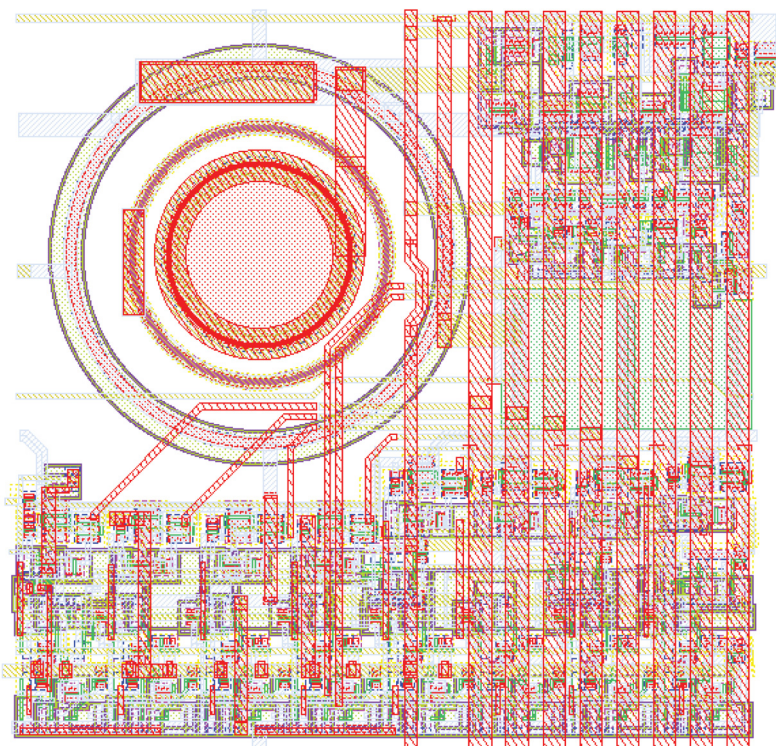


Two-Dimensional SPAD Imaging Camera for Photon Counting

Volume 2, Number 5, October 2010

Fabrizio Guerrieri
Simone Tisa
Alberto Tosi
Franco Zappa, Senior Member, IEEE



DOI: 10.1109/JPHOT.2010.2066554
1943-0655/\$26.00 ©2010 IEEE

Two-Dimensional SPAD Imaging Camera for Photon Counting

Fabrizio Guerrieri,¹ Simone Tisa,² Alberto Tosi,¹ and
Franco Zappa,^{1,2} *Senior Member, IEEE*

¹Politecnico di Milano, Dipartimento Elettronica e Informazione, 20133 Milano, Italy

²Micro Photon Devices SrL, 39100 Bolzano, Italy

DOI: 10.1109/JPHOT.2010.2066554
1943-0655/\$26.00 © 2010 IEEE

Manuscript received July 2, 2010; revised August 2, 2010; accepted August 4, 2010. Date of publication August 9, 2010; date of current version September 3, 2010. Corresponding author: F. Zappa (e-mail: franco.zappa@polimi.it).

Abstract: Many demanding photonic applications require the acquisition of images at very low light-level conditions and at high speed. Advanced imagers available on the market are generally not able to provide both performances in one detector. We present a 2-D imager based on a 32×32 array of “smart pixels,” each comprising a single-photon avalanche diode detector, an analog front end, and a digital processing electronics, which provides single-photon sensitivity, high electronic noise immunity, and high readout speed. The imager can be operated at a maximum of about 100 000 frame/s with negligible blind time between frames, provides high Photon-Detection Efficiency in the visible range and dynamic range, and low Dark-Counting Rate, even at room temperature. To easily integrate the imager into different applications, we developed a complete single-photon camera system, which fully operates the array simply through a USB 2.0 link and user-friendly software to configure camera parameters and operating modalities, as well as to perform readout.

Index Terms: Single-photon avalanche diode, imager, photon counting, high frame-rate cameras.

1. Introduction

Many applications require the acquisition of very weak optical signals, generally composed by few photons mostly in the visible and near infrared (400 nm–850 nm) wavelength range. In order to detect such low intensity signals, a very low noise detector is necessary. In photodetection systems, sensitivity is mostly limited by the noise introduced by the very first electronic stage sensing the detectors. If single-photon sensitivity is required, either a drastic decrease of the front-end electronic noise or an extremely long integration time is required. Since the latter is not generally compatible with most applications, efforts must aim to reduce circuit noise or to somehow amplify the electric signal generated by the detector before it reaches the first noisy electronic component. The typical example is the Photo-Multiplier Tube (PMT), where a series of dynodes are exploited to provide an internal mechanism to amplify the single photoelectron into a macroscopic signal of thousands of electrons. The disadvantages of PMTs reside into the limited number of independent “pixels” (i.e., spatial resolution), its overall dimensions, and the required high operating voltages (some kV). Micro-Channel Plates (MCPs) are composed by many “channels,” but as PMTs, they suffer from magnetic field interferences, require high voltages, and are fragile devices.

A solid-state device characterized by an internal amplification process is the Avalanche Photo-Diode (APD), which is a p–n junction that is reversely biased slightly below breakdown to get an amplification which boosts the photo-generated electron-hole pair signal. Unfortunately, the obtained

amplification is neither high enough (the amplification is generally much lower than 1000) nor noiseless (excess noise) thus preventing the detection of single photons.

Imagers manufactured in standard charge-coupled device (CCD) [1] and Complementary Metal–Oxide–Semiconductor (CMOS) Active Pixel Sensor (APS) [2] technologies are usually organized as arrays of pixels, each one delivering an “analog” electrical signal proportional to the light incoming on the detector, during defined integration time slots. Usually, the best sensitivity is attained by CCD imagers, where carriers photogenerated within the photosensitive area are transported over the entire CCD line, eventually reaching the readout node in a serial mode, one photogenerated packet of carriers after the other. This is the reason why CCD are typically slow (low frame-rates) and why there is usually a delay between the arrival of the first and the last packet. In the readout node, a charge-to-voltage conversion is performed before providing such analog information to the off-chip electronics.

Advanced CCDs can reach single-photon sensitivity, but their speed is fundamentally limited by the way the CCD is read out and noise. In the Electron-Multiplying CCD (EM-CCD) [3], [4], internal gain is exploited through impact ionization within a multiplication register added to the output register, with a total gain of about 1000–2000. The output register bandwidth is increased to increase the frame rate, but it comes at the expense of increased noise floor. Although single-photon counting may be possible, the ultimate limit depends on amplified clock-induced charge.

2. Single-Photon Avalanche Diodes

Single-Photon Avalanche Diodes (SPADs) [5], [6] are p–n junctions that are reversely biased well above breakdown. In that way, the detector works in the so-called Geiger-mode, as the absorption of a single photon is enough to generate a macroscopic (tens of mA), standard, and easily detectable current. Simple analog or digital electronics can be used to detect the avalanche triggering, and no readout noises affect the measure [7], [8]. Ignitions caused by thermal generation or tunneling effects within the semiconductor are called dark counts, and their variance represent the SPAD intrinsic noise. Moreover, ignitions due to carriers trapped first (during an avalanche current flow) and then released by trapping centers are called afterpulses and impair detection linearity and boost detector noise [5], [6].

The triggered avalanche is a self-sustaining phenomenon; therefore, a proper front-end electronics, which quenches and resets the detector, is necessary in order to operate the SPAD. The front-end electronics is generally referred as “quenching” circuit [5], [8]. State-of-the-art SPADs are fabricated in custom technologies, have spectral efficiency tailored to the wavelength of interest, and provide the best performances in terms of dark counts and afterpulsing. Unfortunately, the process used to make custom SPADs cannot be easily employed to integrate on-chip micro-electronics too. Therefore, in SPAD arrays fabricated with custom technologies, the sensor chip must be connected to off-chip quenching electronics. Such bonding wires and routing add parasitic capacitance that increase avalanche charge [9], detector afterpulsing, optical crosstalk, and power consumption. Eventually, SPAD arrays in custom technologies are limited to a maximum number of tens or hundreds of pixels at most.

