SPAD imagers for super resolution microscopy

Antolovic, Michel

DOI
10.4233/uuid:cc76e95c-b82e-4555-9110-348ad9989705

Publication date
2018

Document Version
Final published version

Citation (APA)
Antolovic, M. (2018). *SPAD imagers for super resolution microscopy*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:cc76e95c-b82e-4555-9110-348ad9989705

SPAD imagers for super resolution microscopy

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T. H. J. J. van der Hagen;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 23 januari 2018 om 10.00 uur

door

Ivan Michel ANTOLOVIĆ
Magistar inženjer elektrotehnike i informacijske tehnologije,
Sveučilište u Zagrebu,
uit Kroatië

This dissertation has been approved by the promotor and copromotor:
promotor: Prof. dr. ir. E. Charbon
copromotor: Dr. R. A. Hoebe

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof. dr. ir. E. Charbon	Delft University of Technology
Dr. R. A. Hoebe	University of Amsterdam

Independent members:

Dr. T. Anhut	Carl Zeiss Microscopy GmbH
Prof. dr. R. Henderson	The University of Edinburgh
Dr. S. Stallinga	Delft University of Technology
Prof. dr. ir. A. J. P. Theuwissen	Delft University of Technology
Prof. dr. H. P. Urbach	Delft University of Technology

Contents

Nomenclature

ADC	Analog to digital converter
APD	Avalanche photodiode
CAI	Charge accumulating imager
CF	Concentration factor
CMOS	complementary metal oxide semiconductor
DCR	Dark count rate
DSNU	Dark signal nonuniformity
DSP	Digital signal processing
dSTORM	Direct stochastic optical reconstruction microscopy
DTI	Deep trench isolation
E(·)	Expectation value
ECR	Excess count rate
EMCCD	Electron multiplying charge coupled device
FCS	Fluorescence correlation spectroscopy
FLIM	Fluorescence lifetime imaging microscopy
FPGA	Field programmable gate array
fps	Frames per second
FWHM	Full-width at half-maximum
GSDIM	Ground-state depletion and single-molecule return
HDR	High dynamic range
IC	Integrated circuits
ICG	Indocyanine green
LET	Light emission test
NA	Numerical aperture
PDE	Photon detection efficiency
PDP	Photon detection probability

PMT	Photo multiplier tube
PRNU	Photon response nonuniformity
PSF	Point spread function
QIS	Quanta image sensor
sCMOS	Scientific complementary metal oxide semiconductor
SEM	Scanning electron microscope
SIM	Structured illumination microscopy
SMLM	Single molecule localization microscopy
SNR	Signal-to-noise ratio
SPAD	Single-photon avalanche diodes
STED	Stimulated emission depletion microscopy
STI	Shallow trench isolation
TDC	Time to digital converter

1 Introduction

Magnifying very distant or very small objects is a fundamental research method. Star movement was observed and atoms were theorized before modern science could prove them, as it was very often influenced by mythopoetic imagination. During early modern science, magnifying lenses were built in telescopes and microscopes to develop astronomy and make micrometer objects visible.

Theories about magnifying lenses were introduced in the Middle Ages, and the first telescopes and microscopes were built in the 17th century. For a long time, it was believed that microscopes could reach any desired magnification, and thus any spatial resolution, just by improving the lens quality. It was later found that the resolution of optical microscopes is fundamentally limited by diffraction.

1.1 Spatial resolution of light microscopes

Lenses are used to magnify and focus light. Early geometrical optics theorized that we can reconstruct an image of a nanometer spot with the same size as the initial spot. However, light diffracts traveling away from the nanometer spot. The light widens and eventually hits a target area larger than the initial spot. The imaged light intensity forms wave shapes called Airy disks (Figure 1.1). The point spread function (PSF) describes this behavior of light mathematically. The light intensity over one lateral cross-section x is defined by:

$$I(x) = I_0 \left(\frac{2J_1(\pi a x / \lambda z)}{\pi a x / \lambda z} \right)^2, \tag{1.1}$$

where $J_1(x)$ is the first order Bessel function, I_0 the peak intensity, λ the light wavelength, a the lens (aperture) size, and z the distance between the lens and the imaging plane. The first order Bessel function is similar to a *sinc* function. A rectangular aperture would, in contrast to circular aperture, yield a lateral cross-section with a *sinc* function.

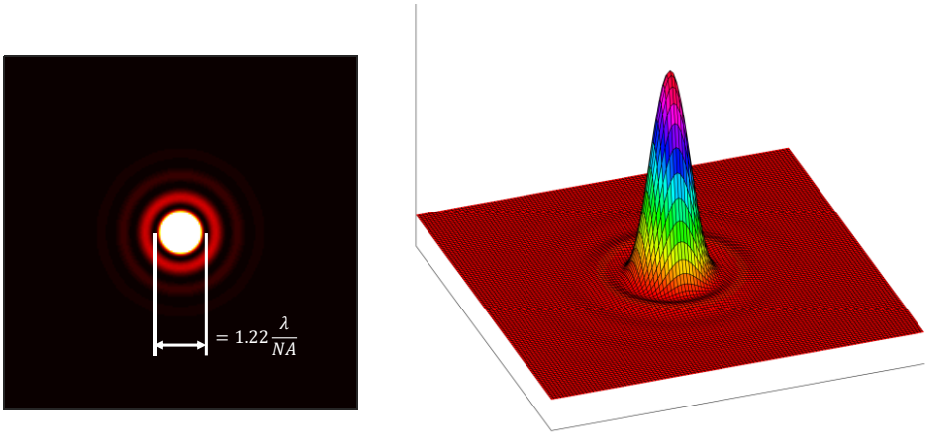


Figure 1.1. Airy disk or PSF of a circular aperture lens system.

The size of a PSF (the first circle in Figure 1.1, high intensity) is proportional to $\lambda z/a$. Z is the focal length and a the lens diameter. We then replace a and z with the numerical aperture of the lens: NA, where NA=$a/2z$. This yields a PSF size of 1.22 λ/NA, as shown in Figure 1.1. The numerical aperture has a limited range, with a usual maximum of 1.4. Thus, it is not possible to reduce the PSF far beyond the λ.

Light diffraction thus limits the smallest resolvable distance of, for example, two nanometer spots. Two spots that are closer than this resolvable distance will be seen as a single blurred spot, as shown in Figure 1.2c. Abbe defined the optical resolution as the largest spatial frequency that contributes to the image, and extracted it to be $\Delta x = \lambda/2NA$[1]. Rayleigh defined a more intuitive optical resolution[2]; $\Delta x = 0.61 \times \lambda/NA$, arguing that two spots are spatially resolvable if the peak of one PSF is positioned at the first dark ring of the adjacent PSF. This is shown in Figure 1.2. Figure 1.2b shows the Rayleigh limit, where the two spots are still separable.

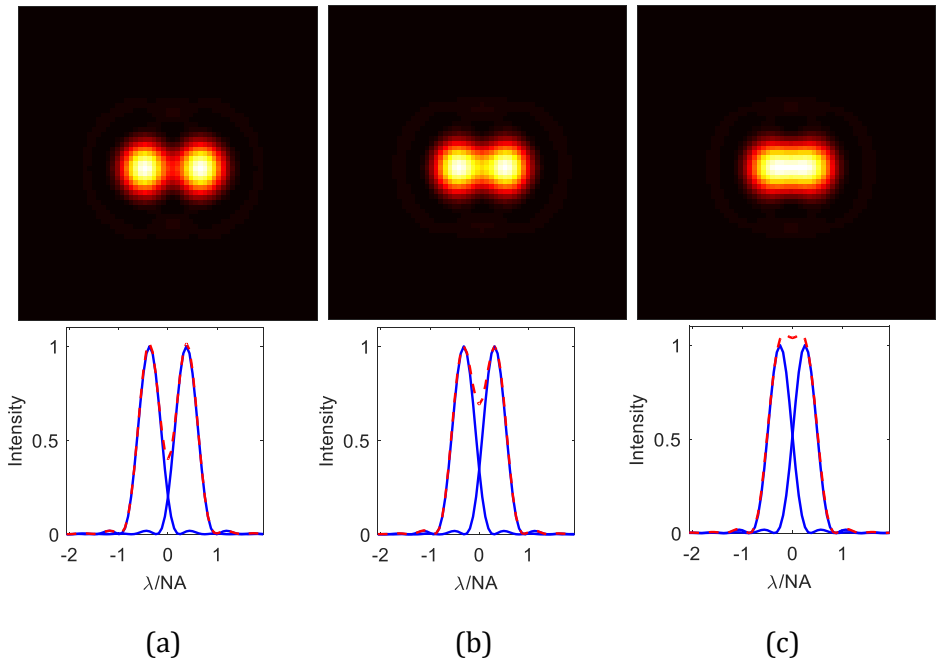


Figure 1.2. Rayleigh criterion for the minimal resolvable resolution. (a) Shows clearly resolved spots, where (b) is at the Rayleigh limit. (c) Shows two merged spots that cannot be resolved.

The same resolution limitation applies when imaging more general structures. An example is shown in Figure 1.3. We cannot distinguish two line structures that are closer than the resolution from a single line structure. Both the single and double line will appear as a single line. In microscopy, biologists usually color cell structures with bright fluorescent dyes to label specific structure of interest or to increase the contrast. This technique is called fluorescence microscopy, because we label structures with bright dyes, fluorophores. Fluorophores are single molecules that can target specific molecules, proteins and other structures of interest. If a microscope detects an isolated fluorophore, it will be shown as a single PSF, a diffraction limited spot.

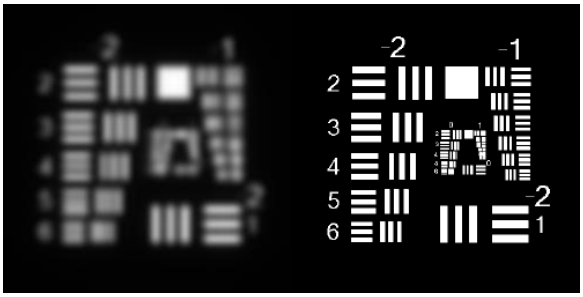


Figure 1.3. An example of a resolution limited observation, compared to an original structure.

Several techniques were developed to overcome the diffraction limitation and achieve a resolution higher then Abbe's/Rayleigh's resolution[3]. This quest has led to the Nobel Prize in Chemistry in 2014. One type of techniques is single molecule localization microscopy (SMLM), which separates the fluorophores in time by means of blinking. A small number of fluorophores will be active on average, ensuring that single fluorophore PSFs are separated. The position of each fluorophores can be then estimated with sub-diffraction resolution by localizing the peak intensity. By subsequently imaging and localizing sparse fluorophores, a super resolved pointillistic image is formed[4]. This thesis uses one particular SMLM technique, known as ground-state depletion and single-molecule return (GSDIM[4], also known as direct stochastic optical reconstruction microscopy; dSTORM[5]).

The final resolution of SMLM is determined by the localization uncertainty[6] and the labeling density[7], and reaches as low as 10 nm, while the diffraction limited resolution is limited to around 200 nm. The uncertainty decreases with the increase of the number of captured photons. More gathered statistics increase the localization accuracy. Thompson *et al.* found that the uncertainty specifically decreases with the square root of the number of detected photons[8].

Thus, early SMLM microscopes featured highly sensitive electron-multiplying charge-coupled device (EMCCD) imagers. Their high sensitivity enhanced the uniformity of the PSFs and increased the localization accuracy. Because of limited speed and the presence of excess noise in EMCCDs, a number of researchers started to use faster scientific complementary metal oxide semiconductor (sCMOS) imagers, eventually enabling faster image formation[9]. For this, faster fluorophores were required[9,10], and faster instrumentation became critical for a fluorophore's photophysical characterization and

optimization. To date, the maximal reported frame rate available to SMLM is still limited to 3 000 frames per second (fps)[10].

However, since tens of thousands of images are needed to create a pointillistic super resolved image, the limited imaging speed slows the acquisition speed of super resolution images to multiple seconds and even minutes.

1.2 SPAD technology

To increase the imaging speed we need to introduce operational and architectural changes in detection and accumulation of photo-electrons. Photons could be counted by means of a direct photon to digital transformation, as opposed to accumulating photo-electrons in a fragile environment with thermal and other analog noise sources, as it is done in EMCCD and sCMOS imagers. This would increase the robustness and allow for higher imaging speeds.

There exist a number of devices that perform photon to digital transformations, e.g. photo-multiplying tubes (PMT) and single-photon avalanche diodes (SPAD). PMTs are large sensors that are challenging to integrate in large arrays due to their bulky mechanical structure requiring vacuum to operate correctly. SPADs can be successfully integrated in large arrays since they are very similar to standard microelectronic chips fabricated in complementary metal-oxide semiconductor (CMOS) technology. In fact, SPADs can themselves be fabricated in standard CMOS and CMOS image sensor (CIS) technologies.

A SPAD is a photodiode that is capable of triggering electrical avalanches upon detection of a photon. The generic structure of a SPAD is shown in Figure 1.4. Photons are absorbed (generating an electron-hole pair) by the semiconductor lattice (silicon or other semiconductors). The electric field of the depletion dipole generates current if the photons are absorbed close to the depletion layer.

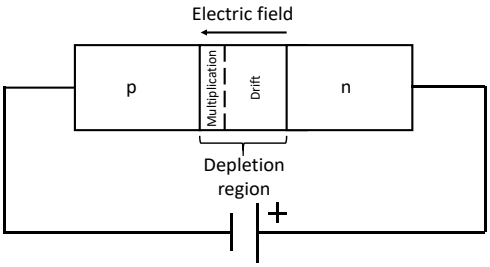


Figure 1.4. Reverse-biased photodiode. We create a depletion region at the p-n junction by applying a reverse bias. At high reverse bias, a multiplication region is formed closer to the more doped layer (in this case the p layer), where carriers are accelerated (and multiplied) to create an avalanche. Photons absorbed in the drift region are creating avalanches after drifting towards the multiplication region.

A high electric field creates current avalanches if SPADs are operated above breakdown voltage V_{BD}, at the operating voltage $V_{OP}=V_{BD}+V_E$ (V_E is the excess bias). Above breakdown, electron-hole pairs are abruptly accelerated and generate impact ionization. The SPADs then operate in so-called Geiger mode, where a detected photo-electron generates an avalanche. In passive quenching regime, the avalanche current is quenched by a ballast resistor. The same resistor is used to sense the current by generating a fast voltage pulse, thus we achieve a photon-to-digital converter. A SPAD usually comprises a junction between a highly doped p+ layer and an n well, shown in Figure 1.5.

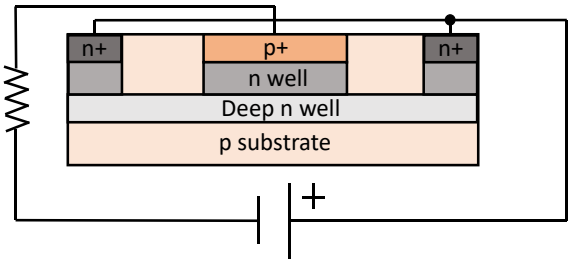


Figure 1.5. A typical implementation of a SPAD in a standard CMOS process, with a ballast resistor in series. The depletion region is in this case formed between the p+ and n well layers. The side and bottom n layers are used for connectivity with the n well.

SPAD pulses usually have a rising edge of a few nanoseconds and a jitter of 30 to 100 ps; the pulse width is between 10 and 100 ns. This implies that at most one photon can be detected every 10 to 100 ns. The maximum imaging speed in these cases is thus 400 kfps (or 40 kfps in case of 100 ns dead time) for a single measurement with 255 different gray values (from black to white) assuming accumulation of subsequent pulses is used to form the gray levels.

SPAD imagers are arrays of SPADs[11] that enable high timing resolution and are thus suitable for time-resolved and high-speed applications[12]. For example, a distance can be measured implicitly by the time taken by light to travel from a laser source to an object and back to the detector. Inherently, SPAD imagers have a high photon response uniformity[13,14] and zero readout noise.

However, the early digitalization of photon counts within the pixel[14] comes at the cost of a somewhat lower fill factor and a lower photon detection probability (PDP). Noise expressed in terms of dark count rate (DCR) is still high in many CMOS SPAD implementations. Single photon outputting and high imaging speed induce extremely high data rates. Major design challenges remain before we can build a SPAD imager exceeding the performance of conventional image sensors.

1.3 Main research question

This brings us to the research question, which is the focus of this thesis:

What are the advantages of SPAD imagers overcoming the main challenges, used in microscopy, compared to conventional EMCCD and sCMOS imager technologies?

1.4 Outline of the thesis

The thesis covers large format SPAD imagers, but it also discusses small format SPAD imagers. A graphical outline is found in Figure 1.6.

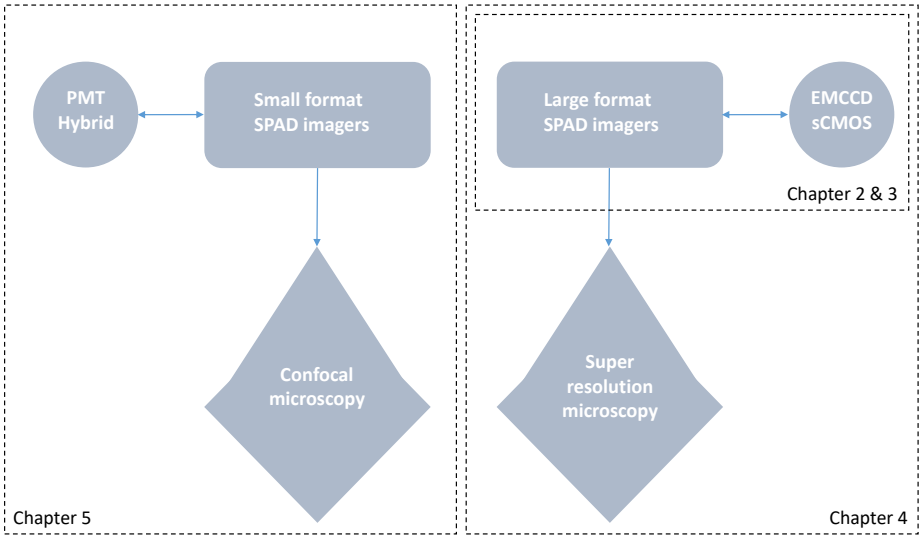


Figure 1.6. Graphical outline of the thesis.

Single photon imagers are introduced in chapter 2 in terms of their basic performance like PDP, DCR, signal-to-noise ratio and dynamic range. The chapter includes large format SPAD, EMCCD and sCMOS imagers. Architecture differences yield significant changes in key performance. We discuss challenges and potential advantages of different state-of-the-art implementations. Chapter 3 plots design challenges with an emphasis on scalability.

Imager differences play a large role in image quality generated by the microscope. In chapter 4, we quantify image quality from different imagers. We show both benefits and drawbacks of using SPAD imagers in super resolution microscopy, and discuss how SPAD advantages can be employed in microscopy. How should we design SPAD imagers to maximize this advantages?

Before concluding, a small format SPAD imager used for confocal microscopy is described and characterized in chapter 5. We present an application where the SPAD design is less challenging, but can lead to major improvements in image quality and detector cost.

1.5 Contributions

The contributions of this thesis are:

1) Modeling of large SPAD image sensors

Signal-to-noise ratio was looked at in the context of dynamic range. An active clock (recharge asynchronous to the SPAD activity) response was experimentally found to deviate from the active event-driven (recharge synchronous to the SPAD activity) response. We found mathematical models and confirmed them with simulation and measurement data (sections 2.2.4, 2.2.5 and 5.1.4). We proposed a quantitative SNR decrease of -3dB as a measure for the extended dynamic range. Active event-driven recharge theoretically extends the dynamic range by ×75 compared to $1/T_{dead}$ (section 2.2.5.1). Active clock recharge extends it by ×2.84.

Nonuniformity analysis was conducted, focusing on fundamental principles. We found that the photon-to-digital transformation offers photon response nonuniformity below 2%. Digital robustness mitigates analog noise sources found in charge accumulating imagers. However, DCR nonuniformity is a main concern for SPAD imagers. Finally, we compared SPAD with EMCCD and sCMOS imagers. Results are found if sections 2.3.2 and 2.4.

Optimum frame time was analyzed. We derived the optimum frame time for localization super resolution microscopy, based on exponentially distributed fluorophore blinking. The optimum frame time is longer than the average blinking time, and depends on the emission and background intensity (section 4.1.1).

2) First super resolution images with a SPAD imager

With consecutive publications in 2015[15], 2016[16] and 2017[17], we showed the first localization super resolution microscopy results acquired with a SPAD imager, gradually improving image quality throughout the publications (section 4.2).

Moreover, based on a previous SPAD implementation with 45% peak PDP, we designed a 55% PDP SPAD array with 100 cps DCR, 0.1% afterpulsing and 120 ps timing jitter to explore the limits of advanced SPADs for super resolution applications. Using a cascode transistor combination for quenching and recharge, the array operates at high 11 V excess bias. Capacitive isolation allowed us to reduce the afterpulsing to 0.1%. Detailed results are found in section 5.1.

3) Widefield exploration of fluorophore blinking

In addition to super resolution images, we were able to perform first widefield blinking analysis with timing resolution in the µs range (section 4.3).

4) Methodological/design contributions
   - A simplified measurement procedure that gives a concentration factor for the microlenses relative to an f-number of 1.8
   - Quantified recommendations on imager architecture to yield a maximum data reduction (section 3.2)
   - Co-design (pixel array and event-driven bus) of two event-driven SPAD imagers with 144×252 and 32×32 pixel resolution

2 Single photon imagers

As mentioned in the introduction, SPADs offer the possibility to combine large format integration (large number of pixels) with high-speed imaging and/or high-timing resolution. In this chapter, we include imager alternatives and compare them. We discuss how far other technologies like electron multiplying charge-coupled device (EMCCD) and scientific CMOS (sCMOS) are from implementing a practical single photon imager, and whether SPAD imagers have a competitive advantage.

Section 2.1 discusses the general requirements for single photon imaging, while section 2.2 presents key parameters for large format imagers with an emphasis on the current drawbacks of each imager technology. The three main imager technologies are shown in Figure 2.1.

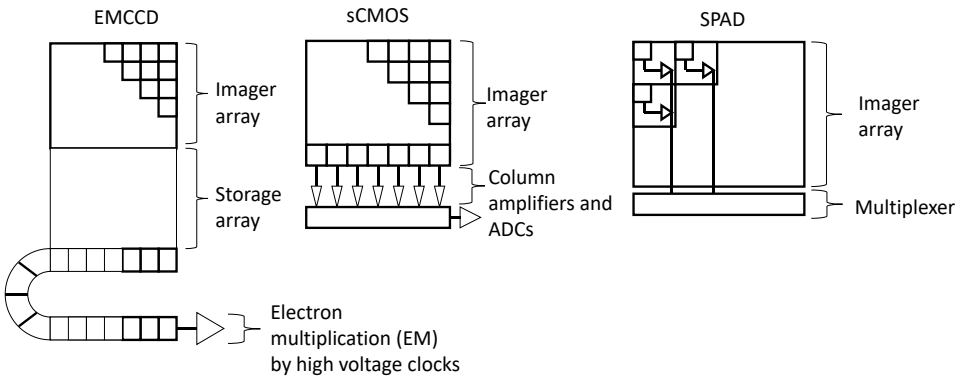


Figure 2.1. The three main photon-counting imager technologies. EMCCDs are highly sensitive and uniform imagers. Their readout speed is low due to the serial readout architecture. sCMOS imagers parallelize the output and increase speed. Although, they are not photon counters per se, sCMOS imagers have reached a de facto photon-counting status due to their low noise that makes it possible to distinguish single-photon events with a certain accuracy. SPAD imagers offer direct photon-to-digital transformation. Operating in the digital domain, these imagers enable DSP like operations operating directly on pixels and at a very high speed.

We track the rapid evolution from single SPADs through SPAD arrays to SPAD imagers in section 2.3. We present different types of SPAD imagers and concentrate on a particular implementation – SwissSPAD.

Other imager technologies have also been steadily improving in performance, due to rapid technology advancement. Thus, Section 2.4 compares high-end imagers in different technologies with the aim of differentiating between

fundamental and technological drawbacks. Technological drawbacks, in contrast to fundamental ones, can be and will be overcome, in time.

2.1 Requirements for single photon and photon counting imaging

Let us begin with the importance of single-photon imaging. Single-photon detection is required for measurements on time distribution, e.g. when measuring lifetime of a first molecular excited state, or any distribution related to a time distribution, like in 3D imaging (distance distribution). Single photon detection is also needed for measurements including quantum photon correlations.

There is a fundamental difference between sensitivity and single-photon detectability. While sensitivity relates the number of incident photons with the number of generated electrons (as in the case of modern detectors using the photoelectric effect), *single-photon detection* requires that one is able to *read out the electron generated by a single photon*. The main constrain preventing single-photon detection is readout noise.

In practice, single electrons are charge amplified to measurable values (defined with a conversion gain V/e^-). The main objective of a single-photon detector is to have a low effective (input referred) readout noise. Higher readout noise yields a higher probability of false detection P_{fd}[18]:

$$P_{fd} = \frac{1}{2} erfc\left[\frac{1}{\sqrt{8}\sigma_{read}}\right], \tag{2.1}$$

where σ_{read} is the readout noise. For example, a readout noise of 0.5 e^- would yield a probability of a P_{fd} of 16%, whereas 0.3 e^- yields a P_{fd} of 5%.

Single-photon imagers rely on either highly amplifying the electron charge (as to mitigate the effect of readout noise) or reducing the readout noise below 0.3 e^-. EMCCD imagers use electron multiplication whereas sCMOS imagers are trying to reach a readout noise lower than 0.3 e^-. SPAD imagers on the contrary, rely on impact ionization (with theoretically infinite electron amplification) to generate pulses in the order of few volts.

Let us consider the case in which we want to evaluate quantum correlation; we assume a cross correlation measurement of a single-photon emitter which photons are detected by a small array. The single-photon emitter will release at most one photon in a fraction of time (frame time set to be same as laser

period), and the cross correlation with lag time 0, $G^2(T=0)$, should be 0. Either no detection occurred, the first pixel detected, or the second pixel detected a single photon. The two pixels cannot detect a photon at the same time. At a large lag time, the cross correlation between two pixels $G^2(T=\Delta t)$ will be P_{det}^2, where P_{det} is the probability of detecting a photon. If we however have a $P_{fd}\neq 0$, $G^2(T=0)= P_{fd} \times P_{det}$. The larger the $G^2(T=\Delta t)$ with respect to $G^2(T=0)$, the better the measurement SNR. Since P_{det} is usually in the order of 0.01, a cross correlation measurement puts a strong requirement on P_{fd} and the readout noise.

Photon-counting enforces a stronger limitation on the CMOS imager readout noise σ_{read}. The readout noise spreads the output voltage for one electron to a Gaussian distribution. Photon counting thus needs to take into account both the voltages below 0.5 e^- and above 1.5 e^-. P_{fd} thus doubles for photon counting[19], compared to single-photon detection (checks if there is no photon or one/multiple photons). For SPADs, dead time limits photon counting at high photon counting rates. If two photons arrive within the dead time, a SPAD will count just one. Limitations enforced by the dead time will be discussed in section 2.2.4.

2.2 Key parameters of single-photon imagers

Section 2.2 describes the key parameters of single photon imagers. The introduced parameters are related to underlying principles. For example, SPADs usually have peak sensitivity at 450 nm wavelength because of a shallow p+ to n well junction. The underlying principles will facilitate the state-of-the-art comparison and the differentiation between fundamental and technological limitations of different single photon imagers.

2.2.1 Quantum efficiency, photon detection probability and fill factor

Sensitivity of a photodetector is the ratio between the input photon number and the number of generated carriers (in case of charge accumulating imagers like EMCCD and sCMOS) or pulses (in case of SPADs) that are detected in a given area and time. Usually, sensitivity in charge accumulating imagers (CIA) is given with quantum efficiency (QE). Equation (2.2)[20] gives the QE for a given wavelength for a junction from 0 to depth z_{end}.

$$QE(\lambda) = \int_{z_{end}}^{0} T(\lambda)\mu(\lambda)e^{-z\mu(\lambda)}p_d(z)dz \tag{2.2}$$

$T(\lambda)$ is the transmittance spectra including interference effects stemming from reflections from the metals, inter metal dielectrics and passivation present on top (and sides) of the junction[21]. $\mu(\lambda)$ is the absorption coefficient, $\mu(\lambda)e^{-z\,\mu(\lambda)}$ the probability of being absorbed at depth z and p_d the probability the generated carrier is detected in the depletion region. Note that $p_d(z)=1$ if z is in the depletion region, while it monotonically approaches 0 both at $z=0$ and $z=z_{end}$, as to account for diffusion towards the depletion region.

In SPADs, each carrier that is detected in the depletion region might not necessarily generate an avalanche pulse. We thus denote SPAD sensitivity with photon detection probability (PDP), as opposed to QE. The avalanche probability (a carrier generating an avalanche) is governed by the ionization coefficient. The PDP equation (2.3) is similar to (2.3), whereas p_d is replaced by p_a.

$$PDP(\lambda) = \int_{z_{end}}^{0} T(\lambda)\mu(\lambda)e^{-z\mu(\lambda)} p_a(z)dz \tag{2.3}$$

p_a is the probability a carrier generated at depth z generates an avalanche. It includes p_d, and can be approximated as $p_a=C\times p_d$, and thus $PDP=C\times QE$. C is here the avalanche probability related to the ionization coefficient that increases with the applied electric field (when increasing the operating voltage), but eventually saturates. SPAD PDP thus saturates after increasing the operating voltage VOP.

A higher sensitivity implies a higher p_a over a wider (deeper) semiconductor volume. Mixing of carriers because of different absorption depth[22] increases timing jitter. However, a wider depletion region reduces the depletion capacitance and the diffusion region resistance[23], leading to improved SPAD timing performance. A higher PDP due to a wider depletion region should thus have no major negative effects of the timing.

In addition to sensitivity, the ratio between the sensitive and total pixel area, or fill factor, is not necessarily 100%. EMCCDs usually have a fill factor of 100%, whereby FSI CMOS transistors limit it to around 50%[21]. SPADs, in addition to restricting the fill factor with transistors, feature a guard ring of 1 to 2 μm. The guard ring prevents a lateral breakdown. The overall sensitivity including the fill factor is denoted as photon detection efficiency (PDE).

2.2.2 Noise sources

Noise sources in imaging sensors are offsets in photon signal that change over time and are unpredictable per integration time *t*. A time-invariant offset is not regarded as noise source and can be easily subtracted. Usually though noise appears as a combination of time-invariant and time-varying offsets and it is often non-trivial to separate the two components to remove time-invariant noise.

2.2.2.1 Shot noise

Shot noise is the inevitable noise source in every particle based measurement, also called Poisson noise; $\sigma_{shot\ noise}=\sqrt{n_i}$, where n_i is the number of detected photons.

2.2.2.2 Dark noise and dark count rate

Dark noise is the generation of carriers without incident light. In SPADs, dark noise is characterized as the mean dark counts per second or dark count rate (DCR). There exist a discrepancy between the dark noise for CAIs expressed in dark current and for SPADs expressed in DCR. A carrier might not generate an avalanche, like discussed in section 2.2.1. DCR is an average offset added to the photon count rate. As we defined noise, DCR change in time should be taken as noise, i.e. $\sigma_{DCR}=\sqrt{(\mathrm{DCR}\times t)}$, where t is the integration time. The main sources of dark noise are trap-assisted thermal generation, band-to-band tunneling and trap-assisted tunneling. All are found to increase with electric field[24] (that is high in SPADs). Tunneling is practically not present in CAIs with usual electric fields[25], whereas in SPADs tunneling increases the DCR substantially[26]. Both tunneling and trap-assisted noise reduction in SPADs were demonstrated with additional techniques[27,28].

2.2.2.3 Clock induced charge noise

The EMCCD technology introduces uncorrelated noise by charge transfer clocking. Carriers are generated close to the surface interface between Si and SiO_2. This effect is usually found to be lower than 1% per transfer clock cycle[29]. In cases of electron multiplication, it can give rise to ionization enhanced clock induced charges. Clock induced charge noise is reduced when increasing the clock frequency. However, a faster clock frequency reduces the charge transfer efficiency and lowers the effective full well capacitance[30].

2.2.2.4 Multiplication noise

Multiplication noise is an EMCCD feature, due to variable gain in the multiplication. The multiplication noise factor γ asymptotically reaches 1.4 for high gain values[31]. This factor is then used to multiply all the noise contributions that are present before electron multiplication. Readout noise is added after electron multiplication.

2.2.2.5 Readout noise

Readout noise is due to shot, thermal, and 1/f noise of charge amplifiers and output amplifiers. If high bandwidth is needed, thermal noise from the source follower acting as the charge amplifier is the major readout noise source both in EMCCD and sCMOS[21]. A major difference in readout emerges due to architectural differences between EMCCD and sCMOS. EMCCD has a higher readout noise because of the serial readout that requires higher bandwidth, where all pixels share the same charge amplifier. Although EMCCDs have a higher readout noise, the electron multiplication can effectively reduce the readout noise by a factor of κ, the multiplication gain. sCMOS implements charge amplification within the pixel and column parallel readout, reducing the bandwidth and the readout noise. If the bandwidth can be reduced, the dominant readout noise source in sCMOS image sensors is the in-pixel source follower 1/f noise[32], proportional to:

$$\sigma_{1/f}^2 \propto \frac{K(C_P + (C_{gse} + C_{gde})W + \frac{2}{3}C_{ox}WL)^2}{C_{ox}^2 WL}, \tag{2.4}$$

where K is a process related parameter bound to 1/f noise, C_P the parasitic capacitance at the source follower gate, W and L the source follower sizes, C_{gse} and C_{gde} the gate to drain/source capacitance per width, and C_{ox} the gate oxide capacitance per area. The 1/f noise can be reduced by reducing the source follower size or increasing C_{ox}. Boukhayma *et al.* used a thin oxide (to increase C_{ox}) PMOS transistor to decrease 1/f noise[32,33]. PMOS was chosen due to a low K factor. This implementation led to an input referred readout noise lower than 0.5 e^-.

However, per-pixel source followers (and column amplifiers) lead to a nonuniform distribution of readout noise over the array. Important to note is that the readout noise changes while changing the frame rate due to increased thermal noise bandwidth. A higher frame rate implies a higher readout noise.

SPADs do include similar noise sources, however they are irrelevant because the pulse amplitude is usually 10 000 higher than the readout noise (1 V with respect to around 100 μV).

2.2.2.6 Afterpulsing and crosstalk

Afterpulsing and crosstalk are secondary pulses or electrons that occur after photon carriers or noise generated carriers. Afterpulsing is practically not present in CAIs. SPADs have afterpulsing values below 10%[34] and vary greatly depending on conditions such as dead time and excess bias.

Crosstalk is the measure of secondary pulses from adjacent pixels, and can be divided into optical and electrical crosstalk. Electrical crosstalk occurs when secondary carriers diffuse to adjacent pixels. The lower the doping concentration, the higher the diffusion length and crosstalk probability. This effect can be enhanced in SPADs due to the large amount of avalanche-generated carriers. Optical crosstalk occurs due to impact ionization based photon emission. Photons detected by adjacent pixels lead to optical crosstalk. Larger amounts of avalanche carriers increase the optical crosstalk probability.

Afterpulsing and crosstalk can be measured by gathering information on inter-arrival times between counts in the same pixel in the case of afterpulsing, and between counts of two adjacent pixels in the case of crosstalk[14,35,36]. If these inter-arrival times show a distribution different from a single-exponential, the sensor features either afterpulsing and/or crosstalk. Researchers also use pulsed lasers to evaluate avalanche probabilities in between two laser pulses that are not stemming from the laser[37].

Authors usually report crosstalk that occurs at higher incident angles, when photons cross the optical stack at one pixel but then generate a carrier at the adjacent pixel, or close to the adjacent pixel and then the carrier diffuses to the adjacent pixel[38]. Crosstalk is in this case not the result of secondary carriers; it reduces the primary signal of one pixel and increases it for the adjacent one. We do not regard this as true crosstalk. The final effect of this kind of crosstalk is degradation in space resolution.

2.2.3 Timing

In CMOS CAIs, timing is limited by the minimum accumulation time, usually between 1 and 10 μs[39]. CAI designers can implement additional time windowing. In one particular implementation, researchers used time

windowing of 125 ns with simulations showing pinned photodiode to pinned storage diode transfer in 0.18 ns[40]. They showed lifetime (of a first molecular excited state) extraction down to 2 ns. In burst mode (with less than 1000 frames in total), CMOS CAIs reach more than 16 Mfps[41], and project to be able to achieve up to 1 Gfps[42].

SPADs create fast rising pulses used as precise time references. The rising edge deviation in time is referred to as timing jitter. Full-width at half-maximum is used as a measure for the timing jitter. SPADs have a timing jitter ranging between 30 and 200 ps[43–45].

SPADs with larger active areas have a larger timing jitter due to a larger depletion volume. Deeper and wider depletion regions increase the electron carriers mixing time due to different absorption depth and thus the timing jitter[22]. Because of this effect, triggering the SPAD with lasers with a shorter wavelength generally decreases timing jitter. However, a wider depletion region reduces the depletion capacitance[23] and thus requires less carriers to reach the timing threshold. This effect can counter effect the mixing effect. Ghioni *et al*. achieved a low timing jitter below 50 ps for a large area SPAD by using a low threshold at the SPAD output[46].

2.2.4 Count suppression in SPADs

Let the SPAD be exposed to a photon count rate i and let it detect a count rate n, with $n=PDE\times i$. If we include the effects of DCR and afterpulsing, $n=(1+P_a)(PDE\times i+DCR)$, where P_a is the afterpulsing probability. It is however not possible to *measure* the same count rate m ($n\neq m$) exactly. If the SPAD creates a photon-carrier during an existing avalanche process or during the phase when the SPAD is biased below breakdown voltage, the generated photon-carrier is not detected. This time during which the SPAD cannot detect photons is called the dead time T_{dead}. Thus, m can be approximated as n only for low count rates. We should here note that this count suppression is tightly linked to the pile-up effect[47,48]. However, due to the fact that pile-up is majorly associated with timing measurements and the tendency to detect the first (fastest) photon within one laser clock, we used the term count suppression. Count suppression is in the context of this thesis evaluated in measurements using continuous light with Poisson distribution.

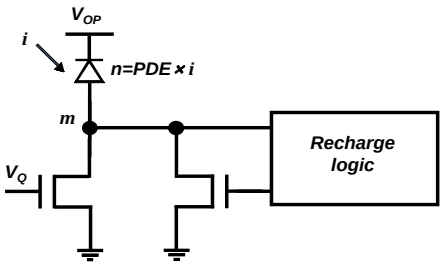


Figure 2.2. A SPAD with recharge logic.

When n is comparable with $1/T_{dead}$, then m will feature a nonlinear response to n.

2.2.4.1 SPAD imagers with active event-driven recharge (architecture synchronous to the SPAD activity)

Figure 2.3 shows the difference between synchronous event-driven and asynchronous clock recharge. If we recharge the SPAD synchronously with the detections, we implement the classical active event-driven recharge.

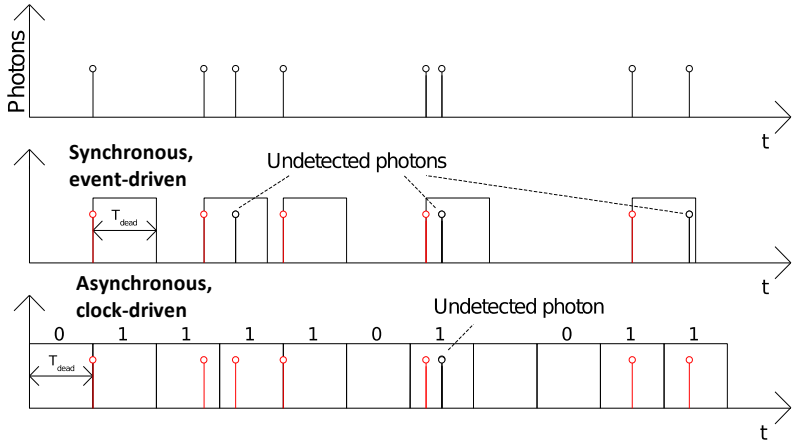


Figure 2.3. Example of event-driven recharge and clock-driven recharge.

The effect of the SPAD dead time with active event-driven recharge as a discrete diode sensor is well understood[49]. Dead time affects the measured count rate m of a sensor, as described by the following equation:

$$m = \frac{n}{1 + nT_{dead}} \tag{2.5}$$

The measured count rate m changes in SPAD imagers when there is no counter or accumulator implemented in each pixel. Another *dead time* is then present influencing the count rate m measured at the imager output. It is defined by the system dead time, i.e. the minimal readout time. In case of an event-driven readout architecture[50]:

$$m_{im} = \frac{m}{1 + mT_{readout}} \tag{2.6}$$

This model is only applicable if each SPAD has its own readout path. If $T_{readout} \gg T_{dead}$:

$$m_{im} \approx \frac{m}{1 + mT_{readout}} = \frac{n}{1 + nT_{readout}}, \tag{2.7}$$

using the annotation n as count rate detected by the SPAD. Eq. (2.7) was derived in the following way[51]: $m_{im} \times T_{readout}$ is the fraction where no information of counts can be read out and $n \times m_{im} \times T_{readout}$ the total rate where no counting occurs. The difference between the detected rate and the measured rate is then $n - m_{im} = n \times m_{im} \times T_{readout}$. If there is a shared readout line and all pixels have the same detection rate n, the rate where no counting occurs is $n \times m_{im} \times T_{readout} \times N$, where N is the number of pixels sharing one readout line. This leads to a different response:

$$m_{im} = \frac{n}{1 + NnT_{readout}} \tag{2.8}$$

Eq. (2.8) was confirmed with simulation results.

