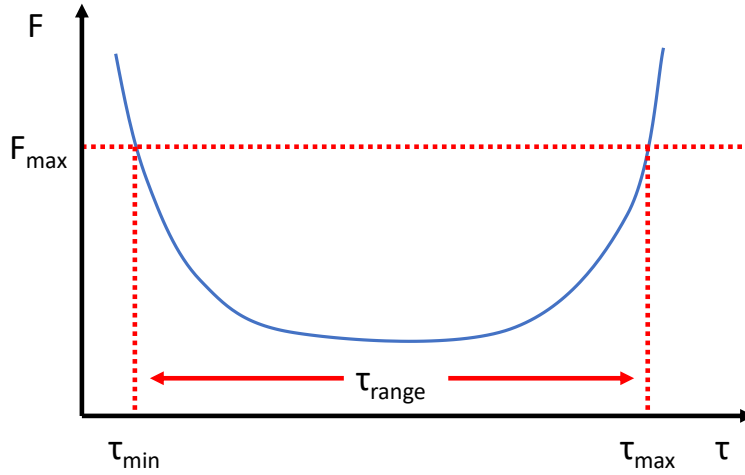


Figure 2:14 A diagram illustrating the lifetime range of a FLIM method. The blue curve is the F value distribution of a particular FLIM method for a range of lifetimes. The lifetime range, τ_{range} , is defined as the difference of the maximum and minimum lifetimes which can be estimated by the method with an F value no higher than F_{max} , denoted by τ_{max} and τ_{min} , respectively. F_{max} is a user-defined value based on minimum required photon economy of a particular application, which can also be derived by the target precision and photon count (N) values, from the relation expressed in Equation (2.31).

2.5.4 Dynamic range

Dynamic range is defined as the ratio between the maximum and minimum number of photons that can be detected within a single frame. Mathematically, this definition can be expressed as

$$DR = 20 \log \left(\frac{n_{i \max}}{n_{i \min}} \right), \quad (2.32)$$

where $n_{i \max}$ and $n_{i \min}$ are the maximum and minimum detectable counts, respectively [202]. High dynamic range is required in gated FLIM for several reasons. Firstly, due to the exponential distribution of the emission profile, the signal intensity at multiple gate positions can vary by up to several orders of magnitude. Secondly, in widefield FLIM, it is common to have a large spatial variation in the fluorescence intensity due to various factors including excitation illumination non-uniformity and differences in concentration, depth or quantum yield of the fluorophores. If the sensor does not have sufficient dynamic range, a sample with a large spatial and temporal intensity variation can only be captured in multiple stages under different exposure settings, resulting in a speed penalty. Finally, within the readout speed limitations of the detector, the dynamic range indicates the maximum number of photons that can be collected for the analysis of a decay. Since the precision of the lifetime determination is influenced by the total photon count, the dynamic range can be also considered as a critical parameter for the precision of FLIM.

The determination of $n_{i\max}$ highly depends on the architecture of the imager. In conventional CMOS imagers, where the maximum number of detected photons is determined by full well capacity, the measured photon count m_i increases linearly until full electron or hole well capacity (EWC or HWC) is reached. Therefore, $n_{i\max}$ is defined as $n_{i\max} = EWC$. On the contrary, in SPAD imagers, due to the absence of such a fixed maximum photon count limit, $n_{i\max}$ is estimated by the photon count which yields the maximum acceptable signal-to-noise ratio (SNR) deviation from its corresponding shot-noise-limited value, as discussed in detail in [202] and summarized below. In this case, the estimated detected photon count is denoted by \tilde{n}_i .

In large-format digital SPAD imagers, a frame is constructed by the sum of the counts of multiple binary frames, whose exposure start and stop times are typically controlled by a global clock signal (active clock-driven recharge), as opposed to photon counts themselves (event-driven recharge). In an active clock-driven recharging architecture, with gate frame bit depth of b , the number of binary frames is equal to $2^b - 1$, which corresponds to the maximum range for the measured photon count (m_i). The maximum value of detected photon counts (n_i), which is defined as the photon count that would have been measured by the camera in the absence of pile-up, can be estimated (\tilde{n}_i) from the following equation [203]:

$$\tilde{n}_i = -\ln\left(1 - \frac{m_i}{2^b - 1}\right)(2^b - 1). \quad (2.33)$$

According to Equation (2.33), the ratio of the missed photons, expressed as $1 - (m_i/n_i)$, increases with n_i , with n_i , as m_i converges to a saturation value. This relation between m_i and n_i is shown in Figure 2:15 for SPAD imagers with active clock-driven recharge, where the saturation value of m_i is equal to $2^b - 1$ [203]. This non-linear relation, which often allows a SPAD sensor to have higher dynamic range than a conventional image sensor with an electron well [204], is caused by the pile-up effect. The pile-up effect describes the phenomenon in which, when the interarrival times of the photons are shorter than the overall dead time of the imaging system, the photon that arrives later is missed. This causes a distortion of the photon timing statistics, leading to both systematic and random errors in the ultimately estimated parameters.

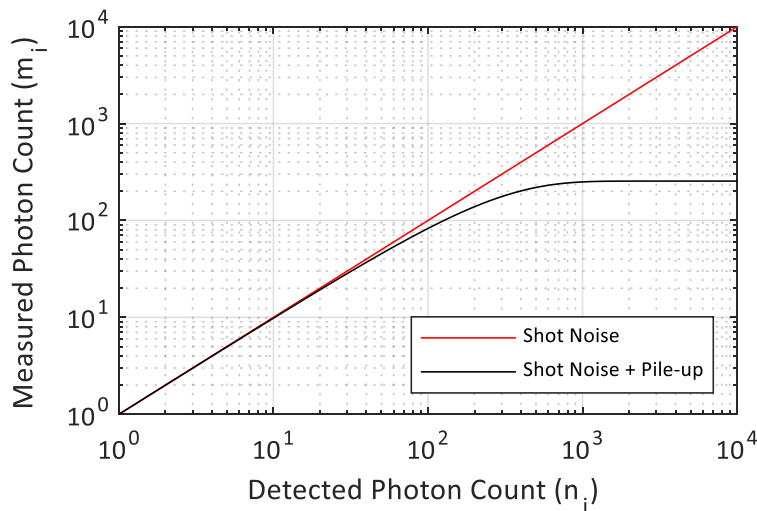


Figure 2:15 The theoretical relation between the detected and measured photon counts in a SPAD imager with an active clock-driven recharge and 8-bit output image format. The red curve represents an ideal shot-noise-limited system with no pile-up or saturation mechanism, where the measured and detected photon counts are equal. The black curve represents a SPAD imager with the aforementioned characteristics, in which the pile-up effect introduces an upper bound to the maximum detectable photon count. The analytical expression for the black curve was retrieved from [203], and expressed in Equation (2.33).

While the systematic error due to pile-up can be corrected using Equation (2.33), the reduced recorded photon statistics leads to an SNR drop due to increased uncertainty, whose deviation from the shot-noise level increases non-linearly with n_i . Using the error propagation method, $\sigma_{\tilde{n}_i}$ can be expressed as a function of σ_{m_i} as [202]

$$\sigma_{\tilde{n}_i} = \sigma_{m_i} \frac{\partial \tilde{n}_i}{\partial m_i}. \quad (2.34)$$

From Equation (2.33), $\partial \tilde{n}_i / \partial m_i$ can be expressed as

$$\frac{\partial \tilde{n}_i}{\partial m_i} = \frac{1}{1 - \frac{m_i}{2^b - 1}}. \quad (2.35)$$

In addition, σ_{m_i} is expressed as a function of n_i for active clock-driven recharge-based SPAD imagers as [202]

$$\sigma_{m_i} = \sqrt{(2^b - 1)(1 - e^{-\left(\frac{n_i}{2^b - 1}\right)})e^{-\left(\frac{n_i}{2^b - 1}\right)}}. \quad (2.36)$$

As $\sigma_{\tilde{n}_i}$ is larger than $\sqrt{\tilde{n}_i}$, as will be shown in Chapter 3, Subsection 3.4.5, it can be inferred that pile-up leads to an SNR drop with respect to the shot-noise level SNR. Using Equations (2.34)-(2.36), the theoretical SNR of the detected photon count, SNR_{n_i} , can be expressed as

$$\text{SNR}_{n_i} = 20 \log \left(\frac{n_i}{\sigma_{n_i}} \right) = 20 \log \left(\frac{n_i \left(1 - \frac{m_i}{2^b - 1} \right)}{\sqrt{(2^b - 1)(1 - e^{-\left(\frac{n_i}{2^b - 1}\right)})e^{-\left(\frac{n_i}{2^b - 1}\right)}}} \right). \quad (2.37)$$

In [202], the estimations of $n_{i \max}$ and $n_{i \min}$, denoted by $\tilde{n}_{i \max}$ and $\tilde{n}_{i \min}$, respectively, were derived for active clock-driven recharge-based SPAD imagers, based on the difference between the SNR in Equation (2.37) and the shot-noise-level SNR. $n_{i \max}$ was estimated for imagers with 100% duty cycle as

$$\tilde{n}_{i \max} = F_{cs} \times T_{int} / T_{dead}, \quad (2.38)$$

where F_{cs} is the count saturation factor, which represents the relation between n_i and m_i for a recharging mechanism, T_{int} is the frame integration (exposure) time, and T_{dead} is the pixel dead time [202]. F_{cs} represents the impact of pile-up on SNR, which depends on the recharging mechanism of the SPAD imager and the minimum acceptable SNR drop from shot-noise level. The values of F_{cs} for three different SNR criteria and three different SPAD recharging mechanisms are provided in [202].

Unlike $n_{i \max}$, $n_{i \min}$ is defined irrespective of the detector architecture as the minimum photon count which yields at least 0 dB SNR (i.e. $n_{i \min} = \sigma_{n_{i \min}}$). Therefore, $n_{i \min}$ of a SPAD imager, in which dark counts dominate the overall noise excluding shot noise, is estimated as

$$\tilde{n}_{i \min} = \sigma_{DC} = \sqrt{T_{int} \times DCR}, \quad (2.39)$$

where σ_{DC} is the standard deviation of the dark count and DCR is the dark count rate (in the order of 10 cps/pixel for SwissSPAD2/3).

The SNR of the frames (i.e. gate photon counts in FLIM) that is used to calculate the dynamic range of a gate image is different from the SNR of the lifetime determination, which is the main indicator of the lifetime

determination precision. The development of an analytical expression linking these two parameters is non-trivial, and is not discussed here.

Further discussion on dynamic range in SPAD imagers and trade-offs between dynamic range and other FLIM performance parameters is available in Chapter 4, Section 4.1.

2.5.5 Accuracy

Accuracy is defined as the relative proximity of a measurement to its actual value. The quantitative expression of lifetime estimation accuracy is given by

$$A = 100 \times \frac{|\Delta\tau|}{\tau_{real}} = 100 \times \frac{|\tau_{meas} - \tau_{real}|}{\tau_{real}}, \quad (2.40)$$

where τ_{meas} and τ_{real} are measured and real values of the lifetime, respectively. In contrast to precision which defines the statistical dispersion of a measurement, accuracy represents the reproducible, systematic errors.

In time-gated phasor-FLIM, where the fluorescence response is recorded at multiple gate positions, accuracy is determined by the factors that affect the recorded photon count of different gate positions non-uniformly. In other words, accuracy is primarily influenced by correlated noise sources. In the context of this discussion, correlated noise is defined as the noise that results from a signal whose magnitude as a function of time is correlated to the magnitude of the signal of interest as a function of time.

Pile-up, which is discussed in Subsection 2.5.4, is considered to be one of the most important sources of correlated noise in FLIM for SPAD imagers. Pile-up is defined as the loss of photons whose delay from a previously detected photon is lower than the dead time of the system. In SPAD imagers with active clock-driven recharge, the ratio between the photons detected by the SPAD² (n_i) and measured by the pixel (m_i) increases with n_i , as shown in Equation (2.33). Therefore, the measured response is compressed by a larger factor at the gates with high photon counts. This effect flattens the decay curve, thereby causing the apparent lifetime to be higher than the actual lifetime. The correction of pile-up for SPAD imagers with different recharge mechanisms was discussed in [202], and was generally applied to the imagers presented in this thesis.

There are several other correlated noise sources related to the imager, that could affect accuracy. Crosstalk, which is defined as the detection of a photon that impinges on a pixel by a nearby pixel, results in the superposition of the decay profiles of the target pixel and its neighboring pixels, the latter weighted by the crosstalk percentage of each neighboring pixel [205, 206]. The widening of the spatial pixel response due to crosstalk is represented by the pixel response function (PRF) [207], which contributes to the overall pixel response s along with the point spread function (PSF) based on the relation [208, 209]:

$$s(r; x) = PSF(x) * PRF(r; x), \quad (2.41)$$

where $*$ indicates the spatial convolution product of the PSF and PRF, r and x are the pixel index and the optical spot position, respectively, in a single dimension. Crosstalk can be compensated by the characterization and deconvolution of the function s in a reference measurement. However, owing to the low crosstalk levels of state-of-the-art SPAD imagers, crosstalk often does not introduce significant levels of accuracy loss

² This term indicates the number of photons that would have been detected by the SPAD in the absence of dead time. Its value is equal to the product of the number of impinging photons on the SPAD and the PDE.

for FLIM, rendering the correction of its effects unnecessary. The crosstalk characteristics of SwissSPAD2 are presented in Chapter 3, Subsection 3.4.3.

Another factor that can affect accuracy is the drift of DCR and PDP during the detector operation due to heating. This problem can be overcome by starting the data recording after the stabilization of the sensor temperature, or by the characterization and correction of the detector sensitivity as a function of time during the detector operation. The former method creates an idle time window at the start of the detector operation, although its impact may be insignificant in relative terms for long operation times. On the other hand, the latter method is complicated by the fact that the PDP and DCR in SPADs may be affected differently by temperature.

Lifetime determination accuracy in FLIM can also be affected by factors related to the scene. An important factor in this category is autofluorescence, when it is not the signal of interest. Autofluorescence is defined as the fluorescence that occurs due to the intrinsic properties of a biological structure [210]. The intensity of autofluorescence is correlated to the fluorescence signal of interest, and varies with the gate position, making it a correlated noise source. Autofluorescence presents one of the most significant challenges in *in vivo* imaging, as the tissue in front of the target cell/protein often exhibits autofluorescence [211]. In the phasor method, autofluorescence can be characterized by measuring the autofluorescent sample without the sample of interest. Subsequently, the autofluorescence signal can be subtracted from the measured signal that contains both autofluorescence and the signal of interest. Another possible cause of accuracy is photobleaching, which is the loss of fluorescence intensity of the sample due to high excitation power [45]. This is a particularly severe drawback of time-gated analysis with a single gate channel, which requires sequential acquisition of the multiple gates comprising a data set.

Due to their deterministic properties, all of the sources of accuracy can be corrected, if properly characterized [74, 201]. However, it must be noted that most of the aforementioned noise sources include shot-noise components, which cause a loss in precision that cannot be recovered by the accuracy correction methods.

Chapter 3 SwissSPAD2/3: Large-format time-gated SPAD imagers

The material presented in this chapter is based on [117].

3.1 Motivation

In recent years, SPAD imagers have gained popularity in time-resolved imaging applications, thanks to their sub-ns temporal resolution at room temperature, single-photon counting ability, low voltage operation and compatibility with integrated electronics. However, SPAD imagers still suffer from performance limitations which pose serious obstacles to their widespread adoption. Particularly in life sciences, the most important obstacles include the trade-off between photon sensitivity and spatial resolution, and high DCR. In time-gated architectures, which are common for large-format SPAD imagers, an additional drawback is the trade-off between photon economy and duty cycle, due to limited number of gate channels.

To overcome these drawbacks of SPAD technology, SwissSPAD was developed by Burri *et al.* in 2011 [120]. With its 512×128 pixels, SwissSPAD achieved the highest spatial resolution in SPAD arrays at the time. However, its native fill factor (5%), modest combined PDP/DCR performance, and high crosstalk (~3.5%) required improvements to make this architecture more attractive for biomedical applications. In our research, we aimed to design the first SPAD imager with more than 100 kpixels, using a SPAD model with a verified superior performance in terms of PDP and DCR. We also aimed to maximally exploit the high PDP and operating voltage insensitivity of the SPAD model by achieving high excess bias, in order to ensure high spatial pixel sensitivity uniformity. Finally, we aimed to overcome the limitations of SwissSPAD in terms of photon economy and duty cycle, by improving the lowest gate width and number of gate channels, respectively.

SwissSPAD2 and SwissSPAD3 are two SPAD image sensors designed to achieve these targets. While sharing the same SPAD design, pixel size and process technology, and having a similar spatial resolution, these two imagers exhibit different properties in pixel architecture and gate characteristics.

3.2 SPAD structure

The SPAD design used in SwissSPAD2 and SwissSPAD3 was developed by Veerappan *et al.*, and presented in [212]. Its structure is illustrated in Figure 3:1. The main junction of the SPAD is a vertical p-i-n junction, as opposed to p-n junctions of conventional SPAD designs. The low-doped epitaxial (epi) layer placed between the p-well and buried n-well generates a high depletion region width in the junction, which brings several advantages. Firstly, the distribution of the electric field across a large depth results in a wide multiplication region when operating in Geiger mode. Since the photons absorbed in a wide range of depths in the main junction have a high probability of impact ionization, the variation of the absorption coefficient with the photon wavelength has a reduced influence on the PDP. As a consequence, these devices have a wider spectral response than the designs with narrow junction widths [213]. Secondly, the distribution of the electric field

across a wide depletion region also decreases the rate of tunneling noise, thereby reducing the total dark count rate [214]. The guard ring was generated by using p-well lateral diffusion and a low-doped p-epi layer outside the p-well. Since the doping concentration of the p-well declines gradually towards the edges, the electric field of the lateral junction is lower than the electric field of the vertical junction, which prevents premature edge breakdown.

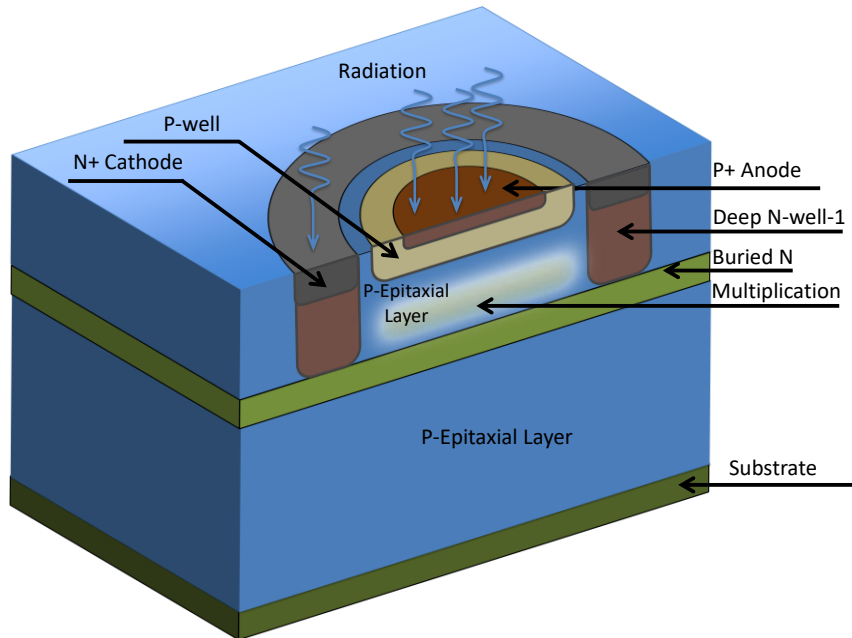


Figure 3:1 Cross-section of the p-i-n SPAD structure used in SwissSPAD2 and SwissSPAD3 [212].

An additional advantage of this SPAD model is the low sensitivity of the PDP to excess bias voltage variations, as will be seen in Subsection 3.4.1. Towards the higher values of the excess bias voltage range in which the device can operate, the PDP reaches a compression point, in which the sensitivity of the PDP to excess bias sharply declines. This feature presents an important advantage in imaging sensors with high spatial resolution, such as SwissSPAD2 and SwissSPAD3. In such large-format arrays, the excess bias voltage can have spatial and temporal variations due to two main reasons. Firstly, a pixel-to-pixel breakdown voltage variation can lead to a spatial excess bias voltage variation since all SPADs are supplied with a common operation voltage at their cathode terminals. Secondly, fluctuations can occur in the SPAD operation voltage in situations where a high percentage of the SPADs generates an avalanche simultaneously. An important reason for these fluctuations is the lack of high-density decoupling capacitors available in the CMOS process for the voltage range of the cathode. The MOS capacitors, the highest density capacitor type in the available process, cannot operate above 3.3 V due to oxide breakdown.

While wide spectral response and high PDP are clearly desirable features for a SPAD, a high operation voltage requirement for achieving high PDP is a disadvantage, particularly in large-format imagers. The two main reasons for avoiding high operation voltages are high power consumption, and fill factor penalty due to the shielding requirements of the wires carrying higher voltages. These effects may introduce a trade-off in the SPAD and imaging sensor design processes between photon sensitivity, power consumption and fill factor. As high current consumption makes the supply voltage more vulnerable to IR drop (voltage drop that occurs during current flow through a resistor), it requires a larger supply wire width for the reduction of parasitic resistance. The fill factor penalty resulting from the large wire width and higher capacitive coupling risk due to high voltage [215] can be mitigated by placing the high voltage wire next to the ground wires or the wires transmitting

the signals that are least vulnerable to capacitive coupling. Considering this trade-off, two variants of the device were designed by Veerappan *et al.* [212] with different p-well depths and doping concentrations, offering two different combinations of spectral width and breakdown voltage. The peak PDP wavelengths of these two models are also different as this parameter depends on the electric field distribution across the junction depth. Despite its wider spectral response, variant 1 exhibits excessive DCR at even moderate levels of excess bias, possibly due to unexpected tunneling noise contributions from the p+ contact layer. Therefore, variant 2 was used in our imager design.

3.3 SwissSPAD2/3: Design evolution

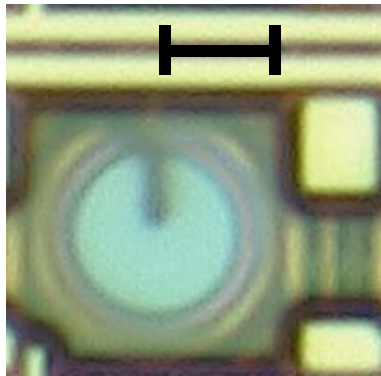
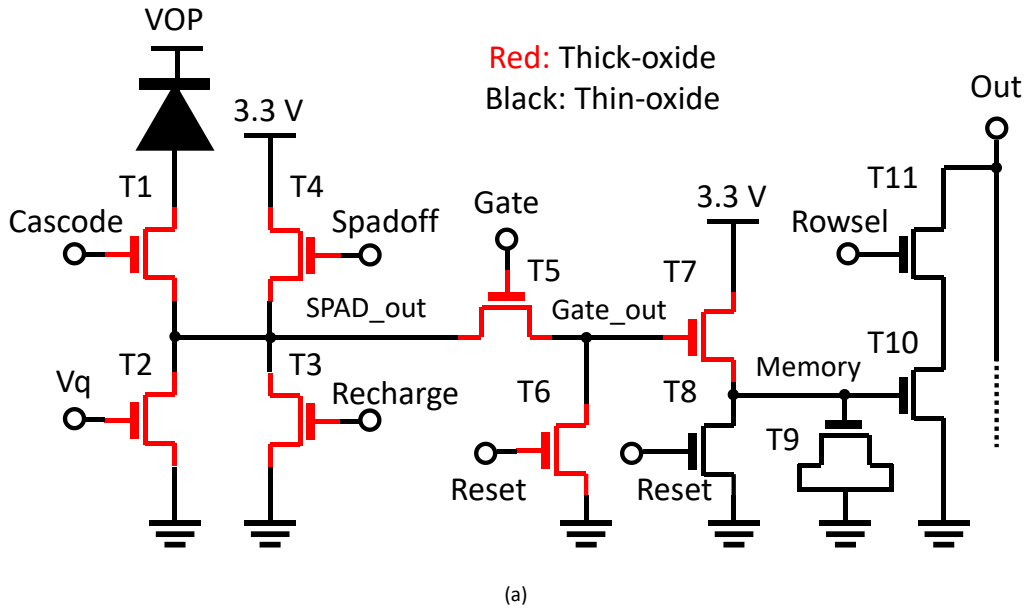
The SwissSPAD2/3 series was designed as time-gated widefield SPAD imagers for time-resolved imaging applications. Inspired by the architecture of its predecessor SwissSPAD [120], SwissSPAD2 aimed to improve its various specifications such as spatial resolution, fill factor and maximum excess bias voltage. Another goal of this effort was to mitigate several device-related shortcomings of SwissSPAD. Subsequently, SwissSPAD3 was designed to extend the duty cycle of SwissSPAD2 to 100% by implementing two gate channels, and to achieve superior gate characteristics through improvements in the gate signal tree architecture and the power distribution network layout. Both chips were designed in 2D front side illumination (FSI) 0.18 μm CMOS process technology, using the SPAD design described in Section 3.2 [212].

3.3.1 Pixel architecture

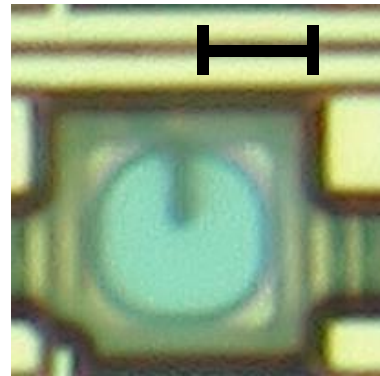
In SwissSPAD2, two pixel variants were designed using a round SPAD and a square SPAD with rounded corners, whose photomicrographs are shown in Figure 3:2(b) and (c), respectively. The pitch of this pixel is 16.38 μm in both versions, with a native fill factor of 10.5% and 13% for the round and square variants, respectively. The 3 μm SPAD active area radius was selected considering the targeted spatial resolution and the trade-off between the size and fill factor of the SPAD. This trade-off arises from the fact that the width of the inactive layers of the SPAD including the guard ring does not scale with the width of the active area at the center. In this pixel, the minimum allowed pixel pitch for the chosen SPAD dimensions was fully determined by the minimum required spacing between adjacent SPADs. No additional fill factor penalty was introduced by the placement of transistors or wires in the pixel.

The pixel schematic of SwissSPAD2 is shown in Figure 3:2(a). The pixel comprises 11 NMOS transistors, including both thick and thin oxide transistors. The functions of the transistors are as follows. $T1$ is a cascode transistor which increases the excess bias voltage range of the pixel, thus increasing the maximum achievable PDP [216]. $T2$ is responsible for passive quenching; quenching speed can be controlled by tuning voltage V_q , which determines the on resistance of $T2$. The pixel employs clock-driven active recharge. In gated mode, recharging is performed by $T3$ with the enabling of the *RECHARGE* signal before each gate opening. In intensity mode in which the gate is permanently open, recharge is performed by $T6$ with the enabling of *RESET* after the readout of the pixel. The *SPADOFF* signal, which controls $T4$ reduces the excess bias voltage of the SPAD while the gate is closed, thus reducing afterpulsing and power consumption caused by avalanche events that occur outside the gate window. $T5$ is the gate transistor which ensures that the memory only detects the avalanche events that occur inside the gate window. $T7$ is a source follower that prevents premature reset of the memory when the SPAD is recharged before each gate window. $T6$ and $T8$ are reset transistors that reset the *Memory* and the *Gate_out* nodes, after the readout of the pixel. $T9$ increases the capacitance between the *Memory* node and ground, which is required for *Memory* to hold the logic high state from the photon

detection until the pixel readout. T_{10} and T_{11} are responsible for the readout of the pixel, when signal *ROWSEL* is asserted.



(b)



(c)

Figure 3:2 (a) Pixel schematic of SwissSPAD2. The pixel consists of 11 NMOS transistors, 7 with thick-oxide and 4 with thin-oxide gate. The pixel stores a binary photon count in its memory capacitor. The in-pixel gate defines the time window, with respect to a 20 MHz external trigger signal, in which the pixel is sensitive to photons. (b-c) Pixel micrograph of SwissSPAD2 with (b) round SPAD and (c) square SPAD with rounded corners. Scale bar: 5 μm .

The required size of T_9 for holding the logic high state for 1 ms was determined using simulation tools. In the simulation, the *Memory* node was first charged through T_7 . Subsequently, T_7 was turned off and the voltage in *Memory* was left to decrease through the leakage current at T_8 , whose gate terminal voltage was set to 0 V throughout the simulation. After 1 ms, *ROWSEL* was asserted for 20 ns, which is the output bus (*OUT* node) pulldown duration required for the 97.7 kfps target binary frame rate operation. After *ROWSEL* was de-asserted, *OUT* voltage was registered by a D flip-flop. If the resulting output signal of the flip-flop is low, the photon detection is considered successful. For the reliability of the design, the size of T_9 was selected in such a way that this simulation yields successful results at all four process corners, i.e. taking into account the fabrication-related transistor performance variation.

The width and spacing of signal and power wires in the pixel were selected in accordance with several performance and space requirements. The wires of the three signals with critical timing constraints, *RECHARGE*,

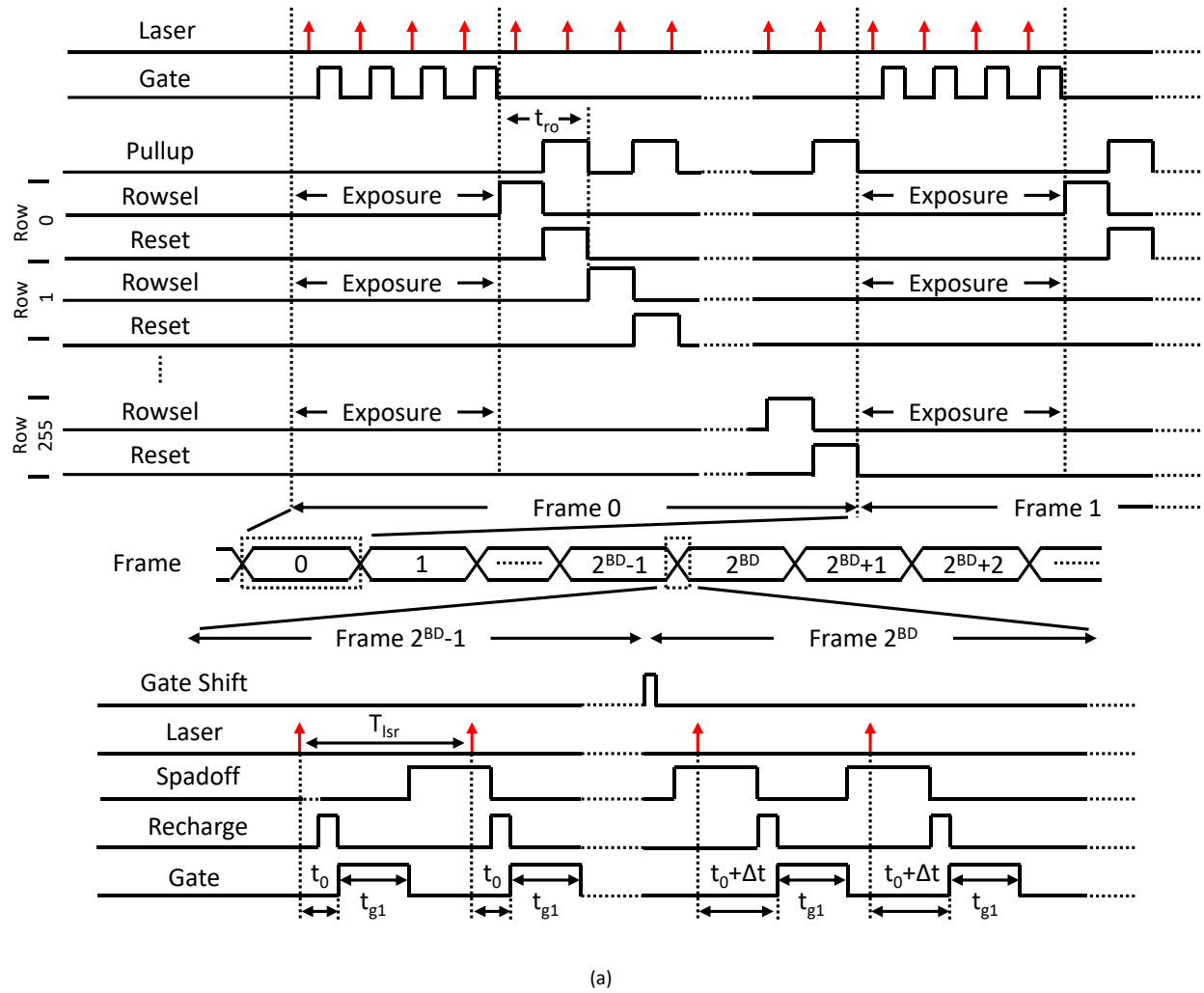
GATE and *OUT*, were sized and separated with the goal of minimizing their propagation times across the column. The relation between the wire properties and the signal propagation time is simulated based on the distributed RC model. In the simulation, each pixel was modeled as a segment of the RC chain with identical parasitic resistance and capacitance. The parasitic resistance and capacitance of the wires in a pixel were extracted from the pixel layout using standard extraction tools in the CAD software. Within the area constraints of the pixel layout, the width of the wires and the spacing between two adjacent wires were chosen in such a way the product of the parasitic resistance and capacitance is minimized, which in turn minimizes signal propagation time. Aside from signal propagation speed, an additional constraint for the placement and sizing of the wires was introduced by optical requirements. The constraint was to have no physical obstruction for the photons impinging on the SPAD active area with an angle of incidence up to 45° . This angle was selected to make the sensor compatible with lenses with a wide range of numerical apertures (*NA*), including a relatively large error margin to compensate for possible deviations of the metal heights from the available stack information.

SwissSPAD2 pixel can operate both in global and rolling shutter modes³. The timing diagrams of the SwissSPAD2 pixel in global and rolling shutter mode operations are shown in Figure 3:3. Figure 3:3(a) illustrates the global shutter operation. During the exposure window, which starts and ends in the entire array simultaneously, the gate window opens in a synchronized fashion with respect to the laser pulse. Subsequently, in the readout window where the gate is fully closed, the *Rowsel* and *Reset* signals of the rows are activated sequentially. Between *Rowsel* pulses of the adjacent rows, the global *Pullup* signal is activated to reset the output bus to high voltage. After the readout of the last row, which is Row 255 in one half of the array, the exposure for the next frame starts. The readout speed of a row, t_{ro} , is equal to the sum of the pulldown and pullup times of the output bus, which are indicated by the width of the *Rowsel* and *Pullup* signals, respectively. To operate the camera in the intensity mode in global shutter, the gate can be fully opened during the exposure window, and fully closed in the readout window. The details of the gate operation in the gated mode are illustrated at the bottom part of the diagram. After each sequence of $2^b - 1$ binary frames⁴, where b is the bit depth of a gate position count, the delay between the laser pulse and the entire gate sequence is shifted by Δt , which represents the gate step (G_S) in FLIM operation.

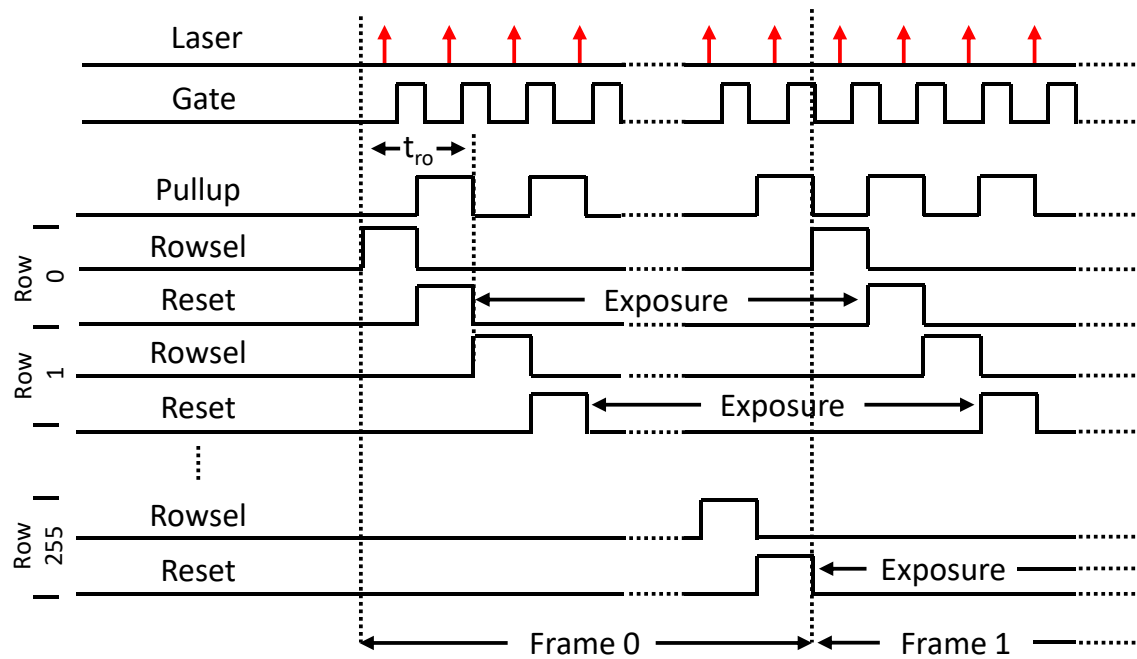
Figure 3:3(b) illustrates the rolling shutter operation. In this operation, exposure and the readout occur simultaneously. The readout structure is similar to the operation described above; the only difference is that the readout of Row 0 starts immediately after the readout of Row 255. The exposure of a pixel starts and ends with two adjacent *Reset* pulses. Rolling shutter allows higher frame rate and duty cycle compared to global shutter. On the other hand, its drawback is the presence of distortions in the images of dynamics faster than the frame readout duration ($\sim 10 \mu s$), since the pixels across the array capture different time frames.

³ In global shutter mode, the exposure time of each pixel starts and ends simultaneously. In rolling shutter mode, the start and stop times vary between rows or columns. The main advantage of rolling shutter is a larger duty cycle, since readout and exposure can occur simultaneously. On the other hand, the main advantage of global shutter is the absence of artifacts between different rows/columns during the imaging of fast phenomena.

⁴ In the currently available firmware, 1,020 binary frames were used to construct a 10-bit image instead of 1,023, due to hardware constraints.



(a)



(b)

Figure 3:3 Timing diagram of the operation of SwissSPAD2. (a) Global shutter: upper panel: gate window and readout signal sequence for two consecutive binary frame time windows. Lower panel: gate shifting mechanism during the transition between two b-bit frames. (b) Rolling shutter: gate window and readout signal sequence for two consecutive binary frame time windows. The gate shifting mechanism described in (a) also applies here.

The pixel of SwissSPAD3 shares similar characteristics with SwissSPAD2 in terms of process technology, SPAD properties and pixel pitch. The SPAD in SwissSPAD3 is identical to the round SPAD used in SwissSPAD2. The reason for using the same SPAD dimensions as SwissSPAD2 was to minimize the probability of its failure, as the main target of this imager design was to improve the pixel-level performance. Since the pixel pitch was the minimum value allowed by the spacing requirements of the selected SPAD size, it also remained constant. An additional advantage of not changing the SPAD and pixel dimensions was the compatibility of SwissSPAD2 microlens design for SwissSPAD3.

SwissSPAD3 mainly differs from SwissSPAD2 in its gate properties. While SwissSPAD2 has only a single gate channel, SwissSPAD3 employs two contiguous gate channels, which cover the entire exposure period. Its readout scheme, while having the same architecture as SwissSPAD2, also has two channels for reading the two gate channels sequentially. The schematics of the SwissSPAD3 pixel is shown in Figure 3:4. The pixels comprise 14 transistors, excluding 2 transistors which are used as decoupling capacitors between power supply and ground. The higher transistor density requirement due to the addition of an extra gate channel to the pixel could only be achieved by redesigning the blocks that occupy the largest pixel area in SwissSPAD2. As previously mentioned, the memory transistor ($T9$) in SwissSPAD2 occupies the largest area in order to minimize the voltage decay speed of the memory due to leakage. Reducing the size of this transistor without suffering from premature memory reset would only be possible by also reducing the leakage current of the reset transistor. This was achieved by replacing the thin oxide reset transistor ($T8$ in SwissSPAD2) with a thick oxide NMOS ($T9$ in SwissSPAD3). The transistor $T7$ in SwissSPAD2, whose function was to prevent the discharging of *Memory* when the SPAD is recharged, was eliminated in SwissSPAD3. The reason for this is that in SwissSPAD3, the SPAD is not recharged after photon detection until the pixel readout. Since the elimination of $T7$ removes one intermediate node (*Gate_out*), the transistor $T6$ in SwissSPAD2 was eliminated, and $T3$ was repurposed as a reset transistor after readout, in SwissSPAD3. $T12$ is a thick-oxide transistor, since the high-level voltage of *Out_1* is above 1.8 V, the operation voltage of a thin-oxide transistor in this technology. $T12$ also acts as memory for *Out_1* and holds the voltage until the end of the frame.

In order to ensure that the two gates are contiguous in all pixels across the array, the second gate was implemented as the difference between the unmodulated SPAD output (*Out_1*) and the gated SPAD output (*Out_2*). The relation between the gate outputs and the pixel outputs can be expressed as

$$Gate_1 = Out_1 - Out_2, \quad (3.1)$$

and

$$Gate_2 = Out_2. \quad (3.2)$$

Furthermore, *Out_1* is also equal to the total intensity of the binary frame. The calculation of *Gate_1* from the pixel outputs is performed using Equation (3.1) outside the chip.

The operation of the SwissSPAD3 pixel is as follows (see again Figure 3:4). The exposure of the pixel starts with the assertion of *Reset*. When a detected photon generates an avalanche, *Out_1* reaches high voltage. Upon the activation of *Out_1*, *Gate_M* is set to zero by the NOR gate which is implemented by $T4 - T7$. If the previous state of *Gate_M* is high during the rise of *Out_1*, *Out_2* is set to high voltage through $T8$ before $T8$

is turned off by the NOR gate. Otherwise, *Out_2* remains low. The state of *Gate_M* before the avalanche is determined by *Gate*. When *Out_1* is zero, the state of *Gate_M* is the opposite of *Gate*, since a NOR gate behaves as an inverter when one of its inputs is zero. After the detection of the first photon, the values of both *Out_1* and *Out_2* are fixed until the pixel readout; therefore, the pixel is insensitive to all subsequent photons. The behavior of *Out_1* and *Out_2* upon photon detection by the SPAD is illustrated in Figure 3:5. In the figure, the photons highlighted in green and red represent the photons detected and missed by the pixel, respectively.

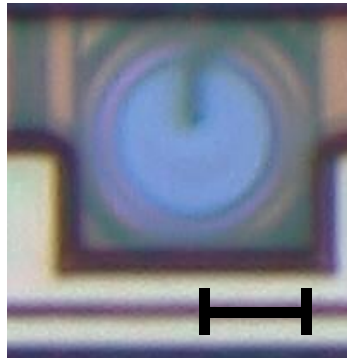
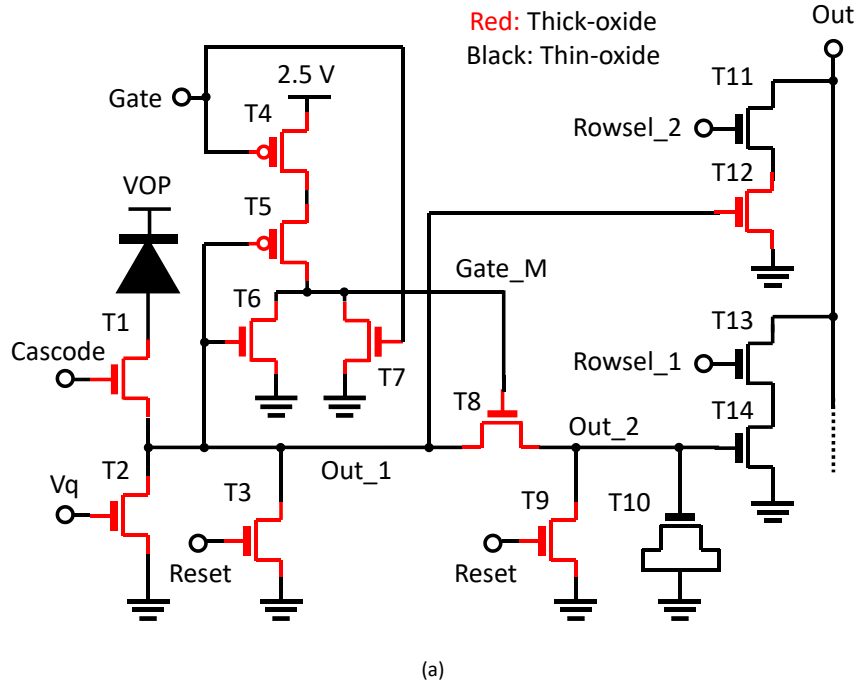


Figure 3:4 (a) Pixel schematic of SwissSPAD3. (b) Pixel micrograph of SwissSPAD3, with a round SPAD. Scale bar: 5 μm .

The readout timing diagram is shown in Figure 3:6. During the readout, *Rowssel_1* and *Rowssel_2* are enabled sequentially to read *Out_1* and *Out_2* through a shared data bus, respectively. Between *Rowssel_1* and *Rowssel_2*, the *Out* bus is reset by a pullup transistor dedicated to the column, which is located outside the pixel. After the end of *Rowssel_2*, *Reset* is asserted again to start a new exposure. In each frame, the exposure starts and ends with *Reset*. Contrary to SwissSPAD2, t_{ro} , the readout duration of a row, is equal to twice the sum of the pullup and pulldown durations in SwissSPAD3. Under the fastest operation settings, t_{ro} is equal to 80 ns.

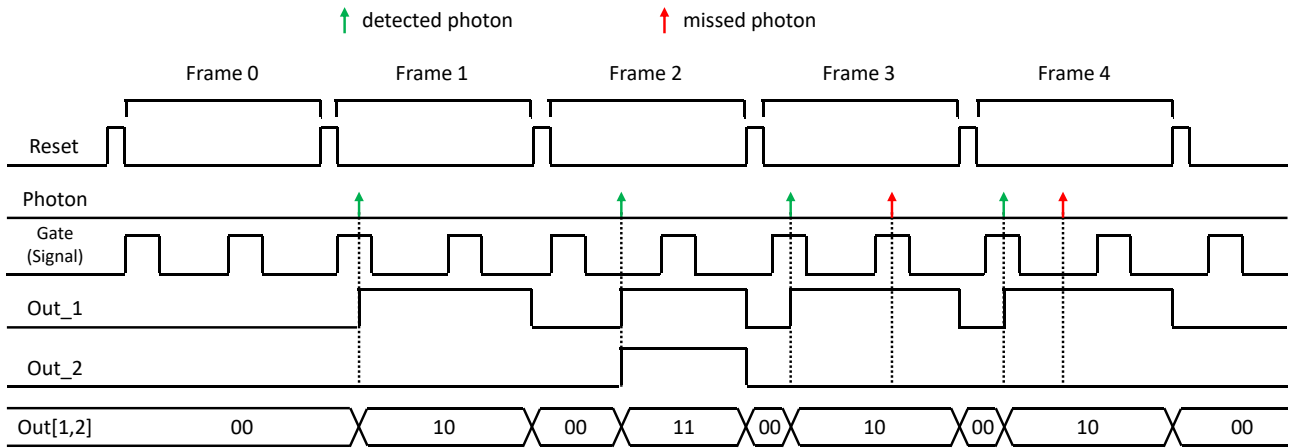


Figure 3:5 Timing diagram showing the SwissSPAD3 pixel output for different photon detection scenarios. The photons in the diagram represent the photons that are detected by the SPAD. *Out_1* is activated if at least one photon is detected inside the frame. *Out_2* is activated if the first photon in the frame is outside the gate signal pulse window, and not activated if it is inside. Both *Out_1* and *Out_2* are insensitive to subsequent photons detected in a frame. Note that the 01 state is not possible and represents thus a fault if it occurs.

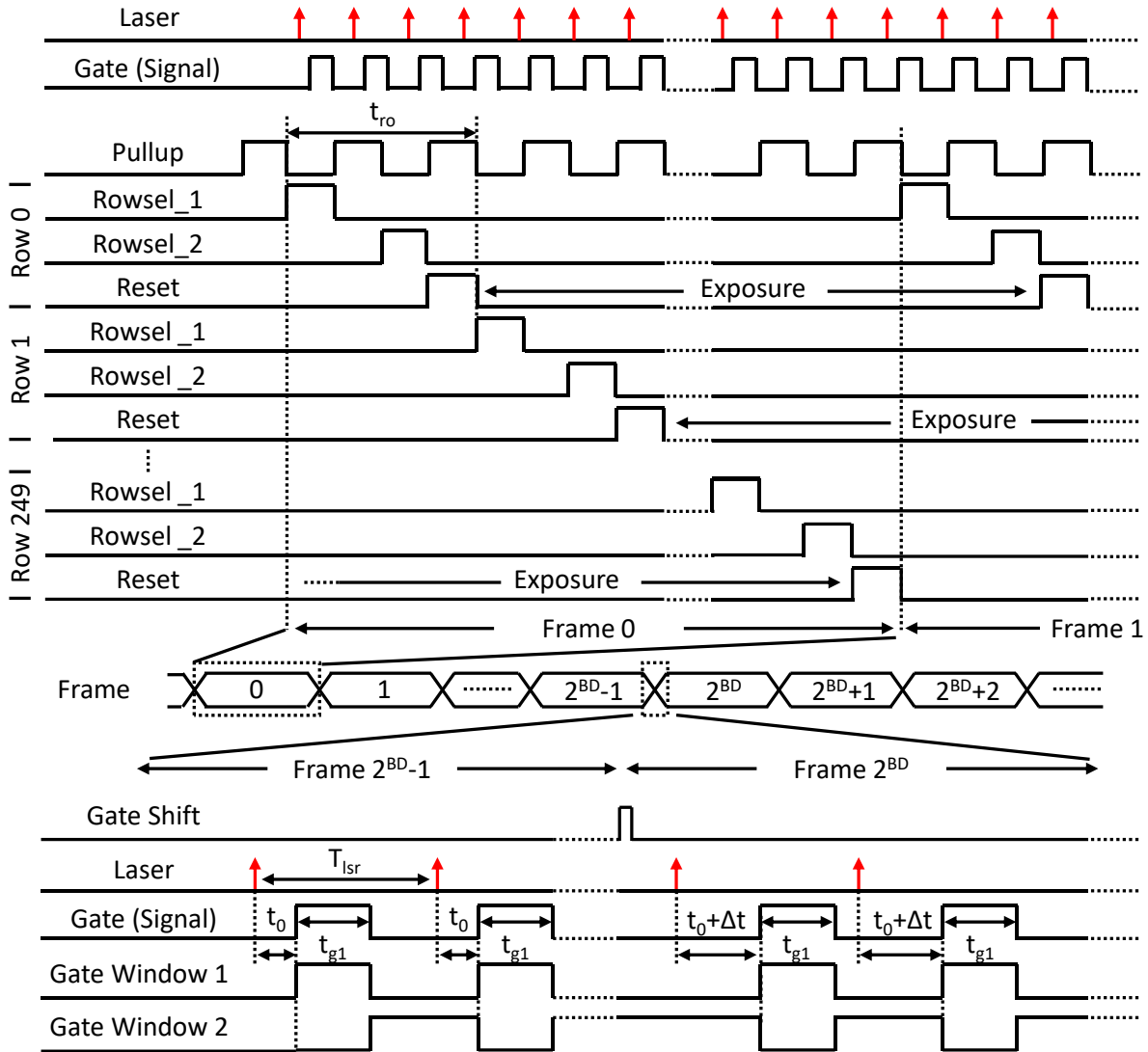


Figure 3:6 Timing diagram of the operation of SwissSPAD3 in rolling shutter mode.