In recent years, some groups reported on SPADs fabricated in standard high-voltage [10], [11] or even low-voltage [16] CMOS processes with sufficiently high degree of purity. Such CMOS compatibility opened the way to the development of arrays of SPADs and to the monolithic integration of SPADs with quenching/reset circuits [17] and with additional electronics [18]. Fig. 1 shows the cross section of a CMOS SPAD fabricated in standard 0.35 μm high-voltage CMOS technology, based on a 0.8- μm structure previously reported in [18]. The active $p^+ - n$ junction is formed between an n-well (e.g., a standard high-voltage p-type MOS (PMOS) well) and a shallow p^+ implantation (e.g., a PMOS source/drain region). A deeper diffused guard-ring (e.g., a high-voltage n-type MOS well) raises the breakdown voltage at the edge of the junction far above that of the useful active area (V_{BD}). In our CMOS technology, the former is 60 V, while the latter is 24 V. We chose an SPAD diameter of 20 μm , which provides a good tradeoff between low dark counts (which increase with the depleted volume) and large active area. The p-substrate is shared with the electronics; hence, it must be kept to ground, in order to avoid direct biasing of the p-substrate/n-well junction.

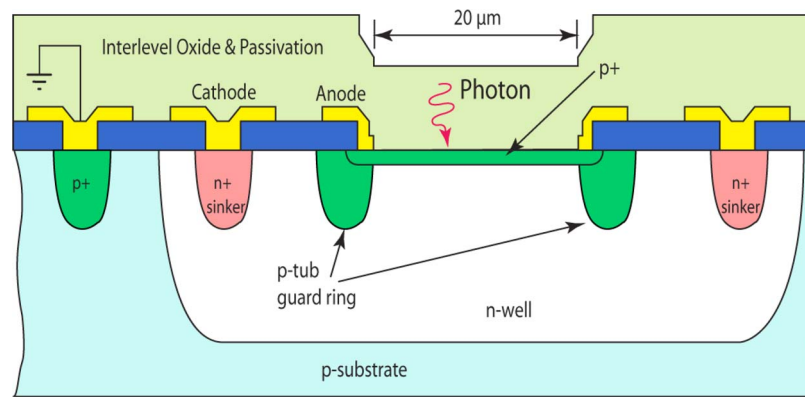


Fig. 1. Cross section of the CMOS SPAD detector. The sensitive region is within the annular guard-ring diffusions, where the electric field can trigger an avalanche multiplied current.

3. Single-Photon Avalanche Diode Array

Early CMOS SPAD arrays reported in the literature were either composed by a relatively low number of pixels operated in parallel or by a quite large number of multiplexed detectors [11], [19]. Both solutions are unsuitable for most applications. In particular, a multiplexed access impairs the high-frame rate capability of SPAD arrays. We present an expandable architecture for photon-counting arrays, based on compact smart pixels comprising both detector and electronics. In every pixel, there is a 20- μm CMOS SPAD, which is operated by a very compact quenching circuit we designed in order to maximize the performances while reducing the electronics overall dimensions. All array pixels are read out in parallel, with negligible (20 ns) deadtime among integration time-slots (i.e., between adjacent frames).

Compared with EM-CCDs, the proposed SPAD array has the major drawback of larger pixel dimension (100 μm instead of few micrometers), lower fill-factor (if no microlens arrays are employed), and a limited number of pixels (some thousands instead of millions). However, the advantages are real single-photon sensitivity, all-digital in-pixel processing, global-shutter acquisition mode, no readout noise, and very high frame-rate.

3.1. Pixel Architecture

We conceived the pixel with three goals: to maximize the detector active area; to reduce transistor count and sizes; to act as the basic building block for easily expandable array architectures. Each pixel of the array comprises both the SPAD detector and the analog front end and the digital counting electronics. Fig. 2(a) shows a simplified diagram of the pixel. The in-pixel SPAD's cathode is connected to a high-voltage $V_{\text{REV}} = V_{\text{BD}} + V_{\text{EX}}$, where V_{EX} is the SPAD excess voltage. Depending on the application, V_{EX} can range from 3 V up to 6 V. All pixel electronics work with a single supply of 3.3 V. The SPAD anode is connected to a variable-load quenching circuit (VLQC) [20], which senses the avalanches, quenches, and then resets the SPAD. The reason why we decided to operate the SPAD from the anode can be understood looking at Fig. 1; the wide depleted region across the p-substrate/n-well junction makes the cathode parasitic capacitance definitely larger than the anode one. Since the capacitance connected to the terminal used to operate the SPAD sets the charge flowing through the junction during every avalanche process [9], the smaller the terminal capacitance, the shorter the avalanche duration, and, hence, the related issues. Moreover, in order to further minimize the parasitic capacitance, the VLQC is designed with the minimum number of transistors connected to the SPAD [20].

A Linear Feedback Shift Register (LFSR) counter follows the quenching circuit and is used to count detector ignitions, hence, the number of detected photons and dark counts. A latch-stage [also shown in Fig. 2(a)] feeds the data bus and is employed to temporary store the photon-counts for reading out the array while the pixel starts acquiring a new frame (see next section).

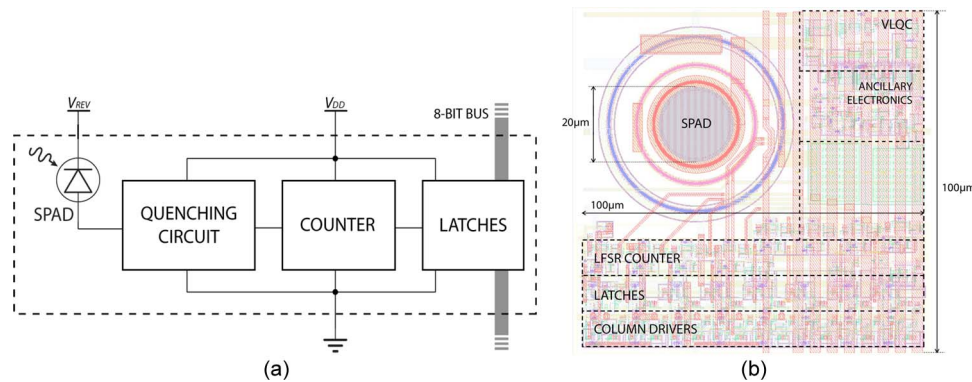


Fig. 2. (a) Block diagram of each pixel and (b) its layout. Note the SPAD detector, the analog sensing and quenching front-end (VLQC), and the digital counter and latch register. The output is a single 8-bit bus.

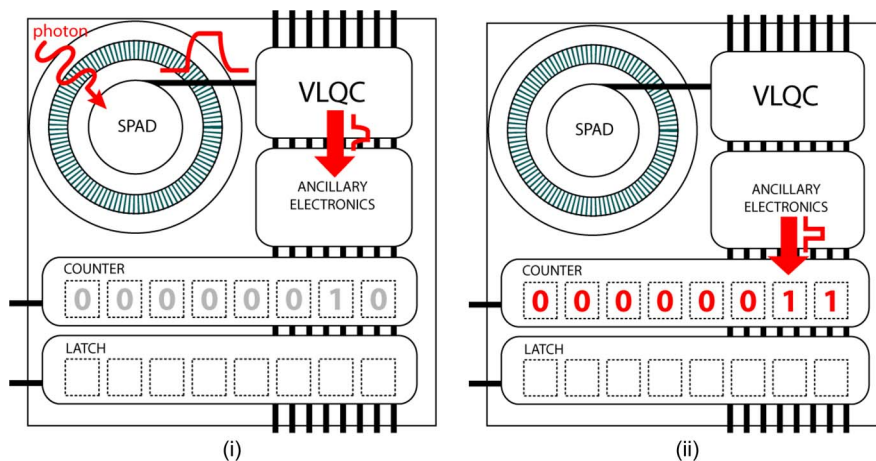


Fig. 3. Pixel operations at each SPAD ignition (e.g., due to a detected photon). (i) The SPAD ignition is sensed by the analog VLQC stage, which quenches the avalanche current and provides a pulse. (ii) The following electronics increments the in-pixel counter.