2.2.4.2 SPAD imagers with passive recharge

If we turn off the recharge logic of Figure 2.2 and use V_Q to bias the transistor around the threshold voltage, we obtain a synchronous passive recharge. After avalanche quenching, the excess bias voltage V_E (above the breakdown voltage V_{BD}) is gradually increased. A change in V_E implies a change in PDP. Thus, the PDP is very low at the start of recharge and recovers to a static value after T_{dead}. Although with low probability, the SPAD can fire right at the start of recharge, prolonging the effective dead time. It increases the effective dead time by a factor of e and yields a paralyzable response[49]. The measured count rate is then[49]:

$$m = ne^{-nT_{dead}} \tag{2.9}$$

2.2.4.3 SPAD imagers with clock recharge (architecture asynchronous to the SPAD activity)

SPADs can also be recharged asynchronously (clock based) with respect to the detections. In SPAD imagers, if the readout windows (or frames windows) are not synchronized with time-of-arrival of photons, the response curve will not follow (2.7). This is the case in most classical readout architectures, where all pixels are read out periodically (clock-driven architecture). However, $T_{readout}$ of a event-driven readout is usually shorter than $T_{readout}$ of a classical, clock-driven readout mechanism. Event-driven readout is also used in applications where n is different for each pixel, and most of the pixels have $n \approx 0$ in one $T_{readout}$.

In clock-driven architectures, imagers have an "exponential" ($1\text{-}e^{-x}$) response to light. This response can also be mathematically derived from the Poisson probability density function. It follows that the probability of detecting one or more photons in a pixel is $p_1(counts=1,2,\ldots,\infty)=1-e^{-\lambda}$, where λ is the expected number of counts, or detected photons, per pixel. If we have a detected count rate n in counts per second, then n would be λ within one second. Due to the imager architecture of SwissSPAD, we are interested here in the probability of detecting one or more photons per pixel, in one dead time T_{dead} or one readout time $T_{readout}$ (frame time). This probability would then be defined as:

$$p_{frame}(counts = 1,2,\ldots,\infty) = 1 - e^{-nT_{readout}} \tag{2.10}$$

The estimated measured count rate becomes:

$$E[m] = \frac{1 - e^{-nT_{readout}}}{T_{readout}} \tag{2.11}$$

where m is the measured count rate in *cps* at the output of our SPAD imager. Solving (2.11) with respect to n, we obtain:

$$E[n] = -\ln(1 - mT_{readout}) / T_{readout} \tag{2.12}$$

Eq. (2.12) shows the correction formula estimating n if only m is available. This correction becomes very important in super-resolution microscopy, for instance, where the response linearity influences the precision of the estimation of the point spread function (PSF). The exponential count loss was

indeed observed experimentally[14]. This signal response is similar to one observed by Sbaiz *et al.*[52] and Fossum in the quanta image sensor (QIS) approach[18], where more small binary pixels were used to build a larger virtual pixel with higher dynamic range. Figure 2.4 shows the response curves for the three different recharge cases. Note that the curves start to diverge from a linear response at around 10-20% of $1/T_{dead}$.

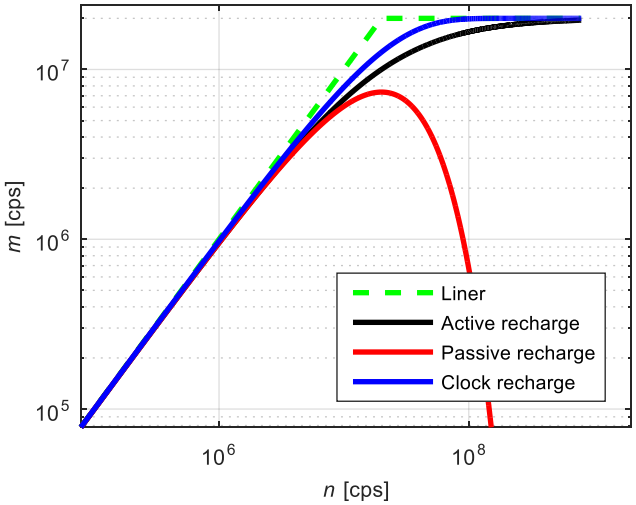


Figure 2.4. Modeled response curves for active event-driven (black), passive (red) and clock recharge (blue) with T_{dead}=50 ns, compared to a linear response that saturates at 20 Mcps.

2.2.4.4 Shot noise change due to count suppression

The SPAD saturates at a photon detection rate higher than $1/T_{dead}$ or $1/T_{readout}$ because of count suppression. Due to count suppression, the measured count number m_i=m×t, where t is the integration time, has a lower shot noise. The variance for active event-driven recharge is[53]:

$$\sigma_{m_i}^{\ 2} = \frac{nt}{\left(1+nT_{dead}\right)^3} + \left(1+\frac{2}{3}nT_{dead}+\frac{1}{6}\left(nT_{dead}\right)^2\right)\frac{\left(nT_{dead}\right)^2}{\left(1+nT_{dead}\right)^4}, \quad (2.13)$$

where n×t=n_i, the detected count number.

The variance for passive recharge is[53]:

$$\sigma_{m_i}^{\ 2} = mt(1-2mT_{dead}+\frac{mT_{dead}^2}{t}) \quad (2.14)$$

For the clock based recharge, the variance is given by the binominal distribution[18]:

$$\sigma_{m_i}{}^2 = \frac{t}{T_{dead}}(1-e^{-nT_{dead}})e^{-nT_{dead}} \tag{2.15}$$

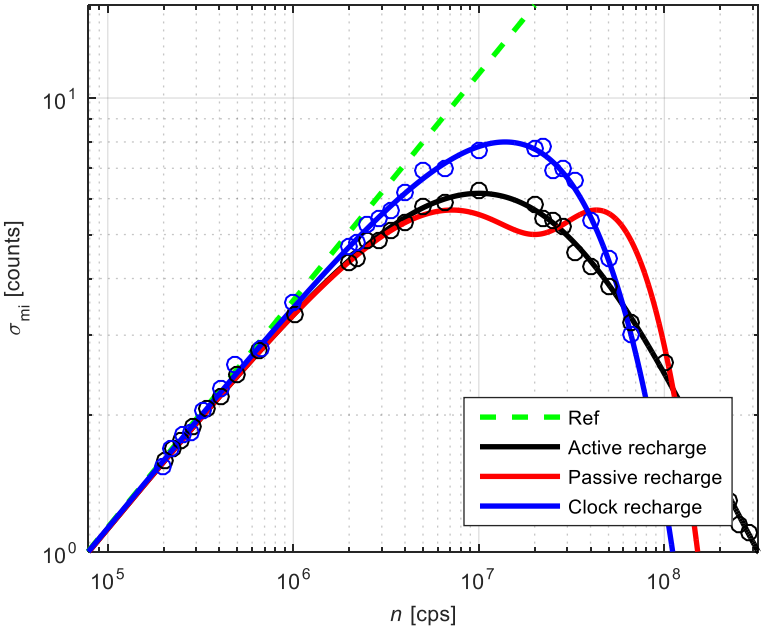


Figure 2.5. Modeled standard deviation of the measured count number σ_{mi} due to shot noise and count suppression, T_{dead}=50 ns and t=12.8 µs. Green shows the reference, i.e. standard deviation due to shot noise. Circles indicate simulated data.

Figure 2.5 shows σ_{mi} of the three different recharge mechanisms. σ_{mi} is a meaningless noise figure. The detection referred variance $\sigma_{ni}{}^2$ is the relevant noise figure, and it can be derived as:

$$\sigma_{n_i}^2 = \left(\sigma_{m_i}\frac{\partial n_i}{\partial m_i}\right)^2 \tag{2.16}$$

Figure 2.6 shows the increase of noise due to count suppression in SPAD detectors. Passive quenching has the largest noise increase with respect to the shot noise figure (green). Clock based recharge increases the noise before the $1/T_{dead}$ (20 Mcps) region less than the active event-driven recharge, but increases more at high detection rates.

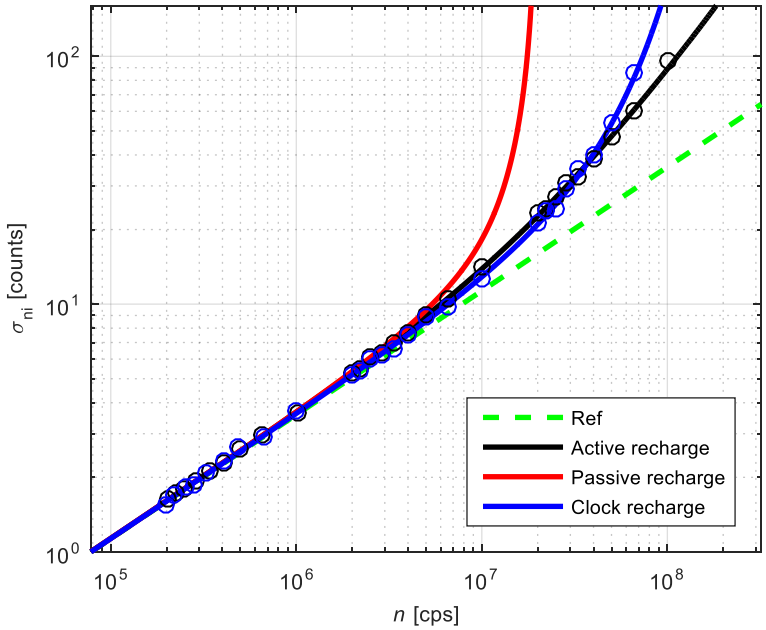


Figure 2.6. Modeled standard deviation of the detected count number σ_{ni} due to shot noise and count suppression, T_{dead}=50 ns and t=12.8 µs. Green shows the reference, i.e. standard deviation due to shot noise. Circles indicate simulated data.

2.2.5 Signal-to-noise ratio over dynamic range

SNR changes over different incident photon fluxes. A theoretical maximum SNR is governed by shot noise, i.e. $SNR_{max}=n_i/\sqrt{n_i}=\sqrt{n_i}$, where n_i is the number of detected photons during integration time t. Each of the imager technologies will have additional noise. For EMCCD, the SNR is going to be[54]:

$$SNR_{EMCCD} = \frac{n_i}{\sqrt{\gamma^2 \times (n_i + \sigma_{dark}^2 + \sigma_{CIC}^2) + \frac{\sigma_{read}^2}{\kappa^2}}}, \qquad (2.17)$$

where n_i is the number of detected photons, t the integration time, γ the multiplication noise factor, σ^2_{dark} the dark count number during integration time, σ_{CIC} the clock induced charge noise and σ_{read} the readout noise, usually given in e^- rms. κ is the electron multiplication gain. As it can be seen, EMCCD minimizes the readout noise while it suffers from multiplication noise. EMCCD is thus regarded to have a SNR advantage for low light imaging, where readout noise highly contributes to lowering SNR. sCMOS does have a higher effective readout noise, but doesn't suffer from multiplication noise. sCMOS SNR is:

$$SNR_{sCMOS} = \frac{n_i}{\sqrt{n_i + \sigma_{dark}^2 + \sigma_{read}^2}} \tag{2.18}$$

At high n_i, with shot noise dominated SNR, sCMOS has a higher SNR compared to EMCCD. SPADs have an advantage both for low and high light imaging because of the absence of readout and multiplication noise. However, relatively high DCR limits the SNR at low light imaging, while count suppression limits it at high light imaging. Due to count suppression, the SNR reduces when detected count rate n is comparable to $1/T_{dead}$. The detection referred SNR is:

$$SNR_{SPAD} = \frac{n_i}{\sqrt{\sigma_{n_i}{}^2 + \sigma_{dark}^2}} \tag{2.19}$$

A more detailed explanation of the count suppression phenomena is given in section 2.2.4. Figure 2.7 compares the SNR of the three imagers at different integration times. It is clear that EMCCD has a higher SNR (compared to sCMOS) at low number of detected photons due to a lower effective readout noise (1.3 e^- for sCMOS[55]). SNR for EMCCD and sCMOS is plotted up to the full well capacity, that is typically 180 000 and 30 000 electrons for EMCCD and sCMOS, respectively[55,56]. For SPAD, we used a DCR of 100 cps and a dead time of 100 ns. DCR degrades the SNR at longer integration times (33 ms) at the start of the dynamic range. The count suppression, on the other side, degrades the SNR at shorter integration times (1 ms) at the end of the dynamic range. Figure 2.7d shows the SNR comparison with respect to the incident number of photons. We used a QE of 100% and 70% for EMCCD and sCMOS, respectively. The SPAD SNR is plotter for a 40% PDE. The lower overall sensitivity translates the SNR curve horizontally towards the right, decreasing the SNR. Fill factor and the temporal aperture ratio particularly affect SPAD imagers and cause major design challenges.

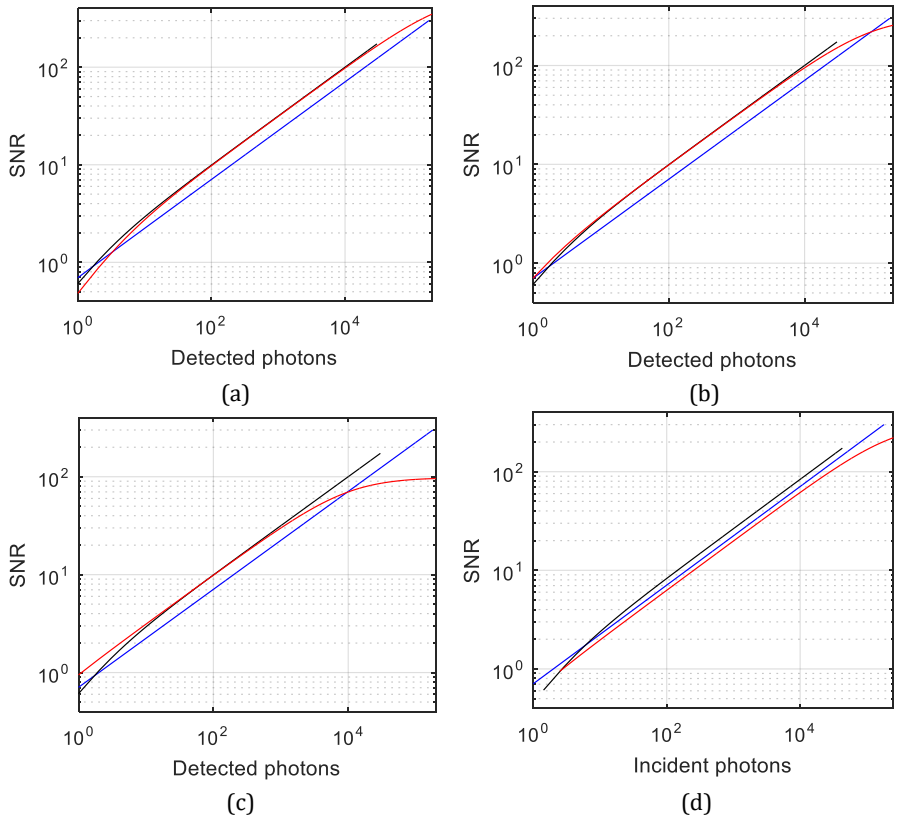


Figure 2.7. SNR for the three imagers; EMCCD (blue), sCMOS (black) and SPAD (red). SNR for (a) 33, (b) 10 and (c) 1 ms integration time. (d) Shows the SNR with respect to the incident number of photons, including the differences in sensor sensitivity. Reducing the dead time to 10 ns (1 ms integration time) increases the SPAD SNR at high illumination.

It is important to mention that the CAI dynamic range is independent on the integration time, limited by the readout noise (constant over integration) at the minimum signal and by the full well capacity at the maximum signal:

$$DR_{CAI} = \frac{\text{full well capacity}}{\sigma_{read}} \tag{2.20}$$

If σ_{read} is lower than 1 e^-, the minimum signal is still going to be limited by 1 e^-. We will nevertheless use (2.20) as figure for DR.

The SPAD has integration time dependent dynamic range, limited by the DCR at the minimum and the maximum count rate at the maximum signal. We should however note that although the maximum measured count rate m is $1/T_{dead}$, the related maximum detected count rate n is extended at the cost of a lower SNR.

Figure 2.7c shows a SNR (n related) for an integration time t=1 ms. One could expect a maximum count number of t/T_{dead}=1 ms/100 ns=10^4. This holds for m_i, while n_i is extended further. For SPAD, the dynamic range is given by:

$$DR_{SPAD} = \frac{F \times t / T_{dead}}{\sigma_{dark}} = \frac{F \times t / T_{dead}}{\sqrt{t \times DCR}}, \tag{2.21}$$

where F is the extension factor due to count suppression in SPADs. F should be chosen such that the SNR is still acceptable up to Ft/T_{dead}. Note that F is here the correction factor depending on the recharge mechanism used, usually ranging between 0.4 (passive recharge) and 75 (active event-driven recharge with corrections). Please see the following subsection for details. Dynamic range can be also limited by the period of the 1-bit memory reset signal, denoted as $T_{readout}$. Note that the minimum count number n_{min}=$\sqrt{(tDCR)}$. SNR is under this condition:

$$SNR = \frac{n_{\min}}{\sqrt{n_{\min} + tDCR}} = \frac{n_{\min}}{\sqrt{n_{\min} + n_{\min}^2}} \approx 1, \tag{2.22}$$

or 0 dB (SNR is in imaging taken with $20\log_{10}$). If $tDCR$ is lower than 1, we should replace (2.21) with:

$$DR_{SPAD} = F \times t / T_{dead} \tag{2.23}$$

In addition to this discussion, we should distinguish between the intrinsic and extrinsic dynamic range limitation. The intrinsic dynamic range is discussed in the previous paragraph (limited by the readout noise, full well capacity; DCR and maximum count rate n). The extrinsic dynamic range is, on the other hand, limited by the ADC in CAI, and by the counter in SPADs. The ADC usually has a 16 bit resolution and can further limit the CAI dynamic range or worsen the SNR due to quantization.

Also, the full dynamic range using multiple exposures (sometimes referred to as optical dynamic range) is different for CAI and SPAD. The maximum photon count rate for CAI is:

$$n_{\max,CAI} = \text{full well capacity}/t_{\min} \tag{2.24}$$

The full well capacity is usually between 10^4 and 10^5, and t_{min} (minimum integration time) between 1 μs and 10 ms. For SPAD, the maximum photon count rate is given with:

$$n_{\max,SPAD} = F / T_{dead} \tag{2.25}$$

T_{dead} is generally in the range between 10 and 100 ns.

2.2.5.1 High dynamic range imaging

High dynamic range (HDR) imaging enables to image a scene with both very dark and very bright details. True HDR imagers measure a scene during one integration period and achieve HDR due to very large full well capacity and high number of bits in the used ADCs, without compromising the SNR. The imaging community also implements HDR by binning multiple frames with different integration times, but trades the SNR in this case[57,58]. Fossum proposed the concept of QIS as a parallel development with pixel shrinking sizes and reduction of noise and full well capacity[18,59]. In QIS, the dynamic range is extended by a factor of 4.6 due to the same mechanism as the SPAD count suppression with clock based recharge (section 2.2.4.3). This comes with a cost of oversampling both in space and/or in time. If we assume a typical 16 bit ADC, QIS needs to use oversampling by 65 536. This could be achieved by oversampling 1024 times in time, and 8 by 8 oversampling in space. In addition to the single-photon readout noise requirement, oversampling enforces strong requirements for speed and/or pixel size. The benefit is the extension of the dynamic range with a lower SNR for n_i>65 536 (detected counts). It should be noted that increasing the speed also increases the readout noise, and the oversampling counteracts against low readout noise.

Researchers started to use SPAD imagers to implement QIS due to high speed and single-photon detection capabilities, measuring the QIS phenomena[14,60,61]. It is however not clear why clock based recharge is used, as opposed to active event-driven recharge that extends the dynamic range further with a higher SNR.

Figure 2.8 shows the SPAD SNR with active event-drive, passive and clock based recharge. The reader can choose what should the SNR be compared to; it can be compared to an ideal imager with unlimited dynamic range (green circles) or to the SNR a linear imager reaches at saturation (green crosses), in this case 16. T_{dead} is here set to 50 ns and the accumulation time t to 12.8 μs. The maximum

measured count number m_i is then 256 and the maximum SNR for a linear imager that saturates at $1/T_{dead}$ is 16.

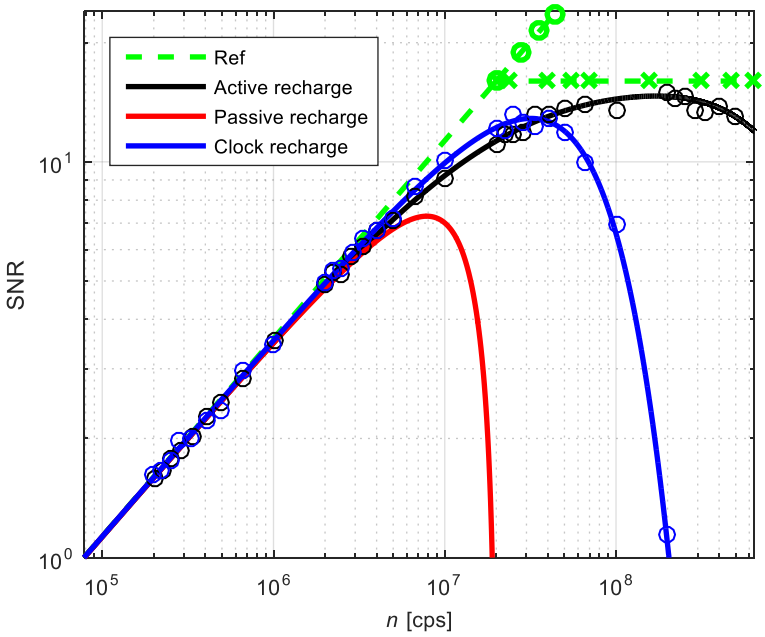


Figure 2.8. SNR_{ni} of the detected count number n_i is for all recharge mechanisms comparable to the classical shot noise $\sqrt{n_i}$ up to 20 Mcps (10% of $1/T_{dead}$). Due to count suppression, the SNR figures diverge significantly. It can be seen that the active event-driven recharge increases the dynamic range far more than the clock based recharge used in QIS. T_{dead}=50 ns and t=12.8 µs. Green shows the reference, i.e. shot noise limited SNR.

Interestingly, the active event-driven recharge SNR is higher than the standard clock based QIS SNR for exposure count rates n higher than $1.8/T_{dead}$ (36 Mcps, see Figure 2.8), due to the fact that dn_i/dm_i is very large for the clock based recharge at $n>1.8/T_{dead}$ (m is increasing very little with n, see Figure 2.4). When compared to the maximum SNR of a linear sensor with n_i=256, i.e. SNR=$\sqrt{n_i}$, the active event-driven recharge approach enables additional 4 bits of dynamic range with a drop of SNR of -1.05 dB ($n=16/T_{dead}$=320 Mcps, Figure 2.8). The clock based SNR is -2.35 dB and -45.40 dB at $n=1/T_{dead}$ and $n=16/T_{dead}$ respectively.

The dynamic range will eventually be defined by the application. We define three commonly used limits to the dynamic range. The first limit is set by n at which m loses 10% of n. One can also take n at which m saturates, i.e.

m=0.99/T_{dead}, like done in the QIS theory. However, we think that a quantitative SNR measure is a better criterion. We used a maximum SNR drop of -3 dB as acceptable. Table 1 gives theoretical values for F for different recharge mechanisms.

Table 1. Theoretical F factors to estimate the maximum n, for different recharge mechanisms and criteria. Factors are provided as $F/T_{dead}=n_{max}$. Please note that (2.13) *introduces a SNR dependent on t/T_{dead}, thus F (SNR decrease) for event-driven recharge is indicated for typical t/T_{dead} ratios (255 to 1023, as to reach 8 to 10 bit image depth).*

Recharge mechanism	10% nonlinearity[a]	Response saturation[b]	SNR decrease[c]
Passive	$0.105/T_{dead}$	$0.864/T_{dead}$[d]	$0.39/T_{dead}$[e]
Event-driven	$0.111/T_{dead}$	$99/T_{dead}$	$36/T_{dead}$ to $75/T_{dead}$[f]
Clock	$0.214/T_{dead}$	$4.605/T_{dead}$	$2.842/T_{dead}$

[a] n at which $m=0.9n$
[b] n at which $m=0.99/T_{dead}$
[c] n at which SNR_{ni} drops by 3dB compared to $\sqrt{(t/T_{dead})}$
[d] n at peak $m=1/eT_{dead}$, since passive recharge never reaches $0.99/T_{dead}$
[e] n at peak SNR_{ni}, since passive recharge never reaches $\sqrt{(t/T_{dead}/2)}$
[f] for t from $255\times T_{dead}$ to $1023\times T_{dead}$, to yield an 8- to 10-bit image

2.2.6 Maximum frame rates

We consider the maximum frame rates in continuous mode and will exclude the frame rate in burst mode[62] or event-driven readout in SPADs[63]. In burst mode, it is possible to integrate the signal with a high rate, and read out at a later stage with a low rate. In this way, readout can be very slow as to lower the noise. In CAI continuous readout, there is a tradeoff between the speed and the noise. Higher speed requires a higher bandwidth and increases the noise.

In EMCCDs, the maximum frame rate is defined as[64]:

$$FR_{EMCCD} = \left(h \times w \times (\tau_{\mathrm{ADC}} + \frac{b}{p}\tau_{\mathrm{IO}}) \right)^{-1}, \qquad (2.26)$$

where h is the row number, w the column number, τ_{ADC} the ADC conversion time, b the number of bits, p the number of I/O pins and τ_{IO} the I/O clock period. If we however include a memory element after the ADC, the equation reduces to:

$$FR_{EMCCD} = \min\left[\left(\frac{h \times w}{f_{\mathrm{ADC}}}\right)^{-1}, \left(h \times w \times \frac{b}{p}\tau_{\mathrm{IO}}\right)^{-1}\right], \tag{2.27}$$

where f_{ADC} is the ADC sample rate. sCMOS usually has a column parallel readout and thus:

$$FR_{sCMOS} = \min\left[\left(\frac{h}{f_{ADC}}\right)^{-1}, \left(h \times w \times \frac{b}{p}\tau_{\mathrm{IO}}\right)^{-1}\right] \tag{2.28}$$

In SPAD imagers, this equation reduces to:

$$FR_{SPAD} = \left(h \times w \times \frac{b}{p}\tau_{\mathrm{IO}}\right)^{-1} \tag{2.29}$$

High speed CAI increases the sensor readout noise[33] and reduces the dynamic range (or increases quantization noise) due to tradeoff between sample rate and number of bits in the ADC[65]. For example, while increasing f_{ADC} in an EMCCD from 1 MHz to 10 MHz, the readout noise increases[56] from 21 to 49 e^-. CAIs have a fundamental tradeoff between frame rate and total readout noise.

2.2.7 Nonuniformity sources

Serial readout in EMCCDs mitigates the readout noise nonuniformity. sCMOS include charge pixel and column amplifiers, and ADCs. The pixel and column based electronics increase the nonuniformity[66]. The nonuniformity can be reduced by correlated double-sampling (CDS)[67].

In SPADs, dark noise nonuniformity is emphasized due to higher electric fields. In addition to dark noise, afterpulsing is also highly dependent on traps randomly distributed over the array. However, if the afterpulsing is on average low, it will not have a significant effect on nonuniformity[14]. The photon response nonuniformity (PRNU) is manly caused by the breakdown nonuniformity that translates into PDP nonuniformity. Breakdown voltage standard deviation is in modern CMOS processes usually below 100 mV. Due to PDP saturation at higher operating voltages, the breakdown voltage deviation causes PDP nonuniformity below 2%[14].

2.3 SPAD evolution from single SPAD to imagers

SPADs have been implemented before the 1980s[68–71] in custom made processes and have emerged as one of the main single point detectors in confocal microscopy, optical range finding, proximity sensors and quantum cryptography. With the use of CMOS, SPADs increased integration capability enabling SPAD arrays[72]. SPADs are now implemented with down to 65 nm CMOS technology[73].

Single SPAD sensors are implemented either in custom or in standard CMOS processes. Most of the commercially available single SPAD sensors are built in a custom process and exhibit a higher PDP[34]. However, they have limited reproducibility and integration capability, both in the number of SPAD pixels and in the functionality on chip. A PDP comparison between custom and CMOS SPADs can be found in Figure 2.9.

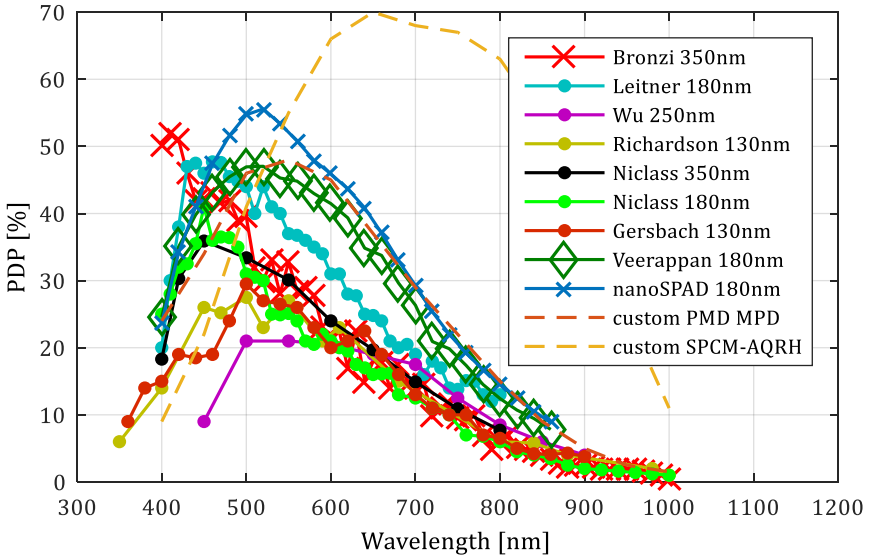


Figure 2.9. State-of-the-art SPAD PDP comparison.

CMOS SPADs can either be designed in substrate isolated or substrate shared mode. Substrate shared SPADs usually have a deeper n-well to p-substrate junction acting as the detection area, and are thus more sensitive to longer wavelengths[74]. However, the SPAD anode has either a small or no potential barrier to p-substrate, which is shared with surrounding SPADs and electronics; this construction may have a negative effect on DCR, afterpulsing, and crosstalk. Substrate isolated SPADs are the best candidates to build SPAD

arrays; they require additional design care to extend the spectral responsivity towards the red, since the detection area is usually formed at a shallow p+ to n-well junction. This is why most of the current CMOS SPADs have a peak PDP at around 450 nm. A PDP comparison between substrate shared and isolated SPADs can be found in Veerappan, 2014[45] (Figure 14).

CMOS SPADs' most important specifications are DCR, PDP, timing jitter, afterpulsing and dead time (see section 2.2). Typically, DCR ranges from 10^{-1} to 10^{4} counts per second per micrometer square at room temperature[34,44]. Peak PDP is comprised between 20% and 50%, with high diversity in spectral responsivity[44]. Timing jitter is around 100 ps. Afterpulsing with active recharge is usually below 10% with T_{dead} between 20 and 100 ns[34].

2.3.1 Types of SPAD imagers

Large 2D SPAD arrays are predominantly implemented with substrate isolated SPADs (currently up to 340×240 and 512×512 pixels)[60,75]. Pulses generated by SPADs can be counted in a digital or analog counter. They can be also used as time-resolved signals for time-to-digital converters (TDCs). Digital counters and TDCs are built at pixel[76] or column level[77]. When analog counters are used, the corresponding voltage is digitized by analog-to-digital converters (ADCs).

2.3.1.1 Counting based SPAD imagers

Counting based SPAD imagers output photon count based frames. The pixels usually also include time gating as short as 700 ps[78], but do not output time-explicit time-stamps. Time-gated arrays throw away photons outside the gate in case only one gate is implemented. The benefit is photon counting ability and potentially very high speed.

Counting imagers can have digital or analog counters. 1-bit digital counters have been firstly used to maximize the fill factor, and accumulation of bits is done in the FPGA[13]. This enables a continuous frame rate mainly limited by the I/O and FPGA speed (see section 2.2.6). Analog counters are, as well, used to maximize fill factor, and increase the full well capacity while keeping single photon capability. However, these imagers require ADCs that decrease the frame rate.

Imagers with multi-bit digital counters have been lately implemented with 3D stacking technology, and enable fill factor of 45%[73]. The chip output data rate is reduced by in-pixel accumulation. For example, a 512×512 pixel array with

10 µs SPAD recharge/reset period and 12-bit counter reduces the data rate from 512×512/10 µs=26 Gbps to 512×512×12/([2^{12}-1]×10 µs)=77 Mbps, while keeping the same maximum measurable count rate. Of course, one could reduce the recharge period in the multi-bit counter to have a higher frame rate and higher maximum count rate. These imagers could also have a variable counter bit width, as to cover both higher speed imaging (at lower number of bits) and high dynamic range imaging (at higher number of bits with lower frame rate).

2.3.1.2 TDC based SPAD imagers

TDC based SPAD imagers output explicit timing data, where the counting space-resolved information is less important. The data generation is increased due to the timing information. Each count carries a timestamp with the number of bits related to the TDC resolution. The data stream is related to the laser clock, and the activity can be controlled by changing the laser intensity.

TDCs were implemented both per pixel or column level. Per pixel TDCs are known to reduce the fill factor, typically below 5%[79,80]. Column level TDCs increase the fill factor (28% in an implementation by Lindner *et al.*[81]), but increasing routing complexity. Pulse collisions can occur if a shared bus is used[81], thus techniques must be devised to suppress or at least to detect these collisions.

2.3.2 SwissSPAD - a 512x128 time-gated SPAD image sensor

SwissSPAD is a 512×128 counting based SPAD imager with 1-bit in-pixel memory, fabricated in a high-voltage CMOS process[13]. Figure 2.10 shows the block diagram and pixel circuit. The pixel circuit can be divided in three sections – the photo-sensor, a 1-bit memory, and the readout. The first section includes the SPAD and passive quenching. A bias voltage V_{OP} is applied to the cathode of the SPAD reaching the Geiger mode of operation at $|V_{BD}|+V_E$, where V_E is the excess bias voltage. In this architecture, the access to the memory can be time-limited by using a time gate (T4), which enables photosensitive windows as short as 4 ns with a skew better than 150 ps across the entire array[13]. The photosensitive window is determined by the recharge (T2) to gate (T4) falling edge time points. The corresponding gate rise time, fall time, and gate duration distributions have also been fully characterized over the array[13], featuring values of 20 ps, 101.3 ps and 314.8 ps FWHM, respectively. A 1-bit memory (T5-8) is used to hold the photon hit information, which is subsequently read out using row select transistor T11.

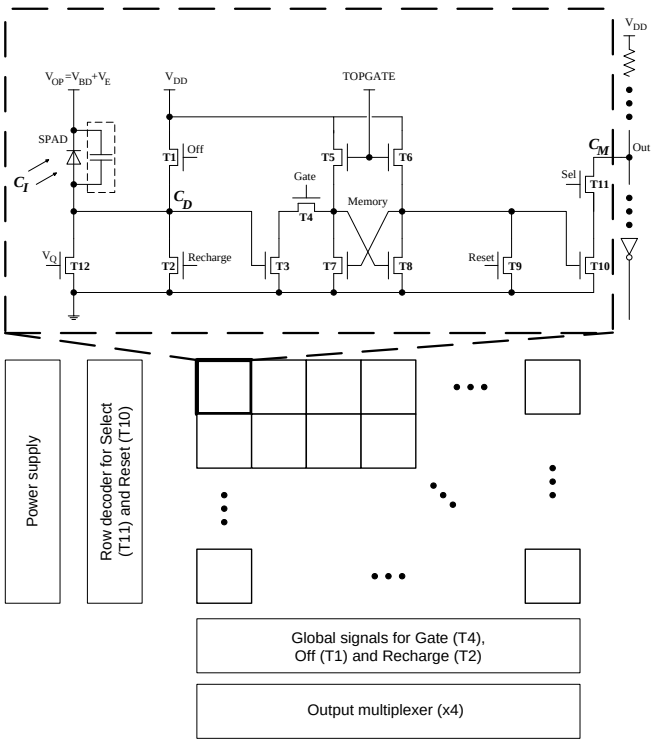


Figure 2.10. SwissSPAD block diagram with pixel circuit. Readout is performed in a rolling shutter manner. T12 is used for passive quenching, T1 and T2 for switching the SPAD on and off. T4 is used for global gating in time-resolved applications. T7 and T8 form a memory loaded by T5 and T6. T9 is a reset transistor recharging the memory output at the end of each readout of the row. T11 is used to select the specific array row and transfer the memory value to the output line.

The timing sequence of readout is shown in Figure 2.11. RS shows the gate signal at T9, while OE shows the gate signal at the select transistor T11.

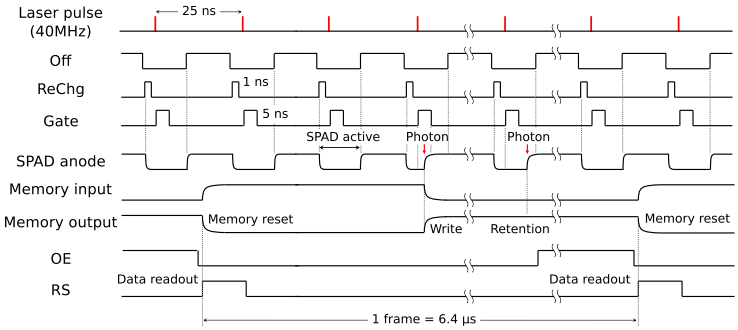


Figure 2.11. Timing diagram of the readout. Courtesy of Samuel Burri.

Readout takes place in rolling shutter mode with a 6.4 µs frame period. The 1-bit information is sent to a data acquisition board based on a field programmable gate array (FPGA), where multiple bits can be accumulated so as to acquire gray level images. Figure 2.12 shows the chip micrograph.

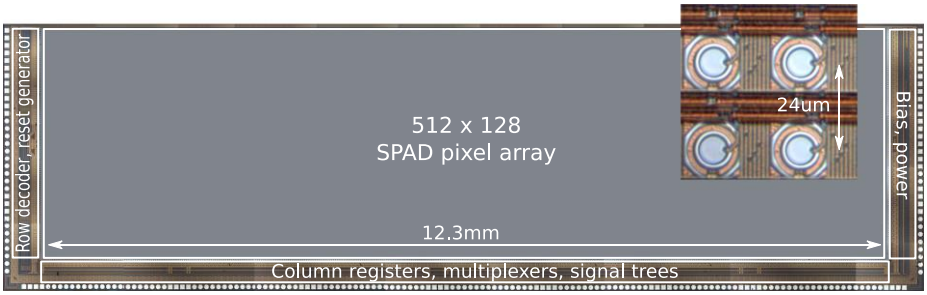


Figure 2.12. The SwissSPAD 65k pixel array surrounded by logic for row selection, gating, voltage distribution and readout. The pixel pitch is 24 µm and the logic is designed in a 0.35 µm CMOS technology. The native fill-factor (i.e. without microlenses) is 5%. Courtesy of Samuel Burri.

Figure 2.13 shows a physical appearance of the SPAD camera.

Figure 2.13. Physical appearance of the camera based on SwissSPAD; a daughterboard hosting the chip is electrically connected to a motherboard hosting two Xilinx-IV that process the raw frames generated by the chip and format them to send them through a USB2 link.

2.3.2.1 Photon detection probability

The PDP curve has a peak at 450 nm due to the shallow p+ to n well junction, with a PDP>5% between 400 and 700 nm.

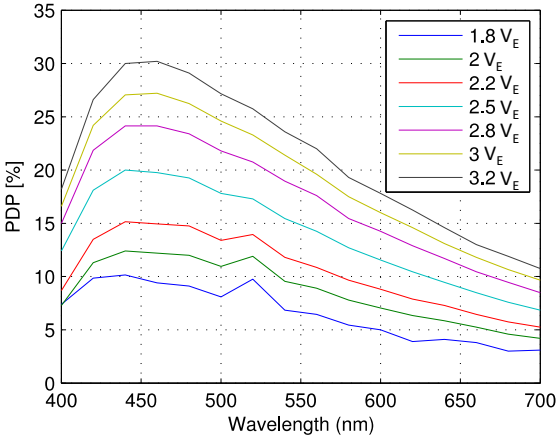


Figure 2.14. PDP spectra at different excess bias V_E.

SPADs are biased at a constant V_{OP}. As mentioned earlier, V_{BD} nonuniformity, mainly caused by photodetector area/doping differences, will directly lead to V_E and PDP nonuniformity. Higher V_E causes higher PDP. The distribution of the breakdown voltage and the Gaussian fit using the excess count rate (ECR) method[20] are presented in Figure 2.15. In this method, the ECR is measured under constant illumination over different V_{OP}. V_{OP} is used at higher voltages to generate pulses above V_{th}, the threshold voltage of transistor T3. The ECR versus V_{OP} curve can be interpolated to determine the V_{OP} value at which ECR=0. We measured an average V_{BD} value of 19.53V with a very low standard deviation of 0.067 V. A variation of 67 mV in V_{BD} will cause a relative count change of less than 1.8% (decreasing when using higher V_{OP}). Because the uncertainties of V_{BD} and V_{th} contribute to the measured uncertainty, we expect the relative count change to be even lower.

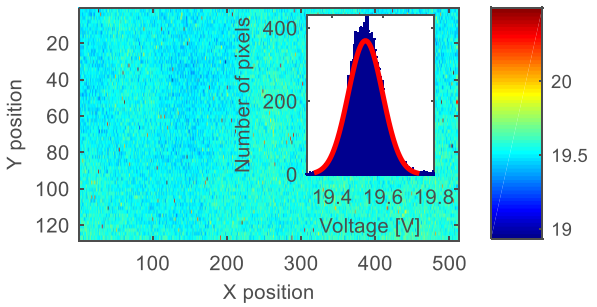


Figure 2.15. SwissSPAD breakdown voltage distribution at 25°C. Breakdown voltage differences cause most of the PDP nonuniformity in SPAD imagers. However, in SwissSPAD they are negligible.

2.3.2.2 Noise

The DCR distribution over an array (Figure 2.16) can in general be modeled with a normal distribution. High DCR pixels (*hot pixels* or *screamers*) however, do generally fall outside the Gaussian curve profile. Thus, the actual DCR distribution is slightly right-skewed due to hot pixels. When parameterizing it with a gamma distribution, skew decreases for a chip operating at lower temperatures. The skew factor is defined as $2/\sqrt{k}$, k being the shape factor of the gamma distribution; it ranges from 0.46 at 40 ˚C to 0.3 at -10 ˚C. Smaller SPAD structures yield a narrower DCR distribution over the array[82].