Unlike in SwissSPAD2, the exposure window start and end times of the pixel are not controlled by the global *Gate* signal in SwissSPAD3. Instead, they are controlled by *Rowsel_1*, *Rowsel_2* and *Reset*, none of which are global signals. As a result, SwissSPAD3 only operates in rolling shutter mode. The addition of global shutter feature to this pixel requires two extra elements: a global reset/recharge signal which is distributed to all pixels with minimal skew through a signal tree, and an additional global gating between the SPAD and *Out_1*. The implementation of these features increases the number of transistors and signal distribution wires in the pixel, thus increasing the demand for inactive pixel area and potentially causing a failure to meet the propagation time requirements of the critical signals for the desired frame rate and timing performance. Considering the inherent duty cycle penalty of the global shutter, the aforementioned performance degradations caused by increasing the pixel complexity, and the emphasis on maximizing duty cycle for FLIM which requires rolling shutter, the global shutter feature was not added to SwissSPAD3.

3.3.2 Time gate implementation

In SwissSPAD2 and SwissSPAD3, the physical mechanisms that are responsible for the gating exhibit a major difference. The relation between the gate signal edges and the gate window edges in the two sensors are shown in Figure 3:7. In the figure, the solid lines represent the gating signals, whereas the dotted lines represent the gate window, i.e. the relative detection probability of the photons as a function of the photon arrival time. In the SwissSPAD2 gate sequence, which is shown in Figure 3:7(a), *Recharge* and *Gate* signals are asserted back-to-back. A photon arriving inside the *Recharge* pulse can only be detected if the delay between the photon arrival time and the *Recharge* signal falling edge is shorter than the total time required for the quenching and the recharge of the SPAD. The maximum delay that meets this requirement is called “gate window rising edge delay” (d_r). As shown in the figure, this requirement determines the actual position of the gate window rising edge. In addition, for the photon to be detected, the voltage of *Gate_out* must be higher than the threshold voltage of *T7* by the time *T5* is switched off. This constraint determines the position of the gate window falling edge. The minimum delay between a photon arrival time and the gate signal falling edge that meets this condition, i.e. the gate window falling edge delay (d_f), is determined by the quenching time of the SPAD required for the *Gate_out* voltage to reach the level to activate *T7*.

Even after the exclusion of SPAD jitter, *Gate/Recharge* signal jitter and the laser pulse width, the transition times of the gate window are still expected to be non-zero. An important contribution of the rise and fall times of the gate window stems indeed from the temporal variation of d_r and d_f , which can be represented by their respective standard deviations, σ_{d_r} and σ_{d_f} . As explained above, d_r is determined by the sum of quenching and recharge times, whereas d_f is only determined by the quenching time. Assuming that the uncertainty in quenching and recharge times are uncorrelated, σ_{d_r} and σ_{d_f} can be expressed as

$$\sigma_{d_r} = \sqrt{\sigma_q^2 + \sigma_{rc}^2}, \quad (3.3)$$

and

$$\sigma_{d_f} = \sigma_q, \quad (3.4)$$

where σ_q and σ_{rc} are the standard deviations of quenching and recharge times, respectively, and the two processes are statistically independent. σ_q primarily results from the variation in the transient behavior of the avalanche current, $I_{av}(t)$, due to the random nature of the impact ionization process. On the other hand, σ_{rc}

is affected by two main factors [120]. The first factor is the drain current noise of the transistor $T3$, primarily consisting of thermal noise and flicker noise. The second factor is the probability of a new photon detection during the recharging process. The SPAD is sensitive to photons during recharge, as it has a positive excess bias voltage immediately after the start of recharging. During this process, the SPAD is therefore sensitive to a secondary avalanche due to a subsequent absorbed photon or afterpulsing. In either case, an avalanche immediately triggers a second quenching and recharge cycle, thereby contributing to σ_{rc} . From this description, it can also be seen that unlike σ_{df} , σ_{dr} is influenced by the photon flux as well. Furthermore, the quenching time is longer during the rising edge than the falling edge due to the lower quenching resistance when $T3$ is on, causing σ_q to be higher in the rising edge than the falling edge. For all these reasons, the gate window rise time is expected to be longer than the fall time in SwissSPAD2.

In Figure 3:7(a) and the paragraph above, the mapping of the gate window rising/falling edges to the gating signals was done assuming that the laser pulse shifts forward with respect to the gate sequence. When the shift is performed in the opposite direction, the rising and the falling edges of the gate window are swapped, with no other changes in the gate window characteristics.

In SwissSPAD3, the gate window is opened and closed only by the *Gate* signal (Figure 3:7(b)). The quenching speed of the SPAD is constant throughout the gate sequence owing to the absence of *Recharge*, which results in constantly high quenching resistance. This leads to d_r and d_f being equal, which in turn yields a symmetrical gate window with both edges having similar characteristics to the faster edge in SwissSPAD2.

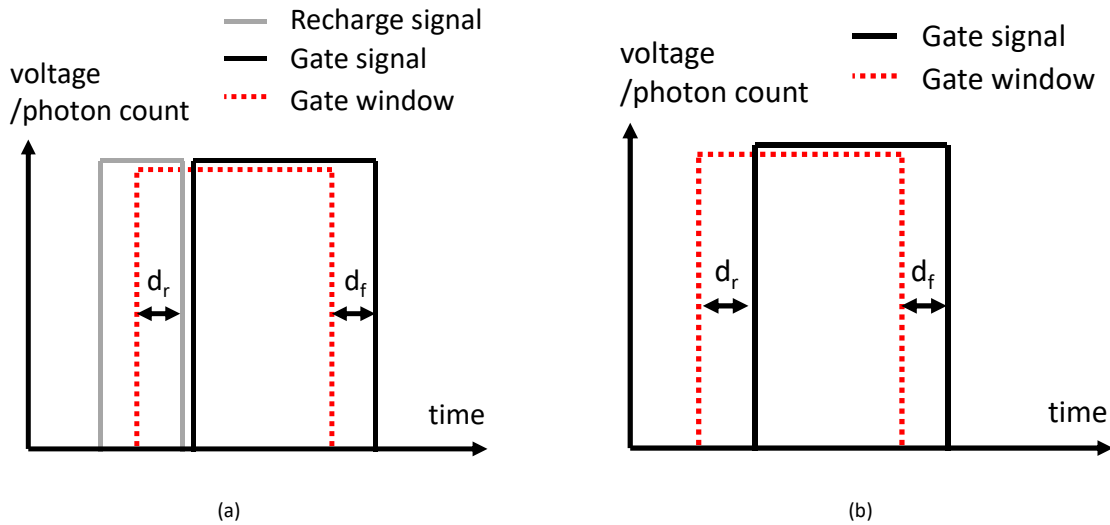


Figure 3:7 Conceptual diagram illustrating the relationship between the gate window and the gate signals in (a) SwissSPAD2 and (b) SwissSPAD3.

As discussed in Chapter 2, the distribution of the gate signals to the pixels is of key importance for the achievement of fast gate window transition time and narrow gate width, both of which are significant parameters for time-resolved imaging performance. For this reason, the complexity of the distribution network of the gate signals from the pad to the pixels is significantly higher than the rest of the input and output signals of the sensor. The gate signals are *Gate*, *Recharge* and *Spadoff* in SwissSPAD2, and *Gate* in SwissSPAD3. These signals are generated in the FPGA by a clocking wizard IP from an external laser trigger, and sent to the pixels through an analog pad and the distribution network.

In image sensors, there are several established ways to distribute critical signals to multiple destinations. Three of these architectures are conceptually illustrated in Figure 3:8. The first option is to propagate the signal through a single wire across all columns, and separate it into branches for the distribution to pixels in each column. Owing to the absence of a signal tree, this distribution scheme has the simplest design complexity. It also occupies minimal area and consumes low power due to the shortest total wire length. The main downside of this method is that it causes a skew in both horizontal and vertical directions. The second architecture includes a signal tree which distributes the signal with equal propagation delay from the pad to the first pixel of every column. The signal is then distributed to the pixels in each column using parallel wires. Since this architecture allows the placement of the tree outside the pixel array, it only demands additional pixel area for a single wire, like in the first architecture. However, the significant increase in the total wire area introduced by the tree causes higher power consumption compared to the first approach. In addition, since the tree does not extend to the pixels within a column, the skew can only be eliminated in a single direction (i.e. horizontally). The third option is to use a signal tree to distribute the signal to all pixels in the array. The H-tree, which is illustrated in Figure 3:8(c), is a commonly used example of such architectures. While H-tree is an architecture which minimizes the skew in both directions, it poses several implementation challenges particularly in image sensors implemented in a 2D process technology. Firstly, the placement of the tree branches in the array requires the addition of buffers inside the pixel area. The area required by these buffer blocks and their power rails introduces restrictions to the fill factor and potentially increases the parasitic capacitance of both gate and output wires, since the total area dedicated to the wires in a pixel is fixed. Secondly, it disrupts the uniformity of the pixel layout. Since the position of the tree branch inside the pixel varies with the location of the pixel in the array, the implementation of the tree requires either multiple pixel layout variations, or the allocation of space for the tree and buffer on all four sides of the pixel. The former option disrupts the modularity of the chip, thereby significantly increasing the complexity of the chip design process. On the other hand, the latter option further decreases the fill factor. For both SwissSPAD2 and SwissSPAD3, the second option was selected for the gate signal distribution. The main reason for not choosing the H-tree was the large pixel area demand that it would cause, considering that the chips were designed in 0.18 μm 2D CMOS process technology.

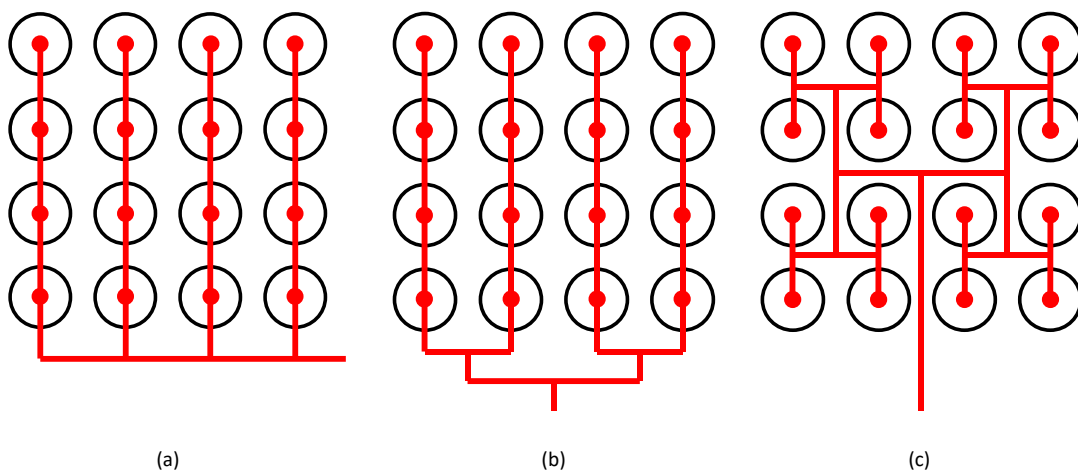


Figure 3:8 Three alternative gate signal distribution schemes in a pixel array: (a) No signal tree, (b) column-level signal tree, (c) pixel level H-tree.

In SwissSPAD3, the reduction of the number of gate signals from 3 to 1 and the reduction of the number of rows from 512 to 500 enable the allocation of a larger area for a single gate signal tree compared to SwissSPAD2. As a result, several design improvements were implemented in the signal tree of SwissSPAD3 to

achieve better signal integrity. Firstly, the spacing between the wires of multiple tree levels was increased. Due to the structure of the tree, the density of metals is higher in the central regions compared to the edges. This results in a non-uniform spacing between the metals, causing a variation in the wire parasitic capacitance and the propagation delay between the different signal routes. The physical separation of the wires mitigates this problem, improving the horizontal gate skew. Secondly, the capacitance of the decoupling capacitors around the trees was increased. Due to the high density of wires and buffer blocks, the tree is the block with one of the highest power consumption levels in the entire chip. This makes the power rails around the tree susceptible to temporary voltage drop during the propagation of the signal edges. The minimization of this voltage drop through decoupling capacitors is crucial to achieve low gate skew and gate width. Together with the elimination of the slow gate edge mechanism, the supply voltage drop mitigation enables two symmetric and fast gate edges, which determine the minimum achievable gate width without pulse shrinking. Finally, additional buffers were placed in the first two levels of the tree. The earliest levels of the tree have the longest wire length, which results in the largest load capacitance for the buffers. The placement of extra buffers in the middle of these wires reduces the load capacitance of each, ensuring faster signal transition.

3.3.3 Readout and system architecture

The block diagram of the SwissSPAD2 image sensor is shown in Figure 3:9(a). Its 512×512 pixel array is divided into two identical halves of 512×256, which operate independently. These halves are mirrored with respect to the dotted line in the diagram. The spacing between the adjacent rows on different sides of the division line is equal to the pixel pitch, thereby making the pixel pitch uniform across the full array.

The readout control electronics dedicated to each row is located on the left side of the chip. An 8-bit binary code is used to select a row from the 256 rows on one half of the array. The first stage of the row selection block is the row decoder. The decoder is formed by a 9:1 AND gate, in which the input consists of the 8-bit code and an *Enable* signal, and the output is the *Rowsel* signal of one of the rows in the array. The value of the code which enables *Rowsel* is determined by whether the normal or inverted version of each bit is connected to the decoder. This architecture, in which the selection of each row is independent, allows the customization of the frame size. For instance, by reading only 24 rows out of 256, up to ~1 Mfps binary frame rate can be achieved. This flexibility comes at the expense of increased number of I/O pins for the control of the readout. The second stage of the row selection block is the reset generator. In this stage, the *Rowsel* signal and its delayed and inverted version are connected to the inputs of a NOR gate, which outputs the *Reset* signal of a row. This block asserts the *Reset* signal upon the falling edge of *Rowsel*, and keeps it high for the duration of the delay between the two NOR gate inputs.

The *Gate*, *Recharge* and *Spadoff* signals, which are transmitted to each column with equal delay, are distributed to each half independently via signal trees. The architecture of these trees, which are located at the top and bottom of the array, is discussed in the previous subsection. The pullup network of the column readout buses and the column output registers are also located at the top and bottom. A PMOS pullup transistor dedicated to each column output bus is used to reset its voltage to logic high after each readout event. The output register of each column, which is implemented by D flip flops, is used to register the state of *Out* at the end of the readout, slightly before the activation of *Pullup*. Finally, 4:1 multiplexers, which are shared by every group of four adjacent columns, are used to read out these four columns sequentially through a shared pad. The number of output channels was limited to 128 due to the restrictions on the available number of FPGA user I/O pins.

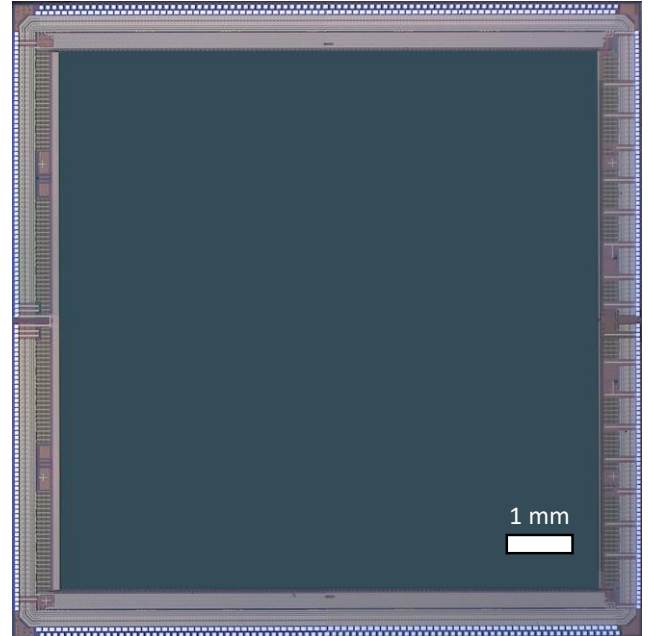
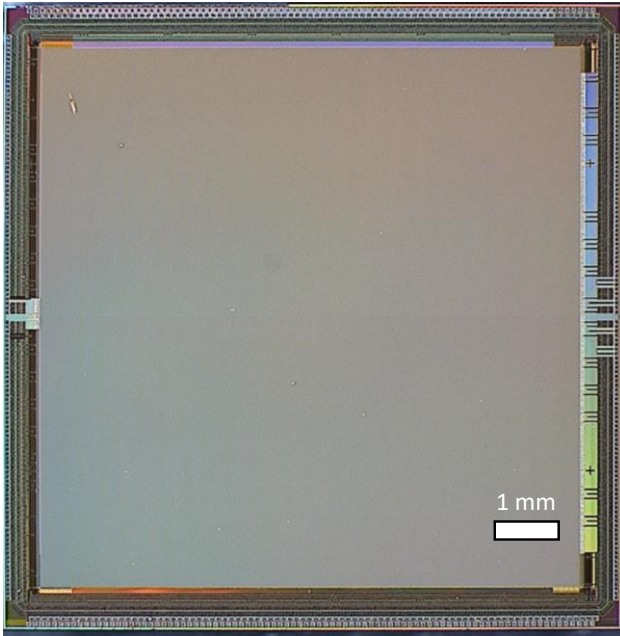
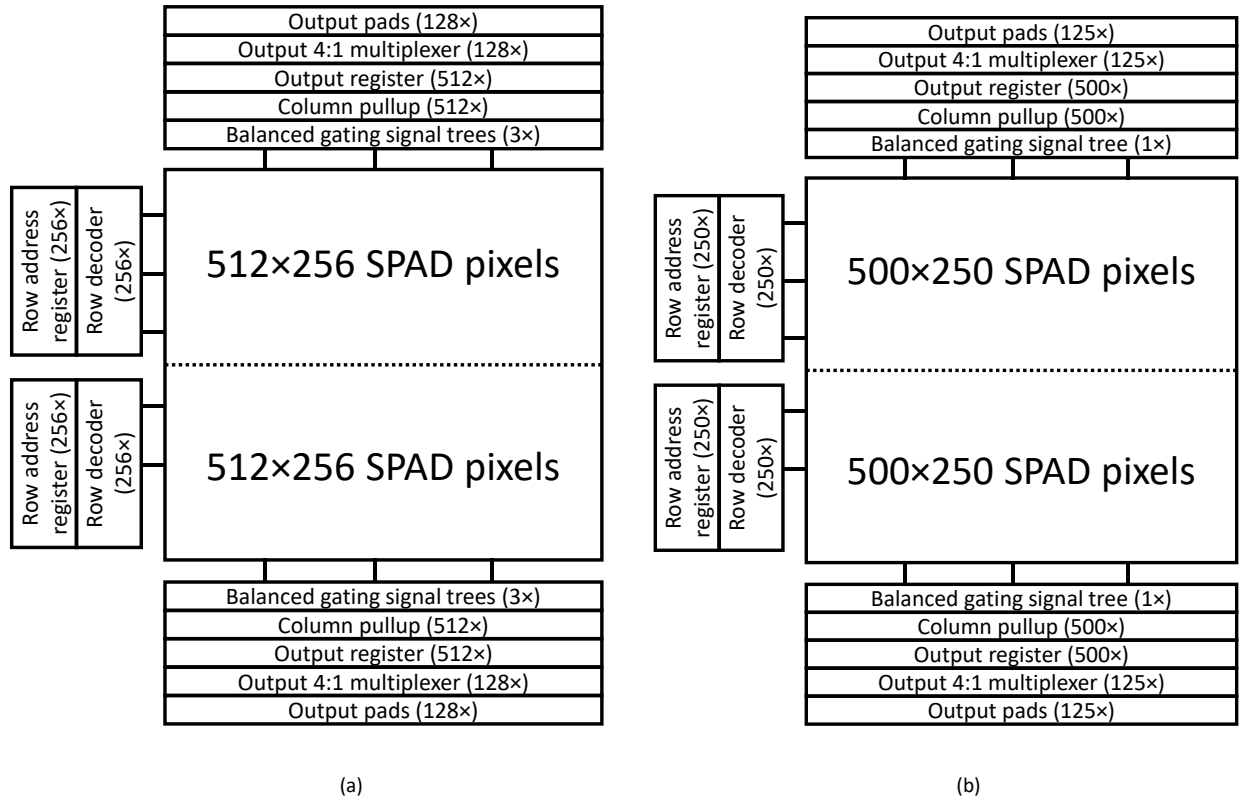


Figure 3:9 Sensor architecture block diagram of (a) SwissSPAD2 and (b) SwissSPAD3. The photomicrographs of (c) SwissSPAD2 and (d) SwissSPAD3.

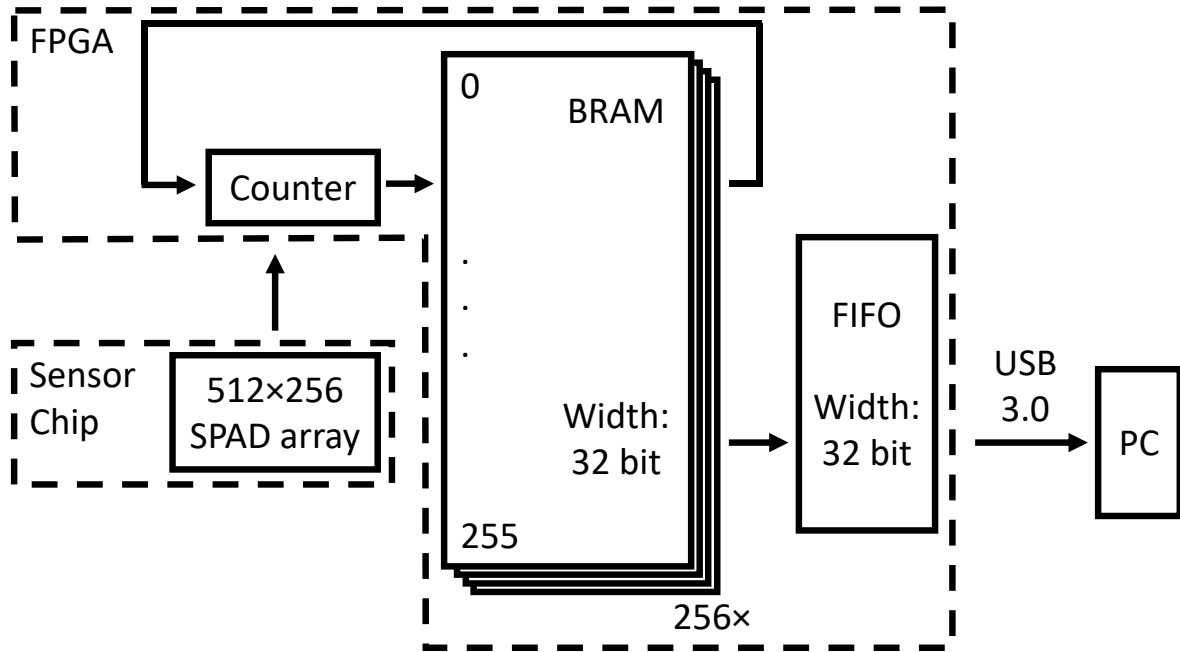
The block diagram of SwissSPAD3 is shown in Figure 3:9(b). The general architecture of the chip is similar to that of SwissSPAD2, with minor differences in various parameters. Firstly, the number of pixels was reduced to 500×500 . Secondly, the number of signal distribution trees was reduced from three to one, since the signals *Readout* and *Spadoff* are removed from SwissSPAD3. Both of these changes allow the allocation of more area for decoupling capacitors for the gate signal tree, which in turn enable shorter gate width and lower skew.

The SwissSPAD2 chip is operated by a commercial FPGA integration module (XEM7360, Opal Kelly, USA), which contains an FPGA (Kintex-7 XC7K160T, Xilinx, USA), 2 GB DDR3 RAM and a USB 3.0 connector. In the currently available system, one half of the array with 512×256 pixels is controlled by a single FPGA. As previously mentioned, SwissSPAD2 can generate binary frames at a maximum speed of 97.7 kfps. Under these conditions, the data generation speed of one half of the chip is equal to ~1.49 GB/s. This value is significantly larger than the 340 MB/s maximum data transfer speed of the USB 3.0 connection. As a consequence, the continuous operation of the camera at its maximum speed requires data compression on the FPGA. Considering this limitation, two operation modes were developed for SwissSPAD2: continuous mode and burst mode.

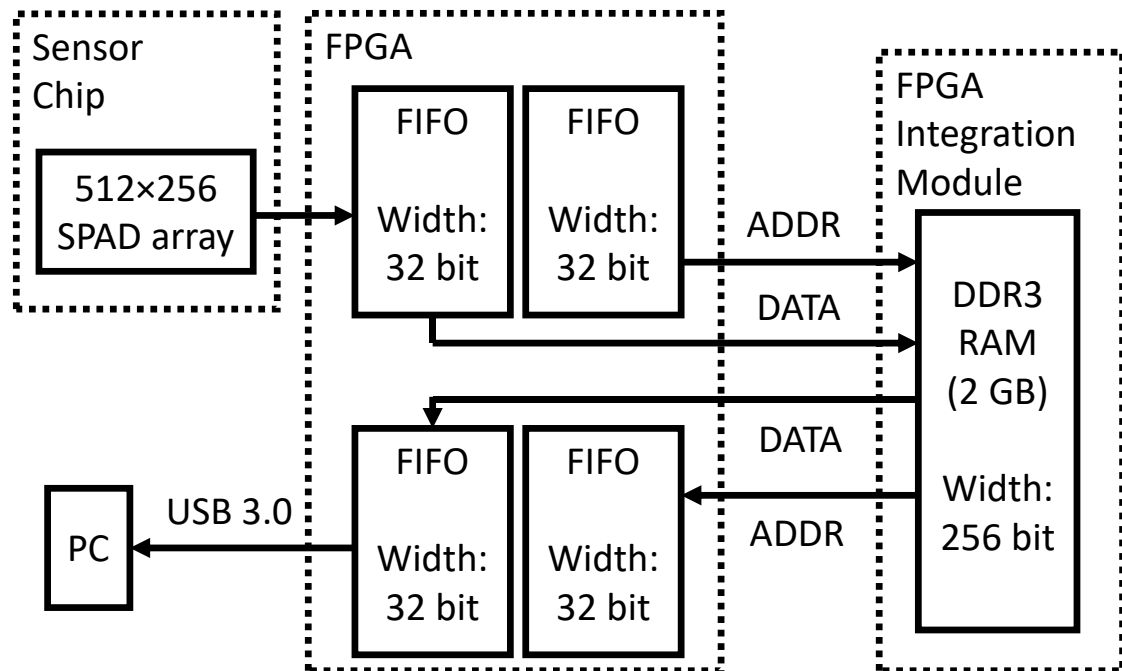
The architecture of the continuous mode is shown in Figure 3:10(a). In this mode, every sequence of 255 sequential binary frames generated by the chip is converted to a single 8-bit frame. This operation was performed in the FPGA, using block RAM (BRAM). In the BRAM space, an 8-bit word is assigned to each pixel to store its count. When a binary frame is registered by the FPGA, the 8-bit count of a pixel is incremented if the value of its binary counter is one; it remains the same if it is zero. After the end of the process, the 8-bit counts on the BRAM are transferred via USB to the PC, and then reset for the next sequence. To ensure continuous data acquisition of the sensor chip, the BRAM blocks are duplicated in such a way that the writing and reading processes run in parallel. While the 8-bit frame index i is being accumulated on the first block (block A), the frame index $i - 1$ is transferred to PC from the second block (block B), then block B is reset. After the accumulation of frame i , the roles of the blocks A and B are reversed for the frame $i + 1$. Since the combined duration of the BRAM readout and reset is shorter than the duration of the accumulation of 255 binary frames even at the fastest speed of the chip, this configuration enables continuous data acquisition. The counter blocks, which are implemented using the lookup tables (LUT) on the FPGA, increment the 8-bit count of the BRAM upon photon detection. Since the FPGA and the USB microcontroller are operated by two asynchronous clock signals, a dual-clock FIFO was used as a buffer to ensure the transfer of data without any loss. When the binary counts are converted to an N-bit format, the data compression ratio is calculated by

$$Comp = \frac{2^N - 1}{N}. \quad (3.5)$$

According to Equation (3.5), the 8-bit conversion compresses the data by a factor of 31.9. As a result, the required data transfer rate through USB in this mode is equal to 46.7 MB/s, which is below the maximum capacity of the USB 3.0 connection. The main advantage of this compression, which is perfectly acceptable for certain applications such as FLIM, is that it enables the continuous operation of the entire 512×256 array at the highest binary frame rate that is within the sensor chip constraints. However, its downside is the loss of temporal resolution by a factor of $2^N - 1$, which can be of key importance for certain dynamic imaging applications, such as motion blur correction [217]. The relations between compression bit depth and compression ratio, frame rate and data rate are shown in Figure 3:11. According to this figure, in theory, the USB 3.0 throughput allows the continuous operation of SwissSPAD2 at its fastest frame rate when the output image format is 5 bits or higher.



(a)



(b)

Figure 3:10 System architecture block diagram of SwissSPAD2 for (a) 8-bit continuous mode imaging and (b) 1-bit burst mode imaging.

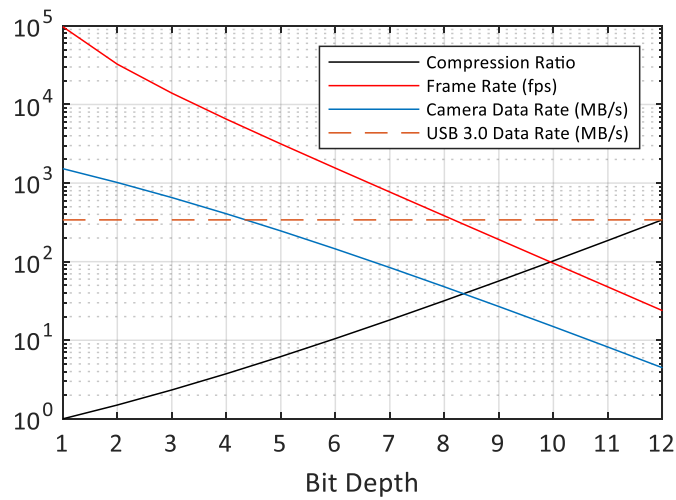


Figure 3:11 Relations between bit depth and compression ratio, frame rate and camera data rate of SwissSPAD2. Frame rate and camera data rate are calculated for the maximum binary frame rate of 97.7 kfps. It can be observed that for continuous operation at the fastest binary frame rate, a minimum of 5-bit data compression is required.

In burst mode, the writing and reading of data in the RAM occur sequentially. During the reading of the data from the RAM, the image sensor is insensitive to photons. For this reason, in this mode, instead of the BRAM, the data is stored in the DDR3 RAM in the integration module, outside the FPGA. The architecture of the burst mode imaging is shown in Figure 3:10(b). During the operation of the camera, the first stage is data acquisition. The binary output of the chip is written to the RAM through a FIFO. Simultaneously, the address on the RAM which corresponds to each data block is generated inside the FPGA and transferred to the RAM via another FIFO. The data width of the RAM is equal to 256 bits, and the data of each row of the sensor chip transferred in four groups of 128 columns sequentially. Therefore, the RAM address is incremented after reading every two groups of 128 columns. With its 2 GB capacity, the RAM can theoretically store up to 131,072 binary frames with 512×256 pixels; this corresponds to 1.34 s continuous acquisition time at the fastest sensor frame rate of 97.7 kfps. After the acquisition is finished, the data in each address of the RAM is sequentially read and transferred to a FIFO, from which it is streamed to the PC via USB, thus at a slower rate.

Both system architectures are also compatible with SwissSPAD3. However, several modifications are needed for the adaptation of these configurations to SwissSPAD3. From the readout standpoint, the most important difference of SwissSPAD3 from SwissSPAD2 is that it generates two bits of data per frame per pixel instead of one. In the continuous mode, this configuration requires two parallel 8-bit counting operations per pixel. This can be achieved by extending the utilized address space of the BRAM from 256 to 512, which is the full capacity of a BRAM unit in the FPGA of choice. This change allows the duplication of the frame data without increasing the number of BRAM units to be used, which prevents the potential implementation challenges arising from doubling the utilized BRAM capacity. Unlike the continuous mode, no significant modification is required for the adaptation to SwissSPAD3 in the burst mode. However, due to the increased number of gate channels per frame and the minor difference in the spatial resolution (Table 3:1), the number of frames per burst changes to 68,719 in SwissSPAD3.

3.4 Sensor characterization

3.4.1 Photon detection probability (PDP)

Photon detection probability (PDP) in SPADs is defined as the probability of a photon impinging on the active area to trigger an avalanche. Quantitatively, PDP is expressed as [218]

$$PDP(\lambda) = \int_{z_{end}}^0 T(\lambda)\mu(\lambda)e^{-z\mu(\lambda)}p_a(z)dz \quad (3.6)$$

where λ is the wavelength of the light, $T(\lambda)$ is the transmittance [219], $\mu(\lambda)$ is the absorption coefficient, $\mu(\lambda)e^{-z\mu(\lambda)}$ is the probability of absorption at depth z , and $p_a(z)$ is the probability of a photon absorbed at depth z to trigger an avalanche. The boundaries of the junction are defined by $z = 0$ and $z = z_{end}$. The first contributor to the PDP profile is the change of absorption probability with wavelength. The other main contribution comes from the excess bias voltage, due to the fact that the avalanche probability p_a is significantly affected by the impact ionization probability, which depends on the kinetic energy of the free charge carriers generated upon photon absorption. Given that the kinetic energy increases when the electric field in the multiplication region gets larger, the PDP also varies with excess bias voltage.

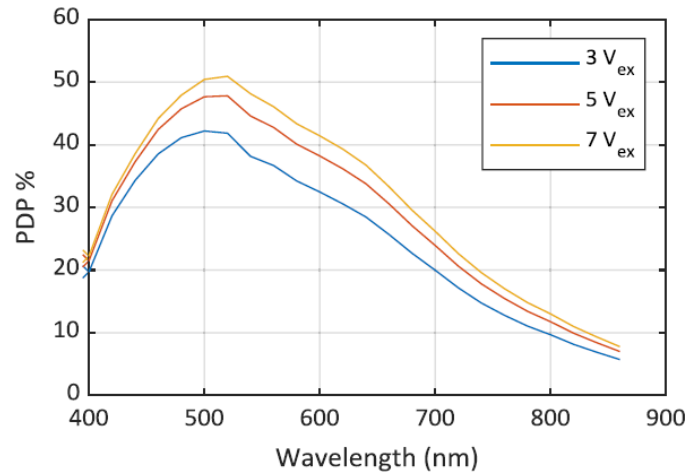


Figure 3:12 Photon detection probability (PDP) characterization of the p-i-n SPAD used in SwissSPAD2 and SwissSPAD3, for wavelengths between 400 and 860 nm [202].

To characterize the dependence of the PDP on these two parameters, the PDP measurement results for different wavelengths ranging from 400 nm to 850 nm for different excess bias voltages are provided in Figure 3:12. These results were measured by Antolovic *et al.* [202] on a version of the SPAD with a 5.85 μm active area radius. The PDP can be slightly reduced in the SPAD version with 3 μm radius, which is used in SwissSPAD2/3, as a result of a possible difference between the real fill factor and the layout-based fill factor. This difference can be attributed to the “border effect” described in [220]. The peak sensitivity of the SPAD is achieved at around 520 nm. Since the upper boundary of the multiplication region is deeper than the surface of the epitaxial layer, the PDP declines at lower wavelengths in the visible spectrum. At these wavelengths, absorption most likely occurs close to the substrate surface above the multiplication region; this leads to a reduced impact ionization probability. The PDP decline at larger wavelengths is due to the lower boundary of the multiplication region depth, which is limited by the depth of the buried n-well, determined by process

constraints. Although the excess bias of the SPAD can exceed 10 V when used as a standalone device [212], the maximum excess bias in the SwissSPAD2/3 pixels is determined by the maximum voltage that the oxide of the transistors can tolerate without breaking down. For this reason, only the values at the excess bias range that can be achieved in the SwissSPAD2/3 pixels were shown in Figure 3:12 (to date, SwissSPAD2 was operated at maximum $V_{\text{ex}} = 6.5$ V).

3.4.2 Dark count rate (DCR)

Dark counts, which are counts recorded by the pixel in the absence of any absorption of a photon, are one of the main sources of noise for image sensors in general. In the specific case of digital SPAD imagers whose readout noise is negligible, dark counts also determine the overall noise level of the camera. As a result, they are one of the major factors that determine the dynamic range, along with the photon count rate. For all these reasons, the characterization of dark count rate (DCR) is essential in SPAD imagers.

The DCR measurement results of SwissSPAD2 are shown in Figure 3:13. In this experiment, we aimed to measure two parameters: average DCR and hot pixel percentage. The average DCR indicates the overall noise level for the majority of the pixels in the array, whereas the hot pixel percentage indicates the percentage of the unusable pixels in the array due to extremely high DCR. The presence of too many hot pixels randomly scattered around the array causes a degradation in the effective spatial resolution. During this experiment, we captured a sequence of 1,024 8-bit images after placing the camera in a dark environment. The exposure time was tuned to capture sufficient statistics within reasonable data size limitations, while not entering the pile-up region which would render the reliable determination of the actual photon count challenging due to quantization noise. The DCR distribution was calculated by summing all of the images in the sequence and dividing the final count map by the total exposure time. The population plot was generated from the cumulative distribution function of the DCR values of the pixels. Hot pixels, also known as “screamers”, were defined in this work as the pixels that exhibit at least 100 times higher DCR than the median value. To this date, in the literature, there is an absence of a unified criterion for hot pixels in SPAD imagers [221]. This results from the variation in the dark count threshold deemed sufficient for entirely discarding a pixel from the analysis, which highly depends on the application and the operating conditions of the imager.

Figure 3:13(a) shows the DCR map of a SwissSPAD2 sample. This figure shows that the hot pixels are randomly scattered across the array instead of being concentrated in a particular region. The population distribution of DCR is shown in Figure 3:13(b). The value that corresponds to 50% of the population denotes the median DCR of the array. In this measurement, the median DCR at 6.5 V excess bias was measured as 7.5 cps/pixel, or 0.26 cps/ μm^2 . We can also observe a sharp increase in DCR above the 95th percentile of the pixels. The threshold that defines hot pixels is reached at around the 99th percentile; therefore, it can be concluded that only around 1% of the pixels are classified as hot pixels. Due to the reasons mentioned in the previous paragraph, it is non-trivial to interpret this number from the perspective of overall performance. Nevertheless, it is noteworthy that the slope of the population curve of SwissSPAD2 is generally lower than that of the other imagers of comparable format, such as [119]. Furthermore, the inflection point, which marks the percentage at which the DCR abruptly increases, is located at a higher percentage in SwissSPAD2. In this plot, it can also be seen that DCR increases with excess bias voltage. Figure 3:13(c) shows the DCR distribution of five different camera samples. The profiles of these samples indicate that the DCR distribution of the pixels is consistent among multiple samples, with minor variation.

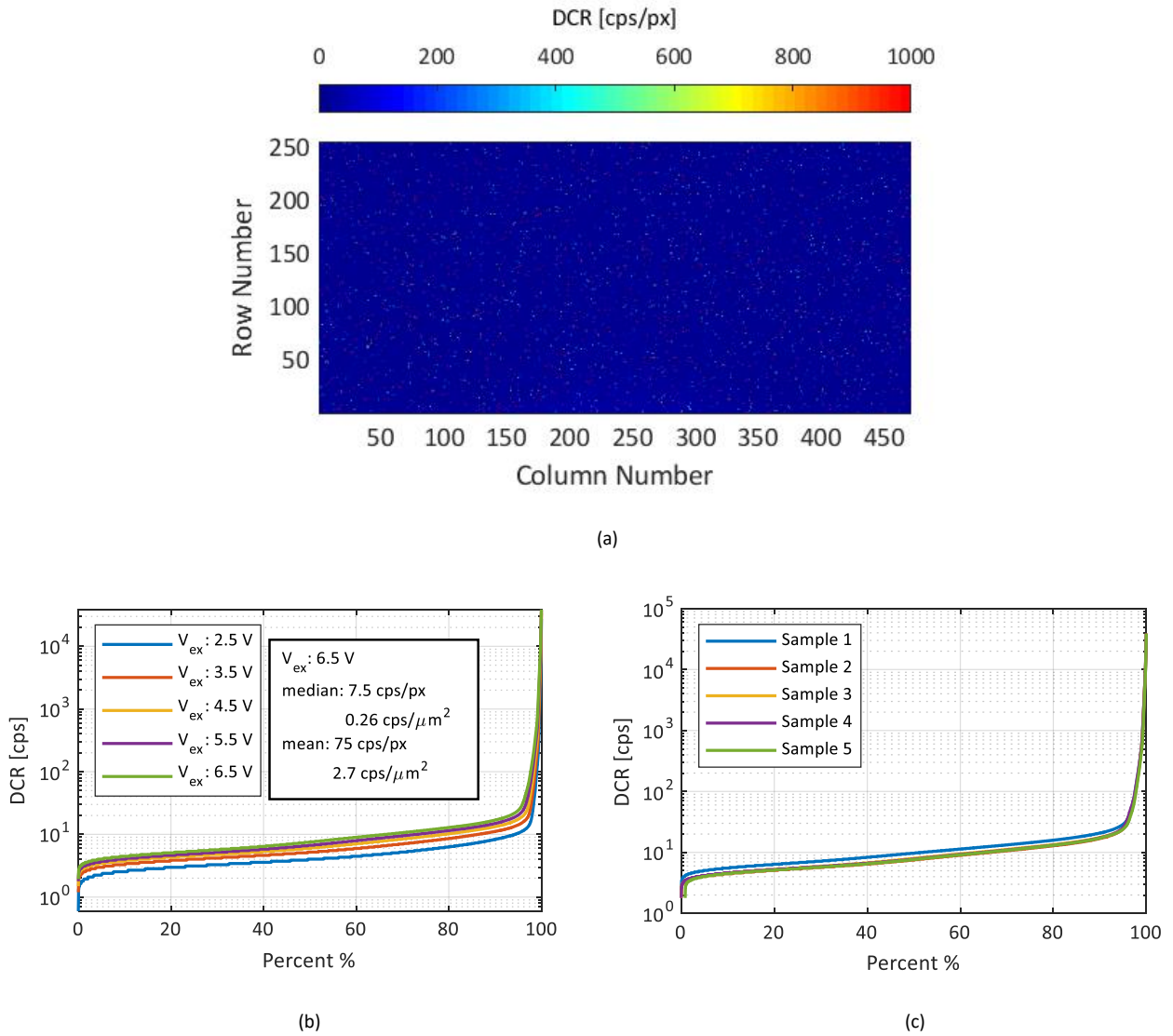


Figure 3:13 Dark count rate (DCR) characterization of SwissSPAD2. (a) 2D DCR map. This map shows that the hot pixels are randomly scattered across the field of view. (b) Population distribution of DCR at various excess bias voltages. The disparity between mean and median DCR is due to the high counts of the hot pixels. (c) The population distribution of DCR for five image sensor samples. Spatial resolution: 472×256 .

Figure 3:14 shows the DCR characterization results of SwissSPAD3. Compared to SwissSPAD2, the number of pixels in this analysis is approximately twice as high, with a 499×498 spatial resolution as opposed to 472×256 . Figure 3:14(a) shows the 2D DCR map of the full array for 6 V excess bias voltage. From this image, it can be seen that the hot pixels display a highly uniform spatial distribution, with no visible difference between the distribution characteristics of the top and bottom halves of the array. Figure 3:14(b) shows the population distribution of DCR for a single sensor sample at five different excess bias voltages. The mean and median values at $V_{ex} = 6$ V are reported in the figure. In terms of the average values and the percentile corresponding to a sharp increase, the DCR characteristics of SwissSPAD3 and SwissSPAD2 display a high degree of similarity. The difference between the median values of the two sensors is within the sample-to-sample variation, as shown in Figure 3:13(c). Differences in the degree of ambient light shielding between the two experiments and an increase in the average DCR with long-term use are two additional potential contributors to this difference.

In the literature, long-term DCR increase in SPADs has been observed under various effects, such as artificially introduced gamma-ray [222, 223, 224], X-ray [225], proton [226, 223, 224] and neutron [225] irradiation. The probable physical mechanisms behind the DCR increase were identified as the reduction of the breakdown voltage or an increase in afterpulsing due to radiation-induced defects or traps in the substrate [222, 225]. Due to their underlying mechanisms, these effects are generally permanent, and can only be mitigated through processes such as deep cooling or thermal annealing [226]. The possibility of SwissSPAD2 to have undergone such an effect during its operation under ambient CW light or picosecond pulsed laser remains to be investigated.

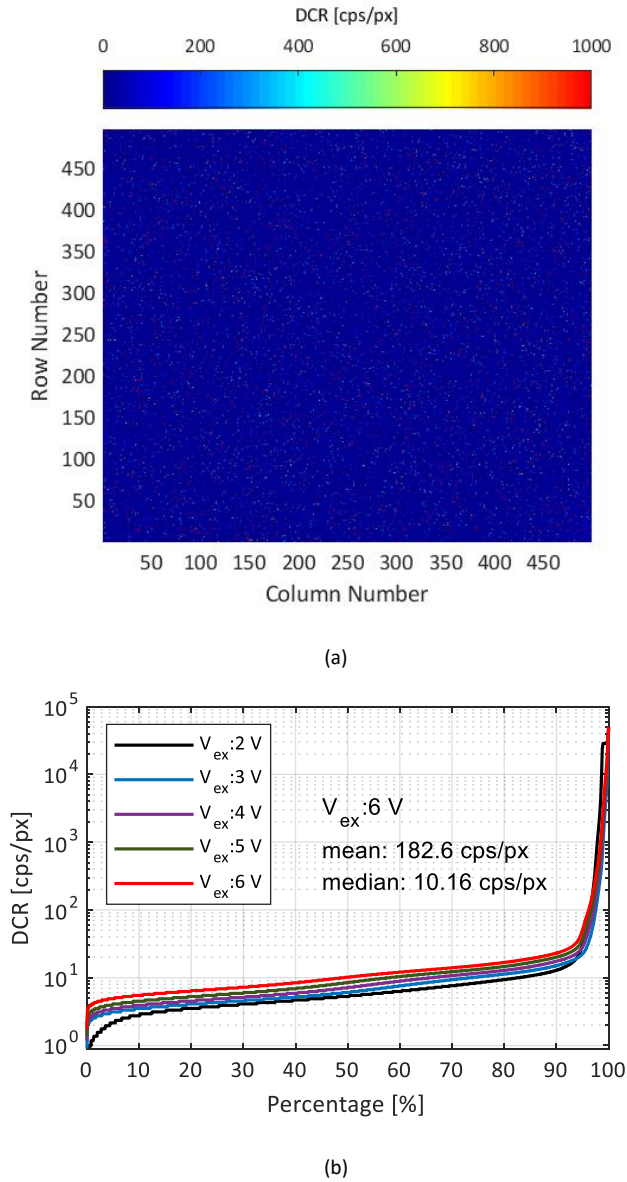


Figure 3:14 Dark count rate (DCR) characterization of SwissSPAD3. Spatial resolution: 499×498. (a) 2D DCR map. This map shows that the hot pixels are randomly scattered across the field of view. $V_{ex} = 6$ V. (b) Population distribution of DCR at various excess bias voltages. The difference between mean and median DCR is due to the high counts of the hot pixels.

3.4.3 Crosstalk

In an image sensor, crosstalk is defined as a false photon detection event caused by another photon detection event in a neighboring pixel. Being a source of correlated noise, crosstalk is an undesired effect as it negatively affects both SNR and the spatial resolution of the sensor.

Crosstalk in SPAD imagers can be measured in multiple ways. Firstly, it can be measured from the distribution of inter-arrival times between two neighboring SPADs [203, 227]. It can also be estimated by the comparison between the dark count of a neighbor of a hot pixel with the average dark count [119]. The crosstalk of SwissSPAD2 was measured using the latter approach. Firstly, the sensor was placed in a dark environment with negligible ambient light. Under these environmental conditions, a sequence of 1,024 8-bit dark images was captured, in which the photon counts mainly consist of dark counts. To evaluate the crosstalk percentage (CT), the counts of the pixels adjacent to the hot pixels (C_{NP}) were compared to the counts of the hot pixels themselves (C_{HP}), using the equation below:

$$CT = 100 \times \frac{C_{NP}}{C_{HP}}. \quad (3.7)$$

To distinguish between counts due to crosstalk and dark counts, the median dark count of the images was subtracted from all pixels before the crosstalk calculation, using the equation below:

$$C_{CT} = C_{TOTAL} - C_{MDC} \quad (3.8)$$

where C_{CT} is the crosstalk count, C_{TOTAL} is the total count, and C_{MDC} is the median dark count. In order to avoid an overestimation of the crosstalk, the pile-up correction equation (2.33) [203] was applied to all pixels, which is a required step to estimate the true photon count of the hot pixels that generally operate in the pile-up region. The lower photon count limit for the hot pixels (LL_{HP}) was determined in such a way that a sufficiently high number is achieved for both of the following parameters: the number of pixels to be analyzed in the array, and the photon count of an average hot pixel. By finding the right balance between these two parameters, the error due to the pixel-to-pixel variation of DCR, which is significant in the case of two adjacent hot pixels, and the error due to shot noise are both kept at a low level. Furthermore, an upper photon limit for hot pixels (UL_{HP}) was also set to exclude pixels in extreme pile-up region (i.e. saturated pixels), whose real photon counts cannot be accurately recovered even by using the pile-up equation. In our analysis, the lower and upper photon count limits of the hot pixels are defined as:

$$LL_{HP} = \text{median}(C_{PUC}) \times 500 \quad (3.9)$$

and

$$UL_{HP} = (2^b - 1) \times N_{fr} \times 2, \quad (3.10)$$

where C_{PUC} is the photon count after pile-up correction, b is the bit depth of a frame, and N_{fr} is the number of frames in the data set.