Fig. 2(b) shows the layout of the pixel, fabricated in a standard $0.35\text{-}\mu\text{m}$ high-voltage CMOS technology, with the different block highlighted. Pixel pitch is $100\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$.

3.2. Pixel Operations

Fig. 3 shows the operations performed by the pixel every time the SPAD is ignited: (i) The VLQC senses the detector's ignition, swiftly quenches it, and delivers a pulse to the following electronics; then, after an user-selectable hold-off time (40 ns–800 ns), it resets the SPAD; (ii) the ancillary electronics forms a standard pulse suitable to increment the 8-bit counter.

Fig. 4 shows the operations eventually performed at the end of every frame: (iii) The array global electronics generates a Stop pulse, which transfers the number of counted photons to the temporary storage register; (iv) finally, the global electronics generates a Start pulse which resets the counter, thus marking the beginning of the next frame.

3.3. Array Architecture

The pixel described in the previous section is actually a basic building-block for SPAD arrays with many pixels. We devised an array architecture capable of fully parallel operation of all pixels, thus allowing free-running imaging acquisition at very high frame-rates.

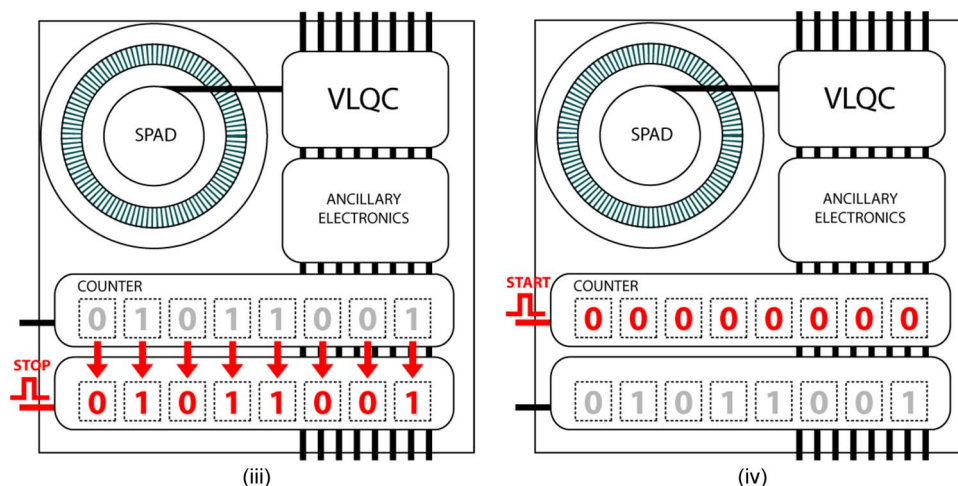


Fig. 4. Pixel operations at the end of each frame. (iii) A global Stop pulse (common to all pixels) forces the copy of the counters' contents into the respective latch registers. Then (iv), a global Start pulse resets all counters.

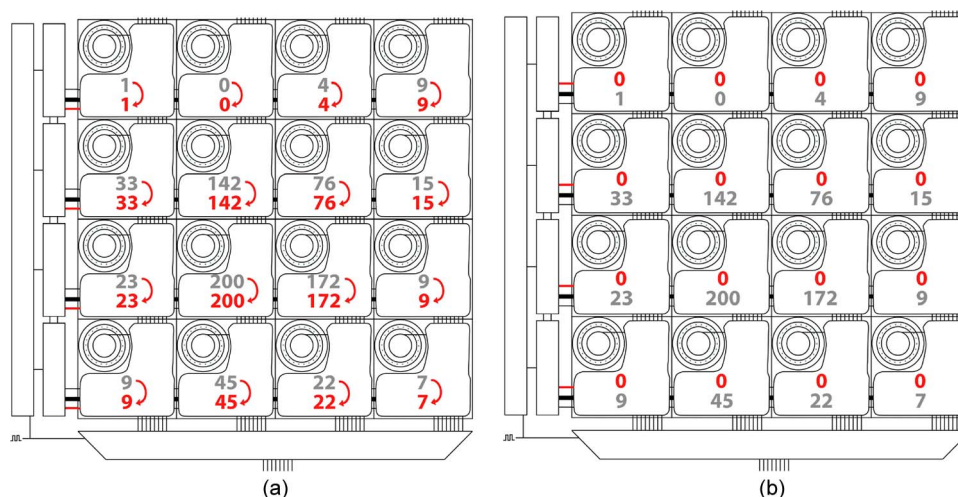


Fig. 5. Array operations at each frame end. The Stop pulse stores the accumulated counts into the local registers (a), and then, the Start pulse resets all counters for starting a new frame (b).

Fig. 5 describes the architecture of the array, which closely recalls an electronic memory with 8-bit buses running vertically along the columns, while control signals (e.g., Start and Stop pulses and power supply lines) travel horizontally. The output of the imager is a single 8-bit data bus (see at the bottom of both Figs. 5 and 6). Frame acquisition begins with a global Start pulse, sent to all array pixels, which resets the pixels internal counter. During the integration frame time, all pixels work in photon-counting mode and operate independently; each counter accumulates the number of photons (and dark counts) detected by the corresponding SPAD. At the end of the frame, a global Stop pulse is applied to all pixels to store the pixel actual count to the pixel memory register [see Fig. 5(a)]. After a very short time delay (about 20 ns) in order to avoid time overlapping, a new Start pulse resets all pixel counters thus marking the beginning of a new frame [see Fig. 5(b)]. The stored data are available during the whole next frame before a new Stop pulse updates the registers with the new frame data.

The data of the previous frame are read out during the acquisition of the new frame. The array global electronics sequentially addresses all pixels, as shown in Fig. 6, as in a standard memory

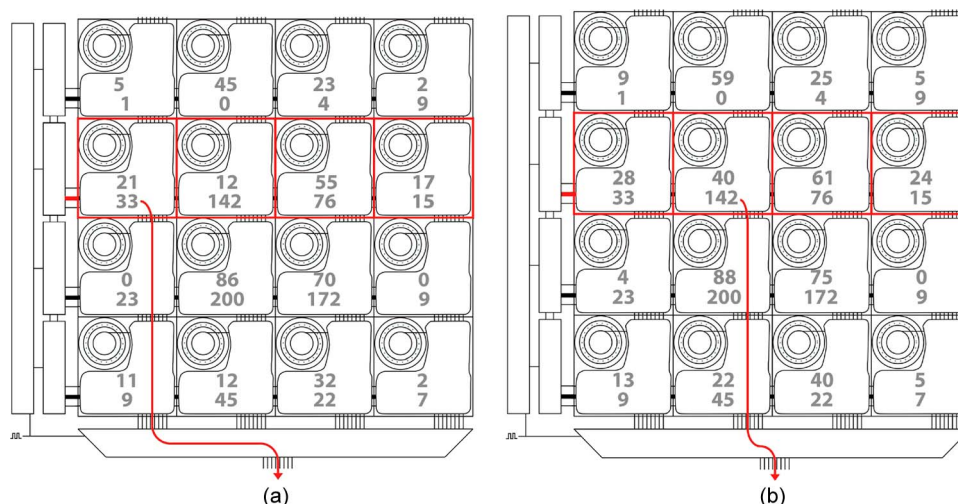


Fig. 6. Array operations during readout, data is read out sequentially by selecting the storage registers in a row-by-column addressing scheme. Note a single 8-bit data bus is output from the imager.

array. Compared with CCDs, the impact of data transfer overhead to the maximum achievable frame rate is dramatically reduced. The minimum integration time (the maximum frame rate) is set by the time necessary to download all data. With a system clock of 100 MHz, each pixel is read out in 10 ns, and hence, a complete 32×32 pixels array (1024 pixels) is read out in just about $10 \mu\text{s}$, giving a frame rate of about 100 000 frames/s. Such an acquisition scheme acts as an electronic global shutter, which avoids any image smearing or rolling shutter related problem. Note that the proposed architecture allows reaching even faster frame rates by simply placing more output 8-bit data buses, instead of only one, as in the fabricated chip. In fact, with four data buses instead of just one, the frame-rate will increase four times.