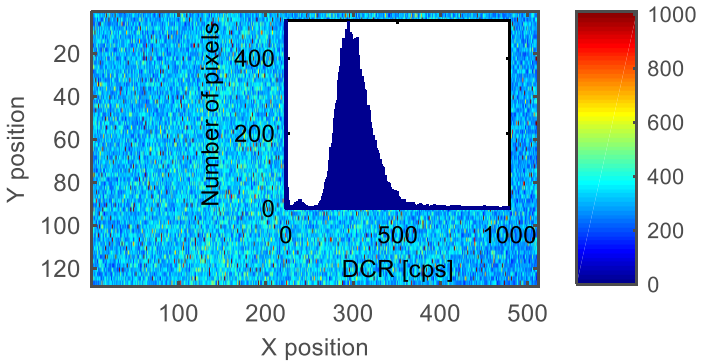


Figure 2.16. SwissSPAD DCR distribution at 40˚C, and V_{OP}=23. The color map shows DCR in cps per pixel. The Gaussian distribution is lightly right-skewed with an average of 1169 cps and a median of 302 cps. The average is highly dominated by hot pixels.

The DCR map of the chip was measured in a temperature controlled chamber at V_{OP}=23 V. The single bit frame outputs were accumulated for 255 cycles, so as to achieve a total exposure of 255×6.4 µs = 1.632 ms, resulting in 8-bit pixel counting resolution. 1000 images were taken at temperatures over the range [-10, 0, 10, 20, 30, 40] ˚C. Figure 2.17 shows DCR histograms at different temperatures. Similar DCR histograms were presented by Niclass *et al.* for a 0.35 µm CMOS chip[83]. The dark signal nonuniformity (DSNU, defined as DCR standard deviation divided by the DCR mean, σ/µ) is here >30% if hot pixels are excluded. If hot pixels are included, the DSNU rises above 1 000 %.

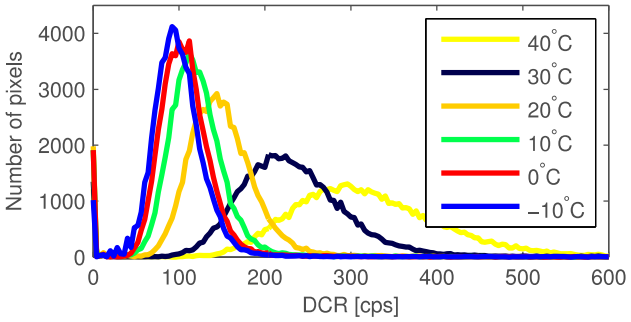


Figure 2.17. SwissSPAD DCR histograms at different temperatures. The median DCR decreases and the distribution becomes narrower and less skewed.

The average DCR, dominated by hot pixels, decreases rapidly with temperature decrease (Figure 2.18), which indicates that hot pixels are trap-assisted (both thermal generation and trap-assisted tunneling)[20].

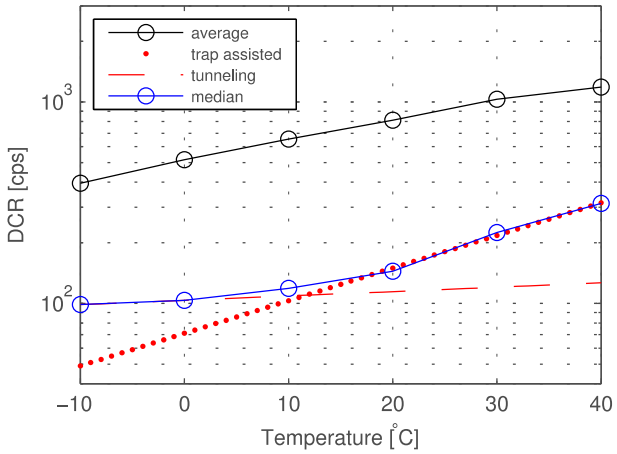


Figure 2.18. SwissSPAD measured DCR over temperature. The crossing of the dashed and dotted line indicates the cut-off temperature, at which band-to-band-tunneling DCR becomes dominant over trap-assisted DCR. However, the average, dominated by hot pixels, indicates that hot pixels are trap assisted and that the number of hot pixels can be reduced with a temperature decrease.

SPAD imagers benefit if temperature control is implemented. It eases image correction and lowers the number of hot pixels. A good operating point is the cut-off temperature where the dashed and dotted lines from Figure 2.18 cross over. The DCR at temperatures lower than the cut-off temperature will not change drastically, since band-to-band tunneling becomes the dominant DCR mechanism.

While the chip approaches thermal equilibrium, the DCR increases to stabilize after a certain time (Figure 2.19). Cooling lowers this equilibrium point and stabilizes the DCR faster, in addition to lowering it.

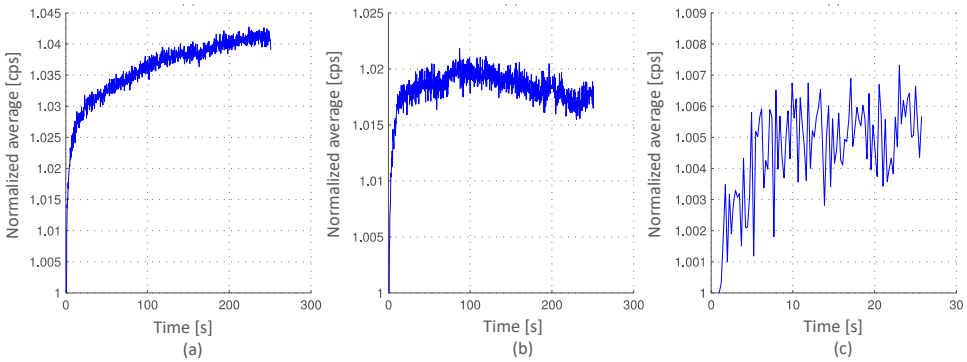


Figure 2.19. SwissSPAD normalized average DCR over time at room temperature (a) without cooling and (b) with air cooling, and 23 V_{OP}. Lower DCR increase and faster stabilization can be achieved with cooling. (c) DCR at 0˚C shows negligible heating effects on DCR when operating at lower temperatures.

Figure 2.20 shows inter-arrival time histograms. Afterpulsing and crosstalk are measured under constant illumination of 2 kcps in a temperature controlled chamber, and are found to be lower than 0.3%, too low to introduce nonuniformity. Low afterpulsing is a consequence of the long frame time of 6.4 µs.

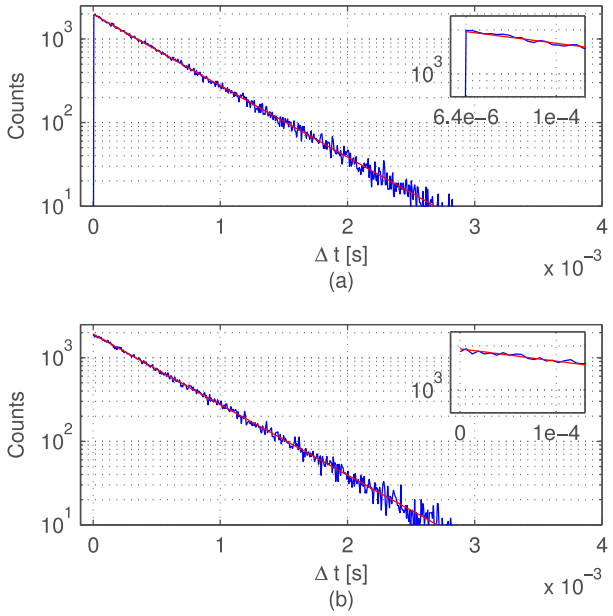


Figure 2.20. (a) SwissSPAD afterpulsing and (b) crosstalk histograms, for a representative pixel (inter-arrival times vs. number of occurrences). (a) The integral of the excess of the exponential fit from the dead time (which is the 6.4 µs frame time) to infinity represents the afterpulsing probability. Afterpulsing generally appears at short inter-arrival times. (b) Crosstalk is also measured using inter-arrival times, but between two adjacent pixels. Note that the integral starts at zero in this case, since there is no dead time between pixels, in principle. Both distributions fit very well an exponential, implying negligible nonuniformity due to either afterpulsing or crosstalk.

2.3.2.3 Count suppression

The count suppression (section 2.2.4) was experimentally observed with SwissSPAD. For the SPAD imager response measurement, a DC white light adjustable lamp, monochromator, integrating sphere and reference diode were used to generate a uniform light distribution across the sensor. n is calculated using the reference photodiode. The measured count rate m response following (2.11) is shown in Figure 2.21. It widens the dynamic range[18] by 4.6.

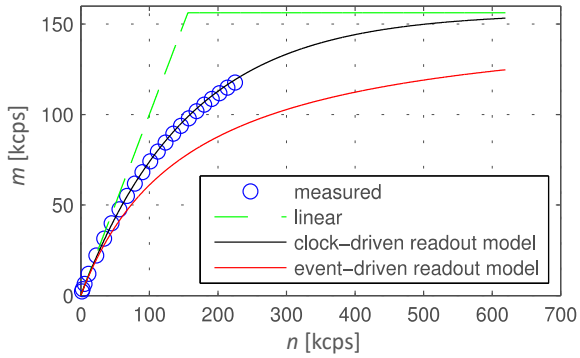


Figure 2.21. SwissSPAD signal response for the useful dynamic range of SwissSPAD with clock-driven readout architecture. The clock-driven readout signal response is modeled with (2.11).

Count suppression shot noise changes were observed as expected (see section 2.2.4.4). σ_{mi} and σ_{ni} are shown in Figure 2.22 and Figure 2.23. Blue circles show measured data, higher than expected by modeling due to thermal related changes in count number (see section 2.3.2.2).

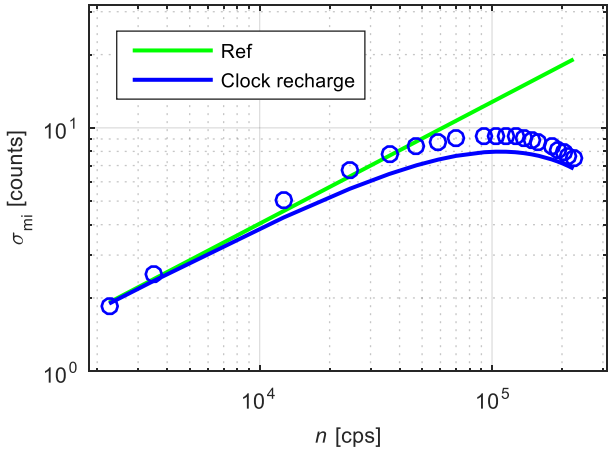


Figure 2.22. Measured count number m_i related shot noise σ_{mi} decreases while n is approaching 165 kcps, i.e. $1/T_{readout}=1/T_{dead}$. The green curve is the classical shot noise of $\sqrt{n_i}$. The blue curve is modeled with (2.15), circles indicated measured data.

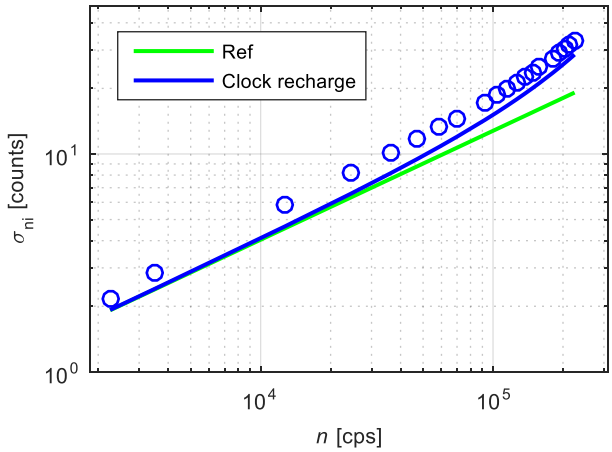


Figure 2.23. Detected count number n_i related shot noise σ_{ni} increases while n is approaching 165 kcps, i.e. $1/T_{readout}=1/T_{dead}$. The green curve is the classical shot noise of $\sqrt{n_i}$. The blue curve is modeled with (2.16), circles indicate measured data.

2.3.2.4 Count suppression, PDE and DCR correction

The PDE and DCR corrections are dependent on the temperature and excess bias voltage. An example of different responses is shown in Figure 2.24. The photon flux i is converted into the detected count rate n=PDE×i+DCR. The calibration shown here is performed for a V_{OP} of 22 V at room temperature, corresponding to an excess bias voltage V_E of 2.5 V. The same experimental setup from the last section was used. Single bit frames were accumulated for 255 cycles, corresponding to an integration duration t of 1.632 ms and 8-bit frames. To minimize shot noise, an average image over 1000 8-bit frames under constant illumination was used for the calculation.

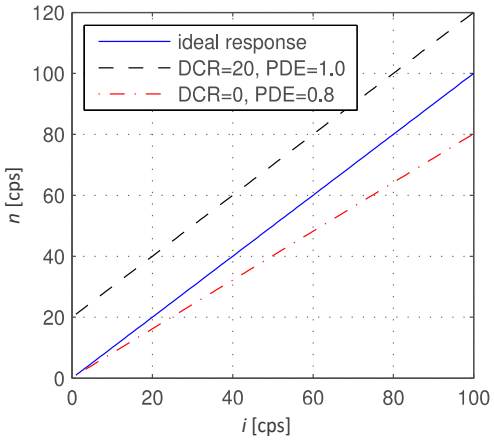


Figure 2.24. Example of different responses. Detected count rate n is shown against incident photon flux i.

m is firstly corrected for count suppression with (2.12). We then corrected it for the DCR offset and PDE around the normalized average value. Note the difference between the impinging photon flux *i*, the detected count rate *n* and the measured count rate *m.* The final correction for *m* is given by:

$$m_{corr} = \frac{\left[-\ln(1 - m \times T_{readout}) / T_{readout}\right] - \mathrm{DCR}}{\mathrm{PDE}_n}, \tag{2.30}$$

where each pixel has its own $\mathrm{PDE_n}$(normalized around the average PDE), DCR and *m* values.

The linearity *L*, displayed in Figure 2.25, is defined as:

$$L = 1 - \frac{A_1}{A_2} \tag{2.31}$$

where A_1 is the area between the (PDE and DCR) corrected measured count rate m_{corr} and a linear response, i.e. $\int \left| m_{corr} - linear \right|$ and A_2 the area between the linear response and the x axis, i.e. $\int linear$. The curves are integrated from n=0 to 230 kcps.

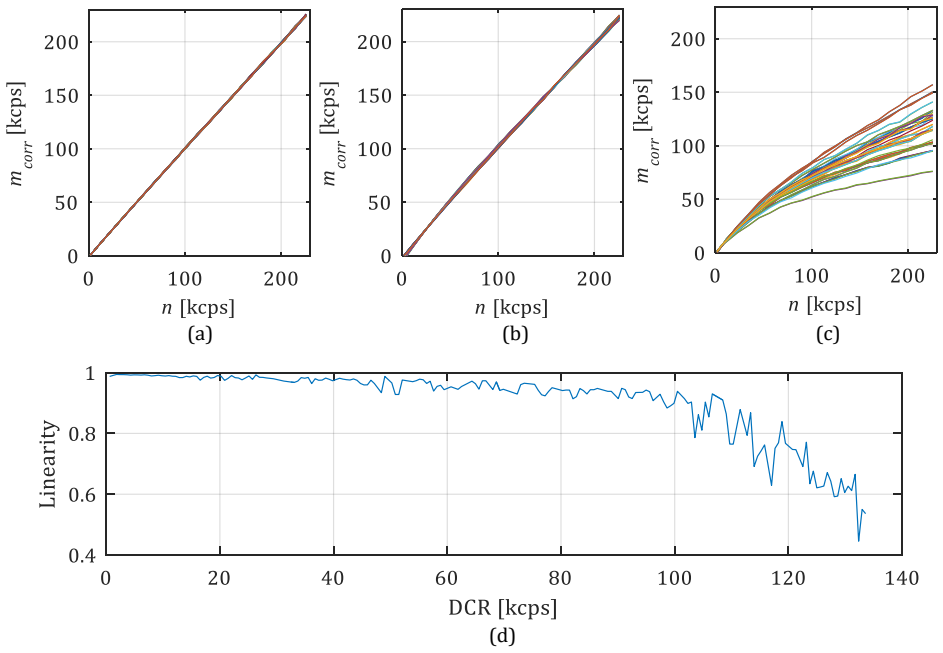
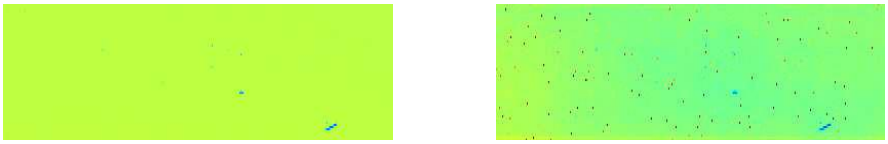


Figure 2.25. SwissSPAD m_{corr} for pixels with different DCR as a function of detected count rate n measured at a reference diode. m_{corr} is extended over $1/T_{readout}$=156 kcps, increasing the dynamic range. V_{OP}=22 V and λ=600 nm. Pixels were corrected for count loss, gain (PDE), and offset (DCR) differences. Plots showing (a) pixels with low DCR, (b) pixels with high DCR, and (c) pixels with very high DCR (screamers). (d) Shows linearity defined with (2.31) as a function of DCR.

Most of the pixels can actually still be used since only a few of them reach high count suppression due to DCR. In numbers, there are 1.8% noisy pixels with more than 6127 cps DCR that can be used (if the signal intensity is high enough to be observable in the time variability of DCR) by correcting for PDE differences and DCR. The 6127 cps figure is used as a corner DCR of a cumulative DCR graph, where DCR starts to increase rapidly (see Figure 2b in the paper by Burri *et al.*[13]). Only 0.3% of all pixels, comprised within the aforementioned 1.8% fraction, are in the very high count suppression region ($m \sim 1/T_{readout}$). Those pixels will respond to light insignificantly and nonlinearly even after recovering the count suppression. They need to be replaced, e.g. by the average of adjacent pixels. An example of the correction is given in Figure 2.26.

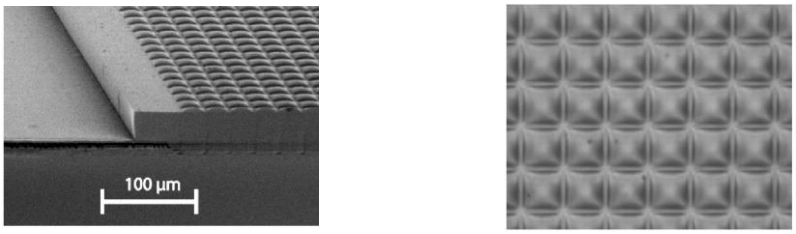


(a) (b)

Figure 2.26. (a) Corrected array response compared to a (b) raw response.

2.3.2.5 Fill factor increase

SwissSPAD native fill factor is 5%. Low fill factor can be partially recovered in SPAD imagers by using microlenses[13,84–86]. Microlenses are miniature lenses placed on each pixel to concentrate light onto the sensitive area, so as to increase the overall PDE. Figure 2.27 shows a SEM scan and micrograph of microlenses deposited on chip.



(a) (b)

Figure 2.27. (a) SEM scan of the microlens array deposited on the sensor. (b) Optical micrograph of the microlens array.

The curvature and height of microlenses with respect to the sensitive layers are usually optimized for a range of f-numbers at the image plane, i.e. for a degree of collimation of light impinging. The f-number is defined as $f_2/2r$, where f_2 is the focal length of the imaging lens and r is the aperture radius (or lens radius). The numerical aperture $NA=r/f_2$ is often used interchangeably. In our experiments, we repeated the characterization conducted by Mata Pavia *et al.*[86]. We firstly used a USAF pattern as target to verify the focus accuracy of the setup, schematically shown in Figure 2.28. The object plane was then replaced with a white target to give uniform light distribution over the whole sensor. While maintaining the original light source, the objective setup was changed to increase the f-number (decrease the aperture stop) and the light intensity was recorded. The maximum incident angle is in the example in Figure 2.28 shown to decrease from α to β, eliminating high angles not supported by the microlens. Knowing the relative change in light intensity, i.e. proportional to $r^2\pi$, we

estimate the change in the microlens effectiveness. Note that Figure 2.28 shows a telecentric lens system. If non-telecentric objectives are used, auxiliary pixels will feature a central ray that is not parallel to the optical axis and reduce microlens effectiveness.

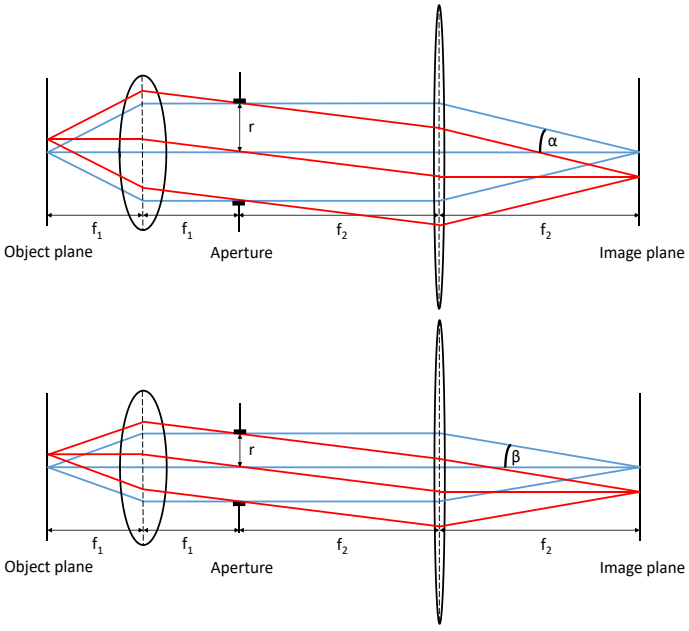


Figure 2.28. A 2-lens telecentric system showing the microlens concentration factor measurements. The aperture was narrowed to reduce the maximum image plane incident angle and increase microlens effectiveness.

As the level of collimation increases, the concentration becomes more efficient (see Figure 2.29). Results of the concentration factor (CF) are presented in Figure 2.29 and show an improvement compared to results by Burri *et al.*[13]. The average heights are given by the manufacturer. The simulation/measurement difference for 37 µm height microlenses is probably due to a different height of the central microlenses, closer to 45 µm. This improvement was achieved by optimizing the microlenses height. Since this relative measurement method compares the CFs at different f-numbers with the CF at f/1.8, the reference CF at f/1.8 was calculated using the same simulator used in[86].

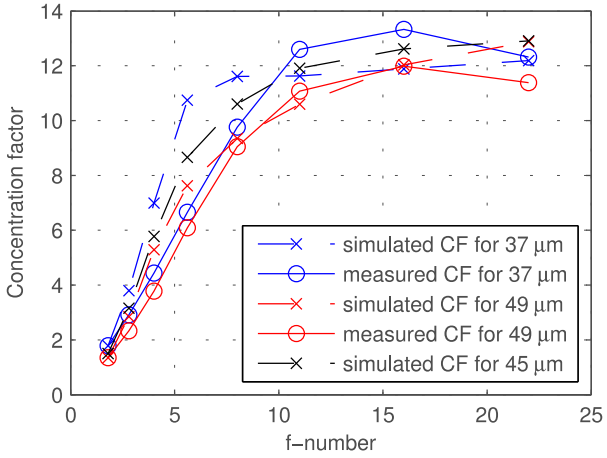


Figure 2.29. Average SwissSPAD concentration factor over different f-numbers for a central 20×20 pixel region for chip#1 with 37 µm microlenses height and chip#3 with 49 µm microlenses height.

When increasing the f-number of the optical system, the light becomes more collimated and the microlenses more effective, but less uniform, mainly due to the telecentric error[86], as shown in Figure 2.30. Height variations and misalignments become more critical at higher f-numbers.

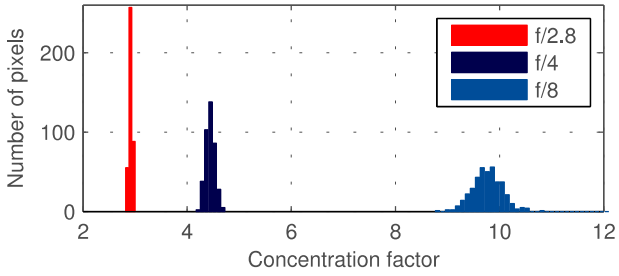


Figure 2.30. Shows the CF distribution for SwissSPAD chip#1 at f/2.8, CF at f/4 and CF at f/8.

The most critical manufacturing step is the alignment of the microlenses with respect to the detector array. Chip surface nonplanarities do also affect the height uniformity of the microlenses. Microlense height measurements for three different chips show typical intra-chip nonuniformities of 37.4 +/− 1.8 µm for chip#1, 49 +/− 2.1 µm for chip#2 and 51.6 +/− 8.4 µm for chip#3. The result is represented as an average with the maximum span. Heights were measured at six different points.

When the manufacturing microlenses height was set to an optimal height of 45 μm, 14 different chips yielded an average height of 43.4 μm with 8.1 μm standard deviation, which would result in inter-chip CF changes of 2.4 at f/8, estimated using simulations.

2.3.2.6 High speed imaging

SwissSPAD lacks multi-bit digital counters within the chip. The data rate from the chip is thus given by 512×128/6.4 μs, and yields 10 Gbps spread over 128 data pins. If the imager had multi-bit digital counters, it would decrease the data rate as discussed in section 2.3.1.1 and possibly increase the maximum frame rate. However, the data transfer bottleneck is the USB2/USB3, the link between the FPGA and PC. In practice, USB3 does not achieve 10 Gbps.

To account for the data bottleneck, we implemented two solutions. The first includes a 1 GB DDR2 memory. The video is firstly stored in the memory and then read out at a lower speed. The memory then limits the video length. The second solution includes FPGA based counters. For example, a 2-bit counter reduces the data rate to 2×512×128/(3×6.4 μs)=6.8 Gbps. A 4-bit counter reduces it further to 2.7 Gbps and an 8-bit count to 321 Mbps.

On the other hand, the frame rate is drastically reduced from 165 kfps (1-bit) to 613 fps (8-bit) due to the accumulation of SPAD counts on multi-bit digital counters, thereby losing time granularity. A tradeoff thus emerges between dynamic range and frame rate, as shown in Figure 2.31. The top picture shows either white or black pixels at a high frame rate of 165 kfps. The bottom picture distinguishes 65 536 grayscale values, but has a frame rate of 2.4 fps.

With FPGA-based counters, the multi-bit frame rate is limited by the 1-bit frame rate. If we implement chip-based counters, changing the dynamic range (accumulate more 1-bit data from a SPAD with shorter dead time) wouldn't have drastic changes in the frame rate. If we assume a limiting data bandwidth between the chip and FPGA, increasing the dynamic range from 1 bit to 8 bits would decrease the frame rate just by a factor of 8.

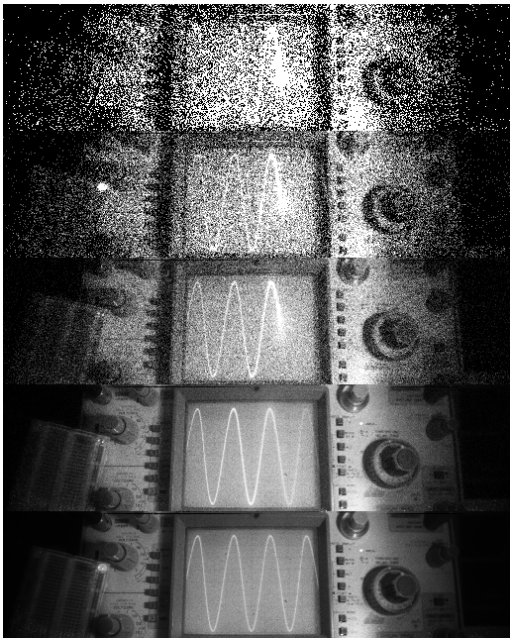


Figure 2.31. SwissSPAD imaging at different frame rates. Top image is a 165 kfps 1-bit frame, bottom image is a 2.4 fps 16-bit frame. At a higher frame rate, one can observe the head of the sinus signal. The oscilloscope frequency is in this case lower than the frame rate. Courtesy of Samuel Burri.

2.3.2.7 Intensity fluorescence

We measured fluorescence intensity using a Leica SR GSD super resolution microscope (Leica Microsystems, Wetzlar), with SwissSPAD at one optical port. Figure 2.32 shows a widefield image of a cellular cluster magnified 10×. The microlens boosted PDE was here critical to enable high quality images.

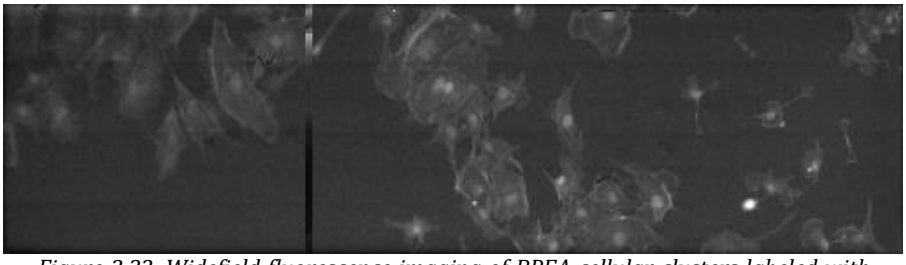


Figure 2.32. Widefield fluorescence imaging of BPEA cellular clusters labeled with MitoTracker Red CMX Ros, Alexa Fluor 488, and DAPI dyes (magnification = 10×). The dead column in the image is due to a false connection between the imager and the FPGA.

Figure 2.33 reports fluorescence intensity images of samples stained with Safranin and Fast Green, having peak excitation wavelengths of 530 nm and 620

nm respectively. A wideband lamp was used as the excitation source. Filtering was used in two subsequent exposures of the same sample and software based recomposition was then applied. It should be noted that Figure 2.33 and a 320×240[60] array represent the state-of-the-art in SPAD image quality when compared with images taken with other recent SPAD imagers featuring lower resolution, uniformity and/or fill factor[87–89].

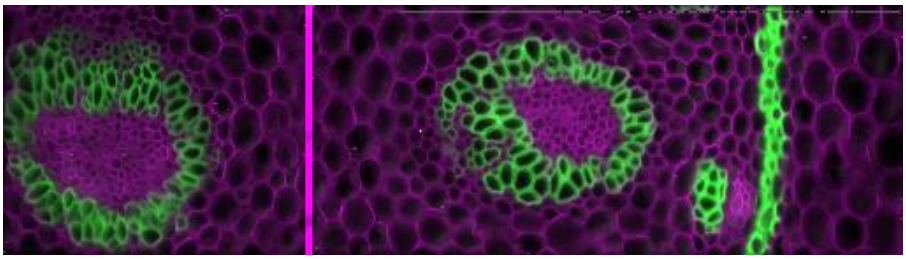


Figure 2.33. Composite fluorescence image of a thin slice of a plant root stained with a mixture of fluorescent dyes. The picture was obtained with two color filters centered at 530 and 640nm, respectively. The dead column in the image is due to a false connection between the imager and the FPGA.

2.3.2.8 Fluorescence lifetime

SwissSPAD uses time gating to extract lifetime information of a fluorophore first molecular excited state. A pulsed laser is used to excite a fluorescent molecule, a dye labeled to structure (a nucleus within a cell for example) or specific target molecules. While returning to the ground state, molecules emit a photon and the delay between the laser excitation and the emission is exponentially distributed. The distribution parameter, the lifetime τ, is molecule and environment specific. One can use a time-gated (or windowed) sensitivity and shift it with respect to the laser to extract the lifetime τ. Figure 2.34 shows the measurement procedure. If the shift is performed forward in time, with steps in the order of 25 ps, up to the laser period. The maximum shift is limited by the SwissSPAD frame time. The output photon number will be a convolution between the time gate (top graph) and emission photon delay distribution (middle graph), where t is now the time shift defined by the gate (bottom graph). The user can decide whether to conduct a deconvolution or to use the falling edge of the convolution output to extract the lifetime. A steep rising edge of the time gate reconstructs a replica exponential function at the output falling edge. A sharper rising edge of the time gate thus increases the lifetime precision.

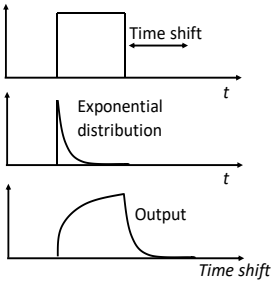
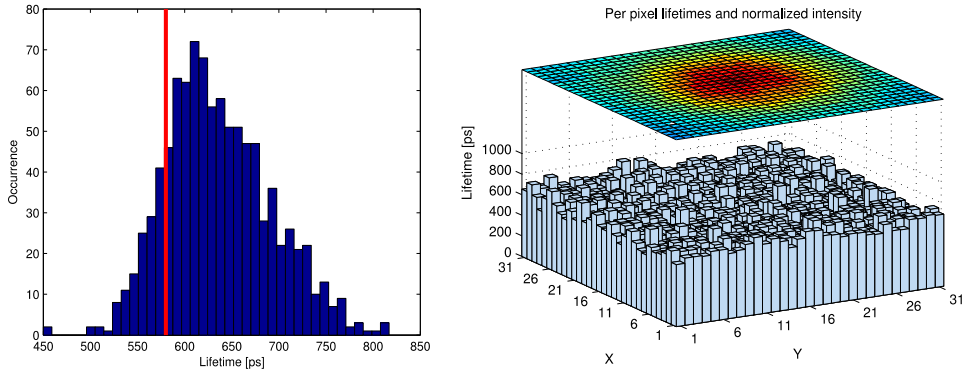


Figure 2.34. A time gate (top) can be shifted from t=0 to t=6.4 ns in steps of 25 ps. The output photon number (bottom) is then a convolution between the time gate (top) and the exponential distribution (middle).

We demonstrated potential lifetime imaging in a lab setup using point detection. Indocyanine green (ICG) in milk with a concentration of 40 µM was excited using a 790 nm laser with 55 ps pulse width and 100 MHz repetition rate synchronized with the SwissSPAD gating. Fluorescence intensity from the excited spot was measured for 512 gate windows offset by a fraction of the repetition period (25 ps). Figure 2.35 shows lifetime extracted per pixel and normalized in intensity over the excited spot. The extracted lifetimes with µ=636 ps and σ=56 ps overestimate the 580 ps reference lifetime given in literature[90]. Homulle *et al.* showed[91] how the accuracy of lifetime extraction from gated measurements can be improved through refinement of the modeling and simulation.



(a) (b)

Figure 2.35. FLIM results show extracted lifetimes distribution of 31x31 pixels compared to reference lifetime of 40µM ICG in milk (red). Figure (b) shows the comparison of intensity and lifetime per pixel. Courtesy of Samuel Burri.

2.4 State-of-the-art EMCCD, sCMOS and SPAD imagers

In order to compare the SwissSPAD imager with microlenses and an EMCCD, we used the dual port Leica SR GSD microscope placing an Andor iXon3 897 BV EMCCD on one port and SwissSPAD on the other, as shown in Figure 2.36.

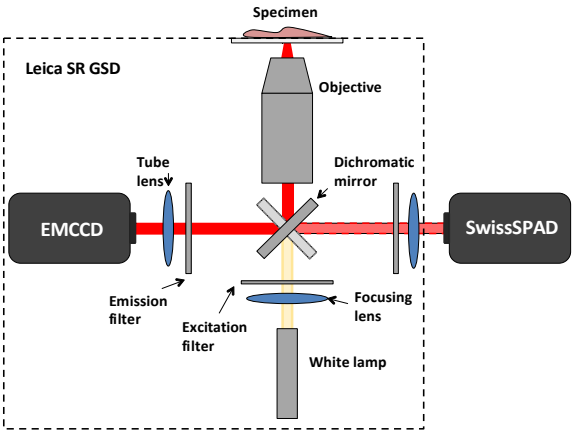


Figure 2.36. The measurement setup for comparison.

The EMCCD camera comprises 512×512 pixels, with a pitch of 16 µm; it is cooled to -75°C, while SwissSPAD operates at room temperature. The field of view that SwissSPAD is covering is about ¼ of that of the EMCCD. The field of view is extended due to differences in pixel size (16 µm *versus* 24 µm).

We used BPAE cells labeled with MitoTracker Red CMX Ros, Alexa Fluor 488 and DAPI dyes. SwissSPAD was used at V_{OP}=24. The EMCCD raw intensity image was converted to a photon count image using $counts_D = (d - b) \times g_{amp}/g_{EM}$, where d is the digital intensity value, b the bias offset, g_{amp} the preamp gain value and g_{EM} the EM gain. Due to pixel size differences, 2×2 SPAD pixels and 3×3 EMCCD pixels were binned to obtain counts for the same area. MATLAB software was used to find the overlapping area of the two images and compare the intensities.

The SwissSPAD to EMCCD intensity ratio in the region of interest was found to be 12% comparing counts per area. This ratio is a substantial improvement compared to the ratio measured without microlenses. The compared regions of interest and histograms are shown in Figure 2.37 and Figure 2.38.

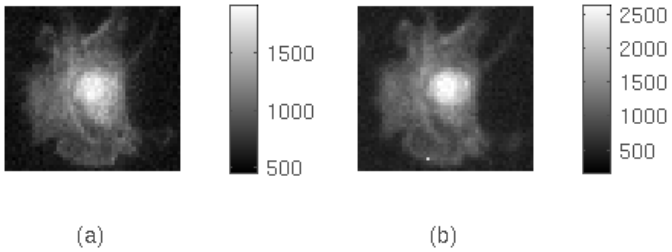


Figure 2.37. Compared region of interest with binned pixels, so as to match the size. The EMCCD (a) and SwissSPAD (b) exposure times are 10 ms and 83.2 ms, respectively, to match the average number of collected photons. The lateral bar shows the number of collected photons per exposure.

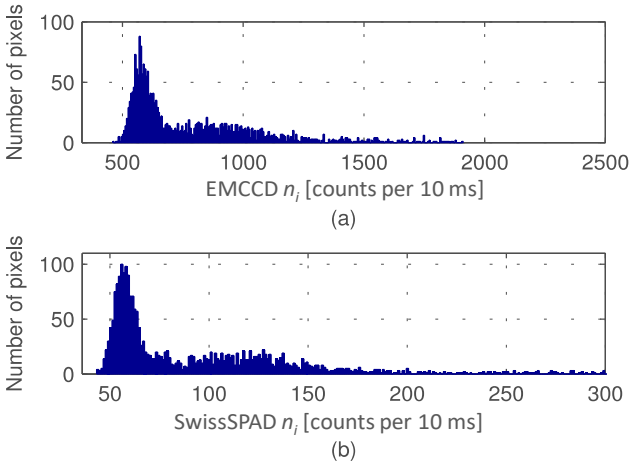
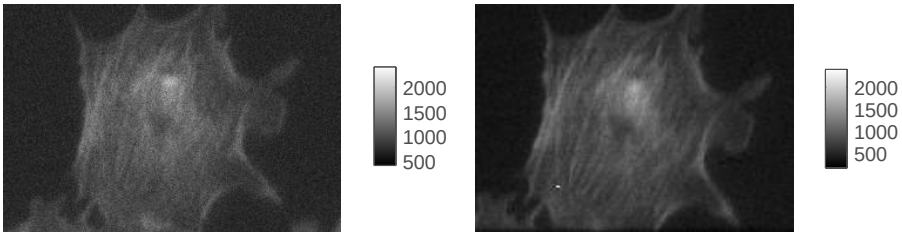


Figure 2.38. Histograms of compared region of interest with binned pixels. (a) EMCCD histogram, (b) SwissSPAD histogram.

Figure 2.39 shows images of the same cell obtained with EMCCD (a) and SwissSPAD (b); the exposure times were 10ms and 73.4ms, respectively, to match the number of collected photons. The scale shows the number of collected photons per exposure.



(a) (b)

Figure 2.39. BPAE cell imaging. The cells were labeled with MitoTracker Red CMX Ros, Alexa Fluor 488, and DAPI dyes and imaged in a microscope using two ports matching the photon counts per pixel in each sensor. (a) shows the EMCCD and (b) the SwissSPAD images obtained with 10ms and 73.4ms exposure, respectively. The scales indicate photon counts per frame.

SwissSPAD was corrected for hot pixels and count suppression. EMCCD seems to exhibit somewhat noisier images (Figure 2.38 and Figure 2.39), presumably due to multiplication noise and ADC nonlinearities. The noise is however very uniformly distributed over the array. SwissSPAD Figure 2.39 shows horizontal pattern that is assumed to originate from microlenses. The main SwissSPAD disadvantages remain the lower PDE and the number of hot pixels (not present on images due to corrections).

Table 2 lists a comparison between commercial high-end EMCCD and sCMOS microscopy cameras, SwissSPAD and another SPAD used for super resolution microscopy[92].

Table 2. A comparison table between state-of-the-art commercial EMCCD and sCMOS, and two SPAD imagers with demonstrated application measurements[17,92].