The measurements were performed using 5 different samples with round SPADs and no microlenses, at 6.5 V excess bias voltage. The crosstalk values of Sample 1 are shown in Figure 3:15. The average crosstalk of the camera was measured to be below 0.06% for the nearest pixels, and below 0.04% for the nearest diagonal pixels. The results represent the median of 478 hot pixels, which exhibit at least 500 times more DCR than the

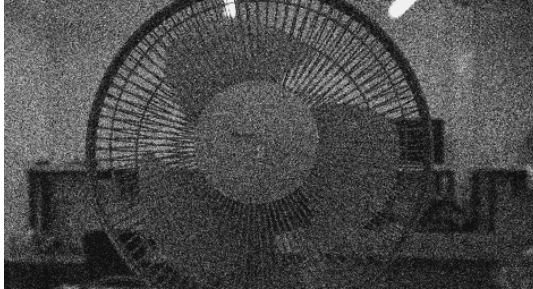
median value in the array. These values indicate low crosstalk compared to the state-of-the-art in large-format SPAD imagers. By comparison, the crosstalk percentage for the nearest pixel is 0.17% for [122], up to 3.5% for [120], and up to 4.3% for [119]. The crosstalk measurement results of the other 4 camera samples, which display similar characteristics as Sample 1, are presented in Appendix A.

-0.001	0.007	0.007	0.006	-0.001
0.004	0.034	0.057	0.032	-0.003
0.000	0.056	100	0.059	0.001
0.001	0.030	0.053	0.029	0.003
-0.001	0.006	0.004	0.000	0.001

Figure 3:15 Crosstalk percentage values of a SwissSPAD2 sample (Sample 1 of 5) [81].

3.4.4 Frame rate

Frame rate is a critical performance parameter of SwissSPAD2, which determines its dynamic range and ability to capture fast phenomena. The maximum frame rate of the SwissSPAD2 sensor chip is determined by the minimum readout duration of a single row, which is equal to 40 ns. Based on this limitation, the time required to read out the entire 512×256 sub-array is equal to $40 \text{ ns} \times 256 = 10.24 \mu\text{s}$. In the experiments reported below, the high-speed performance of SwissSPAD2 system was demonstrated.



4-bit, 4.6 kfps

(a)



8-bit, 273.5 fps

(b)

Figure 3:16 Grayscale images of a rotating fan captured in global shutter mode at (a) 4.6 kfps for 4 bits (b) 273.5 fps for 8 bits. The binary frame rate is 69.7 kfps. Spatial resolution: 472×255.

Figure 3:16 shows two grayscale images of a rotating fan captured by SwissSPAD2. The images were captured in global shutter mode, in which the exposure and readout windows occur sequentially for each binary frame. During the acquisition of a frame, the first step is a $4.02 \mu\text{s}$ exposure window in which a 95 ns gate window is opened 10 times with a 400 ns period. This exposure window is then followed by a $10.32 \mu\text{s}$ readout window, during which the camera is blind. Under these settings, the total acquisition time of a binary frame is equal to $14.34 \mu\text{s}$, which corresponds to a binary frame rate of 69.7 kfps. The N -bit grayscale images were constructed inside the FPGA by the summation of $2^N - 1$ binary frames. According to this relation, the frame rates of the 4-bit and 8-bit grayscale images in Figure 3:16 correspond to 4.6 kfps and 273.5 fps, respectively.

The frame rate of the camera was also tested in burst mode with binary images. In this test, a similar rotating fan was captured with a binary acquisition time of $10.335\ \mu\text{s}$, corresponding to 96.8 kfps frame rate. The gate is open throughout the operation, enabling full duty cycle in a rolling shutter configuration. Using the settings above, a sequence of 110,000 images was captured with 512×256 spatial resolution. The total data size of the sequence is 1.68 GB, which is equal to 84% of the capacity of the 2 GB on-board DDR3 RAM. Figure 3:17 shows nine selected images from this sequence, which capture a full rotation cycle of the fan. There are two numbers added to each image. The number on the top right corner is the frame index in the sequence, and the number on the top left corner is the elapsed time since the first frame in the sequence. From the examination of the full sequence, it was concluded that the fan finishes its full period at the frame number 4,431, whose delay from the frame number 1 is equal to $10.335\ \mu\text{s} \times (4431 - 1) = 45.8\ \text{ms}$. Therefore, the rotation frequency of the fan is equal to $1/(45.8\ \text{ms}) = 21.8\ \text{Hz}$ or 1308rpm.

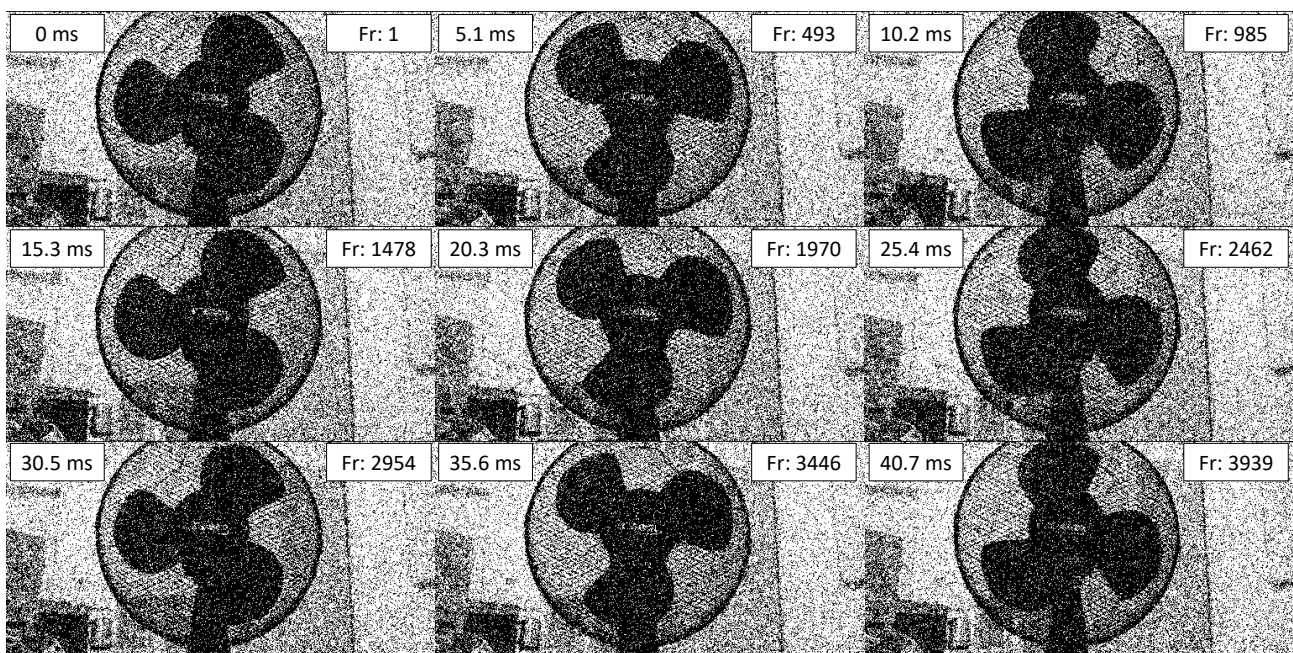
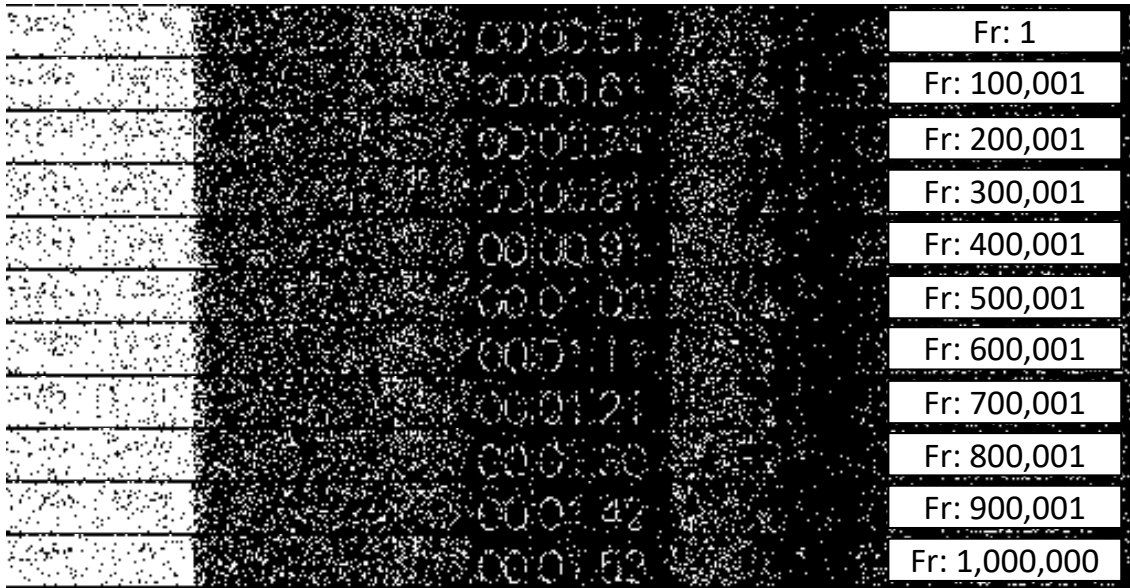
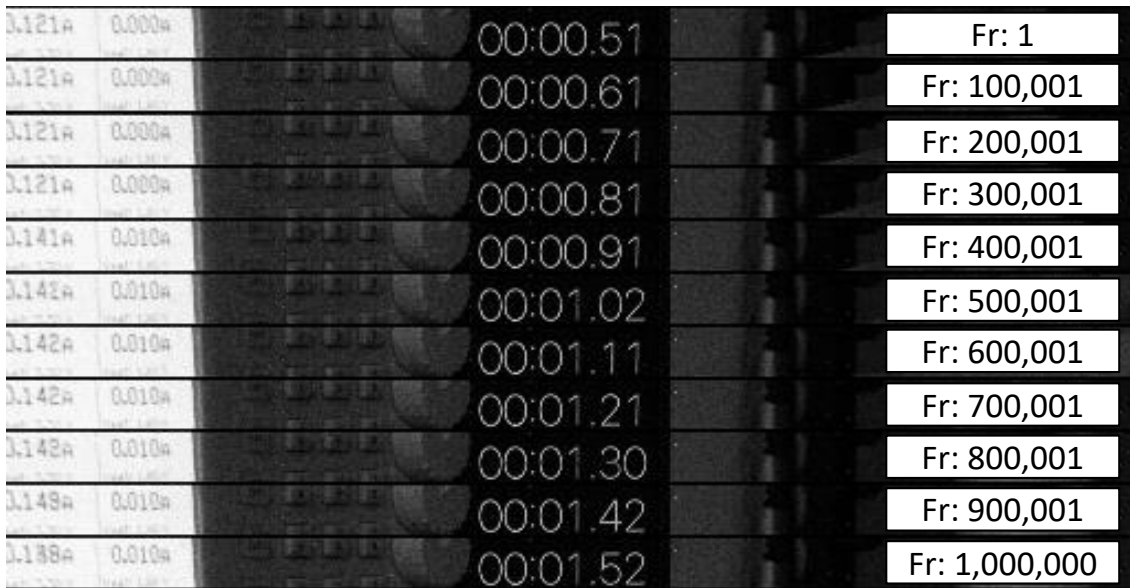


Figure 3:17 Binary images of a rotating fan captured at 96.8 kfps in rolling shutter mode. Spatial resolution: 512×256 .

In SwissSPAD2, it is also possible to increase the frame rate by reducing the number of rows in the frame. This feature was demonstrated in Figure 3:18. In this experiment, a digital stopwatch was captured with the camera in rolling shutter with 512×24 spatial resolution and a fully open gate. In burst mode, a binary sequence of 1,000,000 frames was captured with $1.015\ \mu\text{s}$ binary acquisition time and 985 kfps frame rate. Figure 3:18(a) shows several selected frames in the sequence; the time delay between each selected frame is approximately 100 ms. The total acquisition time of the sequence is equal to 1.015 s, and the memory space occupied by the entire sequence is equal to 1.43 GB, 71.5% of the on-board DDR3 RAM capacity. The 8-bit versions of the images are shown in Figure 3:18(b) as reference. The 8-bit images were constructed by averaging the annotated frame numbers and 254 more frames adjacent to them. The inequalities in the time steps between the frames is due to the irregular incrementation steps in the least significant digit of the stopwatch, and is unrelated to the operation of SwissSPAD2.



(a)



(b)

Figure 3:18 A sequence of 1,000,000 binary frames capturing a digital stopwatch at 985 kfps in rolling shutter. (a) Original binary frames. (b) 8-bit images formed by the average of 255 adjacent frames including the index displayed in each image. Spatial resolution: 512×24.

For the comparison of this performance with the state-of-the-art, various methods can be employed. Our proposed method is to compare the readout duration of a single row, assuming an inversely proportional relation in all imagers between the frame rate and the number of rows in a frame, as for SwissSPAD2. This comparison, while being imperfect due to its ignoring the number of pixels in a single row (i.e. the number of columns), can be justified by assuming the favorability of an aspect ratio closer to 1:1. Given this proposed method, the frame rate per row can be obtained from Table 3:1. From this table, it can be seen that the frame rate per row of SwissSPAD2 is similar to [119], despite having twice as many columns, and significantly higher

than the rest of the SPAD images with a similar architecture, including [120] (~25% higher) and [122] (~100% higher).

The generation of binary images at fast frame rates without readout noise is an attractive feature for various applications focusing on image reconstruction. Some examples for these applications include SNR enhancement in high-speed imaging with sparse photon counts [228, 229], and motion blur correction [217]. A discussion of [217], which was conducted using SwissSPAD2 in binary imaging mode, is available in Chapter 4, Section 4.4.

In all three experiments, there are relatively minor discrepancies between the maximum frame rate allowed by the chip readout speed (97.7 kfps) and the measured frame rate of the camera. Aside from the additional exposure window in global shutter mode, another reason for this discrepancy is the transition time between the exposure and readout states in the state machine of the firmware. This transition time of the order of 80 ns, which is independent from the sensor chip properties, can be eliminated with further optimization of the firmware architecture.

3.4.5 Dynamic range

The dynamic range is defined as the ratio between the maximum and minimum number of photons that can be detected in a single frame. Mathematically, dynamic range is expressed as

$$DR = 20 \log \left(\frac{n_{i \max}}{n_{i \min}} \right), \quad (3.11)$$

where $n_{i \max}$ and $n_{i \min}$ are the maximum and minimum detectable counts, respectively. For conventional imagers, this value is typically unaffected by the frame rate since readout noise and electron well capacity, which are the main parameters that determine the dynamic range, are both time-independent. On the other hand, in SwissSPAD2, the dynamic range is influenced by the dark count and the bit depth of a frame; both of which are time-dependent parameters. For this reason, the dynamic range in a digital SPAD imager is a function of the exposure time and the bit depth.

In SwissSPAD2, the maximum photon count is determined by the gradual deviation of the SNR from the shot noise level with increasing photon count due to the pile-up effect. As proven by the quanta image sensor (QIS) and Gigavision concepts [204, 230], the non-linear response function of these sensor types leads to a drop in photon sensitivity even in the early saturation (pile-up) region. While the counts of the photons detected by the SPAD (also called “detected photon count (n_i)”) can be estimated from the photon count measured by the camera (also called “measured photon count (m_i)”) using an analytical expression derived from the rules of Poisson distribution, the SNR drop caused by pile-up cannot be fully recovered due to reduced statistics. As discussed in [202], for accurate SNR calculation in SwissSPAD2, the mean and the statistical dispersion of the detected photon count must be used instead of the measured photon count. As the detected photon count and its standard deviation are estimates rather than the actual values, we denote them as \tilde{n}_i and $\sigma_{\tilde{n}_i}$, respectively.

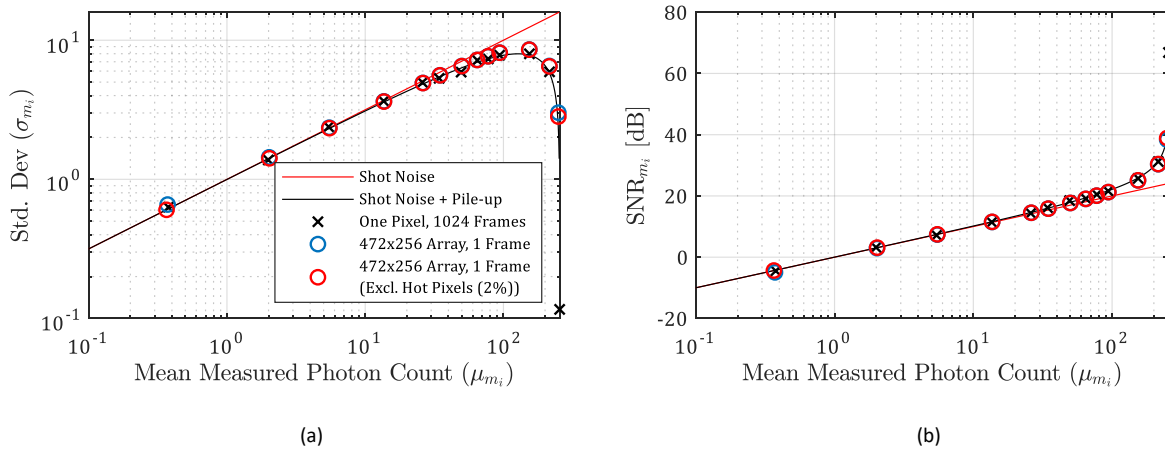
The dynamic range characterization results of SwissSPAD2 are presented in Figure 3:19, for a bit depth of 8 and an 8-bit exposure time of 1.63 ms. In this experiment, we calculated the SNR of the captured images in the temporal and spatial domain for multiple light levels (LL_0 to LL_{12}). Spatial light level uniformity is achieved by illuminating the camera with an integrating sphere. The SNR, which was calculated either in the spatial or

temporal domain from the mean total photon count (μ_C), the mean dark count (μ_{DC}), and the standard deviation of the total photon count⁵ (σ_C) as

$$SNR = 20 \log \left(\frac{\mu_C - \mu_{DC}}{\sigma_C} \right), \quad (3.12)$$

was then compared to the maximum achievable SNR limited by shot noise, in which the standard deviation is the square root of the mean photon count. In addition, a theoretical model that includes both shot noise and pile-up effects, based on Equation (2.37), was also plotted. Since hot pixels are expected to distort the SNR by increasing the dispersion at low light levels, the SNR calculation in the spatial domain was performed both including and excluding hot pixels. Figure 3:19(a-b) and Figure 3:19(c-f) show the standard deviation and SNR based on measured and detected photon counts, respectively. As previously shown in [202], the results based on the detected photon count are more accurate, with a loss of SNR occurring in the pile-up regime (panels d and f), as opposed to an unrealistic increase beyond the fundamental shot-noise-limited SNR (panel b). It can also be observed that the SNR of the sensor is governed by the shot noise and pile-up limitations for approximately three decades of photon counts, both in temporal domain and spatial domain. In spatial domain, the removal of hot pixels, which amount to around 2% of the array, improves the SNR consistency with the shot noise and pile-up limited model, particularly at low light levels.

The shot-noise limited behavior of SwissSPAD2 in the photon-counting regime is further demonstrated in Figure 3:19(g). In this figure, the photon count distribution of the uniformly illuminated pixels in single-photon level illumination was plotted against the Poisson-fitted values. The agreement between the fitted and measured values shows that the camera is able to detect single photons with no other significant noise contribution other than shot noise. Figure 3:19(h) shows the SNR comparison with shot noise at higher photon count levels, below the deep pile-up region. A general agreement with the measured and fitted values can still be observed despite a small mismatch, likely caused by the mild pile-up effect which is also theoretically expected to exist.



⁵ Spatial photon count variation is the photon count variation of all pixels in the array in a single frame. Temporal photon count variation is the photon count variation of multiple frames captured sequentially by a single pixel. Standard deviation of the total photon count is calculated after the subtraction of the mean dark count from the mean total photon count. This subtraction is performed in each pixel individually.

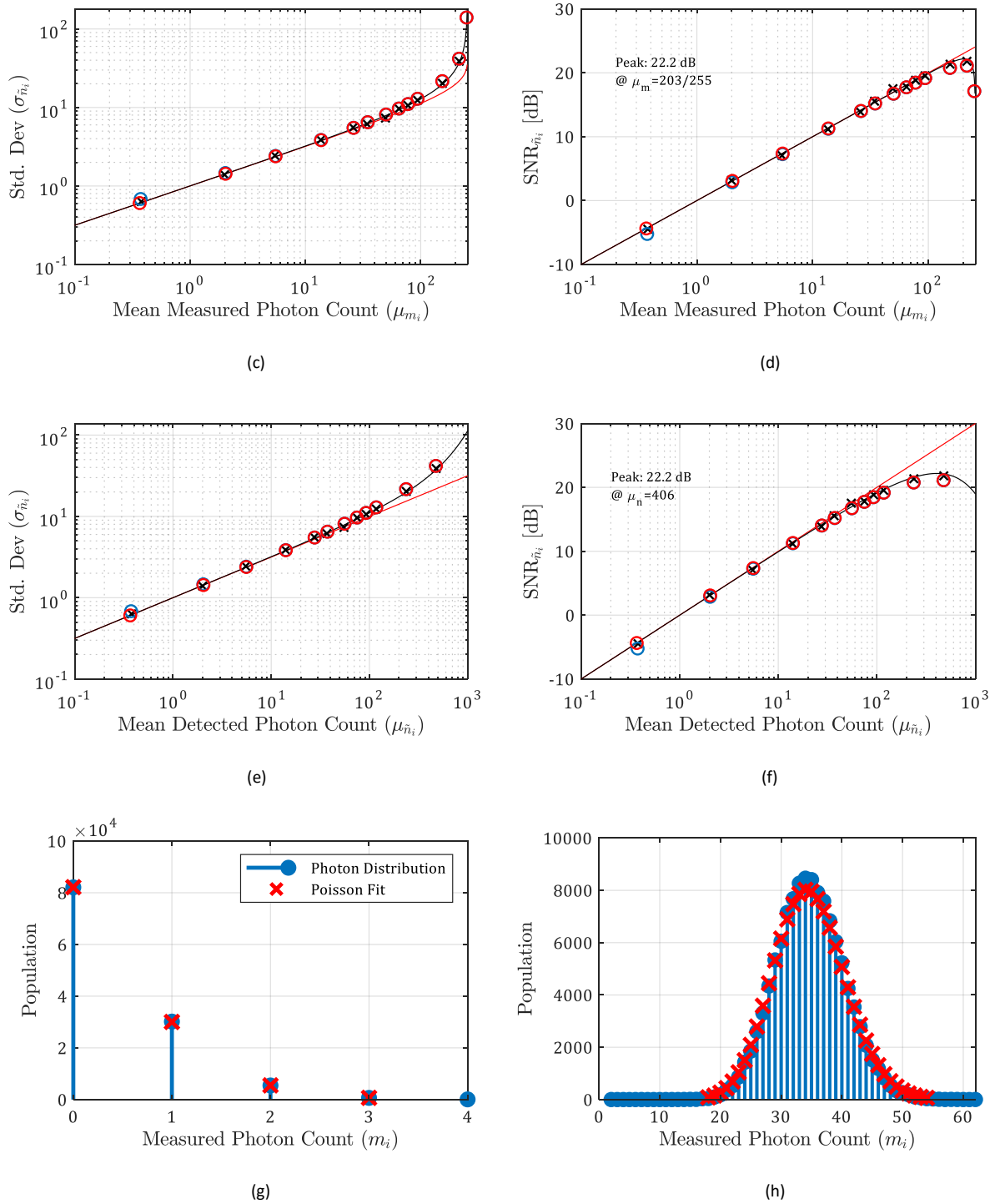


Figure 3:19 Dynamic range characterization of SwissSPAD2 for 8-bit images. Number of pixels: 472×256. The images are captured under 13 different light levels ($LL_0 - LL_{12}$). Exposure time: LL_{0-10} : 1.63 ms, LL_{11} : 3.26 ms, LL_{12} : 6.52 ms. (a) Standard deviation and (b) signal to noise ratio (SNR) of the *measured photon count* m_i as a function of the mean measured photon count. (c) Standard deviation and (d) SNR of the *estimated detected photon count* \tilde{n}_i as a function of the mean measured photon count. (e) Standard deviation and (f) SNR of the *estimated detected photon count* as a function of the estimated mean detected photon count. There is a consistent match between the measured SNR values and the theoretical SNR values based on the model that includes both shot noise and pile-up. (g) Measured photon count spatial distribution under low light level. The match between the measured and Poisson fit values demonstrates the single-photon counting ability of the imager. (h) Measured photon count distribution under higher light level. A good agreement with the Poisson fit values can still be observed, despite the mildly visible effects of pile-up, as expected. The distributions in (g) and (h) were generated after the removal of hot pixels which consist of approximately 2% of the entire array.

As it is now proven that the active clock-driven recharge pile-up model discussed above accurately represents the SNR of SwissSPAD2, $n_{i\max}$ and $n_{i\min}$ required for the determination of the dynamic range can be estimated based on this model. In [202], $n_{i\max}$ for SPAD arrays is estimated by

$$\tilde{n}_{i\max} = F_{cs}(2^b - 1), \quad (3.13)$$

where F_{cs} is a factor determined by the readout architecture type and the SNR drop tolerance in the determination of the maximum photon count. Using the value of F_{cs} determined in [202] for active clock-driven recharge and a 3 dB SNR decrease criterion, i.e. $F_{cs} = 2.842$, and a bit depth $b = 8$, $n_{i\max}$ can be estimated as $\tilde{n}_{i\max} = 724.71$. For $\tilde{n}_{i\min}$, the estimated value of $n_{i\min}$, considering that it is mostly dominated by dark noise, the following expression can be used

$$\tilde{n}_{i\min} = \sqrt{T_{int} \times DCR}, \quad (3.14)$$

where T_{int} is the total integration (exposure) time of the frame, and DCR is the dark count rate. Using $T_{int} = 1.63$ ms and $DCR = 7.5$ cps, $\tilde{n}_{i\min}$ is found as $\tilde{n}_{i\min} = 0.11$. Inserting the values of $\tilde{n}_{i\max}$ and $\tilde{n}_{i\min}$ in Equation (3.11), the dynamic range of SwissSPAD2 in this experiment can be estimated as $DR = 76.3$ dB. The upper bound of the dynamic range is determined by the minimum T_{int} that is achievable by the system. The expression of T_{int} for 8-bit frames in the intensity mode (i.e. when no periodic gate asynchronous with the readout is present) is given by

$$T_{int} = 255 \times T_{exp} \quad (3.15)$$

in global shutter and

$$T_{int} = 255 \times t_{row} \times N_{row} \quad (3.16)$$

in rolling shutter, where T_{exp} is the binary exposure time, t_{row} is the readout duration of a row, and N_{row} is the number of rows in the frame. In the rolling shutter mode, the minimum value of T_{int} is equal to 2.61 ms for $t_{row} = 40$ ns and $N_{row} = 256$. In the global shutter mode, assuming a minimum binary exposure time of $T_{exp} = 10.8$ ns, the minimum T_{int} is equal to 2.75 μ s. Following the methodology described above, the upper bound of the dynamic range for rolling shutter and global shutter is determined as 74.3 dB and 104.1 dB, respectively. If the condition which yields the lowest dynamic range is considered as $\tilde{n}_{i\min} = \tilde{n}_{i\max}$, the lower bound of the dynamic range can be determined as 0 dB.

3.4.6 Gate characteristics

As discussed in Chapter 2, the gate profile characteristics of a gated image sensor have a significant influence on its timing performance. For this reason, the gate characteristics of both SwissSPAD2 and SwissSPAD3 were measured. The experiments were conducted according to the following methodology. The collimated fiber output of a 790 nm pulsed laser (PiLas, A.L.S., Germany; pulse width (FWHM): ~ 40 ps, PRF: 20 MHz) is directed towards the camera sensor without any objective lens. The beam width was set to a size that illuminates the entire sensor active area. Other than ensuring that all pixels receive sufficient signal intensity to achieve acceptable SNR and no part of the array is saturated, no further attempt was made to achieve spatial uniformity of the light level. The trigger signal which synchronizes the laser and the camera was generated by the laser controller. During the acquisition of the data, the gate – whose width is fixed throughout the entire experiment – was shifted by the minimum allowed gate step over more than one full laser period. Subsequently, several

parameters that determine the time gating performance were calculated from the raw data. These parameters are the rise/fall times and gate width together with their standard deviations, and gate skew. The definitions and performance impacts of these parameters are detailed in Chapter 2, Sections 2.4 and 2.5. A detailed list of experimental parameters, which may show differences between the testing of the two cameras, is included in the caption of the figures representing a particular experiment.

Figure 3:20 represents the gate characteristics of SwissSPAD2. The experiment conducted to characterize the gate profile is illustrated in Figure 3:20(a-b). In this experiment, the gate signals were shifted with respect to the laser pulse with equal steps of G_s , for a measurement window larger than the laser period ($T_{l_{sr}}$) (Figure 3:20(a)). The plot of the recorded counts of the gate positions represents the gate window profile (Figure 3:20(b)). The shape of the measured gate window profile is shown in Figure 3:20(c). In this plot, the photon counts of the pixels in the array are normalized to their average high intensity levels before being superimposed, to visually show the uniformity of the gate characteristics across the array. The gate width, rise time and fall time are annotated inside the plot. The gate width, which is equal to 13.1 ns in this particular data set, can be defined by the user within the range of 10.8-22.8 ns. The fall time of the gate (617 ps) is significantly slower than the rise time (378 ps). Considering that in this data set the gate moves forward with respect to the laser pulse, the falling edge of the gate window is determined by the recharge signal, whereas the rising edge of the gate window is determined by the falling edge of the gate signal. This behavior is consistent with the description of the gating structure of SwissSPAD2 in Subsection 3.3.2, where the causes of this discrepancy were discussed.

The rising and falling edge position distributions of the gate window are plotted in Figure 3:20(d) and (e), respectively. The histograms of the edge positions are shown in the main windows of the figures, whereas the 2D color maps are displayed in the insets. For both edges, horizontal and vertical gradients in the edge positions are visible. The vertical gradient, which is a result of the signal propagation delay on the column wires, was theoretically expected. According to distributed RC wire model, the gate delay between the two adjacent pixels is indeed expected to be non-uniform throughout the wire, and to increase non-linearly as the signal propagates towards the farthest pixels. On the other hand, a significant horizontal gradient was not expected, as the skew between the columns was targeted to be minimized by means of signal trees.

The subsequent examination of the chip layout showed that the cause of the horizontal skew was the voltage drop in the central region of the signal trees due to a relatively high parasitic resistance between the DC supply pads of the chip and the power wires of the buffers in the control signals tree. Due to the structure of the layout, this parasitic resistance was highest at the center and lowest on both sides of the tree. As a consequence, the two central outputs of the first level of the tree exhibit a slower propagation time than the two outputs situated away from the center. This effect can be clearly observed in Figure 3:20(b). In the 2D map, there are two distinct regions with a clear boundary between them. The first and the fourth vertical quarters of the array (columns 0-127 & 384-511) receive the gate signal falling edge earlier than the second and the third quarters (columns 128-383). Therefore, their gate window rising edge positions are at a later time point. This inequality in the propagation time of the central and peripheral branches also applies to later levels of the tree. However, due to the decline of the signal wire length in the later tree levels, the magnitude of the skew is not as large as in the first level. The skew of the second tree level is also visible in the inset of Figure 3:20(b). Each large vertical color band consisting of one quarter of the array is also divided into 4 equally sized color bands with clear boundaries, but relatively smaller contrast.

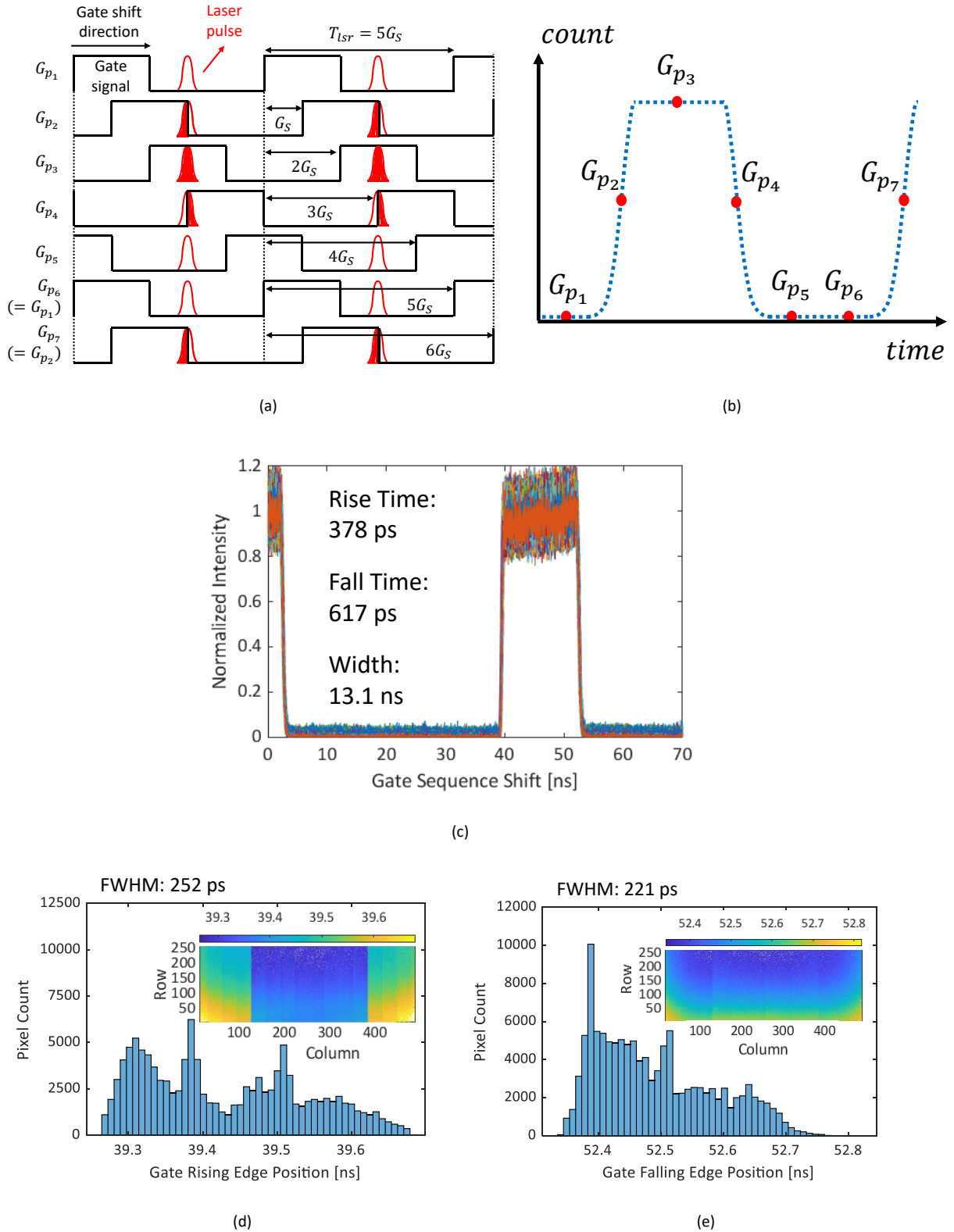


Figure 3:20 Gate characteristics of SwissSPAD2. (a-b) The conceptual illustration of the experiment which was conducted to generate the gate profile. G_p : gate position, T_{LSR} : laser period, G_S : gate step. For visualization purposes, $T_{LSR} = 5G_S$ was selected, as opposed to $T_{LSR} = 2,800G_S$, which represents the actual experimental settings. (c) The gate window profile. The transition times and the gate width are annotated in the figure. The gate width is user-programmable, and the minimum gate width in the internal laser trigger mode is 10.8 ns. The response of every 4th pixel is plotted. (d-e) Spatial distribution of the (d) rising edge and (e) falling edge position of the gate window. The insets show the 2D color map of the gate edge positions. Experimental parameters: Laser controller trigger mode: internal, laser PRF: 20 MHz, array size: 472×256, shutter mode: global shutter, bit depth: 10, gate step: 17.86 ps, number of gate positions: 4,000, integration window: 71.4 ns.

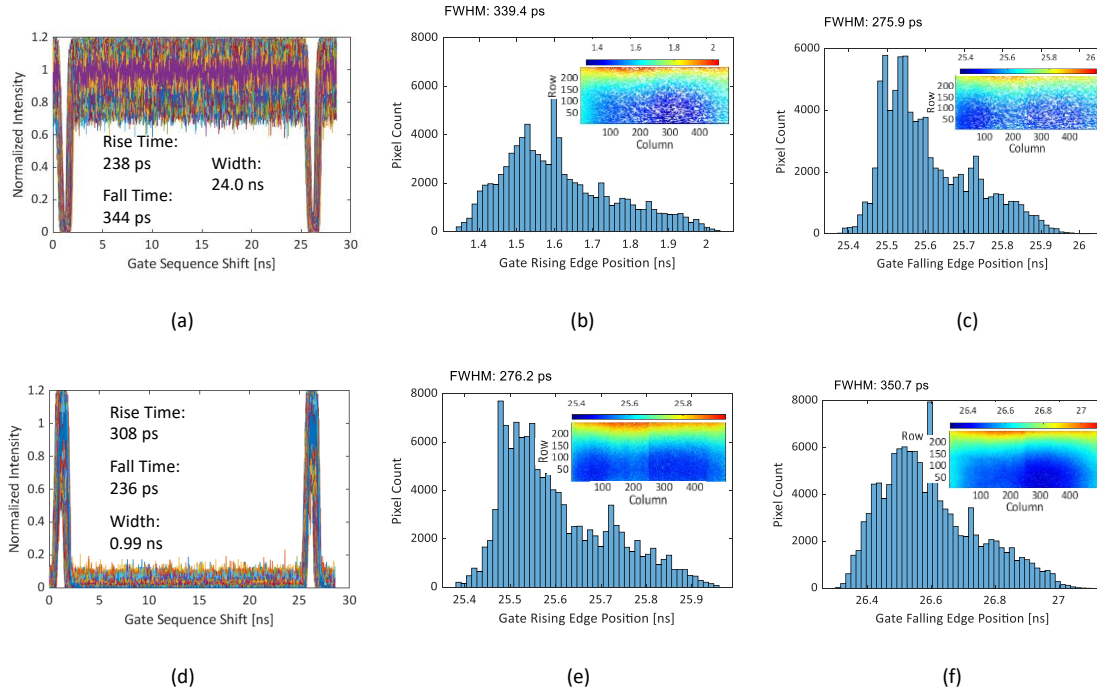


Figure 3:21 Gate characteristics of two gate channels of SwissSPAD3 for the minimum achieved gate width. (a-c) and (d-f) represent the characteristics of the first (G1) and the second gate channel (G2), respectively. (a, d) Gate window profile. The transition times and the gate width are annotated in the figure. Percentage of plotted pixels: G1: 4%, G2: 21%. (b-e, f) Spatial distribution of the (b, e) rising edge and (c, f) falling edge position of the gates. The insets show the 2D color map of the gate edge positions. Experimental parameters: Laser wavelength: 637 nm, laser controller trigger mode: internal, laser PRF: 40 MHz, array size: 499×249, binary frame rate: 49.8 kfps, shutter mode: rolling shutter, bit depth: 8, gate step: 17.9 ps, number of gate positions: 1,600, integration window: 28.6 ns, V_{ex} : 6 V.

The gate characteristics of SwissSPAD3 were measured using a similar methodology and experimental setup to SwissSPAD2. The main difference was that two output bits were recorded for each pixel instead of one as in SwissSPAD2, in order to capture both gate channels. The second gate output was calculated in post-processing by subtracting the gate output (*Out_2*) from the intensity output (*Out_1*). As explained in the earlier sections, the gate architecture of SwissSPAD3 ensures that the two gate channels are contiguous, non-overlapping and achieve 100% duty cycle. In order to observe any possible gate performance effects of simultaneous, fast readout, the chip was operated in rolling shutter, at 40 MHz laser PRF, and 49.8 kfps binary frame rate. During the calculation of the frame rate, the readout duration of both outputs of a single frame is considered to be the binary acquisition time. Therefore, the maximum frame rate is equal to approximately half of SwissSPAD2. The position of the separator between the two gates, which determines the ratio of their widths, can be controlled prior to the experiment with a step of 17.9 ps from 0 to the laser period.

Figure 3:21 displays the characteristics of the two gate channels of SwissSPAD3 under the settings which generate the shortest achievable gate width of 0.99 ns. The rise and fall times of the gates are significantly more symmetrical compared to SwissSPAD2. In addition, the values of the transition times are closer to the faster edge of SwissSPAD2 than the slower edge. These results are consistent with the expected gate behavior of the two cameras based on their respective gating architectures, which was described in Subsection 3.3.2. While the achievement of a short gate width is an important figure of merit for the imager, it is also possible to generate two contiguous gate channels with approximately equal widths. This gating configuration was demonstrated in Figure 3:22. In this configuration, the width of the gate channel 1 (*G1*) and channel 2 (*G2*) are 12.1 ns and 12.9 ns, respectively. Though, it is possible to achieve a smaller gate width difference between *G1* and *G2* by adjusting the size of *G2* with a resolution of 17.9 ps.

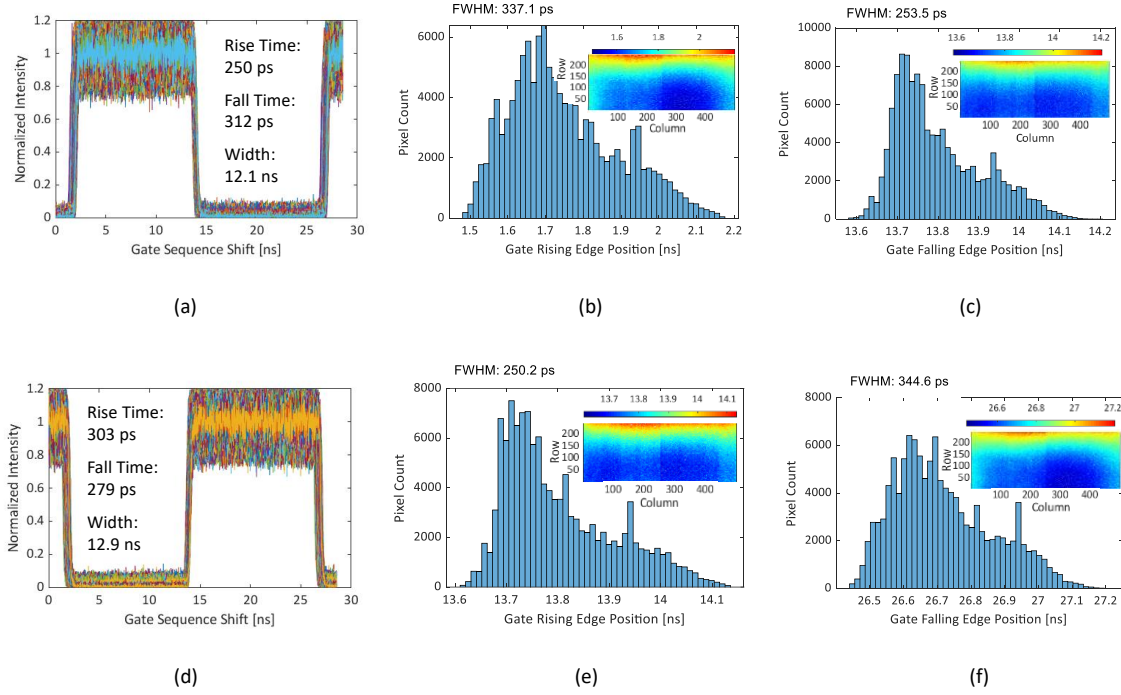


Figure 3: Gate characteristics of two gate channels of SwissSPAD3 with comparable gate widths. (a-c) and (d-f) represent the characteristics of the first (G1) and the second gate channel (G2), respectively. (a, d) Gate window profile. The transition times and the gate width are annotated in the figure. Percentage of plotted pixels: G1: 11%, G2: 10%. (b-c, e-f) Spatial distribution of the (b, e) rising edge and (c, f) falling edge position of the gates. The insets show the 2D color map of the gate edge positions. Experimental parameters: Laser wavelength: 637 nm, laser controller trigger mode: internal, laser PRF: 40 MHz, array size: 499×249, binary frame rate: 49.8 kfps, shutter mode: rolling shutter, bit depth: 8, gate step: 17.9 ps, number of gate positions: 1,600, integration window: 28.6 ns, V_{ex} : 6 V.

Compared to SwissSPAD2, a significant decrease in the horizontal skew was observed in SwissSPAD3. The FWHM of the rising/falling edge position distribution of the pixels in a single row (Row 125) ranges between 81.2 ps and 134 ps. While the reported skew measurement results of imagers do not always include the skew characteristics in a single dimension, and the skew varies with the pixel format and the size of the array, it can be concluded that the overall skew performance for SwissSPAD3 is on-par if not better than the published state-of-the-art. This assessment can be supported by three examples from the prominent large-format SPAD imagers in the literature. In [119], the horizontal gate skew is in the order of several hundreds of picoseconds for the falling edge, and close to a nanosecond for the rising edge. In [120], the FWHM of the skew for the 512×128 array is 139 ps. Due to the relatively lower number of rows, which reduces the RC delay across the signal wire, the contributions of the horizontal and vertical skew are closer to each other than SwissSPAD3, indicating a FWHM of the horizontal skew that is not significantly different from 139 ps. This conclusion was reached based on the pixel distribution maps of the gate edges that are provided in the paper, in the absence of a quantitative representation. Finally, in [122], the FWHM of the gate position variation for a 1024×500 array was measured as 410 ps, with seemingly comparable contribution of horizontal and vertical skew, according to the 2D distribution map.

Low skew in SwissSPAD3 was achieved by reducing the parasitic resistance between the supply/ground pads and the signal tree via layout improvements, and by reducing the disparity of parasitic capacitance between multiple branches of the tree via increasing the spacing between the wires. The vertical distribution of the gate edges is in agreement with the distributed RC chain model, where the delay of the signal is proportional to the second power of the length of the wire [231]. According to this model, the propagation delay between two adjacent pixels increases with the distance of the pixels from the signal tree. This effect manifests itself

as the long tail in the gate edge position distribution histograms in Figure 3:21 and Figure 3:22 (b, c, e, f). By comparing Figure 3:21 and Figure 3:22, it can also be observed that the gate transition times and the skew have low dependence on the gate width. This behavior proves that the IR drop in the voltage supply wires of the gate tree is not sufficiently high to adversely influence the gate performance, unlike in SwissSPAD2.

Figure 3:23 shows the photon count distribution characteristics of the fully open gate positions in SwissSPAD3, for both gates. The main objective of the analysis represented by this figure was to quantitatively compare the gate shape to an ideal rectangular pulse, in the region where the gate is fully open. The rising and falling edges of the gate window were excluded from the analysis as they clearly deviate from a rectangular shape, due to the various factors discussed in Chapter 2, Subsection 2.4.2. In this comparison, which was performed independently for each pixel due to the spatial light level non-uniformity, the statistical dispersion of the photon counts of multiple open gate positions was compared to its theoretical minimum limit. The open gate positions were identified for both $G1$ and $G2$ from the gate profiles shown in Figure 3:23(a-b). The two factors that determine this theoretical limit are shot noise and pile-up. The expression for the standard deviation of photon count caused by these two effects, $\sigma_{\tilde{n}_i}$, is provided in Equations (2.34)-(2.36) [202]. For an accurate comparison of the theoretical and experimental values, the experimental value of $\sigma_{\tilde{n}_i}$ was calculated after converting the measured gate photon counts, m_i , to the estimated detected photon counts, \tilde{n}_i , using the pile-up correction method provided in Equation (2.33) [203]. Finally, the theoretical and experimental SNR were calculated using Equation (3.12). As the objective of this analysis was to characterize the gate shape rather than to determine the SNR of a single gate image, the mean dark count (μ_{DC}) was not subtracted from the mean total photon count (μ_C). While theoretical and experimental $\sigma_{\tilde{n}_i}$ were found using separate methods as explained above, the mean of the same \tilde{n}_i was used for the calculation of both SNR values.

The results of this experiment are summarized in Figure 3:23(c-f). The similarity of the gate window to a rectangular gate was estimated by analyzing the difference of the theoretical and experimental SNR values in decibels. This difference is expected to be zero for a perfectly rectangular gate, and positive for a gate which deviates from the rectangular shape. Figure 3:23(c-d) show the spatial distribution maps of the SNR difference for the two gates, and Figure 3:23(e-f) show their histograms. The results show that the average SNR of $G1$ and $G2$ for 499×498 pixels is only 0.12 dB and 0.06 dB lower than the theoretical limit, respectively. Therefore, these results suggest that the gates of SwissSPAD3 exhibit very small deviation from a rectangular shape.

In the histogram, it can be seen that for some pixels, the difference of the theoretical and experimental SNR appears to be negative, indicating a noise level lower than shot noise and pile-up limit. A probable cause of this unrealistic result is the fact that the mean value of n_i is obtained by estimation, rather than measurement, thereby resulting in an error. A second probable cause is that the pile-up correction formula in Equation (2.33) is valid in SwissSPAD3 only when one of the gates has a multi-bit photon count of zero. Due to the uncorrelated background signal and dark counts, the gates have non-zero counts even when the laser pulse is fully outside the gate window; thereby causing a deviation in the pile-up behavior.