Note that the pin-out count is very limited: The 32×32 SPAD array with 1024 pixels requires just the single 8-bit data bus, few synchronization signals (master clock, frame start, frame stop), and the power supplies +3.3 V for the logic electronics and V_H to bias the SPADs.

4. Gate Modes

Some experiments require the possibility to gate the imager on and off very quickly. In the SPAD camera, we implemented two different gating modes.

4.1. Software Gating

The software gating is shown in Fig. 7(a). Counts are integrated by the counter only when the GATE signal is at a logic high level. Basically, an electronic switch enables the VLQC output signal to reach the counter. The term “software” refers to the fact that the counter is disabled during the gate-off time period, even though the SPAD keeps running, i.e., gets ignited due to incoming photons or dark counts, since both the SPAD and the front-end electronics are always enabled.

4.2. Hardware Gating

The hardware gating works directly on the SPADs bias voltage: During the gate-on period, the voltage is kept above breakdown at $V_H = V_{BD} + V_{EX}$, while during gate-off, V_H is swiftly pulled and kept below breakdown. Therefore, no avalanches can be ignited during gate-off. Both raising and falling edges last few nanoseconds; hence, it is possible to promptly gate-on the imager right after a laser pulse excitation, without the problem of saturating the array with the intense laser stimulus. Note that this is not an issue in fluorescence measurements, since excitation light can be separated from emitted light through proper wavelength filtering. Instead, when excitation and emission

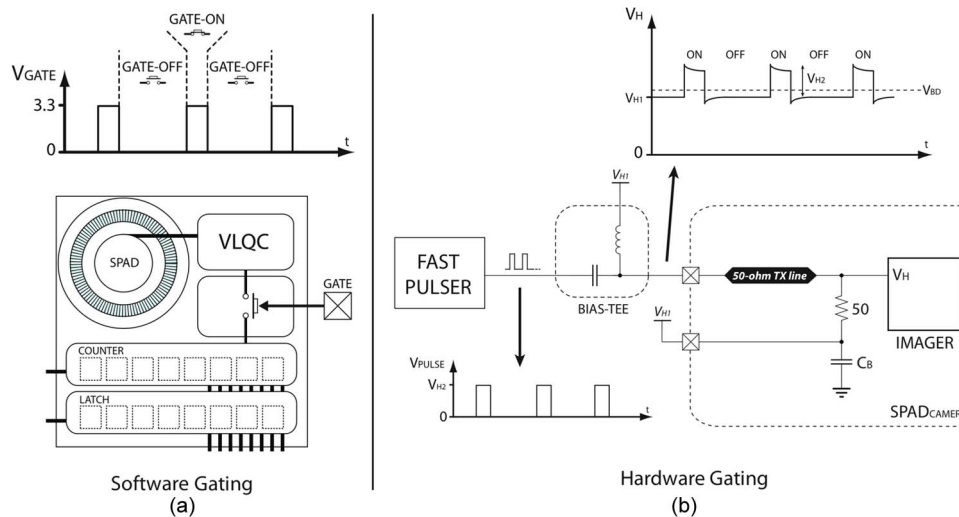


Fig. 7. (a) Block diagram of the software gating within each pixel: ignitions are counted by the counter only when the GATE pin is at a high logic level. (b) Block diagram of the hardware gating of the whole SPAD imager.

wavelengths are the same (as in photon migration measurements and near-infrared spectroscopy), this problem can be effectively solved with fast-gating.

Fig. 7(b) shows the circuit used for hardware gating the SPAD imager. The bias-tee component is composed by an inductor and a capacitor. When the fast pulse generator output is still, the capacitor is an open circuit, the inductor behaves as a short-circuit thus the power-supply V_{H1} feeds the imager through the V_H pin. Since V_{H1} is lower than the SPAD breakdown voltage V_{BD} , photons are not able to trigger the array pixels. On the contrary, when the pulse generator applies the fast rising edges, the capacitor acts as a constant voltage level shifter, while the inductor becomes an open circuit, the result being that the voltage at the bias-tee output is the sum $V_{H1} + V_{H2}$. Hence, the SPAD is enabled above breakdown. An impedance-matched transmission line terminated with a $50\text{-}\Omega$ resistor avoids reflections of the steep rising edge. The capacitor C_B filters the voltage V_{H1} at the resistor terminal.

During both gate-on and gate-off the voltage V_H tends to decay to V_{H1} , depending on the bias tee capacitor size and imager current consumption. Therefore, with the hardware gate, the gate-on time width cannot be arbitrarily long, and the gate-off duration should be long enough to allow the voltage V_H to recover to V_{H1} .

4.3. Gate-Mode Selection Criteria

There are reasons to prefer one modality, depending on the specific application. The parameters to be taken into account are the gate-on duration T_{on} and the gate duty-cycle (defined as the ratio between the gate-on time duration and gate period), compared with the deadtime T_{dead} , i.e., the time the SPAD is held off after each ignition by the electronics.

4.3.1. Very Short Gate-On and High Gate-Off Durations

The situation when $T_{on} \ll T_{dead}$ is the best scenario for hardware gating. That is true for several reasons: first of all because with software gating, a photon that triggered the SPAD during gate-off may keep the system “dead” during the adjacent gate-on window, thus decreasing the overall detection efficiency. In applications such as functional spectroscopy [12], but also in optical mammography [13] and molecular imaging [14], [15], there is the need to discriminate between early- (often due to reflections and scattering in shallow layers of the sample under investigation) and late- (which traveled much deeper in the sample) arriving photons to increase the contrast of measured scattering and absorbing perturbations within the sample. The strong presence of the so-called

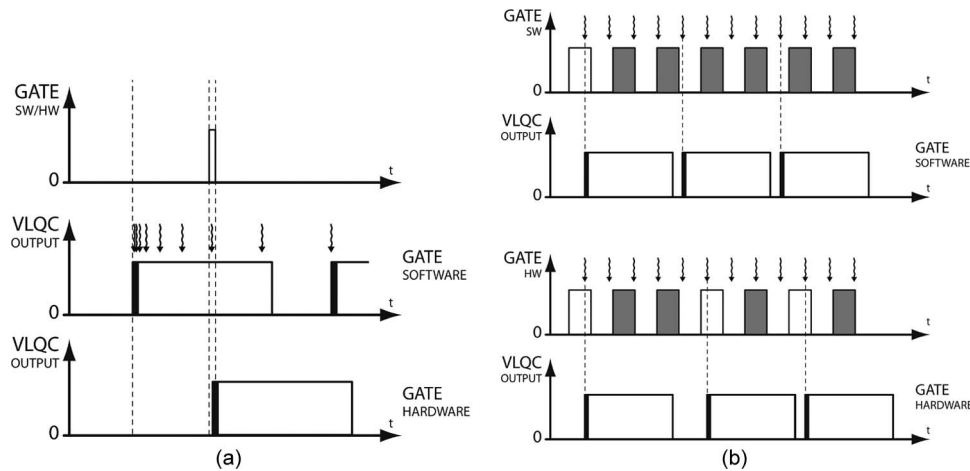


Fig. 8. (a) Different effect of the deadtime after the SPAD ignition in software and hardware gating, when $T_{on} \ll T_{dead}$. In software gating, if the SPAD gets ignited slightly before the gate-on window, photons within those windows cannot be counted. (b) Effect of the deadtime after the SPAD ignition in software and hardware gating when $T_{on} \ll T_{dead}$ and the duty-cycle is about 50%. Most gate-on windows (those highlighted in gray that are completely masked by the deadtime) are ineffective to count photons, because the SPAD is “dead” during the entire deadtime.

early photons” severely limits system performance since their number overcomes (often of a few orders of magnitude) that of the “late photons” [21]. Thanks to hardware gating, the SPAD cannot be triggered when gated-off; hence, every gate-on window is indeed available for detecting photons, as shown in Fig. 8(a). Moreover, the avalanche triggered during one gate-on window does not mask the next gate-on window because of the longer gate-off duration.