Camera	EMCCD[56]	pco.edge 4.2 sCMOS[93]	SwissSPAD[13,14]	QVGA SPAD analog pixel[60]
Peak PDP	> 90%[a]	> 70%	30%	39.5%
Fill factor	100%	unknown	Native 5%, effective 60%	26.8%
Pixel pitch	16 µm	6.5 µm	24 µm	8 µm
Array size	512×512	2048 x 2048	512×128	320×240
Max. frame rate	100 fps, 512×512 pixels, 14 bit	100 fps (full chip), 853 fps (QVGA)	156 kfps, 1 bit	16 kfps, 1 bit

Data rate	367 Mbps	6.7 Gbps	10.2 Gbps	1.22 Gbps
Dynamic range	30 000[a,b], 133 333[c]	37 500	9 232[d]	~1 000[e], 1 950[d]
Readout noise	< 1 e-[c]	1.3 e-	N.A.	<0.15 e- in analog mode
Dark counts	0.001 cps at -85°C	0.08 e-/pixel/s at -15°C	~200 cps at 25 °C	47 cps at 25 °C
Afterpulsing	N.A.	N.A.	<0.3%	unknown
Crosstalk	unknown	unknown	<0.3%	unknown
PRNU	unknown	< 0.2 %	<1.8%[f]	unknown
DSNU	unknown	<0.3 e- rms	>30%[g]	unknown
Min. gating duration	N.A.	N.A.	4 ns	unknown
Gate skew	N.A.	N.A.	<150 ps	unknown
Operating temperature	-75 °C	5 °C	25 °C	25 °C

[a]In CAI imagers as full well capacity/readout noise
[b]Without multiplication
[c]With multiplication
[d]In 1-bit SPAD imagers $4.6t/T_{readout}/\sqrt{(tDCR)}$, 4.6 for the QIS extention[18], t=33 ms, see (2.21)
[e]Defined as for CAI (analog memory full well capacity/readout noise)
[f]Without corrections
[g]without corrections and excluding hot pixels, defined as standard deviation divided by the mean DCR

2.5 Conclusion

The main challenges in SPAD array development remain the PDP, fill factor, noise nonuniformity and overall data rates. Pixel sizes can be expanded to increase fill factor, however DCR and DCR nonuniformity increase, since larger areas imply a larger probability that pixels will include silicon traps. The effective fill factor can be increased by depositing a microlens array on the SPAD sensor. We have shown an increase in effective fill factor by a factor of 12 (from 5% native to 60% effective fill factor for the 512×128 pixel array SwissSPAD[14]),

applying extensive software tools to optimize microlenses for different light collimation levels[86]. We demonstrated a large format SPAD array for high speed, fluorescence intensity and lifetime measurement. Forthcoming 3D integration technologies will allow separated tiers for sensors and electronics[94]. In this way, complex electronics will be built without compromising the native pixel fill factor. SPADs will however always feature a sensitive area smaller than the pixel area due to a guard ring, usually 1 to 2 μm in width, which needs to be deployed to prevent premature lateral edge breakdown.

CMOS SPAD peak PDP reaches 50% today, however we expect that it will rise towards custom made PDPs (>70%) with technology advances.

3 Design considerations for high-end implementations

Chapter 2 introduced specific benefits of SPAD imagers compared to other implementations. However, complex integration raises questions about power sustainability, signal skew and integrity, and enormous data rates. Additionally, there remain challenges of single SPAD fill factor and PDP. Sections 3.1 and 3.2 details the large format integration challenges and presents solutions based on experimental results. In section 3.3 we emphasize what other specific applications (could) benefit from SPAD imagers overcoming the technological drawbacks.

3.1 Nonuniformity

SwissSPAD showed low PRNU, also due to a long dead time that limits the average SPAD array current below 1 mA. However, shorter dead times with integrated counters might increase the power consumption and IR drop. The SPAD power consumption is composed of three parts; the SPAD leakage current (usually neglected), the capacitive load current and the transient current during SPAD quenching (when SPAD has a resistance R_D of around 100 Ω[23]). Lee *et al*. reported a 190 fF junction capacitance C_D for a 30×30 μm² avalanche photodiode implemented in a 180 nm CMOS process[23] and 35 fF and 140 fF junction capacitance for 10×10 μm² and 20×20 μm² areas, respectively, in a 130 nm CMOS[95]. We here assume a junction capacitance, C_D, of 1-10 fF for a 2×2 μm². For example, for a SPAD capacitance of 10 fF, a V_E of 3.3 V, an array size of h×w=1024×1024 and a T_{dead}=100 ns, the capacitive load power at the array count rate saturation will be:

$$P = V_E^2 C_D fhw = \frac{V_E^2 C_D hw}{T_{dead}}, \tag{3.1}$$

in this case 1.14 W. This is, of course, a worst-case scenario with a fully bright scene and a chip operating at the end of the dynamic range. One should however note that CMOS image sensors usually have a total power consumption of less than 300 mW[96–98]. Large format arrays do both increase the array current consumption and the power line resistance. Longer power lines are required to reach to central pixel, and fill factor maximization requests lines as thin as possible.

Figure 3.1 shows an example of a 1D SPAD array where R_P is the per pixel parasitic resistance. V_1 is the effective V_{OP} for the first left SPAD. If all SPADs fire

at the same time, it can be expressed as: $V_1=V_{OP}-R_P\times w\times I_A$. V_2 is then $V_{OP}-R_P wI_A-R_P(w-1)I_A$.

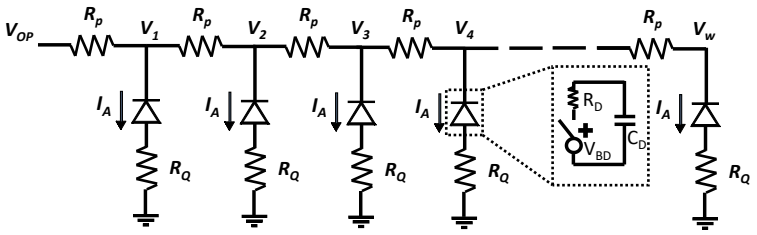


Figure 3.1. An example of a 1D SPAD array with parasitic resistance at the VOP line. The SPAD is modeled with a parallel $R_D||C_D$ impedance with a breakdown V_{BD} bias at the R_D node. During quenching, the switch in the model is shorted.

In general, we can write:

$$V_i = V_{OP} - \sum_{j=1}^{i} R_P (w - j + 1) I_A \tag{3.2}$$

V_w will then be:

$$V_w = V_{OP} - \sum_{i=1}^{w} R_P i I_A = V_{OP} - \frac{w(w+1)}{2} R_P I_A \tag{3.3}$$

The typical R_P is 50 mΩ/μm^2. For a typically used V_{OP} line of 1 μm width, and a pixel pitch of 3 μm, R_P is 150 mΩ. The average capacitive load current can be estimated by $V_E^2 C_D/T_{dead}$. Peak transient current (beginning of quenching) can reach as high as 10 mA[99]. The SPAD transient current at the end of quenching will be V_E/R_Q[100], usually below 100 μA[101]. For w=1024, and an average current of 0.33 μA per pixel, the voltage drop ΔV is 26 mV. One can also use I_A= V_E/R_Q=3.3/100 kΩ=33 μA to yield a ΔV of 2.6 V. The overestimated ΔV figure shows the higher requirements a large format SPAD imager (in this case for a high-speed imager fully deploying its dynamic range) lays on the power design. If just the last SPAD detects a photon, the voltage drop is:

$$\Delta V = w R_P I_A \tag{3.4}$$

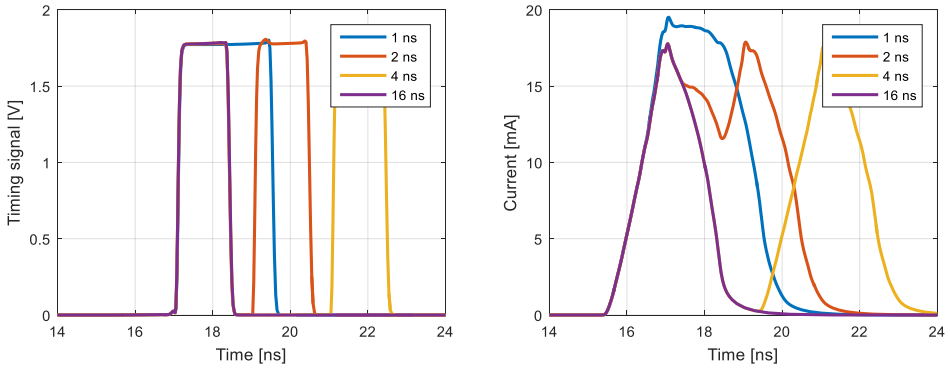
For w=1024, and taking a current of I_A=33 μA, ΔV=5 mV. The ΔV effect of PRNU is investigated with SwissSPAD, and shows that ΔV=67 mV gives a PRNU=1.8%. Important to note is the w^2 (simultaneous firing array) and w (separate firing SPAD) dependency of the IR drop. Note that equation (3.3) is more appropriate for the average current consumption and equation (3.4) for the peak current

consumption. Inadequate design can directly lead to PDP nonuniformity. In this case, wider V_{OP} lines and a meshed grid need to be used.

3.1.1 Nonuniformity of timing signals

If TDCs are built outside the SPAD array to increase the fill factor, like in the chip reported by Lindner *et al.*[81], the pulses generated by the SPADs propagate through the whole array height/width before detected by the TDC. If we use the timing of SPAD pulses, repeaters are built to improve the rise time and signal integrity.

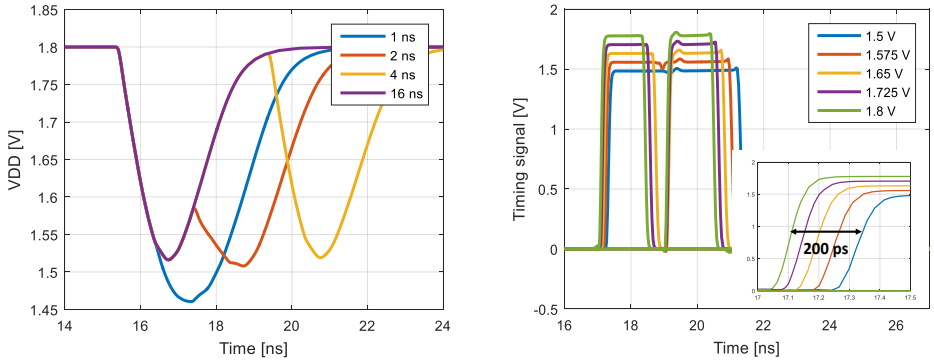
Figure 3.2 shows post-layout SPAD timing pulses at the end of a 126 pixel column (a), and the repeater power consumption (b). The pulses propagate through the whole column with 8 repeaters per line, with 9 address lines and a timing line. Different colors in the figures indicate the SPAD pulse period (1, 2, 4 and 16 ns separation between pulses).



(a) (b)

Figure 3.2. (a) Post-layout timing signals at the bottom of the column with different separations of consecutive pulses. (b) Maximum repeater current given for a SPAD firing at the top of the column and propagating to the bottom. Different colors in the figures indicate the SPAD pulse period (1, 2, 4 and 16 ns separation between pulses).

Figure 3.3 shows IR drop effects with different SPAD pulse periods. It shows the worst-case scenario, when power is supplied through a single vertical power line connected to a voltage source at the bottom of the column. The consequence is a skew in the timing signal. In this case, an IR drop of 300 mV led to a rising edge skew of 200 ps.



(a) (b)

Figure 3.3. (a) Post-layout repeater worst case V_{DD} IR drop of a 126 SPADs event-driven column with repeaters, at different pulse separation. (b) Timing skew due to voltage drops in the repeater power lines. Different colors in the figures indicate the SPAD pulse period (1, 2, 4 and 16 ns separation between pulses).

3.1.2 Load balancing

When designing timing critical signal paths, load balancing is important to ensure high uniformity. SPAD imager designers try to avoid repeaters through the array as to maximize the fill factor. The signal integrity is then compromised and the pixels further from the timing signal tree will have a higher gate jitter. Figure 3.5 shows the SwissSPAD gate width uniformity over the array. The gate width is defined by the falling edge of the recharge signal and falling edge of the gate signal (see section 2.3.2 and Figure 3.4).

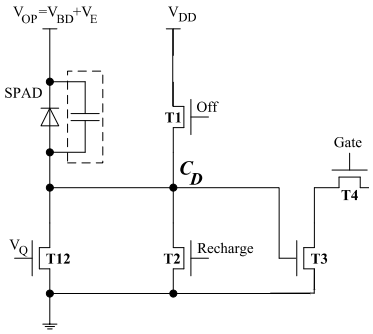


Figure 3.4. Pixel schematic with recharge and gate transistors. The recharge transistor was designed larger to be enable prompt SPAD discharge.

In this design, the recharge transistor is larger than the gate transistor, leading to a higher load at the recharge signal path. The recharge and gate signals propagate from the bottom to the top. A higher recharge load thus delays the

recharge signal and the sensitive gate becomes shorter at the top of the array (see Figure 3.5).

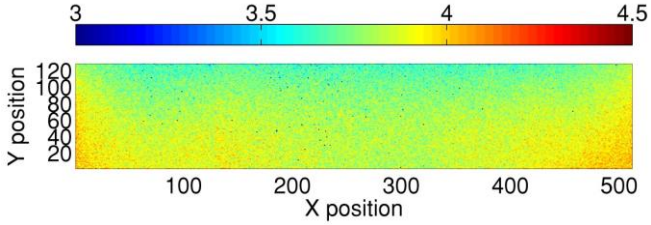


Figure 3.5. Sensitive time gate distribution over the SwissSPAD array, give in nanoseconds. Courtesy of Samuel Burri.

In a newer version of SwissSPAD, known as SwissSPAD-2, we balanced the recharge and gate transistor and thus expect better gate uniformity over the array. Figure 3.6 shows a post-layout simulation of the two signals (recharge and gate) at the bottom (no load) and top of the array. The photon sensitive time window at the top will be shifted compared to the bottom. However, the time widow width will be uniform (in contrast to the implementation in Figure 3.5).

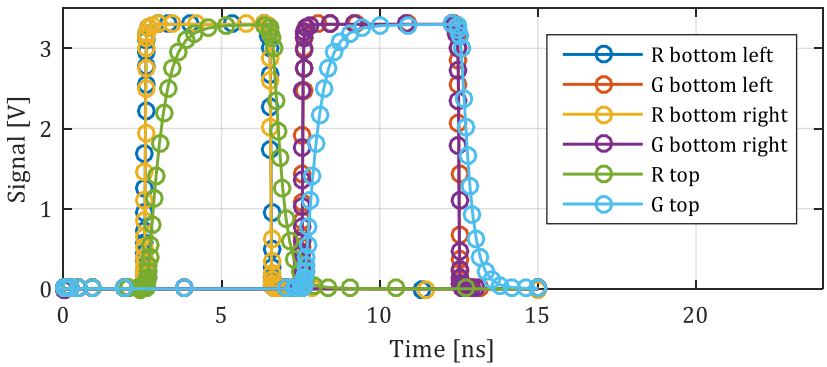


Figure 3.6. Post-layout simulation of the recharge (R) and gate (B) signal at the bottom left/right (no load, to estimate the horizontal mismatch) and top (delayed) array. Timing signals were fed through a signal tree to horizontally balance the parasitic capacitance (left right). Vertically, timing signals exhibit classical RC delay, but recharge and gate are matched.

3.2 Handling large data rates

SPAD imagers generate large data rate because of the absence of photon accumulation or counters. Counters exponentially increase the time (denominator) in the data rate equation:

$$D_{rate} = \frac{hwb}{2^{b-1}T_{bit}}, \tag{3.5}$$

where D_{rate} is the data rate, h is the vertical pixel number, w the horizontal pixel number, b the counter number of bits and T_{bit} the time a SPAD can count one photon. In case a counter is implemented in the pixel, T_{bit} is equal to T_{dead}. Thus, the data rate decreases with $b/2^{b-1}$.

In low light conditions, the number of photon events is infrequent and sparse, thus we suggested the use of event-driven readout mechanisms. In this case, we send the address of the firing pixel and its count or timestamp, rather than the whole array information. When the scene is uniformly illuminated, the data amount is then the probability of one per pixel event P multiplied by $log_2(h)\times log_2(w)\times h\times w$. P can be expressed as the number of fired pixels k divided by $h\times w$. A low data amount is achieved just with a low event probability. Single events can be read out faster than the whole array, if speed is the requirement. Event-driven readout is beneficial for a low photon rate and sparse scene illumination. If we compare the event-driven readout with a frame-based readout, the data amount is smaller for:

$$P\log_2(h)\log_2(w)hw < hw \tag{3.6}$$

$$P < \frac{1}{\log_2(h)\log_2(w)} \tag{3.7}$$

$$k < \frac{hw}{\log_2(h)\log_2(w)} \tag{3.8}$$

For a 1024×1024 imager, this average per pixel photon detection rate should be 1%. For example, for time-correlated single-photon counting, the photon probability is set to less than 10% to avoid pile-up effects. This needs to be further multiplied by the fraction of pixels actually receiving photons (dark parts of image are not generating events), easily ending at lower than 1% pixel activity.

Compression mechanisms can be used to lower the 1-bit frame data amount as it was done in FPGA in SwissSPAD. Under specific light conditions and known 0/1 sequence probability, we can apply Huffman coding. This then, however requires dynamic Huffman coding for dynamic scenes. Figure 3.6 shows the data amount (*not* the data rate) for different compression and readout mechanisms at different illumination levels. Run length coding is shown for

uniform illumination. Here, the bit width of the run length code is kept constant for one illumination point and is $\log_2(\text{length}_{max})$. At extremely low and high illumination the data is send with few lengths with a large bit width. At medium illumination, few outliers increase the bit width while there is a large amount of interleaving lengths. The data amount is thus high for medium illumination (n_i from 0.1 to 2). We developed a modified run length coding, a tradeoff between effectiveness and complexity. Data compression was implemented on a FPGA platform. It compresses the 1-bit data stream generated by SwissSPAD and lowers the data rate towards the PC or memory. A hardware implementation puts limits on the maximum complexity level. Complex operations need to be pipelined.

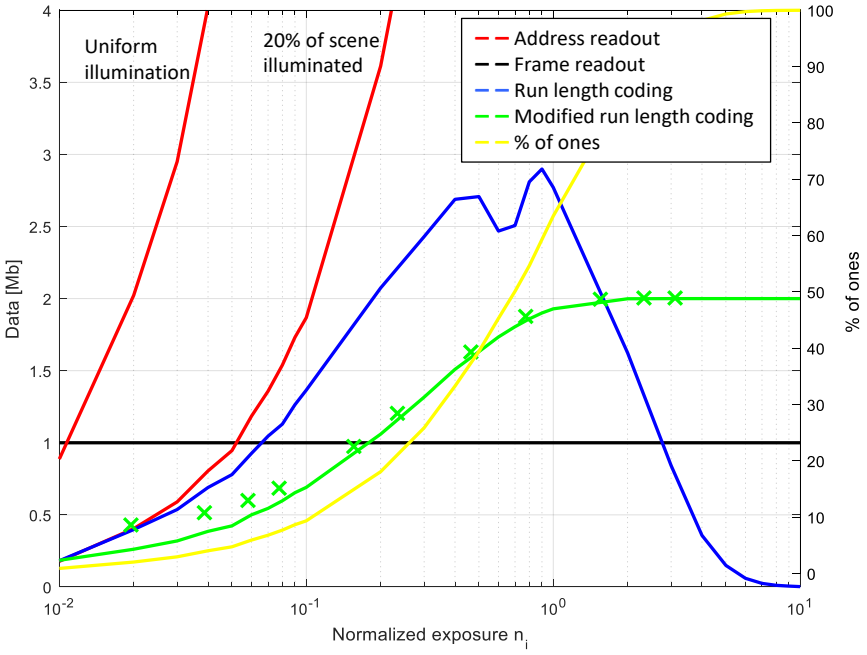


Figure 3.7. Data per frame of a 1024×1024 1-bit SPAD imager (simulated). Frame readout has a fixed 1024×1024 bits at any illumination, while address readout increases the data amount with $P \times \log_2(h) \times \log_2(w) \times h \times w$, where P is shown as % of ones. Crossed data points show measured data for a 512×128 imager and scaled to a 1024×1024 imager data. Data compression was implemented on a Virtex 4 FPGA.

The modified run length code compresses a sequence of 10 zeros as binary 00, a sequence of 4 zeros as binary 01, a 0 as 10, and a 1 as 11. Theoretically, such coding can compress the data to 20% or expand to 200%. This coding was chosen so as to match the average zero length present in our experimental data (microscopy).

Figure 3.8 shows the schematic of the compression algorithm, implemented in a Virtex 4 FPGA. The input data width N can be expanded to 2N in the worst case, when the input is a vector of interleaving ones and zeros. The pipeline array thus needs to have double the bit width, and it can take a maximum of N clock cycles to expand the data. Thus, we need an N×2N array. During the compression, we use combinational logic to create the next vector. We also need to track the number of compressed bits and the vector width for the next stage. The N different stages are then pipelined and the final data has a variable data width. The output data width needs to be interfaced with a DDR memory or USB, and have a fixed data width. In this stage, we need to use a 6N data vector. We firstly shift down the 2N data that was copied in the previous cycle to the *width fixer*, but by a variable shift depending on the content in the 6N vector. Afterwards, we copy the 2N vector to a fixed position. When the bottom content length of the 6N vector exceeds the output data width, we latch the compressed data. Note that each stage needs to implement a variable-per-clock shifter, an operation very costly in hardware. The FPGA synthesizer fails to compile the VHDL code if copying and shifting is written within the same clock cycle. Such a design already occupies 27% of FPGA slices and 25% of 4 input LUTs in a Virtex 4.

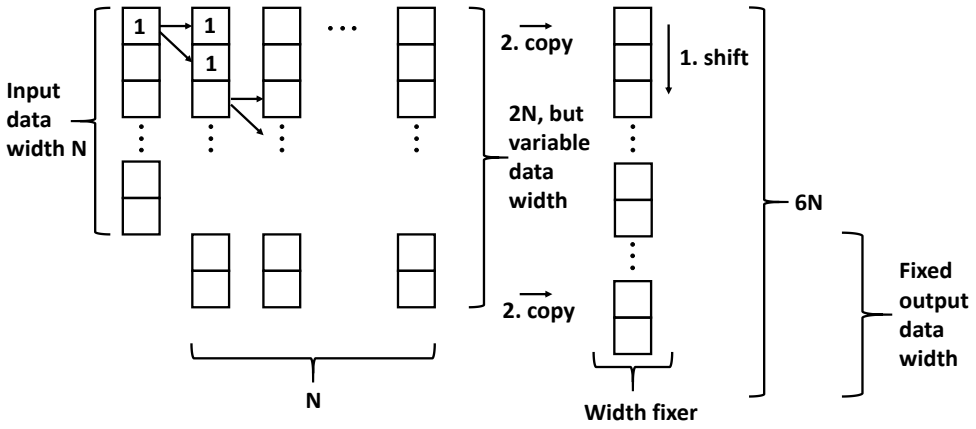


Figure 3.8. A simplified scheme of the modified run length compression algorithm.

We measured our compression scheme for a 1-bit SPAD array illuminated at different light intensities. The measured data match the simulated data, with a deviation in the low light regime, where hot pixels artificially increase pixel activity.

In the case of TDC based SPAD imagers, the data rate increases by the TDC bit resolution. Previous system implementation used data histogramming to reduce the data rate. Gershbach *et al.* indicated a possible data reduction by a factor of 1024, for TDCs with 10-bit resolution[102]. For example, a TDC resolution r will give 2^r histogram bins. If the bin width (or visually height) is z, the histogram contains $z \times 2^r$ bits of data. If the full histogram was used, it would contain $k=2^{r+z}$ events, with r bits (TDC value). The potential compression ratio is $r \times 2^{r+z}/(z \times 2^r)$, with $r=10$, equal to 1024. However, this is an overestimation. In the best case, only one histogram bin will be fully employed, and many pixels will have just DRC counts.

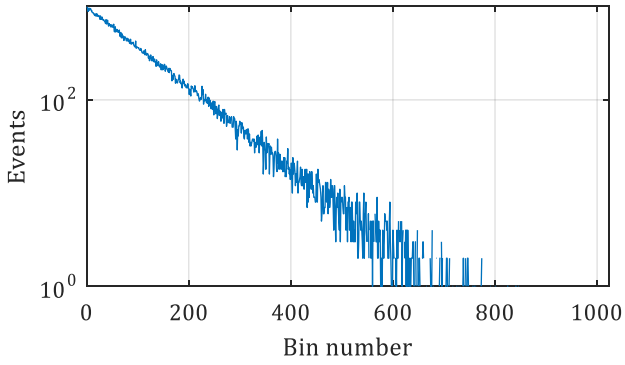


Figure 3.9. An example of exponential distribution with τ=2.5 ns, 1024 bins and a maximum of 1023 events per bin (2^z-1), with totally k_p=100 000 pixel events.

Figure 3.9 shows an example of an exponential time distribution (fluorophore lifetime of 2.5 ns) with k=100 000 pixel events, r=10 and z=10. The compression ratio is here given by:

$$\frac{rk}{z2^r}, \tag{3.9}$$

where k is the number of detected events per pixel. The compression ratio is in this case 98. To effectively use histogramming, the average per pixel number of events k should be:

$$\bar{k} > \frac{z2^r}{r}, \tag{3.10}$$

in our example k>1024. This will, of course, depend on the imaging scene. A high percentage of pixels with no illumination will lower the compression effectiveness.

3.3 Potential benefits of large format SPAD arrays

If SPAD CMOS technology advances to PDP>50% and native fill factor>50% due to 3D integration, SPAD imagers have the opportunity to combine HDR, high speed and photon counting possibility. In this chapter, we discussed some of the remaining challenges, related to PDP uniformity, power sustainability and data rates.

We propose a SPAD imager with a 12-bit counter per pixel, shown in Figure 3.10, and additional 4 bits (or more) generated while correcting for the count suppression mechanism of active event-driven recharge (see sub-section 2.2.4.1). A 65536:1 dynamic range will be achieved using T_{dead}=1 µs. To have a maximum of 2^{12} counts, the frame time is 4.1 ms, frame rate 244 fps. The chip output data rate is 2.9 Gbps. Although the recharge is event-driven, per pixel counters enable fast accumulation with event-driven increment and clock-wise readout. Thus the readout is clock based. The recharge period or hold-off should be entangled with the readout speed by: $T_{readout}=T_{dead}(2^{\text{bits}}-1)$. To enable a faster frame rate with the same imager, we include an additional readout mode with 8-bit counter resolution. In this mode, the maximum frame rate is 3.9 kfps, and the output data rate 31.2 Gbps. All the electronics would be designed in the bottom tier IC, while the SPAD would be placed in the top tier. Abbas *et al.* demonstrated 12-bit counters in 40 nm CMOS with a pixel pitch lower than 10 µm. We enforce the SPAD array power consumption to be below 500 mW, the V_{OP} voltage drop below 100mV. We set the V_{OP} to 25 V, with an excess bias V_E of 5 V.

The SPAD capacitance C should be thus smaller than $P/(V_E^2 fwh)= PT_{dead}/(V_E^2 wh)$=19 fF. This capacitance sets the SPAD diameter/pitch to lower than 9.5 µm, considering a typical capacitance of 190 fF for a 30 µm photodiode pitch[23]. With a maximum power of 500 mW, the SPAD array average current consumption is 100 mA, thus the maximum per pixel consumption is 95 nA. Given the required average voltage drop of below 100 mV, the per pixel parasitic resistance needs to be $R_P<\Delta V/(w(w+1)I_A)\times 2$=2 Ω (see equation (3.3)), for SPADs operating simultaneously. For the peak IR drop of 100 mV, we take into account the peak current consumption from section 3.1 (V_E/R_Q), to be 33 µA. This requires $R_P<\Delta V/(wI_A)\times 2$= 6 Ω. Since CMOS metals feature a resistance below 50 mΩ/square, the parasitic resistance requirement should be fulfilled. Figure 3.10 shows a top level schematic of the proposed SPAD imager.

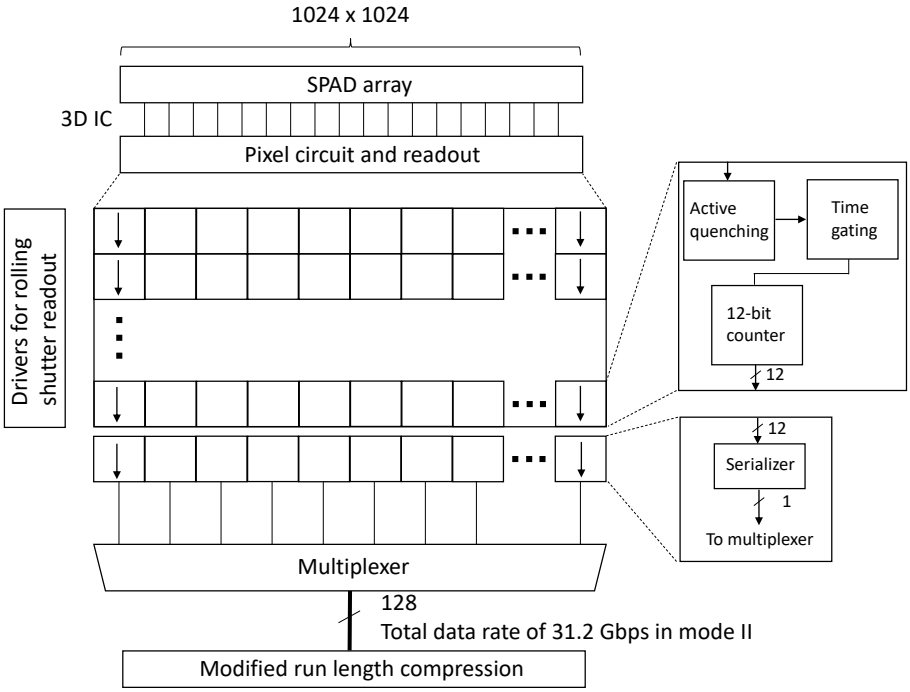


Figure 3.10. Proposed architecture of a 12-bit counter 1024×1024 SPAD array with extended dynamic range due to count suppression. If the dead time of the SPADs is set to 1 µs, 12 bits yield 244 frames per second, while the QIS adds 4 bits (totally 16 bits). Mode II enables a lower dynamic range at an higher frame rate; 8 bits frames are now send with a frame rate of 3.9 kfps.

Table 3 compares two most recent SPAD imagers with specialized CMOS imagers and the proposed SPAD imager. With new 3D integration possibilities, SPAD imagers offer the opportunity to design high dynamic range (counters in bottom tier) imagers with no readout noise. Trading dynamic range (reducing counter bits) for a higher frame rate does not change the SPAD performance. The PDP will advance further reaching custom SPAD designs with PDP>70%. The fill factor already reached 45% due to 3D IC integration. SPAD imagers, however, have a fundamental disadvantage in higher DCR and DCR nonuniformity. They are thus not suitable for long exposures.

CMOS imagers are suitable for longer exposures due to lower DCR. PDP is approaching >90% and fill factor is compensated with microlenses. With shrinking CMOS nodes, readout noise will fall further below 0.27 e^- peak rms[103], but also reduce the full well capacitance. Reducing noise requires lower speeds. SPAD can thus seek competitive advantage in combining multiple modes (HDR, no readout noise, high speed) into one imager. CMOS imagers have a

fundamentally higher optical dynamic range due to the multi-photon full well capacity (which can be reset at very high speeds) and low DCR.

Table 3. A state-of-the-art comparison of most recent SPAD imagers with three CMOS imagers specialized for high speed, high dynamic range and low noise. The last column describes a proposed SPAD imager for high speed and high dynamic range (dual mode).

Camera	Digital counters[73]	512x512 1-bit SPAD[75]	CMOS high speed[62,104]	CMOS HDR[105]	CMOS low noise[33]	Proposed SPAD imager
Peak PDP	27.5%	>50%	unknown	unknown	>75%	>25%
Fill factor	45%	10.5%	unknown	74%	40%	>45%
Pixel pitch	7.83 µm	16.38 µm	32 µm	11 µm	6.5 µm	<10 µm
Array size	128×120	512×512	400×256	2048×2048	640×480	1024×1024
Max. frame rate	500 fps, 12 bit	100 kfps, 1 bit	7.8 kfps	24 fps	80 fps	244 fps[d], 3.9 kfps[e]
Data rate	92 Mbps	26.2 Gbps	11.1 Gbps[a]	1.6 Gbps[b]	246 Mbps	2.9 Gbps[d], 32 Gbps[e]
Dynamic range	11 066	11 817[g]	2 000	63 095	13 335	65 536[d,f], 4 096[e,f]
Readout noise	N.A.	N.A.	5 e^-	1.47e^-	0.48 e^-	N.A.
Dark counts	11 kcps	50 cps	unknown	0.15 cps	5.6 cps	<100 cps
Afterpulsing	unknown	<1%	N.A.	N.A.	N.A.	<1%
Crosstalk	unknown	<0.5%	unknown	unknown	unknown	<0.5%
PRNU	< 2%	unknown	unknown	<0.1%	0.77%	< 2%
Min. gating duration	1 ns	4 ns	N.A.	N.A.	N.A.	-

[a]Assuming a 14 bit ADC

[b]Assuming a 16 bit ADC

[d]12 bit counter with 1 µs dead time

[e]8 bit counter with 1 µs dead time for high speed imaging

[f]Dynamic range is here defined by the counter bit width and conservatively extended due to count suppression by 16; DR=16×2^{bits}. See section 2.2.5.1 for details.

[g]In 1-bit SPAD imagers $4.6t/T_{readout}/\sqrt{(tDCR)}$, 4.6 for the QIS extention[18], t=33 ms, see (2.21)

3.4 Conclusion

We have discussed different nonuniformity issues and included hand-calculations for power consumption, IR drop, and data rates.

The imaging architecture needs to be tailored for specific applications. Quantum correlation applications usually feature a low detection count rate and can benefit from event-driven readout. The same holds for fluorescence lifetime measurements. Classical imaging has no benefit in using the event-driven approach due to medium/high (and relatively uniform) illumination scenes. The photon probability needs to be very low to yield a reduction in data generation. Considering data reduction, 3D imaging also has little benefit from an event-driven approach given the high background illumination.

Event-driven readout is architecturally interesting for sharing resources, but implementing per pixel counters substantially reduces the data rate by increasing the timing denominator in the data rate equation.

Applications with timing distributions, like 3D imaging and fluorescence lifetime, can benefit from histogramming if the average number of pixel events is high. In our example, there should be more than 1024 events per pixel for histogramming to yield any data reduction. Scenes with large amounts of dark sections can actually increase the data amount with histogramming, especially when the accumulation time is short (yields less than 1024 events per pixel). For example, a Picoquant scanning fluorescence lifetime imaging system does not histogram data due to short per-pixel accumulations and large areas of the image that do not contain any fluorescence. Such systems usually collect less than 1000 events per pixel, and histogramming increases the data rate. However, we should emphasize that the reason for a short per-pixel accumulation time; scanning decreases the frame time. Thus, if one can use a pixel array, the per-pixel accumulation can be prolonged to collect more events and histogramming might become an effective tool for data reduction.

3D IC integration will expand the (analog-less) digital potential to combine single photon with high speed and extended dynamic range. New SPAD CMOS technology enables combining multiple modes into one imager, with HDR, photon counting, no readout noise and high speed. One such imager is proposed in this chapter.

4 Large format SPAD imagers for localization super resolution microscopy

Microscopy resolution is conventionally limited by diffraction. In the last 25 years, numerous techniques emerged to surpass this limitation[3]. Three main approaches are[106]:

1) Structure illumination microscopy (SIM)
2) Simulated emission depletion microscopy (STED)
3) Single molecule localization microscopy (SMLM)

Demmerle *et al*. conducted an insightful spatial frequency comparison between the three different super resolution aproaches[107].

SIM uses the Moiré effect to alias higher spatial frequencies and effectively increases the spatial resolution[108,109]. This approach generates images with ~100 nm resolution after sub-second to second acquisition time. The samples do not need to be specially prepared to apply this technique.

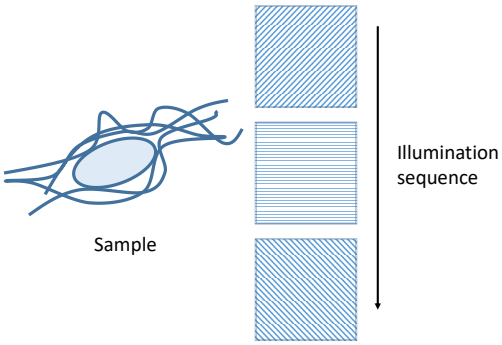


Figure 4.1. Example of SIM image creation.

STED uses two lasers to effectively narrow the PSF. The outer part of the emission (generated by an excitation laser) is depleted by a donut shape laser[110–112]. An example of STED PSF engineering is shown in Figure 4.2. One can achieve a lateral spatial uncertainty as low as 20 nm[113], but at a cost of a high power laser that induces sample bleaching (see Figure 3 by Heller *et al.*[114] for typical numbers). STED is a scanning technique; the acquisition time can be in the range of minutes (higher depletion power, higher resolution, and lower photon rate) due to the *depleted* photo rate (see Figure S4 by Gottfert *et al.*[115] for typical photon numbers). Although STED achieves a high resolution, high

spatial frequencies have lower amplitudes when compared to other super resolution techniques[107].

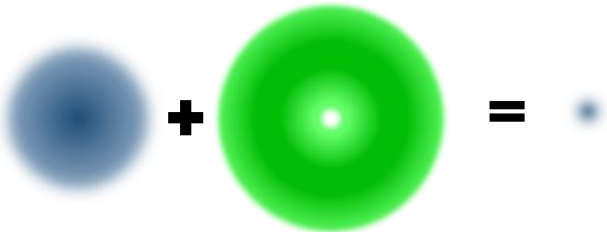


Figure 4.2. STED emission is depleted by a donut shape laser effectively narrowing the PSF and achieving high spatial resolution.

SMLM sparsely activates single molecules, fluorophores, within one frame time. Single molecule PSFs are then fitted using a Gaussian 2D model, and the positions localized. Combining multiple frames with a large amount of *localized single molecules,* we create a super resolved pointillistic image. The sample needs to be prepared in a special way to allow for sparse activation of single molecules. The amount of required frames is 10^3-10^5, and the acquisition time spans to minutes[116]. sCMOS can reduce this to 10 seconds. Within SMLM, molecules are sparsely activated with different techniques: PALM[117,118], STORM[119], dSTORM[5], GSDIM[4]. We used the GSDIM technique in this chapter.

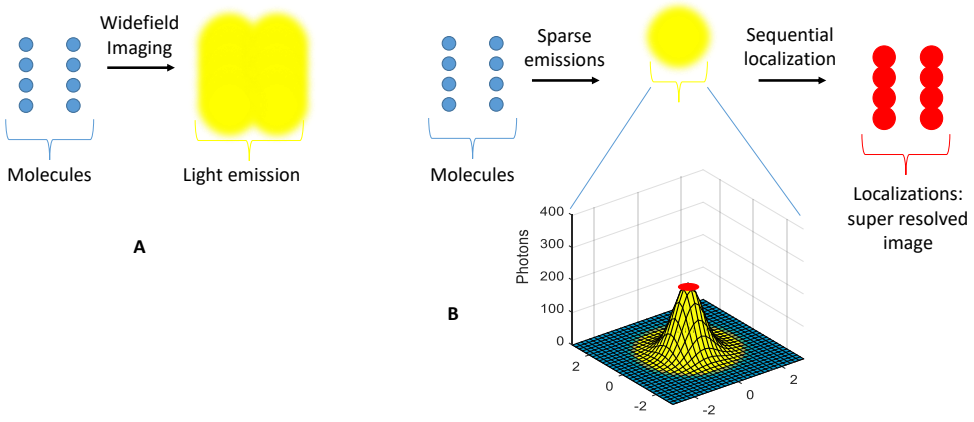


Figure 4.3. An example of SMLM. In classical widefield microscopy (A), all fluorescent molecules emit light simultaneously. In localization super resolution (B), molecules are sequentially (or sparsely per frame) activated to emit light and localized. Grouped localizations form a pointillistic super resolved image.

Some of the fundamental advantages of SPAD imagers discussed in chapter 2 are employed in SMLM, which is a sensitivity critical application. Section 4.1

problematizes the long acquisition time in localization super resolution. The fill factor improvement allowed us to acquire the first super resolution images with a SPAD imager (section 4.2) and add photo-physical information usually not available with other imagers. We show that by overcoming the fill factor drawback and increasing the overall sensitivity, we could use the high imaging speed of a SPAD imager, SwissSPAD, to investigate photo-physical properties of fluorophores while recording super resolution images (section 4.3). We then show two possible super resolution enhancements: using a saturation method to extract fluorophore lifetime (section 4.4) and calculating quantum correlations to allow for denser fluorophore activation (section 4.5).

The experimental setup (shown in Figure 4.4) is composed of a dual port Leica SR GSD super resolution microscope (Leica Microsystems, Wetzlar, Germany), and initially an Andor iXon3 897 BV EMCCD (Andor Technology, Belfast, UK) with a pixel pitch of 16 µm on one port and SwissSPAD with a pixel pitch of 24 µm on the other. We later used a pco.edge 4.2 sCMOS imager (PCO, Kelheim, Germany) in combination with SwissSPAD. The emission light can be directed to either the EMCCD/sCMOS or SwissSPAD. The objective and the tube lens (HCX PL APO 160×/1.43 Oil CORR GSD) magnification is 160×, and the effective EMCCD and SwissSPAD pixel sizes are therefore 100 nm and 150 nm, respectively. When using sCMOS, a demagnificator was used to yield 100 nm pixel size.

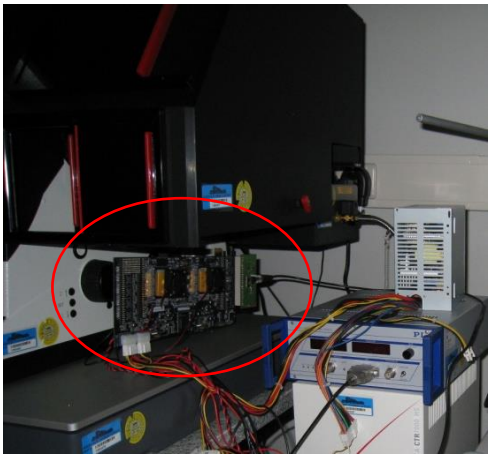


Figure 4.4. Experimental setup with SwissSPAD on the left hand side of the image, connected the Leica SR GSD microscope.

For SMLM measurements, we used human bone osteosarcoma epithelial cells (U2OS) stained for microtubuli with Alexa 647, in Vectashield (Vector Laboratories, Burlingame, USA) embedding resin or MEA buffer[120], as well as

fibroblast cells stained for actin with Alexa 647 together with an OxEA buffer[121], and GATTAquant nanorulers[122,123] (GATTAquant, Braunschweig, Germany).

4.1 Speed of super resolution acquisition

A major drawback of SMLM is the long acquisition time. During 10^3 to 10^5 frames, we localize between 5 and 100 molecules per frame and create a pointillistic image with 50 000 to 1 000 000 total localizations. Lin *et al.* showed that higher laser intensities increase the molecule blinking rate and suggest that it increases the frame rate and potential number of active (emitting) molecules per time (and still keep the spatial sparsity requirement). However, blinking speed and frame rate should be synchronized so achieve optimum results. Higher frame rates may decrease the acquisition time at a cost of suboptimal performance. Here, we investigate the frame rate speed given and certain blinking distribution.