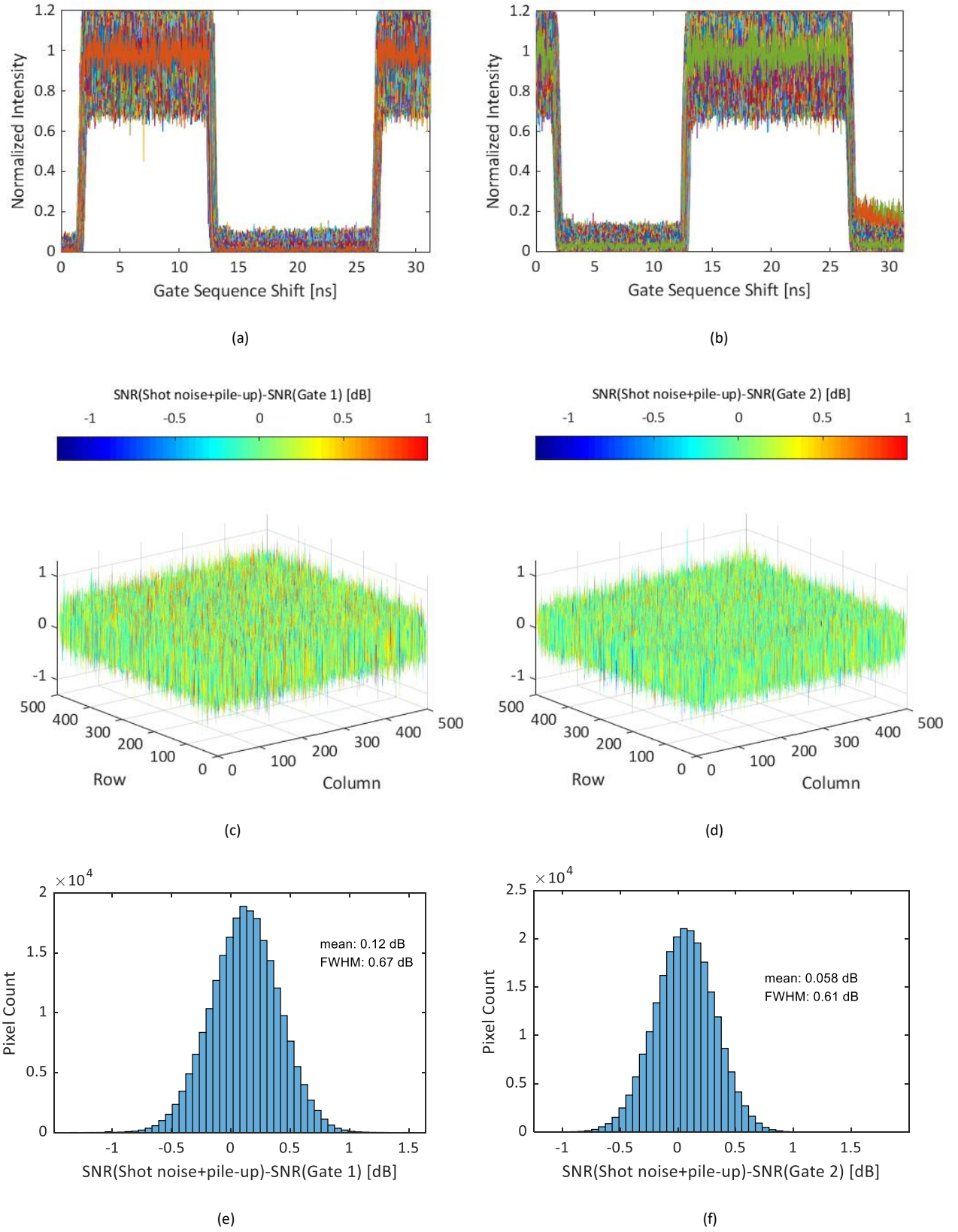


Figure 3:23 Photon count distribution characteristics of the fully open gate positions in SwissSPAD3. (a-b) Gate window shape for (a) gate 1 (G1) and (b) gate 2 (G2). Plotted pixel percentage: G1: 10%, G2: 14%. (c-d) Theoretical and experimental SNR difference of the detected photon counts of the fully open gate positions of (c) G1 and (d) G2. The difference between the theoretical SNR which is limited by the shot noise and pile-up, and the measured SNR, displayed in (c-d), was calculated for each pixel separately. (e-f) Spatial distribution of the difference between the theoretical and

experimental SNR, for (e) $G1$ and (f) $G2$. As the SNR of $G1$ and $G2$ is only 0.12 dB and 0.058 dB lower than the theoretical values, respectively, it can be concluded that the photon counts of the fully open gate positions are independent of the gate positions themselves, and limited by shot noise and pile-up. This proves that in the flat parts of the gate window, there is negligible deviation from the characteristics of a rectangular gate. Experimental parameters: Pixel format: (a-b): 499×249, (c-f) 499×498, laser PRF: 40 MHz, laser wavelength: 510 nm, number of fully open gate positions: $G1$: 480, $G2$: 580, V_{ex} : 6 V.

3.5 Microlens design and characterization

3.5.1 Motivation and theory

As discussed in Chapter 1, one of the main obstacles to the widespread adoption of SPAD technology as the detector of choice for scientific imaging is its shortcomings in PDE. Particularly in large-format imagers, a degradation in PDE is observed due to a decrease in fill factor as the number of pixels increases in the array. The contributors of the fill factor decrease in small pixels include the guard ring and the minimum spacing between adjacent SPADs due to crosstalk minimization purposes and fabrication design rules. For imagers designed in 2D CMOS technology, spacing requirements for pixel electronics and signal and power transfer wires may introduce additional fill factor limitations, whose importance often increases with the number of transistors in the pixel. For all of the reasons above, fill factor improvement for SPAD sensors is an essential part of the design process. A widely used method of fill factor improvement is the use of microlenses. By being deposited on each pixel in the array, microlenses concentrate the incoming light on the pixel active area, thereby improving the effective fill factor of the pixels.

The most important performance indicator of microlenses is the ability to concentrate the impinging photons on the active area. This ability is expressed by a parameter called concentration factor. This term, first introduced within the framework of SPAD-based imagers in [232], was originally defined as the ratio of the output and input irradiance of an optical concentrator, expressed as

$$CF = E_o/E_i, \quad (3.17)$$

where E_o and E_i are output and input irradiance, respectively. Based on similar principles, the concentration factor can also be defined as the ratio of the photon counts upon a detector with microlenses and without microlenses, under the same illumination and operating conditions [221]. Due to its convenience in experimental measurements, we used the latter definition in our work, which can be quantitatively expressed as a function of the angle of incidence (θ) and the wavelength (λ) as

$$CF(\theta, \lambda) = \frac{C_{\mu l}(\theta, \lambda)}{C_{no_ul}(\theta, \lambda)}, \quad (3.18)$$

where $C_{\mu l}$ and C_{no_ul} are the photon count rates with and without microlenses, respectively. It is often practical to use an effective fill factor for a microlensed pixel, $FF_{\mu l}$, which can be calculated using the equation below

$$FF_{\mu l}(\theta_{max}, \lambda) = FF_{no_ul} \times CF_{avg}(\theta_{max}, \lambda), \quad (3.19)$$

where θ_{max} is the maximum angle of incidence in a given optical setup, FF_{no_ul} is the native fill factor and CF_{avg} is the average concentration factor of a range of incidence angles θ between 0° and θ_{max} . The definition of CF_{avg} that is used in this work is expressed as

$$CF_{avg} = \frac{2CF(\theta = \varphi = 0^\circ) + CF(\theta = \theta_{max}, \varphi = 0^\circ) + CF(\theta = \theta_{max}, \varphi = 45^\circ)}{4}, \quad (3.20)$$

where θ is the angle of incidence and φ is the azimuth angle (Figure 3:24). The term “concentration factor” used throughout the section refers to CF_{avg} defined in Equation (3.20), for a particular θ_{max} or numerical aperture (NA).

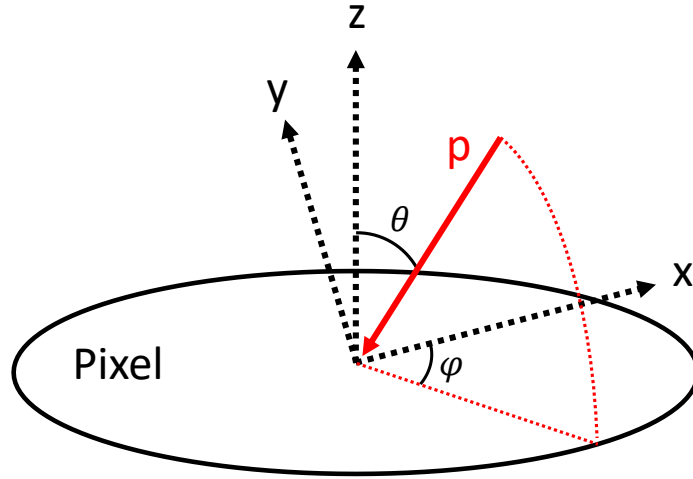


Figure 3:24 Illustration of the angles used during the discussion of microlenses. p is the photon trajectory, θ is the angle of incidence, and φ is the azimuth.

An ideal convex lens can focus collimated light impinging at normal incidence ($\theta = 0^\circ$) on a single point, called the focal point. Therefore, such a lens structure can theoretically increase the fill factor to 100%. However, this goal cannot be achieved in reality due to the minimum gap requirement between adjacent microlenses, which is a fabrication constraint. This minimum gap typically does not scale with the pixel pitch, and therefore causes a more significant fill factor degradation in smaller pixels. The expression of the maximum fill factor as a function of the pixel pitch (PP) and the minimum microlens gap (G_{min}) for square microlenses at $\theta = 0^\circ$ is given by

$$FF_{\mu l \max} = \left(\frac{PP - G_{min}}{PP} \right)^2. \quad (3.21)$$

Microlenses can also improve the transmittance of the optical stack of the system, defined as the percentage of the incident photons that arrive from the outermost surface of the optical stack to the SPAD multiplication region. The highest contributor to the loss in transmittance is Fresnel loss, defined as the loss due to reflection at an interface between two media with different refractive indices. The Fresnel loss, R , between two media with refractive indices n_1 and n_2 is expressed by

$$R = \left| \frac{n_1 - n_2}{n_1 + n_2} \right|^2 \quad (3.22)$$

for $\theta = 0^\circ$. When absorption and scattering are neglected, the transmittance of an interface i is equal to the refracted light percentage at the interface, expressed as $T_i = 1 - R_i$. For an optical stack with multiple media, the overall transmittance is given by

$$T = \prod_{i=1}^n T_i, \quad (3.23)$$

where n is the interface number. The list of the media in a typical camera with microlenses is illustrated in Figure 3:25.

Air (n_1)
Anti-reflective coating (n_2)
Microlens (n_3)
Passivation layer (n_4)
Middle layers (n_5)
Silicon (n_6)

Figure 3:25 A typical optical stack for CMOS/SPAD cameras that shows the media in the trajectory of the photons until their detection.

Anti-reflective (AR) coating can optionally be used between two media as an intermediate layer to minimize the reflections between them. According to Rayleigh, the optimal refractive index of the AR coating is the geometric mean of the refractive indices of the surrounding media, given by

$$n_2 = \sqrt{n_1 n_3}. \quad (3.24)$$

The passivation layer, whose refractive index is in general significantly higher than its surrounding media, can be removed, at least partially, to further improve the overall transmittance. The middle layers represent the full stack of the CMOS process. For simplicity, they are represented, during the microlens development process, as a single layer with an average refractive index. In the microlenses developed for SwissSPAD2, AR coating was not used, and the passivation layer was not removed.

After the inclusion of the transmittance ratio, the maximum fill factor expression can be rewritten as

$$FF'_{\mu l \max} = \frac{T_{\mu l}(\lambda)}{T_{no-\mu l}(\lambda)} \left(\frac{PP - G_{min}}{PP} \right)^2. \quad (3.25)$$

Note that the value of $FF'_{\mu l \max}$ can exceed 100%. This results from the fact that while a decrease in Fresnel loss (i.e. increase in the overall transmittance of the optical stack) should by definition be considered as a gain in PDP, it is conventionally considered as a gain in fill factor. When this is the case, it can be understood that $FF'_{\mu l \max} > 100\%$ does not violate the fundamental laws of physics, as $PDE'_{\mu l \max} = PDP \times FF'_{\mu l \max}$ never exceeds 100%.

Deviations in the fill factor of a microlensed pixel from the value defined by Equation (3.25) can occur due to various elements which have adverse effects on the concentration factor. These elements include non-zero maximum angle of incidence and microlens fabrication constraints. The concentration factor is highly dependent on the range of photon angles of incidence. Since this range is determined by the properties of the optical setup in a particular experiment, the target application is an important factor in the microlens design process.

For most applications, this range is determined by the numerical aperture (NA) of the lens facing the image sensor (also called tube lens in microscopy setups), which is calculated from the following equation:

$$NA = n \sin(\theta_{max}), \quad (3.26)$$

where n is the refractive index of the medium, which is approximately 1 for air in the case of tube lens. Since CF_{avg} decreases as θ_{max} increases, it also decreases with higher NA. Due to this relation, in a microscopy setup, it is desirable to minimize the NA of the tube lens. On the other hand, the microlens fabrication constraints can introduce two effects: mismatch of the focal point with the multiplication region, and optical aberrations. Aberrations indicate the inability to concentrate the light on a single point due to errors in the lens shape and multiple wavelengths of incident light. The main cause of aberrations in a single lens system are the imperfections in the curvature of the lens. Furthermore, the square shape of the lens, which is selected to cover the largest possible area of a square pixel, introduces additional aberrations, compared to a circular lens.

3.5.2 Design of the microlenses in SwissSPAD2

The microlenses of SwissSPAD2 were separately designed for two different optical setups optimized for microscopy and photography. In microscopy, owing to the relatively large available space and the lack of need for a large angle of view, the optical setup can be configured as a telecentric lens system, which is illustrated in Figure 3:26(a). In this system, all chief rays coming out of the tube lens are parallel to the optical axis. As a result, the average angle of incidence is zero for all pixels in the sensor. In addition, the focal length (therefore, also the NA) of the objective lens and the tube lens can be selected independently. Typically, the objective lens is selected with high NA in order to collect the maximum possible amount of light from the sample, and the tube lens is selected with low NA to transmit the light to the camera sensor at a small angle of incidence.

On the other hand, the most common lens system in photography is a fixed focal length lens, which generally displays the characteristics of a single convex lens (even though compound lenses can be used, their main purposes being to minimize optical aberrations). A conceptual representation for this setup is shown in Figure 3:26(b). The reasons for this choice include the angle of view requirement for the addition of perception of depth, and compactness. In these lenses, high NA is typically preferred for the ability to collect more light, resulting in higher SNR, and shallower depth of field. For this reason, the maximum angle of incidence on the sensor in a photography setup is generally higher than in microscopy. Additionally, the average angle of incidence increases as the position of the pixel moves away from the center of the array towards the edges (see also below).

In our design process, the microlenses were targeted for microscopy and designed for low NA, which allows the highest concentration factor. In this design, NA of 0.02 was selected based on the specifications of commonly used widefield microscopy setups, as it was originally given to us by Stefan Hell in 2005. From Equation (3.26), the maximum angle of incidence can be calculated as 1.14 degrees. For the microlens configuration that is designed for photographic lenses, the NA choice was 0.25, which corresponds to an f-number of 1.9, and maximum angle of incidence of 14.5 degrees.

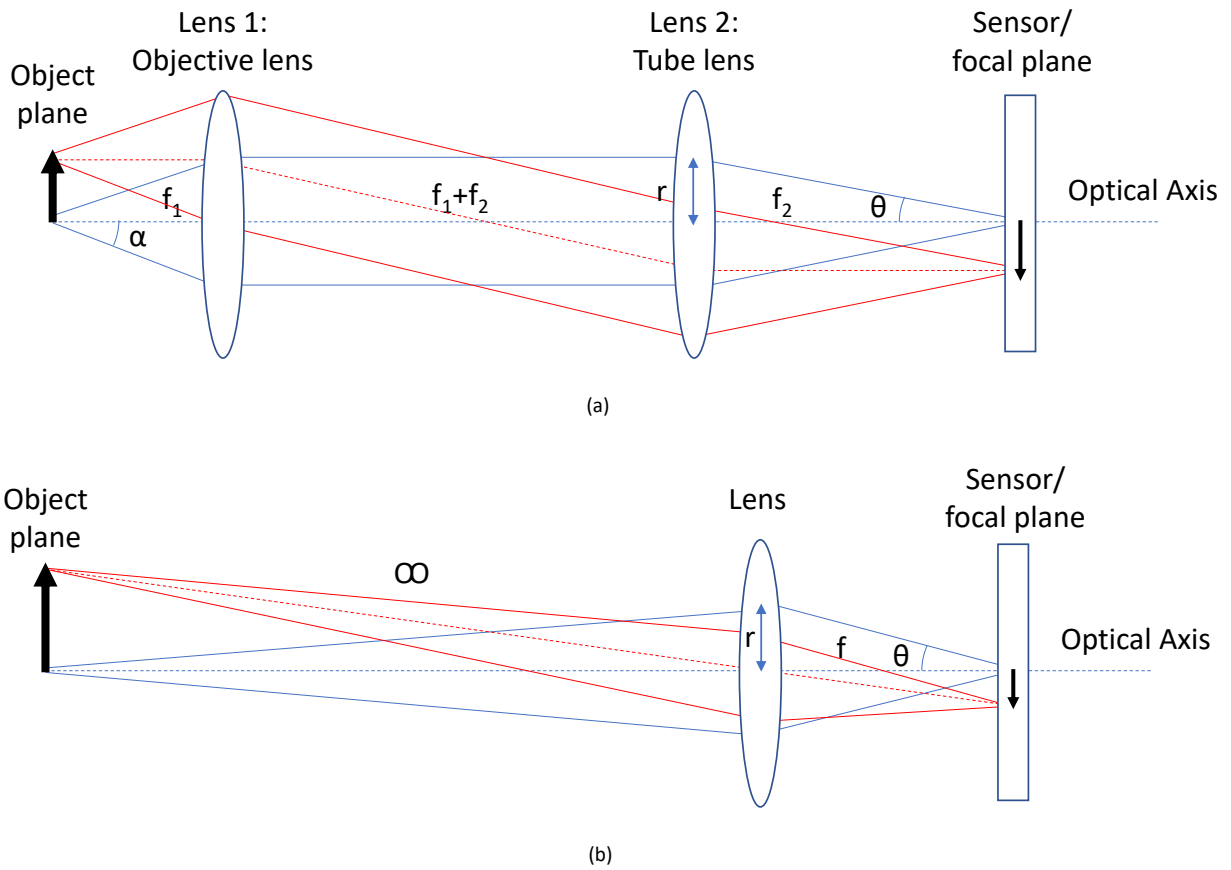


Figure 3:26 (a) Diagram of a telecentric optical system, often used in microscopes. High α is desirable for high SNR and resolving power, whereas the ability to reduce θ allows the maximization of the concentration factor of microlenses. The two downsides of this approach, zero angle of view and the large focal length requirement for tube lenses, do not introduce any negative effect to the system due to the lack of need for the sense of depth in microscopy and the relatively higher availability of space for the tube lens compared to a photography setup. (b) Diagram of a conventional single-lens optical system typically used in photography. High θ is desirable to improve photon sensitivity by collecting more light from the object, and minimize depth of field. In addition, concerns for compactness introduce an upper limit for f , further restricting the lower bound of θ . These constraints limit the maximum achievable concentration factor for microlenses.

In fixed focal length lens setups where the average angle of incidence deviates from zero in the peripheral pixels, the area where the microlenses concentrate the light also deviates from the center of the pixel active area. In severe cases in which the concentration area of light falls fully outside the pixel active area, significant sensitivity loss is observed in the peripheral pixels, causing an undesired effect called vignetting. The severity of vignetting depends on two parameters: the size of the image sensor and the NA of the lens. For a $NA = 0.25$, vignetting correction was applied due to the significance of this effect. In this correction method, the offset between the central points of the active area and light concentration area is simulated for each pixel in the array, and a new microlens pitch that is smaller than the pixel pitch is determined such that the offset is corrected for all pixels. In this method, the difference between microlens pitch and pixel pitch was selected to be 11 nm, which corresponds to a 2.8- μm -offset between microlens and pixel at the outermost pixels in the array.

The design parameters of the microlenses, which can be optimized for the highest performance, are shown in Figure 3:27 [233]. The diameter of the lens is limited by the pitch and the minimum required gap between the adjacent lenses. This gap is introduced by fabrication constraints and is a major contributor of the maximum theoretical fill factor. The sag is the thickness of the curved part of the lens, and is expressed as a function of

its diameter and the radius of curvature. The relation between the sag, diameter and radius of curvature is given by [233]

$$Sag = R_{curvature} - \sqrt{R_{curvature}^2 - \left(\frac{Diameter}{2}\right)^2}. \quad (3.27)$$

The residual height is the distance from the bottom of the microlenses to the substrate surface. This parameter must be set in such a way that the focal point of the lenses overlaps with the multiplication region of the SPAD, therefore maximizing the avalanche probability of the absorbed photons. In SwissSPAD2 microlens design, the microlens pitch is equal to the pixel pitch in the low-NA version, and shorter by 11 nm in the high-NA version with vignetting correction feature. The gap was set to approximately 1.5 μm , the minimum value allowed by the fabrication process. The sag and the residual height that yield the highest concentration factor were calculated using ray tracing simulations.

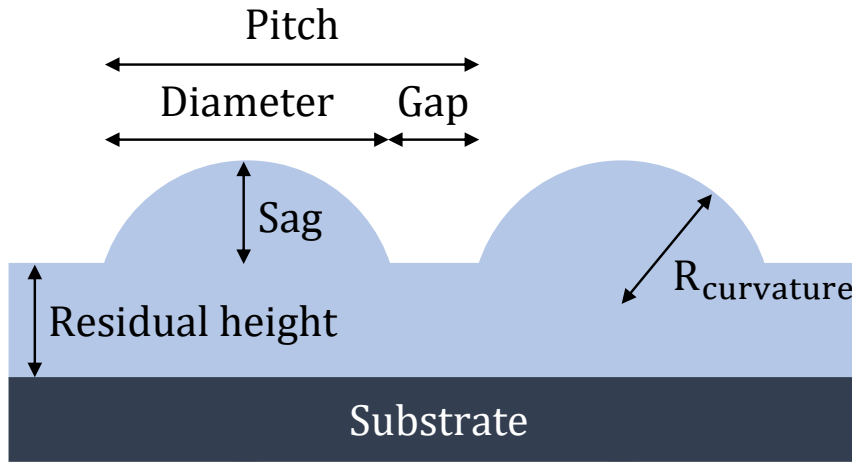


Figure 3:27 Microlens design parameters. Adapted from [233].

The description of the SwissSPAD2 microlens fabrication process is available in [233].

3.5.3 Measurement results

As theoretically expected, the ray tracing simulation results show that the low-NA configuration yields higher concentration factor than the high-NA configuration. For the round SPAD, for $NA = 0.02$, the highest simulated concentration factor is equal to 7.78, which yields a fill factor of 81.7%. Considering that the theoretical maximum fill factor for a microlens gap of 1.5 μm and pixel pitch of 16.38 μm is equal to 82.5%, excluding the transmittance ratio, the gap can be regarded as the primary fill factor limitation. The highest simulated concentration factor drops to 5.32 for $NA = 0.25$, corresponding to a fill factor of 55.9%. For the square SPAD, the highest simulated concentration factor for $NA = 0.02$ and $NA = 0.25$ are 6.29 and 5.11, respectively.

A SwissSPAD2 sample with microlenses designed for the low-NA optical setup, whose photomicrograph is shown in Figure 3:28, was tested with a fluorescence microscope (IX81, Olympus, Japan), where the emission light is directed at the sensor with approximately normal incidence. Testing was performed using a *convallaria majalis* sample that exhibits autofluorescence behavior and low photobleaching rate. Two SwissSPAD2 cameras with round SPADs were used for the experiment. The images were captured sequentially using the same camera port of the microscope. The exposure time was selected in such a way that ensures sufficient SNR and

avoids high pile-up region for both cameras. After the data acquisition, the mismatch between the field of view of the two images was removed by cropping the images in post-processing. The pile-up correction operation, described in Equation (2.33), was applied to both images [203]. In addition, the hot pixels, which account for around 1% of the total number of pixels, were removed from the images of both cameras, and dark count correction was performed by subtracting a reference dark image from the signal image in both cases.

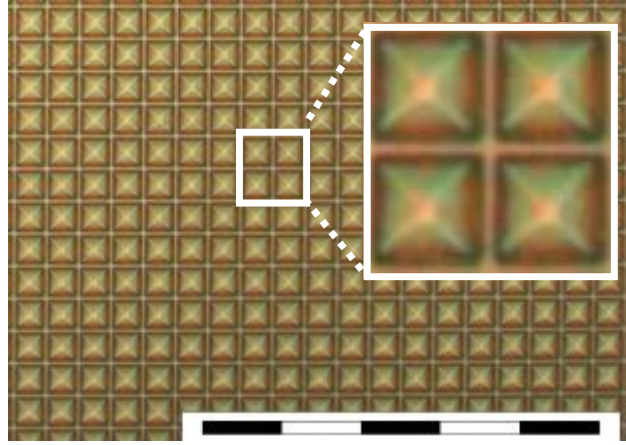


Figure 3:28 Photomicrograph of a group of pixels in SwissSPAD2 with microlenses. Scale bar: 200 μm .

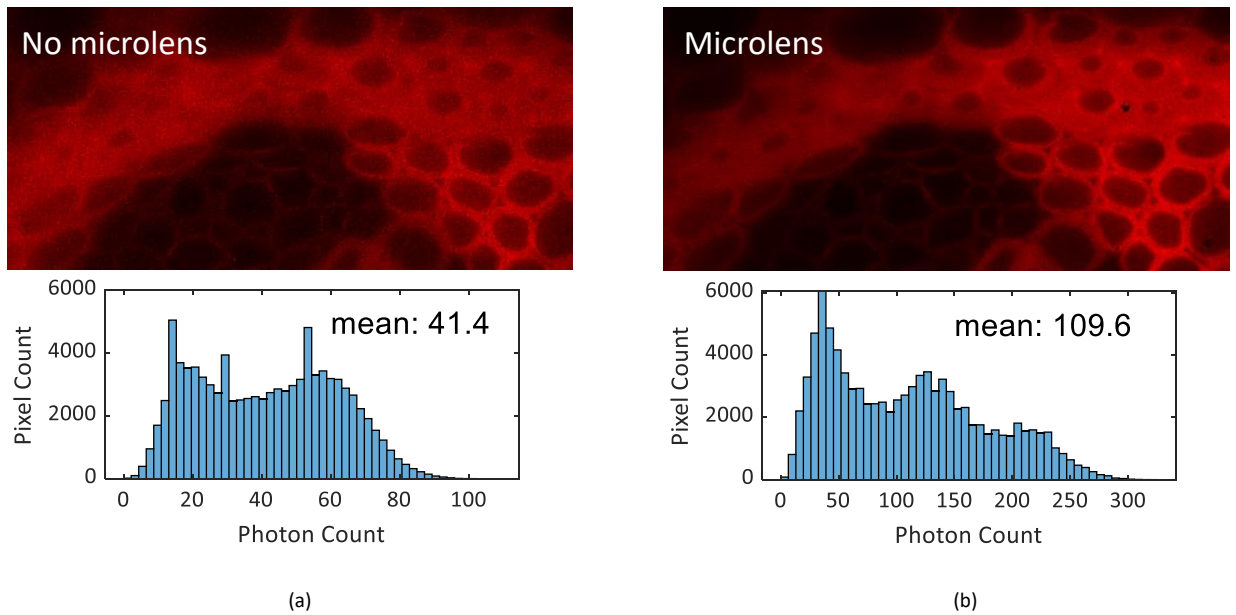


Figure 3:29 Fluorescence intensity image of a *convallaria majalis* sample using a SwissSPAD2 (a) without and (b) with microlenses [81]. Experimental parameters: V_{ex} : 6.5 V, array size: 453×210, bit depth: 10, integration time: 3.21 ms, $\lambda_{emission}$: 607 nm, pile-up correction: on. Hot pixels with 1% highest dark count rate in the array were corrected using an interpolation method based on setting their intensity values to the mean of the four nearest-neighbor pixels. Different color brightness scales were used in the two images to achieve maximum contrast.

The results of this experiment presented in Figure 3:29(a) and (b) show a concentration factor of 2.65, which corresponds to an effective fill factor of 27.8%. This value is lower than the simulation results for normal incidence. Further investigation on the fabricated microlens structure has revealed that the main cause of this loss was optical aberrations. The source of this aberration was the distortion in the microlens curvature that occurred during the thermal reflow phase of the fabrication process.

To verify that the distorted lens shape was the main cause of the low concentration factor, a new experiment was conducted with the goal of measuring the maximum concentration factor that can be achieved with the camera at any angle of incidence. The description of the experimental setup is as follows. A camera is placed 3.67 m away from the light source. The light beam was directed at the camera with a size that is sufficient to uniformly illuminate the entire 512×256 array. Considering that the total active area of the 512×256 sub-array of the sensor is equal to $8.4 \times 4.2 \text{ mm}^2$, the pixel-to-pixel variation of the angle of incidence of the light is less than 0.13 degrees. For the light source, a continuous wave collimated LED (M590L3-C, Thorlabs, USA) with a peak wavelength of 590 nm and bandwidth of 18 nm was used. No objective lens was placed in front of the sensor. The camera was mounted to a PCB holder that permits its rotation in two orthogonal directions. For each sample, the first step was the calibration stage, in which the angle of the camera was optimized for highest photon sensitivity under fixed illumination settings. The tuning of the angle of incidence for highest CF was necessary due to the aforementioned distortions in the microlens shape which occurred during fabrication, and the resulting lateral shift in the focal point. To achieve that, the camera was operated in live mode during the calibration process, thus generating 8-bit intensity outputs at video rate and returning the average intensity of all pixels. After the identification of the optimal angle, a sequence of intensity images was captured at this angle. The pile-up correction formula described in Chapter 2 was applied to find the corrected photon counts [203]. To subtract dark counts and background illumination, an additional series of images of the same scene was captured with no LED illumination. In this experiment, the values of the optimal angles of incidence were not measured.

The microlensed sensors used for this experiment consist of 5 chips with round SPADs and 2 chips with square SPADs. In addition, two sensors with no microlenses (one with round SPADs, one with square SPADs) were used as reference. All sensors were tested sequentially, with separate calibration processes for finding the optimal angle of each of them. To test the possible influence of the excess bias voltage (V_{ex}) on the concentration factor, the measurements were repeated at four V_{ex} values. The measurement results of the chips with round and square SPADs are shown in Figure 3:30(a) and (b), respectively. At 6.5 V_{ex} , which represents the preferred operation mode, the CF of round pixels ranges between 3.65 and 4.2, whereas the CF of the square pixels ranges between 3.2 and 3.6. Compared to the CF of 2.65 that was measured with a round SPAD in the previously reported experiment, these results indicate a smaller disparity between the simulation and measurement results. This difference is even smaller when the measurement results are instead compared to revised simulation results which take into account the aforementioned distortions in the fabricated lens curvature. According to these results, for a sag of $4.4 \text{ }\mu\text{m}$ and gap of $1 \text{ }\mu\text{m}$, the revised CF is $\sim 25\%$ lower than the ideal lens shape at a residual height of $10 \text{ }\mu\text{m}$, and $\sim 55\%$ lower at a residual height of $20 \text{ }\mu\text{m}$. The CF is largely unaffected by V_{ex} in the range between 4.5 and 6.5 V. At 3.5 V_{ex} , a considerable increase or decrease was observed in some samples. A possible cause of this deviation is the breakdown voltage difference between multiple samples. Since the PDP dependence on V_{ex} decreases at higher values of V_{ex} in this SPAD design, as shown in Figure 3:12 [212, 202], a possible CF error due to PDP disparity between two sensors is less significant at higher V_{ex} , which leads to more uniform CF values.

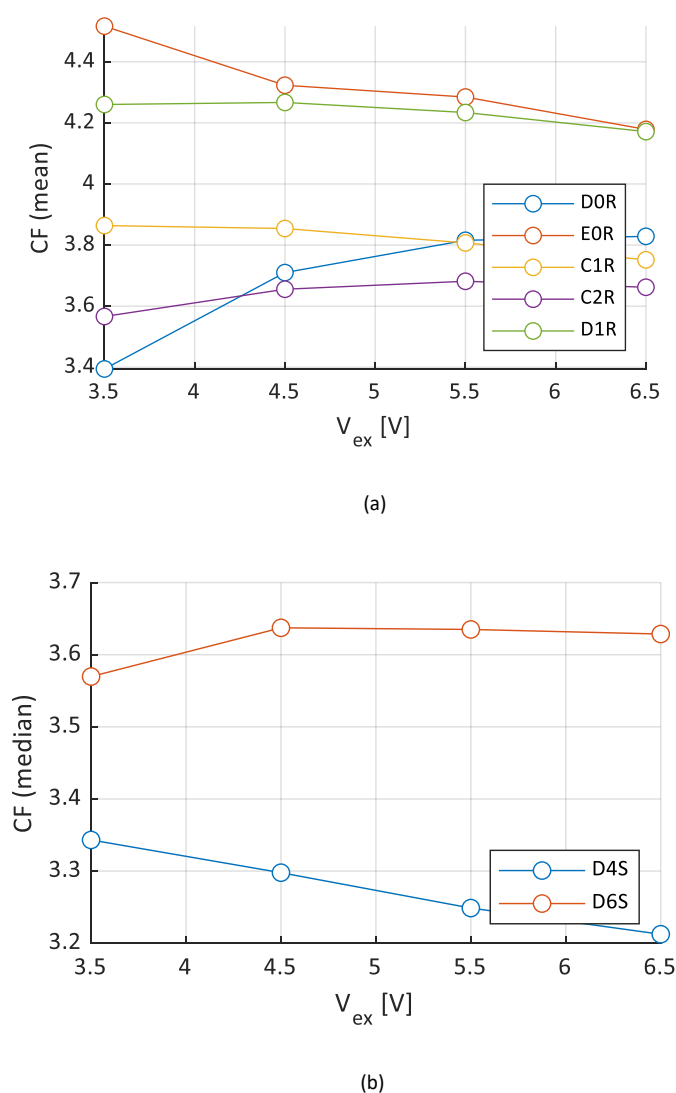


Figure 3:30 Concentration factor characterization of multiple SwissSPAD2 samples with microlenses at various excess bias voltages. SPAD shape: (a) round, (b) square with rounded corners. Measurements were done at optimal angle of incidence for each sample; the tuning of the angle was performed during the calibration stage prior to the measurements.

The spatial distribution of the measured CF for two SwissSPAD2 samples is presented in Figure 3:31. The SPAD types employed in the samples in Figure 3:31(a) and (b) are round and square with rounded corners, respectively. The selected samples are the ones with highest measured CF according to Figure 3:30, for their respective SPAD types. These results were obtained at $V_{ex} = 6.5$ V, and 16×16 binning was applied for visualization purposes. While the results are promising, further experiments with more controlled ambient light level and angle of incidence are required for a more comprehensive quantitative assessment.

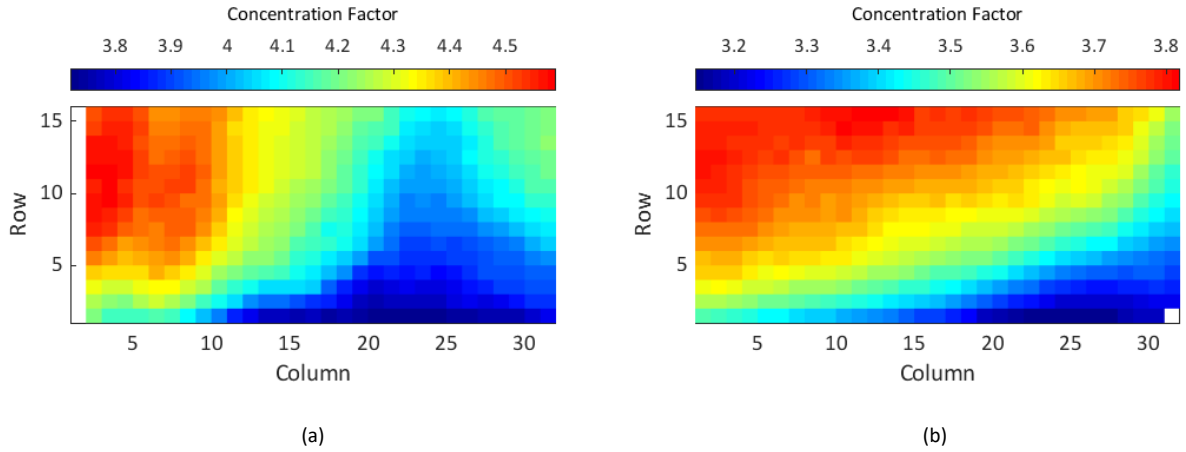


Figure 3:31 Concentration factor spatial distribution of two SwissSPAD2 samples with the highest measured average CF for their respective SPAD types. SPAD shape: (a) round (sample: E0R), (b) square with rounded corners (sample: D6S). Measurements were done at optimal angle of incidence for each sample; the tuning of the angle was performed during the calibration stage prior to the measurements. $V_{ex} = 6.5$ V, Binning: 16×16 .

3.6 State-of-the-art comparison

The comparison of the SPAD model used in SwissSPAD2 and SwissSPAD3 with the state-of-the-art SPADs designed in a standard CMOS process is given in Figure 3:32. Only the SPADs fabricated in a standard CMOS process were included in the list, for the purpose of evaluating only the devices that are compatible with large-format image sensors. The comparison results show that this SPAD achieves one of the best combinations of PDP and DCR.

The comparison of the SwissSPAD2/3 imager specifications with several state-of-the-art large-format SPAD imagers is given in Table 3:1. SwissSPAD2 achieved the largest array format and lowest median DCR compared to the imagers in its category at the time of its introduction; it later got surpassed in both categories by MegaX [122]. The two relative advantages of SwissSPAD2 compared to MegaX are its higher PDP and frame rate. The higher PDP is partially caused by the higher excess bias allowed in SwissSPAD2; in MegaX, the lack of cascode transistor due to pixel area constraints limits the excess bias. The second possible cause of the PDP difference is the discrepancy between the active area indicated in the layer drawings and the real active area, whose effect becomes more significant as the pixel pitch gets smaller. The characterization of this discrepancy, which is likely caused by the p-well lateral diffusion in the p-i-n SPAD used in both cameras [212], requires further investigation. The higher frame rate is caused by the fact that the total throughput of both image sensors is similar (256 total I/O blocks with 100 MHz data transfer rate per block), while MegaX contains approximately 4 times more pixels.

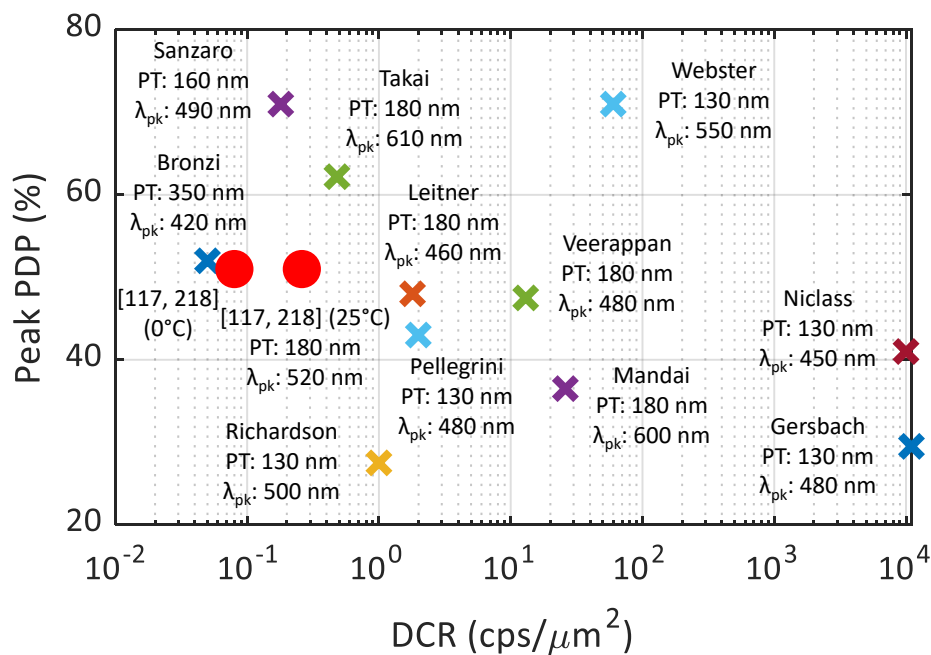


Figure 3:32 Photon detection probability (PDP) and dark count rate (DCR) comparison between the p-i-n SPAD used in SwissSPAD2/3 and state-of-the-art standard CMOS-based FSI SPADs [117]. DCR represents the median value. The PDP and 0° C DCR results are based on nanoSPAD [218], a chip that employs a variant of the p-i-n SPAD design used in SwissSPAD2/3 with 107.5 μm^2 active area as opposed to the 28.3 μm^2 active area in the case of SwissSPAD2/3.

SwissSPAD3, while generally exhibiting similar performance parameters as SwissSPAD2, differs in several categories, particularly in gate characteristics and frame rate. To the best of the author's knowledge, SwissSPAD3 is the only large-format SPAD imager which achieves 100% duty cycle with its two contiguous gate channels covering the entire laser period. In addition, its minimum achievable gate width, 0.99 ns, is the shortest compared to other large-format SPAD imagers.

Table 3:1 Performance comparison of SPAD image sensors with high spatial resolution [117] [81].

	SwissSPAD2 [117, 81]	SwissSPAD3	MegaX [122]	[123]	[121, 234]	[181]	[235]	[118]	[119]	[120]	[57]
Process Technology	180 nm CMOS	180 nm CMOS	180 nm CMOS	65 nm CMOS	130 nm CIS	350 nm HV CMOS	350 nm HV CMOS	130 nm CIS	130 nm CIS	350 nm HV CMOS	40 nm CMOS
Array Format	512×512 (512×256 used so far)	500×500	1024×1000	1200×900	320×240	160×120	32×32	256×256	256×256	512×128	192×128
Pixel Pitch	16.38 μm	16.38 μm	9.4 μm	6 μm	8 μm	15 μm	25 μm	8 μm	16 μm	24 μm	18.4 μm (H) 9.2 μm (V)
Fill Factor (Nominal)	10.5%	10.5%	A: 7.0% B: 13.4%	-	26.8%	21%	20.8%	19.6%	61%	5%	13%
Fill Factor (With Microlenses)	28%-47%	-	-	-	50%	-	-	-	-	60%	42%
Chip Size	9.5×9.6 mm ²	9.6×9.7 mm ²	11×11 mm ²		3.4×3.1 mm ²	3.42×3.55 mm ²	-	3.5×3.1 mm ²	5×5 mm ²	13.5×3.5 mm ²	3.2×2.4 mm ²
Maximum PDP	~50% @ 520 nm (V _{ex} = 6.5 V)	-	A: 10.5% @ 520 nm B: 26.7% @ 520 nm (V _{ex} = 3.3 V)	-	39.5% @480 nm (V _{ex} = 1.5 V)	-	-	-	39.5% @480 nm (V _{ex} = 3 V)	46% @490 nm (V _{ex} = 4 V)	-
Median DCR	7.5 cps/px 0.26 cps/ μm^2 (V _{ex} = 6.5 V)	-	A: 0.4 cps/px 0.06 cps/ μm^2 B: 2.0 cps/px 0.17 cps/ μm^2 (V _{ex} = 3.3 V)	-	47 cps/px 2.7 cps/ μm^2 (V _{ex} = 1.5 V)	580 cps/px 12.3 cps/ μm^2 (V _{ex} = 3 V)	500 cps/px 3.8 cps/ μm^2 (V _{ex} = 5 V)	50 cps/px 4.0 cps/ μm^2 (V _{ex} = 2 V)	6.2 kcps/px 40 cps/ μm^2 (V _{ex} = 1.5 V)	366 cps/px 12.7 cps/ μm^2 (V _{ex} = 4.5 V)	25 cps/px 1.14 cps/ μm^2 (V _{ex} = 1.5 V)
Readout Noise	0	0	0	-	0.168e ⁻	-	-	-	Negligible	0	-
Uniform SPAD Pitch	Yes	Yes	A : Yes B : No	Yes	No	No	-	-	No	Yes	No
Maximum Frame Rate*	97.7 kfps (1 bit)	49.8 kfps (1 bit)	24 kfps (1 bit)	450 fps	16 kfps (1 bit)	486 fps (5.4 bit)	50 fps (1.5 V analog output)	4 kfps (3-bin histogram)	100 kfps (1 bit)	156 kfps (1 bit)	18.6 kfps (1 bit)
Number of Gate Channels	1	2	1	1	1	1	1	1	1	1	1

* For cameras with multiple gates, a frame includes the images of all gates.

3.7 Future work

Despite their capabilities in time-resolved imaging and high overall performance among large-format SPAD imagers, there is significant room for improvement in the SwissSPAD2/3 architecture. These improvements are crucial for increasing the competitive advantages of SPAD imagers in both FLIM and life sciences in general, compared to more established widefield detector technologies such as ICCD, position sensitive detectors and lock-in pixels.

One of the most important drawbacks of SwissSPAD2/3 is low PDE. The native fill factor of 10.5%, peak PDP of 45-50% and microlens concentration factor of 4.2 yield a peak PDE of $\sim 20\%$. There are several possible measures that can be taken to improve PDE. Firstly, the removal of the protective layer on the chip surface can increase the transmittance of the materials above the SPADs. Secondly, a pixel architecture which supports higher SPAD excess bias voltages than the currently available cascode scheme can further increase the PDP by increasing the avalanche probability. Thirdly, the guard ring size could be reduced by finding the minimum width of the epi layer which prevents premature edge breakdown. The increase in DCR due to higher electric field caused by some of the measures above can be mitigated by reducing the temperature of the sensor. Furthermore, by optimizing the microlens design and fabrication process, the concentration factor of the microlenses can be improved to a level approaching its theoretical limitation.

It would also be desirable to extend the dynamic range of the camera with minimal sacrifice from the frame rate or spatial resolution. Currently, the acquisition time of a frame with a given bit depth is limited by the minimum required time to acquire a binary frame, which is equal to $\sim 10.2 \mu\text{s}$ for all 256 rows in a half array, i.e. 40 ns per row. Due to the binary nature of the in-pixel memory, the dead time of the detector system is also equal to $\sim 10.2 \mu\text{s}$. While this value is acceptable for widefield systems owing to the fact that they compensate for the slower local count rate by capturing thousands of pixels in parallel, this value is far from the local photon count rate potential of the SPAD technology, which is limited by the dead time of a SPAD. The throughput limitations of SwissSPAD2/3 are introduced by the total number of available I/O blocks on the chip and the FPGA, the bandwidth of a single I/O block, and the pullup/pulldown times of the output signal wires. Considering these factors, the most feasible method to reduce the dead time of a SPAD image sensor is to add an in-pixel counter. For a given sensor throughput, an N-bit counter reduces the overall dead time by a factor of $(2^N - 1)/N$. The important drawbacks of this approach are the loss of temporal resolution for a given binary frame rate, and decrease in fill factor due to the area occupied by the counters and the additional output signal wires for the extra bits.

Another feature of interest is global shutter operation without sacrificing duty cycle. The inherent advantage of global shutter is the ability to capture fast phenomena without the distortions of the rolling shutter mode caused by non-uniform exposure windows. However, in global shutter operation of SwissSPAD2, the camera must be insensitive during readout. The most common method to overcome this problem is to duplicate the available memory space in each pixel such that the count of a pixel can be shifted between the two memory blocks in a single clock cycle between the acquisition of two frames. This method allows the first memory block to capture a frame in the global shutter mode while the previous frame is read from the second memory block. The main challenge of this method is the requirement of additional electronics, potentially causing a fill factor penalty.

An additional performance improvement direction is the number of gate channels in a pixel. In SwissSPAD3, the number of channels was increased from one to two. In contiguous gate channel implementations, for a given total measurement window, the number of required gate channels to achieve 100% duty cycle is inversely proportional to the gate width. Therefore, a further increase in the number of gate channels can improve photon economy.

Global shutter feature can also be added to SwissSPAD3. As discussed in Subsection 3.3.1, the implementation of global shutter to the SwissSPAD3 pixel requires additional global recharge/reset and gate signals, to recharge the SPAD before the exposure, and start/stop the exposure, respectively. The impact of these two additional signals on the fill factor and the skew of the other critical signals can be potentially minimized by layout optimizations.

Most of the listed performance improvements require an increase in the density of the electronics in a pixel. Without causing a decrease in the fill factor, the most effective method to address this requirement is to design the new generation imager with the above features monolithically in a more advanced technology node, or using 3D stacked CMOS technology.

Chapter 4 Application results

The material presented in this chapter is based on [81] (Section 4.2), [211] (Section 4.3) and [217] (Section 4.4).

4.1 Sensor performance trade-offs in the target applications

In Chapter 3, the architectures and technical specifications of the image sensors developed in this thesis were presented. Another important stage of this research is the use of these imagers in target applications. The primary target application for SwissSPAD2/3 is widefield time-domain FLIM using the phasor approach. This chapter focuses mainly on the demonstration of this application with SwissSPAD2. In addition, the assessment of performance trade-offs is an essential step during the investigation of the feasibility of a detector for a given application. Therefore, the trade-offs of SwissSPAD2 which have a significant impact on the widefield FLIM performance are also discussed.