A further advantage is that the detection system gets completely rid of afterpulsing. Afterpulsing is a detrimental effect due to charge trapping in the detector depleted region during the avalanche current flow. When those charges are released after the avalanche ends, they can trigger again the SPAD. To avoid such a problem the common solution is to enforce a long deadtime after each SPAD ignition, thus increasing the probability of trap release during the hold-off period, hence lowering the probability of afterpulses. A deadtime that is too long is impractical with software gating, because the longer the deadtime, the higher the probability of blinding gate-on periods. Instead, when the SPAD is hardware gated with long gate-off durations between gate-on windows, trapped charges have sufficient time to get released without impairing the next gate-on windows.

4.3.2. Very Short Gate-On and Gate Off Durations

The situation when again $T_{on} \ll T_{dead}$ but the duty-cycle gets close to 50% (i.e., T_{off} gets similar to T_{on}) yields to a worsening effect of deadtime over the following T_{on} windows, as shown in Fig. 8(b). The longer the deadtime, the higher the number of gate-on windows that are kept “dead” by the previous SPAD ignition. For software gating, this applies for photons detected during both gate-on and gate-off periods. Instead, for hardware gating, this applies only for photons detected during the gate-on periods. In both situations, it is not possible to lengthen the deadtime very much to strongly cancel afterpulsing without causing a decrease in the overall detection efficiency.

4.3.3. Very Long Gate-On Duration

If the deadtime is negligible compared with the gate-on period, i.e., $T_{on} \gg T_{dead}$, the software gating is the best solution. Photons detected during the gate-off period are not counted, and the enforced deadtime can be long enough to definitely reduce afterpulsing. Moreover, hardware gating is not suitable, because of the discharge of the bias-tee capacitor [shown in Fig. 7(b)] when multiple pixels get ignited during the long T_{on} duration: The effective bias applied to the SPAD imager would no longer have the standard value, as shown in Fig. 9.

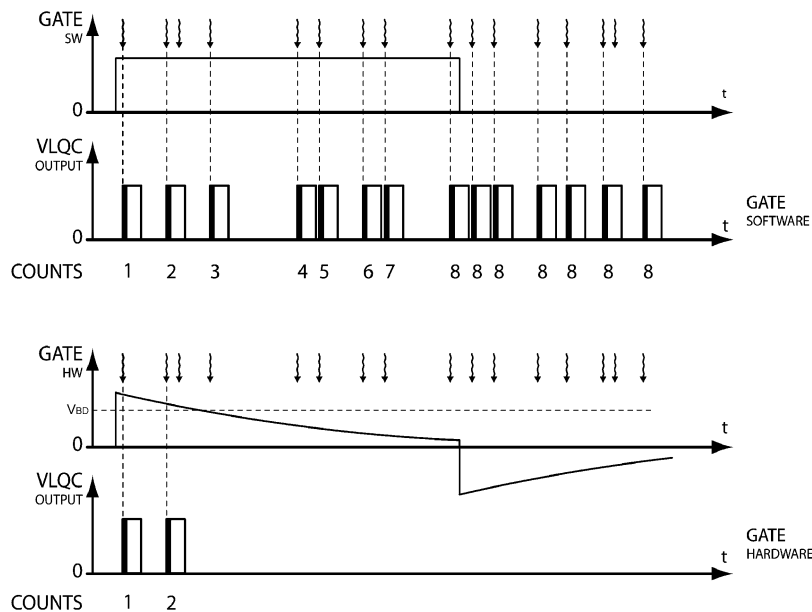


Fig. 9. Effect of long gate-on durations when $T_{on} \gg T_{dead}$. The AC coupling of hardware gating could impair response uniformity, due to the slowly decaying excess bias (bottom).

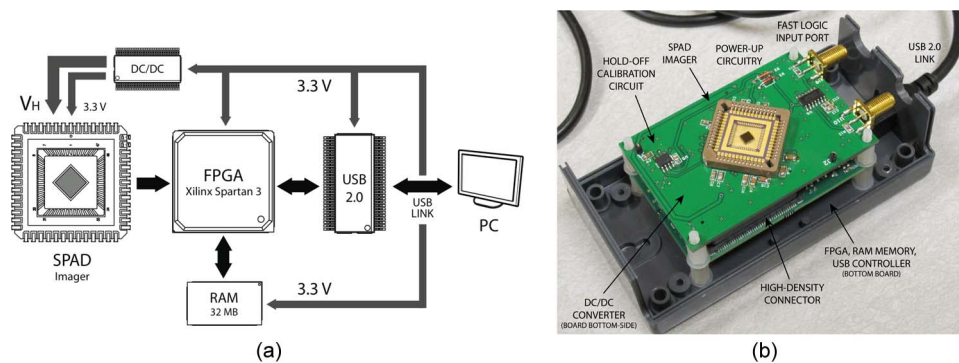


Fig. 10. (a) Block diagram of the SPAD array imaging camera, based on an FPGA for parameter settings and data readout. (b) SPAD camera boards and USB cable.

5. Single-Photon Avalanche Diode Camera

We developed a complete camera module with system electronics, optics, and software to be used in high-speed imaging applications. As shown in Fig. 10(a) the SPAD imager is read out by a fast Field Programmable Gate Array (FPGA) device that is able to communicate with the chip at 100 MHz and store data on a 32-Mbyte SDRAM memory. A Cypress USB 2.0 controller is in charge of communicating and transferring data to and from a remote PC at high speed (up to 480 Mbit/s). Power supply voltages are all derived from the USB link, with no need for external power supplies. The imager biasing voltage V_H (about 29 V, depending on the desired excess voltage) is generated from the 3.3 V by means of a DC/DC converter. FPGA, RAM memory, and a USB controller are placed on a high-performance FPGA-development board by OpalKelly (XEM3010), whereas the imager and the power supply electronics are housed in piggy-back-mounted board; the two boards communicate through high-density connectors [see Fig. 10(b)].

The imager chip is glued into a 52-pin plastic leaded chip carrier package, which is in turn connected to the rest of the circuit through a socket. The camera can be provided with a standard C-Mount connector or with off-the-shelf optical objectives.

Flexibility is one of the key features of the SPAD camera, in order to allow high configurability to different applications. On the camera side, depending on the target application, the Very high speed integrated circuit Hardware Description Language (VHDL) of the FPGA can be modified to perform the required imager setting and data processing. Also on the computer side, we developed an interface software that allows the user to acquire and present only the important information out of the large data sets the camera can acquire. Eventually, the right combination of FPGA firmware code and PC software can be developed to maximize the performances of the camera in any application. In order to address the most typical experimental needs, we developed three firmware-software modalities.

5.1. Single-Shot Modality

In this modality, the camera works at the fastest speed and data are stored into the 32-MB RAM memory available on-board. No communication with the PC is occurring during data acquisition, as the camera-to-computer link speed would not allow the camera to work at top performances. In fact, data-rate from the imager can reach 102 Mbytes/s (one byte for each one of the 1024 pixels at 100 000 frames/s), which is several times the maximum USB 2.0 speed. The user is required to configure the integration time (frame duration) and the number of frames to acquire (up to the RAM memory capabilities). The PC retrieves all data when the acquisition is completed.