4.1.1 Optimal frame time

When increasing the speed of the molecule blinking to achieve fast SMLM acquisitions, it is not *a priori* clear what frame duration should be used.

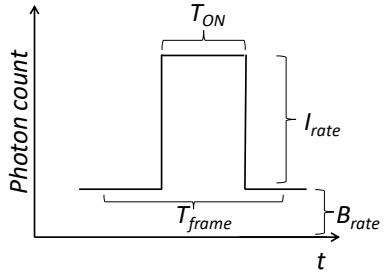


Figure 4.5. Schematic representation of blinking of a fluorophore in time. B_{rate} includes the imager noise and sample background noise.

Figure 4.5 shows a fluorophore photon count rate I_{rate} in the presence of a background B_{rate}, corresponding to the background accumulated by all pixels, which are contributing to I_{rate}. Note that more than one pixel contributes to the PSF. The blinking time is assumed to be T_{ON} and is here assumed to be entirely embedded in a frame. In the general case, the SNR as a function of T_{frame} is computed as:

$$SNR = \begin{cases} \dfrac{I_{rate}T_{frame}}{\sqrt{I_{rate}T_{frame} + B_{rate}T_{frame}}}, \text{ if } T_{frame} \le T_{ON}, \\ \dfrac{I_{rate}T_{ON}}{\sqrt{I_{rate}T_{ON} + B_{rate}T_{frame}}}, \text{ if } T_{frame} > T_{ON}. \end{cases} \quad (4.1)$$

For a constant T_{ON} value, the maximum SNR is reached for T_{frame}=T_{ON}. This changes when T_{ON} is an exponential random variable with a decay constant τ_{ON}, which is the case for the emission time of the fluorophores. First, the estimated *E(SNR)* with the above assumption for the probability density function of T_{ON}, $pdf\left(T_{ON}\right) = e^{\frac{-T_{ON}}{\tau_{ON}}} / \tau_{ON}$, is calculated as:

$$E(SNR) = \int_{-\infty}^{\infty} SNR(T_{ON})\, pdf\,(T_{ON})\, dT_{ON}$$

$$= \int_{0}^{T_{frame}} \frac{I_{rate}T_{ON}}{\sqrt{I_{rate}T_{ON} + B_{rate}T_{frame}}}\, pdf\,(T_{ON})\, dT_{ON} + \int_{T_{frame}}^{\infty} \frac{I_{rate}T_{frame}}{\sqrt{I_{rate}T_{frame} + B_{rate}T_{frame}}} pdf\,(T_{ON})\, dT_{ON}$$

$$= \sqrt{T_{frame}B_{rate}} - e^{-\frac{T_{frame}}{\tau_{ON}}} \left(\frac{B_{rate}\sqrt{T_{frame}}}{\sqrt{I_{rate} + B_{rate}}} \right)$$

$$+ e^{\frac{T_{frame}B_{rate}}{\tau_{ON}I_{rate}}} \left[erf\left(\sqrt{\frac{T_{frame}(B_{rate} + I_{rate})}{\tau_{ON}I_{rate}}} \right) - erf\left(\sqrt{\frac{T_{frame}B_{rate}}{\tau_{ON}I_{rate}}} \right) \right] \left[\frac{\sqrt{\pi I_{rate}\tau_{ON}}}{2} - \frac{\sqrt{\pi}B_{rate}T_{frame}}{\sqrt{I_{rate}\tau_{ON}}} \right]. \quad (4.2)$$

An example of SNR as a function of T_{frame} when T_{ON} is an exponential random variable with a typical decay constant τ_{ON} =10 ms is shown in Figure 4.6.

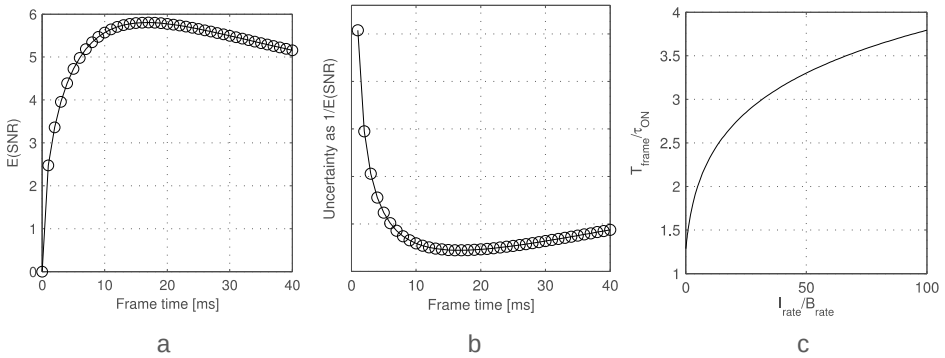


Figure 4.6. (a) SNR of blinking emission as a function of T_{frame}, with $I_{rate}/B_{rate}=2$ and $\tau_{ON}=10$ ms. SNR_{max} is reached in this case for $T_{frame}=1.67\times\tau_{ON}$. (b) The inverse of the SNR is proportional to the expected localization uncertainty. It represents an approximation from the result obtained in[124]*, where the uncertainty is proportional to the inverse square root of the number of photons in the PSF for the shot noise limited case ($B_{rate}<<I_{rate}$), and to the inverse of the number of photons for the background noise limited case. (c) Optimal T_{frame}/τ_{ON} ratio as a function of the I_{rate}/B_{rate} ratio. T_{ON} is an exponential random variable with a decay constant τ_{ON}. For measurements with lower background, the maximal SNR is reached at longer times, to cover long emissions. For measurements with low fluorophore emission intensities, the maximal SNR is reached close to the average emission duration τ_{ON}.*

The optimal T_{frame} is found solving:

$$\frac{d\left(E(SNR)\right)}{dT_{frame}} = 0\,, \tag{4.3}$$

which does not yield a close-form T_{frame}. The optimal T_{frame} is a function of I_{rate} and B_{rate} and can be seen in Figure 4.6c. If B_{rate} = 0, T_{frame} should be chosen so that T_{frame} > max(T_{ON}), however, since T_{ON} is a random variable, the optimal T_{frame} should be infinite. On the other hand, for B_{rate} > I_{rate}, T_{frame} should be chosen so as to minimize the noise effect, thus T_{frame}= τ_{ON}.

SwissSPAD achieves a short frame time without negative effects on the imager performance, such as clock induced charge noise and readout noise. The fast frame time can be used to analyze the optimal T_{frame}. To the best of our knowledge, this analysis has not been carried out before, also due to the fact that the performance of charge accumulating imagers changes as a function of readout speed. Figure 4.7a and Figure 4.7b show ThunderSTORM[125] super resolution results of simulated data with τ_{ON}=10 ms. The background per pixel is monotonically increasing with increasing T_{frame}. The fluorophore intensity is also increasing but starts saturating after $T_{frame}>\tau_{ON}$. The number of

localizations decreases rapidly until $T_{frame}=\tau_{ON}$ (Figure 4.7a), while the uncertainty curve reaches its minimum at $T_{frame}>\tau_{ON}$ (Figure 4.7b). The result of Figure 4.7b is in good agreement with the plots of Figure 4.6b. Figure 4.7c and Figure 4.7d show experimental SwissSPAD data. The experimental curves approximates the simulated curves, although there is an additional effect of rejecting dim emissions (and partial merging of localizations) causing an artificial uncertainty decrease. Merging close localizations in consecutive frames had insignificant influence on the results presented in Figure 4.7, since ThunderSTORM averages the positions, imaged sizes, and backgrounds, and adds the intensity of the individual localizations. Adding photons from multiple frames and then calculating the position and uncertainty yields a decrease in uncertainty[126], but it requires intense processing, currently not implemented in super resolution software. Oversampling with SPAD imagers (no negative effect due to readout noise) could be additionally effective due to per-molecule-blink minimization of background noise, which is experimentally found to range between 5 and 100 photons (Poisson average). If background noise is minimized, emission Poisson noise is the dominant noise source. Gyongy *et al*. demonstrated a 20% reduction in localization standard deviation with experimental data on GATTAquant-PAINT nanorulers[92]. Simulated results suggest a reduction of 50%[92]. During the work of this thesis, we experimented with similar algorithms on experimental cell data. We could not achieve an increase in spatial resolution. In addition to the differences in the algorithms, a thicker cell sample (in comparison to the flat and relatively sparse nanorulers) might complicate the sectioning both in space and time.

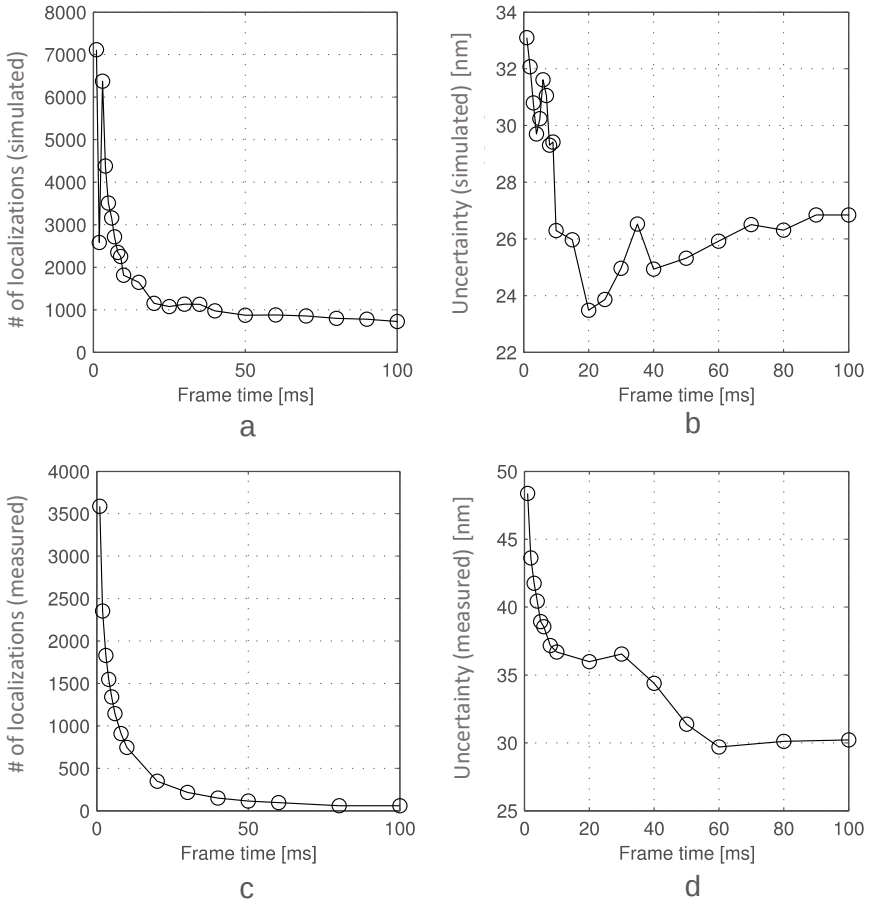


Figure 4.7. (a) Simulated data show how the number of localizations decreases with increasing frame time. With short frame times, the emissions are oversampled and single emitters detected as multiple ones. With longer frame times the emissions begin to be rejected. (b) The localization uncertainty reaches a minimum of 24 nm at 20 ms. τ_{ON} was set to be 10 ms, and I_{rate}/B_{rate} was to 3.6, yielding a theoretical optimum between 18 and 19 ms. The operating point also represents a tradeoff between the number of localizations and the localization uncertainty. (c) Measured number of localizations in SwissSPAD over different frame times (1, 2, 3, 4, 5, 6, 8, 10, 20, 30, 40, 50, 60, 80, and 100 ms). (d) The localization uncertainty starts saturating to reach around 35 nm after T_{frame}=10 ms, but decreases again artificially below 35 nm, most likely due to localization rejections.

Figure 4.8 shows simulated SMLM results on emissions in a circle using three different T_{frame}, where the theoretical optimum is between 18 and 19 ms. Short frame times result in larger number of localizations (also faulty ones) with high localization uncertainty, while long frame times result in localization rejection and worse localization uncertainty.

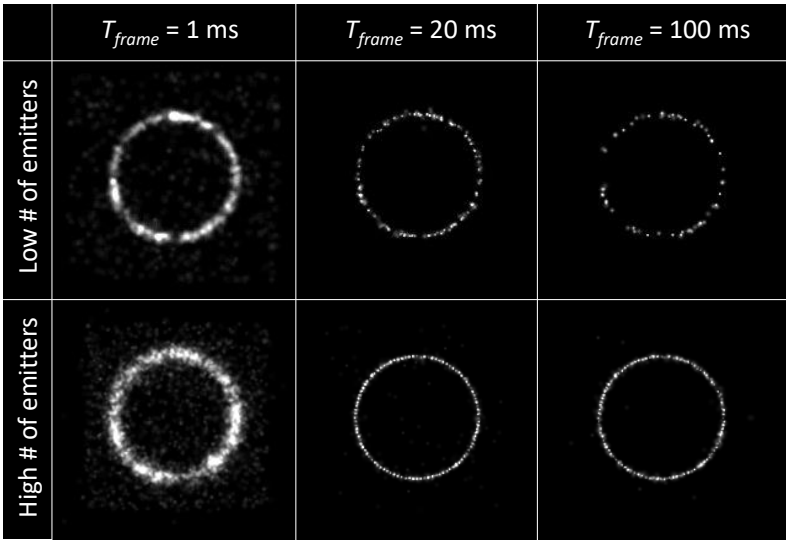


Figure 4.8. Simulated super resolution images of molecules randomly blinking at positions in a circle, obtained with different frame times. τ_{ON}=10 ms. If the blinking is oversampled, the uncertainty is high. If the blinking is undersampled (100 ms), the image will both yield a lower number of localizations (Low # of emitters) and higher localization uncertainty (High # of emitters). A T_{frame} of 20 ms represents the optimum. I_{rate}/B_{rate} was set to be 3.6, yielding a theoretical optimum between 18 and 19 ms.

4.2 First SPAD super resolution images

Figure 4.9 presents the first SPAD localization super resolution images, presented in 2015[15]. We imaged U2OS cells stained with Alexa 647 embedded in a Vectashield resign, and compared the EMCCD and SPAD results. The key was implementing microlenses that increased the overall SwissSPAD PDE by 12×.

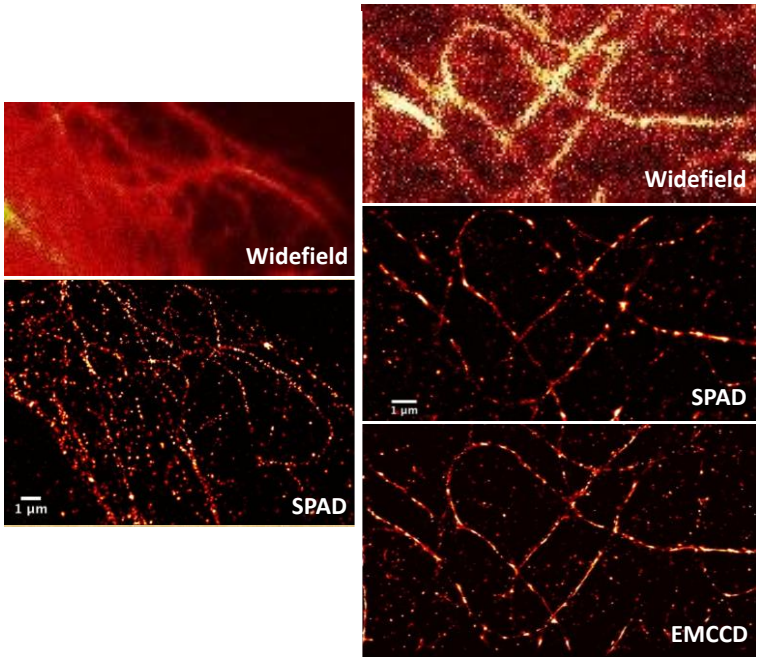


Figure 4.9. First SMLM results acquired with a SPAD imager, compared to EMCCD widefield and super resolution images. The SPAD imager field of view is rotated by few degrees compared to the EMCCD due to mechanical mounting restrictions.

The super resolution community emphasized two main EMCCD drawbacks, the multiplication noise (see section 2.2.2.4) and limited speed (see section 2.2.6). This led to multiple publications about sCMOS used in SMLM[9,127–130], and the main microscopy companies switched to sCMOS due to lowered readout noise, increased noise uniformity and sensitivity in the newest sCMOS models. When comparing SwissSPAD and EMCCD/sCMOS we alternately recorded frames with the two imagers. To compare SwissSPAD and sCMOS, we also used GATTAquant PAINT 80R nanorulers, where three emitters in each nanoruler are separated by 80nm[123]. The SwissSPAD has a PDE of around 9% at the emission wavelength of Alexa Fluor 647 and Atto 655.

The SwissSPAD video was first pre-processed in MATLAB to correct for the sensor's nonlinear photon response and DCR, on a pixel-by-pixel basis, employing the following correction scheme[14]:

$$C = \left[-\ln(1 - C_M T_{readout}) / T_{readout} \right] - \mathrm{DCR}, \tag{4.4}$$

where C represents the corrected count rate, C_M the measured count rate, and $T_{readout}$=6.4 µs the dead time of the pixel.

The recorded 1-bit, 6.4 μs frames were then binned to form 10 ms frames, background was subtracted[131] and images were analyzed with ThunderSTORM[125]. Figure 4.10a shows a super resolved image of actin in a fibroblast sample, stained with Alexa Fluor 647, used with an OxEA buffer. Figure 4.10b shows the corresponding widefield image. Figure 4.10c-f shows GATTAquant PAINT 80R nanoruler data of SwissSPAD and sCMOS, respectively. Images where reconstructed using 5000 frames of 10 ms duration. When zoomed, one can note the resolvability of 80 nm distances between single emitters, where sCMOS has a 2-3 times finer resolution. Gyongy *et al.* showed the benefits of aggregating short frames with a single GATTAquant nanoruler imaged with a SPAD imager[92,126], but without super resolution images.

Figure 4.10g-i show microtubuli from a fibroblast sample, which was stained with Alexa Fluor 647, and prepared in OxEA buffer. The sCMOS and SwissSPAD images contain 40 000 localizations. Although the effect is minimized by preprocessing (see Eq. 8), some of the noisy pixels distort localizations. Highly noisy pixels, known as screamers, constitute 2% of the overall pixel population[14]; interpolation is used to minimize localization distortions. The interpolation of hot pixels caused loss of data and a localization uncertainty increase. This effect was verified by the software, resulting in a bias of the localization position (away from the true position).

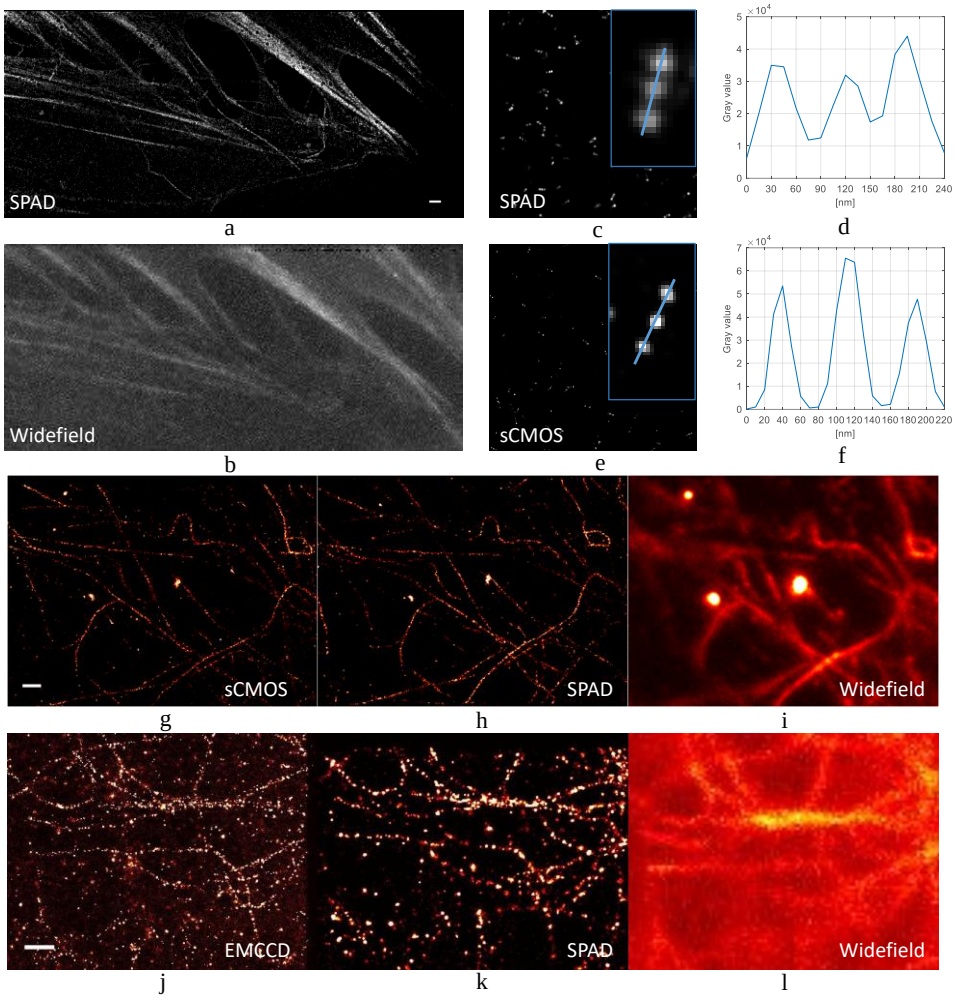


Figure 4.10 (a) SwissSPAD super resolution image of actin labeled with Alexa 647 in an OxEA buffer compared to (b) the widefield image. The white bar shows 1 µm. (c,d) Shows SwissSPAD GATTAquant PAINT 80R nanorulers where emitters are separated by 80 nm. (e,f) Shows the nanoruler imaged with sCMOS. (g) sCMOS and (h) SwissSPAD super resolution images of microtubuli compared with (i) a widefield image taken with a sCMOS. The white bar shows 1 µm. (j-l) Show an U2OS cell stained with Alexa Fluor 647, in Vectashield.

Figure 4.10j-l show a SwissSPAD super resolution image of microtubuli in an U2OS cell stained with Alexa Fluor 647, in Vectashield. It was obtained during a 70 second exposure and contains 90 000 localizations.

The sCMOS typically collected 800 photons with a localization uncertainty of 10 nm (for Figure 4.10g), while SwissSPAD had 100 photons collected with 20 nm uncertainty (for Figure 4.10h). For the measurements with EMCCD, the typical

estimated number of collected photons was 1800, and the typical estimated localization uncertainty about 15 nm (Figure 4.10j), while SwissSPAD had 200 photons collected and 30 nm uncertainty (Figure 4.10k). The EMCCD and SPAD distributions are shown in Figure 4.11. The emphasis should be placed on the comparison between the SPAD and one of the two other imagers. A direct comparison between the sCMOS and EMCCD images and uncertainty results is not completely fair since the sCMOS imaged samples with OxEA, and the EMCCD imaged samples with Vectashield. The localization uncertainty is estimated using ThunderSTORM, employing the Thompson *et al.* formula[124] for sCMOS and SPAD, and Quan *et al.* formula[130] for EMCCD (as to include the multiplication noise).

It is worth mentioning that, although the EMCCD collected 10 times more photons, the localization uncertainty is only about twice better, while 10 times more photons should yield $\sqrt{10}=3.16$ better localization uncertainty, but the excess noise lowers this by a factor of $\sqrt{2}$[9], resulting in $\sqrt{5}=2.23$ better localization uncertainty. CMOS SPAD structures reaching a PDP of 40% between 440 and 620 nm have been published[45], but not yet implemented as SPAD imagers. A theoretical analysis did actually show that SPAD imagers with the same sensitivity as EMCCD and sCMOS imagers will feature superior localization accuracy because of absence of excess and readout noise[132], assuming that the dark noise uniformity is similar to conventional CMOS imagers.

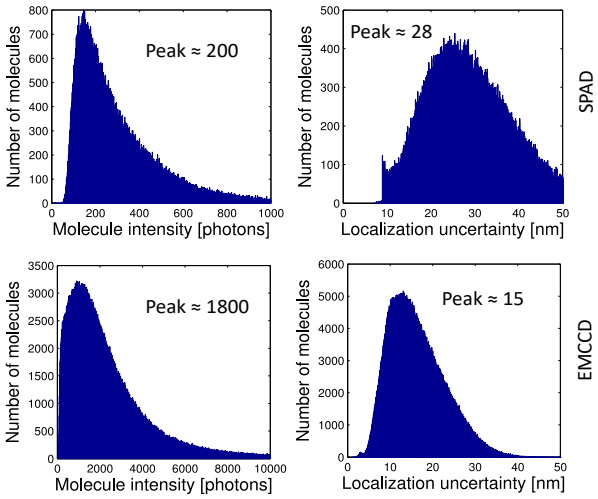


Figure 4.11. The distribution of the number of collected photons (per molecule blink) and localization uncertainty for SPAD and EMCCD.

4.3 Blinking analysis of fluorophores

Figure 4.12 shows measured examples of molecule emissions with diversity in emission duration and intensity. Even two locally identical emissions can show different intensities. One should note that the event extraction is a complex process that should be ideally performed in space *and* time domain. A lower molecule intensity may be actually caused by a position offset. Thresholding just in the time domain can thus lead to false blinking duration extraction. On the other hand, a space extraction of localization forces a longer frame time that integrates more photons and eases distinguishing the molecule from the background. A space *and* time extraction would start with space localization, using a long frame time. It would than localize the molecules and cut the subspaces (usually 3×3 or 5×5 pixels). Then, one would use oversampled frames (1-bit SwissSPAD frames) to perform dynamical time binning. The shortest frame time would be limited by the shot noise limited signal-to-noise ratio. The background highest count number should be lower than the molecule lowest photon number. We used this procedure for the results presented at Photonics West in 2016[16], while Gyongy *et al*. published an arxiv paper later in 2016 on a similar procedure with a mathematical expression of the exact number of 1-bit frames that should be binned in time[126].

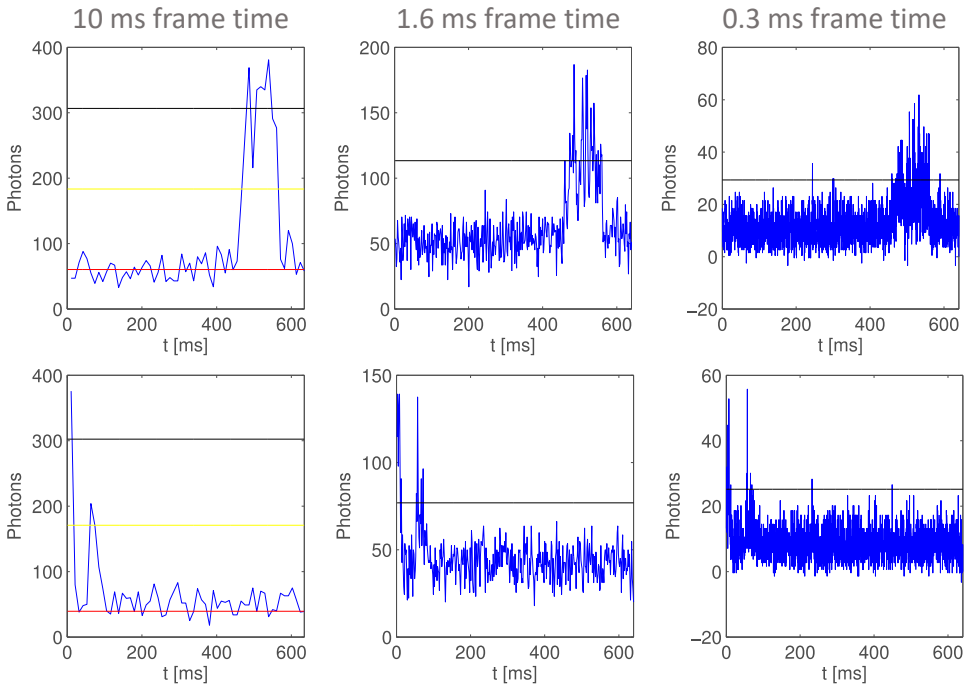


Figure 4.12. Examples of measured fluorophore blinking activity at different frame times. The red line indicated the average background level, the black shows the average estimated blink level. The yellow line in the first column and the black line in the second and third column represent the proposed threshold level.

Estimating blinking duration is also important if one wants to use a fixed frame rate, but optimized given a blinking distribution (like discussed in section 4.1.1). Figure 4.13a and Figure 4.13b show examples of blinking without and with additional fast blinking.

To estimate the *on* and *off* averages as well as the corresponding emission and background bands (see Figure 4.13a), we first used the whole signal sample length to estimate A, where:

$$\min(signal) = \lambda_1 - A\sqrt{\lambda_1} \tag{4.5}$$

and thus λ_1, i.e. the average value of the background noise as random variable with Poisson distribution, where:

$$\sigma_{background} = \sqrt{\lambda_1} \tag{4.6}$$

Parameter *A* can be chosen to be 3 and includes the standard 3σ of a Poisson distribution. Note that the whole signal includes background and emission signals, and that min(*signal*) and max(*signal*) are easily found. The emission sample length, i.e. the total length of the signal above max(*background*), is then used to find *B*, where:

$$\max(signal) = \lambda_2 + B\sqrt{\lambda_2} \tag{4.7}$$

and thus compute λ_2, i.e. the average value of the emission signal and background noise as random variable with Poisson distribution.

In the presence of additional fast blinking (due to a triplet state or additional dark states), the photon response will indeed not follow Poisson statistics with a constant photon rate, as it is clear from Figure 4.13b. Additional fast blinking makes the time thresholding more complicated (Figure 4.13b). The binning and thresholding expression that Gyongy derived[126] cannot be used in this case. One needs to use a longer frame time to achieve background and molecule signal separation.

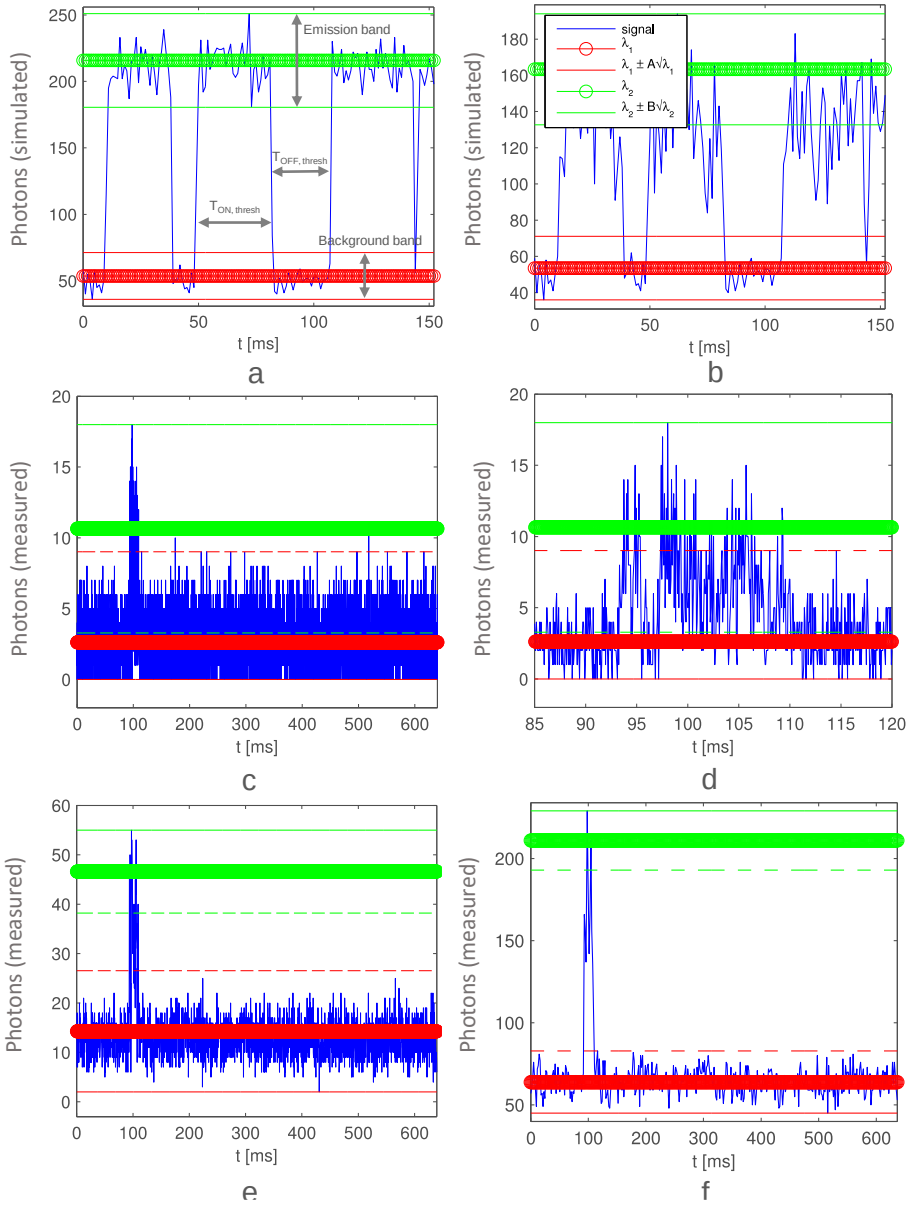


Figure 4.13. Stochastic blinking, where the molecule switches between an on and an off state: simulated photon response without fast blinking (a); simulated photon response with fast blinking, where the signal exhibits many outliers situated outside the estimated Poisson band (b). SwissSPAD-measured molecule blinking with 0.064 ms (c, d), 0.3 ms (e), and 1.6 ms (f) frame time. In particular (d) shows the zoomed blinking, at around t=100 ms, with 0.064 ms frame time. Note that the emission and background bands overlap in (c, d). We used a U2OS sample with MEA buffer. Also note that if additional fast blinking is present, the photon response may be larger than expected from Poisson statistics alone. The upper and lower estimated boundaries for the photon response are marked with thin green (emission) and red

(background and noise) lines, while the estimated averages are shown with thicker lines.

Examples of measured blinking with different frame times, with estimated emission and background bands, are shown in Figure 4.13c-f. From the figures it can be ascertained that emissions indeed have two (or more) blinking components: slow and fast. The number of emitting fluorophores per frame is sparse due to the slow component ($\tau_{OFF,thresh}$ and $\tau_{On,thresh}$ in Figure 4.13a). The fast component is not an application necessity, and is induced by the triplet state or other dark states. The additional fast blinking can lower the overall molecule emission.

To analyze such a fast blinking phenomena, shorter frame times must be used. However, due to the decrease of frame time, the SNR is reduced as well, leading to the overlapping of signal and background bands. One can use two methods for molecule blinking analysis: a thresholding method, which is suitable for signals with two distinguishable signal intensities (molecule *on* and molecule *off*), where the bands are well separated, and an autocorrelation method, which can be used for continuous signals, whenever the SNR is not high enough to clearly distinguish *on* from *off* states[133].

We recorded images with three different laser intensities and firstly used the thresholding method to extract $\tau_{ON,thresh}$ and $\tau_{OFF,thresh}$ of the slow blinking component. Fitted monoexponential distributions for the *on* time have $\tau_{ON,thresh}$=6.1 ms, $\tau_{ON,thresh}$=3.1 ms and $\tau_{ON,thresh}$=2.0 ms for laser intensities of 5.7 kW/cm^2, 8.5 kW/cm^2 and 11.3 kW/cm^2 respectively. $\tau_{ON,thresh}$ decreases with increase of laser intensities. Although *off* times over a wide range of laser intensities should be analyzed as the sum of three exponential distributions[10], we fitted *off* data monoexponentially as an indication of changes over different laser intensities. The distributions for the *off* time have $\tau_{OFF,thresh}$=13.4 ms, $\tau_{OFF,thresh}$=5.7 ms and $\tau_{OFF,thresh}$=3.3 ms for laser powers of 5.7 kW/cm^2, 8.5 kW/cm^2 and 11.3 kW/cm^2 respectively. The threshold was set at the upper noise bound because of possible multiple emissions with different photon intensities and additional fast blinking (see Figure 4.13). Single outliers over the upper noise bound were rejected. A fast sCMOS camera can extract $\tau_{ON,thresh}$ and $\tau_{OFF,thresh}$ in the range of 1-10 ms, thus faster blinking (in the μs range) cannot be properly investigated, as in SwissSPAD.

Due to variations in the fast blinking behavior (data presented later), a fixed thresholding expression is not always feasible. We thus used space localization

with the ThunderSTORM program. The frame time was fixed to 1.6 ms, composed of 255 (2^8-1) 1-bit frames. The 8-bit data was formed by accumulating the original 1-bit data externally. We then returned to the 1-bit data with 6.4 μs time resolution and used autocorrelation to explore if additional fast blinking is present. Since we expect the blinking to be exponentially distributed, the autocorrelation curve is fitted to extract a decay constant of the exponential autocorrelation curve τ_{auto}. The drawback of using 1-bit data is that the emission and noise bands (as shown in Figure 4.13) are not easily distinguishable and τ_{auto} will be a combination of the fast *on* and *off* lifetimes (if present), i.e.:

$$\frac{1}{\tau_{auto}} = \frac{1}{\tau_{ON,auto}} + \frac{1}{\tau_{OFF,auto}} \tag{4.8}$$

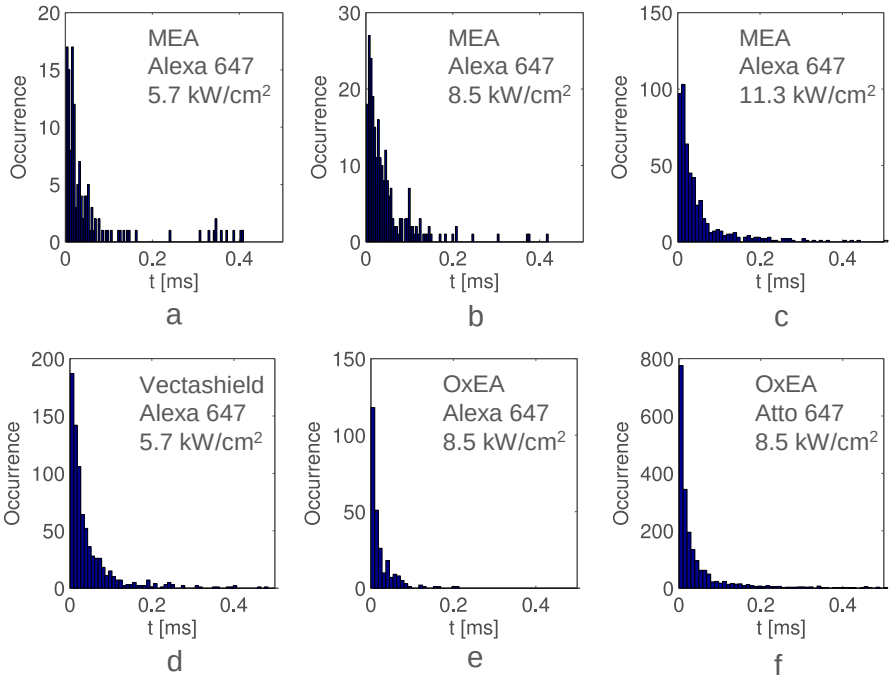


Figure 4.14. SwissSPAD measured autocorrelation decay distributions for different laser intensities of 5.7 kW/cm² (a), 8.5 kW/cm² (b) and 11.3 kW/cm² (c), of Alexa 647 with MEA buffer. (d) Shows the distribution of Alexa 647 with Vectashield. Plots (e) and (f) show the distributions of Alexa 647 and Atto 647 with OxEA buffer, respectively.

In our U2OS sample with microtubuli stained with Alexa 647 and with MEA buffer, we measured 68%, 74% and 70% emissions with $0<\tau_{auto}<1$ms. The rest

of the blinks feature a τ_{auto} = -∞, indicating no blinking (autocorrelation is flat), and a minor number of outlying τ_{auto} because of the low number of emission photons. The averages τ_{auto} of this fast blinking were 58 μs, 48 μs and 54 μs for laser intensities of 5.7 kW/cm^2, 8.5 kW/cm^2 and 11.3 kW/cm^2 respectively (Figure 4.14a, b and c). The U2OS sample stained with Alexa 647, but with a Vectashield embedding resin, had 46% / 53 μs (Figure 4.14d), 55% / 58 μs and 41% / 57 μs for laser intensities of 5.7 kW/cm^2, 8.5 kW/cm^2 and 11.3 kW/cm^2 respectively. A Fibroblast sample stained with Alexa 647 in an OxEA buffer[121] had an average τ_{auto} of 28 μs, where 20% emissions had 0<τ_{auto}<1ms (Figure 4.14e). When using Atto 647 in an OxEA buffer, we detected 27% emissions with 0<τ_{auto}<1ms, with an average of 49 μs (Figure 4.14f).

Data show that laser power has a small effect on fast blinking. On the contrary, buffers change blinking parameters significantly, indicating that the fast blinking is caused by a triplet state. The two different dyes also have a different blinking behavior. The extracted τ_{auto} (Figure 4.14) - with an unprecedented time resolution range down to 6.4 μs - can be used for the photophysical analysis of a dye and its optimization. Timing parameters could also be used to estimate pH or concentration values[134,135].

4.4 Saturation method for lifetime extraction

Fluorophore lifetime of the first excited state is a molecular property used to distinguish between different fluorophores, or the fluorophore environment. Usually, it is extracted using a pulsed laser source. We measure the delay from excitation (laser clock) to emission, so as to reconstruct the distribution of the delay and to fit exponential decay parameter τ. The delay is going to be a random exponential distribution:

$$I(t) = \begin{matrix} 0, \text{ for } t < T_0 \\ I_0 e^{-(t-T_0)/\tau}, \text{ for } t \geq T_0 \end{matrix}, \tag{4.9}$$

where *I(t)* is the number of occurrences with delay *t*, I_0 the peak number of occurrences, and T_0 the constant delay due to the electrical and excitation path.