The first significant application trade-off at detector level is frame rate vs. spatial resolution. In SwissSPAD2, the maximum frame rate is determined by the readout speed of a single row. Since one frame is acquired by the sequential readout of up to 256 rows in one half of the chip – which can be carried out in parallel to the other half – the frame rate decreases as the number of rows in the frame increases. The relation between the binary frame rate and the number of rows is given by

$$f_{gs} = (T_{read} + T_{exp})^{-1} = (t_{row}N_{row} + T_{exp})^{-1} \quad (4.1)$$

in global shutter and

$$f_{rs} = T_{read}^{-1} = (t_{row}N_{row})^{-1} \quad (4.2)$$

in rolling shutter, where f_{gs} and f_{rs} are the frame rate in global shutter and rolling shutter, respectively, t_{row} is the readout duration of a row, N_{row} is the number of rows selected for readout, T_{exp} is the exposure time in the global shutter mode, and T_{read} is the readout time. According to Equations (4.1) and (1.1), the frame rate can be improved by reducing the number of rows; however, this improvement comes with a loss in spatial resolution. The frame rate can be also partially improved by reducing the number of columns. However, unlike rows, all of which are read sequentially, only four adjacent columns sharing an output pad are read sequentially, as shown in Figure 4:1.

Therefore, column-level frame rate improvement can only be achieved by reducing the number of columns read out by each of the 128 output pads. By using this method, the frame rate improvement can only reach a factor of 4. While this method does not affect the field of view, it reduces the fill factor by turning off the pixels that are skipped, thereby increasing the percentage of the photo-insensitive area in the pixel array. It also reduces the spatial resolution by a factor that is equal to the frame rate improvement. For this reason, column-

level frame rate improvement is not preferred in applications requiring high photon sensitivity and high spatial resolution, such as FLIM.

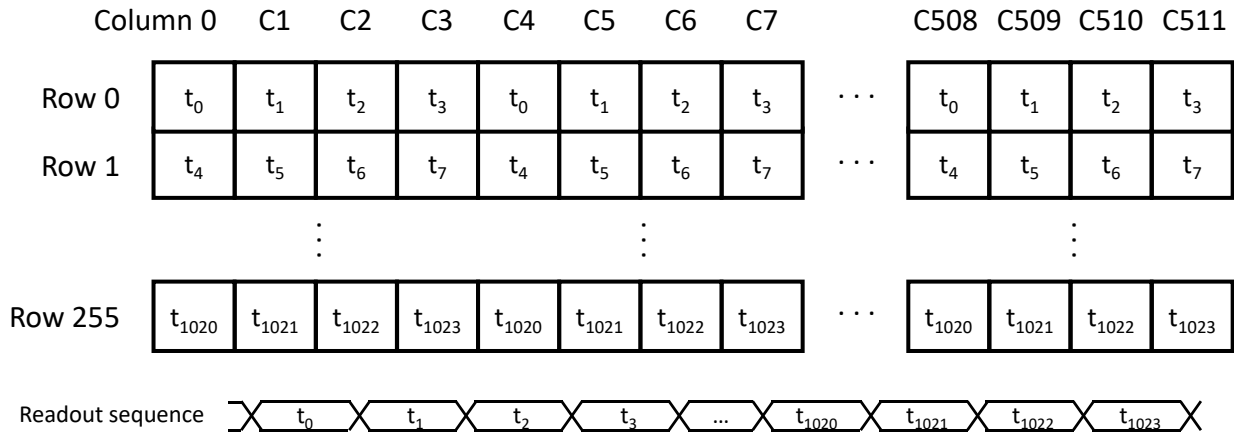


Figure 4:1 Readout sequence of the pixels in one half of SwissSPAD2 with 512x256 pixels. From this diagram, it can be observed that while the total frame readout time is directly proportional to the number of rows that are selected, it can only be reduced by up to a factor of four by reading out fewer columns.

In rolling shutter mode, since f_{rs} is proportional to N_{row}^{-1} , the gain in frame rate is equal to the loss in spatial resolution when adjusting N_{row} . In global shutter mode, the same rule only applies when T_{exp} decreases proportionally to N_{row} . The role of the exposure time in the trade-off between the frame rate and spatial resolution is linked to a separate trade-off between frame rate and duty cycle. Indeed, unlike rolling shutter in which the duty cycle (also termed temporal aperture) is always 100%, the duty cycle Y when operating in global shutter is given by

$$Y = \frac{T_{exp} \times G_W}{(T_{exp} + T_{read}) \times T_{lsr}}, \quad (4.3)$$

where G_W is the gate width and T_{lsr} is the laser period. In the intensity mode, the value of G_W is equivalent to T_{lsr} . Equations (4.1) and (4.3) indicate that for a given T_{read} , an increase in the exposure time increases the duty cycle, while decreasing the frame rate. This trade-off has a significant role in the operation of SwissSPAD2 for FLIM, as the camera can only operate in global shutter when using the gated mode.

Figure 4:2 illustrates the trade-offs between the duty cycle and other important FLIM parameters, in global shutter mode. Figure 4:2(a) shows that for a given gate width, an increase in binary exposure time results in an increase in duty cycle, but a decrease in frame rate. On the other hand, as shown in Figure 4:2(b), when the exposure time is fixed, the increase in the gate width improves the duty cycle, at the expense of higher F-value, indicating lower photon economy.

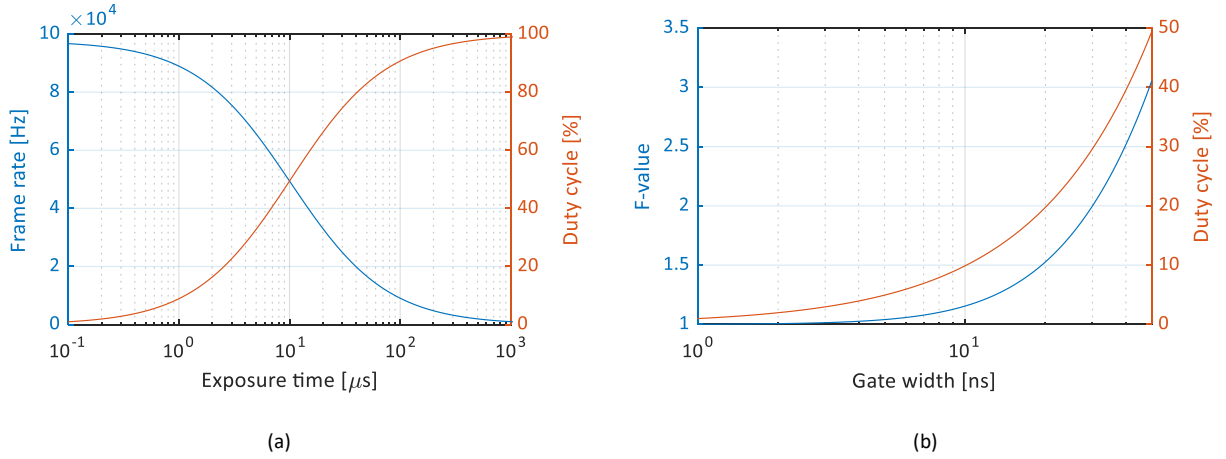


Figure 4:2 (a) Trade-off between frame rate and duty cycle as a function of the binary exposure time, T_{exp} , in global shutter mode. The calculations are based on Equations (4.1) and (4.3). $T_{read} = 10.24 \mu s$, $T_{lsr} = 50 ns$, $G_W = 50 ns$. (b) Trade-off between F-value and duty cycle as a function of the gate width, G_W , in global shutter mode. The calculations are based on Equations (2.27) and (4.3). $T_{exp} = 10 \mu s$, $T_{read} = 10.24 \mu s$, $T_{lsr} = 50 ns$, $\tau = 5 ns$.

Another major trade-off is between dynamic range, spatial resolution and frame rate. Particularly in fluorescence microscopy where the photon budget of the samples can be low to avoid photobleaching, it is common to apply spatial binning to a number of adjacent pixels and group them into a single region of interest (ROI), in order to improve the signal-to-noise ratio (SNR) without sacrificing the frame rate. SNR of fluorescence lifetime determination is expressed as

$$SNR = 20 \log \left(\frac{\mu_\tau}{\sigma_\tau} \right), \quad (4.4)$$

where μ_τ and σ_τ are the mean and the standard deviation of the measured lifetime, respectively. This method, while improving the dynamic range by increasing the statistics collected by a ROI in a given time window, also causes a degradation in spatial resolution by decreasing the number of ROIs in the frame and increasing the effective pixel pitch.

As discussed in Chapter 2, Subsection 2.5.4, the dynamic range of a camera is defined as the ratio between the maximum and minimum number of photons that can be detected by the camera in a single frame. Its quantitative expression is given in Equation (3.11). In [202], $\tilde{n}_{i max}$ and $\tilde{n}_{i min}$ are derived for different SPAD imager architectures based on their SPAD recharging mechanism. In the case of SwissSPAD2, the beginning and the end of the photo-sensitive window are determined by a clock which is independent from the photon arrival time; the sensor is therefore categorized as an imager with a clock-driven recharge. In binary SPAD imagers with active clock-driven recharge, $\tilde{n}_{i max}$ can be expressed as [202]

$$\tilde{n}_{i max} = F_{CS} \times (2^b - 1), \quad (4.5)$$

where b is the bit depth of the frame and F_{CS} is a coefficient which determines the maximum allowed SNR drop due to pile-up. In [202], several guidelines are presented for choosing the value of F_{CS} . $\tilde{n}_{i min}$, on the other hand, is determined by the noise level of the detector. In SPADs, the dominant source of noise is dark noise, which is proportional to the exposure time and independent of the number of frames or readout events. Considering that the dark counts exhibit a Poisson distribution, the standard deviation of the dark count, σ_{DC} ,

is equal to \sqrt{DC} , which is defined as $\sqrt{T_{exp} \times (2^b - 1) \times DCR}$. Plugging the definition of $\tilde{n}_{i\ max}$ and $\tilde{n}_{i\ min}$ in Equation (3.11), the definition of the dynamic range can be rewritten as

$$\begin{aligned} DR &= 20 \log \left(\frac{F_{CS} \times (2^b - 1)}{\sqrt{T_{exp} \times (2^b - 1) \times DCR}} \right) = 20 \log \left(\frac{F_{CS} \times \sqrt{2^b - 1}}{\sqrt{T_{exp} \times DCR}} \right) \\ &= 20 \log \left(\frac{F_{CS}}{\sqrt{T_{exp} \times DCR}} \right) + 20 \log (\sqrt{2^b - 1}). \end{aligned} \quad (4.6)$$

According to Equation (4.6), the dynamic range increases non-linearly with the bit depth, provided that T_{exp} , DCR and F_{CS} are constant. For instance, an increase in the bit depth from 8 to 10 in a frame increases the dynamic range by approximately 6 dB, or a factor of 2. On the other hand, the same bit depth increase from 8 to 10 reduces the frame rate by a factor of 4. This relation highlights the trade-off between the dynamic range and frame rate in SPAD imagers.

When $N \times N$ binning is applied to the pixels, the expression for the dark count can be rewritten as

$$DC = N^2 \times T_{exp} \times (2^b - 1) \times DCR. \quad (4.7)$$

In addition, $\tilde{n}_{i\ max}$ also increases by a factor of N^2 . Therefore, in a binned frame, the dynamic range is given by

$$DR = 20 \log \left(\frac{N^2 \times F_{CS} \times (2^b - 1)}{\sqrt{N^2 \times T_{exp} \times (2^b - 1) \times DCR}} \right) = 20 \log \left(\frac{N \times F_{CS} \times \sqrt{2^b - 1}}{\sqrt{T_{exp} \times DCR}} \right). \quad (4.8)$$

The comparison between Equations (4.6) and (4.8) shows that $N \times N$ binning improves the dynamic range by a factor of N . Obviously, its downside is the decrease in the spatial resolution also by a factor of N , in both directions.

4.2 Phasor-based widefield FLIM

This section presents the results of time-domain phasor-based widefield FLIM experiments conducted using SwissSPAD2. In this work, the performance of our widefield FLIM system was evaluated in terms of the lifetime estimation accuracy and precision under different data acquisition parameters, such as gate width, gate step and number of gate positions. Furthermore, the results obtained in the measurements were compared to the values derived from analytical models and simulations. Finally, the achieved performance was compared to the state-of-the-art FLIM systems in the literature, and the performance limitations of the detector and the setup were discussed.

4.2.1 Methodology

4.2.1.1 Overview of the experimental setup

The SwissSPAD2 camera module used in these experiments allows access to a 472×256 sub-array of the imager. This limitation is introduced by the number of available I/O pins in the FPGA integration module used in

the camera. The pixels are equipped with the round variant of the SPAD with 10.5% native fill factor and no microlenses. The reason for choosing the round SPAD variant was that at that time, the imager with the square SPAD variant was not fully characterized.

The characteristics of a typical gate window used in the experiments are shown in Figure 4:3(a). The gate displays a square shape with sharp edges, whose transient times are annotated in the figure. The gate width, which is 13.1 ns for the particular gate in the figure, is user-controlled, and can vary between 10.8 ns and 22.8 ns with minor changes in the shape or the transition times of the gate window. In addition to the characteristics of the camera, the gate in the figure also reflects the characteristics of the experimental setup. For instance, the transition times are partially determined by the laser pulse width and the jitter of the trigger signal.

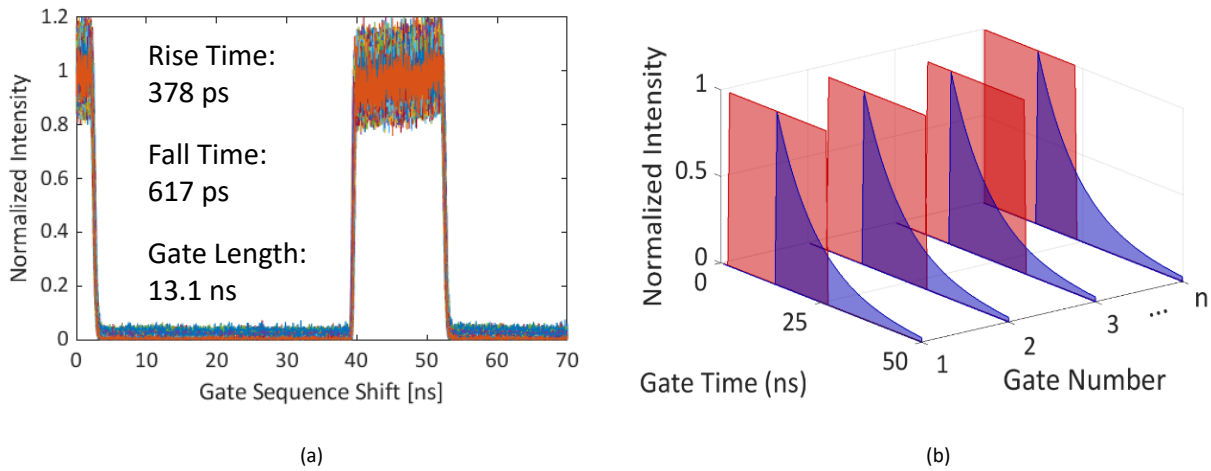


Figure 4:3 (a) Gate window characteristics of SwissSPAD2. The response of every 4th pixel in the 472×256 array is plotted. While the width of the plotted gate is 13.1 ns, the minimum achievable gate width is 10.8 ns. (b) Conceptual illustration of scanning the fluorescent decay with the time gate. The areas in red and blue indicate the gate window and the fluorescence decay, respectively. The scanning is performed with a uniform gate step, across the 50 ns laser period.

The description of the microscopic setup in this experiment, illustrated in Figure 4:4, is as follows. The excitation light is transmitted by the laser output to the microscope in free space. In accordance with the principles of widefield microscopy, the entire field of view is illuminated uniformly and at normal incidence by focusing the excitation light on the back focal plane of the objective lens of the microscope. The excitation light which is reflected by the sample is suppressed at the camera output using both a dichroic mirror and a long-pass emission filter. The emitted fluorescence light from the sample passes through the dichroic mirror and is focused on the image sensor through a tube lens. The pulsed laser used for the excitation of the sample has a wavelength of 532 nm and a PRF of 20 MHz. The laser wavelength is chosen according to the absorption spectrum of the fluorescent dyes used in the experiment, while the PRF is determined by the operation restrictions of the imager.

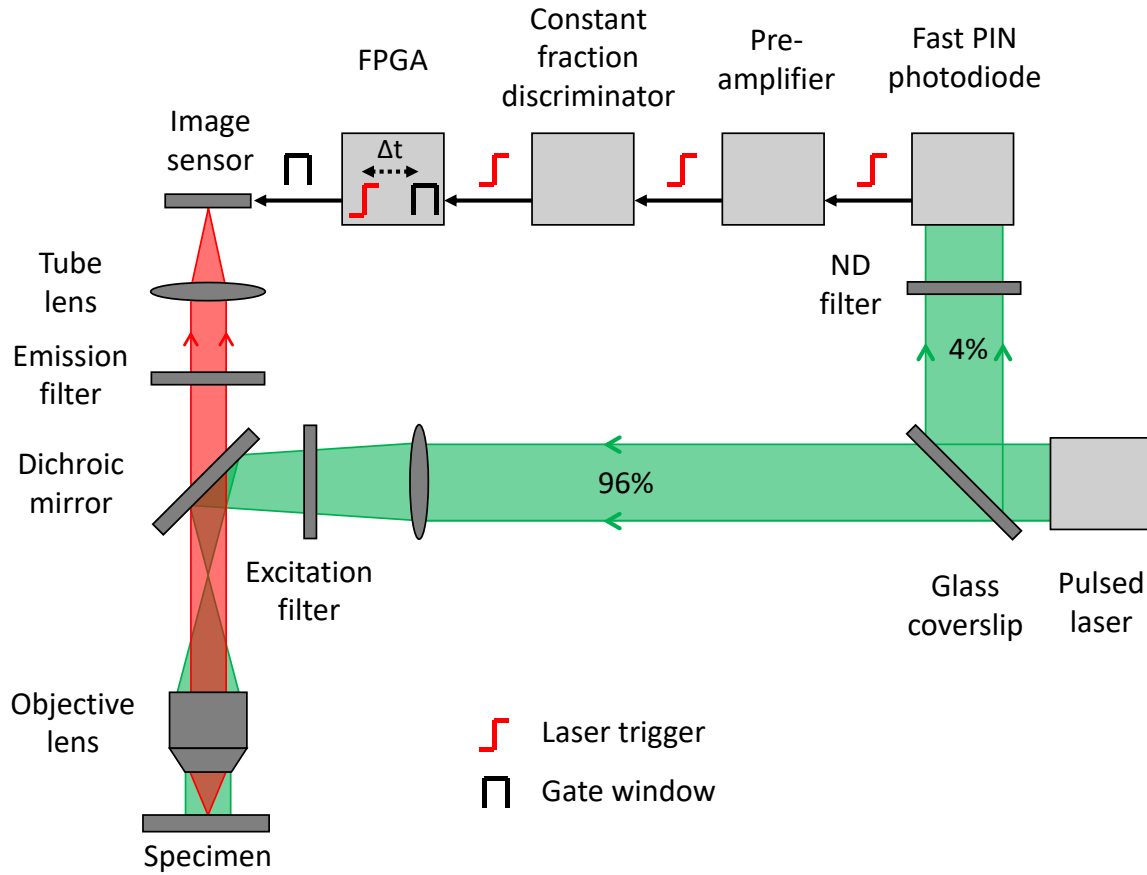


Figure 4:4 Conceptual diagram of the widefield FLIM setup, used in the experiments reported in this chapter.

While the sample is being illuminated, the camera operates in the continuous data acquisition mode. The gate signals of SwissSPAD2, *Recharge* and *Gate*, are generated by the FPGA from the rising edge of the external laser trigger, which is generated from the laser pulse using a fast PIN photodiode. The laser trigger is first detected by a constant fraction discriminator (CFD), which improves its jitter by generating a new trigger based on the precise maximum point of the initial trigger. The CFD output is in turn transmitted to the FPGA. The delay between the laser trigger and the gate signal sequence, Δt , is controlled by the FPGA, with a resolution of $1/56$ ns. The gate scanning process that is used in this experiment is illustrated in Figure 4:3(b). In the figure, the square shaped window in red represents the gate, and the exponential decay in blue represents the fluorescence emission profile of the sample. The FPGA scans the gate across the laser period incrementing the gate delay during the data acquisition after every N frames. Several gate parameters can be set by the user prior to the operation, in order to control the trade-offs. These parameters include the gate width, gate step, number of gate positions, the delay of the first gate position, the bit depth of the gate counts, and the exposure time of a frame.

As photon sensitivity has higher priority than the uniformity of the exposure start and stop times across the array in FLIM, rolling shutter operation ideally would have been the preferred mode of operation, given that in SwissSPAD2, global shutter operation causes a drop in duty cycle. However, due to the unavailability of rolling shutter in gated mode of SwissSPAD2, the global shutter operation is selected. As discussed in Section 4.1, the primary trade-off in the global shutter mode is between the duty cycle and the frame rate. In order to set the duty cycle to approximately 50% according to Equation (4.3) when not factoring in G_W and T_{lsr} , T_{exp}

is set to 10 μs , which is close to $T_{\text{read}} = 10.2 \mu\text{s}$, i.e. the minimum possible readout duration for a frame with 256 rows. Under these settings, the binary frame rate of the camera is equal to 49.4 kfps. The camera continuously streams 8 or 10-bit gate outputs to the PC via USB during the acquisition. Using this data, the phasor analysis and lifetime estimation are performed by the CPU of the PC using a MATLAB-based software, in post-processing.

While a general overview of the camera operation is presented above, several parameters may vary between the experiments, or deviate from these general settings. In these cases, such parameters are mentioned in the sections describing the experiments.

4.2.1.2 Background correction

After the completion of pile-up correction, background correction was performed. Dark counts constitute the majority of the uncorrelated background, while the sources of correlated background include scattering and autofluorescence. For this set of experiments, only uncorrelated background, which exhibits uniform intensity across all gate positions, was corrected. The mean value of the uncorrelated background signal influences the modulation of a decay phasor, but not its phase. This effect can be explained as follows. When the detected emission signal is expressed as the sum of fluorescence and uncorrelated background signals, its overall detected phasor z_{ov} can be expressed as

$$z_{ov} = \frac{I_{ucb} \times z_{ucb} + I_{fl} \times z_{fl}}{I_{ucb} + I_{fl}}, \quad (4.9)$$

where z_{ucb} is the average phasor of the uncorrelated background, I_{ucb} is the total detected photon count of the uncorrelated background, z_{fl} is the average phasor of the fluorescence emission signal, and I_{fl} is the total detected photon count of the fluorescence emission signal. Considering ideal Dirac delta function-shaped IRF or phasors after IRF correction, according to Equations (1.14) and (1.15), for $D = n/f$, $n \in N$, where D is the measurement window, n is an integer and f is the phasor frequency, and uniform gate width and gate step, $g_{avg_{ucb}}$ and $s_{avg_{ucb}}$ are equal to zero in the case of uncorrelated background. Therefore, according to Equation (1.13), z_{ucb} is also equal to zero, in which case Equation (4.9) can be rewritten as

$$z_{ov} = \frac{I_{fl}}{I_{ucb} + I_{fl}} z_{fl}. \quad (4.10)$$

According to Equation (4.10), the overall phasor has therefore the same phase φ , and lower modulus m compared to the “pure” fluorescence phasor. Given that the same rule also applies to the calibration sample, z_{ov} after IRF calibration can have higher or lower m , but still the same φ as z_{fl} , resulting in an unchanged phase lifetime τ_φ (Equation (1.16)).

When the gate positions are non-uniformly distributed, $g_{avg_{ucb}}$ and $s_{avg_{ucb}}$ are no longer equal to zero; therefore, Equation (4.10) is not valid and τ_φ becomes dependent on uncorrelated background.

The dependence of the modulation of the overall phasors on uncorrelated background needs to be corrected in order to improve the visual representation of the phasors and to improve the accuracy of fraction estimation in mixture analysis (see Subsection 4.2.5). However, the correction process does not improve the single-exponential lifetime estimation accuracy and precision, because the shot noise of the background cannot be removed, and the mean value, which can be corrected, does not influence the phase. As the intensity ratio

between the uncorrelated background and the fluorescence signal is dependent on the sample, uncorrelated background correction cannot be performed by IRF deconvolution alone, thereby requiring an additional step for this process.

A common method of uncorrelated background correction in FLIM, which is often used in TCSPC, is to make a reference measurement with a non-fluorescent sample under the same experimental conditions as the sample of interest, and subsequently subtract the two measurement responses [35]. In our experiments, we followed similar principles. However, we also aimed to estimate the background from the IRF calibration sample response, rather than to perform an additional calibration measurement only for background correction. The method that we used, which is a first order approximation neglecting decay wrap around, and whose limitations are discussed at the end of this subsection, was introduced in [90] and is illustrated in Figure 4:5(a). To find the background photon count in a gate window, we averaged the counts of multiple gate positions that contain negligible fluorescence signal, especially in the case of lifetimes that are significantly shorter than the laser period. The selection of multiple gate positions improves the precision of the background estimation by decreasing shot noise, since in a given data set the selection of more gate positions in an analysis increases the photon count included in the analysis. The calculated uncorrelated background count, which is measured using a sample with short lifetime, is subsequently subtracted on a pixel-per-pixel basis from all samples, as the dark count contribution is independent of the sample characteristics.

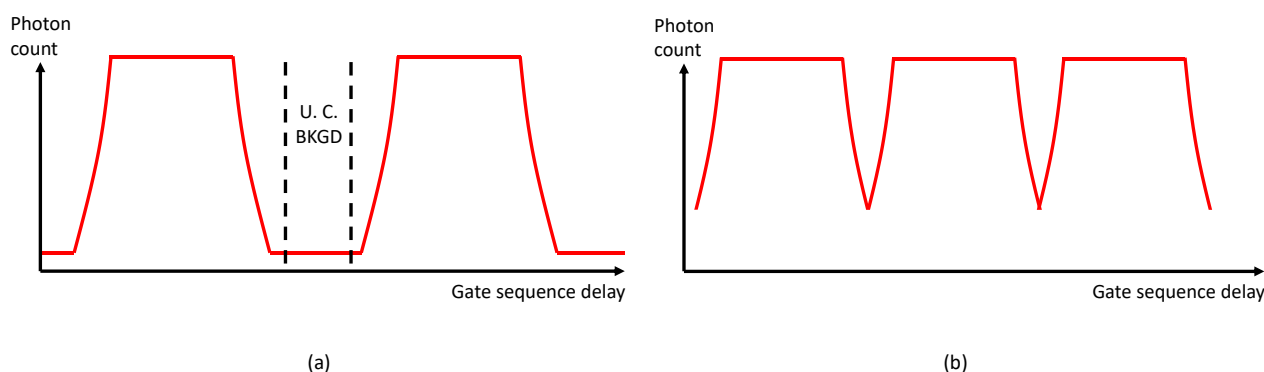


Figure 4:5 (a) Illustration of the uncorrelated background correction method employed in the FLIM experiments reported throughout this chapter. The red curve shows the gated fluorescence response of the reference sample, and the gate positions located between the two vertical dashed lines are selected to calculate the mean of the uncorrelated background signal, on a pixel-per-pixel basis. It is desirable to include as many gate positions as possible in this selection, in order to calculate the mean background signal with minimum shot noise. (b) Illustration of a fluorescence emission response where two adjacent periods have a significant overlap. In the scenarios where the calibration sample displays these characteristics, our background correction method cannot be applied to the calibration sample.

In the absence of noise, single-exponential fluorescence decay phasors are located on a semicircle with a radius of 0.5 and center of (0.5, 0), inside the first quadrant, also called the universal semicircle. As seen in Figure 4:6, the uncorrelated background correction technique corrects the modulation of the phasors, bringing single-exponential decay phasors closer to the universal semicircle. Besides the phasor location shift, the background signal also increases the dispersion of the photon count and phase lifetime due to shot noise, and causes the pixels to approach the pile-up region by increasing the recorded photon count. The background correction method employed in this experiment does not mitigate these two effects.

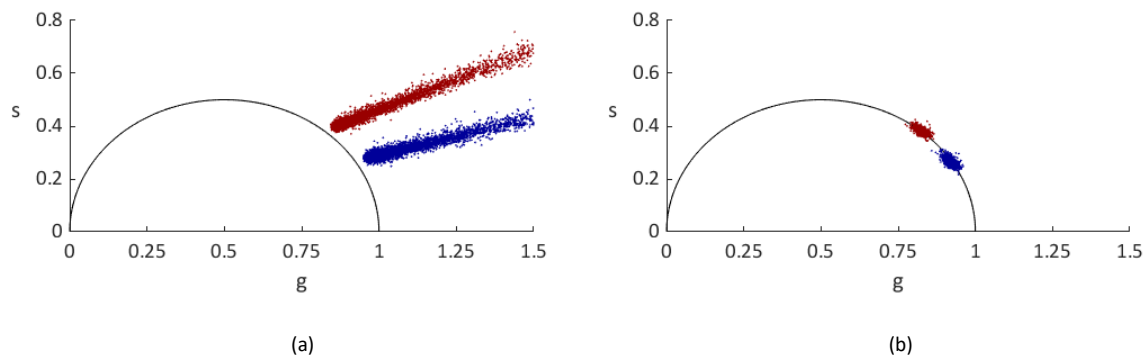


Figure 4:6 Phasors of two single-exponential uniform fluorescent dye solutions (a) before and (b) after uncorrelated background correction. More information on the particular data set and the experiment is available in Subsection 4.2.2 and Figure 4:9.

The use of the aforementioned technique was enabled, in our experiments, by exploiting the otherwise undesired large ratio between the laser period and the sample lifetime, which is caused by the laser PRF range limitations of the detector. This effect leads to significant separation between adjacent fluorescence emission response periods, as shown in Figure 4:5(a). In experimental setups where this ratio becomes smaller in order to achieve higher photon efficiency, all gate positions include considerable fluorescent signal, as shown in Figure 4:5(b). Therefore, this technique cannot be used. In this case, a more suitable background correction method is a 3-point background estimation method [211], which can estimate the background from the photon counts of the gates with fluorescent signal. However, this technique is only valid for single-exponential decays; therefore, its application in mixture analysis is limited. In conclusion, background sample subtraction must be used whenever possible.

4.2.1.3 Noisy pixels consideration

The removal of hot pixels is performed using the following stages. A 2D ‘hot pixel map’ matrix is generated, in which pixels with an intensity above a certain percentile threshold (typically 98-99% for SwissSPAD2) are set to ‘NaN’ and the pixels below this threshold are set to ‘0’. Subsequently, this matrix is added to each gate image of the samples of interest. After the operation, in the entire 8 or 10-bit gate image sequence generated by the detector, the hot pixels counts are converted to ‘NaN’, and the counts of the rest of the array are unaffected. In the following stages, all pixels with ‘NaN’ entries are excluded from the analysis. In the experiments with spatial binning, hot pixel removal is performed before the binning operation. This allows the rest of the pixels in an ROI containing hot pixels to be included in the phasor analysis, thereby preventing a significant loss in spatial resolution. The hot pixel map is generated from the gate image of the calibration sample with the minimum total recorded photon count, which is located in the uncorrelated background window shown in Figure 4:5(a). Since the map is generated from the calibration data of a particular experiment, the accuracy of the hot pixel removal process is not affected by any long-term change in the hot pixel distribution in the sensor.

A popular method to “recover” hot pixels in a camera is to estimate the recorded data of the hot pixel from the neighboring pixels using various interpolation techniques, which makes the resulting image more visually appealing by eliminating the gaps caused by the deleted pixels. This technique was not employed in our experiments for two reasons. Firstly, the use of spatial binning allows all ROIs to display valid lifetime estimation, as the removal of a hot pixels inside a ROI does not prevent the summation of the photon counts of the rest of the pixels of the ROI. Secondly, any pixel recording generated via interpolation could have distorted the quantitative analysis results, such as F-value.

4.2.1.4 Binning

Another optional feature in the analysis is pixel binning. The binning process combines a group of $N \times M$ adjacent pixels in the array. The photon count of the newly formed large pixel, also called the region of interest (ROI) in FLIM, is equal to the sum of the counts of all its contributing pixels. As discussed in Section 4.1, binning improves SNR by increasing the photon statistics captured per ROI at the expense of spatial resolution (the corresponding image is “blurred” accordingly). For shot-noise limited pixels with uniform photon flux, binning improves the SNR of the photon count by a factor of $\sqrt{N \times M}$. In the experiments presented in this chapter, 4×4 or 8×8 binning was used, depending on the sample characteristics of the particular experiment. The binning parameters of each experiment are annotated in their corresponding sections.

4.2.1.5 Frame rate definition

In the experiments presented in this chapter, the bit depth of the gate images is set to 8 or 10, depending on the required trade-off between the acquisition speed and lifetime estimation precision. Due to the memory capacity constraints of the FPGA and the 8-bit word size requirement in the USB data transfer protocol, 10-bit gate images are generated in post-processing by the sum of four consecutive 8-bit images generated by the FPGA. For this reason, the 10-bit images are formed by only 1,020 binary images, instead of 1,023.

As already mentioned in Section 4.1, the detector operates in gated mode only when global shutter is on, due to hardware limitations of the sensor chip. In this mode, the exposure and readout are performed sequentially, and the detector is insensitive during the readout window. In this mode, the data acquisition frame rate of a complete series of G gate images (f_{decay}) is defined as:

$$f_{decay} = \left((T_{read} + T_{exp}) \times y \times G \right)^{-1}, \quad (4.11)$$

where T_{read} is the binary frame readout time, T_{exp} is the binary frame exposure time, y is the number of binary frames corresponding to a gate position, and G is the number of gate positions acquired in a decay.

The minimum value of T_{read} is determined by the readout duration of a row (t_{row}); this relation can be expressed as $T_{read} = N_{row} \times t_{row}$, where N_{row} is the number of rows in the (Equation (4.1)). For the readout of 472×256 frames at the fastest permitted speed, the values of N_{row} and t_{row} are 256 and 40 ns, respectively. Therefore, the minimum value of T_{read} is 10.24 μ s. In the firmware version that is used in the experiments reported in this chapter, the values that T_{exp} can take are determined by:

$$T_{exp} = n \times c_{fw_1} - c_{fw_0}, \quad (4.12)$$

where $c_{fw_0} = 50$ ns and $c_{fw_1} = 400$ ns are firmware constants, and n is an integer which is a user-defined parameter. In this relation, T_{exp} is determined by an on-FPGA counter which increments every 400 ns, and has a 50 ns offset, causing the first increment to occur only after 350 ns. The most typical value of n used in our experiment is 25, which corresponds to $T_{exp} = 9.95$ μ s. This value, combined with $T_{read} = 10.24$ μ s, yields close to 50% duty cycle, based on the duty cycle definition in Equation (4.3), if the influence of gating is excluded.

The relation between y and the bit depth (b) is defined as $y = 2^b - 1$. However, since 10-bit images are constructed in this camera from four individually saved 8-bit images, y is equal to 1,020 in 10-bit mode, instead of 1,023. In the experiments, all data sets were acquired in 10-bit format, in order to have flexibility to choose

either 8-bit or 10-bit format in a single data set. As a result, the actual frame rate of the 8-bit data sets in these experiments is approximately four times slower than the frame rate of the 8-bit acquisition mode. In the experiment reported in Subsection 4.2.3, where the effect of frame rate on lifetime determination precision is analyzed (Figure 4:10), the frame rate is calculated as the time required to acquire the number of gate positions which are used in the analysis, as opposed to the acquisition time of all gate positions in the raw data set. In that sense, it is a “virtual” frame rate, which could be achieved using 8-bit acquisition rather than the 10-bit one used in these experiments.

The firmware version available in this experiment was only able to capture up to 250 10-bit images in a single sequence. In order to capture larger data sets, a pause with a duration of ~ 1.5 s was required between consecutive 250-gate sequences, in which the state machine operating the FPGA was reset. The duration of this pause, which adversely affects the frame rate, is not included in the frame rate formula presented in Equation (4.11). The currently available firmware version, which was developed after the completion of the experiments presented in this chapter, allows the continuous acquisition of data sequences of more than 4,000 10-bit images, which is sufficient to individually capture all data sets used in these experiments in a single, uninterrupted sequence.

4.2.1.6 F-value dependence on gate width

In SwissSPAD2, which is a time-gated imager, the largest contribution to the deviation from $F = 1$ is the width of its gate. A first-order approximate analytical expression for the F-value and σ_τ was derived as a function of the gate width G_W , in order to theoretically estimate the F-value of SwissSPAD2 under different gating settings. In this method, the gate is modeled as a TCSPC bin, in which the gate width G_W is considered analogous to the bin width. This approach assumes that the photon time-of-arrival distribution within a gate is uniform, which becomes inaccurate as the value of the expression G_W/τ , i.e. the ratio between the gate width and fluorescence lifetime, increases. The expressions of σ_τ and F-value as functions of gate width G_W are shown in Equation (2.26) and (2.27), respectively. Their derivations are available in the supporting information of [81].

4.2.1.7 Dye mixture analysis

An important feature of the phasor method is its ability to analyze samples with multiple fluorophore components. This feature is essential for FLIM-FRET analysis, which requires the recovery of the fractions of two lifetime components of a mixture.

The concept of multicomponent phasor analysis is illustrated in Figure 4:7. The two red dots in the figure represent the phasors of two single-exponential fluorescent dyes, and the green dot represents the phasor of a mixture of the two dyes. The g and s coordinates of the mixture are the weighted averages of the phasors of its components. The relative distance of the mixture phasor to the phasors of the components indicates the percentage of the recorded photons for each component in the mixture, also called phasor ratio. In theory, the mixture’s phasor is located on the line drawn between the component phasors, dividing the line into two segments. The relation between the phasor ratio and the lengths of these line segments is expressed as [77, 7]:

$$r_1 = \frac{d_2}{d_1 + d_2}, \quad (4.13)$$

where r_1 is the phasor ratio of component 1. In practice, the mixture phasor may not lie exactly on the line due to the statistical dispersion of the component phasors; in this case, the distances are calculated from the projection of the mixture's phasor on the line.

The phasor ratio of the second component, r_2 , can be calculated from $r_2 = 1 - r_1$. The volume fraction of the components can be derived from the phasor ratio by [90]

$$r_1^{-1} - 1 = (\mu\chi)^{-1}(v_1^{-1} - 1), \quad (4.14)$$

where v_1 is the volume fraction of component 1, μ is the initial concentration ratio and χ is the ratio of the product of extinction coefficient and quantum yield. According to Equation (4.14), the relation between r_1 and v_1 is linear only if the product $\mu\chi = 1$. In our mixture experiment, the value of $\mu\chi$ is constant for all mixtures.

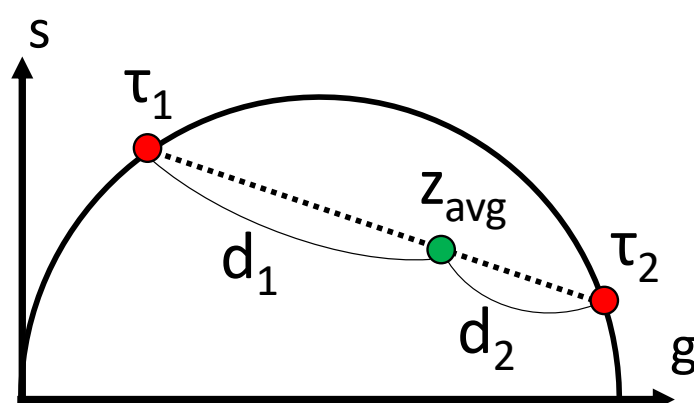


Figure 4:7 Conceptual illustration of multi-exponential decay analysis. The phasor of the mixture (colored green) is located on the line segment between the phasors of its components (colored red). The volume fraction of the components can be derived from the phasor ratio using Equation (4.14), which is calculated from the ratio of the distances of the mixture phasor to its components (Equation (4.13)).

4.2.2 Phasor analysis of fluorescent solutions

The measured fluorescence decay profiles of the four commercially available fluorescent dyes are shown in Figure 4:8. These decay profiles are captured with a gate width of 13.1 ns and gate step of 17.86 ps, using 2,800 gate positions across the 50 ns laser period. The displayed functions represent the measured emission profiles without IRF deconvolution. ATTO 550, Cy3B and Rhodamine 6G (R6G) samples (Figure 4:8 (a-c)) are aqueous solutions sandwiched between two glass coverslips separated by a 1 mm thick rubber gasket. Figure 4:8(d) shows the decay profile of a quantum dot (QD) sample (Qdot585 Streptavidin, ThermoFisher Scientific, $\sim 1 \mu\text{M}$), left to dry out on a glass coverslip. This method causes the formation of random non-uniform density patterns, which render the sample suitable for imaging, contrary to the first three samples with spatially uniform emission intensity. These patterns also display non-uniform average phase lifetimes. The three dye solutions have different concentrations, ranging between 10 nM-1 μM in aqueous buffer. As seen from Figure 4:8(a), ATTO 550 is the dye with the lowest concentration, which leads to the highest shot noise compared to the other samples. This variation between different samples provides an opportunity to evaluate the influence of the photon count on the lifetime determination performance.

In Figure 4:9, the phasors of two single-exponential dye solution samples are plotted on the phasor plot, using the phasor analysis method described in Chapter 1, Section 1.3. In this experiment, three fluorescent dyes with similar excitation and emission spectra (absorption peak around 550 nm, emission peak around 570 nm),

but distinct lifetimes were used: Cy3B, R6G and ATTO 550 (literature values: $\tau = 2.8$ ns , 4.08 ns and 3.6 ns, respectively). The intensity of the fluorescent signal is highest in R6G and lowest in ATTO 550. All dye solutions were excited with a 532 nm 20 MHz pulsed laser characterized by ~ 100 ps pulse width (LDH-P-FA-530XL, PicoQuant, Germany). The bit depth was set to 10 for ATTO 550, the calibration sample, and 8 for the samples of interest. Since ATTO 550 exhibits the lowest fluorescence intensity, longer exposure was selected for this sample to collect comparable statistics to the rest of the samples, thereby ensuring that the calibration stage does not introduce excessive statistical dispersion (Equation (2.28)).

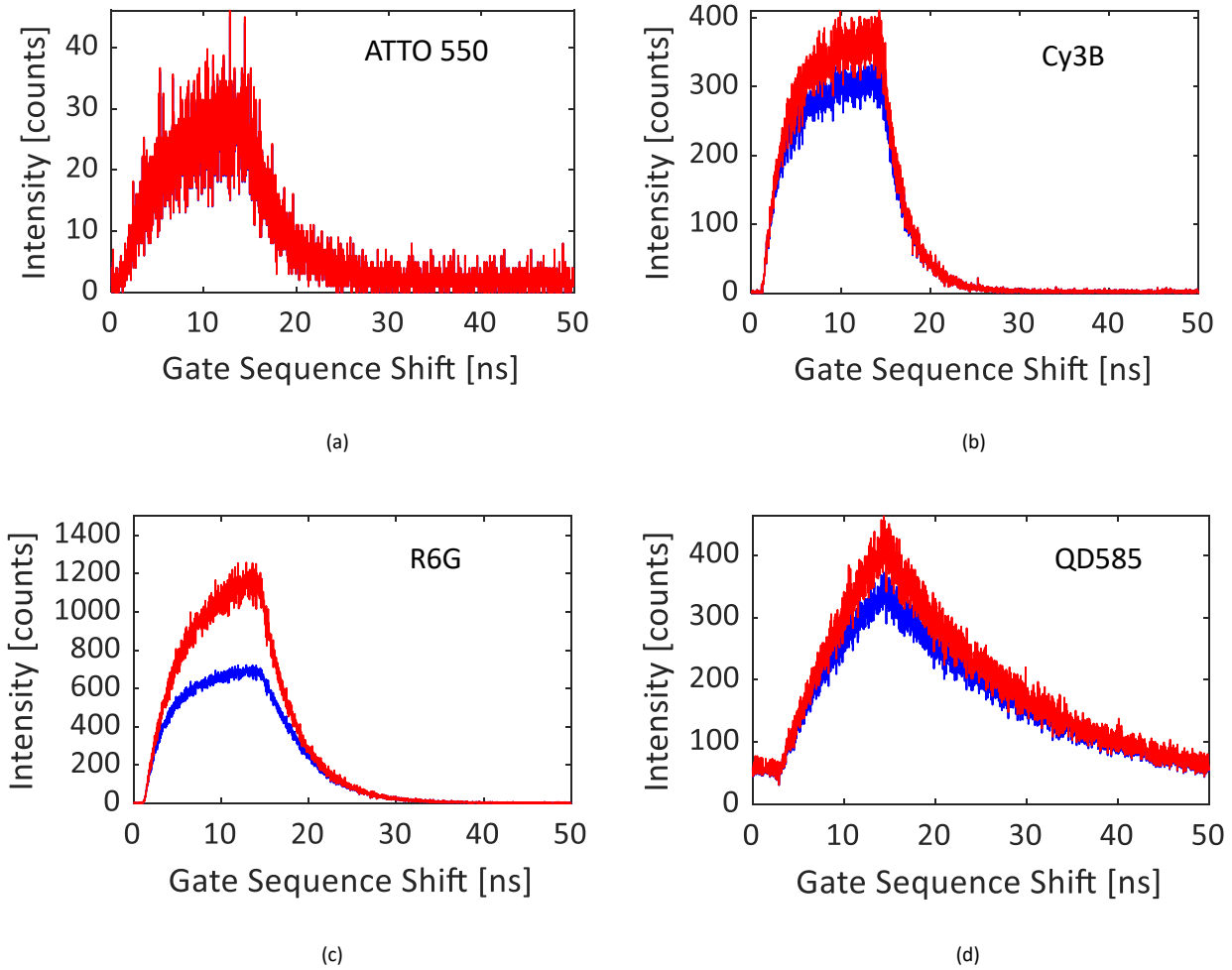


Figure 4:8 Gate response of the detector to four different fluorescent samples with various lifetimes and emission intensities, which are used in the experiments throughout this chapter. The samples are: (a) ATTO 550, (b) Cy3B, (c) Rhodamine 6G (R6G) and (d) Quantum Dot (QD) 585. The blue and red signals indicate the gate photon counts before and after pile-up correction, respectively. The signals represent the convolution of the real emission profile and the IRF. The plots represent the response of a single pixel with the coordinates of (193, 190). Experimental parameters: laser PRF: 20 MHz, gate width: 13.1 ns, bit depth: 10, background correction: off.

The three plots displayed Figure 4:9 were plotted with 16, 140 and 2,800 gates. The phasors of ATTO 550 are represented with a black dot. All ATTO 550 phasors fully overlap with each other on the phasor plot since the IRF deconvolution process was performed separately for each ROI, in order to improve the overall accuracy [86]. It can be observed from the three figures that the dispersion of the phasors increases as G decreases, since the number of recorded photons is proportional to the number of gate positions. The phasor clusters of the two dyes are still visually separable at 16 gates, as seen in Figure 4:9(a), at a virtual acquisition frame rate

of 12.1 fps. This result indicates the ability to resolve two fluorescent dyes with a lifetime difference of 1.4 ns in real-time.

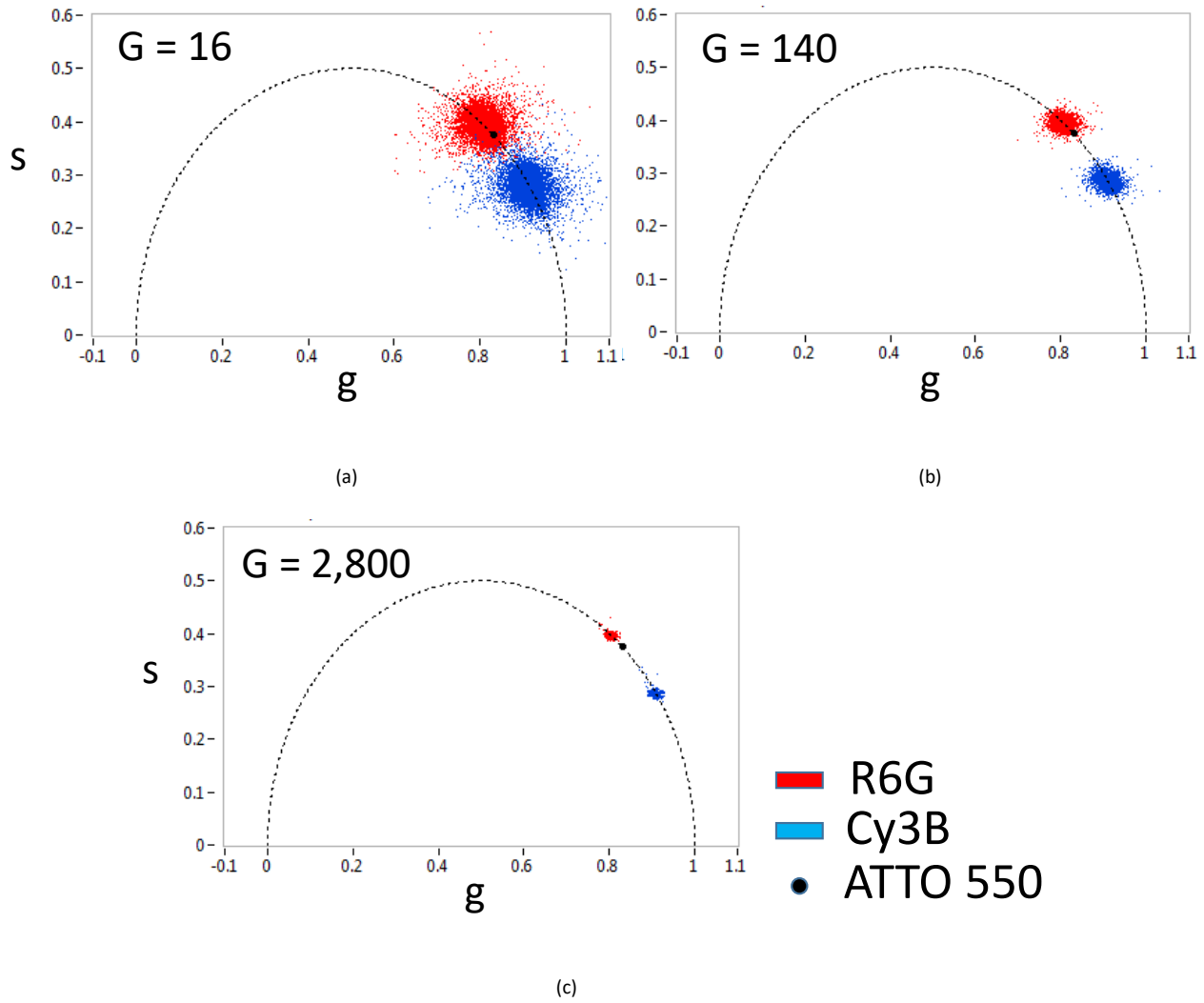


Figure 4:9 Phasors of Rhodamine 6G (R6G) ($\tau = 4.08$ ns) and Cy3B ($\tau = 2.8$ ns) solutions using data sets with (a) 16, (b) 140 and (c) 2,800 gate positions. The ATTO 550 ($\tau = 3.6$ ns) sample was used for calibration. Each phasor point represents a 4×4 ROI. All ATTO 550 phasor locations are identical, since IRF deconvolution was performed separately for each ROI. A decrease in the number of gates leads to a proportional decrease in the recorded number of photons, which increases the dispersion of the phasors of the ROIs due to shot noise. The separability of R6G and Cy3B ($\Delta\tau_{real} = 1.28$ ns, $\Delta\tau_{meas} = 1.4$ ns) is achieved down to 16 gates, corresponding to an acquisition frame rate of 12.1 fps. Experimental parameters: laser and phasor frequency: 20 MHz, gate width: 13.1 ns, array size: 472×256 , binning: 4×4 , bit depth: 8 (R6G & Cy3B), 10 (ATTO 550), pile-up correction: on, background correction: on, percentage of removed pixels: 0% (R6G, Cy3B), 0.5% (ATTO 550).