5.2. Free-Running Modality

The free-running modality is thought of for experiments where the data to be acquired are larger than the 32 MB available on-board; hence, a camera-to-computer communication is required during image acquisition. In this scenario, the USB communication is the bottleneck determining the overall system speed: On average, if the imager generates more data than the link can sustain, data will be lost. In addition, the USB link does not guarantee a constant isochronous data-rate, but as opposed to this, it is characterized by high-speed data bursts, followed by quite long delays. The RAM memory is used to solve the latter problem, working like a data buffer, while the former problem can be solved only by reducing the generated data-rate readout from the imager, which is given by

$$DR_{\text{IMAGER}} = \frac{B \cdot N_R \cdot N_C}{8 \cdot 2^{20}} \cdot \frac{1}{T_{\text{INT}}} \text{ [Mbytes/s]}$$

where B is the number of bits readout per pixel, N_R and N_C are the number of rows and columns, and T_{INT} is the integration time (frame duration, i.e., the inverse of the frame-rate). In order to avoid data loss, DR_{IMAGER} must be lower than DR_{LINK} , i.e., the USB average data-rate. We experimentally found that the maximum DR_{LINK} the USB connection on the OpalKelly board can sustain is about 25 Mbytes/s. Therefore, the only way to reduce the imager data-rate is to either lengthen T_{INT} (decrease the imager frame rate) or to reduce B (read out less than 8 bits per pixel), N_C , or N_R . However, the integration time is generally application-dependent; hence, it should not be modified. Instead, in applications requiring short integration times or with very faint optical signals, a dynamic range of 8 bits (corresponding to a maximum count of 255 photons per frame) may be too much. In some cases 4 or even 2 bits will suffice. For this reason, the camera can be configured via software from the computer to use a lower number of bits: The pixel internal counter is always 8 bits, but depending on the internal FPGA routing, it is possible to collect only some of the least significant bits, thus avoiding wasting bandwidth to send a lot of useless zeros. For example, the 32×32 SPAD imager running at 100 000 frames/s can be operated at the DR_{LINK} limit of 25 Mbytes/s by reading out just 2 bits, corresponding to an average photon flux of 400 000 photons/s. Such a feature is also available for the *single-shot* modality, and it allows extension of the maximum number of frames, which can be packed into the 32-MB RAM memory.

Finally, the on-chip global electronics improves imager flexibility by allowing the user to arbitrary select any desired Region-of-Interest (ROI) of the overall array. This makes it possible to further reduce imager data-rate as well as to increase the maximum camera frame-rate, which linearly depends on the number of pixels in the ROI. The camera configuration is performed again through

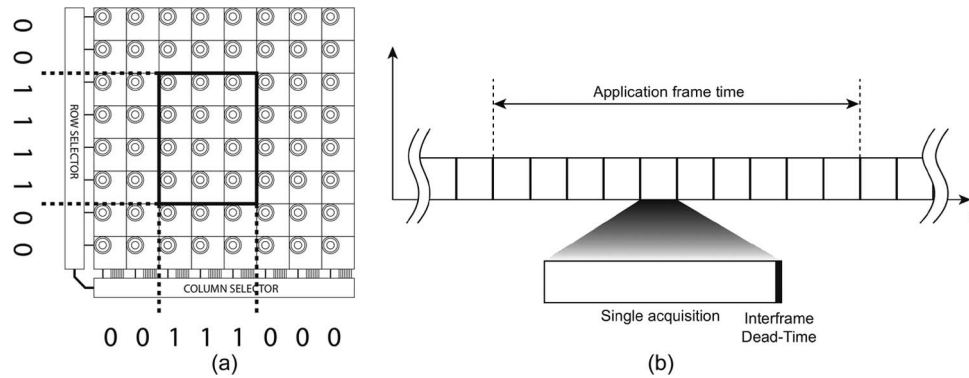


Fig. 11. (a) Imager configuration when just a subset of pixels must be read out from a specific Region-of-Interest. (b) How to obtain the required frame-rate by accumulating several acquired images. The dynamic range is correspondingly increased.

the computer software: Depending on the pixels the user wants to read-out, a proper setup is loaded into the chip at system startup, as shown in Fig. 11(a).

5.3. Live Modality

Having an experimental setup properly running can sometimes be a time-consuming and difficult task: A perfect alignment of optical components plays a fundamental role in the overall experiment performances. For this reason, the *live* modality operates the SPAD camera as a low-speed real-time imaging system (mostly like an ultra-sensitive webcam), thus allowing the user to have a “live” look at what the imager is focused on. As in the *free-running* modality, during the acquisition, data are transferred to the computer, but single frames are treated in a different way. Since a slow (e.g., movie-like, at about 25 frames/s) frame-rate is now required, several acquisitions, are accumulated into the FPGA in order to increase the movie signal-to-noise ratio. The user can select how many frames the camera has to accumulate as well as the movie frame-rate. No data are saved on the computer.

This in-hardware accumulation of images is not peculiar of the *live* modality and is available in all the working modalities described so far as an efficient way to improve the imager dynamic range, as discussed in the next section.

6. Dynamic Range

One of the problems that may impair photodetection is pixel saturation. For instance, in CCDs, a maximum number of electrons can be stored into the potential well, and as this limit is approached, electrons begin to escape the well, thus producing image artifacts such as saturation and blooming effects. In CMOS APS, saturation is reached when the photodiode junction capacitance gets completely discharged by the photocurrent. For SPAD pixels, the saturation is set by the SPAD detector itself and by the following counting electronics.

Concerning the detector, the deadtime T_{dead} imposed by the quenching electronics after every ignition, results in a maximum saturated counting rate equal to $1/T_{\text{dead}}$. SPAD saturation may have a strong effect on measurement linearity when photon flux approaches the saturation level. In fact, if the measured photon-rate is N_{meas} , a better estimation of the real photon-rate N_{real} is

$$N_{\text{real}} = \frac{N_{\text{meas}}}{1 - N_{\text{meas}} \cdot T_{\text{dead}}}$$

For example, with a deadtime of 40 ns, giving a saturated maximum count of 25 MHz, if the SPAD gets ignited 1 MHz (counts/s), the real average photon flux should be 1.042 MHz. The corrective

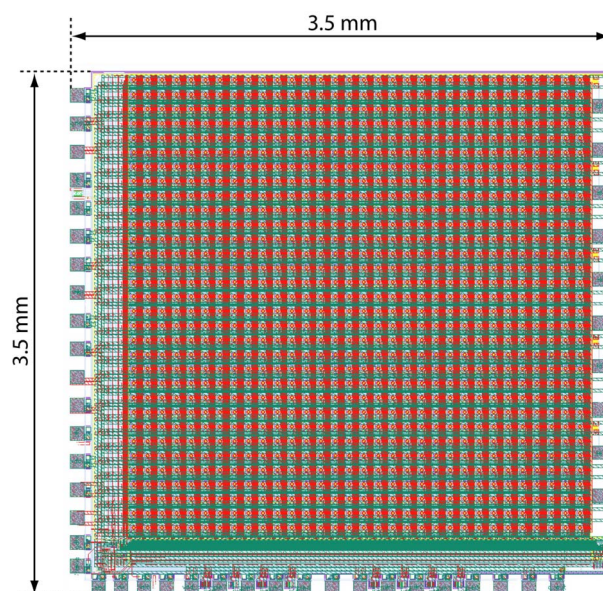


Fig. 12. SPAD imager chip layout and overall dimensions. The pixel pitch is $100\ \mu\text{m}$, and the diameter of the SPADs is $20\ \mu\text{m}$.

factor gets significant, let's say 5%, when the measured count approach 5% of $1/T_{\text{dead}}$, i.e., at about 1.25 MHz. Higher single-photon counting rate can be attained only by reducing the deadtime.

Concerning the counting electronics, the maximum measurable count is limited by the number of available bits of the in-pixel counter. In our case, the 8 bit counter limits the count to $2^8 - 1 = 255$. If the application requires long integration times, the maximum light intensity is going to be limited. However, with the SPAD camera, there is an easy way to solve this problem: Since the camera is designed to work at very high frame-rates, the solution is to keep integrating the image at the maximum speed (shorter frame-time) and then to reconstruct the requested long frame duration by accumulating several acquired images, as shown in Fig. 11(b). Such an accumulation can be performed off-board on the remote computer or directly on-board, at hardware level in the FPGA, depending on user's needs. Because of the very short minimum acquisition durations (down to $10\ \mu\text{s}$ in the presented SPAD imager), the corresponding increase in measured dynamic range can be dramatic. For a required frame-rate of 100 frames/s (i.e., a frame-time of 10 ms), the integration of 1000 frames of 8 bit each will bring to a dynamic range of 255 000, i.e., of 108 dB.

In principle, the same strategy could be applied to CCDs and CMOS APS sensors; however, for the technique to work properly, it is necessary to have no-read-out noise together with negligible deadtime between consecutive acquisitions and, possibly, also very short minimum acquisition time in order to boost the achievable dynamic range improvement.