SMLM needs to apply a high laser power to force the majority of the fluorophores into the dark state, so as to fulfill the sparsity requirement. One could think of applying a pulsed laser (instead of a continuous laser) set to have an equivalent average power. For example, using a 1 ns Alexa Fluore 647 lifetime[136], we could use a 10 ns laser period with a 100 ps pulse width. The required peak intensity P_0 would then be 10^2×5 kW/cm^2 or 500 kW/cm^2

(considering 5 kW/cm^2 usually used with continuous lasers[10,17]). The laser beam is a Gaussian shape spot with $\sigma_x=\sigma_y=15$ μm. The total peak laser power P can be calculated as $P=P_0\times 2\pi\sigma_x\sigma_y=7.06$ W, with $P_0=500$ kW/cm^2. This high output power raises concerns on objective damage, especially in widefield configurations were illumination is focused at the back focal plane of an objective.

However, SMLM has an appealing feature: fluorophores are usually single-photon emitters[137,138], sparsely distributed over space. If the fluorophore is isolated and excited with a strong continuous laser, its maximum photon flux will saturate and its inter-photon time will be governed by its lifetime.

Here, we investigate the required photon flux exciting the single emitter, so as to achieve lifetime governed inter-photon times, and the required sensor dead time to be able to detect the emitted photon flux.

Taking an extinction coefficient of 270 000 Lmol^{-1}cm^{-1} for Alexa Fluor[139], we can calculate the absorption cross-section as[140] 270 000 Lmol^{-1}cm^{-1} × 3.82×10^{-21} cm^{3}molL^{-1}molecule^{-1} = 1.0314^{-15} cm^2/molecule. Given typical SMLM laser intensities[10,17] of 5-15 kW/cm^2, we can calculate the excitation photon flux to be between 1.62×10^{22} and 4.88×10^{22} photons/s/cm^2, for a 647 nm excitation laser. Considering the cross-section, each molecule will be excited at a rate between 16.77×10^6 and 50.32×10^6 photons/s, or with an average period between 59.61 and 19.87 ns. With the calculated excitation flux and a total microscope collection efficiency to be around 10%[141], and the Alexa Fluor 647 quantum yield[136] of 0.33, we estimate the number of collected photons per 10 ms to be around 5 280. This figure is comparable with an average of 1 800 photons per 10 ms we collected in SMLM (with an EMCCD), considering that fluorophores also switch to the triplet and dark state and emission and frames can be misaligned.

The inter-arrival time distribution will be determined jointly by the distribution of the excitation photons, the electron decaying from the excited state to emit a photon, and the dark states (including triplet state). Here, we neglect the dark state effect and consider it to have a lifetime longer than 10 μs. The fluorophore will be excited at a rate $1/\tau_{exc}$ (before calculated to be between 16.77×10^6 and 50.32×10^6 photons/s) to the first excited state S_1 and decay to the ground state S_0 with a rate $1/\tau_{em}$, where τ_{em} is the Alexa Fluor 645 lifetime, taken to be 1 ns[136]. The fluorophore can be excited after it returns to S_0. The average ΔT between emitted photons is going to be the summation of two random variables $\mathbf{X}_{\text{exc}}$ and

$\mathbf{Y}_{em}$, $\mathbf{X}_{exc}$ being exponentially distributed (excitation photon arrival) and $\mathbf{Y}_{em}$ being exponentially distributed (electron decay time from S_1 to S_0, the fluorophore lifetime, while emitting a photon). Inter-arrival time is a random process (summation of two variables with exponential distributions) due to the memory-less feature of exponential distributions; the probability of an excitation photon after ΔT from the previous excitation photon is governed by the same distribution as the probability of an excitation photon after ΔT from a fluorophore photon emission. Thus:

$$\Delta\mathbf{T} = \mathbf{X}_{exc} + \mathbf{Y}_{em} \tag{4.10}$$

The probability density function for $\mathbf{X}_{exc}$ is:

$$f_X(\mathrm{t}) = \frac{1}{\tau_{exc}} e^{\frac{-t}{\tau_{exc}}} \tag{4.11}$$

And the probability density function for $\mathbf{Y}_{em}$ is:

$$f_Y(\mathrm{t}) = \frac{1}{\tau_{em}} e^{\frac{-t}{\tau_{em}}} \tag{4.12}$$

The inter-arrival probability density function $f_{\Delta T}(t)$ can then be calculated as:

$$\begin{aligned} f_{\Delta T}(\mathrm{t}) &= \int_0^t f_X(\mathrm{t}-\mathrm{y})\, f_Y(\mathrm{y}) dy = \int_0^t \frac{1}{\tau_{exc}} e^{\frac{-(t-y)}{\tau_{exc}}} \frac{1}{\tau_{em}} e^{\frac{-y}{\tau_{em}}} dy \\ &= \frac{e^{\frac{-t}{\tau_{exc}}}}{\tau_{exc}\tau_{em}} \int_0^t e^{\frac{(\tau_{em}-\tau_{exc})y}{\tau_{em}\tau_{exc}}} dy = \frac{e^{\frac{-t}{\tau_{exc}}}}{\tau_{em}-\tau_{exc}} \left(e^{\frac{(\tau_{em}-\tau_{exc})y}{\tau_{em}\tau_{exc}}} \Big|_0^t \right) \\ &= \frac{e^{\frac{-t}{\tau_{exc}}}}{\tau_{em}-\tau_{exc}} \left(e^{\frac{(\tau_{em}-\tau_{exc})t}{\tau_{em}\tau_{exc}}} - 1 \right) = \frac{1}{\tau_{em}-\tau_{exc}} \left(e^{\frac{-t}{\tau_{em}}} - e^{\frac{-t}{\tau_{exc}}} \right) \end{aligned} \tag{4.13}$$

If we assume the two random variables to be generated by independent random processes, the estimated ΔT and variance are going to be:

$$E[\Delta T] = \tau_{em} + \tau_{exc} \tag{4.14}$$

$$\mathrm{var}[\Delta T] = \tau_{em}^2 + \tau_{exc}^2 \tag{4.15}$$

The inter-arrival time ΔT is going to be biased away from the fluorophore lifetime τ_{em}, indicating that the excitation period τ_{exc} should be $\tau_{exc} << \tau_{em}$ in order for the ΔT to be governed by the fluorophore lifetime τ_{em}. This requirement is 2 orders of magnitude away from the current excitation rate (and the used fluorophore with 1 ns lifetime), and suggests that the high power requirement is not feasible for widefield illuminated microscopy. This could be however achieved with concentrated illumination, usually used in scanning or small field of view microscopy approaches. There, photons are emitted and detected at a high pace, and we should estimate the requirements for the SPAD array dead time.

Photons go through the microscopy setup and are detected with an efficiency α of around 10% (see supplementary data by Israel *et al.*[141]). The average ΔT is thus increasing by $1/\alpha$:

$$f_{\Delta T}(\mathrm{t}) = \frac{\alpha}{\tau_{em} - \tau_{exc}} \left(e^{\frac{-\alpha t}{\tau_{em}}} - e^{\frac{-\alpha t}{\tau_{exc}}} \right) \tag{4.16}$$

Furthermore, taking an approximated Gaussian spatial distribution of the PSF with $\sigma = 0.21\lambda/\mathrm{NA}$[142], we assume an emission wavelength λ=665 nm, NA=1.25, to yield a standard deviation of σ=110 nm. Using an effective pixel size of 100 nm, we estimate the percentage of photons β hitting the central pixel with:

$$\beta = \frac{1}{2\pi\sigma^2} \int_{-50}^{50} \int_{-50}^{50} e^{\frac{-(x^2+y^2)}{2\sigma^2}} = 0.1228 \tag{4.17}$$

The central pixel receives a maximum photon flux, and adds another β=0.12 coefficient to α. Thus:

$$E\left[\Delta T_{central}\right] = \frac{\tau_{em} + \tau_{exc}}{\alpha\beta} \approx \frac{\tau_{em}}{\alpha\beta}, \tau_{em} >> \tau_{exc} \tag{4.18}$$

To avoid count suppression (see section 2.2.4), the SPAD dead time T_{dead} needs to be smaller than $E[\Delta T_{\mathrm{central}}]$, the average photon detection period. We shall allow a maximum of 5% of suppressed counts; the measured photon rate m is thus $0.95n$, where $n = 1/E[\Delta T_{\mathrm{central}}]$ is the detected photon rate. Taking a passive recharge approach modeled with (2.9) and $m = 0.95n$:

$$0.95n = ne^{-nT_{dead}} \tag{4.19}$$

$$T_{dead} = \frac{-\ln 0.95}{n} = -E[\Delta T_{central}]\ln 0.95 = \frac{-\tau_{em}}{\alpha\beta}\ln 0.95 \qquad (4.20)$$

Simplifying (4.20):

$$T_{dead} = 4.27\tau_{em} \qquad (4.21)$$

This seems to be a hard requirement SPADs may not easily reach for τ_{em}=1 ns (T_{dead} is usually in the order of 50 ns). Solutions may include higher optical magnification or smaller pixel sizes. Figure 4.15 shows examples of emitted, detected and measured inter-arrival times compared to the fluorophore lifetime distribution, with three different excitation laser intensities. If the excitation laser intensity is too low (Figure 4.15a), the emission inter-arrival time is governed by the excitation laser. The photons go through the optical setup and reduce the photon flux by 10×. Finally, the short inter-arrival times are cut out due to SPAD dead time, as shown in Figure 4.15c (red versus green curve). Fitting the histogram to an exponential curve mitigates the dead time problem. If averaging of the inter-arrival times is used, we bias the τ_{em} estimator towards longer inter-arrival times.

Experimentalists could use two fluorophores with different lifetimes and evaluate the inter-arrival times at different laser intensities. Finally at higher laser power, the inter-arrival times from the two fluorophores will show a different distribution governed by the lifetimes. At lower laser powers, inter-arrival times are going to be similarly distributed, depending on the quantum yield. Note however, that at any operating point, one can gain additional information from the inter-arrival times, since the final fluorophore intensity is not directly related to the average photon flux, since it includes rates to the triplet and dark states.

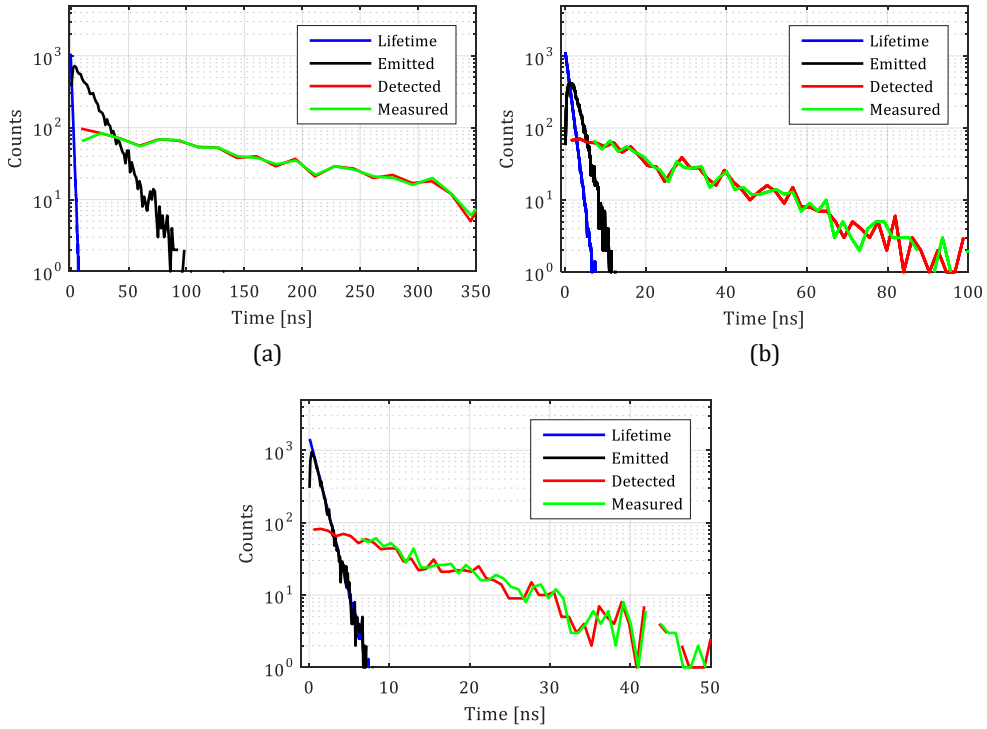


Figure 4.15. Examples of inter-arrival times at different excitation periods τ_{exc}. (a) Shows τ_{exc}=15 ns, (b) τ_{exc}=1.5 ns, and (c) τ_{exc}=0.15 ns, compared to τ_{em}=1 ns. The emission inter-arrival distribution (black) is dominated by the larger of the two (τ_{exc}, τ_{em}). Red shows detected inter-arrival, whereas green shows the same distribution cut-off because of the SPAD dead time at T_{dead}=50 ns×β, using β=0.12 from (4.17).

4.5 Quantum correlations of single molecules

Many fluorophores are single-photon emitters[137,138]. One can use advanced statistical methods to enrich the information about the sample, if single-photon emitters are paired with single-photon detectors. For example, Israel *et al*.[141] calculated pixel cross-correlations to estimate the number of active emitters per diffraction volume. They then used time sections where just one emitter was active, and applied PSF localization to achieve sub-diffraction resolution. This method could be also applied in SMLM to allow for denser molecule activation.

SPAD imagers with event-driven or 1-bit frame-based readout are good candidates for similar setups. The first prerequisite for practical measurements is the overall sensitivity. Israel *et al*. reported joint fiber and SPAD sensor sensitivity of 36% (see supplementary data by Israel *et al.*)[141]. The overall

system speed is the second critical aspect, and is usually linked with the laser clock, operating between 10 and 100 MHz. It was however shown that lower clock periods (below 1 MHz) decrease the photobleaching effect because molecules are given time to relax from the triplet state[143]. If time considerations allow, lower repetition rates substantially increase the total number of collected photons before bleaching, albeit with longer acquisition times.

4.6 Conclusion

Following the community drive to decrease the acquisition time (of super resolution localization microscopy) by using shorter frame times, we investigated the optimal frame time and concluded that it is longer than the average blinking time, and dependent on the emission intensity and background intensity ratio.

SPAD imagers have been regarded by some as unsuitable for applications where sensitivity is critical. This belief was driven by the lack of high fill factor sensors. With the introduction of SPAD imagers with improved fill factor, we demonstrated first super resolution localization images taken with a SPAD imager, with an estimated localization uncertainty of 20 nm and resolution better than 80 nm. We showed that super resolution localization microscopy can exploit the high timing resolution provided by this type of imagers. We also presented the instrumentation and methodology for a systematic widefield blinking analysis. We believe that the characterization of very fast blinking is critical for future developments in super resolution, with the goal of high acquisition speeds for the best possible localization uncertainty.

Finally, we presented data where the MEA buffer allowed for 70% of emission with additional fast blinking in the µs range, whereas OxEA allowed for 20%. To the best of our knowledge, this is the first comprehensive widefield analysis of blinking with microsecond timing resolution and the first performed on a SPAD imager.

State-of-the-art SPAD imagers have a competitive advantage when combining spatial and temporal resolution, but still don't reach the same SMLM resolution as EMCCD and sCMOS due to the lower PDE. Further developments of the SPAD imagers will yield a higher PDE and a smaller pixel pitch, which is expected to further increase the noise uniformity and lower the number of hot pixels. This goal will likely be achieved in 3D ICs, where SPADs are operating in backside

illumination and the electronics in the bottom tier chip will perform complex operations on individual pixels and clusters of pixels.

Further, we conducted a feasibility study on using photon inter-arrival statistics to extract lifetime information (saturation method). SPAD arrays offer a unique capability to study single photon statistics spread over a widefield area. For example, one approach used photon quantum correlations to estimate the emitter number and achieved sub-diffraction limited resolution. Large format imagers with integrated on-chip functionality will facilitate further development in this field.

5 Small format SPAD imagers for confocal microscopy

While chapters 2, 3 and 4 focus on large format SPAD imagers, the present chapter 5 deals with small format SPAD imagers. The reasons for this is that small arrays usually stream SPAD pulses directly to the output, easing usage (plug and play) and giving flexibility to build systems for different applications. Small format SPAD imagers mitigate many challenges presented in previous chapters, however they do require special design care. So far, the dynamic range of small format SPAD imagers was limited when compared to a standard and largely used PMT technology. We present a small format imager called nanoSPAD in section 5.1, and specially focus on different modes of recharge mechanisms and correction models to increase the dynamic range in subsection 5.1.4. We include a short comparison between competing technologies for confocal microscopy in section 5.2.

5.1 nanoSPAD

nanoSPAD is a small format SPAD array with 23 pixels, designed for image scanning microscopy[144–147], a special type of confocal microscopes. The small format array is scanning the sample over a matrix to reconstruct a full image. A SPAD array replaces a single point detector and increases both the light collection and lateral resolution. Each pixel in the SPAD array is operating as a pinhole, with the central pixel representing the classical pinhole aligned with the optical axis. The pixels that are not aligned with the optical axis collect less light, but have a narrower PSF. When summing contributions from all 23 pixels, we have an increased light collection and lateral resolution[147,148].

The pixel schematic is shown in Figure 5.1. The SPAD is based on a previous low noise SPAD design[44], implemented in a 180 nm standard CMOS technology. The quenching and recharge is implemented with a cascode transistor combination[149] to enable high excess bias V_{EX}. While the SPAD fires, the anode voltage increases to V_{EX} and the inverter input reaches V_{EX}-V_{T1}, where V_{T1} is the T1 threshold voltage. The body effect allows for $V_{EX} = 2\times V_{Tmax} = 7.2$ V with reliable operation. We have however measured continuous SPAD operation with V_{EX}=11 V for over 170h (one week) with no measured degradation. SPAD pulses exhibit a fast rise time with exponentially decaying fall time, limiting the time the anode voltage is >7.2 V.

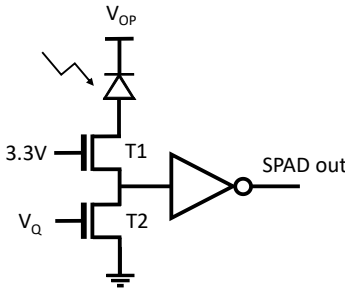
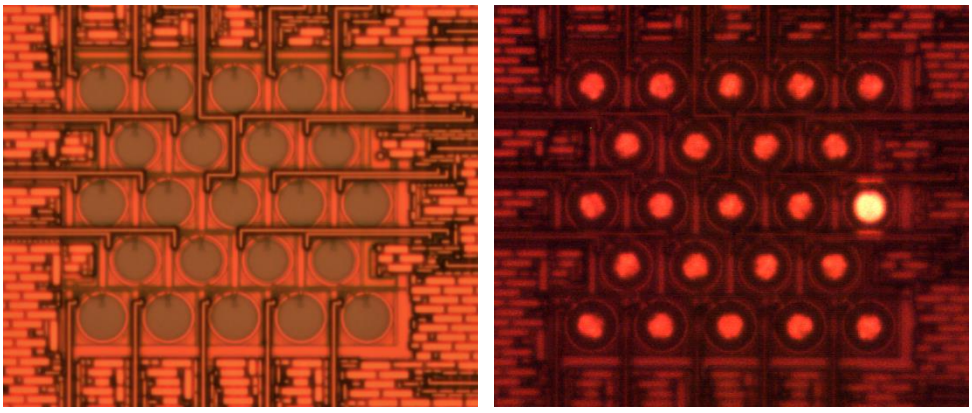


Figure 5.1. nanoSPAD pixel schematic with cascode transistors to allow for high VOP.

V_Q in the figure is a global signal. Passive recharge is implemented with a constant bias, and other recharge mechanisms (event-driven or clock based) can be performed by driving V_Q with a FPGA through a DAC. The inverter performs capacitive isolation and should reduce afterpulsing and timing jitter. 'SPAD out' signals are fed to digital outputs with programmable output strength. Pixels are arranged in a hexagonal manner to accommodate a round shape (desirable in image scanning microscopy) and to increase the fill factor. The radial pixel pitch is 23 µm, and the radius of the active area is 5.85 µm. This translates into a native fill factor of 23.5%.

Chip micrographs are shown in Figure 5.2. The right hand side image shows a uniform light emission test. The pixels do not show edge breakdown.



(a) (b)

Figure 5.2. (a) nanoSPAD micrograph with illumination. (b) nanoSPAD micrograph with SPADs operating with low quenching resistance to enable simultaneous avalanching and light emission. Light is emitted from the active area, in pixel centers. The bright outlier on the right is a "hot" pixel.

5.1.1 Photon detection probability

Figure 5.3 shows PDP measurements, with the peak PDP increased by 10% compared to previous designs[44]. Multiple sources could be the cause of the increase. This design was implemented in a new tapeout run, opening a possibility for a process change. We used anode connections in the border of the SPAD active area, whereas the previous design had a 6×2 µm metal shielding on top of the 113 µm² active area. Additionally, p-well diffusion might increase the effective active area by a constant, thus yielding falsely higher PDP for smaller active areas (original 6 µm compared to 5.85 µm in the current design). If p-well diffusion is significantly contributing to the PDP increase, microlenses will not yield the expected increase in effective fill factor. One can also estimate this effect by measuring the DCR of SPADs with different active areas.

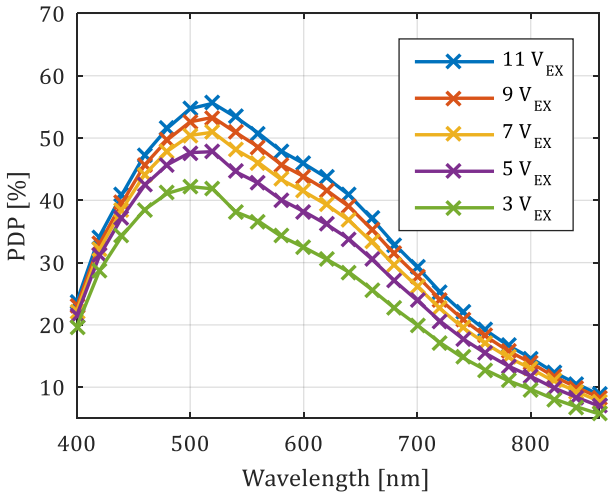


Figure 5.3. nanoSPAD PDP measurements at different V_{EX}. The peak PDP at 520 nm is 55%, while PDP is higher than 40% between 440 and 640 nm.

We tested the breakdown uniformity with the excess count rate method. Results are presented in Figure 5.4. The V_{OP} was changed from 25.32 V in steps of 20 mV to 25.78 V, and the count rate linearly extrapolated to the intercept point on the horizontal axis. Measurements show an average of 21.95 V_{BD} with a standard deviation of 56 mV. Since the PDP is not linear with respect to V_{OP} (but compressing at higher V_{OP}), we underestimate the breakdown voltage V_{BD}. Figure 5.4b also reveals a narrower distribution of counts at around 23 V_{OP}, which was measured to be the breakdown voltage for a testing SPAD without integrated electronics (using pulse emergence voltage as the breakdown). A

56 mV breakdown voltage change yields a PDP change of 0.17% considering a 5.96% change in PDP when changing the V_{EX} from 3 to 5 at 520 nm.

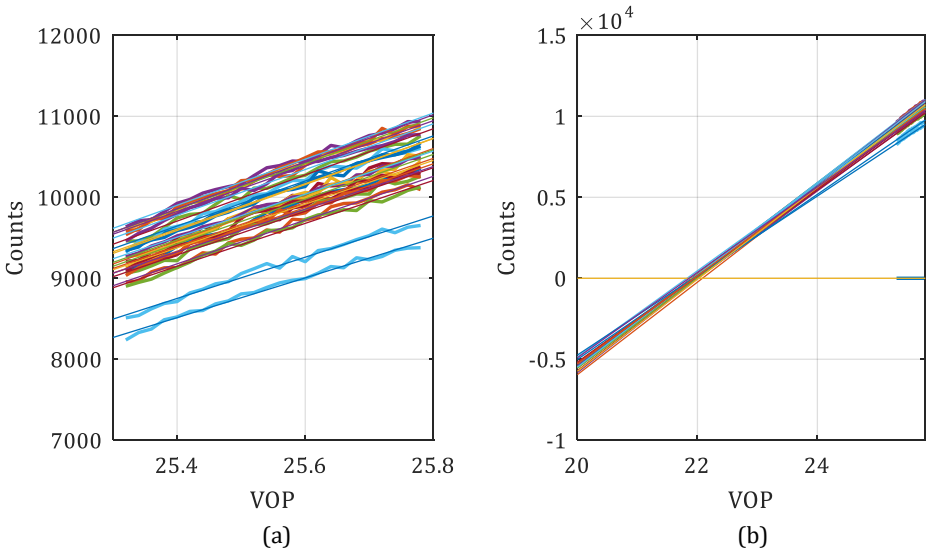


Figure 5.4. Excess counts versus V_{OP} to extract the breakdown voltage. (a) Shows a zoomed part with measured data points. One can note the nonlinear response of counts to V_{OP}. The data is linearly extrapolated to extract V_{OP} at which counts cross zero, yielding the breakdown voltage V_{BD}.

The photon response nonuniformity (PRNU) was tested using a white lamp with an integrating sphere. At 1.5% of the dynamic range (to avoid count suppression due to dead time, operating with passive recharge) and 7 V_{EX}, we measured a PRNU of 1.6%, expressed as the ratio between the standard deviation and the average count number. Active clock recharge yields the same results.

5.1.2 Noise sources

We measured uncorrelated (DCR) and correlated noise (afterpulsing and crosstalk) sources finding low noise performance. Due to low noise, count output is shot-noise-limited and maximizes the signal-to-noise ratio.

5.1.2.1 Dark count rate

Figure 5.5 shows multiple pixels DCR at different V_{EX}. The majority of the pixels features a DCR lower than 100 cps at V_{EX}=9. There are two major noise generation mechanisms: band-to-band tunneling and trap assisted generation.

Band-to-band tunneling increases with higher electric fields. Trap assisted generation is highly dependent on temperature. A low DCR versus V_{EX} dependency indicates trap assisted DCR generation. At V_{EX}=8, DCR starts to roll-off towards high values due to increased electric field and band-to-band tunneling. Figure 5.5b indicates a higher band-to-band tunneling for V_{EX}=11 with respect to V_{EX}=9. While trap assisted generation dominates the DCR at temperatures higher than 20 ˚C, band-to-band tunneling is limiting a further DCR decrease at temperatures below 0 ˚C.

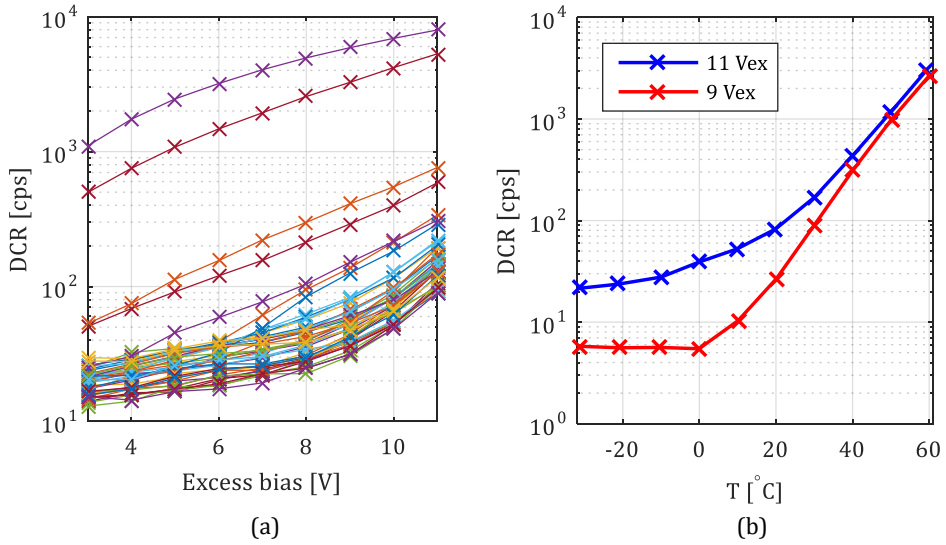


Figure 5.5. (a) Room temperature DCR of 46 different pixels at excess bias voltage ranging from 3 to 11 V. The majority of the pixels have a DCR lower than 100 cps at 9 V_{EB}. (b) DCR reduces exponentally with temperature, but reaches a plateau at around 0 ˚C. DCR then becomes dominantly generated by band-to-band tunneling, which is less effected by cooling.

Figure 5.6 represents the quantified percentage of “hot” pixels. At 3.3 V_{EX}, less than 4% of pixels have a DCR higher than 100 cps, whereas less than 2% have a DCR higher than 1 kcps.

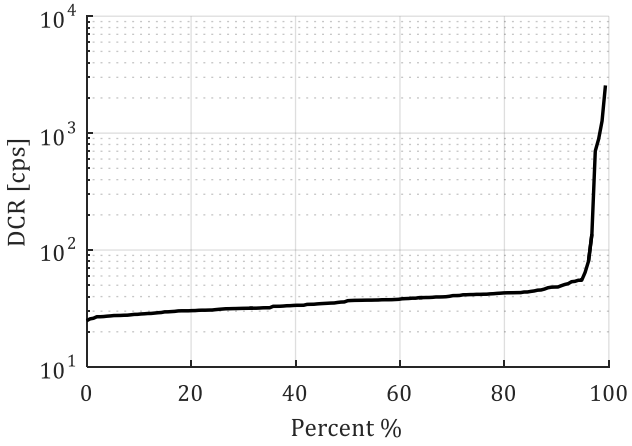


Figure 5.6. DCR distribution of 7 combined chips, at V_{EX}=3.3. Better DCR uniformity is achieved by using smaller active areas or by cooling[14].

5.1.2.2 Afterpulsing and crosstalk

Afterpulsing is found to be 0.1% with 50 ns dead time. The SPAD was operating at 11 V_{EX} and 0.75 V_Q. Crosstalk between two adjacent pixels was measured to be 0.09% under the same conditions. A low afterpulsing is achieved due to capacitive isolation of the SPAD anode through an inverter. Deep trench isolation limited the crosstalk, but also reduced the fill factor. Increased correlations (above red curve) are present only in the first ΔT bin, indicating optical coupling. If electrical coupling was be present, increased correlations would be featured at longer ΔT as well (diffusion to adjacent pixels).

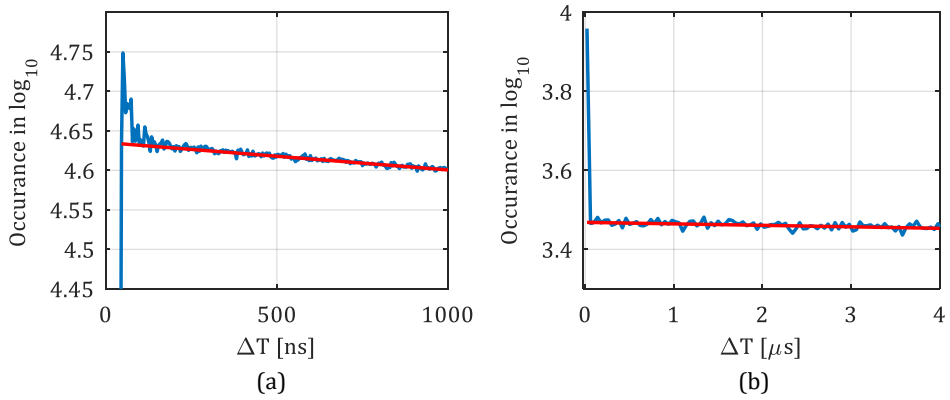


Figure 5.7. Inter-arrival time distributions between consecutive counts (a) in the same pixel and (b) between adjacent pixels.

5.1.3 Timing jitter

The timing jitter measurements are presented in Figure 5.8, featuring a 134.32, 128.29 and 123.12 ps FWHM for 7, 9 and 11 V_{EX}, respectively. When compared to the previous implementation, the timing jitter increased from 100.8 ps[44] to the current 123.12 ps. In addition to the fixed inverter threshold (see Figure 5.1), a cascode transistor in series could increase the output resistance and influence the timing as well. The measurements are performed with a 40 MHz laser, a SPAD dead time of 100 ns and a count rate at 100 kcps.

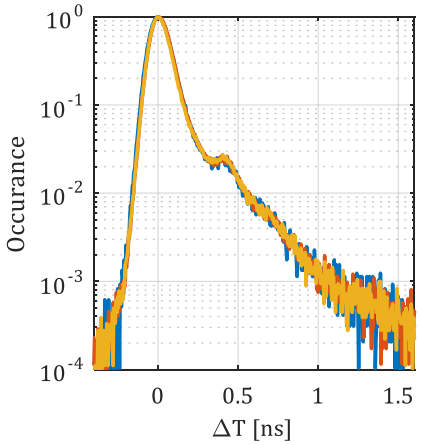


Figure 5.8. Timing jitter histogram for three different V_{EX} (7, 9, 11 V), using a red laser (637 nm). The right hand side features a diffusion tail.

5.1.4 Recharge mechanisms

For the inter-arrival time histograms presented in this subsection, we used in-FPGA histogramming with a time resolution of 10 ns. Figure 5.9a shows inter-arrival histograms of all pixels. The dead time was estimated by looking at the peak inter-arrival time, yielding in this case an average dead time of 225 ns and standard deviation of 57 ns for passive recharge with a constant V_Q bias. Due to variations in the threshold voltage of T2 V_{T2}, a constant bias yields different recharge resistances and dead times. Considering a rather large spread of dead times, we expect pixel-to-pixel variations in the response (measured count rate *m* versus detected count rate *n*). Responses should following (2.9), all with a different T_{dead}.

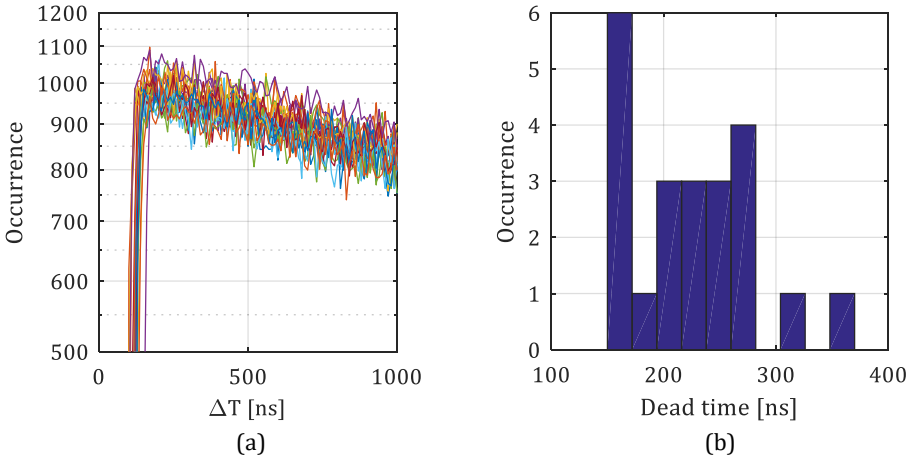


Figure 5.9. (a) Inter-arrival time histograms for each pixel in the array. (b) Histogram of the dead time through the array.

The response of one pixel at 7 V_{EX} and 50 ns dead time is shown in Figure 5.10. The measured count rate *m* is plotted versus the detected count rate *n* extracted by using a reference photodiode. *m* is increasing with *n* until *m* reaches $1/(e \times T_{dead})$. Further increasing *n* decreases *m* due to pulse overlapping at the output. Due to the non-monotonic nature of the response, *m* to *n* corrections cannot be performed without a priori knowledge of the detection environment.

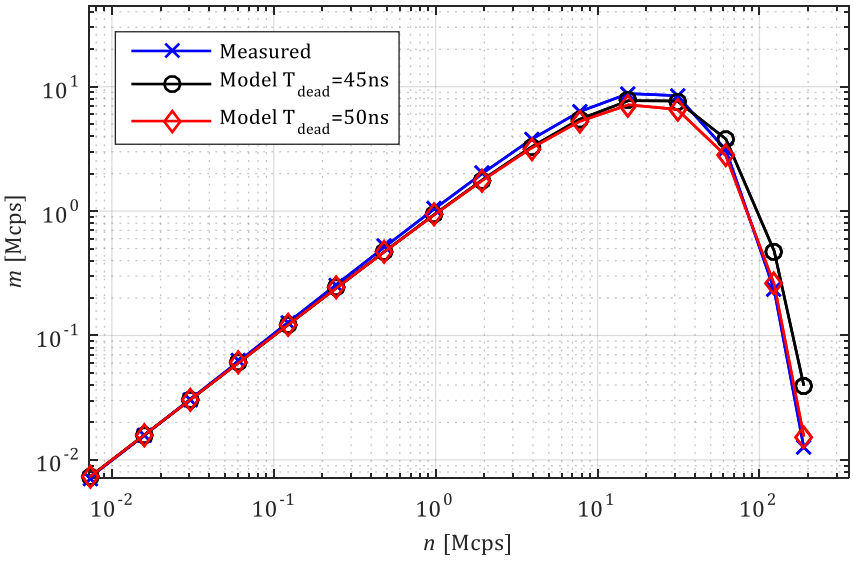


Figure 5.10. Measured per pixel count response m with respect to the detected count rate n. m increases to reach a maximum of $1/(e \times T_{dead})$, in this case 7.6 Mcps.

We define m_i as the measured count and n_i as the detected count number. Further, $n_i=n \times t$, where t is the integration time. For standard deviation and signal-to-noise measurements, count rate data is not applicable. The m_i referred shot noise and signal-to-noise ratio are presented in Figure 5.11. The deviation from the theory is likely caused by rare oscillation effects with an oscillation period of 20 ns (example shown in Figure 5.20a). Oscillations occur only at high impinging photon rates.

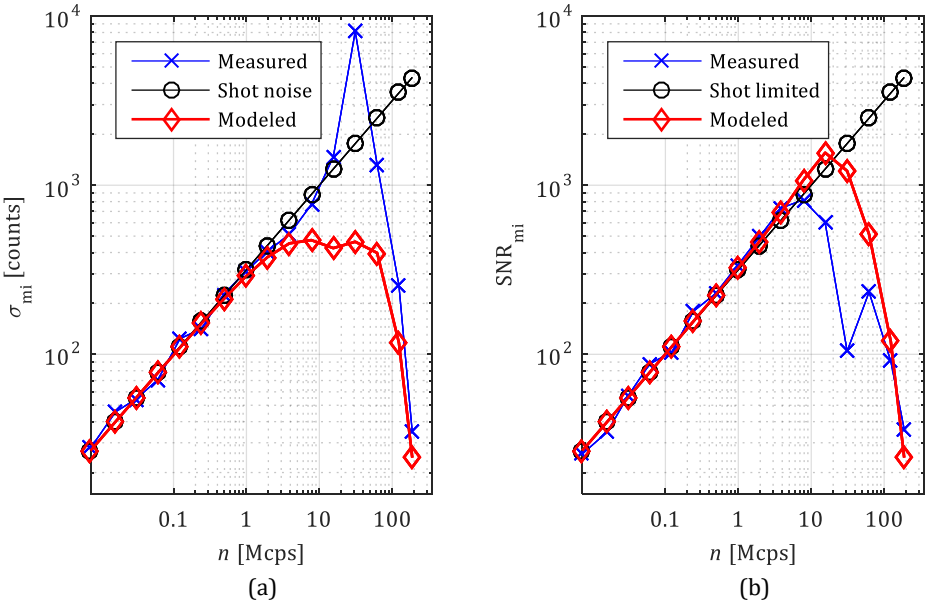


Figure 5.11. (a) Measured standard deviation of the counts compared to the Poisson shot noise and the model with (2.14). (b) Shows the equivalent output SNR.

Due to the dead time spread over the array (Figure 5.9) and the non-monotonic photon response (Figure 5.10), the useful dynamic range is in this case limited to 7.8 Mcps of detected count rate *n.* At higher detection count rates, the SNR decreases and the response over the array is highly nonuniform due to dead time variations (Figure 5.12).

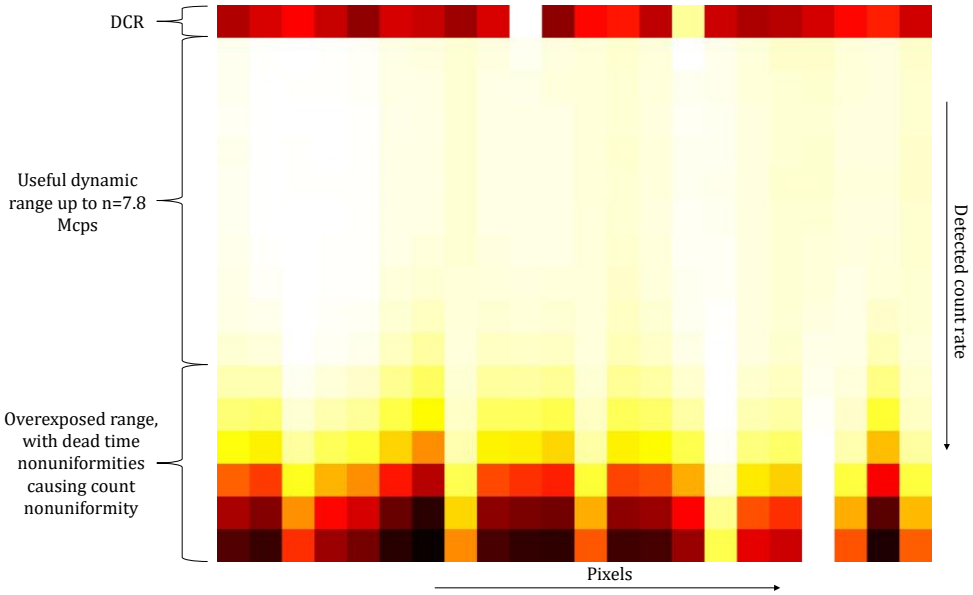


Figure 5.12. Measured count rate (indicated by color, normalized with respect to the maximum count rate over the row, i.e. there is one white pixel per row) over the array (horizontal pixels) when increasing the detected count rate n. The first row represents the DCR. The following 10 rows are the useful dynamic range with uniform photon response over the array. Last six rows feature nonuniform photon response due to differences in T_{dead}, where each pixel has a different measured count rate (m) versus detected count rate (n) response (Figure 5.10).