The phase lifetimes calculated from the phasor values in Figure 4:9 using Equation (1.16) are listed in Table 4:1. The table shows the mean and the standard deviation of the calculated phase lifetimes for 16, 140 and 2,800 gate positions. The measured lifetimes are smaller than the literature values (300 ps or 10% for Cy3B, 200 ps or 5% for R6G), but are consistent with the values obtained by a reference TCSPC measurement (data not shown). Their standard deviation scales with $G^{-1/2}$; this behavior is consistent with shot-noise-limited systems as the recorded photon count is proportional to the number of gate positions.

Table 4:1 Phase lifetime and standard deviation (in ns) obtained from Figure 4:9. The measured phase lifetimes are slightly smaller than the literature values (Cy3B: 2.8 ns, R6G: 4.08 ns) and the standard deviation scales with $G^{-1/2}$.

# Gates (G)	16	140	2,800
Cy3B	2.43 ± 0.23	2.48 ± 0.08	2.49 ± 0.02
R6G	3.89 ± 0.25	3.88 ± 0.08	3.89 ± 0.02

4.2.3 Influence of frame rate on phase lifetime precision

To demonstrate the capability of our method to rapidly acquire sufficient data for good lifetime estimation precision, we investigated the influence of the data acquisition frame rate on the precision. In this experiment, the phasor analysis of single-exponential fluorescent dyes is performed with multiple data sets using a wide range of data acquisition frame rates. The objective of this experiment was to find the maximum frame rate that achieves lifetime estimation with acceptable accuracy and precision. Furthermore, as the number of recorded photons scales with the number of gate positions in our method, this experiment also allows to investigate whether photon economy, which is discussed in Chapter 2, Subsection 2.5.2, remains indeed independent of the total photon count, as it should be in a shot-noise-limited system.

The frame rate, f , is adjusted by setting the number of gate positions, G , given that f is proportional to G^{-1} . An important challenge of this experiment is to generate data sets at multiple frame rates under constant sample photophysical properties, primarily due to photobleaching. To overcome this challenge, we generated all data sets from a single original data set captured with $G = 2,800$ and $G_s = 17.86$ ps. This approach significantly decreases the total exposure time of the sample, thereby reducing the risk of photobleaching. When G was artificially decreased to generate a data subset, the total measurement window ($T_{lsr} = 50$ ns), the gate width ($G_w = 13.1$ ns), and the uniformity of the gate step throughout the gate scanning range were retained. For data availability reasons, as explained in Subsection 4.2.1, the data set was acquired in 10-bit format, and converted to 8-bit by retaining only one out of every four consecutive frames representing a gate position. In each binary frame, which is acquired in global shutter mode, a $10\ \mu\text{s}$ exposure time is followed by a $10.2\ \mu\text{s}$ readout time, during which the detector is blind. The exposure time was tuned in such a way that the duty cycle is around 50%. In this configuration, the 8-bit frame rate of the detector is 194 fps, and the acquisition time of the full original data set with $G = 2,800$ is 14.42 s. The value of G in the subset with the lowest number of gates is 8. The stated values for the frame rate and acquisition time reflect the scenario where the data is acquired independently under detector settings optimized for the corresponding data set. In the experiment, ATTO 550 and Rhodamine 6G (R6G) solutions are used as samples of interest, and the Cy3B solution is used as the calibration sample. 4×4 binning is applied to improve the SNR, thus decreasing the spatial resolution from 472×256 to 118×64 .

The measurement results of the target FLIM performance parameters as a function of the frame rate and number of gate positions are shown in Figure 4:10 [81]. Figure 4:10(a) shows the measured lifetime values of the two samples of interest. The solid data points and error bars represent the mean and the standard deviation of the ROIs for each value of G , respectively. The dashed lines indicate the literature values of the lifetimes. The accuracy is better than 6.6% for ATTO 550 and 2.2% for R6G for the frame rate of 24.3 fps ($G = 8$). Figure 4:10(b) shows the standard deviation of the measured lifetimes of the ROIs in the array, σ_t , for both samples. Considering that the lifetimes of the samples display good spatial uniformity across the field of view, and IRF deconvolution is performed at pixel level, this standard deviation can be considered to be an indicator of the lifetime estimation precision of a single ROI. The measurement of the standard deviation in the temporal

domain, in order to exclude a possible spatial lifetime non-uniformity of the sample, was omitted. The reason for this was the excessive time that would have been required to collect a sufficient number of samples, possibly resulting in additional variation due to photobleaching. As expected, σ_τ is inversely proportional to $G^{-1/2}$, which can be confirmed by the dashed line in the plot, generated by fitting the data points.

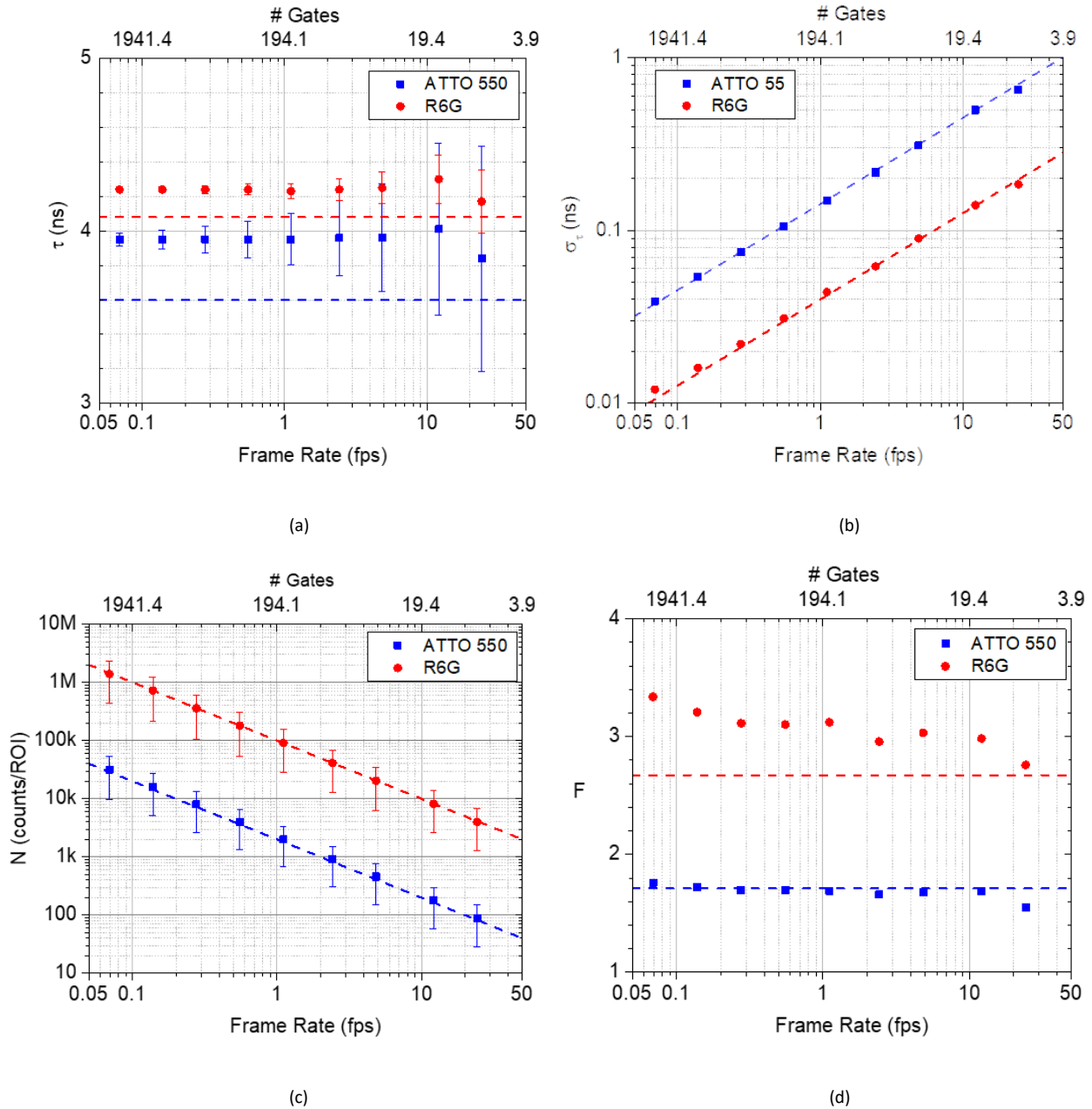


Figure 4:10 Influence of acquisition frame rate on (a) lifetime accuracy, (b) lifetime standard deviation, (c) number of recorded photons, and (d) F-value, based on FLIM experiments on ATTO 550 and R6G solutions ($\# \text{ Gates} = G$). (a) The dashed lines show the expected lifetimes of the dyes obtained from the literature. Error bars represent the standard deviation of the lifetimes of all ROIs. (b-c) The dashed lines are the fitted lines to the data points. They indicate that σ_τ and N are proportional to $G^{-1/2}$ and G , respectively. (d) The dashed lines show Monte Carlo simulation results. A possible cause of the discrepancy between the Monte Carlo and measurement results of R6G is the SNR drop due to pile-up, which is excluded from the Monte Carlo simulation. Experimental parameters: laser and phasor frequency: 20 MHz, gate width: 13.1 ns, array size: 472x256, binning: 4x4, bit depth: 8, pile-up correction: on, background correction: on.

Figure 4:10(c) displays the number of gates and frame rate as a function of the number of recorded photons per ROI. The dashed line shows that G is proportional to N . Combined with Figure 4:10(b), it can be concluded

that σ_τ is proportional to $N^{-1/2}$. An important indicator of precision is the relative standard deviation (σ_τ/τ), rather than σ_τ itself. The reason for this is the general requirement of lower standard deviation for the discrimination of shorter lifetimes. σ_τ/τ was measured as 4.5% for R6G and 18.1% for ATTO 550 for $G = 8$. The contribution of the calibration sample uncertainty to the overall error can be confirmed by the relative standard deviation results. The relative standard deviation of R6G is only 4.02 times lower than ATTO 550, whereas the square root of their photon count ratios is equal to 6.7. This difference proves that the uncertainty in the calibration sample measurement contributes to the overall precision more significantly for R6G than for ATTO 550. This effect can be mitigated by increasing the ratio between the photon count of the calibration sample and each sample of interest.

Figure 4:10(d) shows the F values of the measurements of the two samples as a function of the frame rate. Since σ_τ is proportional to $N^{-1/2}$ in Equation (2.26), the F value is expected to be independent of the number of gates or the frame rate. It must be noted that this equation only applies when the nanotimes of the recorded photons are independent and identically distributed (i.i.d.) random variables, i.e. the time of arrival of each recorded photon is independent of the previously recorded photon. In Figure 4:10(d), a slight increase in the F value can be observed with an increase in the number of gates for both samples. Other than unidentified systematic errors, a possible cause of this behavior is the presence of a noise source that generates a correlation between the time of arrival of two consecutive recorded photons, thereby making the photon detection events non-i.i.d. In SPAD imagers, one phenomenon that can result in this effect is afterpulsing. Despite the relatively high dead time of tens of microseconds which significantly mitigates afterpulsing, it cannot be fully eliminated due to active-clock driven recharging scheme present in the imager [202]. Both the expected and measured F values of the R6G sample are higher than the ATTO 550 sample. This is caused by the higher contribution of the calibration sample to the overall standard deviation in R6G than ATTO 550, due to their respective photon counts. According to Figure 4:10(d), the F values of the system at 24.3 fps is equal to 1.55 for ATTO 550 and 2.76 for R6G. This shows that compared to a shot-noise limited ideal system, where the IRF is equal to a Dirac delta function, $1.55^2 = 2.4$, and $2.76^2 = 7.6$ times higher recorded photon counts are required for ATTO 550, and R6G, respectively, to achieve the same precision. The measured F value is in general agreement with the Monte Carlo simulation results for ATTO 550, the latter being shown with the dashed lines. In the case of R6G, a difference between the measurement and Monte Carlo results can be observed. This difference is suspected to be caused by the SNR drop due to pile-up, which was not included in the Monte Carlo simulation. This hypothesis is supported by the fact that the photon count of R6G is significantly higher than for the rest of the samples, leading to a higher pile-up related SNR drop than ATTO 550.

4.2.4 Influence of gate width on phase lifetime determination

In this experiment, the lifetime analysis was performed with seven different gate widths ranging from 10.8 ns to 22.8 ns, the former being the minimum achievable value for SwissSPAD2. For each gate width, 2,800 gate positions with 17.86 ps delay between consecutive gates sample the entire laser period of 50 ns ($PRF = 20$ MHz). The analysis was performed on two subsets of this data set. The subsets, which consist of 140 and 16 gate positions, were generated by selecting one in every 20 and 175 gate positions, respectively, while retaining the uniformity of the gate step throughout the subset. In these two new data sets, the gate step increases to 357 ps and 3.13 ns, respectively. To achieve higher SNR, 4×4 binning was applied, which reduces the effective spatial resolution from 472×256 to 118×64. ATTO 550 and R6G were used as samples of interest, and Cy3B ($\tau = 2.8$ ns) was used as calibration sample.

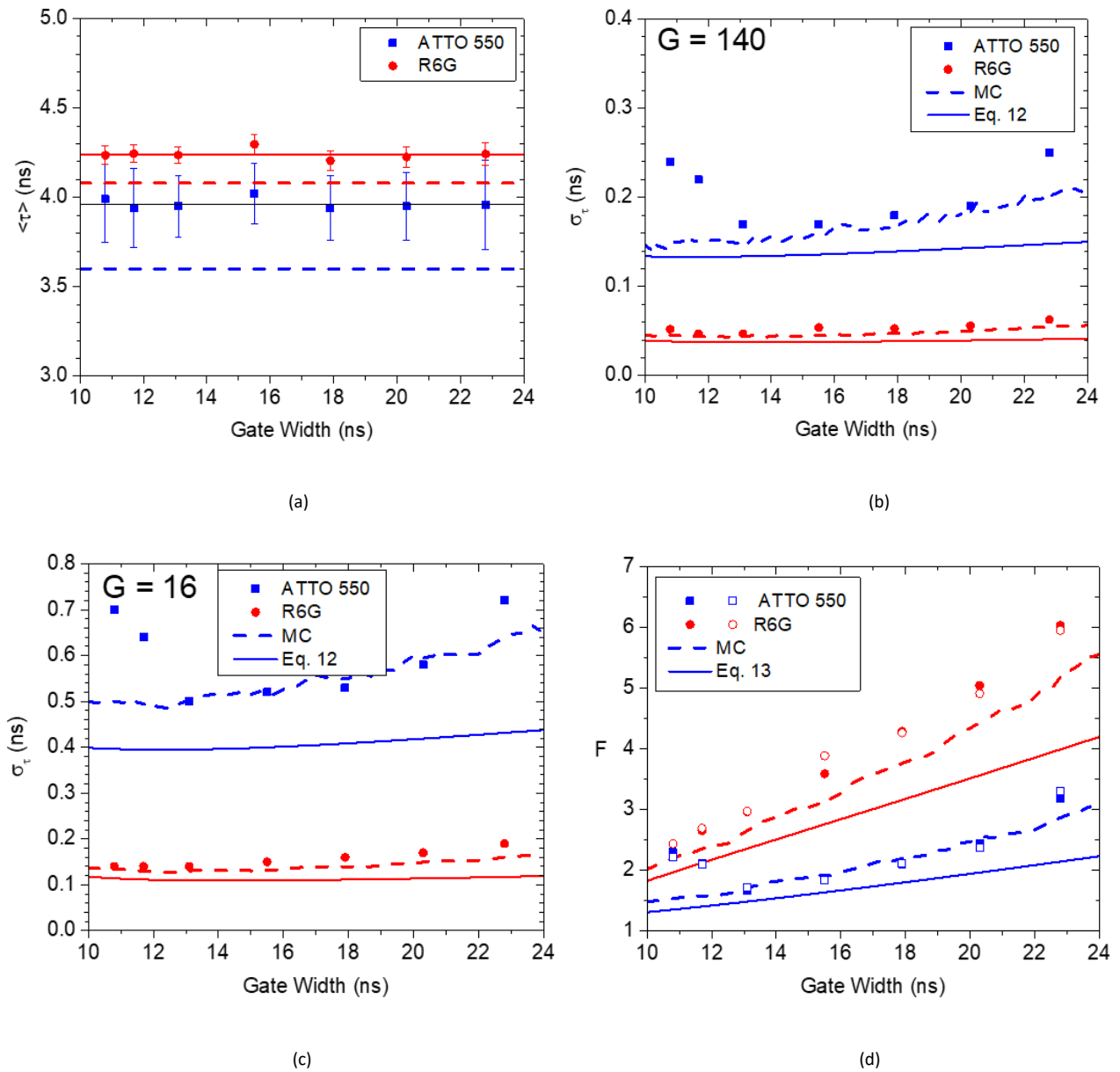


Figure 4:11 Influence of gate width on the phase lifetime determination performance, using ATTO 550 and Rhodamine 6G (R6G) dye solutions. (a) Accuracy as a function of gate width. The dashed lines show the expected lifetimes of the dyes obtained from the literature. Error bars show the standard deviation over all ROIs. (b-c) Standard deviation as a function of gate width, for (b) 140 and (c) 16 gate positions. (d) F-value, an indicator of photon economy, as a function of gate width. In (b-d), the plain and dashed lines represent the analytical shot-noise model and Monte Carlo simulation results, respectively. In (d), filled and open dots represent $G = 16$ and $G = 140$, respectively. These figures show that while the accuracy remains relatively constant at the available gate width range, the photon economy improves as the gate width decreases. Experimental parameters: laser and phasor frequency: 20 MHz, gate width: 13.1 ns, array size: 472×256, binning: 4×4, bit depth: 8, pile-up correction: on, background correction: on.

The results of the analysis are summarized in Figure 4:11. Figure 4:11(a) displays the extracted lifetimes for both samples of interest. The error bars indicate the standard deviation of the results of all ROIs in the array. The average of all data points is marked with solid lines, and the dashed lines show the literature values of the lifetimes of the two dyes. In this figure, it can be observed that the extracted lifetime is independent from the gate width, with a relatively constant accuracy across the plotted gate width range. A discrepancy can however be observed between the literature values and measurement results of both samples (ATTO 550: 3.94 ns vs. 3.6 ns, R6G: 4.24 ns vs. 4.08 ns). Since a deviation from the literature values is also present in the reference

TCSPC measurement (data not shown), the cause of this deviation is considered to be related to the particular samples under analysis rather than the measurement method or the system performance.

Figure 4:11(b) and (c) display the standard deviation of the lifetime estimation of the ROIs for 140 and 16 gate positions, respectively. The data points show the measured values, the solid lines show the theoretically expected values based on the simple shot-noise model described in Equations (2.26) and (2.28), while the dashed lines show the Monte Carlo simulation results. According to the approximate model, the standard deviation is dominated by shot noise, along with the photon time-stamping uncertainty due to the width of the gate, which is considered analogous to a TCSPC bin. The model also assumes that the overall variance of the measurement is the sum of the variances of the measurements of the sample of interest and the calibration sample (Equation (2.28)). The lower overall standard deviation in Figure 4:11(b) compared to Figure 4:11(c) is caused by the fact that larger statistics are captured with more gate positions. Figure 4:11(d) displays the F value of the measurements. The solid line shows the theoretical results based on the shot-noise model given by Equation (2.27), and the dashed line shows the Monte Carlo simulation results. The filled and open data points show the measured results of $G = 16$ and $G = 140$, respectively. The proximity between the filled and open data points indicates that the F value is not highly influenced by the recorded photon count, and the standard deviation is proportional to $N^{-1/2}$ as suggested by the approximate model. In this figure, it can also be observed that the F value improves as the gate width becomes shorter. This trend indicates that from a photon economy point of view, it is preferable to decrease the gate width as much as possible.

In Figure 4:11(b-d), it can be generally observed that the Monte Carlo simulation achieves a closer estimate of the measured results than the approximate model, as expected. Indeed the analytical model does not take into account the details of the phasor analysis process as opposed to the Monte Carlo simulation; instead, it uses a more conventional center of mass method (CMM) to estimate the standard deviation. Furthermore, the difference between the Monte Carlo and measurement results in the R6G sample may be due to the fact that Monte Carlo simulation does not include the pile-up effect, which affects R6G more than ATTO 550, as discussed in Subsection 4.2.3.

In conclusion, in the absence of a more comprehensive error model which takes the entire data acquisition and phasor analysis process into consideration, the Monte Carlo simulation provides an accurate estimation of the precision and photon economy for samples that are not captured under severe pile-up effect.

4.2.5 Dye mixture analysis

After the demonstration of the single-exponential lifetime estimation performance of our system, the next objective is to test its multicomponent analysis capability. This test is required for the assessment of the feasibility of our method for FLIM-FRET analysis, a target application of our research. For this reason, our experiment is designed to emulate a typical FLIM-FRET analysis, whose primary requirement is the extraction of the relative fractions of two components of a multi-exponential fluorescent mixture with known lifetimes [90].

In this experiment, Cy3B ($\tau = 2.8$ ns) and Rhodamine 6G ($\tau = 4.08$ ns) solutions were used as single-exponential dyes, and five mixtures of these dyes with different volume fractions and unknown concentrations were prepared. The volume fractions of Cy3B and R6G in the mixtures were set to 10-90%, 30-70%, 50-50%, 70-30% and 90-10%. Using SwissSPAD2, we measured the phasors of these two single-exponential dye solutions and five mixtures. Cy3B was selected both as a sample of interest and the calibration sample, using two sequentially acquired data sets for each role. This approach was chosen to add the shot noise of both dyes to

the analysis, which would simulate a realistic FLIM-FRET scenario where the calibration sample is independent of the sample of interest. The phasor analysis was performed after the array was cropped to 248×160 by selecting the brightest pixels in the field of view, and 8×8 binning was applied in order to improve SNR.

The phasors of the samples are plotted in Figure 4:12(a). The green and red dots in the plot represent the phasors of the mixtures and their single-exponential components (dyes), respectively. As expected, the phasors of the dyes are located on the universal semicircle, whereas the phasors of the mixtures are located inside the semicircle, between the phasor clusters of both dyes. After the generation of the phasors, the phasor ratio distribution of ROIs of each mixture is calculated using Equation (4.13) [90]. The parameters d_1 and d_2 in the equation are calculated from the projection of a mixture phasor on the line segment drawn between the mean phasor coordinates of the two dyes. In Figure 4:12(b), the mean and the standard deviation of the phasor ratios of ROIs are plotted separately for each mixture, against the known volume fractions. In the figure, the data points and the error bars indicate the mean and the standard deviation, respectively.

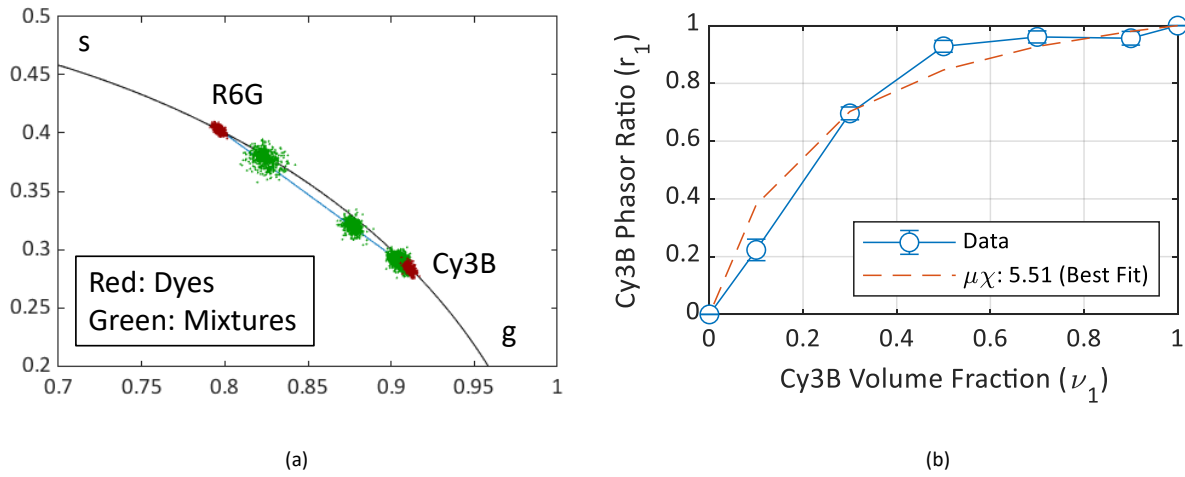


Figure 4:12 Mixture analysis results of Cy3B and Rhodamine 6G (R6G) dyes. (a) Phasor plot of R6G, Cy3B, and five different mixtures of the two dyes with different volume fractions. Two separately acquired Cy3B data sets were used for calibration and actual measurement, in order to include the shot noise introduced by all samples. Calibration was performed based on the Cy3B lifetime of $\tau = 2.5$ ns, which was separately measured in a reference TCSPC setup. (b) Measured Cy3B phasor ratios of the mixtures as a function of their volume fractions. The dashed line represents the relation between the phasor ratio and volume fraction, shown in Equation (4.14) [90], using the value for the parameter $\mu\chi$ obtained by fitting. The data points and error bars represent the mean and standard deviation of the calculated phasor ratios of the ROIs, respectively. Experimental parameters: laser and phasor frequency: 20 MHz, number of gate positions: 234, gate width: 22.8 ns, array size: 248×160, binning: 8×8, bit depth: 10, pile-up correction: 10, background correction: on, percentage of removed pixels: 1%.

The precision of this measurement is calculated by fitting Equation (4.14) to the data points on the plot, and evaluating the quality of the fit for the most suitable set of intrinsic parameters. The result of the fitting yields $\mu\chi = 5.51$ and $\sigma_{\mu\chi} = 1.77$. The fitted curve is shown with the red dashed line in Figure 4:12(b). Considering that $\chi = 0.51$ is the ratio of the product $\Phi\epsilon$ (Φ : quantum yield, ϵ : extinction coefficient) for both dyes [90], the initial concentration ratio between Cy3B and R6G can be calculated as $\mu = 5.51/0.51 = 10.8$. This significant difference between initial concentrations is the reason for the fact that the mixture phasors in Figure 4:12(a) are highly skewed towards the Cy3B phasor cluster, and almost indistinguishable for the mixtures with 50% to 90% Cy3B volume fractions.

Considering the achieved fitting accuracy of the analytical relation between the volume fraction and phasor ratio on the measured data, it can be concluded that the capability of our system to perform mixture analysis is verified.

4.2.6 Phase lifetime map for complex samples

Aside from its speed advantage, an important feature of the phasor method is its graphical representation of fluorescence decay profiles. This feature allows the user to make rapid visual assessments on the lifetime distribution of the sample in the spatial or temporal domain. Since the phasor does not depend on the signal intensity, the phasor representation *de facto* filters out information such as tissue thickness (if the tissue does not result in scattering) and fluorophore concentration, which are in general secondary and uninteresting properties of a sample. Finally, since each phasor belongs to a particular pixel in the camera, this representation enables the display of the phasors of a selected area in the field of view, or inversely, the examination of the visual features corresponding to particular phasor cluster in the plot.

To demonstrate these practical advantages of phasors, a sample with higher complexity than a single-exponential dye solution is necessary. The required complexity can involve the presence of a wide range of lifetimes and heterogeneous spatial features across the sample. Quantum dots are selected as the sample for this experiment since they meet this complexity requirement. Furthermore, the lifetimes of quantum dots are relatively longer than the dye solutions used in this experiment, which introduces additional challenges in data processing due to a significant overlap between the emissions corresponding to two consecutive laser periods. In addition to the inherent characteristics of the quantum dots, the preparation method of the sample, described in Subsection 4.2.1, also contributes to the generation of non-uniform patterns in the field of view.

Figure 4:13(a) and (b) show the intensity and lifetime images of the sample, respectively. The sample consists of multiple shapes with distinct fluorescent intensity and lifetimes. Due to the optical constraints of the microscope, the image of the sample does not occupy the entire field of view, but is confined to a circular area in the center. In the intensity image, the peak magnitude is concentrated at the bottom center of the field of view. This is attributed to the non-uniform excitation light intensity distribution in the optical setup. Owing to the insensitivity of the lifetime to signal intensity, this variation does not appear in the lifetime distribution in Figure 4:13(b), as expected. Although the lifetime is uncorrelated to the intensity, it is observed to be correlated to the shape of the structures in the sample.

The two complementary output formats of the phasor method are presented in Figure 4:13(c-d). Figure 4:13(c) shows the image of the sample where the lifetime and intensity distributions are both represented. The pixels are color-coded based on their lifetimes, and their brightness is set according to the emission intensity. Contrary to the standard lifetime image in Figure 4:13(b), this image format suppresses the areas with low signal intensity, which exhibit low SNR due to high shot noise. The second output format is the phasor plot (Figure 4:13(d)). In addition to the lifetime distribution of the sample, this plot also provides information on the exponential decay characteristics of each ROI. The color coding in the phasor plot represents the phasor density: the areas shown in red and blue indicate the highest and lowest number of ROIs per unit area, respectively. Two conclusions can be inferred from the information on the phasor plot. Firstly, the sample generally exhibits single-exponential characteristics, as most of the phasors are located around the universal semicircle. Secondly, there is a single peak in the lifetime distribution, and the pixel density consistently decreases in both directions away from the peak value. Two reference points for high and low bounds of the lifetime distribution are selected from the phasor plot as $\tau_H = 16.7$ ns and $\tau_L = 13.9$ ns, which are marked in Figure 4:13(d) with green and red dots, respectively. These two reference values are then used for color-coding the FLIM images in Figure 4:13(b) and Figure 4:13(c). From the color bar in Figure 4:13(b), a color is assigned to each pixel based on the phasor ratio of its phasor on the map. The phasor ratio of a pixel is determined by the relative distance of its phasor to the phasors of τ_H and τ_L , as formulated in Equation (4.13) [7].

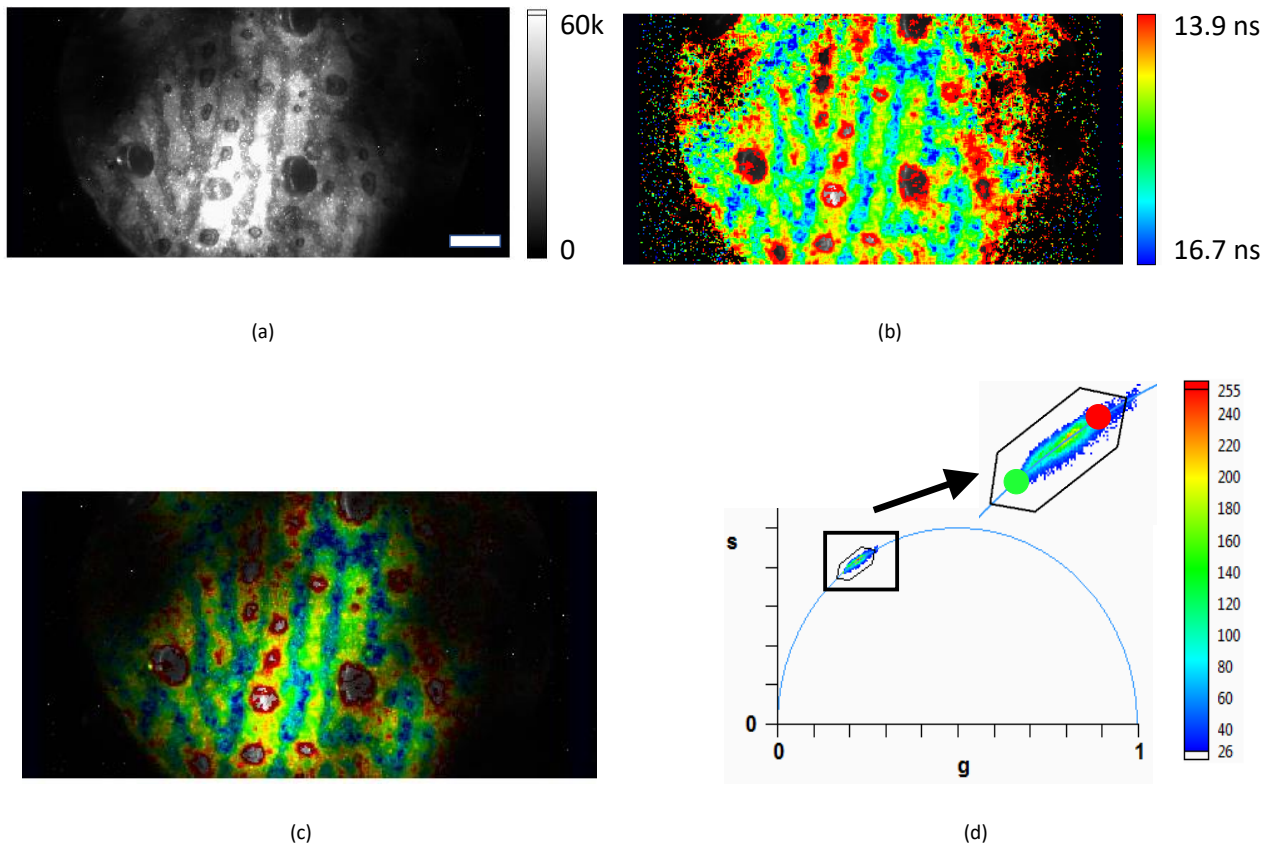


Figure 4:13 Lifetime analysis of a dried quantum dot sample. (a) Fluorescence intensity image. Scale bar: 25 μm . (b) Phase lifetime image. (c) An image that combines both intensity and phase lifetime information. Colors represent the lifetime according to the color spectrum in (b), and the brightness represents the intensity. (d) Phasor distribution of the sample. In the color-coding method in (b-c), two values representing high and low lifetimes were selected from (d) as $\tau_H = 16.7$ ns and $\tau_L = 13.9$ ns which are marked on the phasor plot with green and red dots, respectively. The mapping of the pixels to the color spectrum in (b) was done based on their phasor ratios with respect to the two reference points. Colors represent the phasor density of the areas in the phasor plot. It can be observed that the lifetime distribution is not dependent on fluorescence intensity, as expected.

The results obtained in this experiment demonstrate the capability of the system to perform FLIM on a complex sample with spatially non-uniform intensity and lifetime characteristics. In addition, they also confirm that the lifetimes extracted in this method are independent of the emission signal intensity.

4.2.7 State-of-the-art comparison

Aside from the determination of various parameters such as accuracy, precision and F-value, another crucial component of FLIM performance assessment is the comparison with state-of-the-art FLIM systems. In this subsection, a performance comparison of SwissSPAD2 with selected academic and commercial FLIM systems of various architectures is presented. In Table 4:2, the specifications of SwissSPAD2 that are relevant to FLIM are compared to 9 other systems. These systems can be classified into three main categories: widefield gated imagers, scanning-based TCSPC systems, and widefield TCSPC systems.

Outside of gated SPAD imagers such as SwissSPAD2, widefield gated systems are mainly based on CCD or CMOS image sensors. In this category, the gate can be implemented in two ways. The first way is by means of an image intensifier, e.g. [133]. Despite its capability of reaching single-photon sensitivity via electron multiplication in the image intensifier, this comes with a relatively low dynamic range due to the saturation of the electron well with the multiplied electrons. According to [236], the RMS noise requirement for an imager to be qualified as single-photon sensitive is determined to be 0.15 e^- . Based on this criterion, the multiplication

gain required for Reference [133] for single-photon sensitivity can be calculated as $Gain = 7/0.15 \cong 47$. At this gain, the maximum local photon count rate, defined as the count rate of a pixel, drops to $25,000 \times 10/47 = 5.3$ kcps, which is significantly lower than the other gated cameras in the table. A second disadvantage of ICCD is its single gate channel, introducing duty cycle penalty. The number of gate channels in an ICCD is fundamentally limited by its architecture. Furthermore, its frame rate of 10 fps is not sufficient for video-rate FLIM without binning, particularly if more than two gate positions are required due to the complexity of the sample's fluorescent decay behavior. On the other hand, their minimum gate width can reach down to 200 ps, which is significantly lower than large-format SPAD cameras. In the second architecture, based on lock-in pixels, inherently non-overlapping gates are generated by implementing multiple electron wells in a pixel, and directing the photoelectrons to one of these wells using external control voltages. To this date, the highest number of gates that are employed in a lock-in pixel is four [9]. With their high electron well capacity, these cameras can reach extremely high photon count rates. However, this feature is not useful in settings with low photon budgets. Under these conditions, the number of recorded photons is limited by the sensitivity and duty cycle of the detector rather than its photon count rate. In addition, due to the lack of an electron multiplication mechanism, these cameras do not have the ability to overcome their high readout noise in photon-starved environments, unlike the ICCD. At the time of writing, the only lock-in pixel camera that reaches sub-electron readout noise [9], has significantly lower spatial resolution than SwissSPAD2. (~8 times fewer pixels than the pixel format of the camera module, and 16 times fewer pixels than the pixel format of the sensor chip). In addition, its readout noise is still not sufficiently low to meet the single-photon sensitivity requirement mentioned above [236].

The second category is scanning-based TCSPC. The most important advantage of TCSPC compared to gating methods is temporal resolution. The bin width and the temporal resolution of the two high-grade commercial TCSPC systems in Table 4:2 [237, 238] are significantly superior to that of time-gated systems. The global count rate, defined as the sum of the photon count rates of all pixels in the resulting image that are acquired in parallel, which is the main downside of scanning approach compared to widefield, can be improved in two ways. In architectures where the dead time of the detector is higher than the dead time of the TDC, multiple detectors can be connected to a single TDC to decrease the equivalent detector dead time of an input channel. Secondly, multiple input channels can be connected in parallel to a single histogramming block. Using multi-channel TCSPC hardware, single-point FLIM is conventionally performed by optically splitting the fluorescent signal into multiple channels and directing them to the separate detectors [237, 238]. There are two disadvantages of this method. Firstly, splitting the beam into multiple channels increases the system complexity as the number of channels increases. Secondly, since the dwell time on a single point remains constant, the achievement of higher photon count rates requires a proportional increase in the excitation power, which increases the vulnerability of the sample to photobleaching [45, 7]. Multi-channel TCSPC without increasing photobleaching can be achieved via multifocal FLIM, as demonstrated in [60]. This approach has been discussed and compared to conventional FLIM approaches in Chapter 1. The main challenge of this approach is to generate multiple beamlet patterns with relatively uniform characteristics, which requires a more complex optical setup than a standard confocal microscope.

The third category is widefield TCSPC. In the table, there are three systems that employ widefield TCSPC. The first are position-sensitive anode detectors, such as the H33D detector [239]. The H33D detector can be described as a single-channel TCSPC system with 2D positioning for each recorded photon. Therefore, its primary disadvantage is the limitation of the global photon count rate by the dead time of its TDC, which results in a saturated global count rate of only 700 kcps (or ~2 Mcps in its latest generation [240]). The second and third

systems, MEGAFRAME 128 [10] and QuantiCam [57], employ in-pixel TDCs. In addition to performing time-stamping with its TDCs, QuantiCam also uses a gate to suppress the signal in the unwanted time windows. While MEGAFRAME 128 has achieved an important milestone for SPAD technology by demonstrating the feasibility of in-pixel TDCs in mid-format SPAD imagers, QuantiCam offers better overall performance with its significantly larger fill factor and considerably larger time-stamping range, also thanks to the use of a more advanced technology node (40 nm vs. 130 nm). Its TDCs reach 33 ps timing resolution, which is significantly shorter than any gated imager in any technology. However, since every recorded photon contains a 14-bit time-stamp, the bandwidth of its electronics limits its dead time to 53.76 μ s, which results in a relatively low global photon count rate of 50.7 Mcps.

Among all listed TCSPC systems, the highest global photon count rate reached is 332 Mcps [237]. Considering that SwissSPAD2 can reach a useful count rate of 2.53 Gcps with an F value of ~ 1.7 (Figure 4:10(d)), its precision for the same spatial resolution and target lifetime is equivalent to an ideal shot-noise-limited system ($F = 1$) with a global count rate of $2.53 \text{ Gcps} / 1.7^2 = 875 \text{ Mcps}$. According to this calculation, the lifetime precision of SwissSPAD2 under equal spatial resolution surpasses all TCSPC systems in the table. It is noteworthy to mention that with a calibration stage with higher photon counts, the F value of SwissSPAD2 is expected to decrease even further.

The comparison of SwissSPAD2 and other FLIM systems can be summarized as follows. SwissSPAD2 outperforms most gated lock-in pixels in the table in terms of dynamic range. The 4-tapped lock-in image sensor in [9], while exceeding SwissSPAD2 in dynamic range, exhibits 8 times lower spatial resolution and cannot reach single-photon sensitivity. The ICCD camera in [133] can reach single-photon sensitivity. However, the multiplication gain required for that lowers its photon count rate significantly, to a value below SwissSPAD2. Scanning-based TCSPC systems, while achieving excellent timing resolution, cannot reach SwissSPAD2 in global photon count rate, even after adjusting for the F value for the range of lifetimes that are used in the experiments in this chapter. Similarly, widefield FLIM systems also have lower global photon count rate. On the other hand, SwissSPAD2 exhibits several relative shortcomings in terms of FLIM performance: modulation frequency limitation, long gate width, single gate channel and lack of rolling shutter operation in gated mode. These shortcomings are overcome in SwissSPAD3.

Table 4:2 Performance comparison of state-of-the-art FLIM systems with SwissSPAD2.

	SwissSPAD2 [117, 81]	LaVision PicoStar HR [133]	LIFA FLIM system [109]	pco.flim [113]	[9]	[241, 153]	MEGAFRAME 128 [10]	QuantiCam [57]	H33D Gen I [239]	PicoQuant rapidFLIM (MultiHarp 150 16P) [237, 58]	B&H FASTAC FLIM (SPC-154N) [177]
Detector type	SPAD	Gated ICCD	CCD Lock-in	CMOS Lock-in	CMOS Lock-in	CMOS Lock-in	SPAD	SPAD	Position sensitive anode detector	Hybrid photomultiplier tube detector (PMA Hybrid 06)	Hybrid photomultiplier tube detector (HPM-100-06)
Array format	512×512 (472×256 used)	1,376×1,040	504×512	1,008×1,008	128×128	512×256	160×128	192×128	~Ø400	1	1
Pixel size/spatial resolution	16.38 µm	6.45 µm	24 µm	5.6 µm	22.4 µm	11.2 µm (H) 5.6 µm (V)	50 µm	18.4 µm (H) 9.2 µm (V)	100 µm	-	-
Fill factor (native)	10.5%	~100%	50%	18%	3.5%	16.7%	1%	13%	~100%	~100%	~100%
Fill factor (with µl)	28-47%	-	-	-	26.3%	-	-	42%	-	-	-
Maximum PDP/QE/Sensitivity	~50% @ 520 nm (V _{ex} = 6.5 V)	11% @550 nm (S25)	-	39% @510 nm	-	137 ke-/lux.sec	27.5% (V _{ex} = 1.4 V)	~34% @ 560 nm (V _{ex} = 1 V)	18% @400 nm (S20)	28% @400 nm	28% @400 nm
Frame rate	97.7 kfps (1 bit)	10 fps	30 fps (3)	45 fps (3)	45 fps (3)	12 fps (3)	50 kfps (10-bit TCSPC mode))	18.6 kfps (14-bit TCSPC mode)	-	-	-
Full well capacity	-	25,000 e ⁻	67,000 e ⁻	52,000 e ⁻	5,500 e ⁻	2,700 e ⁻	-	-	-	-	-
rms readout noise	0	7-8 e ⁻	25 e ⁻	48 e ⁻	0.85 e ⁻	1.75 e ⁻	0	0	0	-	-
Dark count rate (median)	7.5 cps/px @27 °C (V _{ex} = 6.5 V)	< 0.1 e ⁻ /s/px	400 e ⁻ /s/px @20 °C	1,220 e ⁻ /s/px	-	-	50 cps/px (V _{ex} = 0.73 V)	25 cps/px (V _{ex} = 1.5 V)	20 kcps (global)	<100 cps (global)	15 cps (global)
Dynamic range	~5,179:1 (74.3 dB) @Bit depth: 8, binary frame rate: 97.7 kfps (1)	> 2,000:1 (66 dB) (5)	2,000:1 (66 dB)	>1,000:1 (60 dB)	6,471:1 (76.2 dB)	1,543:1 (63.8 dB)	18,179:1 (85.2 dB) @ Bit depth: 8 Binary frame rate: 50 kfps	~15,680:1 (83.9 dB) @Bit depth: 8, binary frame rate: 18.6 kfps (1)	-	-	-
Single-photon sensitivity (photon counting capability) (7)	Yes	No (8)	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Number of gate/time channels	1	1	2	2	4	2	-	1	1	up to 16 (1 detector per channel)	4
Minimum gate width	10.8 ns	200 ps	-	-	0.8 ns	-	-	-	-	-	-
Minimum gate step	17.9 ps	5 ps	-	-	-	-	-	-	-	-	-
Minimum bin width	-	-	-	-	-	-	55 ps	33 ps	-	10 ps	405 fs
Temporal resolution	-	-	-	-	-	-	140 ps (FWHM)	208 ps	100 ps (FWHM)	TDC: <45 ps (RMS) Detector: <50 ps (FWHM)	TCSPC: <7 ps (FWHM) Detector: <20 ps (FWHM)
Dead time	10.2 µs	-	-	-	-	-	20 µs	53.76 µs	1.4 µs (TDC)	650 ps (TDC) (Bursts up to 2,048 events per channel at this rate are allowed)	100 ns (TAC)
Maximum saturation local (pixel) photon count rate	97.7 kcps	20 kcps (5)	4.02 Mcps	4.68 Mcps	990 kcps	64.8 kcps	50 kcps	18.6 kcps	700 kcps	(6)	40 Mcps
Maximum saturation global photon count rate	12.8 Gcps	28.6 Gcps (5)	1.04 Tcps	4.76 Tcps	16.2 Gcps	8.49 Gcps	1.03 Gcps	457 Mcps	700 kcps	(6)	40 Mcps

Application results

Maximum useful local (pixel) photon count rate	20.9 kcps (2)	20 kcps (5)	4.02 Mcps	4.68 Mcps	990 kcps	64.8 kcps	5.55 kcps (2) (10-bit TCSPC mode)	2.06 kcps (2) (14-bit TCSPC mode)	350 kcps (4)	332 Mcps (sustained) (6)	20 Mcps (4)
Maximum useful global photon count rate	2.53 Gcps (2)	28.6 Gcps (5)	1.04 Tcps	4.76 Tcps	16.2 Gcps	8.49 Gcps	113.7 Mcps (2) (10-bit TCSPC mode)	50.7 Mcps (2) (14-bit TCSPC mode)	350 kcps (4)	332 Mcps (sustained) (6)	20 Mcps (4)
FLIM method	Widefield/Gated	Widefield/Gated	Widefield/Gated	Widefield/Gated	Widefield/Gated	Widefield/Gated	Widefield/TCSPC	Widefield/TCSPC	Widefield/TCSPC	Scanning/TCSPC	Scanning/TCSPC

(1) The values are obtained from Equation (4.6) [202]. F_{CS} factor was selected as 2.842 for clock-driven recharge, and 36 for event-driven recharge, for the operation mode described in [202] as “SNR decrease”, at which the SNR decreases by 3 dB due to pile-up.

(2) Maximum photon count rate is calculated using the **10% nonlinearity criterion** for active clock-driven architectures in [202] ($F_{CS} = 0.214$ for clock-driven recharge, and $F_{CS} = 0.111$ for event-driven recharge).

(3) A frame consists of all taps in the detector.

(4) 50% of the saturation rate.

(5) These values are lower if the camera is configured for single-photon sensitivity.

(6) Whether this value represents saturation or useful photon count rate is not available.

(7) The maximum readout noise requirement for photon counting capability was determined by Fossum [236] as $0.15 e^-$ RMS, which applies to detectors without electron multiplication.

(8) While it is possible to achieve photon counting capability with ICCD detectors, it comes at the expense of significantly reduced dynamic range. Therefore, under typical operation settings, single-photon sensitivity is not achieved.

4.2.8 Discussion

The research presented in this section aims to investigate the feasibility of SwissSPAD2 for widefield, phasor-based FLIM at video-rate acquisition speed. This work differentiates itself from previous demonstrations of this ability in the literature by the properties of the detector used in the experimental setup. As shown in Table 4:2, large-format SPAD imagers enable a unique combination of high global count rate and single-photon sensitivity. Due to the relatively lower count rates of TCSPC-based architectures among these imagers [57, 10], SwissSPAD2 was designed with a time gating architecture. However, the time gates implemented in large-format SPAD imagers exhibit long gate widths, which are often required to overlap when scanning fluorescence decays in the order of a few nanoseconds. Therefore, the demonstration of picosecond-level lifetime precision at a reasonable photon economy with long gates was the primary objective of this work.

In the experiments presented in this section, it was demonstrated that dye pairs with a lifetime difference of 1.4 ns or less can be separated at an acquisition speed approaching video rate (12.1 fps), at 13.1 ns gate width. This performance was achieved despite the significant loss in photon sensitivity due to the lack of microlenses, and duty cycle limitations caused by global shutter operation requirement and a single gate channel. After the completion of the experiments, microlenses were deposited on SwissSPAD2, improving PDE by a factor of 2.65 under collimated light with normal incidence. When the angle of incidence is optimized for a specific camera sample, this factor can reach up to 4.2. In addition, SwissSPAD3 is expected to achieve close to 100% duty cycle with its two complementary gate channels and rolling shutter operation, thus further improving photon sensitivity. In SwissSPAD3, this improvement comes at the cost of approximately 50% decrease in the maximum photon count rate. However, this can be compensated in the case of short lifetimes by the ability to reach 80 MHz PRF, as opposed to the 20 MHz PRF of SwissSPAD2. The decrease of the ratio between the laser period and target lifetime avoids acquiring gate images with negligible signal, thereby increasing the average photon count rate of the measurements. As a result, a given recorded photon count can be achieved with fewer gates, which leads to an increased frame rate for decay measurements.