7. Camera Characterization

We fabricated the imager using a standard $0.35\text{-}\mu\text{m}$ high-voltage CMOS technology. The 32×32 SPAD array overall dimension is $3.5\ \text{mm} \times 3.5\ \text{mm}$, whereas the active area measures $3.2\ \text{mm} \times 3.2\ \text{mm}$. The chip layout is shown in Fig. 12. The high-voltage required to bias the SPAD above breakdown ranges from 27 V to 30 V, depending on the required excess bias, given the SPADs breakdown voltage of 24.6 V. The rest of the on-chip electronics work at 3.3 V.

The higher the excess bias, the higher the pixels Photon Detection Efficiency (PDE) and Dark-Counting Rate (DCR) [1], [6]–[8]. In fact, Fig. 13(a) shows the typical PDE of the pixels, measured at different excess voltages, while Fig. 13(b) shows the cumulative distribution for dark-counting rates of all the 1024 pixels. Measurements were performed at room temperature and at 4 V excess bias. As can be seen in Fig. 13, about 75% of pixels have DCR lower than 4 kHz, while 92% are below

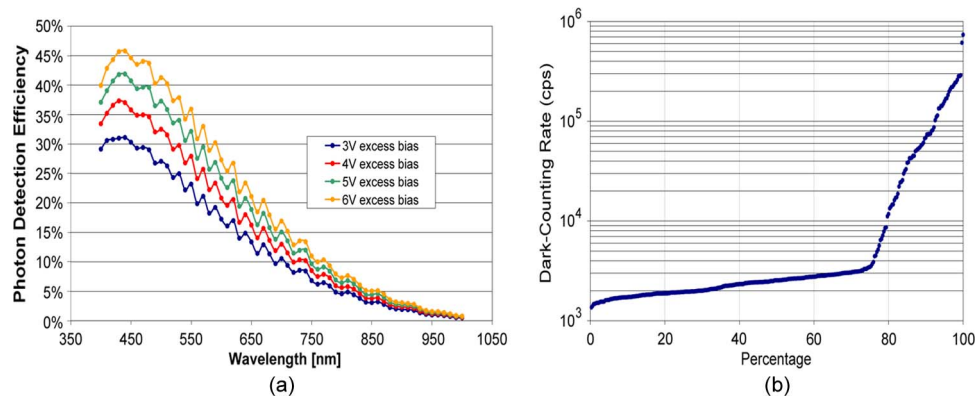


Fig. 13. (a) Typical pixel Photon Detection Efficiency (PDE). (b) Pixel cumulative distribution of dark-counting rates measured at room temperature and at 4 V of excess bias.

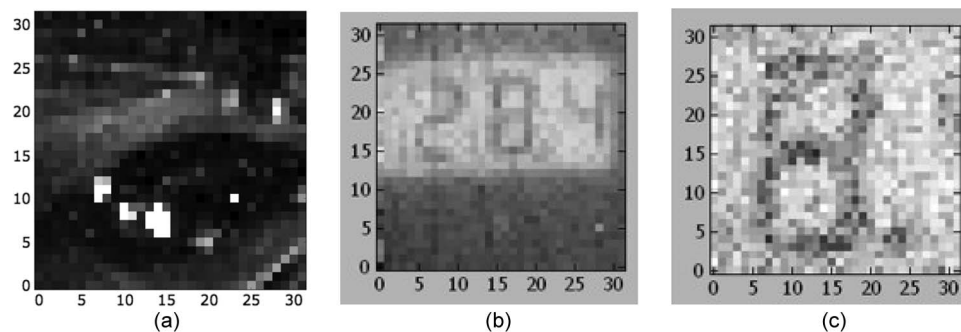


Fig. 14. (a) Night-time acquisition of a street taken with the SPAD camera running at 500 frames/s, with an integration time of 2 ms. Image acquisitions of an LCD display, with the SPAD camera running at 1000 frames/s (b) and at 100 000 frames/s (c).

100 kHz. This means that, when running the camera at 100 000 frames/s, 92% of pixels will count (on average) at most one dark count (out of the available 255 counts every integration frame-time). Even the “hottest” pixel with 700 kHz will accumulate a maximum of seven dark counts for every 10- μ s frame-time.

We characterized also crosstalk among pixels, which can be either electrical (due to spurious capacitive and substrate coupling among pixels electronics) or optical (due to photons emitted by hot-carriers from the ignited SPAD that could trigger neighboring SPADs). We measured a residual crosstalk probability of 3.53×10^{-5} between the closest pixels (100 μ m apart) and 4.33×10^{-6} for those 200 μ m apart, i.e., much less than one spurious ignition every 100 000 ignitions of the neighboring SPAD. Such value is indeed negligible for most applications.

8. Applications

8.1. Low Light Imaging

The high-sensitivity combined with the high-speed of the SPAD camera make it an ideal instrument in security surveillance applications and safety traffic monitoring applications, where no illumination light is present, e.g., at night. In order to test the camera capabilities, we acquired several pictures at night on in a poorly illuminated street. Fig. 14(a) shows one frame acquired by the SPAD camera with a frame-time duration of 2 ms when running the real-time movie at 500 frames/s. It is possible to recognize a car passing by; the three bright spots are the car's tail lights and the license plate.

Fig. 14 shows an electronic liquid crystal display acquired at different speeds (i.e., exposure times) by the SPAD camera. Figures are still clearly visible, even when acquiring frames at 10 μs (i.e., 100 000 frames/s).

8.2. Fluorescence Correlation Spectroscopy (FCS)

The basic concept in single-molecule FCS is to excite and collect light from a very small volume of solution and to work in a concentration regime, which results in rare burst-like events corresponding to the transit of a single molecule.

In order to work properly, this technique requires high-sensitivity (possibly at the single-photon level) photodetectors and long acquisition times in order to accumulate many events and achieve the required statistical accuracy. Although long acquisition times may be acceptable in basic research, they reduce the interest in the technique for clinical or biopharmaceutical environments. An obvious way to overcome such limitation is to acquire data from multiple identical spots. The SPAD camera presented in this paper is an ideal candidate for such an application, combining both extreme-sensitive detectors with possibility of parallel acquisition (up to 1024 parallel detection channels) [22].

8.3. Other Applications

Any applications demanding single-photon sensitivity in the visible wavelength range, combined with high imaging speed find in the SPAD camera a perfect candidate. Among them, it is possible to mention Adaptive Optics, which is an application where the SPAD array have already been tested and used in [23]. Also, Quantum Imaging [25], Confocal Microscopy [26], and 3-D optical ranging [27] will take advantage of SPAD arrays, as proved by preliminary measurements.

9. Future Developments

At present, the SPAD camera detector is equipped with no antireflection coating: further improvements in PDE can be expected with proper tailoring of the air-silicon interface of the chip. Furthermore, at present, the imager fill-factor is just 3.14%, due to pixel diameters of 20 μm and 100 μm pitch.

It is worth saying that not all applications require high pixel fill-factor; for instance, biological applications requiring confocal imaging or multi-spot FCS [22] require that pixels be kept sufficiently apart to work properly. Instead, many other applications require a fill-factor as close as possible to 100%. To solve this limitation, many approaches are available. The first trivial possibility is to somehow reduce the dimension of the pixel, i.e., of the in-pixel electronics. However, if the goal is to keep increasing the complexity and “smartness” of in-pixel processing, there are only a few available degrees of freedom. For instance, the number of bits can be reduced, but for the same chip technology, only a 5% filling-factor can be achieved. For a given technology, SPAD quality decreases for larger diameter as a result of the bigger active volume, which causes higher dark-counting rates and lower yield. With a practically feasible 50- μm diameter SPAD, the filling factor would be 16% for the same amount of electronics.

The use of a more scaled electronics is not successful because if the electronics become smaller, the quality of SPADs will greatly decrease, resulting in higher dark-counting rates and afterpulsing [28].