Figure 5.13 shows inter-arrival time histograms for different detection count rates. A steeper curve with a high occurrence at short ΔT indicates a higher detection count rate. Histograms show very low afterpulsing rates and indicate a dead time (ΔT with highest occurrence) of 50 ns.

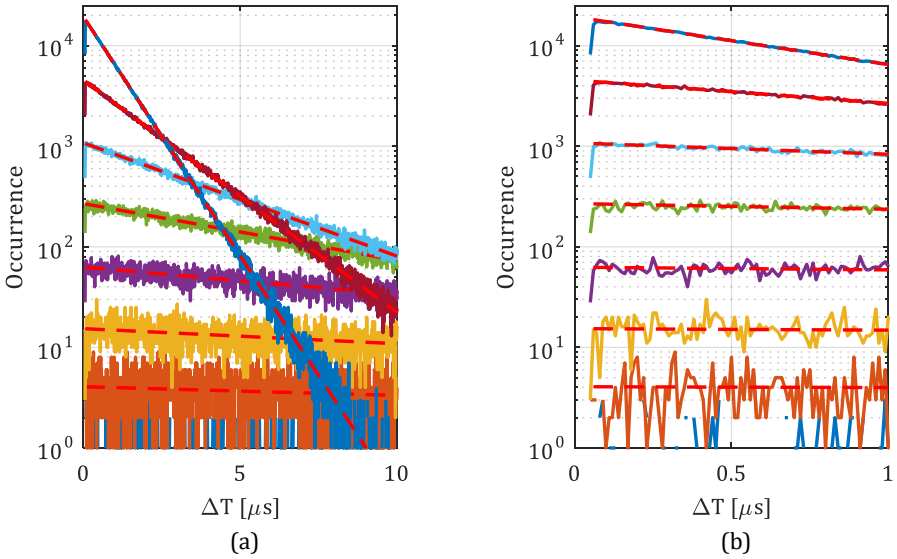


Figure 5.13. Inter-arrival times for passive quenching with dead time of 50 ns. The red discontinuous lines represent exponential fitting to estimate afterpulsing.

The SPAD power consumption is similar to that of a CMOS switch. It is composed of three parts; the SPAD leakage current, the capacitive load current equivalent to V_{EX}^2Cf (3.1), and the transient current during the SPAD quenching (when SPAD has a resistance of around 100 Ω). The nanoSPAD current consumption (23 SPADs) over the dynamic range is shown in Figure 5.14. The figure is important to investigate the trend in large format SPAD arrays. A megapixel design with equivalent pixel design will consume 0.18 A and 3.5 A at 1 Mcps and 200 Mcps, respectively. Thus, we need to use smaller SPADs for the high dynamic range imager proposed in section 3.3, combined with meshed and wide V_{OP} metal lines.

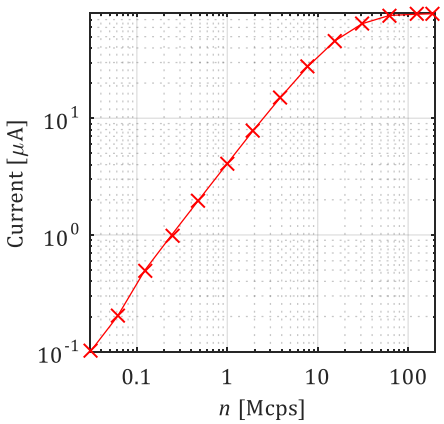


Figure 5.14. nanoSPAD current consumption when using passive recharge over the detected count rate range. At 1 Mcps per pixel nanoSPAD consumes around 4 μA.

Due to the non-monotonic photon response and array nonuniformity at high photon rates, we investigated additional recharge mechanisms. A FPGA was used to generate signals driving the quenching and recharge transistor T2 from Figure 5.1. We generated recharge pulses both synchronous to a single pixel output (event-driven) and asynchronous to the SPAD activity (clock-driven). Figure 5.15 show measured and corrected count rate responses compared to theoretical models presented in sections 2.2.4.1 and 2.2.4.3 (equations (2.5) and (2.11)). T_{dead} is 250 ns and the integration time t is 100 ms. A longer T_{dead} is used to mitigate the recharge uncertainty with respect to the SPAD pulse to 2%. We used a FPGA logic clock with 10 ns period.

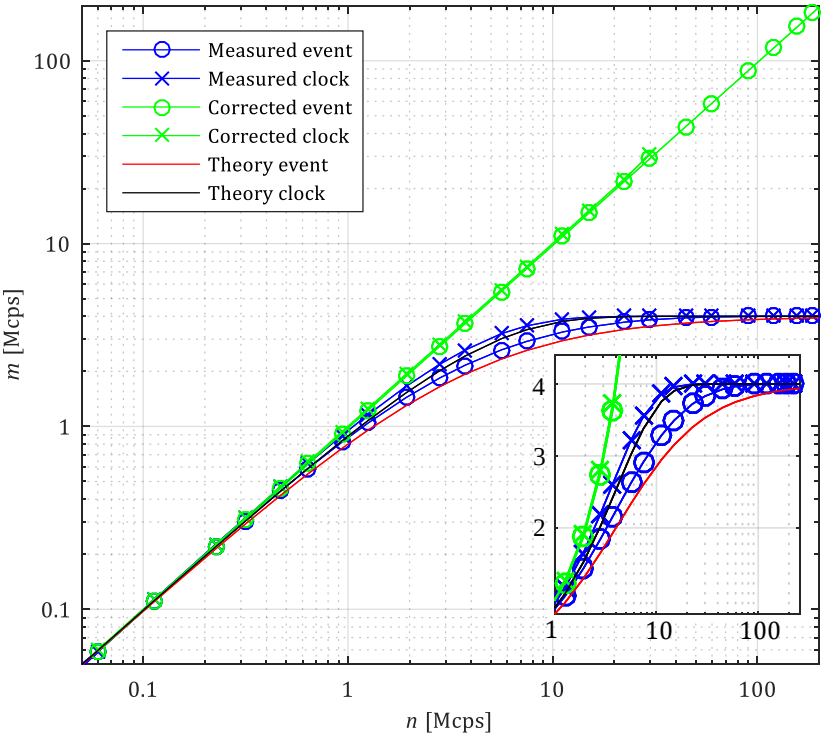


Figure 5.15. Count rate response of event-driven (synchronous) and clock (asynchronous) recharge with 250 ns dead time. The responses are linearized back as to match the detected count rate (green), up to the complete saturation at m=4 Mcps. Clock recharge saturates (reaches m=4 Mcps) at n=44.86 Mcps, and event-driven recharge saturates at n=224.1 Mcps. Black and blue curves show theoretical curves following equations (2.5) and (2.11). The inset graph shows a graph with linear y axis for a closer estimation between the difference of the measured data and theoretical model.

A clear difference between the two recharge mechanisms is also shown in Figure 5.16. The standard deviation (σ_{mi}) of the measured counts m_i is lower than $\sqrt{n_i}$ due to a small change in m_i with respect to n_i (n_i is shot noise governed).

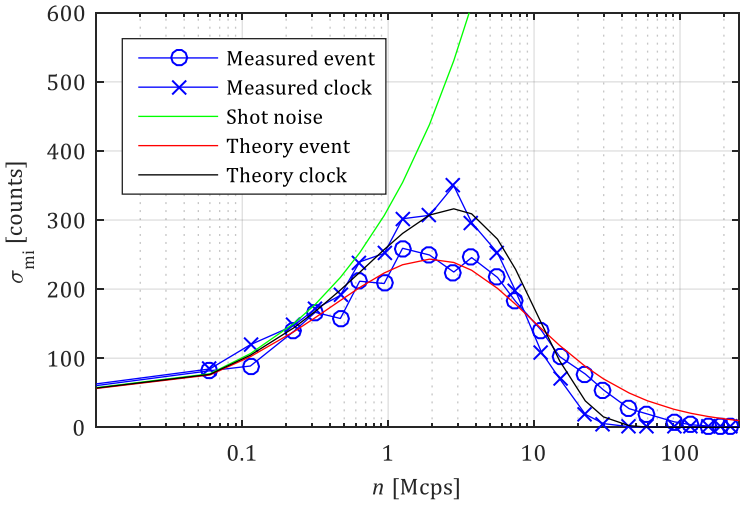


Figure 5.16. Standard deviation of the measured count number σ_{mi} due to shot noise and count suppression, T_{dead}=250 ns and integration time t=100 ms. Green shows the reference, i.e. standard deviation due to shot noise.

However, we are interested in the corrected linear response (input referred, reversing m_i to n_i). The input referred noise (σ_{ni}) is shown in Figure 5.17a. To estimate the dynamic range extension, we used linear interpolation of measured data to extract n with SNR_{ni} drop of -3dB (see section 2.2.5 on discussion for this figure). This SNR is compared to a theoretical sensor with linear response up to $1/T_{dead}$=4 Mcps, with a maximum SNR of $\sqrt{(t/T_{dead})}$= 632.4. The dynamic range extended to n=7 Mcps and n=55 Mcps for clock and event-driven recharge, respectively.

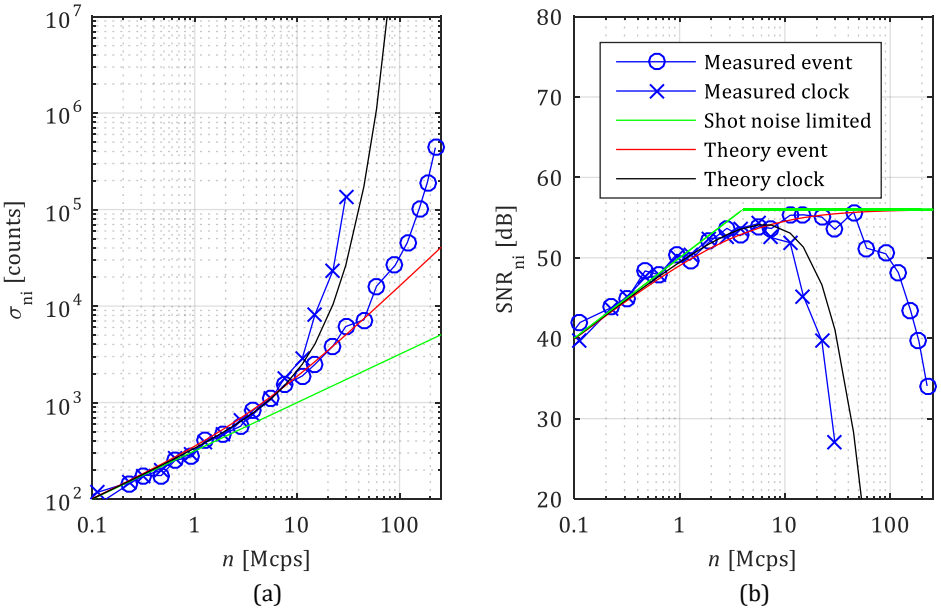


Figure 5.17. (a) Standard deviation of the detected count number σ_{ni} due to shot noise and count suppression, T_{dead}=250 ns and and integration time t=100 ms. (b) SNR_{ni} of the detected count number n_i is for both recharge mechanisms comparable to the classical shot noise $\sqrt{n_i}$ up to n=0.4 Mcps (10% of $1/T_{dead}$). Due to count suppression, the SNR figures diverge significantly. It can be seen that the event-driven recharge increases the dynamic far more than the clock recharge.

Inter-arrival histograms with active event-driven and clock recharge mechanisms are shown in Figure 5.18 and Figure 5.19.

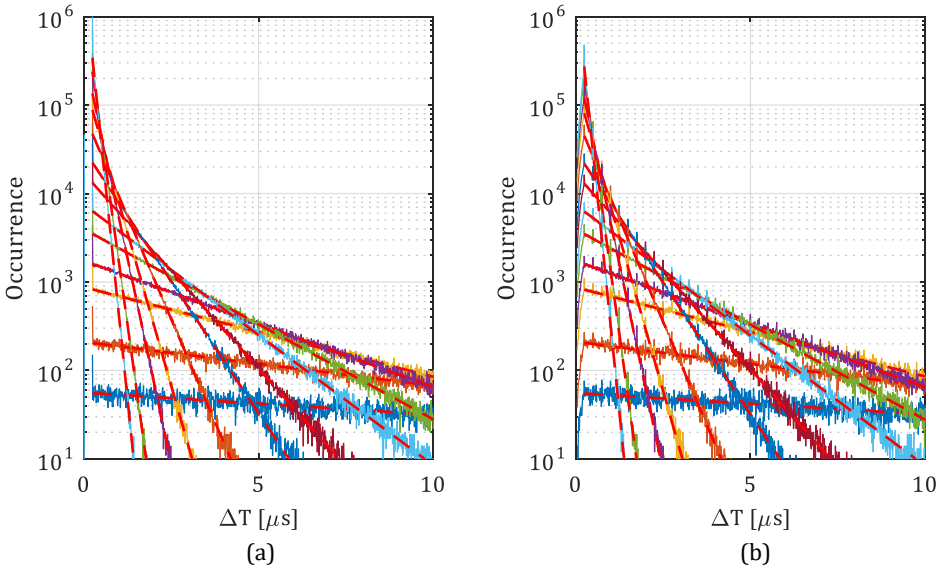


Figure 5.18. Inter-arrival histograms for (a) event-driven and (b) clock recharge mechanisms with 250 ns dead time. Each curve is a histogram of the same pixel at different count rates. The red discontinuous lines represent exponential fitting to estimate afterpulsing.

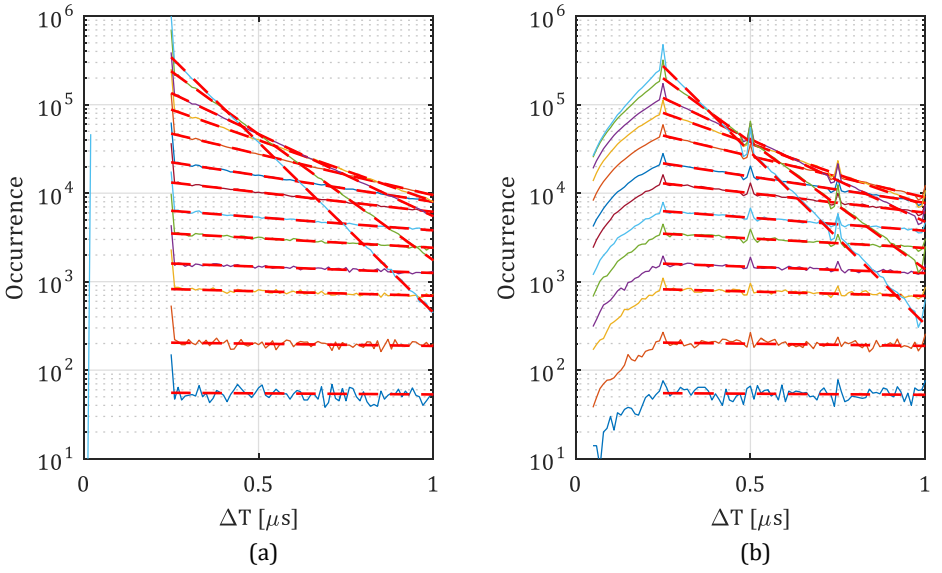


Figure 5.19. Shows a zoomed Figure 5.18, with a clear dead time of 250 ns. At high detection count rates, short pulses in the order of 20 ns emerge (light blue occurrence in part a). The red discontinuous lines represent exponential fitting to estimate afterpulsing.

As it is clear from Figure 5.19, active recharge increases (compared with passive recharge) the occurrence at the bin with ΔT=250 ns, the dead time. We used relatively long recharge pulses of 20 ns. Due to FPGA-based recharge, we used relatively long recharge pulses of 20 ns, thus increasing the probability of a photon detection during the recharge itself (the avalanche not being quenched until the end of the recharge). This effect is also known as the *twilight effect*[150,151]. The photons detected from 230 ns (T_{dead} – recharge pulse width) to 250 ns will be transferred to the ΔT=250 ns bin. The twilight effect falsely induces an increased afterpulsing estimate. True afterpulsing (that is caused by charge trapping) is constant over different detection rates, while twilight "afterpulsing" increases with detection rates. We measured a twilight "afterpulsing" increase from 0.3% to 5.6% for a detection count rate increasing from 0.11 Mcps to 2.75 Mcps for active event-driven recharge.

Interestingly, histograms presented in Figure 5.18 and Figure 5.19 offer an intuitive way of understanding why clock recharge has a better linearity and the response reaches $m=1/T_{dead}$ at a lower n: While event-driven recharge cuts all the inter-arrival times with ΔT<250 ns, clock recharge includes more counts because it allows inter-arrival times with ΔT<250 ns (i.e. the dead time can be as short as the recharge pulse width, in this case 20 ns).

At very high detection count rates we found rare events of 20 ns pulses presented in Figure 5.20a, for all recharge mechanisms (including passive). SPAD current consumption increases due to a clock recharge mechanism (see Figure 5.20b). When compared to passive recharge, the current increased from 4 to 10 μA and from 80 to 300 μA for 1 Mcps and 200 Mcps, respectively.

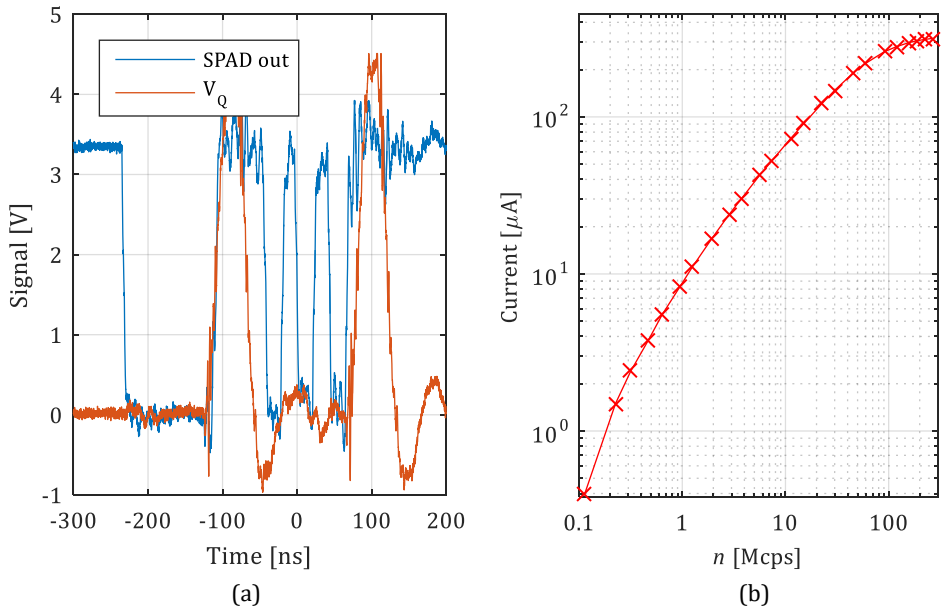


Figure 5.20. (a) The SPAD output at very high detection count rates can feature very short dead times. (b) nanoSPAD current consumption for the whole array when using clock recharge over the detected count rate range. nanoSPAD consumes around 10 µA at 1 Mcps per pixel.

In contrast to passive recharge, active clock recharge features very uniform response over the whole dynamic range, shown in Figure 5.21.

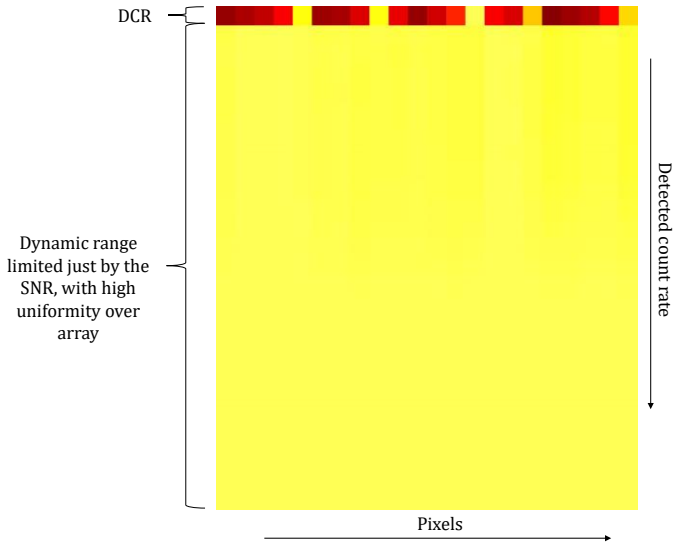


Figure 5.21. Measured count rate (indicated by color, normalized with respect to the maximum count rate over the row) over the array (horizontal pixels) when increasing the detected count rate n. The first row represents the DCR. The following rows show uniform photon responses, with dynamic range solely limited by SNR decrease.

5.2 Performance comparison between SPAD imagers, PMT and hybrid detectors

Current confocal microscopes use mostly PMTs, APDs/SPADs or hybrid detectors. Each of those technologies has advantages and drawbacks. PMTs are a cheaper, mature technology, with high dynamic range (in analog operating mode) and large active area, thus making them easy to align. However, PMTs also exhibit higher multiplication noise and afterpulsing. Hybrid detectors combine the best of the two (at the expense of a higher cost), thus having high QE, high dynamic range and negligible afterpulsing, but also a somewhat higher DCR.

A major drawback of PMTs and hybrid detectors is the fact that they usually don't have a spatially pixelated sensitive area, and are rarely implemented as 2D arrays that are bulky and extremely large. Because of that, systems for image scanning microscopy (that require a 2D array) employ either single-pixel sensors or linear arrays (like GaAsP PMTs) coupled to the optics through fibers and microlenses to form 2D arrays[144].

Table 4. A comparison of detectors used in confocal microscopy.

Sensor	nanoSPAD	SPAD#1	SPAD#2	SPAD#3	SPAD array	GaAsP PMT	hybrid
Description	180 nm CMOS process	Custom process ID100[152]	Custom process PDM[153]	Custom process SPCM-AQRH[154]	CMOS process SPC2[155]	Based on H7422P-40[156]	Based on R10467U-40[157,158]
Detection efficiency maximum	55% @ 520 nm (PDP)	35% @ 500 nm	50% @ 550nm	70% @ 700nm	38% @ 400nm (PDE)	40% @580 nm	45% @500 nm
Spectral range [nm]	400 to 860	350 to 900	400 to 900	400 to 1000	400 to 900	300 to 720	300 to 720
Fill factor	23.5%	-	-	-	3%, but detection efficiency includes fill factor	-	-
Size of active area [μm^2]	107	314-1962	314-7850	25434	314	20e6	7e6
DCR [cps]	<100[a]	<200[a]	<250[a]	<1500	<4000[a]	<300[b]	<700[158,a]
Afterpulsing	0.1%	0.5%	<3%	<1%	5%	High[159,c]	negligible
Multiplication noise	no	no	no	no	no	yes, in analog mode	negligible
Timing jitter	120 ps	40-60 ps	50 ps	350 ps	77.1 ps[28,d]	300 ps[159]	<120 ps[158], 90[157]
Dead time	50 ns	45 ns	77 ns	< 35 ns	77 ns	unknown	unknown
Dynamic range	7.8 Mcps with passive	20 Mcps	12 Mcps	>35 Mcps[e]	12 Mcps	3 Mcps[160,f]	10 Mcps
Shutdown count rate	no	no	no	no	no	unknown	80 Mcps
Integration capability	yes	8 channel increases DCR to 3.5 kcps	no	no	yes	16 channel linear array	no

[a]At room temperature

[b]At 0 °C

[c]Up to 1.5 μs afterpulsing correlations are very high

[d]Assuming this is the same SPAD as in paper by Bronzi *et al.*[28]

[e]Non-linear response included

[f]In photon counting mode, dynamic range is higher in analog mode

nanoSPAD shows an advantage in the detection efficiency. Its noise performance is comparable to SPADs made in custom processes. In contrast to custom SPADs, CMOS SPADs offer integration capability with possible new

modes of operation. In particular, different approaches of parallelization could become possible with these sensors. This will increase the sensitivity and live-cell compatibility in confocal microscopy. Further designs however need to increase fill factor (with use of microlenses) and dynamic range by implementing integrated recharge mechanisms.

5.3 Conclusion

We showed a small format SPAD imager with cascode quenching and recharge architecture. This architecture allowed us to operate the sensor at 11 V of excess bias V_{EX}. The sensor shows a 10% PDP improvement and reduced afterpulsing. The timing jitter however increased by 20%, probably due to introduction of the cascode transistor and a fixed inverter threshold. Further optimization can be achieved by adjusting the inverter and cascode operating point in future designs. A lower inverter threshold can reduce the jitter.

We tested multiple recharge mechanisms investigating the response, shot noise and SNR. We introduced a new paradigm on the dynamic range by quantitative estimating the SNR. Interestingly, we found that clock recharge mechanisms do not increase afterpulsing dramatically. The inter-arrival histograms offer an intuitive way to understand why clock recharge features larger response linearity.

6 Conclusion

The main research question proposed in the introduction, "*What are the advantages of SPAD imagers overcoming the main challenges, used in microscopy, compared to conventional EMCCD and sCMOS imager technologies?",* has been thoroughly answered throughout the chapters. Due to direct photon-to-digital transformation, SPAD imagers count single photons. SPADs eliminate readout noise and tradeoffs between bandwidth and noise, shrinking CMOS nodes and dynamic range (full well capacity). Without readout noise, imagers can be oversampled for better timing resolution. Shrinking CMOS technology nodes potentially extend the dynamic range due to increased integration capability for quenching and recharge. SPAD technology combines single photon detection with picosecond timing resolution and megapixel sizing. A superb microscopy detector should have a frame rate (or parameter based output, i.e. lifetime output) of 30 fps, timing resolution in order of 100 ps, array size between 32×32 and 2048×2048, photon-counting capability for shot noise limited imaging and spectral resolvability.

Charge accumulating CMOS imagers need to reduce the readout noise throughout the array below 0.3 e^- for photon counting. The nonuniform readout distribution[33,103] makes this a hard requirement. SPAD DCR distribution has a similar uniformity challenge. We found 2% of nanoSPAD pixels above 1 kcps. DCR nonuniformity reduces with temperature. Afterpulsing and crosstalk do not increase nonuniformity. A uniform V_{BD} distribution and analog-less detection limits photon response nonuniformity below 2%. Digital robustness mitigates analog noise sources found in charge accumulating imagers. Initially small figures for fill factor and PDP are increased from 1-5% to 45% (3D integration) and 60% (microlenses) for fill factor and to 55% PDP for a substrate isolated CMOS SPAD.

In this thesis, we created new models for SPAD count response, count standard deviation and signal-to-noise ratio for active event-driven, passive and clock recharge

A deviation from the active event-driven response was experimentally observed with SwissSPAD for the first time. We then found differences between *count response* for event-driven and clock recharge, creating mathematical and simulation-based models that had a major impact in the community[14]. We also unexpectedly measured a reduction of count standard deviation below shot noise, and confirmed it with simulations. Later on, we confirmed our *count*

standard deviation and *signal-to-noise ratio* measurement with mathematical models[18,52,53].

Interestingly, we both mathematically and experimentally found that active clock recharge has better linearity compared to active event-driven recharge. However, active event-driven recharge theoretically extends the dynamic range by 75× and 26× (section 2.2.5.1) compared to $1/T_{dead}$ and active clock recharge, respectively. We measured an extension by 14× and 7.9× (section 5.1.4) compared to $1/T_{dead}$ and active clock recharge, respectively. It is not clear why a part of SPAD community continues to use clock recharge to extend the dynamic range[161].

The SPAD community has no standardized quantitative measure for the extension of dynamic range due to the suppressed count response, mainly citing $1/T_{dead}$[162] or linearity of the corrected response[163] as limiting the dynamic range. For this reason, we proposed a quantitative SNR decrease of -3dB as a measure for the extended dynamic range.

We investigated active clock recharge and found small increase in afterpulsing in our particular implementation (section 5.1.4).

nanoSPAD is based on a previous SPAD implementation with 45% peak PDP[44]. Using a cascode transistor combination for quenching and recharge, the array operates at high 11 V excess bias, with 55% peak PDP at 520 nm and >40% from 440 to 640 nm. The array shows a median DCR below 400 cps for excess bias up to 11 V. Capacitive isolation allowed us to reduce the afterpulsing from 7.2% at T_{dead}=300 ns[44] to 0.1% at T_{dead}=50. The timing jitter was however increased from 100 to 120 ps, presumably due to cascode quenching that increases the output resistance and a fixed inverter threshold.

We derived the optimum frame time for localization super resolution microscopy, based on exponentially distributed fluorophore blinking. An optimum resolution was also confirmed with experimental results. With consecutive publications in 2015[15], 2016[16] and 2017[17], we showed first localization super resolution microscopy acquired with a SPAD imager, improving image quality throughout the publications. In addition to the super resolution images, we were able to perform first widefield blinking analyses with timing resolution in the µs range.

Newly designed event-driven SPAD imagers with 144×252 and 32×32 pixels (shown in chip gallery) are good fits for fluorescence lifetime imaging, multi-laser STED and time-correlated light sheet microscopy.

Bibliography

1 Abbe E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch für mikroskopische Anat* 1873; **9**: 413–468.

2 Rayleigh, Lord. On the theory of optical images, with special reference to the microscope. *Philos Mag* 1896; **42**: 167–195.

3 Hell SW. Microscopy and its focal switch. *Nat Methods* 2009; **6**: 24–32. doi:10.1038/nmeth.1291.

4 Fölling J, Bossi M, Bock H, Medda R, Wurm CA, Hein B *et al.* Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nat Methods* 2008; **5**: 943–5. doi:10.1038/nmeth.1257.

5 Heilemann M, van de Linde S, Schüttpelz M, Kasper R, Seefeldt B, Mukherjee A *et al.* Subdiffraction-Resolution Fluorescence Imaging with Conventional Fluorescent Probes. *Angew Chemie Int Ed* 2008; **47**: 6172–6176. doi:10.1002/anie.200802376.

6 Ober RJ, Ram S, Ward ES. Localization accuracy in single-molecule microscopy. *Biophys J* 2004; **86**: 1185–200. doi:10.1016/S0006-3495(04)74193-4.

7 Van de Linde S, Wolter S, Heilemann M, Sauer M. The effect of photoswitching kinetics and labeling densities on super-resolution fluorescence imaging. *J Biotechnol* 2010; **149**: 260–266. doi:10.1016/j.jbiotec.2010.02.010.

8 Thompson RE, Larson DR, Webb WW. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J* 2002; **82**: 2775–83. doi:10.1016/S0006-3495(02)75618-X.

9 Huang F, Hartwich TMP, Rivera-Molina FE, Lin Y, Duim WC, Long JJ *et al.* Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat Methods* 2013; **10**: 653–8. doi:10.1038/nmeth.2488.

10 Lin Y, Long JJ, Huang F, Duim WC, Kirschbaum S, Zhang Y *et al.* Quantifying and Optimizing Single-Molecule Switching Nanoscopy at High Speeds. *PLoS One* 2015; **10**: e0128135. doi:10.1371/journal.pone.0128135.

11 Rochas A, Gosch M, Serov A, Besse PA, Popovic RS, Lasser T *et al.* First fully integrated 2-D array of single-photon detectors in standard CMOS technology. *IEEE Photonics Technol Lett* 2003; **15**: 963–965. doi:10.1109/LPT.2003.813387.

12 Gerritsen HC, Asselbergs MAH, Agronskaia A V., Van Sark WGJHM. Fluorescence lifetime imaging in scanning microscopes: acquisition speed, photon economy and lifetime resolution. *J Microsc* 2002; **206**: 218–224. doi:10.1046/j.1365-2818.2002.01031.x.

13 Burri S, Maruyama Y, Michalet X, Regazzoni F, Bruschini C, Charbon E. Architecture and applications of a high resolution gated SPAD image sensor. *Opt Express* 2014; **22**: 17573–89. doi:10.1364/OE.22.017573.

14 Antolovic IM, Burri S, Bruschini C, Hoebe R, Charbon E. Nonuniformity Analysis of a 65-kpixel CMOS SPAD Imager. *IEEE Trans Electron Devices* 2016; **63**: 57–64. doi:10.1109/TED.2015.2458295.

15 Antolovic IM, Burri S, Bruschini C, Hoebe RA, Charbon E. Super resolution with SPAD imagers. In: *Focus on Microscopy*. Gottingen, 2015http://www.focusonmicroscopy.org/2015/PDF/286_Antolovic.pdf.

16 Antolovic IM, Burri S, Bruschini C, Hoebe R, Charbon E. Analyzing blinking effects in super resolution localization microscopy with single-photon SPAD imagers. In: *Proc. SPIE 9714 Single Molecule Spectroscopy and Superresolution Imaging IX*. International Society for Optics and Photonics, 2016, p 971406 doi:10.1117/12.2211430.

17 Antolovic IM, Burri S, Bruschini C, Hoebe RA, Charbon E. SPAD imagers for super resolution localization microscopy enable analysis of fast fluorophore blinking. *Sci Rep* 2017; **7**: 44108. doi:10.1038/srep44108.

18 Fossum ER. Modeling the Performance of Single-Bit and Multi-Bit Quanta Image Sensors. *IEEE J Electron Devices Soc* 2013; **1**: 166–174. doi:10.1109/JEDS.2013.2284054.

19 Teranishi N. Required Conditions for Photon-Counting Image Sensors. *IEEE Trans Electron Devices* 2012; **59**: 2199–2205. doi:10.1109/TED.2012.2200487.

20 Fishburn MW. *Fundamentals of CMOS Single-Photon Avalanche Diodes*. 2012.

21 Magnan P. Detection of visible photons in CCD and CMOS: A comparative view. *Nucl Instruments Methods Phys Res Sect A Accel Spectrometers, Detect Assoc Equip* 2003; **504**: 199–212. doi:10.1016/S0168-9002(03)00792-7.

22 Etoh T, Nguyen A, Kamakura Y, Shimonomura K, Le T, Mori N. The Theoretical Highest Frame Rate of Silicon Image Sensors. *Sensors* 2017; **17**: 483. doi:10.3390/s17030483.

23 Lee M-J, Kang H-S, Choi W-Y. Equivalent Circuit Model for Si Avalanche Photodetectors Fabricated in Standard CMOS Process. *IEEE Electron Device Lett* 2008; **29**: 1115–1117. doi:10.1109/LED.2008.2000717.

24 Hurkx GAM, Klaassen DBM, Knuvers MPG. A new recombination model for device simulation including tunneling. *IEEE Trans Electron Devices* 1992; **39**: 331–338. doi:10.1109/16.121690.

25 Wang X. *Noise in Sub-Micron CMOS Image Sensors*. 2008.

26 Niclass C, Gersbach M, Henderson R, Grant L, Charbon E. A Single Photon Avalanche Diode Implemented in 130-nm CMOS Technology. *IEEE J Sel*

Top Quantum Electron 2007; **13**: 863–869. doi:10.1109/JSTQE.2007.903854.

27 Richardson JA, Grant LA, Henderson RK. Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology. *IEEE Photonics Technol Lett* 2009; **21**: 1020–1022. doi:10.1109/LPT.2009.2022059.

28 Bronzi D, Villa F, Bellisai S, Markovic B, Tisa S, Tosi A *et al.* Low-noise and large-area CMOS SPADs with timing response free from slow tails. In: *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*. IEEE, 2012, pp 230–233 doi:10.1109/ESSDERC.2012.6343375.

29 Denvir DJ, Conroy E. Electron-multiplying CCD: the new ICCD. In: Johnson CB, Sinha D, Laplante PA (eds). . International Society for Optics and Photonics, 2003, p 164 doi:10.1117/12.457779.

30 Andor manual. https://physiology.case.edu/media/eq_manuals/eq_manual_ixon3_hardware_guide.pdf.

31 Robbins MS, Hadwen BJ. The noise performance of electron multiplying charge-coupled devices. *IEEE Trans Electron Devices* 2003; **50**: 1227–1232. doi:10.1109/TED.2003.813462.

32 Boukhayma A, Peizerat A, Enz C. Temporal Readout Noise Analysis and Reduction Techniques for Low-Light CMOS Image Sensors. *IEEE Trans Electron Devices* 2016; **63**: 72–78. doi:10.1109/TED.2015.2434799.

33 Boukhayma A, Peizerat A, Enz C. A Sub-0.5 Electron Read Noise VGA Image Sensor in a Standard CMOS Process. *IEEE J Solid-State Circuits* 2016; **51**: 2180–2191. doi:10.1109/JSSC.2016.2579643.

34 Bronzi D, Villa F, Tisa S, Tosi A, Zappa F. SPAD Figures of Merit for Photon-Counting, Photon-Timing, and Imaging Applications: A Review. *IEEE Sens J* 2016; **16**: 3–12. doi:10.1109/JSEN.2015.2483565.

35 Humer G, Peev M, Schaeff C, Ramelow S, Stipcevic M, Ursin R. A Simple and Robust Method for Estimating Afterpulsing in Single Photon Detectors. *J Light Technol* 2015; **33**: 3098–3107. doi:10.1109/JLT.2015.2428053.

36 Kindt WJ, van Zeijl HW, Middelhoek S. Optical Cross Talk in Geiger Mode Avalanche Photodiode Arrays: Modeling, Prevention and Measurement. In: *28th European Solid-State Device Research Conference European*. 1998, pp 192–195.

37 Ziarkash AW, Joshi SK, Stipčević M, Ursin R. Comparative study of afterpulsing behavior and models in single photon counting avalanche photo diode detectors. *arXiv* 2017.http://arxiv.org/abs/1701.03783 (accessed 3 Jan2018).

38 Shinohara T, Watanabe K, Arakawa S, Kawashima H, Kawashima A, Abe T *et al.* Three-dimensional structures for high saturation signals and crosstalk suppression in 1.20 μm pixel back-illuminated CMOS image sensor. In: *2013 IEEE International Electron Devices Meeting*. IEEE, 2013, pp 27.4.1–27.4.4 doi:10.1109/IEDM.2013.6724704.

39 Mase M, Kawahito S, Sasaki M, Wakamori Y, Furuta M. A wide dynamic range CMOS image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic A/D converters. *IEEE J Solid-State Circuits* 2005; **40**: 2787–2795. doi:10.1109/JSSC.2005.858477.

40 Li Z, Kawahito S, Yasutomi K, Kagawa K, Ukon J, Hashimoto M *et al.* A Time-Resolved CMOS Image Sensor With Draining-Only Modulation Pixels for Fluorescence Lifetime Imaging. *IEEE Trans Electron Devices* 2012; **59**: 2715–2722. doi:10.1109/TED.2012.2209179.

41 Arai T, Yonai J, Hayashida T, Ohtake H, van Kuijk H, Etoh TG. A 252-V/luxs, 16.7-Million-Frames-Per-Second 312-kpixel Back-Side-Illuminated Ultrahigh-Speed Charge-Coupled Device. *IEEE Trans Electron Devices* 2013; **60**: 3450–3458. doi:10.1109/TED.2013.2276005.

42 Etoh T, Son D, Yamada T, Charbon E. Toward One Giga Frames per Second — Evolution of in Situ Storage Image Sensors. *Sensors* 2013; **13**: 4640–4658. doi:10.3390/s130404640.

43 Bronzia D, Villaa F, Bellisaia S. Figures of Merit for CMOS SPADs and arrays. *… SPIE Vol* 2013.http://reviews.spiedigitallibrary.org/data/Conferences/SPIEP/74220/877304.pdf (accessed 7 Apr2014).

44 Veerappan C, Charbon E. A Low Dark Count p-i-n Diode Based SPAD in CMOS Technology. *IEEE Trans Electron Devices* 2016; **63**: 65–71. doi:10.1109/TED.2015.2475355.

45 Veerappan C, Charbon E. A Substrate Isolated CMOS SPAD Enabling Wide Spectral Response and Low Electrical Crosstalk. *IEEE J Sel Top Quantum Electron* 2014; **20**: 1–7. doi:10.1109/JSTQE.2014.2318436.

46 Ghioni M, Gulinatti A, Rech I, Maccagnani P, Cova S. Large-area low-jitter silicon single photon avalanche diodes. In: Sudharsanan R, Jelen C (eds). . International Society for Optics and Photonics, 2008, p 69001D doi:10.1117/12.761578.

47 Arlt J, Tyndall D, Rae BR, Li DD-U, Richardson JA, Henderson RK. A study of pile-up in integrated time-correlated single photon counting systems. *Rev Sci Instrum* 2013; **84**: 103105. doi:10.1063/1.4824196.

48 Isbaner S, Karedla N, Ruhlandt D, Stein SC, Chizhik A, Gregor I *et al.* Dead-time correction of fluorescence lifetime measurements and fluorescence lifetime imaging. *Opt Express* 2016; **24**: 9429. doi:10.1364/OE.24.009429.

49 Lee SH, Gardner RP. A new G–M counter dead time model. *Appl Radiat Isot* 2000; **53**: 731–737. doi:10.1016/S0969-8043(00)00261-X.

50 Niclass C, Sergio M, Charbon E. A single photon avalanche diode array fabricated in 0.35-µm CMOS and based on an event-driven readout for TCSPC experiments. In: Becker W (ed). *Optics East 2006*. International Society for Optics and Photonics, 2006, p 63720S–63720S–12 doi:10.1117/12.685974.

51 Knoll GF. *Radiation Detection and Measurement*. Wiley, 2000.

52 Sbaiz L, Yang F, Charbon E, Susstrunk S, Vetterli M. The gigavision camera. In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp 1093–1096 doi:10.1109/ICASSP.2009.4959778.

53 Lee SH, Jae M. Non-Poisson counting statistics of a hybrid G–M counter dead time model. *Nucl Instruments Methods Phys Res Sect B Beam Interact with Mater Atoms* 2007; **263**: 46–49. doi:10.1016/j.nimb.2007.04.041.

54 Wenwen Zhang, Qian Chen. Signal-to-noise ratio performance comparison of Electron Multiplying CCD and Intensified CCD detectors. In: *2009 International Conference on Image Analysis and Signal Processing*. IEEE, 2009, pp 337–341 doi:10.1109/IASP.2009.5054588.

55 ORCA Flash 4. https://www.hamamatsu.com/resources/pdf/sys/SCAS0099E_C11440-42U.pdf.

56 Andor iXon 897 Specifications. www.andor.com/pdfs/specifications/Andor_iXon_897_Specifications.pdf.

57 Yang DXD, El Gamal A. Comparative analysis of SNR for image sensors with enhanced dynamic range. In: Blouke MM, Williams, Jr. GM (eds). . International Society for Optics and Photonics, 1999, pp 197–211 doi:10.1117/12.347075.