The comparison of the lifetime precision measured for different recorded photon counts shows that the precision mostly follows an approximate shot-noise model, in which the main contributors are the number of recorded photons and the width of the gate, the latter defining the timing resolution of a single photon time-stamp. While the F value of a lifetime measurement is larger than 1 as a result of non-zero temporal resolution (i.e. IRF that deviates from a Dirac delta function), it has low dependence on the number of photons, indicating independent and identically distributed photon time-of-arrival distribution. It was also observed that the two samples of interest with different fluorescence intensity levels display different F values under the same data acquisition settings, which is contrary to the independence of the F value on the photon count. This F value discrepancy is caused by the contribution of the calibration sample to the precision, as expressed in Equation (2.29). To reach the maximum photon economy that is offered by the detector for a given gate width in future experiments, the calibration must be performed with significantly higher photon count and longer integration times than the measurement of the sample of interest. To prevent a decrease in accuracy under the increased excitation duration, the calibration sample must be selected as a sample with low photobleaching rate. As the calibration process must only be done in a set of experiments, the contribution of the increased calibration duration to the overall speed of the experiment can be considered negligible.

The Monte Carlo simulation estimates the F value more accurately than the analytical model, since it also takes into account additional uncertainty contributions that are caused by the phasor analysis method, as

opposed to the analytical model which uses a simpler CMM method. The accuracy of the analytical model can be improved by developing a more comprehensive version that includes the performance impact of additional factors such as pile-up, dark count rate and the phasor analysis. In both the analytical model and Monte Carlo simulation, the lifetime variation in different areas of the dye solution samples is excluded. Despite the generally uniform biochemical properties of the dye solutions, a possible contribution of spatial non-uniformity to the measured F value cannot be ruled out. The pile-up effect is also excluded in both approaches, which is potentially another reason for the discrepancy between the measured and estimated F values in the R6G sample, which suffers from pile-up more severely than the rest of the samples.

4.3 Phasor analysis in highly scattering medium

The contribution of the author in the work presented in this section is the development of the detector firmware and the generation of the gate images from the raw data.

Three major challenges in *in vivo* fluorescence lifetime imaging of a cell are scattering, absorption and autofluorescence of the tissue in which the cell is buried. Scattering, by changing the trajectory of the excitation/emission photons, increases both the temporal and spatial distribution of the signal. Tissue absorption attenuates the signal intensity, resulting in higher shot noise for a given excitation power. Finally, autofluorescence, when it is an unwanted source of fluorescence rather than the target signal, introduces systematic errors to the lifetime estimation by generating one or more additional exponential decay components, in addition to increasing the shot noise. The intensity of these effects increases with the depth of the cell in the tissue, thereby limiting the maximum depth at which the analysis of the target cell is possible. Conventionally, near-infrared (NIR) (and more recently, short-wave infrared (SWIR)) wavelengths are selected for *in vivo* imaging due to deeper tissue penetration and lower scattering/autofluorescence. On the other hand, visible wavelengths offer advantages including ease of use, low cost instrumentation (silicon photodetectors), and larger selection of fluorescent dyes [211, 242, 243].

In [211], the phasor-based widefield FLIM technique using long, overlapping gates was applied to fluorescence lifetime imaging in a highly scattering medium, using SwissSPAD2 for detection and 532 nm pulsed laser for excitation. The phasor analysis method used in this experiment is similar to the method used in the other experiments reported earlier in this chapter, which is based on [85, 90]. On the other hand, the method used in this experiment uses a different technique for uncorrelated background compensation, and additionally corrects for autofluorescence [83]. The uncorrelated background was compensated by a 3-point correction technique which estimates the background noise floor from the photon counts of the maximum, minimum, and median gate positions. While being only applicable to single-exponential decays, this method avoids the need for the acquisition of additional gate positions containing negligible signal. The autofluorescence correction was performed using a method that simply subtracts the gate response of only the phantom from the gate response of the sample with a phantom and the cells.

Among several experiments conducted in this work, the experiment that provides the most comprehensive assessment of the feasibility of *in vivo* imaging is summarized in Figure 4:14 [211]. This experiment is designed to simulate the identification of a tumor in the subcutaneous tissue, in which the tumor is labeled and therefore exhibits a different lifetime from the healthy cells around it. Figure 4:14(a) is the illustration of the sample structure, which consists of A549 cells (adenocarcinomic human alveolar basal epithelial cells) injected into two parallel capillaries. The cells in the left capillary are transfected with the fluorescent protein (FP) mCyRFP1, whereas the cells in the right capillary are unlabeled control cells. The field of view (FOV) of the detector

includes both capillaries, as shown in the dashed rectangle in yellow. This FOV was captured under two environmental conditions: without phantom (Figure 4:14(b, c)), and with a 1.5 mm thick phantom (Figure 4:14(d, e)). The phasor locations of the sample with no phantom indicate that the transfected cells include exponential components from both cell autofluorescence and fluorescence of the FP, as expected. In the sample with phantom, the fluorescent signal is spread around the FOV due to scattering, which can be seen from the fluorescence intensity and phasor location uniformity of both FOV halves in Figure 4:14(d) and (e). Despite the limited 2D localization ability due to scattering, it is still possible to detect the presence of the transfected cells from the extracted lifetimes. It can be seen from Figure 4:14(f) that the signal of the FP is distinguishable from phantom autofluorescence at up to 1.5 mm thickness, thus demonstrating the ability to detect a FP in tissue up to this thickness, in this kind of set-up.

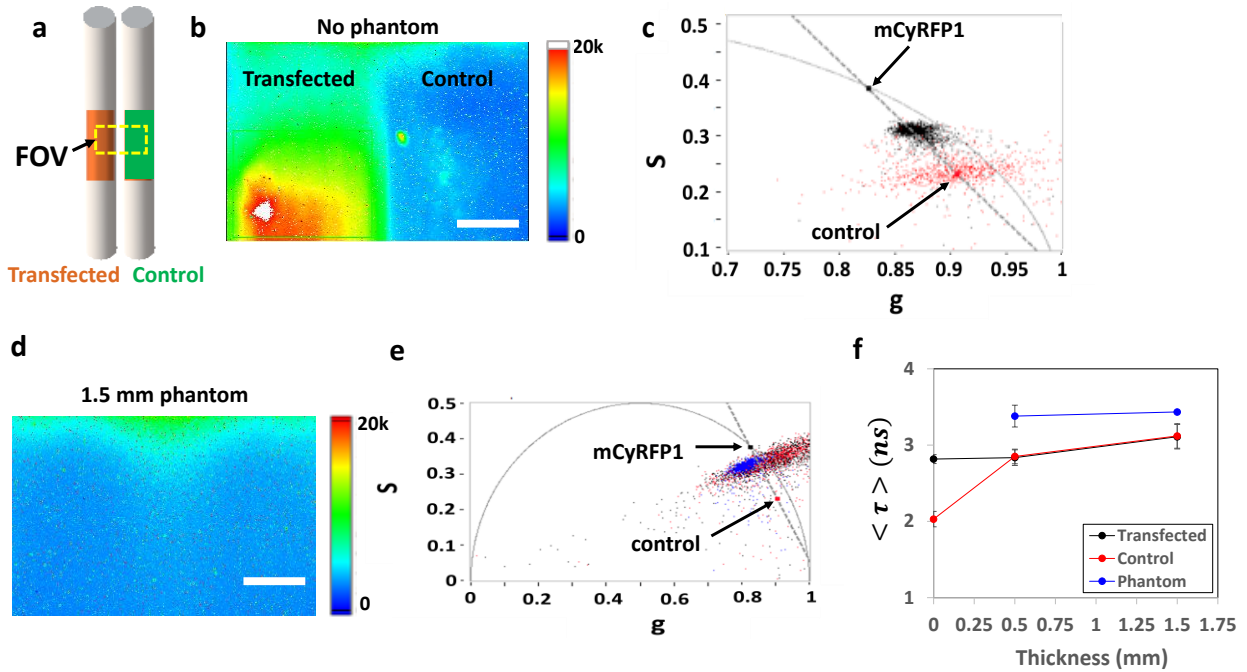


Figure 4:14 Fluorescence lifetime imaging of a cell labeled with a fluorescent protein (FP), behind tissue-like phantom with a thickness up to 1.5 mm [211]. (a) Illustration of the sample structure. A549 cells transfected with mCyRFP1, and unlabeled A549 control cells are placed in two parallel capillaries, which align with left and right halves of the field of view (FOV), respectively. (b, d) Fluorescence intensity images of the two capillaries with (b) no phantom and (d) 1.5 mm phantom. (c, e) Phasor plots of the sample with (c) no phantom and (e) 1.5 mm phantom. (f) Phase lifetimes of transfected cells, control cells and phantom autofluorescence at various phantom thickness values.

The results summarized above are promising as they indicate the compatibility of SwissSPAD2 for various biomedical applications which require *in vivo* imaging of cells through real tissue, such as cancer detection.

4.4 Computational photography for high-speed, low-light-level scenes

The contribution of the author in the work presented in this section is the development of the camera software/firmware that is compatible with the “quanta burst photography” method, and the generation of binary images from raw sensor data.

Single-photon sensitivity and fast frame rate of SwissSPAD2 make it a suitable camera of choice for the imaging of fast phenomena. In conventional CMOS image sensors, readout noise is determined by the number of frames, rather than the exposure time. As a result, particularly in low-light-level scenes, a significant SNR drop

occurs when an integration window is divided into multiple frames. On the other hand, the primary source of noise in SwissSPAD2 is dark noise, which is independent of the number of frames corresponding to a given integration time. This characteristic allows SwissSPAD2 to output binary images at a frame rate of 97.7 kfps, dividing an integration window into 10.2 μs segments, without any penalty in SNR. These sequences of very short time windows provide it the ability to capture and correct fast motion.

In [217], this feature of SwissSPAD2 is used in motion blur correction. The work introduces a motion reconstruction algorithm for single-photon cameras, named “quanta burst photography”, developed by our collaborators in University of Wisconsin-Madison. This method essentially corrects the blur by individually aligning the raw binary frames with a reference frame in the same sequence. The principles of the method are illustrated in Figure 4:15 [217]. The first step is the characterization of the motion of the raw binary frames with respect to the reference frame. Subsequently, each binary frame is aligned with the reference frame based on its motion characteristics. The characteristics of the motion are represented in the form of patch flows assigned to individual binary frames, and the alignment is performed using warping. Finally, all warped binary frames are merged and denoised to construct the final aligned image.

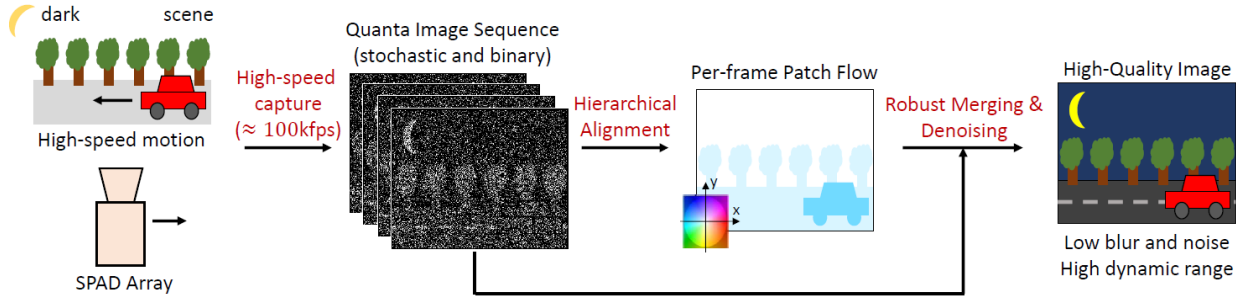


Figure 4:15 Conceptual illustration of the quanta burst photography method [217].

The SwissSPAD2 camera was operated in burst mode with binary output configuration described in Chapter 3, Subsection 3.3.3. With a minimum binary acquisition time (T_{acq}) of 10.32 μs , the camera can collect continuous sequences of up to 131,072 binary frames with a resolution of 512×256 pixels, in a single burst. Since this application entails passive imaging, i.e. with no synchronized light source to the camera, the gate was not used for photon time-stamping. Instead, it was used as an aperture to limit the exposure. This feature was particularly useful to avoid saturation and excessive dark noise in the cases where T_{acq} was increased significantly below its minimum value to extend the total acquisition time of a continuous sequence.

The motion blur correction performance of conventional and quanta burst photography was compared in multiple scenes with different levels of difficulty. The parameters of the scene which determine the difficulty of the motion correction include dynamic range, scene geometry, light level and camera/scene motion complexity. The experiment which best highlights the relative advantages of SPAD imagers is the comparison of conventional and quanta burst photography methods in a scene captured under different light levels, shown in Figure 4:16 [217]. In the experiment, the scene was captured with a sequence of 2,000 binary images using SwissSPAD2. The motion was generated by moving the camera horizontally during the exposure. In the case of conventional burst photography, readout noise and quantization noise were added to the original binary images in order to emulate the output of a conventional CMOS camera. In the case of conventional single photography, the noise was added to the sum of the image sequence. A visual comparison of the first and second columns shows the trade-off between motion reconstruction ability and SNR in conventional CMOS

imagers. On the other hand, as seen in the last column, this trade-off does not exist in quanta burst photography with SwissSPAD2, which combines the higher SNR of conventional single photography with the motion correction ability of conventional burst photography.

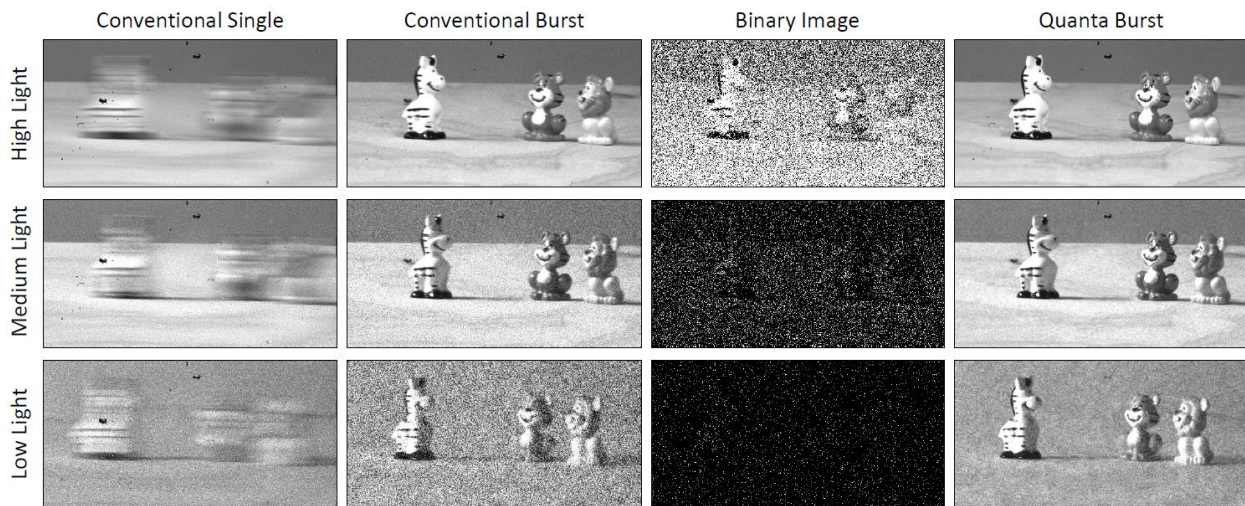


Figure 4:16 Motion correction performance comparison of conventional and quanta burst photography in a scene under high, medium and low light levels [217]. The scene was captured with a 2,000-frame binary sequence using SwissSPAD2. The image sequence used by conventional burst photography was generated by adding read and quantization noise to the binary image sequence after pile-up correction, in order to emulate the raw images captured by a conventional CMOS camera. The image of conventional single photography was generated by adding read noise and quantization error to the sum of the raw binary images. The trade-off between blur correction and SNR in conventional photography, which is particularly visible in the low-light scene, is absent in quanta burst imaging. This is enabled by the ability of SwissSPAD2 to reach high frame rates without additional noise.

Chapter 5 Conclusions

5.1 Summary

Single-photon avalanche diode (SPAD) imagers have become the sensor of choice in many time-resolved imaging modalities, mainly because of their single-photon sensitivity with virtually no readout noise, low timing jitter, and natively digital nature. All these features, along with full CMOS compatibility, make SPAD imagers the ideal system-on-chip implementation of advanced image sensors for bioimaging applications.

The primary objective of this thesis was the study, design, and characterization of large-format SPAD imagers with time-resolved imaging capability, with a special focus on widefield FLIM. Widefield FLIM image sensors demonstrated in this thesis are compatible with real-time operation at high spatial and timing resolutions, owing to high global photon count rates and support of precise timing capabilities. Within this framework, two large-format SPAD camera systems were presented. The design of the pixel, along with global readout and control, leads to the description and characterization of the entire system, including external components, such as the FPGA, data pre-processing, memory management, data communication, and data post-processing.

A time gating architecture was selected to implement time-resolved imaging modalities. This choice was based on the fact that this approach does in general enable higher global photon count rates in large pixel formats, and has a simpler pixel-level implementation, compared to TDC-based architectures. The first imager, SwissSPAD2, employed the largest pixel format of any SPAD camera (512×512 pixels) at the time of its publication. This was made possible by a pixel pitch of 16.38 μm . The imager is capable of generating binary images at frame rates up to 97.7 kfps. To cope with such a massive throughput, we compressed the binary frames in the FPGA using a flexible approach that enabled us to trade-off time granularity and continuous streaming. The SPAD used in SwissSPAD2 offers one of the best combinations of PDP and DCR among all SPADs ever designed in standard CMOS process technologies, enabling a maximum PDP of over 50% and a median DCR of 7.5 cps to be achieved *simultaneously*. One of the most important performance drawbacks of SPAD-based imagers, i.e. their low fill factor, is partially compensated by microlenses, which increase the 10.5% native fill factor by a factor of 4.2 for the optimal angle of incidence of photons. At 20 MHz laser PRF and maximum frame rate, the minimum achievable gate width in SwissSPAD2 is 10.8 ns.

The second imager, SwissSPAD3, was designed with the objective to overcome the limitations of SwissSPAD2, particularly on photon economy and photon sensitivity. SwissSPAD3 features a drastically smaller gate width of 1 ns, while enabling 100% duty cycle at slightly smaller format (500×500 pixels) and an identical 16.38 μm pixel pitch. To the best of our knowledge, SwissSPAD3 achieves the best gate performance in any large-format SPAD imagers ever published. Its gate characteristics were tested primarily at a laser PRF of 40 MHz, while there is a clear potential to operate at much higher PRFs. Importantly, the short gate width came with a significant improvement in gate skew; the FWHM of the rising/falling edge position distribution of the pixels in a single row ranges between 81.2 ps and 134 ps, which is the state-of-the-art at the time this thesis was written.

On the fundamental image science side, it was demonstrated that, when the gate is fully open, the statistical dispersion of the recorded photon count is almost entirely determined by shot noise and pile-up. The SNR drop caused by the gate shape was found to be only 0.05-0.11 dB. In SwissSPAD3, the maximum binary readout speed was measured as 49.8 kfps, approximately half of that of SwissSPAD2. This is mainly due to the representation of a pixel output as two bits as opposed to one bit in SwissSPAD2, considering the approximately constant chip throughput limitation. In the currently available camera system, both halves of the chip can be operated sequentially; while the system hardware supports parallel operation, this feature requires software improvements. The microlensed version of SwissSPAD3 is also available; its characterization is planned for the near future.

Following the design and characterization of these two image sensor chips, the next phase of this research was the demonstration of widefield FLIM. Within the framework of this thesis, a suite of experiments was conducted to characterize the FLIM performance of SwissSPAD2 in a very comprehensive way, in collaboration with UCLA. In these experiments, the phasor approach was selected for its compatibility with real-time data processing and graphical representation feature for rapid, intuitive analysis. The objective of the first experiment was to determine the separability of two single-exponential decays. Using phasor plots, it was shown that the SwissSPAD2 system is capable of distinguishing two lifetimes with a difference as low as 1.4 ns, using only 16 gate positions. The second experiment aimed to evaluate the influence of frame rate on the phase lifetime precision. In this experiment, the mean and dispersion of the measured phase lifetimes of two dyes were recorded for a number of gate positions ranging from 2,800 to 8. The results showed that, down to 8 gates, the variation of the phase lifetime determination precision with the number of gate positions is primarily caused by the decreased photon count; the photon economy is therefore largely unaffected by the number of gates. Under these settings, the acquisition speed of the data set with 8 gates was 24.3 fps, thus demonstrating real time streaming capability.

The influence of the gate width on the phase lifetime determination performance was also investigated. In our experiments, the accuracy, precision, and photon economy of lifetime extraction were calculated for 7 different gate widths ranging between 10.8 ns and 22.8 ns. The measured results were then compared to an analytical model and Monte Carlo simulation. The analytical model includes the effects of shot noise, and the width of an ideal square gate window. It also includes the uncertainty introduced by the calibration stage, which is uncorrelated to the uncertainty of the measurement of the sample of interest. Monte Carlo simulations include the uncertainty introduced by the phasor calculation method. In this experiment, it was shown that narrower gate widths improve photon economy. The discrepancy between the experimental results and the Monte Carlo simulation results is significantly lower than the discrepancy between experimental results and the analytical model. This finding highlights the contribution of the phasor method itself to photon economy. Furthermore, neither the analytical model nor the Monte Carlo simulation include the contribution of the pile-up effect, which may explain the larger measured uncertainty than predicted by both methods.

In addition to single-exponential lifetimes, the capability of the system to analyze a mixture of two single-exponential dyes was also demonstrated. The objective of this experiment was to determine the feasibility of the system for FLIM-FRET analysis. In this experiment, the phasor ratio of five mixtures of two single-exponential dyes with different volume fractions was measured using the phasor method also selected in the previous experiments. The measurement results were then compared with the theoretical values obtained by an equation expressing the relation between the phasor ratio and volume fraction, under a fixed set of intrinsic

fluorescent parameters. The results showed a good match between theoretical and experimental results, thereby demonstrating the suitability of our method for FLIM-FRET.

The final FLIM experiment focused on the imaging aspects of FLIM, which require a more complex sample than a dye solution, with spatial features, a random intensity, and lifetime distribution. A visual comparison between the intensity and lifetime maps of the dried quantum dot solution sample prepared for this experiment indicated no direct correlation between the fluorescence intensity and lifetime, as expected. Furthermore, the distribution of phasors on the phasor plot showed that the sample generally exhibited single-exponential characteristics.

Owing to its capability of operating at high frame rates without any readout-noise related SNR penalty, SwissSPAD2 also offers unique advantages in computational photography. In collaboration with UW-Madison, the camera was employed for motion blur correction, using a technique under the general umbrella of techniques known collectively as quanta burst photography. In the system configuration selected for this application, the camera was operated at 97.7 kfps, collecting up to 131,072 binary frames in burst mode. In this experiment, it was demonstrated that the trade-off between motion reconstruction and SNR in conventional CMOS imagers is largely eliminated. This achievement was largely enabled by single-photon detection and speed capabilities of SwissSPAD2.

5.2 Future work

The future efforts in this research are planned to primarily focus on improving the general specifications of the SwissSPAD2/3 imager architecture and expanding its range of scientific applications.

In terms of system specifications, multiple improvement directions were determined with varying degrees of importance, based on the experiments conducted within the framework of this thesis. The first parameter, which requires improvement is PDE. For a given SPAD model with a fixed PDP, the only way to achieve higher PDE is fill factor improvement. As the fill factor of both imagers is limited by the SPADs rather than the pixel layout, the primary fill factor improvement method is enabled by microlenses. Currently, the measured concentration factor of microlenses is considerably lower than theoretical values, primarily as a result of the discrepancy between the designed and fabricated lens curvature. The improvement of microlens concentration factor is a critical step towards the achievement of competitive PDE levels for these imagers. Alternatively, one could also develop electrical microlensing techniques. These techniques could expand the effective fill factor of SPADs at negligible impact to noise and speed performance. Most importantly, the constraint on maximum incident photon angle could be relaxed or even eliminated.

Another performance parameter that requires improvement is photon economy. In this thesis, it was shown that the photon economy is primarily determined by the temporal resolution of the camera system, whose main contributors are SPAD jitter and gate width. For a given laser PRF, the solution to achieve higher photon economy without a sacrifice in photon sensitivity is to achieve 100% duty cycle by means of multiple parallel gate channels. The primary challenge of increasing the number of gate channels is the reduction in fill factor due to additional transistor area requirements. An effective way to overcome this challenge is to switch to a 3D-stacked process technology, in which the placement of pixel electronics is below the SPAD.

Furthermore, on-FPGA data processing can enable the mitigation of several shortcomings of the current system. Firstly, it is an essential requirement for real-time FLIM for a binary imager with a large pixel format,

which is not supported in the current system. In this system, the phasor calculation and fluorescence lifetime estimation are performed in the CPUs of the PC, where the data of each pixel is analyzed sequentially. Furthermore, the process of sorting the data that is serialized for USB transfer adds an additional computational requirements. To achieve real-time FLIM at video rate, it is a key requirement to perform the calculations of the pixels in parallel, preferably before the serialization of the data of all pixels. Secondly, the implementation of an on-chip data pre-processing stage can also reduce the dead time of the system, which increases the maximum photon count rate. Currently, the dead time of the imagers is significantly higher than the fundamental limitation of the photodiode, which is typically introduced by afterpulsing. This system dead time limitation, introduced by the throughput of the chip, can be overcome by on-chip compression of raw data. With the addition of an option to select between binary and compressed output formats, the proposed system would allow a trade-off between real-time data processing in compressed format and high timing resolution in the binary format (not to be confused with the timing resolution of the delay of a photon relative to a periodic laser pulse); the latter is often not utilized in FLIM.

On the application side, our main objective is to use our FLIM system in more challenging settings, such as *in vivo* imaging. To achieve this target, several performance challenges must be overcome. Firstly, fast image refresh rate is required due to the dynamic nature of the scenes. Secondly, the detection of fluorescence through a tissue introduces SNR challenges due to autofluorescence, scattering and absorption. The transition to a microlensed SwissSPAD3 with an on-FPGA parallel lifetime estimation capability can significantly improve photon sensitivity, photon economy and data processing speed to a level that meets the performance requirements for *in vivo* imaging. This capability can open the way to new applications in life sciences, such as cancer detection, tracking of protein-protein interactions, tracking of cell migration, and FLIM endoscopy. Given our ability in the future to increase the camera frame rate up to Mfps levels by shrinking spatial resolution and image bit depth without any readout noise penalty, the use of our systems can be potentially extended to applications with even more demanding speed requirements, such as action potential tracking in neurons.

Computational photography can also benefit from system level improvements on the detector side. The most critical improvements required in this application include the implementation of a continuous stream of binary data via a PCIe link, and the addition of color filters for enhanced demosaicking capability.

Appendix A: SwissSPAD2 crosstalk characteristics

Sample 1					Sample 2				
-0.001	0.007	0.007	0.006	-0.001	0.004	0.004	0.003	0.001	0.002
0.004	0.034	0.057	0.032	-0.003	-0.001	0.030	0.059	0.031	0.007
0.000	0.056	100	0.059	0.001	0.012	0.056	100	0.073	0.001
0.001	0.030	0.053	0.029	0.003	0.007	0.032	0.056	0.031	0.003
-0.001	0.006	0.004	0.000	0.001	-0.005	0.002	0.008	-0.001	0.000

Sample 3					Sample 4				
0.002	0.000	0.006	0.000	0.002	-0.003	0.006	0.006	-0.001	0.000
0.005	0.028	0.054	0.029	0.005	-0.003	0.026	0.056	0.032	0.004
0.003	0.061	100	0.064	0.009	0.007	0.060	100	0.055	0.007
0.000	0.032	0.071	0.032	-0.002	-0.001	0.026	0.055	0.029	0.002
0.005	0.001	0.008	-0.003	0.002	-0.003	0.004	0.003	0.001	-0.004

Sample 5				
0.006	0.001	0.005	0.006	-0.005
0.005	0.033	0.062	0.032	0.005
0.013	0.067	100	0.065	0.008
0.000	0.035	0.066	0.025	-0.002
0.009	0.003	0.003	0.002	0.003

Figure A:1 Crosstalk characterization results of five SwissSPAD2 samples at 6.5 V excess bias voltage [81]. In each matrix, the values indicate the crosstalk probability of the pixels surrounding the reference pixel in the center. The values are calculated as the normalized median photon counts of multiple hot pixels and their neighbors in a dark image. The median dark count was subtracted from all pixels before the calculation. The results show that the crosstalk probability between nearest neighbors is less than 0.075%.

References

- [1] M.-C. Amann, T. Bosch, M. Lescure, R. Myllyla and M. Rioux, "Laser ranging: a critical review of usual techniques for distance measurement," *Optical Engineering*, vol. 40, no. 1, pp. 10-19, 2001.
- [2] R. J. Walker, J. A. Richardson and R. K. Henderson, "A 128x96 Pixel Event-Driven Phase-Domain $\Delta\Sigma$ Based Fully Digital 3D Camera in 0.13 μm CMOS Imaging Technology," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2011.
- [3] M. Z. Brown, D. Burschka and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993-1008, 2003.
- [4] Y.-J. Chang, S.-F. Chen and J.-D. Huang, "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities," *Research in Developmental Disabilities*, vol. 32, no. 6, pp. 2566-2570, 2011.
- [5] R. Lange and P. Seitz, "Solid-State Time-of-Flight Range Camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390-397, 2001.
- [6] G. A. Morton, "Photomultipliers for Scintillation Counting," *R. C. A. Review*, vol. 10, 1949.
- [7] R. A. Colyer, O. H. W. Siegmund, A. S. Tremsin, J. V. Vallerga, S. Weiss and X. Michalet, "Phasor imaging with a widefield photon-counting detector," *Journal of Biomedical Optics*, vol. 17, no. 1, p. 016008, 2012.
- [8] J. Pearson and H. Cölfen, "ICCD camera technology with constant illumination source and possibilities for application in multiwavelength analytical ultracentrifugation," *RSC Advances*, vol. 8, no. 1, pp. 40655-40662, 2018.
- [9] M. W. Seo, Y. Shirakawa, Y. Kawata, K. Kagawa, K. Yasutomi and S. Kawahito, "A Time-Resolved Four-Tap Lock-In Pixel CMOS Image Sensor for Real-Time Fluorescence Lifetime Imaging Microscopy," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 8, pp. 2319-2330, 2018.
- [10] C. Veerappan, J. Richardson, R. Walker, D. U. Li, M. W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. K. Henderson and E. Charbon, "A 160x128 Single-Photon Image Sensor with On-Pixel 55ps 10b Time-to-Digital Converter," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2011.
- [11] T. Ogawa, H. Sakai, Y. Suzuki, K. Takagi and K. Morikawa, "Pedestrian Detection and Tracking using in-vehicle Lidar for Automotive Application," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, 2011.

-
- [12] M. Bansal, B. Matei, B. Southall, J. Eledath and H. Sawhney, "A LIDAR streaming architecture for mobile robotics with application to 3D structure characterization," in *IEEE International Conference on Robotics and Automation*, Shanghai, 2011.
- [13] D. M. Winker, R. H. Couch and M. P. McCormich, "An overview of LITE: NASA's Lidar In-space Technology Experiment," *Proceedings of the IEEE*, vol. 84, no. 2, pp. 164-180, 1996.
- [14] M. O'Toole, D. B. Lindell and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, vol. 555, pp. 338-341, 2018.
- [15] C. Bruschini, H. Homulle, I. M. Antolovic, S. Burri and E. Charbon, "Single-photon avalanche diode imagers in biophotonics: review and outlook," *Light: Science & Applications*, vol. 8, p. 87, 2019.
- [16] L. Marcu, "Fluorescence Lifetime Techniques in Medical Applications," *Annals of Biomedical Engineering*, vol. 40, no. 2, pp. 304-331, 2012.
- [17] M. Y. Berezin and S. Achilefu, "Fluorescence lifetime measurements and biological imaging," *Chemical Reviews*, vol. 110, no. 5, pp. 2641-2684, 2010.
- [18] X. Liu, D. Lin, W. Becker, J. Niu, B. Yu, L. Liu and J. Qu, "Fast fluorescence lifetime imaging techniques: A review on challenge and development," *Journal of Innovative Optical Health Sciences*, vol. 12, no. 5, p. 1930003, 2019.
- [19] R. K. Henderson, B. R. Rae and D.-U. Li, "Complementary metal-oxide-semiconductor (CMOS) sensors for fluorescence lifetime imaging (FLIM)," in *High performance silicon imaging: fundamentals and applications of CMOS and CCD sensors*, Elsevier, 2014, pp. 312-347.
- [20] L. M. Hirvonen and K. Suhling, "Fast Timing Techniques in FLIM Applications," *Frontiers in Physics*, vol. 8, p. 161, 2020.
- [21] S. Rajoria, L. Zhao, X. Intes and M. Barroso, "FLIM-FRET for Cancer Applications," *Current Molecular Imaging*, vol. 3, no. 2, pp. 144-161, 2014.
- [22] D. L. Bailey, D. W. Townsend, P. E. Valk and M. N. Maisey, *Positron Emission Tomography: Basic Sciences*, London: Springer, 2005.
- [23] J. C. Gore, S. G. Horovitz, C. J. Cannistraci and P. Skudlarski, "Integration of fMRI, NIROT and ERP for studies of human brain function," *Magnetic Resonance Imaging*, vol. 24, no. 4, pp. 507-513, 2006.
- [24] A. Pifferi, D. Contini, A. Dalla Mora, A. Farina, L. Spinelli and A. Torricelli, "New frontiers in time-domain diffuse optics, a review," *Journal of Biomedical Optics*, vol. 21, no. 9, p. 091310, 2016.
- [25] D. J. Gardiner and P. R. Graves, *Practical Raman spectroscopy*, Berlin: Springer, 1989.

- [26] J. Kostamovaara, J. Tenhunen, M. Kögler, I. Nissinen, J. Nissinen and P. Keränen, "Fluorescence suppression in Raman spectroscopy using a time-gated CMOS SPAD," *Optics Express*, vol. 21, no. 25, pp. 31632-31645, 2013.
- [27] H. Chen, E. R. Farkas and W. W. Webb, "In Vivo Applications of Fluorescence Correlation Spectroscopy," in *Methods in Cell Biology*, Amsterdam, Elsevier, 2008, pp. 3-35.
- [28] Y. Bromberg, O. Katz and Y. Silberberg, "Ghost imaging with a single detector," *Physical Review A*, vol. 79, no. 5, p. 053840, 2009.
- [29] H. Defienne, M. Reichert, J. W. Fleischer and D. Faccio, "Quantum image distillation," *Science Advances*, vol. 5, no. 10, p. eaax0307, 2019.
- [30] K. Suhling, P. M. W. French and D. Phillips, "Time-resolved fluorescence microscopy," *Photochemical & Photobiological Sciences*, vol. 4, no. 1, pp. 13-22, 2005.
- [31] J. W. Borst and A. J. W. G. Visser, "Fluorescence lifetime imaging microscopy in life sciences," *Measurement Science and Technology*, vol. 20, no. 10, p. 102002, 2010.
- [32] X. Michalet, A. N. Kapanidis, T. Laurence, F. Pinaud, S. Doose, M. Pflughoeft and S. Weiss, "The power and prospects of fluorescence microscopies and spectroscopies," *Annual Review of Biophysics and Biomolecular Structure*, vol. 32, pp. 161-182, 2003.
- [33] A. Jablonski, "Efficiency of Anti-Stokes Fluorescence in Dyes," *Nature*, vol. 131, pp. 839-840, 1933.
- [34] H. Szmazinski and J. R. Lakowicz, "Optical measurements of pH using fluorescence lifetimes and phase-modulation fluorometry," *Analytical Chemistry*, vol. 65, no. 13, pp. 1668-1674, 1993.
- [35] J. Lakowicz, *Principles of Fluorescence Spectroscopy*, New York: Springer, 2006.
- [36] F. Schaefe, I. Demarco and R. N. Day, "FRET Imaging in the Wide-Field Microscope," in *Molecular Imaging*, Elsevier Inc., 2005, pp. 72-94.
- [37] R. M. Clegg, "Förster resonance energy transfer—FRET what is it, why do it, and how it's done," in *Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier, 2009, pp. 1-57.
- [38] P. J. Alt, "RESOLFT nanoscopy with water-soluble synthetic fluorophores," Georg-August-Universität Göttingen, Göttingen, Germany, 2017. PhD Thesis.
- [39] I. Bugiel, K. König and H. Wabnitz, "Investigation of Cells by Fluorescence Laser Scanning Microscopy with Subnanosecond Time Resolution," *Lasers in the Life Sciences*, vol. 3, no. 1, pp. 47-53, 1989.
- [40] W. Becker, A. Bergmann and S. Smietana, "Fast-Acquisition TCSPC FLIM with sub-25 ps IRF Width," *Proceedings of SPIE*, vol. 10882, p. 1088206, 2017.
- [41] B. Korzh, Q. Y. Zhao, J. P. Allmaras, S. Frasca, T. M. Autry, E. A. Bersin, A. D. Beyer, R. M. Briggs, B. Bumble, M. Colangelo, G. M. Crouch, A. E. Dane, T. Gerrits, A. E. Lita, F. Marsili, G. Moody, C. Pena,

- E. Ramirez, J. D. Rezac, N. Sinclair, M. J. Stevens, A. E. Velasco, V. B. Verma, E. E. Wollman, S. Xie, D. Zhu, P. D. Hale, M. Spiropulu, K. L. Silverman, R. P. Mirin, S. W. Nam, A. G. Kozorezov, M. D. Shaw and K. K. Berggren, "Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector," *Nature Photonics*, vol. 14, pp. 250-255, 2020.
- [42] M. Z. Straayer and M. H. Perrott, "A Multi-Path Gated Ring Oscillator TDC With First-Order Noise Shaping," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1089-1098, 2009.
- [43] M. Lee and A. A. Abidi, "A9 b, 1.25 ps Resolution Coarse–Fine Time-to-Digital Converter in 90 nm CMOS that Amplifies a Time Residue," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 769-777, 2008.
- [44] H. C. Gerritsen, A. Draaijer, D. J. van den Heuvel and A. V. Agronskaia, "Fluorescence lifetime imaging in scanning microscopy," in *Handbook of biological confocal microscopy*, New York, NY, Springer, 2006, pp. 516-534.
- [45] C. Eggeling, J. Widengren, R. Rigler and C. A. M. Seidel, "Photobleaching of fluorescent dyes under conditions used for single-molecule detection: evidence of two-step photolysis," *Analytical Chemistry*, vol. 70, no. 13, pp. 2651-2659, 1998.
- [46] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, Berlin: Springer, 2005.
- [47] W. Zhong, P. Urayama and M. A. Mycek, "Imaging fluorescence lifetime modulation of a ruthenium-based dye in living cells: the potential for oxygen sensing," *Journal of Physics D: Applied Physics*, vol. 36, no. 14, p. 1689–1695, 2003.
- [48] J. Requejo-Isidro, J. McGinty, I. Munro, D. S. Elson, N. P. Galletly, M. J. Lever, M. A. A. Neil, G. W. H. Stamp, P. M. W. French, P. A. Kellett, J. D. Hares and A. K. L. Dymoke-Bradshaw, "High-speed wide-field time-gated endoscopic fluorescence-lifetime imaging," *Optics Letters*, vol. 29, no. 19, pp. 2249-2251, 2004.
- [49] R. W. Cole, T. Jinadasa and C. M. Brown, "Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control," *Nature Protocols*, vol. 6, no. 1, pp. 1929-1941, 2011.
- [50] F. R. S. Lord Rayleigh, "Investigations in optics, with special reference to the spectroscope," *Philosophical Magazine*, vol. 8, no. 49, pp. 261-274, 1879.
- [51] R. H. Webb, "Confocal optical microscopy," *Reports on Progress in Physics*, vol. 59, no. 3, pp. 427-471, 1996.
- [52] T. Wilson and C. J. R. Sheppard, *Theory and Practice of Scanning Optical Microscopy*, London: Academic, 1984.
- [53] S. E. D. Webb, Y. Gu, S. Leveque-Fort, J. Siegel, M. J. Cole, K. Dowling, R. Jones, P. M. W. French, M. A. A. Neil, R. Juskaitis, L. O. D. Sucharov, T. Wilson and M. J. Lever, "A wide-field time-domain

- fluorescence lifetime imaging microscope with optical sectioning," *Review of Scientific Instruments*, vol. 73, no. 4, pp. 1898-1907, 2002.
- [54] C. Buranachai, D. Kamiyama, A. Chiba, B. D. Williams and R. M. Clegg, "Rapid Frequency-Domain FLIM Spinning Disk Confocal Microscope: Lifetime Resolution, Image Improvement and Wavelet Analysis," *Journal of Fluorescence*, vol. 18, pp. 929-942, 2008.
- [55] J. Huysken, J. Swoger, F. del Bene, J. Wittbrodt and E. H. K. Stelzer, "Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy," *Science*, vol. 305, no. 5686, pp. 1007-1009, 2004.
- [56] H. Choi, D. S. Tzeranis, J. W. Cha, P. Clemenceau, S. J. G. de Jong, L. K. van Geest, J. H. Moon, I. V. Yannas and P. T. C. So, "3D-resolved fluorescence and phosphorescence lifetime imaging using temporal focusing wide-field two-photon excitation," *Optics Express*, vol. 20, no. 24, pp. 26219-26235, 2012.
- [57] R. K. Henderson, N. Johnston, F. M. Della Rocca, H. Chen, D. D. U. Li, G. Hungerford, R. Hirsch, D. McLoskey, P. Yip and D. J. S. Birch, "A 192×128 Time Correlated SPAD Image Sensor in 40-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 7, pp. 1907-1916, 2019.
- [58] S. Orthaus-Mueller, B. Kraemer, A. Tannert, T. Roehlicke, M. Wahl, H. J. Rahn, F. Koberling and R. Erdmann, "Rapid FLIM: The new and innovative method for ultra-fast imaging of biological processes," in *SPIE BIOS*, San Francisco, 2017.
- [59] S. P. Poland, A. T. Erdogan, N. Krstajic, J. Levitt, V. Devaughes, R. J. Walker, D. D. U. Li, S. M. Ameer-Beg and R. K. Henderson, "New high-speed centre of mass method incorporating background subtraction for accurate determination of fluorescence lifetime," *Optics Express*, vol. 24, no. 7, pp. 6899-6915, 2016.
- [60] A. Ingargiola, E. Lerner, S. Y. Chung, F. Panzeri, A. Gulinatti, I. Rech, M. Ghioni, S. Weiss and X. Michalet, "Multispot single-molecule FRET: High-throughput analysis of freely diffusing molecules," *PLoS ONE*, vol. 12, no. 4, p. e0175766, 2017.
- [61] C. J. R. Sheppard, "Super-resolution in confocal imaging," *Optik*, vol. 80, no. 2, pp. 53-54, 1988.
- [62] M. Castello, G. Tortarolo, M. Buttafava, T. Deguchi, F. Villa, S. Koho, L. Pesce, M. Oneto, S. Pelicci, L. Lanzano, P. Bianchini, C. J. R. Sheppard, A. Diaspro, A. Tosi and G. Vicidomini, "A robust and versatile platform for image scanning microscopy enabling super-resolution FLIM," *Nature Methods*, vol. 16, no. 1, pp. 175-178, 2019.
- [63] T. Wilson, "Resolution and optical sectioning in the confocal microscope," *Journal of Microscopy*, vol. 244, no. 2, pp. 113-121, 2011.
- [64] C. J. R. Sheppard, S. B. Mehta and R. Heintzmann, "Superresolution by image scanning microscopy using pixel reassignment," *Optics Letters*, vol. 38, no. 15, pp. 2889-2892, 2013.

- [65] C. D. Wilms, H. Schmidt and J. Eilers, "Quantitative two-photon Ca^{2+} imaging via fluorescence lifetime analysis," *Cell Calcium*, vol. 40, no. 1, pp. 73-79, 2006.
- [66] K. Zheng, T. P. Jensen and D. A. Rusakov, "Monitoring intracellular nanomolar calcium using fluorescence lifetime imaging," *Nature Protocols*, vol. 13, no. 1, pp. 581-597, 2018.
- [67] A. Margineanu, J. J. Chan, D. J. Kelly, S. C. Warren, D. Flatters, S. Kumar, M. Katan, C. W. Dunsby and P. M. W. French, "Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM)," *Scientific Reports*, vol. 6, no. 1, p. 28186, 2016.
- [68] M. Maus, M. Cotlet, J. Hofkens, T. Gensch, F. C. De Schryver, J. Schaffer and C. A. M. Seidel, "An Experimental Comparison of the Maximum Likelihood Estimation and Nonlinear Least-Squares Fluorescence Lifetime Analysis of Single Molecules," *Analytical Chemistry*, vol. 73, no. 9, pp. 2078-2086, 2001.
- [69] P. Hall and B. Selinger, "Better Estimates of Exponential Decay Parameters," *The Journal of Physical Chemistry*, vol. 85, no. 20, pp. 2941-2946, 1981.
- [70] G. Wu, T. Nowotny, Y. Chen and D. D. U. Li, "GPU acceleration of time-domain fluorescence lifetime imaging," *Journal of Biomedical Optics*, vol. 21, no. 1, p. 017001, 2016.
- [71] J. A. Jo, Q. Fang and L. Marcu, "Ultrafast Method for the Analysis of Fluorescence Lifetime Imaging Microscopy Data Based on the Laguerre Expansion Technique," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 11, no. 4, pp. 835-845, 2005.
- [72] K. K. Sharman, A. Periasamy, H. Ashworth, J. N. Demas and N. H. Snow, "Error Analysis of the Rapid Lifetime Determination Method for Double-Exponential Decays and New Windowing Schemes," *Analytical Chemistry*, vol. 71, no. 5, pp. 947-952, 1999.
- [73] R. M. Ballew and J. N. Demas, "An Error Analysis of the Rapid Lifetime Determination Method for the Evaluation of Single Exponential Decays," *Analytical Chemistry*, vol. 61, no. 1, pp. 30-33, 1989.
- [74] D. U. Li, R. Andrews, J. Arlt and R. Henderson, "Hardware implementation algorithm and error analysis of high-speed fluorescence lifetime sensing systems using center-of-mass method," *Journal of Biomedical Optics*, vol. 15, no. 1, p. 017006, 2010.
- [75] D. D. U. Li, H. Yu and Y. Chen, "Fast bi-exponential fluorescence lifetime imaging analysis methods," *Optics Letters*, vol. 40, no. 3, pp. 336-339, 2015.
- [76] R. Yao, M. Ochoa, P. Yan and X. Intes, "Net-FLICS: fast quantitative wide-field fluorescence lifetime imaging with compressed sensing – a deep learning approach," *Light: Science & Applications*, vol. 8, no. 1, p. 26, 2019.
- [77] M. A. Digman, V. R. Caiolfa, M. Zamai and E. Gratton, "The Phasor Approach to Fluorescence Lifetime Imaging Analysis," *Biophysical Journal*, vol. 94, no. 2, pp. L14-L16, 2008.

- [78] X. Wang, Y. Wang, Z. Zhang, M. Huang, Y. Fei, J. Ma and L. Mi, "Discriminating different grades of cervical intraepithelial neoplasia based on label-free phasor fluorescence lifetime imaging microscopy," *Biomedical Optical Express*, vol. 11, no. 4, pp. 1977-1990, 2020.
- [79] A. Celli, S. Sanchez, M. Behne, T. Hazlett, E. Gratton and T. Mauro, "The Epidermal Ca²⁺ Gradient: Measurement Using the Phasor Representation of Fluorescent Lifetime Imaging," *Biophysical Journal*, vol. 98, no. 5, pp. 911-921, 2010.
- [80] C. Stringari, A. Cinquin, O. Cinquin, M. A. Digman, P. J. Donovan and E. Gratton, "Phasor approach to fluorescence lifetime microscopy distinguishes different metabolic states of germ cells in a live tissue," *PNAS*, vol. 108, no. 33, pp. 13582-13587, 2011.
- [81] A. Ulku, A. Ardelean, M. Antolovic, S. Weiss, E. Charbon, C. Bruschini and X. Michalet, "Wide-field time-gated SPAD imager for phasor-based FLIM applications," *Methods and Applications in Fluorescence*, vol. 8, no. 2, p. 024002, 2020.
- [82] H. Chen, N. Ma, K. Kagawa, S. Kawahito, M. Digman and E. Gratton, "Widefield multifrequency fluorescence lifetime imaging using a two-tap complementary metal-oxide semiconductor camera with lateral electric field charge modulators," *Journal of Biophotonics*, vol. 12, no. 5, p. e201970014, 2019.
- [83] A. H. A. Clayton, Q. S. Hanley and P. J. Verveer, "Graphical representation and multicomponent analysis of single-frequency fluorescence lifetime imaging microscopy data," *Journal of Microscopy*, vol. 213, no. 1, pp. 1-5, 2004.
- [84] G. I. Redford and R. M. Clegg, "Polar plot representation for frequency-domain analysis of fluorescence lifetimes," *Journal of Fluorescence*, vol. 15, no. 5, pp. 805-815, 2005.
- [85] F. Fereidouni, A. Esposito, G. A. Blab and H. C. Gerritsen, "A modified phasor approach for analyzing time-gated fluorescence lifetime images," *Journal of Microscopy*, vol. 244, no. 3, pp. 248-258, 2011.
- [86] X. Michalet, "Continuous and discrete phasor analysis of binned or time-gated periodic decays," *AIP Advances (in press)*, vol. 11, 2021.
- [87] M. J. Cole, J. Siegel, S. E. D. Webb, R. Jones, K. Dowling, M. J. Dayel, D. Parsons-Karavassilis, P. M. W. French, M. J. Lever, L. O. D. Sucharov, M. A. A. Neil, R. Juskaitis and T. Wilson, "Time-domain whole-field fluorescence lifetime imaging with optical sectioning," *Journal of Microscopy*, vol. 203, no. 3, pp. 246-257, 2001.
- [88] A. D. Elder, J. H. Frank, J. Swartling, X. Dai and C. F. Kaminski, "Calibration of a wide-field frequency-domain fluorescence lifetime microscopy system using light emitting diodes as light sources," *Journal of Microscopy*, vol. 224, no. 2, pp. 166-180, 2006.
- [89] R. A. Colyer, C. Lee and E. Gratton, "A novel fluorescence lifetime imaging system that optimizes photon efficiency," *Microscopy Research and Technique*, vol. 71, no. 3, pp. 201-213, 2008.