Also, advanced array structures where SPADs and electronics lie on different chips then bonded together by the means of 3-D vias are not able to provide 100% of fill-factor, because some basic front-end electronics and routings are still necessary on the SPADs chip. High processing costs make this solution still difficult to adopt.

The only real solution to the fill-factor issue is to adopt an array of microlenses in combination with the imager. The achievable maximum filling-factor can be higher than 80%, and lenses can be tailored to particular wavelengths to achieve even better results [29].

10. Conclusion

We designed and fabricated a high-speed single-photon camera based on a monolithic array of 32-by-32 “smart pixels” fabricated in a standard high-voltage 0.35- μm CMOS technology. Every

pixel comprises all the circuitry required to perform photon counting, from the 20- μm diameter SPAD detector to an analog electronic front end and a digital in-pixel counter and memory. Each pixel is therefore a completely independent photon-counting channel. The camera has been completely characterized and shows a Photon-Detection Efficiency topping 43% at 5 V excess-bias, moderate Dark-Counting Rates at room temperature, and good yield. The imager can be easily gated-on and off by either a standard logic input (software-gate) or by driving the SPAD bias voltage (hardware-gate).

The SPAD camera comprises all the required electronics, optics, USB link to a remote computer, and easy-to-use software. By means of software, it is possible to set the camera integration time, dynamic range, active area, and many other parameters of interest. The integration time ranges from few tens of nanoseconds to milliseconds, while the imager architecture allows a maximum frame-rate of about 100 000 frame/s. Eventually, the SPAD camera is a general-purpose imager for single-photon, high-speed, applications such as confocal microscopy, biological imaging, adaptive optics in astrophysics, 3-D ranging, and low-light spectroscopy.

References

- [1] J. R. Janesick, *Scientific Charge-Coupled Devices*. Bellingham, WA: SPIE, 2001.
- [2] G. C. Holst and T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*. Bellingham, WA: SPIE, 2007.
- [3] M. S. Robbins and B. J. Hadwen, "The noise performance of electron multiplying charge-coupled devices," *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 1227–1232, May 2003.
- [4] [Online]. Available: http://sales.hamamatsu.com/assets/pdf/hpspdf/e_imagemtec.pdf?src=micro-camera
- [5] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, Apr. 1996.
- [6] M. Ghioni, A. Gulinatti, I. Rech, and S. Cova, "Progress in silicon single-photon avalanche diodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 852–862, Jul./Aug. 2007.
- [7] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sens. Actuators A, Phys.*, vol. 140, no. 1, pp. 103–112, Oct. 2007.
- [8] S. Tisa, F. Zappa, A. Tosi, and S. Cova, "Electronics for single photon avalanche diode arrays," *Sens. Actuators A, Phys.*, vol. 140, no. 1, pp. 113–122, Oct. 2007.
- [9] F. Zappa, A. Tosi, A. Dalla Mora, and S. Tisa, "SPICE modeling of single photon avalanche diodes," *Sens. Actuators A, Phys.*, vol. 153, no. 2, pp. 197–204, Aug. 2009.
- [10] C. Niclass, M. Sergio, and E. Charbon, "A single-photon avalanche diode array fabricated in 0.35 μm CMOS and based on an event-driven readout for TCSPC experiments," *Proc. SPIE*, vol. 6372, p. 63720S, 2006.
- [11] D. Mosconi, D. Stoppa, L. Pancheri, L. Gonzo, and A. Simoni, "CMOS single-photon avalanche diode array for time-resolved fluorescence detection," in *Proc. 32nd ESSCIRC*, 2006, pp. 564–567.
- [12] A. Torricelli, A. Pifferi, L. Spinelli, R. Cubeddu, F. Martelli, S. Del Bianco, and G. Zaccanti, "Time-resolved reflectance at null source-detector separation: Improving contrast and resolution in diffuse optical imaging," *Phys. Rev. Lett.*, vol. 95, no. 7, p. 078101, Aug. 2005.
- [13] P. Taroni, G. Danesini, A. Torricelli, A. Pifferi, L. Spinelli, and R. Cubeddu, "Clinical trial of time-resolved scanning optical mammography at 4 wavelengths between 683 and 975 nm," *J. Biomed. Opt.*, vol. 9, no. 3, pp. 464–473, May/Jun. 2004.
- [14] R. Weissleder and U. Mahmood, "Molecular imaging," *Radiology*, vol. 219, no. 2, pp. 316–333, May 2001.
- [15] M. J. Niederre, R. H. de Kleine, E. Aikawa, D. G. Kirsch, R. Weissleder, and V. Ntziachristos, "Early photon tomography allows fluorescence detection of lung carcinomas and disease progression in mice in vivo," *PNAS*, vol. 105, no. 49, pp. 19 126–19 131, Dec. 2008.
- [16] H. Finkelstein, M. J. Hsu, S. Zlatanovic, and S. C. Esener, "Performance trade-offs in single-photon avalanche diode miniaturization," *Rev. Sci. Instrum.*, vol. 78, no. 10, p. 103 103, Oct. 2007.
- [17] F. Zappa, A. Lotito, A. C. Giudice, S. Cova, and M. Ghioni, "Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors," *IEEE J. Solid-State Circuits*, vol. 38, no. 7, pp. 1298–1301, Jul. 2003.
- [18] S. Tisa, A. Tosi, and F. Zappa, "Fully-integrated CMOS single photon counter," *Opt. Express*, vol. 15, no. 6, pp. 2873–2887, Mar. 2007.
- [19] C. Niclass, A. Rochas, P. A. Besse, R. Popovic, and E. Charbon, "A 4 μs integration time imager based on CMOS single photon avalanche diode technology," *Sens. Actuators A, Phys.*, vol. 130/131, pp. 273–281, Aug. 2006.
- [20] S. Tisa, F. Guerrieri, and F. Zappa, "Variable-load quenching circuit for single-photon avalanche diodes," *Opt. Express*, vol. 16, no. 3, pp. 2232–2244, Feb. 2008.
- [21] A. Dalla Mora, A. Tosi, F. Zappa, S. Cova, D. Contini, A. Pifferi, L. Spinelli, A. Torricelli, and R. Cubeddu, "Fast-gated single-photon avalanche diode for wide dynamic range near infrared spectroscopy," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 4, pp. 1023–1030, Jul./Aug. 2010.
- [22] X. Michalet, R. A. Colyer, G. Scalia, T. Kim, M. Levi, D. Aharoni, A. Cheng, F. Guerrieri, K. Arisaka, J. Millaud, I. Rech, D. Resnati, S. Marangoni, A. Gulinatti, M. Ghioni, S. Tisa, F. Zappa, S. Cova, and S. Weiss, "High-throughput single-molecule fluorescence spectroscopy using parallel detection," in *Proc. SPIE Photon. West Conf.*, San Francisco, CA, 2010, vol. 7608, pp. 76082D-1–76082D-12.
- [23] F. Zappa, S. Tisa, and S. Cova, "Single-photon avalanche diode arrays for fast transients and adaptive optics," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 1, pp. 365–374, Feb. 2006.

- [24] F. Guerrieri, L. Maccone, F. N. C. Wong, J. H. Shapiro, S. Tisa, and F. Zappa, "Sub-Rayleigh imaging via N-photon detection," *Phys. Rev. Lett.*, to be published.
- [25] F. Guerrieri, L. Maccone, F. N. C. Wong, J. H. Shapiro, S. Tisa, and F. Zappa, "Sub-Rayleigh imaging via N-photon detection," in *Proc. Quantum Electron. Laser Sci. Conf.*, San Jose, CA, 2010.
- [26] J. B. Pawley, *Handbook of Biological Confocal Microscopy*. New York: Springer-Verlag, 2006.
- [27] G. S. Buller and A. M. Wallace, "Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 1006–1015, Jul./Aug. 2007.
- [28] H. Finkelstein, M. J. Hsu, and S. C. Esener, "STI-bounded single-photon avalanche diode in a deep-submicrometer CMOS technology," *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 887–889, Nov. 2006.
- [29] S. Donati, G. Martini, and M. Norgia, "Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits," *Opt. Express*, vol. 15, no. 26, pp. 18 066–18 075, Dec. 2007.