58 Chen S, Ceballos A, Fossum ER. Digital integration sensor. In: *International Image Sensor Workshop*. 2013.

59 Fossum E. The Quanta Image Sensor (QIS): Concepts and Challenges. In: *Imaging and Applied Optics*. OSA: Washington, D.C., 2011, p JTuE1 doi:10.1364/COSI.2011.JTuE1.

60 Dutton NAW, Gyongy I, Parmesan L, Gnecchi S, Calder N, Rae BR *et al.* A SPAD-Based QVGA Image Sensor for Single-Photon Counting and Quanta Imaging. *IEEE Trans Electron Devices* 2016; **63**: 189–196. doi:10.1109/TED.2015.2464682.

61 Dutton NAW, Gyongy I, Parmesan L, Henderson RK. Single Photon Counting Performance and Noise Analysis of CMOS SPAD-Based Image Sensors. *Sensors (Basel)* 2016; **16**. doi:10.3390/s16071122.

62 Tochigi Y, Hanzawa K, Kato Y, Kuroda R, Mutoh H, Hirose R *et al.* A Global-Shutter CMOS Image Sensor With Readout Speed of 1-Tpixel/s Burst and 780-Mpixel/s Continuous. *IEEE J Solid-State Circuits* 2013; **48**: 329–338. doi:10.1109/JSSC.2012.2219685.

63 Dutton NAW, Gnecchi S, Parmesan L, Holmes AJ, Rae B, Grant LA *et al.* A time-correlated single-photon-counting sensor with 14GS/S histogramming time-to-digital converter. In: *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*. IEEE, 2015, pp 1–3 doi:10.1109/ISSCC.2015.7062997.

64 El-Desouki M, Jamal Deen M, Fang Q, Liu L, Tse F, Armstrong D. CMOS Image Sensors for High Speed Applications. *Sensors* 2009; **9**: 430–444. doi:10.3390/s90100430.

65 Walden RH. Analog-to-digital converter survey and analysis. *IEEE J Sel Areas Commun* 1999; **17**: 539–550. doi:10.1109/49.761034.

66 Mendis SK, Kemeny SE, Gee RC, Pain B, Staller CO, Quiesup Kim *et al.* CMOS active pixel image sensors for highly integrated imaging systems. *IEEE J Solid-State Circuits* 1997; **32**: 187–197. doi:10.1109/4.551910.

67 White MH, Lampe DR, Blaha FC, Mack IA. Characterization of surface channel CCD image arrays at low light levels. *IEEE J Solid-State Circuits* 1974; **9**: 1–12. doi:10.1109/JSSC.1974.1050448.

68 Cova S, Longoni A, Andreoni A. Towards picosecond resolution with single-photon avalanche diodes. *Rev Sci Instrum* 1981; **52**: 408–412. doi:10.1063/1.1136594.

69 Webb P, McIntyre R, Conradi J. Properties of avalanche photodiodes. *RCA Rev* 1974.http://scholar.google.nl/scholar?cluster=852118840736228062&hl=hr&oi=scholarr&sa=X&ved=0ahUKEwiGwqe5t6TKAhXFjg8KHe7fC8IQgAMIJCgAMAA#0 (accessed 12 Jan2016).

70 Webb PP, McIntyre RJ. Large Area Reach-Through Avalanche Diodes for X-Ray Spectroscopy. *IEEE Trans Nucl Sci* 1976; **23**: 138–144. doi:10.1109/TNS.1976.4328228.

71 Webb P, Jones AR. Large Area Reach-Through Avalanche Diodes for Radiation Monitoring. *IEEE Trans Nucl Sci* 1974; **21**: 151–158. doi:10.1109/TNS.1974.4327455.

72 Rochas A, Gosch M, Serov A, Besse PA, Popovic RS, Lasser T *et al.* First fully integrated 2-D array of single-photon detectors in standard CMOS technology. *IEEE Photonics Technol Lett* 2003; **15**: 963–965. doi:10.1109/LPT.2003.813387.

73 Abbas T Al, Dutton NAW, Almer O, Pellegrini S, Henrion Y, Henderson RK. Backside illuminated SPAD image sensor with 7.83μm pitch in 3D-stacked CMOS technology. In: *2016 IEEE International Electron Devices*

Meeting (IEDM). IEEE, 2016, pp 8.1.1–8.1.4 doi:10.1109/IEDM.2016.7838372.

74 Mandai S, Fishburn MW, Maruyama Y, Charbon E. A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology. *Opt Express* 2012; **20**: 5849. doi:10.1364/OE.20.005849.

75 Ulku AC, Bruschini C, Michalet X, Weiss S, Charbon E. A 512×512 SPAD Image Sensor with Built-In Gating for Phasor Based Real-Time siFLIM. In: *International Image Sensor Workshop*. 2017.

76 Veerappan C, Richardson J, Walker R, Li D-U, Fishburn MW, Maruyama Y *et al.* A 160×128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter. In: *2011 IEEE International Solid-State Circuits Conference*. IEEE, 2011, pp 312–314 doi:10.1109/ISSCC.2011.5746333.

77 Mandai S, Jain V, Charbon E. A fully-integrated 780×800μm^2 multi-digital silicon photomultiplier with column-parallel time-to-digital converter. In: *2012 Proceedings of the ESSCIRC (ESSCIRC)*. IEEE, 2012, pp 89–92 doi:10.1109/ESSCIRC.2012.6341263.

78 Maruyama Y, Blacksberg J, Charbon E. A 1024x8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS. *IEEE J Solid-State Circuits* 2014; **49**: 179–189. doi:10.1109/JSSC.2013.2282091.

79 Gersbach M, Trimananda R, Maruyama Y, Fishburn M, Stoppa D, Richardson J *et al.* High frame-rate TCSPC-FLIM using a novel SPAD-based image sensor. In: Dereniak EL, Hartke JP, LeVan PD, Sood AK, Longshore RE, Razeghi M (eds). *SPIE NanoScience + Engineering*. International Society for Optics and Photonics, 2010, p 77801H–77801H–13 doi:10.1117/12.860769.

80 Villa F, Markovic B, Bellisai S, Bronzi D, Tosi A, Zappa F *et al.* SPAD Smart Pixel for Time-of-Flight and Time-Correlated Single-Photon Counting Measurements. *IEEE Photonics J* 2012; **4**: 795–804. doi:10.1109/JPHOT.2012.2198459.

81 Lindner S, Zhang C, Antolovic IM, Pavia JM, Wolf M, Charbon E. Column-Parallel Dynamic TDC Reallocation in SPAD Sensor Module Fabricated in 180nm CMOS for Near Infrared Optical Tomography. In: *International Image Sensor Workshop*. 2017.

82 Webster EAG. Single-photon avalanche diodes in 90 nm CMOS imaging technology with sub-1 Hz median dark count rate. In: *IISW*. 2011.

83 Niclass C, Favi C, Kluter T, Gersbach M, Charbon E. A 128x128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array. *IEEE J Solid-State Circuits* 2008; **43**: 2977–2989. doi:10.1109/JSSC.2008.2006445.

84 Donati S, Martini G, Norgia M. Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits. *Opt Express* 2007; **15**: 18066. doi:10.1364/OE.15.018066.

85 Donati S, Martini G, Randone E. Improving Photodetector Performance by Means of Microoptics Concentrators. *J Light Technol* 2011; **29**: 661–665. doi:10.1109/JLT.2010.2103302.

86 Pavia JM, Wolf M, Charbon E. Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery. *Opt Express* 2014; **22**: 4202–13. doi:10.1364/OE.22.004202.

87 M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, D. Stoppa. A 160x120 Pixel Analog-Counting Single-Photon Imager With Time-Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging. *IEEE J Solid-State Circuits* 2016; **51**: 155–167. doi:10.1109/JSSC.2015.2482497.

88 Luca P, Dutton N, Krstajic N, Calder N, Holmes A, Grant LA *et al.* A 256 x 256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy. Int. Image Sens. Work. 2015.

89 Krstajić N, Poland S, Levitt J, Walker R, Erdogan A, Ameer-Beg S *et al.* 0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays. *Opt Lett* 2015; **40**: 4305–8. doi:10.1364/OL.40.004305.

90 Gerega A, Zolek N, Soltysinski T, Milej D, Sawosz P, Toczylowska B *et al.* Wavelength-resolved measurements of fluorescence lifetime of indocyanine green. *J Biomed Opt* 2011; **16**: 067010. doi:10.1117/1.3593386.

91 Homulle HAR, Powolny F, Stegehuis PL, Dijkstra J, Li D-U, Homicsko K *et al.* Compact solid-state CMOS single-photon detector array for in vivo NIR fluorescence lifetime oncology measurements. *Biomed Opt Express* 2016; **7**: 1797. doi:10.1364/BOE.7.001797.

92 Gyongy I, Davies A, Dutton NAW, Duncan R, Rickman C, Henderson RK *et al.* Smart-aggregation imaging for single molecule localisation with SPAD cameras. *Sci Rep* 2016; **6**: 37349. doi:10.1038/srep37349.

93 pco.edge 4.2 sCMOS datasheet. https://www.pco.de/fileadmin/user_upload/pco-product_sheets/pco.edge_gold_42_data_sheet.pdf.

94 Mata Pavia J, Wolf M, Charbon E. A Dual Backside-Illuminated 800-Cell Multi-Channel Digital SiPM with 100 TDCs in 130nm 3D IC Technology. In: *IEEE Nuclear Science Symposium*. 2014.

95 Lee M-J, Choi W-Y. Area-Dependent Photodetection Frequency Response Characterization of Silicon Avalanche Photodetectors Fabricated With Standard CMOS Technology. *IEEE Trans Electron Devices* 2013; **60**: 998–1004. doi:10.1109/TED.2013.2240684.

96 OmniVision. Omnivision OS02A1Q Product Brief. http://www.ovt.com/image-sensors/2-5-megapixels.

97 Sony 16.35 Mpixel Image Sensor. http://www.sony-semicon.co.jp/products_en/new_pro/december_2013/imx206_e.html.

98 ON Semiconductor 6.6 Mpixel power consumption. https://www.onsemi.com/pub/Collateral/NOII4SM6600A-D.PDF.

99 Lee M-J, Sun P, Charbon E. A first single-photon avalanche diode fabricated in standard SOI CMOS technology with a full characterization of the device. *Opt Express* 2015; **23**: 13200. doi:10.1364/OE.23.013200.

100 Gallivanoni A, Rech I, Ghioni M. Progress in Quenching Circuits for Single Photon Avalanche Diodes. *IEEE Trans Nucl Sci* 2010. doi:10.1109/TNS.2010.2074213.

101 Veerappan C. *Single-Photon Avalanche Diodes for Cancer Diagnosis*. 2016.

102 Gersbach M, Maruyama Y, Trimananda R, Fishburn MW, Stoppa D, Richardson JA *et al.* A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology. *IEEE J Solid-State Circuits* 2012; **47**: 1394–1407. doi:10.1109/JSSC.2012.2188466.

103 Min-Woong Seo, Kawahito S, Kagawa K, Yasutomi K. A 0.27e-rms Read Noise 220-μV/e-Conversion Gain Reset-Gate-Less CMOS Image Sensor With 0.11-μm CIS Process. *IEEE Electron Device Lett* 2015; **36**: 1344–1347. doi:10.1109/LED.2015.2496359.

104 Sugawa S, Kuroda R, Takeda T, Shao F, Miyauchi, Ken, Tochigi Y. A 20Mfps Global Shutter CMOS Image Sensor with Improved Sensitivity and Power Consumption. In: *International Image Sensor Workshop*. 2015.

105 Ma C, Liu Y, Li J, Zhou Q, Chang Y, Wang X. A 4MP high-dynamic-range, low-noise CMOS image sensor. In: Widenhorn R, Dupret A (eds). . International Society for Optics and Photonics, 2015, p 940305 doi:10.1117/12.2083085.

106 Schermelleh L, Heintzmann R, Leonhardt H. A guide to super-resolution fluorescence microscopy. *J Cell Biol* 2010; **190**: 165–75. doi:10.1083/jcb.201002018.

107 Demmerle J, Wegel E, Schermelleh L, Dobbie IM. Assessing resolution in super-resolution imaging. *Methods* 2015; **88**: 3–10. doi:10.1016/j.ymeth.2015.07.001.

108 Heintzmann R, Cremer CG. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. In: Bigio IJ, Schneckenburger H, Slavik J, Svanberg K, Viallet PM (eds). . International Society for Optics and Photonics, 1999, pp 185–196 doi:10.1117/12.336833.

109 Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J Microsc* 2000; **198**: 82–7.

110 Hell SW, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt Lett* 1994; **19**: 780. doi:10.1364/OL.19.000780.

111 Dyba M, Hell SW. Focal Spots of Size λ / 23 Open Up Far-Field Florescence Microscopy at 33 nm Axial Resolution. *Phys Rev Lett* 2002; **88**: 163901. doi:10.1103/PhysRevLett.88.163901.

112 Dyba M, Jakobs S, Hell SW. Immunofluorescence stimulated emission depletion microscopy. *Nat Biotechnol* 2003; **21**: 1303–1304. doi:10.1038/nbt897.

113 Donnert G, Keller J, Medda R, Andrei MA, Rizzoli SO, Lührmann R *et al.* Macromolecular-scale resolution in biological fluorescence microscopy. *Proc Natl Acad Sci U S A* 2006; **103**: 11440–5. doi:10.1073/pnas.0604965103.

114 Heller I, Sitters G, Broekmans O, Farge G, Menges C. STED nanoscopy combined with optical tweezers reveals protein dynamics on densely covered DNA. *Nature* 2013.https://www.nature.com/articles/nmeth.2599 (accessed 27 Jun2017).

115 Göttfert F, Pleiner T, Heine J, Westphal V, Görlich D, Sahl SJ *et al.* Strong signal increase in STED fluorescence microscopy by imaging regions of subdiffraction extent. *Proc Natl Acad Sci U S A* 2017; **114**: 2125–2130. doi:10.1073/pnas.1621495114.

116 Hedde PN, Nienhaus GU. Super-resolution localization microscopy with photoactivatable fluorescent marker proteins. *Protoplasma* 2014; **251**: 349–362. doi:10.1007/s00709-013-0566-z.

117 Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science (80-)* 2006; **313**: 1642–1645. doi:10.1126/science.1127344.

118 Hess ST, Girirajan TPK, Mason MD. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J* 2006; **91**: 4258–72. doi:10.1529/biophysj.106.091116.

119 Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* 2006; **3**: 793–796. doi:10.1038/nmeth929.

120 Olivier N, Keller D, Gönczy P, Manley S. Resolution doubling in 3D-STORM imaging through improved buffers. *PLoS One* 2013; **8**: e69004. doi:10.1371/journal.pone.0069004.

121 Nahidiazar L, Agronskaia A V, Broertjes J, van den Broek B, Jalink K. Optimizing Imaging Conditions for Demanding Multi-Color Super Resolution Localization Microscopy. *PLoS One* 2016; **11**: e0158884. doi:10.1371/journal.pone.0158884.

122 Jungmann R, Steinhauer C, Scheible M, Kuzyk A, Tinnefeld P, Simmel FC. Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami. *Nano Lett* 2010; **10**: 4756–4761. doi:10.1021/nl103427w.

123 GATTAquant paint nanoruler. 2016.http://www.gattaquant.com/products/localization-based/gatta-paint-nanoruler.html.

124 Thompson RE, Larson DR, Webb WW. Precise nanometer localization analysis for individual fluorescent probes. *Biophys J* 2002; **82**: 2775–83. doi:10.1016/S0006-3495(02)75618-X.

125 Ovesný M, Křížek P, Borkovec J, Svindrych Z, Hagen GM. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* 2014; **30**: 2389–90. doi:10.1093/bioinformatics/btu202.

126 Gyongy I, Davies A, Dutton NAW, Duncan R, Rickman C, Henderson RK *et al.* Smart-aggregation imaging for single molecule localization with SPAD cameras. 2016.http://arxiv.org/abs/1602.04191 (accessed 9 Nov2016).

127 Huang Z-L, Zhu H, Long F, Ma H, Qin L, Liu Y *et al.* Localization-based super-resolution microscopy with an sCMOS camera. *Opt Express* 2011; **19**: 19156–68. doi:10.1364/OE.19.019156.

128 Ma H, Kawai H, Toda E, Zeng S, Huang Z-L. Localization-based super-resolution microscopy with an sCMOS camera part III: camera embedded data processing significantly reduces the challenges of massive data handling. *Opt Lett* 2013; **38**: 1769–71. doi:10.1364/OL.38.001769.

129 Long F, Zeng S, Huang Z-L. Localization-based super-resolution microscopy with an sCMOS camera part II: experimental methodology for comparing sCMOS with EMCCD cameras. *Opt Express* 2012; **20**: 17741–59. doi:10.1364/OE.20.017741.

130 Quan T, Zeng S, Huang Z-L. Localization capability and limitation of electron-multiplying charge-coupled, scientific complementary metal-oxide semiconductor, and charge-coupled devices for superresolution imaging. *J Biomed Opt* 2010; **15**: 066005. doi:10.1117/1.3505017.

131 Hoogendoorn E, Crosby KC, Leyton-Puig D, Breedijk RMP, Jalink K, Gadella TWJ *et al.* The fidelity of stochastic single-molecule super-resolution reconstructions critically depends upon robust background estimation. *Sci Rep* 2014; **4**: 231–254. doi:10.1038/srep03854.

132 Krishnaswami V, Van Noorden CJF, Manders EMM, Hoebe RA. Towards digital photon counting cameras for single-molecule optical nanoscopy. *Opt Nanoscopy* 2014; **3**: 1. doi:10.1186/2192-2853-3-1.

133 Yip W-T, Hu D, Yu J, Vanden Bout DA, Barbara PF. Classifying the Photophysical Dynamics of Single- and Multiple-Chromophoric

Molecules by Single Molecule Spectroscopy. *J Phys Chem A* 1998; **102**: 7564–7575. doi:10.1021/jp981808x.

134 Vogelsang J, Steinhauer C, Forthmann C, Stein IH, Person-Skegro B, Cordes T *et al.* Make them blink: probes for super-resolution microscopy. *Chemphyschem* 2010; **11**: 2475–90. doi:10.1002/cphc.201000189.

135 Stein IH, Capone S, Smit JH, Baumann F, Cordes T, Tinnefeld P. Linking single-molecule blinking to chromophore structure and redox potentials. *Chemphyschem* 2012; **13**: 931–7. doi:10.1002/cphc.201100820.

136 Alexa Fluor 647 quantum yield and lifetime. https://www.thermofisher.com/nl/en/home/references/molecular-probes-the-handbook/tables/fluorescence-quantum-yields-and-lifetimes-for-alexa-fluor-dyes.html.

137 Lounis B, Moerner WE. Single photons on demand from a single molecule at room temperature. *Nature* 2000; **407**: 491–493. doi:10.1038/35035032.

138 Fernandez-Suarez M, Ting AY. Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Biol* 2008; **9**: 929–943. doi:10.1038/nrm2531.

139 Alexa Fluor 647 extinction coefficient. https://www.thermofisher.com/nl/en/home/references/molecular-probes-the-handbook/technical-notes-and-product-highlights/the-alexa-fluor-dye-series.html#.

140 Pawley JB. *Handbook of biological confocal microscopy*. Third Edit. Springer, 2006.

141 Israel Y, Tenne R, Oron D, Silberberg Y. Quantum correlation enhanced super-resolution localization microscopy enabled by a fibre bundle camera. *Nat Commun* 2017; **8**: 14786. doi:10.1038/ncomms14786.

142 Zhang B, Zerubia J, Olivo-Marin J-C. Gaussian approximations of fluorescence microscope point-spread function models. *Appl Opt* 2007; **46**: 1819. doi:10.1364/AO.46.001819.

143 Donnert G, Eggeling C, Hell SW. Major signal increase in fluorescence microscopy through dark-state relaxation. *Nat Methods* 2007; **4**: 81–86. doi:10.1038/nmeth986.

144 Huff J. The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. *Nat Methods* 2015; **12**. doi:10.1038/nmeth.f.388.

145 Huff J, Bathe W, Netz R, Anhut T, Weisshart K. The Airyscan Detector from ZEISS. Jena, 2015https://applications.zeiss.com/C125792900358A3F/0/BF16BECDC08849E1C1257ED60029AF8D/$FILE/EN_41_013_105_wp_Airyscan-detector.pdf.

146 Sheppard CJ. Super-resolution in confocal imaging. *Optik (Stuttg)* 1988; **80**: 53–54.

147 Castello M, Sheppard CJR, Diaspro A, Vicidomini G. Image scanning microscopy with a quadrant detector. *Opt Lett* 2015; **40**: 5355. doi:10.1364/OL.40.005355.

148 Sheppard CJR, Castello M, Vicidomini G, Duocastella M, Diaspro A. Microscopy using source and detector arrays. In: Brown TG, Cogswell CJ, Wilson T (eds). . International Society for Optics and Photonics, 2016, p 971302 doi:10.1117/12.2213544.

149 Lindner S, Pellegrini S, Henrion Y, Rae B, Wolf M, Charbon E. A High-PDE, Backside-Illuminated SPAD in 65/40 nm 3D IC CMOS Pixel with Cascoded Passive Quenching and Active Recharge. *IEEE Electron Device Lett* 2017. doi:10.1109/LED.2017.2755989.

150 Polyakov S V., Migdall AL. High accuracy verification of a correlated-photon- based method for determining photoncounting detection efficiency. *Opt Express* 2007; **15**: 1390. doi:10.1364/OE.15.001390.

151 Stipčević M, Christensen BG, Kwiat PG, Gauthier DJ. Advanced active quenching circuit for ultra-fast quantum cryptography. *Opt Express* 2017; **25**: 21861. doi:10.1364/OE.25.021861.

152 ID Quantique ID100 VIS datasheet. http://marketing.idquantique.com/acton/attachment/11868/f-013d/1/-/-/-/-/ID100 Datasheet.pdf.

153 MPD PDM datasheet. http://www.micro-photon-devices.com/Docs/Datasheet/PDM.pdf.

154 Excelitas SPCM-AQRH datasheet. http://www.excelitas.com/Downloads/DTS_SPCM-AQRH.pdf.

155 MPD SPC2 datasheet. http://www.micro-photon-devices.com/Docs/Datasheet/SPC2.pdf.

156 Hamamatsu h7422 GaAsP PMT datasheet. https://www.hamamatsu.com/resources/pdf/etd/m-h7422e.pdf.

157 Hamamatsu hybrid photodetector R10467U-40 datasheet. https://www.hamamatsu.com/resources/pdf/etd/HPD_TPMH1361E.pdf.

158 Picoquant hybrid PMA-40 datasheet. https://www.picoquant.com/images/uploads/downloads/pma_hybrid.pdf.

159 Becker W, Bergmann A. Detectors for High-Speed Photon Counting. ; : 1–22.

160 Gerritsen HC, Bader A, Agronskaia S. Laser Scanning Confocal FLIM Microscopy. In: *Flim Microscopy in Biology and Medicine*. 2009, p 158.

161 Dutton N, Al abbas T, Gyongy I, Henderson R. Extending the Dynamic Range of Oversampled Binary SPAD Image Sensors. In: *International Image Sensor Workshop*. 2017.

162 Eisele A, Henderson R, Schmidtke B, Funk T, Grant L, Richardson J *et al.* 185 MHz count rate 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130 nm CMOS technology. In: *International Image Sensor Workshop*. 2011.

163 Stoppa D, Mosconi D, Pancheri L, Gonzo L. Single-Photon Avalanche Diode CMOS Sensor for Time-Resolved Fluorescence Measurements. *IEEE Sens J* 2009; **9**: 1084–1090. doi:10.1109/JSEN.2009.2025581.

Summary

Microscopy is an integral part of the Dutch scientific tradition. Van Leeuwenhoek was a trailblazer in microscopy, using it as a tool to explore a new world of microorganisms. For a long time, it was believed that microscopes could reach any desired magnification, and thus any spatial resolution, just by improving the lens quality. It was later found that the resolution of optical microscopes is fundamentally limited by diffraction. Ernst Abbe defined the maximum optical resolution to be around 0.2 micrometers. We could track ants, cells and bacteria, but not viruses, proteins and small molecules. However, in 2014 scientists were awarded the Nobel Prize in Chemistry 2014 for bypassing this resolution limit. The microscopes can now delve into the nanoworld, with resolutions down to 10 nanometers. We can now observe individual molecules, viruses and proteins.

The aim of this research is to explore the potential advantages of SPAD imagers used in microscopy. An ideal microscopy detector requires *high sensitivity* (high quantum efficiency QE or photon detection probability PDP), *photon counting* operation, *low noise* (dark current or dark count rate), *timing resolution* in the order of 100 ps, *frame rate* higher than 10 fps, a large enough *pixel resolution* and wavelength resolvability.

High sensitivity is required as to increase the number of collected photons and the signal-to-noise ratio (SNR). High sensitivity enables to reduce the laser/LED power and reduce phototoxicity of a microscopy sample. The SNR is maximized in *photon counting* operation, and equals the shot noise limited SNR=$n/\sqrt{n}=\sqrt{n}$ (if dark noise is neglected), where n is the number of collected photons. To operate in *photon counting* mode, conventional charge accumulating imagers (both CMOS and CCD) require either a very low readout noise or high photoelectron multiplication that effectively reduces the readout noise. In addition to the readout noise, charge accumulating imagers add other analog noise sources.

In CMOS charge accumulating imagers, other analog noise sources majorly cause the photon response nonuniformity (PRNU), stemming from column and pixel mismatches in source followers, column amplifiers and column analog-to-digital converters.

Single-photo avalanche diode (SPAD) imagers are direct photon-to-digital sensors and thus photon counting devices. They mitigate the major noise

source, i.e. the readout noise that limits *photon counting* in conventional charge accumulating CMOS imagers. Due to their digital nature, SPAD imagers limit the PRNU to the breakdown voltage nonuniformity. Since sensitivity is influenced by the voltage applied above the breakdown voltage, variations in the breakdown voltage yield PRNU. The breakdown voltage is found to be very uniform in modern CMOS processes, measuring 67 mV and 56 mV standard deviation for 350 and 180 nm CMOS, respectively. Such variations yield less than 2% PRNU.

Timing resolution is in microscopy required to separate two or more molecules (or molecule environment) with similar emission spectra (color). One can then use the molecule dynamic, i.e. its lifetime, for separation. The lifetime are exponentially distributed with a mean value between 500 ps and 5 ns. SPAD sensors usually feature a timing jitter in the order of 100 ps resolution, a specification satisfying lifetime measurements.

Frame rates above 30 fps are required to follow dynamic cell or molecule movement. SPAD imagers are in frame rate limited by the "dead time" T_{dead}, during which one can detect only one photon. If one needs a 10-bit image with 1024 gray levels, the maximum frame rate is $1/(1023\times T_{dead})$. Considering a typical T_{dead} of 50 ns and 10-bit image depth, one can achieve a very high 20 kfps.

Due to the digital nature of SPAD imagers, the frame rate is not entangled with the noise as a tradeoff. In contrast, high frame rate charge accumulating imagers feature high noise figures.

Photo multiplying tubes and hybrid detectors where also used as photon counting single-pixel detectors, scanning the field of view to construct as image. However, scanning take time and limits the frame rate to 1-10 fps. CMOS SPAD imagers are routinely implemented with high *pixel resolution*, to date up to 512×512.

SPAD imagers potentially represent an ideal microscopy detector. Photon generated SPAD pulses can be counted or/and used as time resolved signals for time-to-digital converters (TDCs). However, previous implementations featured low *fill factor* (ratio between sensitive area and overall pixel area) and *PDP*. Each SPAD imager also incorporates a certain amount of "hot" pixels with very high DCR and increases the *DCR nonuniformity*. The maximum count rate, limiting the *dynamic range*, is stated to be limited by $1/T_{dead}$.

We firstly used SwissSPAD, a 512×128 counting based SPAD imager with 1-bit in-pixel memory, fabricated in a high-voltage 350 nm CMOS process. It can be operated with a global shutter exposure with rolling readout outputting 1-bit frames each 6.4 µs. Depositing microlenses on top of the chip led to a major *fill factor* improvement. We increased the native fill factor from 5% to effectively 60%. A ×12 increase in sensitivity enabled us to acquire the first SPAD based localization super resolution images with at least 80 nm resolution. Due to the fact that SPAD imagers do not change their performance when changing the frame rate, we investigated the optimal frame rate for localization super resolution. We found that the frame rate needs to be adjusted considering the blinking dynamic found in microscopy samples. We mathematically derived an optimal frame rate depending on dye blinking, dye intensity and background light. This was particularly important due to the tendency to increase the frame rate as to shorten the acquisition times. We found a mathematical relationship between the frame rate and achieved resolution.

We further used SwissSPAD timing resolution of 6.4 µs to explore possible triplet-related blinking in the microsecond time range. We investigated different samples and found data indicating that the fast blinking is caused by molecular triplet states. This data was published in our Scientific Reports paper.

Using a standard 180 nm CMOS process and a p-i-n junction, we designed a SPAD pixel with 55% peak *PDP* at 520 nm. This is the *state-of-the-art PDP* figure for substrate isolated standard CMOS SPADs. Large SPAD arrays are majorly implemented with substrate isolated SPADs. Substrate shared SPADs achieve a higher PDP, due to a deeper junction, but increase DCR, afterpulsing and crosstalk due to a small or no potential barrier to p-substrate, which is shared with surrounding SPADs and electronics. Custom made SPADs can feature a PDP as high as 70%, but are not implemented in SPAD arrays due to reproducibility issues.

A cascode recharge schematic was a key implementation that enabled a high operating voltage and consequently a *high PDP* SPAD array. The implemented pixels further feature 0.6 and 3.7 cps/µm^2 DCR at 3 and 11 V excess bias, respectively. Afterpulsing probability is 0.1%, and crosstalk is 0.09%. The timing jitter was found to be 123 ps at 11 V excess bias.

The *DCR nonuniformity* was firstly evaluated with SwissSPAD measuring 1.8% noisy pixels with more than 212/µm^2 cps DCR. With the new 180 nm

technology, we found 4% pixels above 0.9 cps/μm^2 and 2% pixels above 9.3 cps/μm^2 at 3.3 V excess bias. In addition to the SPAD size, we found that the temperature influences the *DCR nonuniformity*. Lower operating temperatures decrease the *DCR nonuniformity*.

The maximum count rate, limiting the *dynamic range*, was thoroughly analyzed. It is well known that the SPAD response to light is not linear, and can potentially increase the dynamic range. Each recharge mode features a different response equation, enabling to linearize the response back. However, the standard deviation of counts due to this linearization was not evaluated before. We derived and measured the SPAD dynamic range considering a maximum acceptable decrease (-3dB) in SNR due to linearization, for three SPAD recharge modes. We evaluated passive and active event-driven recharge and clocked recharge. Active event-driven recharge increases the dynamic range by a factor of ×36 and ×75 for 8-bit and 10-bit image depth, respectively. A direct clocked recharge (asynchronous to the SPAD activity, with a high repetition rate) was never performed before, mostly due to potential increase in afterpulsing. We however found no afterpulsing increase with clocked recharge.

Lastly, the author codesigned a 144×252 and 32×32 event-driven SPAD array with Chao Zhang and Scott Lindner, using the 180 nm CMOS technology. The chips are optimized for a sweet spot between high speed and high fill factor. We partially parallelized TDCs resources at the bottom (and top for the larger array) of the pixel array. The pixel output is connected to a column bus and timing line, all pulled up to V_{DD}. At arrival of a photon, the bus address assigned to the specific pixel is forced to ground. The timing line is designed to enable a fast leading edge and conservation of the timing resolution. The 32×32 and 144×252 pixel array have a total of 128 and 1728 TDCs, respectively. The author designed the pixel array, address coding, pulse propagation and auxiliary signal integrity circuit.

Acknowledgments

From start of my PhD, I was captivated by the new surrounding and the given opportunity to explore science in Delft.

I would like to thank Professor Edoardo Charbon for including me in our current SPAD group and the SPAD community itself. His presence was key in my research maturation. Regular meetings with him sharpened my focus. I remember a very important event, during discussions about the *optimal frame duration*. At the beginning unclear about the underlying principles causing certain shapes of the simulated and measured curves, we used them to develop it into a mathematically strong proof of the *optimal frame duration*. Professor Charbon's persistence during this process was a major lesson I carry further. Even though scientists cannot constantly engage in such endurance processes, an experienced scientist caries the knowledge about it and seizes arising opportunities.

I thank Dr. Ron Hoebe for introducing me to the microscopy community. Dr. Hoebe guided me in microscopy literature, enabled hands-on experience and brainstormed new ideas that led to chip development. Dr. Claudio Bruschini took largely part as my supervisor. His attention to detail and persistence in increasing the quality is a major contribution in this scientific work. Thank you.

I met Dr. Chockalingam Veerappan and Esteban Venialgo firstly during my PhD interview. Chockalingam shared many fundamental SPAD principles, matlab scripts and helped with laboratory measurements. Esteban firstly pointed out that the photon inter-arrival time was exponential, a major fact that I was not aware before, and one I would use throughout my PhD. He helped with many mathematical advices and his adamant discussion style was of great help during my PhD.

My first room colleagues were Dr. Myung-Jae Lee and Dr. Masahiro Akiyama. Chao Zhang joined shortly later. Thanks to Myung-Jae and Dr. Akiyama for a warm welcome. I would like to thank Myung-Jae for his detailed device level advices. We regularly discussed scientific careers, and he shared great insights.

Chao Zhang started his PhD right after me. Chao shared with me a large amount of his wide-ranging engineering knowledge. I would like to thank him for the

great company during our PhDs. His willingness to help both at work and off work is a sign of great kindness.

Augusto Carimato, Claudia Damiani, Dali Zhang, Harald Homulle, Dr. Shingo Maindai and Ting Gong were colleagues from the adjacent room. Dr. Pengfei Sun joined us regularly from the former DIMES, and Scott Lindner skyped from Zurich. They were always willing to help, and added diversity and fun. Augusto's laugh could be heard from our room. Pengfei's sense of humor and creativity came clear at our course "The art of presenting science", where it was required to present in drunk, sad, funny, and angered manner. So clear that I questioned if he should have pursued an acting career.

I thank Dr. Samuel Burri, who designed the chip (SwissSPAD) that I used for the first part of my PhD. Our joint work enabled my first scientific contribution.

I would also like to thank Daisy Picavet, Emilie Desclos, Prof. Fabio Sebastiano and Prof. Kees Jalink. I had good company with Andrea Ruffino, Arin Ulku, Ashish Sachdeva, Augusto Ximenes, Bishnu Patra, Edwin Schriek, Francesco Gramuglia, Jaco Salentijn, Jeroen van Dijk, Jiang Gong, Job van Staveren, Milad Mehrpoo, Pascal 't Hart, Preethi Padmanabhan and Rosario Marco Incandela. Thanks to Antoon Frehe for smoothly managing the computer systems. Thanks to Minaksie Ramsoekh and Joyce van Velzen for managing logistics and sharing advices on how to proceed with Delft administration.

I was glad I could share special moments with my family, a microscopy picture or chip micrograph. Thank you for the support. Thanks to my girlfriend for being my companion during the PhD. Thank you for sharing the enthusiasm and taking part in the presented work.

I could write about many more moments, details and acknowledgements. I, however, let it float free without despoiling it with exactness.

Publications

Journals

I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, "Nonuniformity Analysis of a 65-kpixel CMOS SPAD Imager," IEEE Trans. Electron Devices 63, 57–64 (2016).

I. M. Antolovic, S. Burri, R. Hoebe, Y. Maruyama, C. Bruschini, and E. Charbon, "Photon-Counting Arrays for Time-Resolved Imaging," Sensors 16, 1005 (2016).

I. M. Antolovic, S. Burri, C. Bruschini, R. A. Hoebe, and E. Charbon, "SPAD imagers for super resolution localization microscopy enable analysis of fast fluorophore blinking," Sci. Rep. 7, 44108 (2017).

I. M. Antolovic, C. Bruschini and E. Charbon, "Recharge-dependent dynamic range extension of a 55% PDP SPAD array," *to be submitted to Optics Express.*

Conferences

I. M. Antolovic, S. Burri, C. Bruschini, R. A. Hoebe, and E. Charbon, "Super resolution with SPAD imagers," in Focus on Microscopy (2015).

E. Charbon and I. M. Antolovic, "SPAD Arrays and Digital SiPMs for All-Digital Imaging," in Single Photon Workshop (2015).

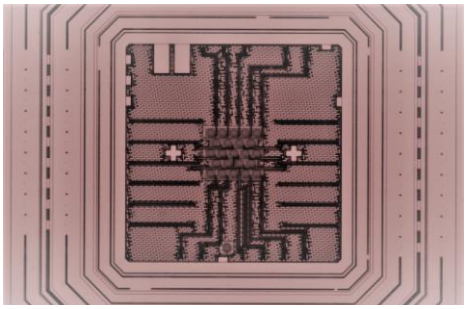
I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, "Analyzing blinking effects in super resolution localization microscopy with single-photon SPAD imagers," in Proc. SPIE 9714 Single Molecule Spectroscopy and Superresolution Imaging IX (International Society for Optics and Photonics, 2016), p. 971406.

R.A. Hoebe, I.M. Antolovic, S. Burri, C. Bruschini, and E. Charbon, "SPAD Imagers for Characterization of Ultra Fast Dyes for Super Resolution Localization Microscopy," in Focus on Microscopy (2016).

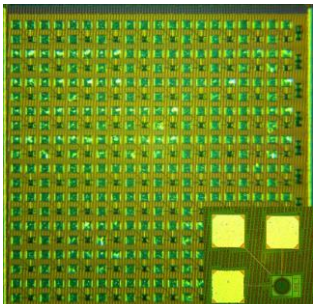
S. Lindner, C. Zhang, I. M. Antolovic, J. M. Pavia, M. Wolf, and E. Charbon, "Column-Parallel Dynamic TDC Reallocation in SPAD Sensor Module Fabricated in 180nm CMOS for Near Infrared Optical Tomography," in International Image Sensor Workshop (2017).

P. M. Wargocki, D. J. Spence, E. M. Goldys, E. Charbon, C. Bruschini, I. M. Antolovic, S. Burri, “Fluorescence lifetime imaging using a single photon avalanche diode array sensor,” in Proc. SPIE 10068 Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XV (International Society for Optics and Photonics, 2017), p. 100680.

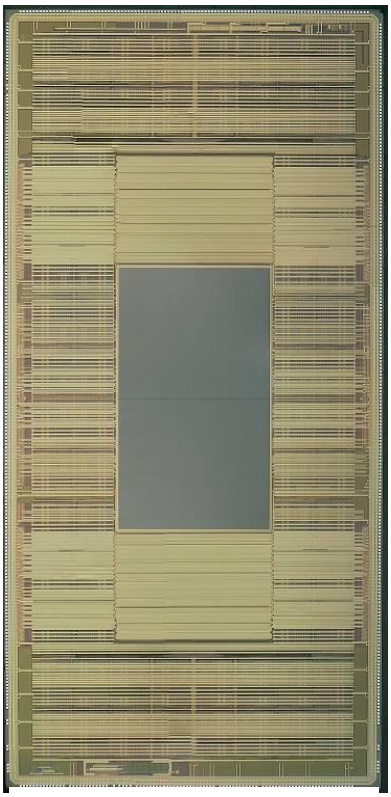
Chip gallery



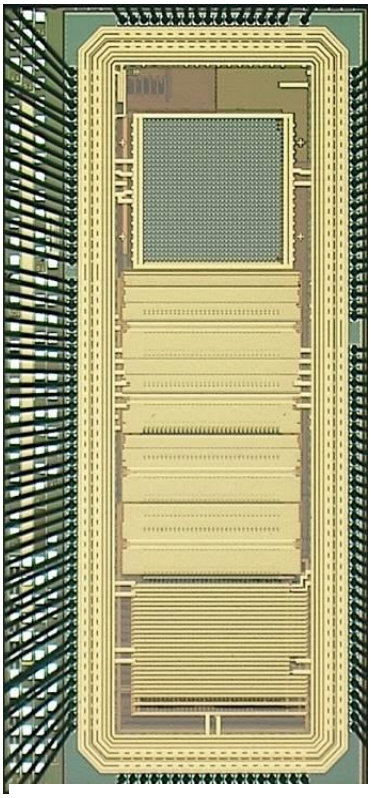
The nanoSPAD chip



A SPAD farm with SPAD inset

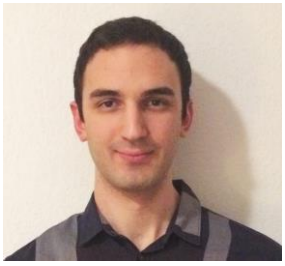


A 252×144 pixel, event-driven based SPAD array (codesigned with Chao Zhang and Scott Lindner; designed the pixel array, address coding, pulse propagation and auxiliary signal integrity circuit)



A 32×32 pixel, event-driven based SPAD array (codesigned with Chao Zhang and Scott Lindner)

About the author



Ivan Michel Antolović received his B.S. and M.S. degree (cum laude) in electrical engineering and information technology in 2010 and 2012 from University of Zagreb, Croatia. During his master, he started working with Hamamatsu's multi-pixel photon counters (MPPC), mainly interested in detection of collagen and estrogen autofluorescence. He was awarded "Josip Lončar" Bronze Plaque for the best student of the field electronic and computing engineering. He enrolled to a PhD at the University of Zagreb while working as a firmware designer at Artronic d.o.o. Since 2013, he continued to pursue a PhD degree in single photon avalanche diode (SPAD) imagers at TU Delft. His interests include large format photon counting SPAD imagers and small format time correlated SPAD imagers for microscopy applications like localization super resolution, confocal and fluorescence lifetime. During his PhD, he worked in collaboration with EPFL, Leeuwenhoek Centre for Advanced Microscopy, Macquarie University, Weizmann Institute. He worked with companies like Leica, LFoundry, NXP, TowerJazz and Zeiss. He was awarded PicoQuant Young Investigator Award in 2016 and Else Kooi Award in 2018.