-
- [90] S. J. Chen, N. Sinsuebphon, A. Rudkouskaya, M. Barroso and X. Michalet, "In vitro and in vivo phasor analysis of stoichiometry and pharmacokinetics using short-lifetime near-infrared dyes and time-gated imaging," *Journal of Biophotonics*, vol. 12, no. 3, p. e201800185, 2019.
- [91] A. Leray, C. Spriet, D. Trinel, Y. Usson and L. Heliot, "Generalization of the polar representation in time domain fluorescence lifetime imaging microscopy for biological applications: practical implementation," *Journal of Microscopy*, vol. 248, no. 1, pp. 66-76, 2012.
- [92] L. M. Hirvonen and K. Suhling, "Wide-field TCSPC: methods and applications," *Measurement Science and Technology*, vol. 28, no. 1, pp. 1-19, 2017.
- [93] S. Ranjit, L. Malacrida and E. Gratton, "Differences between FLIM phasor analyses for data collected with the Becker and Hickl SPC830 card and with the FLIMbox card," *Microscopy Research and Technique*, vol. 81, no. 9, pp. 980-989, 2018.
- [94] C. Zhang, S. Lindner, I. M. Antolovic, M. Wolf and E. Charbon, "A CMOS SPAD Imager with Collision Detection and 128 Dynamically Reallocating TDCs for Single-Photon Counting and 3D Time-of-Flight Imaging," *Sensors*, vol. 18, no. 11, p. 4016, 2018.
- [95] C. Zhang, S. Lindner, I. M. Antolovic, J. Mata Pavia, M. Wolf and E. Charbon, "A 30-frames/s, 252x144 SPAD Flash LiDAR With 1728 Dual-Clock 48.8-ps TDCs, and Pixel-Wise Integrated Histogramming," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137-1151, 2019.
- [96] L. Liu, Q. Yang, M. Zhang, Z. Wu and P. Xue, "Fluorescence lifetime imaging microscopy and its applications in skin cancer diagnosis," *Journal of Innovative Optical Health Sciences*, vol. 12, no. 5, p. 1930004, 2019.
- [97] Z. Gaviola, "Ein Fluorometer, Apparat zur Messung von Fluoreszenzabklingungszeiten," *Zeitschrift für Physik*, vol. 42, pp. 853-861, 1927.
- [98] R. D. Spencer and G. Weber, "Measurements of Subnanosecond Fluorescence Lifetimes with a Cross-correlation Phase Fluorometer," *Annals of the New York Academy of Sciences*, vol. 158, no. 1, pp. 361-376, 1969.
- [99] E. Gratton and M. Limkeman, "A continuously variable frequency cross-correlation phase fluorometer with picosecond resolution," *Biophysical Journal*, vol. 44, no. 3, pp. 315-324, 1983.
- [100] D. M. Jameson, E. Gratton and R. D. Hall, "The measurement and analysis of heterogeneous emissions by multifrequency phase and modulation fluorometry," *Applied Spectroscopy Reviews*, vol. 20, no. 1, pp. 55-106, 1984.
- [101] B. Yuan, S. R. McClellan, B. F. Al-Mifgai, E. A. Gowney and O. A. Komolafe, "A cost-efficient frequency domain fluorescence lifetime measurement system," *American Journal of Physics*, vol. 78, no. 28, pp. 28-34, 2010.

- [102] J. Philip and K. Carlsson, "Theoretical investigation of the signal-to-noise ratio in fluorescence lifetime imaging," *Journal of the Optical Society of America A*, vol. 20, no. 2, pp. 368-379, 2003.
- [103] T. W. J. Gadella, T. M. Jovin and R. M. Clegg, "Fluorescence lifetime imaging microscopy (FLIM): Spatial resolution of microstructures on the nanosecond time scale," *Biophysical Chemistry*, vol. 48, no. 2, pp. 221-239, 1993.
- [104] M. J. Serafino, B. E. Applegate and J. A. Jo, "Direct frequency domain fluorescence lifetime imaging using field programmable gate arrays for real time processing," *Review of Scientific Instruments*, vol. 91, no. 1, p. 033708, 2020.
- [105] C. X. Sheng, T. Z. Li and H. J. Tang, "Research of fluorescence lifetime imaging technology based on heterodyne method of frequency domain," *Proceedings of SPIE*, vol. 7656, no. 1, p. 76566U, 2010.
- [106] P. T. C. So, T. French and E. Gratton, "Frequency domain time-resolved microscope using a fast-scan CCD camera," *Proceedings of SPIE*, vol. 2137, no. 1, pp. 83-92, 1994.
- [107] B. A. Feddersen, D. W. Piston and E. Gratton, "Digital parallel acquisition in frequency domain fluorimetry," *Review of Scientific Instruments*, vol. 60, no. 9, pp. 2929-2936, 1989.
- [108] R. Clegg, B. Feddersen, E. Gratton and T. Jovin, "Time-resolved imaging fluorescence microscopy," *Proceedings of SPIE*, vol. 1640, pp. 448-460, 1992.
- [109] M. Raspe, K. M. Kedziora, B. van den Broek, Q. Zhao, S. de Jong, J. Herz, M. Mastop, J. Goedhart, T. W. J. Gadella, I. T. Young and K. Jalink, "siFLIM: single-image frequency-domain FLIM provides fast and photon-efficient lifetime data," *Nature Methods*, vol. 13, no. 6, pp. 501-504, 2016.
- [110] Q. Zhao, B. Schelen, R. Schouten, R. van den Oever, R. Leenen, H. van Kuijk, I. Peters, F. Polderdijk, J. Bosiers, M. Raspe, K. Jalink, J. G. S. de Jong, B. van Geest, K. Stoop and I. T. Young, "Modulated electron-multiplied fluorescence lifetime imaging microscope: all-solid-state camera for fluorescence lifetime imaging," *Journal of Biomedical Optics*, vol. 17, no. 12, p. 126020, 2012.
- [111] A. Lajevardipour and A. H. A. Clayton, "The Effect of Translational Motion on FLIM Measurements-Single Particle Phasor-FLIM," *Journal of Fluorescence*, vol. 23, pp. 671-679, 2013.
- [112] M. L. Gee, M. Burton, A. Grevis-James, M. A. Hossain, S. McArthur, E. A. Palombo, J. D. Wade and A. H. A. Clayton, "Imaging the action of antimicrobial peptides on living bacterial cells," *Scientific Reports*, vol. 3, p. 1557, 2013.
- [113] R. Franke and G. A. Holst, "Frequency-domain fluorescence lifetime imaging system (pco.flim) based on a in-pixel dual tap control CMOS image sensor," *Proceedings of SPIE*, vol. 9328, p. 93281K, 2015.
- [114] H. Chen, G. Holst and E. Gratton, "Modulated CMOS Camera for Fluorescence Lifetime Microscopy," *Microscopy Research and Technique*, vol. 78, no. 12, pp. 1075-1081, 2015.

- [115] R. G. Bennett, "Instrument to Measure Fluorescence Lifetimes in the Millimicrosecond Region," *Review of Scientific Instruments*, vol. 31, no. 12, pp. 1275-1279, 1960.
- [116] O. J. Steingraber and I. B. Berlman, "Versatile Technique for Measuring Fluorescence Decay Times in the Nanosecond Region," *Review of Scientific Instruments*, vol. 34, no. 5, pp. 524-529, 1963.
- [117] A. C. Ulku, C. Bruschini, I. M. Antolovic, Y. Kuo, R. Ankri, S. Weiss, X. Michalet and E. Charbon, "A 512 × 512 SPAD Image Sensor With Integrated Gating for Widefield FLIM," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 1, p. 6801212, 2019.
- [118] L. Parmesan, N. A. W. Dutton, N. J. Calder, N. Krstajic, A. J. Holmes, L. A. Grant and R. K. Henderson, "A 256 x 256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy," in *International Image Sensor Workshop*, Vaals, Netherlands, 2015.
- [119] I. Gyongy, N. Calder, A. Davies, N. A. W. Dutton, R. R. Duncan, C. Rickman, P. Dalgarno and R. K. Henderson, "A 256×256, 100-kfps, 61% Fill-Factor SPAD Image Sensor for Time-Resolved Microscopy Applications," *IEEE Transactions on Electron Devices*, vol. 65, no. 2, pp. 547-554, 2018.
- [120] S. Burri, Y. Maruyama, X. Michalet, F. Regazzoni, C. Bruschini and E. Charbon, "Architecture and applications of a high resolution gated SPAD image sensor," *Optics Express*, vol. 22, no. 14, pp. 17573-17589, 2014.
- [121] N. A. W. Dutton, I. Gyongy, L. Parmesan, S. Gneccchi, N. Calder, B. R. Rae, S. Pellegrini, L. A. Grant and R. K. Henderson, "A SPAD-Based QVGA Image Sensor for Single-Photon Counting and Quanta Imaging," *IEEE Transactions on Electron Devices*, vol. 63, no. 1, pp. 189-196, 2016.
- [122] K. Morimoto, A. Ardelean, M. L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini and E. Charbon, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica*, vol. 7, no. 4, pp. 346-354, 2020.
- [123] T. Okino, S. Yamada, Y. Sakata, S. Kasuga, M. Takemoto, Y. Nose, H. Koshida, M. Tamaru, Y. Sugiura, S. Saito, S. Koyama, M. Mori, Y. Hirose, M. Sawada, A. Odagawa and T. Tanaka, "A 1200×900 6μm 450fps Geiger-Mode Vertical Avalanche Photodiodes CMOS Image Sensor for a 250m Time-of-Flight Ranging System Using Direct-Indirect-Mixed Frame Synthesis with Configurable-Depth-Resolution Down to 10cm," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2020.
- [124] A. L. Trinh, S. Ber, A. Howitt, P. O. Valls, M. W. Fries, A. R. Venkitaraman and A. Esposito, "Fast single-cell biochemistry: theory, open source microscopy and applications," *Methods and Applications in Fluorescence*, vol. 7, no. 4, p. 044001, 2019.
- [125] W. S. Boyle and G. E. Smith, "Charge Coupled Semiconductor Devices," *Bell System Technical Journal*, vol. 49, no. 4, pp. 587-593, 1970.
- [126] P. Jerram, P. Pool, R. Bell, D. J. Burt, S. Bowring, S. Spencer, M. Hazelwood, I. Moody, N. Catlett and P. S. Heyes, "The LLCCD: low-light imaging without the need for an intensifier," *Proceedings of SPIE*, vol. 4306, pp. 178-186, 2001.

- [127] T. G. Etoh, D. V. T. Son, T. Yamada and E. Charbon, "Toward One Giga Frames per Second — Evolution of in Situ Storage Image Sensors," *Sensors*, vol. 13, no. 4, pp. 4640-4658, 2013.
- [128] H. Kume, H. Nakamura and M. Suzuki, "Gatable photonic detector and its image processing," *Proceedings of SPIE*, vol. 1358, pp. 1145-1155, 1991.
- [129] J. McGinty, J. Requejo-Isidro, I. Munro, C. B. Talbot, P. A. Kellett, J. D. Hares, C. Dunsby, M. A. A. Neil and P. M. W. French, "Signal-to-noise characterization of time-gated intensifiers used for wide-field time-domain FLIM," *Journal of Physics D: Applied Physics*, vol. 42, no. 13, p. 135103, 2009.
- [130] P. Urayama, W. Zhong, J. A. Beamish, F. K. Minn, R. D. Sloboda, K. H. Dragnev, E. Dmitrovsky and M. A. Mycek, "A UV–Visible–NIR fluorescence lifetime imaging microscope for laser-based biological sensing with picosecond resolution," *Applied Physics B*, vol. 76, pp. 483-496, 2003.
- [131] A. Frenkel, M. A. Sartor and M. S. Wlodawski, "Photon-noise-limited operation of intensified CCD cameras," *Applied Optics*, vol. 36, no. 22, pp. 5288-5297, 1997.
- [132] S. Hamilton and E. Conroy, "Improvement of UV on/off ratios of a gated ICCD by simultaneously gating the MCP," *Proceedings of SPIE*, vol. 4876, no. 1, pp. 1221-1228, 2003.
- [133] V. Venugopal, J. Chen and X. Intes, "Development of an optical imaging platform for functional imaging of small animals using wide-field excitation," *Biomedical Optics Express*, vol. 1, no. 1, pp. 143-156, 2010.
- [134] L. Cester, A. Lyons, M. C. Braidotti and D. Faccio, "Time-of-Flight Imaging at 10 ps Resolution with an ICCD Camera," *Sensors*, vol. 19, no. 1, pp. 1-9, 2019.
- [135] J. P. Edgecumbe, V. W. Aebi and G. A. Davis, "GaAsP photocathode with 40% QE at 550 nm," in *SPIE/IS&T Symposium on Electronic Imaging: Science and Technology*, San Jose, 1992.
- [136] T. W. Sinor, J. P. Estrera, D. L. Phillips and M. K. Rector, "Extended blue GaAs image intensifiers," in *SPIE International Symposium on Optical Science, Engineering and Instrumentation*, San Diego, 1995.
- [137] R. A. La Rue, K. A. Costello, G. A. Davis, J. P. Edgecumbe and V. W. Aebi, "Photon Counting III–V Hybrid Photomultipliers Using Transmission Mode Photocathodes," *IEEE Transactions on Electron Devices*, vol. 44, no. 4, pp. 672-678, 1997.
- [138] S. E. Moran, B. L. Ulich, W. P. Elkins, R. J. Strittmatter and M. J. DeWeert, "Intensified CCD (ICCD) dynamic range and noise performance," in *Ultrahigh- and High-Speed Photography and Image-based Motion Measurement*, San Diego, 1997.
- [139] Andor, "An introduction to the ICCD camera," [Online]. Available: https://web.archive.org/web/20210309142555if_/https://andor.oxinst.com/learning/view/article/intensified-ccd-cameras.

- [140] L. van Geest and L. Bosch, "Gated Image Intensifiers And Applications," *Proceedings of SPIE*, vol. 1072, no. 1, pp. 19-29, 1989.
- [141] S. H. Courtney and W. L. Wilson, "Multichannel time-correlated single photon counting: Spectroscopy and time-gated imaging using a resistive anode photomultiplier tube," *Review of Scientific Instruments*, vol. 62, no. 9, pp. 2100-2104, 1991.
- [142] W. G. McMullan, S. Charbonneau and M. L. W. Thewalt, "Simultaneous sub nanosecond timing information and 2D spatial information from imaging photomultiplier tubes," *Review of Scientific Instruments*, vol. 58, no. 9, pp. 1626-1628, 1987.
- [143] D. Rees, I. McWhirter, P. A. Rounce, F. E. Barlow and S. J. Kellock, "Miniature imaging photon detectors," *Journal of Physics E: Scientific Instruments*, vol. 13, no. 7, pp. 763-770, 1980.
- [144] M. C. Chang, S. H. Courtney, A. J. Cross, R. J. Gulotty, J. W. Petrich and G. R. Fleming, "Time-Correlated Single Photon Counting With Microchannel Plate Detectors," *Instrumentation Science & Technology*, vol. 14, no. 3-4, pp. 433-464, 1985.
- [145] N. Teranishi, A. Kohono, Y. Ishihara, E. Oda and K. Arai, "No Image Lag Photodiode Structure in the Interline CCD Image Sensor," in *IEEE International Electron Devices Meeting*, 1982.
- [146] E. Fossum, "Active pixel sensors: are CCDs dinosaurs?," *Proceedings of SPIE*, vol. 1900, pp. 2-14, 1993.
- [147] M. W. Seo, S. Kawahito, K. Kagawa and K. Yasutomi, "A 0.27e-rms Read Noise 220- μ V/e-Conversion Gain Reset-Gate-Less CMOS Image Sensor With 0.11- μ m CIS Process," *IEEE Electron Device Letters*, vol. 36, no. 12, pp. 1344-1347, 2015.
- [148] N. Teranishi, "Required Conditions for Photon-Counting Image Sensors," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2199-2205, 2012.
- [149] T. Spirig, P. Seitz, O. Vietze and F. Heitger, "The lock-in CCD - Two-Dimensional Synchronous Detection of Light," *IEEE Journal of Quantum Electronics*, vol. 31, no. 9, pp. 1705-1708, 1995.
- [150] R. Lange, P. Seitz, A. Biber and S. C. Lauxtermann, "Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging," *Proceedings of SPIE*, vol. 3965, no. 1, pp. 177-188, 2000.
- [151] S. Kawahito, I. Abdul Halin, T. Ushinaga, T. Sawada, M. Homma and Y. Maeda, "A CMOS Time-of-Flight Range Image Sensor With Gates-on-Field-Oxide Structure," *IEEE Sensors Journal*, vol. 7, no. 12, pp. 1578-1586, 2007.
- [152] Z. Li, S. Kawahito, K. Yasutomi, K. Kagawa, J. Ukon, M. Hashimoto and H. Niioka, "A Time-Resolved CMOS Image Sensor With Draining-Only Modulation Pixels for Fluorescence Lifetime Imaging," *IEEE Transactions on Electron Devices*, vol. 59, no. 10, pp. 2715-2722, 2012.
- [153] M. W. Seo, K. Kagawa, K. Yasutomi, T. Takasawa, Y. Kawata, N. Teranishi, Z. Li, I. Abdul Halin and S. Kawahito, "A 10.8ps-Time-Resolution 256 \times 512 Image Sensor with 2-Tap True-CDS Lock-In Pixels for

- Fluorescence Lifetime Imaging," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2015.
- [154] B. Pain, T. Cunningham, B. Hancock, C. Wrigley and C. Sun, "Excess Noise and Dark Current Mechanisms in CMOS Imagers," in *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, 2005.
- [155] J. Y. Kim, S. I. Hwang, J. J. Lee, J. H. Ko, Y. Kim, J. C. Ahn, T. Asaba and Y. H. Lee, "Characterization and Improvement of Random Noise in 1/3.2" UXGA CMOS Image Sensor with 2.8um Pixel using 0.13um-Technology," in *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, 2005.
- [156] Y. Chen, X. Wang, A. J. Mierop and A. J. P. Theuwissen, "A CMOS Image Sensor With In-Pixel Buried-Channel Source Follower and Optimized Row Selector," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2390 - 2397, 2009.
- [157] H. Takahashi, T. Noda, T. Matsuda, T. Watanabe, M. Shinohara, T. Endo, S. Takimoto, R. Mishima, S. Nishimura, K. Sakurai, H. Yuzurihara and S. Inoue, "A 1/2.7 inch Low-Noise CMOS Image Sensor for Full HD Camcorders," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2007.
- [158] C. Lotto, P. Seitz and T. Baechler, "A Sub-Electron Readout Noise CMOS Image Sensor with Pixel-Level Open-Loop Voltage Amplification," in *International Solid-State Circuits Conference*, San Francisco, 2011.
- [159] S. Kawahito, S. Suh, T. Shirei, S. Itoh and S. Aoyama, "Noise Reduction Effects of Column-Parallel Correlated Multiple Sampling and Source-Follower Driving Current Switching for CMOS Image Sensors," in *International Image Sensor Workshop*, Bergen, 2009.
- [160] Y. Chen, Y. Xu, Y. Chae, A. Mierop, X. Wang and A. Theuwsen, "A 0.7e-rms-Temporal-Readout-Noise CMOS Image Sensor for Low-Light-Level Imaging," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2012.
- [161] S. Wakashima, F. Kusuhara, R. Kuroda and S. Sugawa, "A Linear Response Single Exposure CMOS Image Sensor with 0.5e- Readout Noise and 76ke- Full Well Capacity," in *Symposium on VLSI Circuits*, Kyoto, 2015.
- [162] A. Boukhayma, A. Peizerat and C. Enz, "A Sub-0.5 Electron Read Noise VGA Image Sensor in a Standard CMOS Process," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 9, pp. 2180-2191, 2016.
- [163] S. Suh, S. Itoh, S. Aoyama and S. Kawahito, "Column-Parallel Correlated Multiple Sampling Circuits for CMOS Image Sensors and Their Noise Reduction Effects," *Sensors*, vol. 10, no. 10, pp. 9139-9154, 2010.
- [164] A. Boukhayma, A. Peizerat and C. Enz, "A Correlated Multiple Sampling Passive Switched Capacitor Circuit for Low Light CMOS Image Sensors," in *International Conference on Noise and Fluctuations*, Xian, 2015.

- [165] Y. Lim, K. Koh, K. Kim, H. Yang, J. Kim, Y. Jeong, S. Lee, H. Lee, S. H. Lim, Y. Han, J. Kim, J. Yun, S. Ham and Y. T. Lee, "A 1.1e- temporal noise 1/3.2-inch 8Mpixel CMOS image sensor using pseudo-multiple sampling," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2010.
- [166] N. Kawai and S. Kawahito, "Noise Analysis of High-Gain, Low-Noise Column Readout Circuits for CMOS Image Sensors," *IEEE Transactions on Electron Devices*, vol. 51, no. 2, pp. 185-194, 2004.
- [167] S. Cova, A. Longoni and G. Ripamonti, "Active-Quenching and Gating Circuits for Single-Photon Avalanche Diodes (SPADs)," *IEEE Transactions on Nuclear Science*, vol. 29, no. 1, pp. 599-601, 1982.
- [168] R. H. Haitz, "Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes," *Journal of Applied Physics*, vol. 36, no. 10, pp. 3123-3131, 1965.
- [169] S. Cova, A. Longoni and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Review of Scientific Instruments*, vol. 52, no. 3, pp. 408-412, 1981.
- [170] M. Ghioni, S. Cova, A. Lacaita and G. Ripamonti, "New Silicon Epitaxial Avalanche Diode for Single-Photon Timing at Room Temperature," *Electronics Letters*, vol. 24, no. 24, pp. 1476-1477, 1988.
- [171] A. Rochas, A. R. Pauchard, P. A. Besse, D. Pantic, Z. Pijic and R. S. Popovic, "Low-Noise Silicon Avalanche Photodiodes Fabricated in Conventional CMOS Technologies," *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 387-394, 2002.
- [172] A. Rochas, M. Gani, B. Furrer, P.-A. Besse, R. Popovic, G. Ribordy and N. Gisin, "Single photon detector fabricated in a complementary metal oxide semiconductor high-voltage technology," *Review of Scientific Instruments*, vol. 74, no. 7, pp. 3263-3270, 2003.
- [173] E. Charbon, M. Fishburn, R. Walker, R. K. Henderson and C. Niclass, "SPAD-Based Sensors," in *TOF Range-Imaging*, Berlin, Springer, 2013, pp. 11-38.
- [174] C. Niclass, M. Gersbach, R. Henderson, L. Grant and E. Charbon, "A single photon avalanche diode implemented in 130-nm CMOS technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 863-869, 2007.
- [175] B. F. Aull, A. H. Loomis, D. J. Young, R. M. Heinrichs, B. J. Felton, P. J. Daniels and D. J. Landers, "Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging," *Lincoln Laboratory Journal*, vol. 13, no. 2, pp. 335-350, 2002.
- [176] A. Lacaita, M. Ghioni and S. Cova, "Double epitaxy improves single-photon avalanche diode performance," *Electronics Letters*, vol. 25, no. 13, pp. 841-843, 1989.
- [177] W. Becker, *The bh TCSPC Handbook*, Berlin: Becker & Hickl GmbH, 2019.
- [178] X. Michalet, A. Cheng, J. Antelman, M. Suyama, K. Arisaka and S. Weiss, "Hybrid photodetector for single-molecule spectroscopy and microscopy," *Proceedings of SPIE*, vol. 6862, no. 1, p. 68620F, 2008.

-
- [179] M. Gersbach, Y. Maruyama, R. Trimananda, M. W. Fishburn, D. Stoppa, J. A. Richardson, R. Walker, R. Henderson and E. Charbon, "A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 6, pp. 1394-1407, 2012.
- [180] C. Niclass, C. Favi, T. Kluter, M. Gersbach and E. Charbon, "A 128×128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2977-2989, 2008.
- [181] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini and D. Stoppa, "A 160 × 120 Pixel Analog-Counting Single-Photon Imager With Time-Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 155-167, 2016.
- [182] M. Gersbach, J. Richardson, E. Mazaleyrat, S. Hardillier, C. Niclass, R. Henderson, L. Grant and E. Charbon, "A low-noise single-photon detector implemented in a 130 nm CMOS imaging process," *Solid-State Electronics*, vol. 53, no. 7, pp. 803-808, 2009.
- [183] C. Niclass, C. Favi, T. Kluter, F. Monnier and E. Charbon, "Single-Photon Synchronous Detection," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 1977-1989, 2009.
- [184] H. A. R. Homulle, F. Powolny, P. L. Stegehuis, J. Dijkstra, D. U. Li, K. Homicsko, D. Rimoldi, K. Muehlethaler, J. O. Prior, R. Sinisi, E. Dubikovskaya, E. Charbon and C. Bruschini, "Compact solid-state CMOS single-photon detector array for in vivo NIR fluorescence lifetime oncology measurements," *Biomedical Optics Express*, vol. 7, no. 5, p. 1797–1814, 2016.
- [185] X. F. Wang, T. Uchida, D. M. Coleman and S. Minami, "A Two-Dimensional Fluorescence Lifetime Imaging System Using a Gated Image Intensifier," *Applied Spectroscopy*, vol. 45, no. 3, pp. 360-366, 1991.
- [186] E. P. Buurman, R. Sanders, A. Draaijer, H. C. Gerritsen, J. J. F. van Veen, P. M. Houpt and Y. K. Levine, "Fluorescence Lifetime Imaging Using a Confocal Laser Scanning Microscope," *Scanning*, vol. 14, no. 3, pp. 155-159, 1992.
- [187] S. P. Chan, Z. J. Fuller, J. N. Demas and B. A. DeGraff, "Optimized Gating Scheme for Rapid Lifetime Determinations of Single-Exponential Luminescence Lifetimes," *Analytical Chemistry*, vol. 73, no. 18, pp. 4486-4490, 2001.
- [188] C. J. de Grauw and H. C. Gerritsen, "Multiple Time-Gate Module for Fluorescence Lifetime Imaging," *Applied Spectroscopy*, vol. 55, no. 6, pp. 670-678, 2001.
- [189] A. V. Agronskaia, L. Tertoolen and H. C. Gerritsen, "High frame rate fluorescence lifetime imaging," *Journal of Physics D: Applied Physics*, vol. 36, no. 14, pp. 1655-1662, 2003.

- [190] H. Sparks, S. Warren, J. Guedes, N. Yoshida, T. C. Charn, N. Guerra, T. Tatla, C. Dunsby and P. French, "A flexible wide-field FLIM endoscope utilising blue excitation light for label-free contrast of tissue," *Journal of Biophotonics*, vol. 8, no. 1-2, pp. 168-178, 2015.
- [191] R. J. Woods, S. Scypinski, L. J. C. Love and H. A. Ashworth, "Transient Digitizer for the Determination of Microsecond Luminescence Lifetimes," *Analytical Chemistry*, vol. 56, no. 8, pp. 1395-1400, 1984.
- [192] R. K. P. Benninger, O. Hofmann, J. McGinty, J. Requejo-Isidro, I. Munro, M. A. A. Neil, A. J. deMello and P. M. W. French, "Time-resolved fluorescence imaging of solvent interactions in microfluidic devices," *Optics Express*, vol. 13, no. 16, pp. 6275-6285, 2005.
- [193] D. S. Elson, I. Munro, J. Requejo-Isidro, J. McGinty, C. Dunsby, N. Galletly, G. W. Stamp, M. A. A. Neil, M. J. Lever, P. A. Kellett, A. Dymoke-Bradshaw, J. Hares and P. M. W. French, "Real-time time-domain fluorescence lifetime imaging including single-shot acquisition with a segmented optical image intensifier," *New Journal of Physics*, vol. 6, no. 1, p. 180, 2004.
- [194] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, New York: McGraw-Hill, 1969.
- [195] J. McGinty, N. P. Galletly, C. Dunsby, I. Munro, D. S. Elson, J. Requejo-Isidro, P. Cohen, R. Ahmad, A. Forsyth, A. V. Tyillainayagam, M. A. A. Neil, P. M. W. French and G. W. Stamp, "Wide-field fluorescence lifetime imaging of cancer," *Biomedical Optics Express*, vol. 1, no. 2, pp. 627-640, 2010.
- [196] R. Cubeddu, D. Comelli, C. D'Andrea, P. Taroni and G. Valentini, "Time-resolved fluorescence imaging in biology and medicine," *Journal of Physics D: Applied Physics*, vol. 35, no. 9, pp. R61-R76, 2002.
- [197] M. Köllner and J. Wolfrum, "How many photons are necessary for fluorescence-lifetime measurements?," *Chemical Physics Letters*, vol. 200, no. 1,2, pp. 199-204, 1992.
- [198] P. J. Verveer, A. Squire and P. I. H. Bastiaens, "Global Analysis of Fluorescence Lifetime Imaging Microscopy Data," *Biophysical Journal*, vol. 78, no. 4, pp. 2127-2137, 2000.
- [199] S. Pelet, M. J. R. Previte, L. H. Laiho and P. T. C. So, "A Fast Global Fitting Algorithm for Fluorescence Lifetime Imaging Microscopy Based on Image Segmentation," *Biophysical Journal*, vol. 87, no. 4, pp. 2807-2817, 2004.
- [200] H. C. Gerritsen, M. A. H. Asselbergs, A. V. Agronskaia and W. G. J. H. M. van Sark, "Fluorescence lifetime imaging in scanning microscopes: acquisition speed, photon economy and lifetime resolution," *Journal of Microscopy*, vol. 206, no. 3, pp. 218-224, 2002.
- [201] D. U. Li, E. Bonnist, D. Renshaw and R. Henderson, "On-chip, time-correlated, fluorescence lifetime extraction algorithms and error analysis," *Journal of the Optical Society of America A*, vol. 25, no. 5, pp. 1190-1198, 2008.
- [202] I. M. Antolovic, C. Bruschini and E. Charbon, "Dynamic range extension for photon counting arrays," *Optics Express*, vol. 26, no. 17, pp. 22234-22248, 2018.

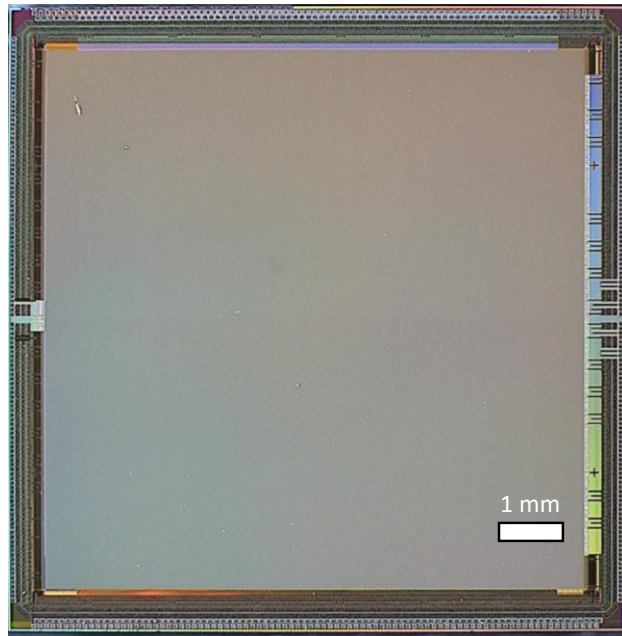
-
- [203] I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe and E. Charbon, "Nonuniformity analysis of a 65-kpixel CMOS SPAD imager," *IEEE Transactions on Electron Devices*, vol. 63, no. 1, pp. 57-64, 2016.
- [204] E. Fossum, J. Ma, S. Masoodian, L. Anzagira and R. Zizza, "The Quanta Image Sensor: Every Photon Counts," *Sensors*, vol. 16, no. 8, p. 1260, 2016.
- [205] A. L. Lacaita, F. Zappa, S. Bigliardi and M. Manfredi, "On the Bremsstrahlung Origin of Hot-Carrier-Induced Photons in Silicon Devices," *IEEE Transactions on Electron Devices*, vol. 40, no. 3, pp. 577-582, 1993.
- [206] I. Rech, A. Ingargiola, R. Spinelli, I. Labanca, S. Marangoni, M. Ghioni and S. Cova, "Optical crosstalk in single photon avalanche diode arrays: a new complete model," *Optics Express*, vol. 16, no. 12, pp. 8381-8394, 2008.
- [207] D. Kavaldjiev and Z. Ninkov, "Influence of nonuniform charge-coupled device pixel response on aperture photometry," *Optical Engineering*, vol. 40, no. 2, pp. 162-169, 2001.
- [208] M. Estribeau and P. Magnan, "CMOS pixels crosstalk mapping and its influence on measurements accuracy for space applications," *Proceedings of SPIE*, vol. 5978, no. 1, p. 597813, 2005.
- [209] D. Kavaldjiev and Z. Ninkov, "Subpixel sensitivity map for a charge-coupled device sensor," *Optical Engineering*, vol. 37, no. 3, pp. 948-954, 1998.
- [210] M. Monici, "Cell and tissue autofluorescence research and diagnostic applications," *Biotechnology Annual Review*, vol. 11, no. 1, pp. 227-256, 2005.
- [211] R. Ankri, A. Basu, A. C. Ulku, C. Bruschini, E. Charbon, S. Weiss and X. Michalet, "Single-Photon, Time-Gated, Phasor-Based Fluorescence Lifetime Imaging through Highly Scattering Medium," *ACS Photonics*, vol. 7, no. 1, pp. 68-79, 2020.
- [212] C. Veerappan and E. Charbon, "A Low Dark Count p-i-n Diode Based SPAD in CMOS Technology," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 65-71, 2016.
- [213] C. Veerappan and E. Charbon, "A Substrate Isolated CMOS SPAD Enabling Wide Spectral Response and Low Electrical Crosstalk," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, p. 3801507, 2014.
- [214] R. K. Henderson, J. Richardson and L. Grant, "Reduction of Band-to-band Tunneling in Deep-submicron CMOS Single Photon Avalanche Photodiodes," in *International Image Sensor Workshop*, Bergen, 2009.
- [215] P. Heydari and M. Pedram, "Capacitive coupling noise in high-speed VLSI circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 3, pp. 478-488, 2005.

- [216] S. Lindner, S. Pellegrini, Y. Henrion, B. Rae, M. Wolf and E. Charbon, "A High-PDE, Backside-Illuminated SPAD in 65/40-nm 3D IC CMOS Pixel With Cascoded Passive Quenching and Active Recharge," *IEEE Electron Device Letters*, vol. 38, no. 11, pp. 1547-1550, 2017.
- [217] S. Ma, S. Gupta, A. C. Ulku, C. Bruschini, E. Charbon and M. Gupta, "Quanta Burst Photography," *ACM Transactions on Graphics*, vol. 39, no. 4, p. 79, 2020.
- [218] I. M. Antolovic, "SPAD imagers for super resolution microscopy," Delft University of Technology, Delft, Netherlands, 2018. PhD Thesis. <https://doi.org/10.4233/uuid:cc76e95c-b82e-4555-9110-348ad9989705>.
- [219] P. Magnan, "Detection of visible photons in CCD and CMOS:A comparative view," *Nuclear Instruments and Methods in Physics Research A*, vol. 504, no. 1-3, pp. 199-212, 2003.
- [220] F. Acerbi, G. Paternoster, A. Gola, N. Zorzi and C. Piemonte, "Silicon photomultipliers and single-photon avalanche diodes with enhanced NIR detection efficiency at FBK," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 912, pp. 309-314, 2018.
- [221] P. W. R. Connolly, X. Ren, A. McCarthy, H. Mai, F. Villa, A. J. Waddie, M. R. Taghizadeh, A. Tosi, F. Zappa, R. K. Henderson and G. S. Buller, "High concentration factor diffractive microlenses integrated with CMOS single-photon avalanche diode detector arrays for fill-factor improvement," *Applied Optics*, vol. 59, no. 14, pp. 4488-4498, 2020.
- [222] M. Marisaldi, P. Maccagnani, F. Moscatelli, C. Labanti, F. Fuschino, M. Prest, A. Berra, D. Bolognini, M. Ghioni, I. Rech, A. Gulinatti, A. Giudice, G. Simmerle, D. Rubini, A. Candelori and S. Mattiazzo, "Single Photon Avalanche Diodes for Space Applications," in *IEEE Nuclear Science Symposium*, 2011.
- [223] L. Carrara, C. Niclass, N. Scheidegger, H. Shea and E. Charbon, "A gamma, X-ray and high energy proton radiation-tolerant CIS for space applications," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2009.
- [224] E. Charbon, L. Carrara, C. Niclass, N. Scheidegger and H. Shea, "Radiation-tolerant CMOS single-photon imagers for multiradiation detection," in *Radiation effects in semiconductors*, Boca Raton, CRC Press, 2011, pp. 31-50.
- [225] L. Ratti, P. Brogi, G. Collazuol, G. F. Dalla Betta, A. Ficarella, L. Lodola, P. S. Marrocchesi, S. Mattiazzo, F. Morsani, M. Musacci, L. Pancheri and C. Vacchi, "Dark Count Rate Degradation in CMOS SPADs Exposed to X-Rays and Neutrons," *IEEE Transactions on Nuclear Science*, vol. 66, no. 2, pp. 567-574, 2019.
- [226] E. Anisimova, B. L. Higgins, J. P. Bourgoin, M. Cranmer, E. Choi, D. Hudson, L. P. Piche, A. Scott, V. Makarov and T. Jennewein, "Mitigating radiation damage of single photon detectors for space applications," *EPJ Quantum Technology*, vol. 4, no. 10, pp. 1-14, 2017.

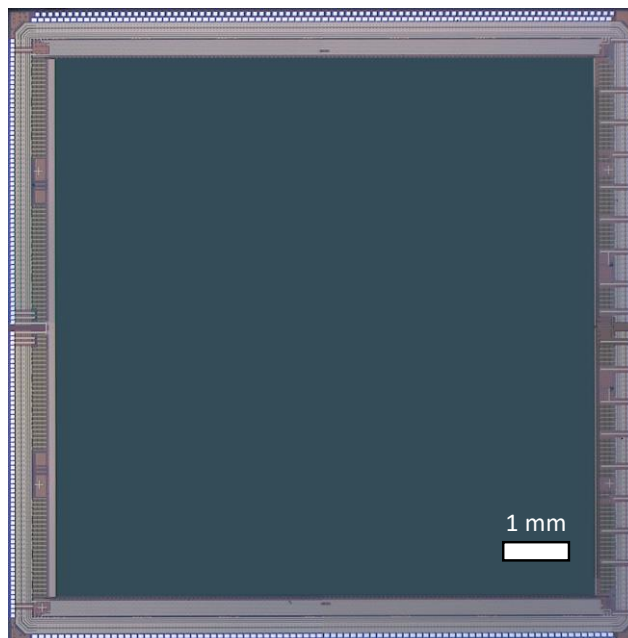
-
- [227] W. J. Kindt, H. W. van Zeijl and S. Middelhoek, "Optical cross talk in Geiger mode avalanche photodiode arrays: modeling, prevention and measurement," in *European Solid-State Device Research Conference*, Bordeaux, 1998.
 - [228] S. H. Chan, O. A. Elgendy and X. Wang, "Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors," *MDPI Sensors*, vol. 16, no. 11, p. 1961, 2016.
 - [229] P. Chandramouli, S. Burri, C. Bruschini, E. Charbon and A. Kolb, "A Bit Too Much? High Speed Imaging from Sparse Photon Counts," in *IEEE International Conference on Computational Photography*, Tokyo, 2019.
 - [230] L. Sbaiz, F. Yang, E. Charbon, S. Susstrunk and M. Vetterli, "The Gigavision Camera," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, 2009.
 - [231] J. M. Rabaey, A. Chandrakasan and B. Nikolic, *Digital Integrated Circuits: A Design Perspective* (2nd Edition), Upper Saddle River: Prentice Hall, 2003.
 - [232] S. Donati, G. Martini and M. Norgia, "Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits," *Optics Express*, vol. 15, no. 26, pp. 18066-18075, 2007.
 - [233] I. M. Antolovic, A. C. Ulku, E. Kizilkan, S. Lindner, F. Zanella, R. Ferrini, M. Schnieper, E. Charbon and C. Bruschini, "Optical-stack optimization for improved SPAD photon detection efficiency," *Proceedings of SPIE*, vol. 10926, p. 109262T, 2019.
 - [234] I. Gyongy, A. Davies, B. Gallinet, N. A. W. Dutton, R. R. Duncan, C. Rickman, R. K. Henderson and P. A. Dalgarno, "Cylindrical microlensing for enhanced collection efficiency of small pixel SPAD arrays in single-molecule localisation microscopy," *Optics Express*, vol. 26, no. 3, pp. 2280-2291, 2018.
 - [235] L. Pancheri, N. Massari and D. Stoppa, "SPAD Image Sensor With Analog Counting Pixel for Time-Resolved Fluorescence Detection," *IEEE Transactions on Electron Devices*, vol. 60, no. 10, pp. 3442-3449, 2013.
 - [236] E. R. Fossum, "Modeling the Performance of Single-Bit and Multi-Bit Quanta Image Sensors," *IEEE Journal of the Electron Devices Society*, vol. 1, no. 9, pp. 166-174, 2013.
 - [237] M. Wahl, T. Röhlicke, S. Kulisch, S. Rohilla, B. Krämer and A. C. Hocke, "Photon arrival time tagging with many channels, sub-nanosecond deadtime, very high throughput, and fiber optic remote synchronization," *Review of Scientific Instruments*, vol. 91, no. 1, p. 013108, 2020.
 - [238] W. Becker, A. Bergmann, G. Biscotti, K. Koenig, I. Riemann, L. Kelbauskas and C. Biskup, "High-Speed FLIM Data Acquisition by Time-Correlated Single Photon Counting," *Proceedings of SPIE*, vol. 5323, no. 1, pp. 27-35, 2004.
 - [239] X. Michalet, O. H. W. Siegmund, J. V. Vallerga, P. Jelinsky, J. E. Millaud and S. Weiss, "Photon-counting H33D detector for biological fluorescence imaging," *Nuclear Instruments and Methods in*

- Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 567, no. 1, pp. 133-136, 2006.
- [240] Y. Kuo, J. Li, X. Michalet, A. Chizhik, N. Meir, O. Bar-Elli, E. Chan, D. Oron, J. Enderlein and S. Weiss, "Characterizing the Quantum-Confined Stark Effect in Semiconductor Quantum Dots and Nanorods for Single-Molecule Electrophysiology," *ACS Photonics*, vol. 5, no. 12, pp. 4788-4800, 2018.
- [241] M. W. Seo, K. Kagawa, K. Yasutomi, Y. Kawata, N. Teranishi, Z. Li, I. A. Halin and S. Kawahito, "A 10 ps Time-Resolution CMOS Image Sensor With Two-Tap True-CDS Lock-In Pixels for Fluorescence Lifetime Imaging," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 141-154, 2016.
- [242] W. L. Rice, D. M. Shcherbakova, V. V. Verkhusha and A. T. N. Kumar, "In Vivo Tomographic Imaging of Deep-Seated Cancer Using Fluorescence Lifetime Contrast," *Cancer Research*, vol. 75, no. 7, pp. 1236-1243, 2015.
- [243] D. Fixler, Y. Namer, Y. Yishay and M. Deutsch, "Influence of Fluorescence Anisotropy on Fluorescence Intensity and Lifetime Measurement: Theory, Simulations and Experiments," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1141-1152, 2006.

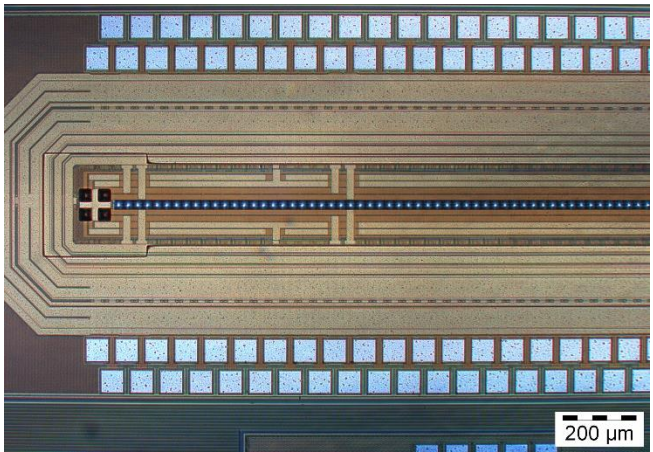
Chip gallery



SwissSPAD2 (Chapter 3)



SwissSPAD3 (Chapter 3)



LinoSPAD2 (not presented in the thesis)

List of publications

Journal Articles as First Author

- 1) **A. Ulku***, A. Ardelean*, M. Antolovic, S. Weiss, E. Charbon, C. Bruschini and X. Michalet, "Wide-field time-gated SPAD imager for phasor-based FLIM applications," *Methods and Applications in Fluorescence*, vol. 8, no. 2, p. 024002, 2020. (*Equally contributing authors)
- 2) **A. C. Ulku**, C. Bruschini, I. M. Antolovic, Y. Kuo, R. Ankri, S. Weiss, X. Michalet and E. Charbon, "A 512×512 SPAD image sensor with integrated gating for widefield FLIM," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 1, p. 6801212, 2019.

Journal Articles as Co-Author

- 3) S. Ma, S. Gupta, **A. C. Ulku**, C. Bruschini, E. Charbon and M. Gupta, "Quanta burst photography," *ACM Transactions on Graphics*, vol. 39, no. 4, p. 79, 2020.
- 4) K. Morimoto, A. Ardelean, M. L. Wu, **A. C. Ulku**, I. M. Antolovic, C. Bruschini and E. Charbon, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica*, vol. 7, no. 4, pp. 346-354, 2020.
- 5) R. Ankri, A. Basu, **A. C. Ulku**, C. Bruschini, E. Charbon, S. Weiss and X. Michalet, "Single-photon, time-gated, phasor-based fluorescence lifetime imaging through highly scattering medium," *ACS Photonics*, vol. 7, no. 1, pp. 68-79, 2020.
- 6) A. Carimatto, **A. Ulku**, S. Lindner, E. Gros-Daillon, B. Rae, S. Pellegrini and E. Charbon, "Multipurpose, fully integrated 128 × 128 event-driven MD-SiPM with 512 16-Bit TDCs with 45-ps LSB and 20-ns gating in 40-nm CMOS technology," *IEEE Solid-State Circuits Letters*, vol. 1, no. 12, pp. 241-244, 2018.

Conference Articles as First Author

- 7) **A. C. Ulku**, C. Bruschini, I. M. Antolovic, S. Weiss, X. Michalet and E. Charbon, "Phasor-based widefield FLIM using a gated 512×512 single-photon SPAD imager," *Proceedings of SPIE*, vol. 10882, p. 108820M, 2019.
- 8) **A. C. Ulku**, C. Bruschini, X. Michalet, S. Weiss and E. Charbon, "A 512×512 SPAD image sensor with built-in gating for phasor based real-time siFLIM," in *International Image Sensor Workshop*, Hiroshima, 2017.

Conference Articles as Co-Author

- 9) I. M. Antolovic, **A. C. Ulku**, E. Kizilkan, S. Lindner, F. Zanella, R. Ferrini, M. Schnieper, E. Charbon and C. Bruschini, "Optical-stack optimization for improved SPAD photon detection efficiency," *Proceedings of SPIE*, vol. 10926, p. 109262T, 2019.
- 10) A. Ardelean, **A. C. Ulku**, X. Michalet, E. Charbon and C. Bruschini, "Fluorescence lifetime imaging with a single-photon SPAD array using long overlapping gates: an experimental and theoretical study," *Proceedings of SPIE*, vol. 10882, p. 108820Y, 2019.
- 11) A. Carimatto, **A. Ulku**, S. Lindner, E. D'Aillon, B. Rae, S. Pellegrini and E. Charbon, "Multipurpose, fully-integrated 128×128 event-driven MD-SiPM with 512 16-bit TDCs with 45 ps LSB and 20 ns gating," in *Symposium on VLSI Circuits*, Honolulu, 2018.

Conference Presentations/Posters as First Author

- 12) **A. C. Ulku***, A. Ardelean*, P. Mos, E. Charbon and C. Bruschini, "SwissSPAD3 – a dual-gate photon-counting SPAD sensor for widefield FLIM imaging (Conference presentation)," in *Focus on Microscopy*, 2021. (*Equally contributing authors)
- 13) **A. Ulku**, A. Ardelean, M. Antolovic, S. Weiss, E. Charbon, C. Bruschini and X. Michalet, "Wide-field time-gated SPAD imager for phasor-based FLIM applications (Poster)," in *Biophysical Society Annual Meeting*, San Diego, 2020.
- 14) **A. C. Ulku**, C. Bruschini, S. Weiss, X. Michalet and E. Charbon, "A time-gated large-array SPAD camera for picosecond resolution real-time FLIM (Conference presentation)," *Proceedings of SPIE*, vol. 10498, p. 104980M, 2018.

Curriculum vitae

Arin Can Ulku

Education

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Ph.D. in Microsystems and Microelectronics

October 2016 – Present

Delft University of Technology, Delft, Netherlands

M.Sc. in Electrical Engineering, Microelectronics Track

September 2014 – August 2016

Sabanci University, Istanbul, Turkey

B.Sc. in Electronics Engineering

September 2010 – June 2014

Research Experience

Advanced Quantum Architecture Laboratory, EPFL

Ph.D. Thesis: Large-format time-gated SPAD cameras for real-time phasor-based FLIM

October 2016 – Present

Delft University of Technology Circuits and Systems Group

M.Sc. Thesis: A single-photon, time-resolved imaging sensor for low-light-level vision

August 2015 – August 2016

Sabanci University Radio Frequency (RF) Circuit Design Group

B.Sc. Thesis: RF integrated circuits for X-band (8-12 GHz) phased-array radar

September 2013 – June 2014